

The Adoption of Physiological Measures as an Evaluation Tool in UX

Vanessa Georges^{1(✉)}, François Courtemanche¹, Sylvain Sénécal¹,
Pierre-Majorique Léger¹, Lennart Nacke², and Romain Pourchon¹

¹ HEC Montréal, Montréal, Canada
vanessa.georges@hec.ca

² University of Waterloo, Waterloo, Canada

Abstract. One of the challenges associated with the use of physiological signals as an evaluation tool in measuring user experience (UX) is their reduced usefulness when they are not specifically associated with user behavior. To address this challenge, we have developed a new evaluation tool which contextualizes users' physiological and behavioral signals while interacting with a system. We have conducted interviews with 11 UX practitioners, from various industries, to evaluate the usefulness of our tool. Through these interviews we gained a better understanding of the challenges facing industry practitioners when using physiological measures and assessed the functionalities provided by our tool.

Keywords: User experience · Interface design · Heatmaps · Eyetracking · Physiological computing · Cognitive load · Affective computing

1 Introduction

User experience (UX) has recently become of strategic importance in the information technology industry [14]. The Tech3Lab is an applied research lab in human-computer interaction at the HEC Montréal business school, specializing in user experience using eyetracking and neurophysiological and behavioral measures. Our research pertains to the development of new evaluation methods, ones that investigate the why instead of the how as information on how users feel about a system, game, or web interface is now a common requirement for all UX evaluation methods [5].

Our recent work with our industry partners has lead us to question a major discrepancy between industry and academic practices: while physiological measures are increasingly used in academia, the adoption of these methods as UX evaluation tools remains uncommon in industry. We have observed a growing demand for more quantitative user research to provide data-driven recommendations for change, which we implement using eyetracking and neurophysiological and behavioral measures. We therefore wanted to understand what can be done to facilitate their adoption in industry. In tackling this issue, we have sought to create a visualization tool that contextualizes physiological and behavioral signals to facilitate their use [4]. The visualization method that we created is UX heatmaps, an integrated visualization tool which contextualizes

physiological and behavioral signals to facilitate the interpretation of these measures [12].

2 Physiological Measures in UX

Traditional evaluation methods other than direct observation, for example questionnaires or interviews, mostly rely on self-reported data to assess the affective and cognitive states of users either during or after the interaction [6]. For example, Hassenzahl et al. have developed a questionnaire to evaluate users' feelings about a system [11]. The results assess the user's reflection on the interaction, but not the interaction itself [13]. Users' emotional and cognitive states can also be inferred using physiological signals, such as electrodermal activity, heart rate, eyetracking and facial expressions (see [2, 3] for reviews). As an evaluation method, electrodermal activity (EDA), which measures the electrical conductance of the skin, can provide practitioners with real-time information as to what the user is experiencing throughout the interaction. EDA is used as an indication of physiological arousal [8], as well as emotions. FaceReader [7], which analyzes facial expressions and infers the probability of seven discrete emotions (happy, sad, angry, surprised, scared, disgusted and neutral) and emotional valence (negative vs. positive) based on facial movements, can provide important temporal information without retrospective or social desirability bias. Furthermore, data is collected without interrupting the user in their authentic interaction.

However, these measures are still difficult to contextualize and interpret, as they are not specifically associated with user behavior or interaction states. Let's take the example of a user asked to browse the product offerings of an e-commerce website and purchase an item. With physiological data, we can infer that the user was frustrated at some point during the interaction, for example during the checkout process, but not the element that caused the negative emotion. We are therefore left wondering what was the button, task or area of the interface which caused the user to feel frustrated or angry. Physiological signals also require a certain degree of interpretation, as the output needs to be processed to transition from raw data to useful actionable insights. To meet these challenges, Kivikangas et al. [15] have developed a triangulation system to interpret physiological data from video game events. Other researchers have also developed tools that allow users' to manually assign subjective emotional ratings on visual interfaces [9] or to visualize emotional reactions using biometric storyboards [10].

