# Is There a Role for Physiological Methods in the Evaluation of Human-Information Interaction?

Mahria Lebow and Heather L. O'Brien*

School of Library, Archival and Information Studies

University of British Columbia, Vancouver, British Columbia

** Corresponding Author Contact Information*

1961 East Mall, Suite 470 Irving K. Barber Learning Centre,

Vancouver, British Columbia, Canada, V6T 1Z1

Telephone: 604-822-6365

Fax: 604-822-6006

Email: h.obrien@ubc.ca

**Please cite this paper as: Lebow, M. & O'Brien, H.L. (2012). Is There a Role for Physiological Methods in the Evaluation of Human-Information Interaction? Winner of the ALISE/Dialog Methodology Paper Competition, Association of Library and Information Science Education, January 2012.**

**Abstract**

In human-information interaction (HII) we are facing new research challenges as we attempt to look more holistically at the relationship between people, technologies, and information objects. These challenges revolve around understanding how the interaction process changes over time and evaluating emotional responses during interaction. The use of physiological measures is becoming prevalent in human-computer interaction (HCI) research to detect emotional responses during technology use and to design responsive computer devices. In this paper, we explore the collection, analysis and interpretation of physiological measures through the research literature and our own experience of employing them in a research study, with the overarching question, "What is the potential for physiological measures in the study of HII?

**Introduction**

Human-Information Interaction (HII) has been described as a "core phenomenon for the information field" (Marchionini, 2008, p. 171). HII is comprised of information objects (e.g., books, videos, etc), people, and technology (e.g., devices, software). Although these elements have always been a focus of information behaviour research, HII stresses the holistic examination of these elements. Marchionini (2008) describes this as moving away from a reductionist approach that focuses on objects, people, and technology as independent entities and towards a more ecological perspective. This shift in viewpoint challenges us to reconsider the ways in which we contemplate information interactions, thinking more broadly than single, information encounters to cumulative information experiences that are influenced by individual differences, dynamic information systems and objects, evolving information needs and contexts, and emotion. Viewing information interactions as experiences that unfold and change over time better reflects the complexity of today's users, technologies, and information objects, but with this complexity comes new challenges, namely, "How do we evaluate information experiences?"

Information scientists use a range of qualitative, quantitative, and mixed methods to measure human information interaction. Kelly (2009) classifies four types of measures: contextual, interaction, performance, and usability. Contextual measures document characteristics about people (e.g., age, level of knowledge about an information domain or tool) and about the information environment (e.g., workplace). Measures of interaction address what took place when a user and a system met at the interface (e.g., how many queries did the searcher pose and what content did they examine?), while performance measures evaluate the objective outcome of the interaction, e.g., did the searcher find information to address an information need? The final group of measures, usability, focuses on users' perceptions of the interaction, such as overall satisfaction or attitude toward a search tool. A number of measures from each of these categories are typically collected in HII research, whether the study is situated in a naturalistic (e.g., library, Web) or experimental (i.e., laboratory) setting, and may be operationalized using questionnaires, interviews, focus groups, observation, eye-tracking, log analysis, etc., all of which have their benefits and drawbacks (Kelly, 2009).

The abundance of methods and measures allows researchers to adapt their approach according to the community of users, technologies, and contexts being investigated. Yet, despite this richness, there are still important measurement challenges to be addressed (Marchionini, 2008; Saracevic, 1997). Given the complexity of HII, we need to develop measures to 1) capture the dynamic nature of information interactions (i.e., how information needs, content, and situations change over the course of an interaction); and 2) link affective responses to specific aspects of experience (Kelly, 2009). At what point during an interaction did the user become frustrated with the system, or encounter information that made him more engaged in the information task?

Responding to these two challenges will modify the way in which we think about evaluation. Typically, we are focused on outcome measures (e.g., time on task, number of documents located, users' overall satisfaction rating), but the need to capture the dynamics of interactions and emotion compels us to think more about process measures with the emphasis on the central but seldom explored "shifts" in interaction (Saracevic, 1997, p. 6). Physiological measures have been suggested as a means of capturing these shifts (O'Brien &

Toms, 2008; Kelly, 2009), because they present a continuous, real-time measure of the user's internal state as they perform a task (Rowe, Sibert & Irwin, 1998).

