Titre: Title:	Emotional Maps for User Experience Research in the Wild
Auteurs: Authors:	Vanessa Georges, François Courtemanche, Marc Fredette, & Philippe Doyon-Poulin
Date:	2020
Type:	Communication de conférence / Conference or Workshop Item
Référence: Citation:	Georges, V., Courtemanche, F., Fredette, M., & Doyon-Poulin, P. (avril 2020). Emotional Maps for User Experience Research in the Wild [Communication écrite]. CHI Conference on Human Factors in Computing Systems (CHI 2020), Honolulu, Hawaï, États-Unis (8 pages). https://doi.org/10.1145/3334480.3383042

Emotional Maps for User Experience Research in the Wild

Vanessa Georges
Polytechnique Montréal
Montréal, QC H3T 1J4, Canada
vanessa.georges@polymtl.ca

François Courtemanche
Marc Fredette
HEC Montréal
Montréal, QC H3T 2A7, Canada
francois.courtemanche@hec.ca
marc.fredette@hec.ca

Philippe Doyon-Poulin
Polytechnique Montréal
Montréal, QC H3T 1J4, Canada
philippe.doyon-poulin@polymtl.ca

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3383042

Abstract
While most traditional user experience (UX) evaluation methods (e.g., questionnaires) have made the transition to the “wild”, physiological measurements still strongly rely upon controlled lab settings. As part of an ongoing research agenda, this paper presents a novel approach for UX research which contributes to this transition. The proposed method triangulates GPS and physiological data to create emotional maps, which outline geographical areas where users experienced specific emotional states in outdoor environments. The method is implemented as a small portable recording device, and a data visualization software. A field study was conducted in an amusement park to test the proposed approach. Emotional maps highlighting the areas where users experienced varying levels of arousal are presented. We also discuss insights uncovered, and how UX practitioners could use the approach to bring their own research into the wild.

Author Keywords
User Experience; Emotion Evaluation; Heat maps; Data Visualization; Physiological Measures; In the Wild Methods.

CSS Concepts
• **Human-centered computing~Visualization**; Visualization systems and tool; Visualization design and evaluation methods; Heat maps.

Introduction

In controlled lab settings, physiological signals can be used to infer users' emotional states during system interaction. Physiological and behavioral signals (e.g., electrodermal activity, heart rate, or facial expressions) can provide UX researchers and practitioners insights as to what users are experiencing without interference [11, 13]. These signals can also be used to uncover emotional states which the user himself is unaware of or cannot recall when asked using traditional evaluation methods, such as questionnaires and interviews [9]. While traditional UX evaluation methods have made the transition to the "wild", physiological measurements have yet to do so. This raises the question: How can we facilitate the transition of physiological measures from lab to their use "in the wild" for user experience evaluation? To meet this challenge, researchers have developed methods to infer users' emotional state and behavior using various sensors, such as GPS, accelerometers and facial expressions. For example, the Feel-o-meter project [18] used a digital camera to capture the facial expressions of passersby to produce a giant smiley face whose expression would reflect users' aggregated data. Also, using facial recognition software, Hernandez et al. [8] developed the Mood Meter, a vision-based computer system which generates affective portraits of various areas around the MIT campus.

Building on prior research contribution, we present a new approach which aims to facilitate the use of physiological measures in field evaluations. While UX heatmaps triangulates users' gaze data with inferred users' cognitive and emotional states to produce user experience (UX) heatmaps [33], the proposed method triangulates GPS and physiological data to create emotional maps, which outline areas where users experienced specific emotional states in outdoor environments. As part of the approach, a recording device and data visualization software were implemented.

Method

In the wild studies often entail a balancing act between experimental control and ecological validity. As such, researchers have argued that the concept of ecological validity should be regarded as a continuum [10]. Therefore, the minimal degree of experimental control that guarantees sufficient data quality should be aimed for.

Physiological measures

To validate the effectiveness of the proposed method arousal, which contrast states of low (e.g., calm) and high (e.g., surprise) intensity [20] was chosen as the physiological state of interest. Therefore, two main criteria guided the selection of physiological signals for our field study: 1) psychophysiological relevance with emotion intensity, and 2) readiness for deployment "in the wild" (motion artefacts robustness and intrusiveness). Respecting those requirements, we selected two signals that are indicators of the autonomic nervous system response to intense emotional state; electrodermal activity (EDA) and electrocardiography (ECG) signals [12].

