New Methods for Measuring Emotional Engagement

Andrew Schall

SPARK Experience, LLC
Bethesdsa, Maryland, USA
andrew@sparkexperience.com

Abstract. Truly understanding the feelings of a user has always been a dream of user experience (UX) researchers. Current methods for understanding emotional response has been limited to self-reporting from study participants or qualitative methods such as surveys or focus groups. New biometric and neurometric devices allow us to collect behavioral data in ways that were not previously practical for user researchers. This paper will provide an overview of these new technologies and how they can be applied to the study of emotional responses during user experience evaluation.

Keywords: Emotion research, emotional design, user experience, physiological measurements, biometric, neurometric, EEG, eye tracking, GSR, facial response analysis.

1 Introduction

Truly understanding the feelings of a user has always been a dream of user experience (UX) researchers. Are they enjoying the experience? Are they frustrated? Are they truly interested and engaged? The broader definition of user experience has grown to extend beyond basic usability. Understanding how a user truly feels in reaction to an experience can help us to optimize specific aspects of the design to exude certain specific emotional states.

Standard user research methodologies rely on either observing the user, or by directly asking the user for input. A common way to address emotional and cognitive aspects in user experience testing today is through retrospective self-report where users are asked to describe or answer questions about their experience after it has been completed, either verbally or through a questionnaire [1]. While these methods are commonplace, they rely too heavily on the highly subjective nature of participant's interpretation and recollection of their emotions. They are also too limited in their capacity to identify changes in emotional or cognitive processing over the course of a test, unless the user is constantly interrupted with questions, which would have a negative impact on the authenticity of the user experience.

The ability to capture biometric and neurometric measurements has existed for over 100 years, predominantly in an academic or clinical setting. Skin conductance,

A. Marcus (Ed.): DUXU 2014, Part IV, LNCS 8520, pp. 347–357, 2014.

respiration, electrical brain activity, pupillary size and cardiovascular activity have all been reported to vary in response to factors such as task difficulty, levels of attention, experiences of frustration and emotionally focused stimuli. Biometric and neurometric measurements have been in use by the cognitive psychology and neuroscience fields for decades, however the extreme complexities in both data collection and analysis have previously made these techniques impossible for those outside of academia or these highly specialized fields. It has been proposed that physiological data might be a valuable tool for user experience testing, as it could help identify significant events in cognitive and emotional behavior [2].

New biometric and neurometric devices, which are practical, reasonably priced, and suitable for UX practitioners, have evoked both a substantial amount of enthusiasm and skepticism. Biometric and neurometric measurements allow us to collect behavioral data in ways that were not previously possible. Researchers can use these measurements when they are interested in understanding the user's emotional reaction at a certain point in time such as when a specific stimulus is displayed, or to catch the overall emotional reaction over a longer period of time that can include the entire interaction with the stimulus. The primary objective of this paper is to provide an introduction to biometric and neurometric tools that can be used by the user experience research community. A secondary goal is to address the specific benefits and challenges of these new tools to accurately and reliably deduce emotional responses in UX research.

2 Physiological Measurements for User Experience Research

There are numerous biometric and neurometric tools and measurements that can be used to gain a deep understanding of human cognition and emotional response. However, many of these such as fMRI (Functional Magnetic Resonance Imaging) MEG (Magnetoencephalography), PET (Positron Emission Tomography) are extremely expensive, highly intrusive, and go well beyond the skillsets of a typical user researcher. This paper focuses entirely on tools that are accessible to those in the UX field including the use of eye tracking, GSR (Galvanic Skin Response), EEG (electroencephalography), and facial response analysis.

2.1 Eye Tracking

Eye tracking is a methodology that helps researchers understand visual attention. Using eye tracking we can detect where users are looking at a point in time, how long they look at something, and the path that their eye follows. Eye tracking has been applied to numerous fields including human factors, cognitive psychology, marketing, and the broad field of human-computer interaction [3].

We are at the beginning of a golden age for eye tracking in user experience research. Most major academic and commercial labs have an eye-tracker, or plan to purchase one in the near future. The primary reasons for this increase in adoption have been ease of use for the researcher and being considerably more participant-friendly. In the past, eye tracking has only been accessible to those with a highly advanced understanding of human physiology, engineering, and computer science. Users of these systems had to have extensive training in order to properly operate the equipment. Making sense of the data was extremely cumbersome and time-consuming, requiring researchers to do analysis by hand.