Physiological measures, specifically, heart rate (HR), eletrocmytogram (EMG), electrodermal activity (EDA), and respiration, are being increasingly used in human-computer interaction (HCI) to design interactive devices responsive to various user groups, e.g., physically disabled (Ahsan, Ibrahimy, & Khalifa, 2009) or situations, e.g., managing interruptions (Chang, McGenere & McLean, 2011), or to detect emotional responses when people are performing tasks, such as listening to music, driving (Healey, 2008), playing a video game (Mandryk, Inkpen & Calvert, 2006), or using different versions of an interface (Mahlke & Minge, 2008). This work has focused on using physiological responses to control computer applications and on the objective evaluation of emotion. Might it provide insight as we attempt to grapple with the complexity of HII, specifically the need to capture the dynamics of interaction and connect components of interaction to emotion?

**The Application of Physiological Methods in HII**

Though there are currently no published studies in information science that utilize physiological measures, researchers have been considering their potential to evaluate emotion (Lopatovska & Arapakis, 2011) and explore engagement with information systems (O'Brien & Toms, 2008). Our interest lies in the latter: documenting changes in user engagement over the course of an interaction. O'Brien and Toms (2008) define engagement as "a quality of user experience with technology" (p. 949) that is comprised of various stages: a point of engagement, period of sustained engagement, disengagement, re-engagement, and non-engagement. The user moves through these stages over the course of an interaction depending on their motivation and interest, the novelty, aesthetic or sensory appeal, and usability afforded by the system, etc. This process model, based on interview data, could be strengthened and validated if we were able to objectively identify movement between these stages. Furthermore, identifying these shifts would tell us more about the experiential needs of users, specifically how to design information applications that meet those needs. For example, is there a design feature that consistently disengages users?

The desire to understand the nature of users' engagement with information systems corresponds with the measurement challenge identified by Kelly (2009) to evaluate dynamic interactions. We turned to physiological measures as a means of supplementing other measures that did not quite provide the whole picture. For example, we considered usability measures (i.e., self-report interviews, questionnaires, etc.) to be a reliable measure of users' perceptions of their own experiences (i.e., engaging, non-engaging). However, they are generally collected at the end of a study, and may not allow us to identify the points during the interaction that contributed to this outcome assessment. In addition, we were not confident that performance measures, such as time on task or number of mouse clicks, would be indicative of engagement, which is less about efficiency and more about experience. Lastly, existing interaction measures of user's actions seemed important for engagement, but which ones? How could we identify behaviours indicative of meaningful engagement?

Our attempts to gain a more holistic picture of engagement led us to physiological measures. Our goal was to isolate points in time during the user-computer interaction that represented users' state of engagement, and to correlate process and outcome measures. In this paper, we document our journey to realize these objectives as we explored user engagement with ten participants browsing an online news website, an interactive experience typical of those studied in HII. Drawing upon our experiences and the research literature, we describe collecting, analysing and interpreting physiological data and the challenges therein. We conclude with some thoughts on the value of physiological methods to our own work and HII more broadly.

**Collecting, Analyzing and Interpreting Physiological Data**

Among the most established and frequently used physiological measures are eletrocmytogram (EMG), electrodermal activity (EDA), heart rate (HR), and respiration. In this section we discuss the collection, analysis and interpretation of physiological data. We begin with an overview of these measures and our rationale for using them in our research.

Eletrocmytogram (EMG) measures the electrical activity associated with the contraction of skeletal muscles (Stern et al., 1980, p. 108). Of particular interest to researchers studying emotion are two facial muscles: zygomaticus major, located along the jaw line, and corrugator supercilli, located near the centre of the forehead; these have been used to gauge “smiling” and “frowning” or “brow furrowing,” respectively (Ravaja, 2004). Corrugator and zygomatic activity are popular and fairly robust for differentiating valence (i.e., categorizing emotions as either negative or positive): corrugator and zygomatic activity have been demonstrated to have a negative and positive association with pleasure, respectively (Cacioppo, Petty, Loch and Kim, 1986; Dimberg, 1990; Lang, 1993). In our study that involved web browsing, we were interested in detecting affective responses to news content or the website.