EDA is composed of two different components: tonic skin conductance level (SCL) and phasic skin conductance response (SCR). The latter is known to reflect short-term responses of the autonomic nervous system and is a reliable indicator of emotional arousal [3]. In our experimental setup, EDA was recorded using

Figure 1: Portable device used to simultaneously record GPS data, electrodermal activity, and heart rate.

The device consists of a Bitalino (r)evolution Freestyle Kit (PLUX Wireless Biosignals S.A.) [1] set into a 3D-printed enclosure box. The enclosure box also includes a GPS module, a Lithium Ion Polymer (lipo) battery and a GPS antenna. Data is recorded on a micro-SD memory card.

two electrodes placed on the wrist of participants' non-dominant hand. Numerous studies showed valid measurements of skin conductance with sensor placement on the wrists and forearm [14]. To obtain clean phasic data, signal preprocessing steps were executed as follows. Data was recorded at 100 Hz and resampled to 25 Hz, before applying a low-pass 2nd order Butterworth filter and a 50 Hz cut-off. Signal was then decomposed in tonic and phasic components using the convex optimization algorithm described in [7].

ECG measures the electrical activity associated with contraction of the heart muscles by recording the potential difference it generates between two electrodes and a ground reference [2]. In our study, ECG was recorded using three electrodes placed on the participants' torso. ECG r-peaks were identified using the Biosppy python library for biosignal processing (*https://biosppy.readthedocs.io/en/stable/*), and interbeat intervals and heart rate (HR) were computed with pyhrv, an open-source Python toolbox for Heart Rate Variability (*https://pypi.org/project/pyhrv/*). To ensure high synchronicity and precise data triangulation, a recording device which acquires GPS and physiological data simultaneously was developed (see Figure 1).

Heatmap Generation

Heatmap generation consists of three main steps: accumulation, normalization, and colorization. In the accumulation step, an empty matrix is first created having dimensionalities corresponding to one of the original data space. In our approach, the accumulation matrix has the size of the image on which the heatmap is rendered, and the original space consist of the actual GPS coordinates of the participants (see Figure 2). The display image, a high-resolution satellite view of the amusement park, was extracted from Google Maps (*http://www.mappuzzle.se*) with a resolution of 2956 x 2585 pixels, corresponding to an actual surface of 635m x 559m. We computed the correspondence between spherical and planar coordinates using a pseudo-Mercator projection (*https://tinyurl.com/yycr85lv*). For small areas, such as the amusement park used in this study, linear curve fitting may have been sufficiently precise for location mapping. However, we used pseudo-Mercator projection to ensure that the method is applicable for studies taking place in larger areas (e.g. state or country level).

In our context, a data point consists of a GPS coordinate obtained at time *t* along with the average HR and phasic EDA values around time *t* (from 500ms before to 500ms after). For each data point, the corresponding matrix cell's *(x, y)* value is incremented by the signals' mean value.

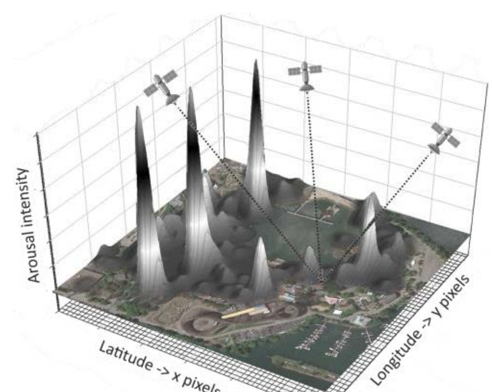

Figure 2: Height map illustrating participants' arousal distribution over the satellite image of the amusement park.

Figure 3: Equipment set-up. The sensor enclosure box was placed inside a belt bag and attached to participants' hip.

GPS data was recorded at 1Hz, collecting one data point per second. As illustrated in Figure 2, after the processing of all data points, the cell's values summation forms a height map with a topology proportional to emotion distribution. High peaks correspond to a high level of arousal.