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). To address these issues, we developed a new visualization method, in the form of heatmaps, which highlights the areas where users were looking when they experienced specific cognitive and emotional states with a higher frequency, called UX heatmaps [12].

2.1 Physiological Heatmaps

To produce physiological heatmaps, different emotional (sadness, happiness, surprise, etc.) and cognitive (cognitive load, stress, etc.) states are first inferred from continuous physiological or behavioral signals. These states are then triangulated with eyetracking data and mapped onto an interface to create heatmaps. In other words, physiological data, for example electrodermal activity and heart rate (HR) are synchronized together, along with eyetracking data. A machine learning model is then used to infer an emotional or cognitive state for each gaze. These are then mapped out onto the interface in the form of heatmaps, which in turn highlight the areas where users tend to emotionally or cognitively react more strongly. Figure 1 illustrates heatmaps generated by participant 01 during our session. On the top interface, a negative valence (red) and positive valence (yellow) heatmaps are shown. The web page below, a cognitive load heatmap is presented.

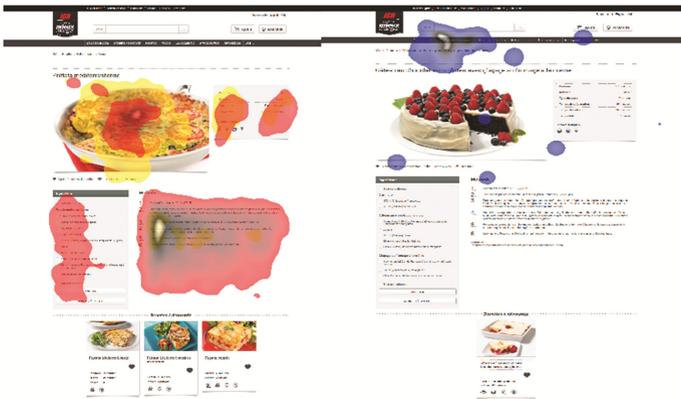


Fig. 1. On the left-hand side, negative valence (red) and positive valence (yellow) heatmaps. On the right, a cognitive load heatmap (blue) is illustrated. (Color figure online)

3 Research Method

For this study, a total of 11 UX practitioners and consultants were recruited over a period of 4 weeks. None of the practitioners interviewed had seen our tool prior to the test. Each interview lasted about 1 h and a half, during which participants were asked to complete a UX evaluation report using the tool following a variation on the think aloud protocol, cooperative evaluation [16]. During the sessions, participants were asked to talk through what they were doing. The interviewer also took on a more active role, by asking questions along the way (e.g. ‘why?’ ‘what do you think would happen?’). Participants were encouraged to ask for explanations along the way.

3.1 Pre-task Interview

We started each session with a preliminary interview to get background information on each participant (see Fig. 2), such as their number of years of experience in UX as well as their title and main functions within their company, to break the ice and assess their level of qualification. We then gathered their thoughts on physiological measures as a UX evaluation method and assessed their level of familiarity with such methods. Participants had between 2.5 and 24 years of experience in UX, for an average age of 8 years. We interviewed UX directors, consultants, ergonomists and strategists, all of which had heard of physiological measures as an evaluation method in user testing before being approached for this experiment; 7 out of the 11 participants had heard about it while in school, validating the predominance of these methods in academia. Out of all the UX practitioners recruited for this experiment, 8 had previously used physiological measures prior to the study. Eyetracking, being the most popular method overall, was mentioned by all; followed by FaceReader with 3 mentions.

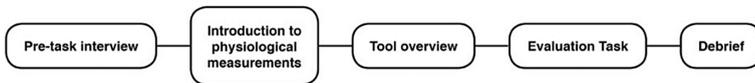


Fig. 2. Experimental procedure.