Electrodermal activity (EDA) refers to the measurement of the amount of sweat produced by the eccrine glands, controlled by the sympathetic nervous system, found on the palm surface of the hands and feet (Stern et al., 1980, p. 198). EDA is used interchangeably with other terms, including galvanic skin response (GSR), skin conductance response (SCR), and skin conductance level (SCL), but EDA is the most complete and technically correct label for this measure (Stern et al., 1980). EDA has been positively correlated with arousal (Lang et al., 1993), and is often used to capture the body's immediate response to a specific, novel stimulus or "orienting response" (Stern et al., 1980), which may be indicative of attention (Frith & Allen, 1983; Cacioppo & Tassinary, 1990). Since novelty and attention are two characteristics of engagement, we wanted to examine EDA with on-screen activity to see if we could determine the point at which participants became engaged with the content or interactive features of the news website.

Heart rate (HR) is the measure of the number of beats the heart makes over time, i.e., beats per minute. The heart muscle is controlled by both the parasympathetic and sympathetic nervous systems (Stern et al., 1980). Researchers examine heart rate, but also look at the change in the interval between consecutive beats or heart rate variability (HRV). HRV is sometimes referred to as the inter-beat interval (IBI) or the R-R interval. The latter abbreviation refers specifically to the interval from one R-wave to the next R-wave (the R-Wave is first upward stroke of the heart beat); however, a complete heartbeat consists of several different waves and, therefore, several intervals. Heart rate can be measured directly by placing electrodes on the participant's chest (ECG) or can be calculated retrospectively from the signal data collected from a sensor that measures blood volume pulse (Berntson et al., 1997). An increase in heart rate has been associated with fear or anger (Levenson, 1992), or increased cognitive demands (Allanson & Fariclough, 2004), i.e., attention (Frith and Allen, 1983) or task difficulty (Carroll, Turner & Hellawell, 1986). HRV has been used to measure cognitive load or stress (Porges & Byrne, 1992), decreasing when mental effort is exerted or an individual is under stress, and increasing when an individual is relaxed or their mental effort is not taxed (Mandryk, Inkpen & Calvert, 2006; Rowe, Sibert & Irwin, 1998). Engagement is a positive state, so indicators of stress during an interaction may indicate that users are disengaging because the task or interface is too challenging, frustrating, etc.

Respiration data may be used to interpret both the participant's frequency of breath (respiration rate) and deepness of breaths (respiration amplitude); these respiratory patterns have been associated with task demand (Allanson & Fairclough, 2004). According to Stern, Ray and Davis (1980), respiration rate is collected primarily to identify artefacts or "noise" in the data of other physiological measures, namely HR and SCR. Therefore, respiration data can be used as a cross check for activity unrelated to the stimulus or condition of interest; this was the reason that we chose to include this measure.

*Collecting Physiological Data*

At its most basic, a physiological response is the electrical signal produced by hundreds of cells responding to a stimulus (Stern et al., 1980, p. 22). Measures of EMG, EDA and HR are collected via an electrode placed on the surface of the participant's skin; respiration data is collected from a band sensor wrapped around the participant's lower chest. The placement of electrodes varies depending on the specific measure. Electrodes do not pierce the skin, but rather are held on via an adhesive or wrap. Depending on the electrode, jelly or paste may be required to reduce impedance of the electrical signal as it travels from skin to sensor. Once in place, the electrode or respiration sensor captures the underlying bioelectric signals and transmits them to a receiving system, typically a computer. A separate electrode is needed for each signal, with some signals needing more than one.

Once the devices are connected and transmitting data to a receiving device, a baseline of physiological activity at a resting state is recorded. In our study, we captured a three-minute baseline at the beginning of the task. In addition to gauging resting state, this is also an opportunity to ensure that the sensors are connected correctly and transmitting properly. It was challenging to find documentation about what values should be seen during the baseline. We relied upon advice from colleagues with experience collecting physiological data, who said the range for EDA should be above 1 and below 20 Hz and that, when an individual is at rest, BVP and heart rate should be about the same, for example. Mandryk, Inkpen and Calvert (2006) recommend collecting several baselines throughout the experimental session and averaging them together retrospectively, in the event that people are anxious or unsettled at the beginning of the experiment when the initial baseline is taken. A limitation of our study is that we did not collect multiple baselines because we had one experimental condition; there was no natural place to collect additional baselines without interrupting participants.