In the case of physiological heatmaps, the normalisation step's goal is to account for inter-subject variations. As physiological signals are subject to significant interpersonal variations, values need to be corrected to account for the subject's baseline [17]. In our approach, the HR and phasic EDA values are normalised using z-score with the following equation:

$$V' = (V - \mu) / s$$

where V is the raw signal value, V' is the normalised value, and μ and s are respectively the participant's mean and standard deviation over the entire recorded period for each signal. Therefore, for a data point at coordinates (x, y), the matrix cell's value is incremented by $V'_{EDA} + V'_{HR}$. With multiple participants, the accumulation matrix is the sum of each participant's z-score at this coordinate. As movement artefacts didn't always occur at the same time in the two signals, if one of the measures was missing for that data point, we use the other one alone. For more details on the accumulation and normalisation algorithms, readers can refer to [6].

The last step in creating a heatmap is colorization. We overlaid on the satellite image a semi-transparent layer that reflects the height of each accumulation matrix's cell and showed the emotional variations. Accumulation matrices can be mapped to different color properties using a colorization function, resulting in various types of visualizations [5]. In this work, a four-colour rainbow gradient (blue, yellow, red and black) was used, where black indicates the highest emotional peak (see Figures 4, 5, and 6.

Experimental Validation

To validate the approach, a 42-participant field study was conducted at a theme park. While an amusement park is an ideal setting to measure a wide range of emotions at various intensity levels, this type of activity involves ample movements causing motion artefacts (e.g., walking from ride to ride, residual vibrations from the rides). Therefore, data from five participants were rejected due to insufficient data quality. Data from 37 participants were thus used in the analyses, of which 21 were female, for an average age of 26. Participants were pre-screened for cardiovascular diseases, epilepsy, motion sickness, vertigo as well as neurological and psychiatric diagnoses. Data was recorded during the amusement park's first four hours of operations in order to avoid overcrowding. During the experiment, participants were asked to complete five pre-selected amusement rides of three distinct thrill levels (i.e., high, moderate and low intensity levels) and of various movement categories (i.e., Ferris wheel, roller coaster ride, pendulum ride, drop tower and swing ride). The order in which to complete these rides was left up to participants. Participants wore lightweight protective gloves and wristbands to ensure electrodes would remain in place throughout the experiment (see figure 3). Electrodes were changed at the halfway point of the experiment to ensure adequate data quality over time. In addition to physiological data, self-reported data was also collected using questionnaires.

	Mean	Max
ECG	8.66 (.183)	4.034 (.055)
EDA	-4.95 (.422)	-3.11 (.223)

Table 1: Correlation between mean and max arousal user ratings for waiting lines and rides.

Due to low turnout, data collected at the Ferris wheel were not included in the analysis. Therefore, data from 4 out of 5 rides were used in the analysis.

P-values were corrected to account for the potential correlation between each repeated measure coming from the same subject by using a mixed linear regression model [16].

Self-reported arousal was obtained using a 9-point SAM scales [4]. Participants were asked to evaluate their experience at three different times: in the waiting line before every ride, during the ride (in low intensity rides) and immediately after each ride. In high and moderate thrill level rides, participants completed the form after the experience, indicating the quality and intensity of the emotion felt during the ride. Beginning and ending questionnaires were also used to assess overall user experience.

Preliminary Results and Discussion

Data was analyzed in order to evaluate the capacity of emotional maps to capture experienced arousal variance over the entire amusement park area We compared the max and mean arousal (area under the curve) of the height maps (see Figure 2) with the corresponding user ratings for both waiting lines and rides. Height maps were generated on a participant basis. The results presented in Table 1 show that participants' experienced arousal levels were significantly correlated to emotional maps' max values using ECG measurements (r=4.034, p-value=.055). However, the mean arousal was not significant (r=8.66, p-value=.183).This could be explained by the context of the experiment itself, i.e. words short and intense emotional experiences, as user ratings were most probably based on the single most intense arousal felt at any given point during the ride as opposed to the overall experience. Results for EDA were non-significant (r=-4.95, p-value=.422 and r=-3.11, p-value=.223), however empirical data are positively correlated. For a qualitative tool in its first iteration, we find these results very encouraging. The remaining questionnaire data have yet to be analyzed.