3.2 Physiological Measure Introduction and Tutorial

After the introductory discussion, all participants were given a short PowerPoint presentation to introduce them to physiological measures, and were given a tutorial on the tool itself. To do so, we presented each participant with the tool, and went through all the functionalities, buttons and features available to them. We wanted the users to have the same basic knowledge and comprehension of the tool and measures before using it in the completion of a UX evaluation report. The interviewer assisted the participant throughout the experiment, as the goal of the session was not to assess the usability of the tool's interface, but the usefulness of its features and functionalities.

3.3 Evaluation Task

During the session, practitioners were asked to complete a user testing evaluation report using our UX heatmaps tool. We therefore provided them with a partially completed PowerPoint report and a 15 participant data set from a previous study. The PowerPoint report included a study summary, a research scenario and qualitative data. We believed this would help UX experts integrate the information on physiological measures quickly and effectively, and also give them a concrete opportunity to use the tool to envision themselves using it in their own practice. First off, participants were briefed on the task at hand, 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 a total of 2 PowerPoint slides. They were asked to: (1) generate and select data visualizations to include in their report using our tool, (2) interpret the results and (3) provide recommendations to the client.

The remainder of the time was used to discuss the advantages and disadvantages of physiological measures as an evaluation method, as well as the tool itself.

4 Results

Participants made interesting comments regarding physiological measures and our tool, which we will address in the following section. We are only reporting comments made by 3 or more participants. Interviewees mentioned the following as the ways in which they would use our tool in their own practice:

- Provide new avenues for research
- Form and confirm research hypotheses
- Guide discussions during interviews
- Confirm and validate findings
- Elaborate evaluation tests

The main contribution of our tool, as stated by 5 participants, is the comparison and the juxtaposition of different emotional and cognitive states. As participant 07 explained, “there are simply no other tools available that make this essential data accessible to us”. Participants also mentioned the collaborative potential of our tool. The visualizations generated could be used to communicate information to the various members of the design team, as well as with clients and management. For example, participant 10 suggested that the visualization generated could be shared with designers for them “to better understand the impact of their creative freedoms on the user”.

4.1 Data Contextualization and Interpretation

Our goal in creating our tool was to address one of the main concerns associated with the use of physiological measures, the interpretation of physiological and behavioral signals. We set out to do these interviews with industry practitioners to find out how we fared at the task. Overall, participants found physiological heatmaps easy to interpret. As six participants mentioned, the visualizations were clear, intuitive and yielded powerful results that facilitated the interpretation of physiological signals.

Users stated that our tool was also easy to understand from a client’s perspective. For example, participant 08 felt that customers would appreciate seeing the emotions generated by problematic areas directly onto the interface, adding “it goes beyond qualitative insight”. Two participants found the interpretation of the data to be difficult without prior knowledge of physiological measures, one practitioner adding “the learning curve is relatively mild; the analysis should become more natural with time”.

As illustrated in Fig. 3, participants were able to make insightful and actionable recommendations based on the visualizations generated with our tool; on the left, a gaze (green), a positive (yellow) and negative valence (red) heatmaps generated by P04. Although the focal element of the page was the text area below, the image clearly elicited positive emotions, while negative emotions or displeasure was experienced by users in correlation to the instructions of the recipe. By comparing regions of negative and positive valence, the

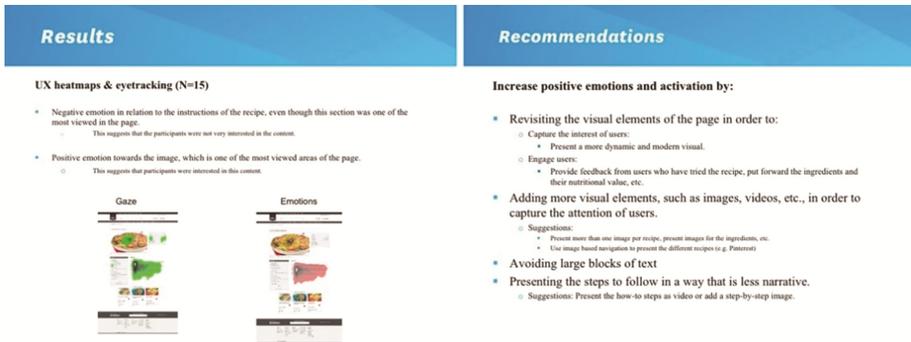


Fig. 3. An example of a completed report by a participant, translated from French to English (Color figure online)