In a research study, measures of physiological activity attempt to capture an individual's response to a stimulus or condition, or phasic activity. However gathering this phasic activity is not as easy as presenting a stimulus and recording a response, as each individual has a unique level of background physiological activity, referred to as tonic activity (Stern, Ray & Davis, 1980, pg 48). An individual's physiological activity - both tonic and phasic - can vary from day-to-day, room-to-room, etc. (Ward & Marsden, 2003). As such, there is not only an inherent difficulty in analyzing and comparing physiological responses across participants, but also within the same person (Ward & Marsden, 2003; Mandryk, Inkpen & Calvert, 2006). In addition to phasic and tonic activity, there is a third type of response known as spontaneous activity, or a physiological response to an unknown stimulus (Stern et. al, 1980, pg 48). Some examples of spontaneous activity include shifting in one's chair, coughing or sighing.   To minimize spontaneous activity, it is important to provide a controlled environment with minimal acoustic or visual distraction (Ward & Marsden, 2003).  Although we conducted the experiment in a quiet room, it was not possible to control for all spontaneous activity, since people naturally change positions, take deep breaths, etc. when performing even a sedentary task.  For this reason, we followed the example of other researchers and recorded the face and upper torso of each participant so that we could retrospectively annotate the study session and note movements or disruptions that may have caused irregularities in the data.

An issue that must be addressed at the data collection stage is informed consent.  We explained to participants where each sensor would be placed and ensured there was no discomfort with being connected to sensors before beginning the study.  Even amongst those who consented, at least one person during the interview commented that the sensors were distracting, particularly the one placed on the forehead to capture the corrugator supercilli muscle.  The respiration band must be worn against the skin (under the participant's shirt); due to the sensitivity of this sensor's placement, the researchers left the room and asked the participant to put the band on in private.  An additional problem that we encountered was not being able to collect data from all participants.  For example, three people maintained below normal SCL readings for the entire task period and did not show signs of an orienting response.  One of these individuals had small, cold fingers and, since the EDA sensor was placed on the finger, this may have been the reason.  Thus, physiological sensors raise ethical and practical considerations that may affect the ability to collect data as well as the analysis.

*Analyzing Physiological Data*

Once a researcher has collected physiological data, how does he analyze it? The first step is preparing the data for analysis. Our data was collected and filtered using a program, Microsoft Visual Studio, and analysed using MATLAB. The shear volume of data makes it challenging to examine or work with the data in some applications. Approximately one gigabyte of data is generated per person over a 20 twenty minute period; the program records a reading from each sensor 256 times per second. Even though we focused on four measures, the sensors collected twenty-eight raw, filtered or normalized variables; we focused on the raw EMG, EDA, HR and respiration variables and proceeded to transform these raw signal numbers into analyzable data. It was difficult to locate clear, standardized procedures for processing physiological data, but we were able to find some direction in the work of Mandryk and Atkins (2007) for smoothing or normalizing the data.

Once the data was smoothed or filtered, we examined the data temporally. During data collection, the researchers placed a "mark" at the beginning and end of the baseline periods, and when participants began and finished browsing the news website. This enabled us to distinguish baseline and experimental task periods. Since we were interested in examining process, it was imperative to be able to sync this timeline with the one created in our screen capture software, Morae, where we recorded on-screen activity and participants' faces. However, there was a discrepancy between the timelines due to the lag between starting the Morae recording and moving across the room to the physiological recording device to place the first marker. To compensate, we wrote a simple program based on the total trial length, as captured by Morae, and the marks collected in the physiological data set; this formula was run for each measure.

As previously mentioned, a key element for success in using physiological measures is to create a tightly controlled environment for data collection and to video record the participant during the experimental session. Therefore, the next step in analysis was identifying occurrences of spontaneous activity (e.g., overt movements, coughing) using the video data in order to identify potential problem areas in the dataset. According to the literature, it is practice to flag and remove these instances of spontaneous activity from analysis, either by having an expert manually and rigorously inspect the data or by using an algorithm to flag responses that are a certain number of quartiles beyond the median value for that particular measure (Berntson, Quigley, Jang, & Boysen, 1990). However, we faced a dilemma in removing noise because our analysis was dependent upon having a physiological timeline that corresponded to the face and on-screen recordings, since we were interested in examining the trajectory of experience. Therefore, we examined the physiological data and the videos to identify artefacts, but did not remove them.