Advancements in remote eye tracker technology now make it possible to calibrate the equipment with the participant's eyes easily in a matter of seconds [4]. Eye trackers today are extremely accurate, can track a diverse population, and retain their calibrations for long periods of time. The operation of eye trackers today requires significantly less training and does not require a dedicated technician during use.

Gone are the days of clamping down a participant's head into a vice and sticking a bite bar into their mouths. Today's eye tracking technology has been miniaturized and integrated into computer monitors (see Figure 1) or as standalone devices no longer physically connected to the participant. The technology is so covert that participants would have no indication that they are being tracked except for the brief calibration that takes place before the beginning of the study session. As researchers, we want the eye tracker to be completely unobtrusive; we want participants to forget that it is even there.

Fig. 1. Tobii T60 Eye Tracking System (Source: Tobii Technology)

Eye tracking is a powerful tool for user researchers and when properly used can provide insights unachievable by other research methods. The most obvious but unique ability of the eye tracker is that it can track the location of a participant's eyes.

The visual hierarchy of an interface dictates what a user will pay attention and when. This sequence of visualizations can be critical for both the usability of a system and consumption of content. Our visual field is constantly bombarded by a variety of stimuli. We are overloaded and overwhelmed by visual information and constantly resort to prioritizing what we choose to pay attention to. In order to measure the effectiveness of content researchers need to determine what users are looking at and what they choose to engage with.

In trying to understand what users decide to pay attention to we can't always rely on the participants to accurately tell us. Participants are terrible at self-reporting where they looked. For the most part, this is due to our eyes often moving involuntarily and the limits of our short-term memory. Guan et al. [5] measured the extent to which participants did not discuss elements that they in fact visually attended to. They labeled these as omissions. Participants had omissions 47% of the time, meaning that almost half of the time they did not mention elements that they looked at. Omissions may have occurred because participants forgot about seeing the elements, or perhaps simply because they just didn't think or care to mention them. It should also go without saying that a researcher can't simply ask a participant if they noticed a certain on-screen element. This action draws the participant's attention directly towards something that they may or may not have originally seen. This inherently and irreversibly biases the participant and no confident answer can be obtained. Eye tracking provides an objective running commentary of where the individual looks without any need for participants to verbalize what they have seen.

Eye tracking is an essential tool to combine with any biometric or neurometric measurements. These measurements are useless unless they are analyzed in context of what the user is observing. Time-locking eye tracking data with these measurements is key to understanding exactly when a participant was looking at something at exactly what they were seeing.

2.2 GSR

Galvanic skin response (GSR) has long been used as to measure physiological arousal [6]. GSR can provide researchers with a spectrum of states from being high aroused/engaged/stressed to a state of noninterest/unengaged/relaxed. This measurement is ideal for detecting situations where a user is having difficulty using an interface and increasingly becomes frustrated and stressed. For applications designed to keep a user actively engaged or interested, GSR can help measure the intensity of their engagement as well as how long it can be sustained. GSR is incapable of representing a broader set of emotional states such as EEG or facial response analysis, which can both detect levels of valence (positive to negative emotions).

The measurement of GSR is dependent on the levels of sweat within the skin. The more sweat produced, the higher the level of electrical conductance that can be measured. In order to obtain the galvanic skin response, a small electrical current is passed through the skin using a pair of electrodes. The soles of the foot or palms are a recommended location for these electrodes due to the higher amount of sweat produced by these areas. However, in user experience research it would be prohibitive to use these areas of the body to measure skin conductance. Newer devices such as the Shimmer3 (see Figure 2) are far less intrusive and allow for total freedom of movement while interacting with devices. According to Shimmer [7] their new GSR module, "brings an effective way to measure activity, emotional engagement and psychological arousal in lab scenarios and in remote capture scenarios that are set outside of the lab." In addition to participant comfort, the device is also fairly stress-free for researchers who want to quickly setup and gather data. The device includes a built-in Bluetooth receiver that can easily be paired with a laptop and software is provided for visualizing the GSR data.