practitioner identified problematic areas of the interface and was able to highlight the graphical elements behind them. Based on these results, the practitioner recommended to increase positive emotions and arousal experienced on the page by adding visual elements, such as videos and pictures, and revising the presentation of the recipe's instructions to avoid superfluous text areas.

When asked about their intent to reuse the tool, 10 out of the 11 practitioners interviewed stated that they would use the tool in their practice. However, when inquired further, 6 of them declared that their use of the tool would depend on the projects, using it only in the assignments where emotions are an important component or if clients specifically requested them to use physiological signals.

5 Discussion

When developing new UX evaluation tools using physiological measures, the ability to locate issues, the ease of use and interpretation and the reduction of analysis time represent important factors. Overall, participants found that physiological signals would be integrated more easily into their practice using our tool. Participants suggested the following improvements to UX heatmaps to further facilitate the adoption of physiological measures their current practice:

- The addition of an event timeline, or replay feature, to better understand overlapping UX heatmaps, to see the order in which the different emotional and cognitive states occurred. This would help with the interpretation of the visualizations.
- The inclusion of supplementary information, collected from traditional UX methods, such as participants' profiles and usability metrics. This would help them to integrate physiological methodologies more easily to the methods they currently use in their practice.
- The automatization of certain functions, such as groups and layer creation, to accelerate the interpretation of the visualizations generated with our tool. This would help them fit this analysis within their short development cycles.

Although our tool makes physiological measures more accessible to UX practitioners by addressing the interpretation of signals, there remains a lot of work to be done regarding some of the more technical aspects of physiological measurements. Participants expressed concerns regarding the time constraints pertaining to the actual experimental setup of such user testing, for example the selection of signals and the placement of sensors, as well as the resources needed to run the experiment. Knowledge of physiological measures is still needed, as the signals used for physiological heatmaps should be selected according to the psychological variables of interest (e.g. emotion, cognitive load, etc.). Physiological measurements still represent important time and financial constraints, as data collection, experimental setup and data extraction still have to be overseen by the UX professional.

As mentioned above, practitioners who use physiological measures are doing so in particular projects only, i.e. projects that require the evaluation of emotions or if these measures are requested by the client. This translates into a steep and ever present learning curve, as practitioners must re-learn how to use the tools and materials associated with the data collection of physiological signals at each use. Therefore, the practitioners are never able to develop an expertise. Unable to justify the financial investment due to sparse usage of such tools, practitioners often end up renting the equipment, which is very costly.

Having practitioners use our UX heatmaps tool in the completion of an actual user testing evaluation report following a cooperative evaluation protocol yielded great results. We would recommend using this method for the evaluation of new tools and methodologies as:

- Participants felt comfortable to criticize physiological methodologies and our tool
- Provided a more relaxed atmosphere where participants could see themselves as collaborators rather than as experimental subjects
- Helped them take ownership of the tool and explore the functionalities it offered
- Helped us get insights as to how this tool would be received in the community.

We had hoped that the interview process would generate new ideas and avenues of research, in addition to potential improvements to our tool. However, this did not occur. We may have had more in-depth insights as to new functionalities had we:

- Interviewed practitioners who were more familiar with or used physiological measures in their current practice
- Had practitioners used the tool over longer periods of time. In the sessions, interviewees had only between 25 to 35 min to use the tool and complete their task

6 Conclusion

The use of physiological measures, in combination with traditional methods, could help practitioners to better measure UX, as they each provide complementary information on how users feel about a system, game, or web interface. [12] While traditional evaluation methods can offer episodic data, i.e. before or after the interaction, physiological measures can provide moment-to-moment information [9]. The addition of

physiological measures can help us identify the cognitive and emotional reactions users experienced using an interface, while a post-task interview can help us delve further, after we have identified these emotions.