One issue that researchers face in the analysis of physiological data is determining the range of acceptable values for a particular measure, given the variation in the measurement units (e.g., microvolts for EMG, beats per minute for HR) and the lower and upper limits. To use EMG data, which capture electrical signals sent by muscles in action, as an example, Van Boxtel (2001) found 20 Hertz (Hz) to be an ideal low-point cut off for the EMG of facial muscles, with values below this being caused by aberrant activity in neighbouring muscles or eye blinking. Fridlund and Cacioppo (1986) discuss a number of different potential upper cut-off points (250, 500 and 1000 Hz) that vary depending on the specific muscle sites monitored by EMG and other factors. They additionally recommend further data reduction beyond filtering in order to have a more manageable data set. A true baseline for EMG is zero (a muscle at rest without stimulus), but there is always some level of background noise present (Fridlund & Cacioppo, 1986). Another issue with the analysis of physiological data is that the detectable response to a stimulus has a lag time following the stimulus presentation, e.g., 1.3 to 2.5 seconds for EDA; often the onset of a response can be quite gradual, making it difficult to confirm when a response began (Stern et al., 1980, p. 206). In addition, there are conflicting opinions about the need to correct or filter some physiological data. For instance, while some researchers maintain that skin conductance data is often naturally normally distributed (Stern, Ray & Davis, 1980, p. 206), others report normalizing their data (Ward & Marsden, 2003). HR, on the other hand, tends not to be normally distributed, and therefore, the median may be more meaningful measure (Berntson et al., 1990).

Once the data has been normalized and examined for artefacts, there are a number of ways to proceed with analysis. One method is to take "snapshots" of phasic activity. Ravaja et al. (2006), for example, analyzed EMG data by calculating the mean of the four five-second chunks during participants' interactions with 32 news messages that varied in valence (positive or negative), involvement (personal relevance) and format (text or video), and then examined these averages in concert with self-reports of mood and involvement. In another study, Ward & Marsden (2003) compared the group mean difference in skin conductance between individuals who encountered poor- versus well-designed web sites. Other researchers may be less interested in averages than in the amplitude or frequency of the signals. This information can be used to assess the intensity of a stimulus and a person's reaction to it, such as calculating the level of skin conductance prior to a response from the peak value of conductance following a response (Stern et al., 1983, p. 205), or how often a response occurs over a period of time (Cacioppo & Tassinary, 1990).

We are taking this latter approach with our data. First, we are determining values one, two, or three standard deviations above and below each individual's average for each metric for the purposes of identifying what values might be abnormal for that participant/metric and represent spontaneous activity. Second, we are examining the videos qualitatively and identifying face (e.g., eye scanning, emotional expressions, etc.) and screen (e.g., mouse movements, scrolling, highlighting or hovering over text in a news story, etc) activities, and the points in time during which they occurred. This is allowing us to see instances of spontaneous activity (e.g., scratching one's eye and causing the sensor on the forehead to move) and behaviours related to the study (e.g., facial expression in response to an error message) and to create timelines with the video and physiological data that may be explored in concert to understand the interaction. Figure 1 shows HR activity for one participant. The dotted red line is experimental average for this person. The graph is annotated with an instance of spontaneous activity derived from the video data.