Emotional Maps

Two emotional maps are presented in Figure 4: the first one generated using participants' data during low traffic days and the second during high traffic. High traffic level may impact user experience in various ways: long queues for rides and concession stands, lack of seating at lunch, parking, etc. Figure 4 clearly shows that participants who visited the amusement park during low traffic days (10 800 visitors on average) experienced higher arousal levels throughout the site

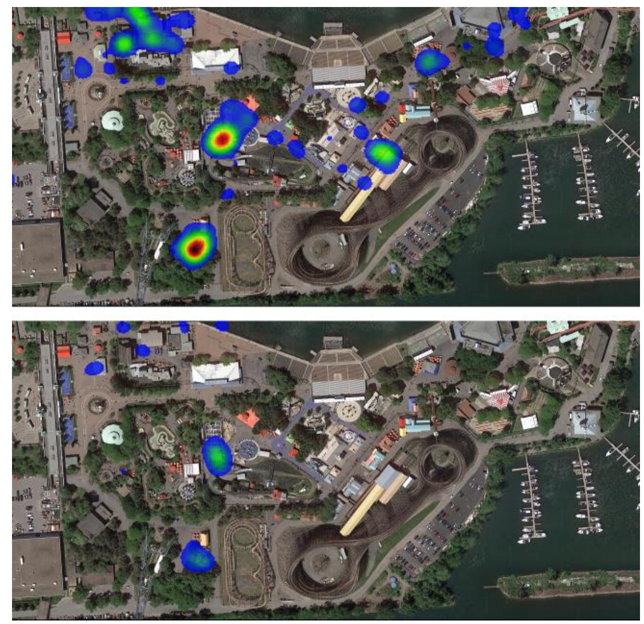

Figure 4: Above, arousal experienced during low traffic days. Below, arousal experienced during high traffic days.

compared to visitors during high traffic days (15 300 visitors on average). Here we used emotional maps to explore different phenomena, and hope to uncover more insights as data analysis continues. This type of

Figure 6: Above, participants who least appreciated this ride. Below, participants who most appreciated this ride.

Participants who rated this ride poorly (n=19, mean rating of 2 out of 7) experienced lower levels of arousal, compared to participants who appreciated the ride (n=18, mean rating of 4.5).

targeted investigation, where practitioners can explore and understand how specific elements (e.g. traffic, population subsets, etc.) can impact user experience, in this case how traffic levels influence experienced arousal, may lead to a better comprehension and development of user experiences.

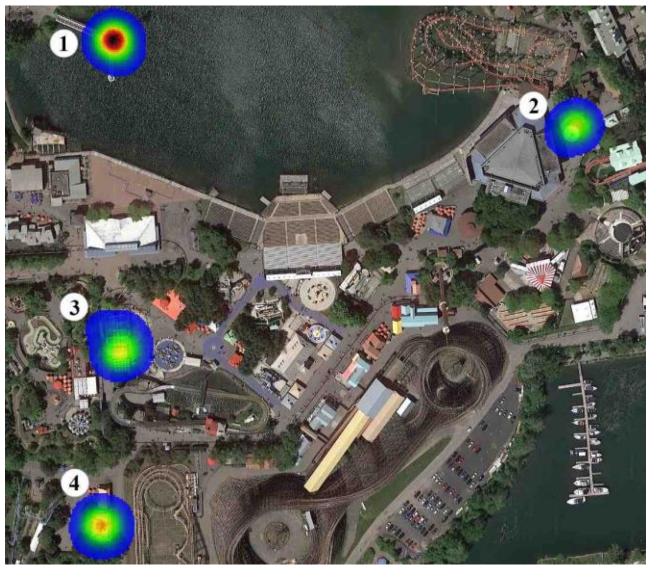

Figure 5: Emotional map generated using triangulated GPS and ECG data from 37 participants at an amusement park.

The robustness of the tool also allowed us to collect reliable emotional data in a context of high-intensity activities, which can be hard to achieve in the evaluation of user experience in the wild. Looking at Figure 5, we can not only locate the rides which generated the highest arousal, but also quantify the intensity of the emotion each ride generated relative to one another. Ride 1 was rated 8.9 by participants on an arousal scale of 9, followed by rides 4 (rated 7), 2 (rated 4.9) and 3 (rated 4.8). These results are consistent with the above visualization. The level of granularity that the approach enables allowed us to dig deeper into various emotional patterns, for example, the relationship between arousal and appreciation of the ride (see Figure 6).