The main research and development activities we undertake at the Tech3Lab aim at facilitating and fostering the adoption of new methodologies, such as eyetracking and physiological measures, in the fields of UX design and research. A first step towards this direction was the development of a physiological heatmaps tool to allow simpler and richer interpretation of physiological signals for UI evaluation. The interviews we conducted with UX practitioners were very helpful, in that they provided guidelines and user requirements insights for us to use in the development of future iterations to facilitate furthermore the adoption of physiological methodologies. Our next step will be to continue to develop our functionalities as well as find ways to simplify the data processing sequence associated with physiological signals, working closely with ergonomists and consultants of the industry to do so.

Acknowledgments. Authors want to thank Brendan Scully for manuscript revision and the UX practitioners who participated to this study. This research was supported by NSERC (National Sciences and Engineering Research Council of Canada).

References

1. Smith, T.F., Waterman, M.S.: Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981). doi:[10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5)
2. Zhihong, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(1), 39–58 (2009)
3. Calvo, R.A., D’Mello, S.: Affect detection: an interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affect. Comput.* **1**(1), 18–37 (2010)
4. Georges, V., Courtemanche, F., Sénécal, S., Baccino, T., Fredette, M., Léger, P.M.: (Forthcoming). UX heatmaps: mapping user experience on visual interfaces. In: 34rd Annual ACM Conference on Human Factors in Computing Systems
5. Hassenzahl, M., Tractinsky, N.: User experience-a research agenda. *Behav. Inf. Technol.* **25**(2), 91–97 (2006)
6. Roto, V., Vermeeren, A.P.O.S., Väänänen-Vainio-Mattila, K., Law, E., 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). ACM (2013)
7. Zaman, B., Shrimpton-Smith, T.: The FaceReader: measuring instant fun of use. In: Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles, pp. 457–460. ACM, Oslo (2006)
8. Boucsein, W.: *Electrodermal Activity*. Springer, Berlin (2012)
9. Huisman, G., van Hout, M., van Dijk, E., van der Geest, T., Heylen, D.: LEMtool: measuring emotions in visual interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 351–360. ACM, Paris (2013)
10. Mirza-Babaei, P., Nacke, L.E., Gregory, J., Collins, N., Fitzpatrick, G.: How does it play better?: exploring user testing and biometric storyboards in games user research. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2013), pp. 1499–1508. ACM, New York (2013)

11. Hassenzahl, M., Burmester, M., Koller, F.: AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität (AttracDiff: A questionnaire to measure perceived hedonic and pragmatic quality). In: Ziegler, J., Szwillus, G. (eds.), *Mensch & Computer 2003. Interaktion in Bewegung*, pp. 187–196. B.G. Teubner, Stuttgart, Leipzig (2003)
12. Courtemanche, F., Léger, P.-M., Frédette, M., Sénécal, S., Georges, V., Dufresne, A.: Method and Product for Visualizing the Emotions of a User, Provisional patent application. US 62/121,552, p. 14 (2015)
13. Virpi, R., et al.: All about UX (2012). <http://www.allaboutux.org/>
14. Riedl, R., Léger, P.-M.: *Fundamentals of NeuroIS*. Springer, Berlin (2016)
15. Kivikangas, M., Nacke, L., Ravaja, N.: Developing a triangulation system for digital game events, observational video, and psychophysiological data to study emotional responses to a virtual character. *Entertainment Comput.* **2**(1), 11–16 (2011)
16. Dix, A., et al.: *Human-computer interaction*. Pearson Education (2004)