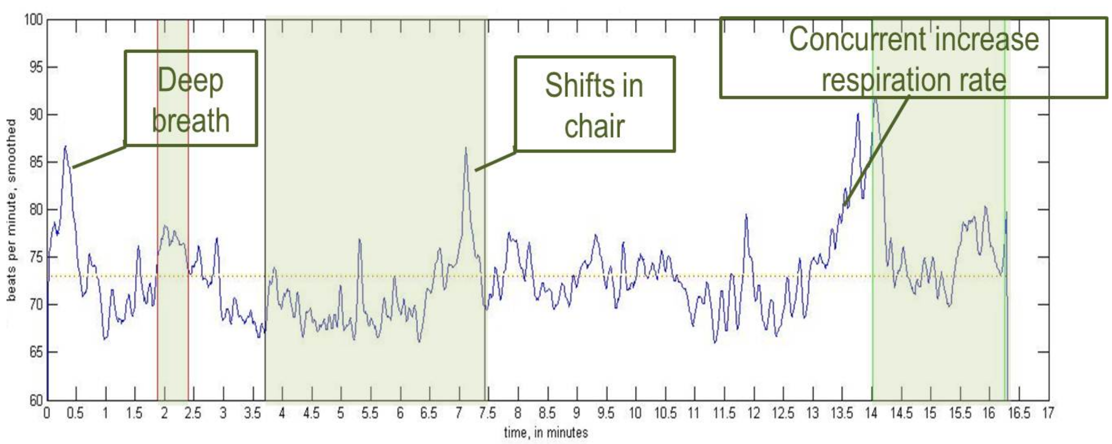


**Figure 1: Heart rate activity for Participant 6**

*Interpreting Physiological Data*

Overall, our approach is time consuming, as we carefully annotate the two video streams (face and screen activity) for each of our participants, and we examine the physiological data, keeping in mind the issues raised such as the individual nature of physiological responses, the delayed responses between the presentation of stimuli and participants' reactions, etc. As we compile our data and attempt to look at it holistically, we acknowledge the difficulties in interpreting the data, which centre around four main issues: generalizability, spontaneous activity, the relationship between physical responses and mental states, and labelling emotions based on physiological indicators.

First of all, there are individual differences in physiological norms and how people react to specific stimuli (Allanson & Fairclough, 2004), making it difficult to generalize across participants. It is for this reason that the Law of Initial Values must be taken into consideration, which states that phasic physiological activity is dependent upon baseline levels (Wilder, 1962). This Law is particularly salient in data analysis, where individuals must be examined with respect to their own baseline and phasic activity rather than across the larger sample. This is the reason that we are looking at standard deviations above and below each person's experimental average to identify each person's range of values, and why comparing averages across participants is not likely to yield meaningful results.

A second issue is spontaneous activity, which may occur when a participant is in a relaxed state or making overt movements unrelated to the experimental task (Stern et al., 1980, p. 48). This activity may be indiscernible from genuine phasic activity in the data set and may be even more pronounced than any actual change in an individual's affective state in response to the experimental stimulus (Healey, 2008). In other words, it can be difficult to discern between a person's physiological response to a stimulus and spontaneous activity, and, in the case of HII research, some of the activities we are interested in observing, such as browsing the web, may not evoke detectable physiological responses. As we continue to examine the data, we must determine whether the physiological changes we observe are significant enough to demonstrate the utility of this measure for studying interactions in which the body may be sedentary but the mind is actively engaged.

Third, there is not a one-to-one relationship between a participant's physiological reaction and mental state (Cacioppo & Tassinary, 1990; Fairclough, 2009). Despite the associations between physiological measures and mental states described earlier, conclusions must be carefully drawn. It is for this reason that multiple measures are used for validation (Lang et al., 1993; Mahlke & Minge, 2008), whether it is examining several physiological measures in concert or combining physiological, behavioural and self report measures. Some researchers have found that physiological measures and subjective measures of self-report validate each other (Mahlke & Minge, 2008; Mandryk, Inkpen & Calvert, 2006), while others have found that these do not always match up (Wilson, 2001). As a result, some researchers have declared that psychophysiological measurements are not as robust as we would like them to be (Ward & Marsden, 2003). As we move deeper into our analysis, we will be comparing not only the physiological and behavioural data (from the videos), but also questionnaire and interview data collected after the browsing session. Will we see correspondence between physiological activity, one-screen behaviours, and self-reports, and what should the strength of this relationship be?

Lastly, interpreting the psychological state of an individual from physiological data is controversial. Although researchers have linked physiological changes to mental states, others have failed to replicate these findings (Dimburg, 1990; Mahlke and Minge, 2008) or caution against using them as indicative of felt emotions (Lopatovska, 2011). For instance, consider the similarities in physical response of going on a first date to accidentally wandering into oncoming traffic – both can cause the heart to race. Fear, happiness and anger have all been reported to be psychological constructs underlying increased heart rate (Levenson, 1992), so what is the "true" emotion experienced by the individual? Thus, we must be careful in interpreting peak responses in the physiological data as positive or negative, and use other data, such as participants' self-reported overall expression of engagement, contextual cues such as what was taking place on the interface (e.g., error message, reading an emotionally charged news story), to understand the trajectory of the interaction.