Fig. 2. Shimmer 3 GSR Unit (Source: Shimmer)

Researchers should be aware that the results from GSR do not correspond with a response to a stimulus in real-time. GSR can produce response latencies between 3 to 6 seconds from the response to a stimulus [8]. Therefore it is not recommended to use GSR to identify the exact moment when a response was triggered. Instead, researchers should analyze the response over a period of time, for example the duration of a task or the presentation of a stimulus, and then compare the result to other such units. For all biometric measurements, but especially with GSR, it is critical to obtain a baseline measurement prior to the presentation of stimuli. Participants will vary in terms of their typical level of sweat output and their emotional state (e.g. feeling anxious) when they arrive at the test facility. By establishing a baseline it provides a point of comparison between their state prior and after a stimulus has been shown.

2.3 EEG

EEG measures electrical activity in the brain by placing electrodes along various points along the scalp. The signals obtained from these electrodes are represented by waveforms reflecting voltage variation over time [9].

EEG raw data is measured at the millisecond level and can be directly attributed to stimuli effects in real-time. EEG units traditionally used in academic settings use

Fig. 3. Emotiv EEG headset (Source: Emotiv)

a skullcap with numerous electrodes connected via wires and require the use of conductive gel. These units, while highly accurate take a long time to setup and are extremely intrusive for the participant. More recent EEG models such as the Emotiv EEG headset (see Figure 2) are completely wireless and use over-the-counter saline solution to provide conductivity for the electrodes. These headsets can be worn comfortably during a user experience test and minimally interfere with a participant's natural behaviors. The trend towards less expensive, lighter weight, and totally wireless solutions will make EEG even more practical for UX researchers within the next few years.

Lee and Tan [10] found through their interactions with other HCI researchers that there is a concern over lack of domain knowledge and because of the high cost of owning and maintaining the EEG equipment. While EEG headsets are not mind reading devices, they can give us an accurate sense of what a participant is feeling. Emotional states can be complex. We often feel a composite of emotions at any given time and those emotions can be internalized and unrelated to what is being shown on the screen.

It is true that correctly interpreting the meaning of EEG waveforms and translating that data into emotional states is extremely complex and likely out of the expertise of a UX researcher. However, new analysis tools such as Emotiv's Affectiv Suite [11] processes the raw EEG data and produce visualizations that correspond with a standardized set of emotional states (e.g. engagement/interest, frustration, happy/sad, etc.)

2.4 Facial Response Analysis

Facial coding is the systematic analysis of facial expressions. Research on facial coding dates back to studies by Charles Darwin who concluded that common facial expressions are universal. In the 1970s, psychologist Paul Ekman's early work identified the universality of six core emotions. He also is well known for popularizing a facial action coding system (FACS) that systematically describes facial expressions and movements12.

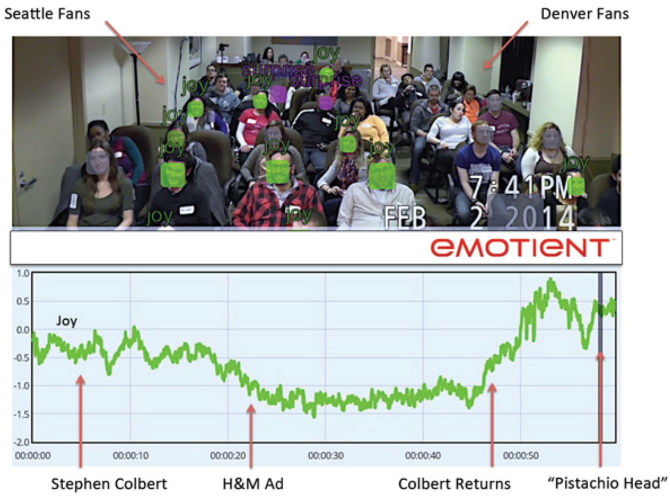

Fig. 4. Facial Analysis Software (Source: Emotient)

Companies like Affdex [12] and Emotient have developed new software that can be used to analyze a user's emotions by examining their facial reactions. These systems use computer algorithms that take video from common webcams as inputs and provide frame-by-frame emotion metrics as outputs [13]. Webcams are already commonly used to capture nonverbal behavior and audio from participants in user experience studies. Using facial analysis is one of the least intrusive methods for capturing emotional reactions in an automated manner. Both companies claim that they are able to capture subtle emotions from only small facial muscle movements called facial action units. The ability to capture these less expressive types of emotions or varying levels of valence is critical for user experience research where participants do not always have a strong outward reaction to stimuli.