The contextualisation of physiological signals and their application to user experience evaluation in the wild can offer some new opportunities to design innovative and engaging experiences. For example, emotional maps can help researchers and practitioners identify problematic areas in the user journey, and help uncover insights as to what users are feeling. Although there is more work left to be done to enable the use of physiological measurements in the wild, we hope this paper exemplifies the potential that these methods have in the evaluation of user experience in outdoor contexts.

Conclusion

This work aimed to develop an approach which would facilitate the use of physiological measures in user experience evaluation in the wild. In the future, we intend to adapt the approach to indoor contexts by acquiring localization data via Bluetooth beacons trilateration [15]. This would open up the approach to new contexts of use, including experiential or immersive environments. The inclusion of other physiological signals to the approach would also allow us to measure a broader range of emotional and cognitive states.

Acknowledgements

Authors would like to thank Ariel Garand and Julianne Chenier. This work was supported by NSERC.

References

[1] Batista, D., da Silva, H. P., Fred, A., Moreira, C., Reis, M., & Ferreira, H. A. (2019). Benchmarking of the BITalino biomedical toolkit against an established gold standard. Healthcare technology letters, 6(2), 32-36.

[2] Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). Handbook of psychophysiology. Cambridge University Press.

[3] Boucsein, W. (2012). Electrodermal activity. Springer Science & Business Media.

[4] Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry, 25(1), 49-59.

[5] Breslow, L. A., Trafton, J. G., McCurry, J. M., & Ratwani, R. M. (2010). An algorithm for generating color scales for both categorical and ordinal coding. Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur, 35(1), 18-28.

[6] Courtemanche, F., Léger, P. M., Dufresne, A., Fredette, M., Labonté-LeMoyne, É., & Sénécal, S. (2018). Physiological heatmaps: a tool for visualizing users' emotional reactions. Multimedia Tools and Applications, 77(9), 11547-11574.

[7] Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., & Citi, L. (2015). cvxEDA: A convex optimization approach to electrodermal activity processing. IEEE Transactions on Biomedical Engineering, 63(4), 797-804.

[8] Hernandez, J., Hoque, M. E., Drevo, W., & Picard, R. W. (2012, September). Mood meter: counting smiles in the wild. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (pp. 301-310). ACM.

[9] Ivonin, L., Chang, H. M., Chen, W., & Rauterberg, M. (2013). Unconscious emotions: quantifying and logging something we are not aware of. Personal and ubiquitous computing, 17(4), 663-673.

[10] Labonte-LeMoyne, É., Courtemanche, F., Fredette, M., & Léger, P. M. (2018). How Wild Is Too Wild: Lessons Learned and Recommendations for Ecological Validity in Physiological Computing Research. In PhyCS (pp. 123-130).

[11] Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. Behaviour & information technology, 25(2), 141-158.

[12] Iris, B. (2009). Mauss & Michael D. Robinson. Measures of emotion: A review, Cognition and Emotion, 23, 209.

[13] Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. Interacting with computers, 22(5), 336-343.

[14] Poh, M. Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. IEEE transactions on Biomedical engineering, 57(5), 1243-1252.

[15] Rida, M. E., Liu, F., Jadi, Y., Algawhari, A. A. A., & Askourih, A. (2015, April). Indoor location position based on bluetooth signal strength. In 2015 2nd International Conference on Information Science and Control Engineering (pp. 769-773). IEEE.

[16] Graybill, F. A. (1976). Theory and application of the linear model (No. QA279. G72 1976.). North Scituate, MA: Duxbury press.

[17] van den Broek, E. L., Janssen, J. H., van der Zwaag, M. D., & Healey, J. A. (2010, January). Prerequisites for Affective Signal Processing (ASP)-Part IV. In Biosignals (pp. 188-193).

[18] Wilhelmer, R., Bismarck, J., & Maus, B. Feel-o-meter: Stimmungsgasometer. (2008).

[19] Georges, V., Courtemanche, F., Senecal, S., Baccino, T., Fredette, M., & Leger, P. M. (2016, May). UX heatmaps: mapping user experience on visual interfaces. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 4850-4860). ACM.

[20] Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.