In summary, in order to use physiological measures, one must appreciate the complexities in collecting, analysing and interpreting the data. This involves recognizing the variability amongst research participants, the differences inherent in each measure (e.g., unit of measurement, range of values), and the need to collect multiple measures to address issues of validity.

**Conclusion**

Despite the various challenges, the use of physiological measures is growing in popularity, with the advantage that it is perceived to be an objective way to evaluate emotion and to capture a steady stream of data without disrupting the flow of an experience (Rowe, Sibert & Irwin, 1998). But are these benefits enough to make HII researchers incorporate physiological measures into their research designs? Furthermore, can physiological measures address the challenges of capturing the process of interaction and human emotion?

Drawing upon our own experience of incorporating physiological measures into a research study that involved ten participants browsing and reading a news website, we are both sceptical and optimistic about their potential utility for HII. Through our attempts to collect, analyze and interpret physiological data, we concur that these measure are complex, require an understanding of physiology, some understanding of electrical signals, and a lack of fear in dealing with large data sets. In addition, we acknowledge that an inability to do so may "lead to serious errors of inference" (Cacioppo & Tassinary, 1990, p. 16).

Although physiological measures may seem more objective than, say, self-report measures, our discussion of issues in the interpretation of this data demonstrates that there are challenges in gleaning conclusions from this type of data as well. We maintain that multiple physiological measures or mixed-methods approaches are needed, with attention to issues of reliability and validity. However, we are still learning how to triangulate physiological, behavioural, and self-report measures with respect to being able to say something meaningful about the interaction process and nature of experience or emotion within this process. This area of inquiry is novel for information scientists (Kelly, 2009) and requires concerted effort to "define and develop the norms" (Lopatovka, 2011, p. 229) that will allow us to robustly address current measurement challenges in HII. As a field, we need to evaluate whether the assumed value added though these measures justifies the level of dedication necessary to employ them, and work towards establishing standards and best practices to promote the effective use of physiological measures in our research designs.

For our own purposes, we will continue to analyze our physiological data in concert with our self-report and behavioural data. We will look for patterns in the data amongst non-spontaneous physiological activity, what the user is reading or interacting with on the screen, and their own assessments of the experience. Our intention to establish relationships amongst these different types of data will help us understand what engages participants and how to identify "shifts" in the interaction where emotion or interactivity appear to change. Focusing on these two areas may give us a sense of how engagement changes over the course of an interaction and what aspects of the information object (e.g., content) and technology (e.g., design features on the interface) influence this engagement.

At this point, we are inching forward, acknowledging the challenges with the realization that there is no clear map to guide us in collecting, analysing and interpreting this data; we are piecing things together from the research literature and knowledgeable HCI colleagues. But we forge ahead with the hope that, by incorporating physiological measures into our methodology, we may be better able to comprehend the process of interaction and develop more dynamic measures of interaction that will allow us to explore timely issues in HII holistically and robustly.

**Acknowledgements**

This research is supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and the Network Centres of Excellence Graphics Animation and New Media (NCE GRAND) grants to O'Brien.

## References


Ahsan, M.R., Ibrahimy, M.I., & Khalifa, O.O. (2009). EMG signal classification for human computer interaction. *European Journal of Scientific Research, 33*(3), 480-501.

Allanson, J., & Fairclough, S. H. (2004). A research agenda for physiological computing. *Interacting with Computers, 16*(5), 857-878.

Berntson, G.G., Quiqley, K.S. Jang, J.F., & Boysen, S.T. (1990). An approach to artefact identification: Application to heart period data. *Psychophysiology, 27*, 586-598.

Berntson, G. G., Bigger, T.J., Eckberg, D. L., et al. (1997). Heart rate variability: Origins, methods, and interpretive caveats. *Psychophysiology, 3*, 623–648.

Cacioppo, J.T., Petty, R.E., Losch., M.E., Kim, H.S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50*, 260–268.

Cacioppo, J.T. & Tassinary, L.G. (1990). Inferring Psychological Significance From Physiological Signals. *American Psychologist, 45*(1), 16-28.