3 Benefits of Physiological Measurement to User Experience Research

User experience testing, which has its grounding in usability testing has traditionally not focused on measuring emotions. Usability testing has often focused on efficiency and easy of use. Researchers didn't always concern themselves with whether the interface was enjoyable to use or caused any other emotional response. More recently as those in the field of HCI have expanded their viewpoints to a more holistic view of user experience, finding a way to measure emotion has become increasingly important. A primary limitation has been the inability to accurately measure emotional response in a practical, accurate, and minimally intrusive way.

UX research is often centered on gaining insights directly from participants to truly understand their experience. However, it is important to recognize that participants (as well as researchers) are not always objective, and fall prey to the weakness of the human mind. Physiological measurements remove the subjectivity of evaluating user experience by relying exclusively on quantitative metrics that are the output of devices that measure primarily involuntary, often subconscious responses to stimuli.

UX researchers frequently need to balance the need for a user to interact with a system without constantly being interrupted, with the need to understand what they are thinking or feeling. Diricana recognizes this need in HCI and states that, “Changes in physiological signals can also be examined for signs of stress arising while users interact with technology, helping detect where the product causes unnecessary irritation or frustration, without having to interrupt the user or record her appearance.” [14] Using methods such as eye tracking and physiological measurements we can gain a deep understanding of what a user is paying attention to and how they are feeling without the need to interrupt a participant during an activity. With certain measurements that can be observed in real-time, we also have the benefit of being able to discuss the output of these tools with participants using a retrospective technique. This may help to validate and augment our research findings based on what the physiological data is telling us and what we can learn from discussing these findings with participants. We still need to have a dialogue with participants because the physiological data tells us what they were looking at and what they were feeling, but it ultimately does not tell us why they were feeling that way.

The benefit of using multiple types of bio/neurometric devices is that we can learn different things from different devices. Valence is a measure of the positive or negative nature of the participant’s experience with the stimulus. Using EEG and facial response analysis we can measure whether the participant is having a relatively good or bad reaction to their experience. GSR and heart rate cannot measure valence, but are a good indicator of a participant’s level of arousal, which depending on the reaction, can tell us whether they are feeling stressed, engaged, or relaxed. Another benefit of using multiple types of measurements is that we can often use them to validate or invalidate each other. For example, if our facial analysis data is strongly indicating that our participant is experiencing great happiness, but the EEG data is indicating high levels of sadness we know that one of the measurements is likely reporting incorrect information.

4 The Challenges of Measuring Emotions

When biometric measures are applied either in a controlled lab or in real environments, there are many issues that must be considered. Conducting these types of studies in a real world environment presents several additional challenges and data quality can be a significant issue [14]. However like any lab-based study, we are faced with the artificially of a controlled environment that ignores all of the potential stimuli a user would likely experience in the real world. Ultimately, it becomes a trade-off for the research team as to what is most important to understand in their study.

Today applications are not only accessed through computers, and are available on a variety of platforms from tablets to smartphones, and even wearable devices. These mobile devices have previously made it difficult to collect physiological data outside of the lab environment. However, great progress has been made in this area over the last few years. Eye tracking vendors such as Tobii and SMI have recently developed wireless glasses that can be used to track participants' eyes. These may not be unobtrusive enough for users to completely forget that they are taking part in a study, but they are practical enough to allow for free body movement during the session. Similar progress has been made in the field of EEG with products such as the Emotiv EEG headset. New GSR units have recently become available including the Shimmer3 unit that measures both electro dermal activity as well as heart rate.

Even with new, more versatile equipment, UX researchers will need to possess the technical competence required to set up and operate advanced equipment and ensure a rigorous process is in place to collect accurate data. Pilot testing is essential to these types of studies to determine if the equipment is properly configured and outputting the expected type of data. Equipment must be carefully calibrated with each participant and baseline measurements should be taken to account for variations between individuals. Physiological studies also require higher sample sizes than typical qualitative research projects.