Carroll, D., Turner, J. R., & Hellawell, J. C. (1986). Heart rate and oxygen consumption during active psychological challenge: The effects of level of difficulty. *Psychophysiology, 23*(2), 174-181.

Chang, J-S., McGrenere, J. & McLean, K. (2011). Audio stream bookmarking with a wristband controller: Exploring the role of explicit commands in an implicit control loop. (UBC CS TR-2011-06). Vancouver, British Columbia: University of British Columbia ICICS/CS, Retrieved from, http://www.cs.ubc.ca/cgi-bin/tr/2011/TR-2011-06.

Dimberg, U. (1990). Facial electromyography and emotional reactions. *Psychophysiology, 27*(5), 481-494.

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers, 21*(1-2), 133-145.

Frith, C. D., & Allen, H. A. (1983). The skin conductance orienting response as an index of attention. *Biological Psychology, 17*(1), 27-39.

Fridlund, A. J. & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology, 23*, 567–589.

Healey, J.A. (2008). Sensing Affective Experience. In J.H. Westerink et al. (Eds.), *Probing Experience: From Assessment of User Emotions and Behaviour to Development of Products* (p.p. 91 – 100). The Netherlands, Springer.

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval, 3*(1-2), 1-224.

Lang, P.J., Greenwald, M.K., Bradley, M.M. & Hamm, A.O. (1993). Looking at pictures: Affective, facial, viseral and behavioral reactions. *Psychophysiology, 30*, 261– 273.

Levenson, R.W. (1992). Autonomic nervous system differences among emotions, *American Psychological Society, 3*(1), 23– 27.

Lopatovska, I. (2011). Researching emotion: Challenges and solutions. In *Proceedings of iConference, February 8-11, 2011, Seattle, WA*, 225-229, ACM.

Lopatovska, I. & Arapakis, I. (in press). Theories, methods, and current research on emotions in library and information science, information retrieval and human–computer interaction. *Information Processing and Management*, DOI: 10.1016/j.ipm.2010.09.001.

Mahlke, S. & Minge, M. (2008). Consideration of multiple components of emotions in human-technology interaction. In C. Peter & R. Beale (Eds.), *Affect and Emotion in Human-Computer Interaction* (pp. 51-62). Berlin: Springer-Verlag.

Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies, 65*(4), 329-347.

Mandryk, R.L., Inkpen, K.M. & Calvert T.W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour and Information Technology, 25*(2), 141-158.

Marchionini, G. (2008). Human-information interaction research and development. *Library & Information Science Research, 30*(3), 165-174.

O'Brien, H.L. & Toms, E.G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science & Technology, 59*(6), 938-955.

Porges, S.W. & Byrne, E.A. (1992). Research methods for measurement of heart rate and respiration. *Biological Psychology, 34*, 93-130.

Ravaja, N. (2004). Contributions of psychophysiology to media research: review and recommendations. *Media Psychology, 6*(2), 193-235.

Ravaja, N., Saari, T., Kallinen, K. & Laarni, J. (2006). The Role of Mood in the Processing of Media Messages From a Small Screen: Effects on Subjective and Physiological Responses. *Media Psychology, 8*(3), 239-265.

Rowe, D.W., Sibert, J. & Irwin, D. (1998). Heart Rate Variability: Indicator of User State as an Aid to Human-Computer Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 480-487). New York: ACM.

Saracevic, T. (1997) The stratified model of information retrieval interaction: Extension and application. In G. S. Schwartz (Ed.), In *Proceedings of the American Society for Information Science* (pp. 313-327). Medford, NJ: Information Today.

Stern, R.M., Ray, W.J. & & Davis, C.M. (1980). *Psychophysiological Recording*. New York: Oxford University Press.

Ward, R.D., & Marsden,P.H. (2003). Physiological responses to different WEB page designs. *International Journal of Human-Computer Studies, 59*, 199–212.

Wilder, J. (1962). Basimetric approach (law of initial value) to biological rhythms. *Annals of the New York Academy of Sciences, 98*, 1211–1220.

Wilson, G.M. (2001). Psychophysiological indicators of the impact of media quality on users. In *Proceedings of CHI 2001 Doctoral Consortium* (pp. 95-95). New York: ACM.