Another challenge lies in the interpretation of data, since the same kind of physiological responses may be observed for different mental states, such as frustration, surprise or increased cognitive effort. Therefore, a correct interpretation requires knowledge of the context in which the data was obtained. In order to better understand the results, it is thus advisable to record additional observations along with the physiological measurements, such as comments, observed behaviors and subjective ratings of events [15].

5 Conclusion

The ability to capture the emotions of our users is becoming an increasingly important aspect of user experience. Users are no longer satisfied with interfaces that simply meet their basic needs in terms of usability. Existing methods of measuring emotional response are flawed and rely too much on self-reporting and other highly subjective measurements. There is an inescapable need to find new ways to objectively measure the complex emotional experiences that result from interacting with digital products. The biometric and neurometric devices discussed in this paper have been identified as having the highest potential for application to HCI research. All of the devices have the capability of being integrated into the existing methods used by user researchers to measure user experience.

The benefits of these tools also come with significant challenges for user researchers. All of these tools originate from unfamiliar fields such as human physiology and neurology, which can be intimidating and potentially risky for those in HCI to adopt. Significant investment is required to purchase the necessary equipment and to employ researchers with a sufficient level of understanding in physiological measurement. Additional time is required to analyze the abundant amount of data that comes from these measurements and then to extract meaning that can be useful for user experience designers.

There are still significant challenges to implementing these new measurements, however the current generation of tools is considerably more economical and practical for UX researchers than ever before, and all indications are that this trend will continue over the next several years. Eventually we will reach a point where collecting physiological data that helps us understand our user's emotions will become commonplace. This eventual enlightenment will bring about interfaces that can be crafted to evoke specific emotional experiences from our users.

References

1. Sherman, P.: How Do Users Really Feel About Your Design? (2007), `http://www.uxmatters.com/mt/archives/2007/09/how-do-users-really-feel-about-your-design.php`
2. Ward, R., Marsden, P.: Physiological responses to different WEB page designs. International Journal of Human-Computer Studies 59, 199–212 (2003)
3. Schall, A., Romano Bergstrom, J.: Eye tracking in user experience design. S.l. Morgan Kaufmann Publisher (2014)
4. Tobii Technology, An introduction to eye tracking and Tobii Eye Trackers (2010), http://www.tobii.com/Global/Analysis/Training/WhitePapers/Tobii_EyeTracking_Introduction_WhitePaper.pdf
5. Guan, Z., Lee, S., Cuddihy, E., Ramey, J.: The Validity of the Stimulated Retrospective Think-Aloud Method as Measured by Eye Tracking. In: Grinter, R., Rodden, T., Aoki, P., Cutrell, E., Jeffries, R., Olson, G. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1253–1262. ACM Press, New York (2006)
6. Picard, R., Scheirer, J.: The Galvactivator: A glove that senses and communicates skin conductivity. Paper presented at Proceedings 9th Int. Conf. on HCI (2001)
7. Shimmer. Motion, Galvanic Skin Response and Optical Pulse captured from one device (December 11, 2013) (press release)
8. Park, B.: Psychophysiology as a Tool for HCI Research: Promises and Pitfalls. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part I, HCII 2009. LNCS, vol. 5610, pp. 141–148. Springer, Heidelberg (2009)
9. Niedermeyer, E., da Silva, F.L.: Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Lippincot Williams & Wilkins (2004)
10. Lee, J.C., Tan, D.S.: Using a low-cost electroencephalograph for task classification in HCI research, p. 81. ACM Press (2006)
11. Emotiv.com (2012), EEG Features, http://emotiv.com/eeg/features.php (accessed: Decmeber 11, 2013)

12. Picard, R.W., Vyzas, E., Healey, J.: Toward machine emotional intelligence: Analysis of affective physiological state. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(10), 1175–1191 (2001)
13. Kanan, E.: Exploring the Emotion Classifiers Behind Affdex Facial Coding. (whitepaper) (2013)
14. Diricana, A., Göktürk, M.: Psychophysiological Measures of Human Cognitive States Applied in Human Computer Interaction. Paper presented at World Conference on Information Technology. Elsevier (2011)
15. Madrigal, D., Mcclain, B.: Testing the User Experience: Consumer Emotions and Brand Success: UXmatters (2009), http://www.uxmatters.com/mt/archives/2009/10/testing-the-user-experience-consumer-emotions-and-brand-success.php (accessed: December 12, 2013)