PHYSIOLOGICAL MEASURES, EYE TRACKING AND TASK ANALYSIS TO TRACK USER REACTIONS IN UGC

Dufresne, Aude
Department of Communication, University of Montreal
Aude.Dufresne@umontreal.ca
Courtemanche, François
Department d'Informatique et Recherche Opérationnelle, University of Montreal
francois.courtemanche@umontreal.ca

Prom Tep, Sandrine
Hautes Études Commerciales Montréal
sandrine.prom-tep@hec.ca
Senechal, Sylvain
Hautes Études Commerciales Montréal
sylvain.senecal@hec.ca

ABSTRACT
In order to improve human-computer interaction, eye-tracking, physiological measures and task recognition are used to assess the content and quality of interaction in an e-commerce application with User Generated Content. Results are analyzed in conjunction with users' verbalization and their subjective assessment of attitudes toward the product and their wish to contribute.
Authors Keywords: *physiological measures, eye tracking, task analysis, e-commerce, User Generated Content*

INTRODUCTION
Understanding affective reactions is essential to improve human-computer interaction and the design of systems. This assessment of user motivational state can be used to evaluate interface design, to define adaptation in systems or to study how interfaces influence the use of systems. We will present our research on the integration of different measures and AI techniques to analyze interactions in the context of Web 2.0 applications where e-commerce integrates user generated content.

CONTEXT
The research focuses on using physiological and eye-tracking measures to assess online consumer behaviors in the context of Web 2.0 sites. Described as a second-stage evolution of the Web, this emerging context is characterized by increasing consumer participation which rely on bi-directional communication tools and technologies, enabling "customers to share their opinions and experiences on goods and services with a multitude of other consumers" [32]. Web 2.0 participative technologies do not equate participation though, and site managers struggle to find new Web 2.0 business models that would lead to profitability [15].
Any form of web site contribution from consumers has shown to be primarily motivated by the desire for social interaction [32] and it is important to assess how the interfaces support the social and confidence dimensions of the user toward the content but also to encourage him to contribute.

Eye-tracking and measures of affective reactions
Eye-tracking measures are being used regularly to analyze what are the points of fixation, the duration of fixations and the scan paths, in order to follow what are the zones of interest and at the opposite, which information or navigation elements may have been missed by users. These measures are used along with usability methodologies to evaluate applications before they are made accessible. In most cases a few subjects are used in a design context, but few precise methodologies or studies are made to define principles that can be used to design or evaluate applications using those measures.
Most eye-tracking applications in human-computer interactions can be separated in mainly two categories: 1) describing the user's state and 2) describing the interactions' state.
The former consist of analyzing data brought by an eye-tracker in order to infer different types of information on the user psychological state. Pupil diameter variations and blink rate can be related to the user cognitive load [4, 29]. Information on the user current amount of mental effort can be used for many different purposes in the context of HCI. For example, Bailey & al. [3] use eye-tracking to identify moments where users can be interrupted at lower cost. These moments usually occur at task boundaries, which are identifiable by a decrease in pupil size (cognitive load). Conati & Merten [11] used eye-tracking for online modeling of students meta-cognitive behaviors (self-explanation and effective exploration) during interaction with a learning environment. Eye-movement and pupil dilation are also used in real context simulations in order to evaluate awareness and vigilance [16]. Pupil size variation can also be used to infer information on the subject affective state. Partala & al. [24], who concluded that pupil enlarge during negative and positive stimuli.
The second category of application of eye-tracking to HCI – describing the interaction's state – has been paid less attention by researchers. Salvucci & Anderson [28] developed different techniques to analyses eye movement protocols in order to infer the user' current position in a cognitive task model. The scan paths have been used to describe more frequent exploration sequences or even to recognize users' activities.
Baccino & Manuta [1] have developed interesting measures to recognize the user's cognitive processes, using eye-

tracking measures, depending on the saccade angles and the distance between fixations.

Physiological measures of affective reactions

In our study we used in addition to eye-tracking data two physiological measures: Galvanic Skin Response (GSR) and blood volume pressure (BVP) to identify user's emotional state. The GSR measures can inform about stress and relaxation [25] and is indicative of emotionally significant HCI events and situations [33]. The heart rate can reflect emotional activity and can help differentiate between negative and positive emotions. It also reflects the amount of mental effort [6, 27], since heart rate variability correlate with the visual display levels of complexity.

The need for Task analysis to integrate context in affect recognition

From an evolutional point of view, primary emotions such as joy, fear or sadness are hardcoded responses to inner or outer environmental changes. More complex emotions such as pride or shame result from the conscious association between primary emotions and the current situation in which they occur [23]. Therefore, taking into account the contextual interaction factors in which physiological and eye-tracking measures are collected is a key step in attaining efficient affect recognition.
Physiological measures, such as skin conductance, are correlated to arousal but cannot give information on emotional valence (positive of negative aspect of emotions). Contextual information can help fill this lack of information. For example, negative affective reactions like stress are associated to understanding difficulties, uncertainty in choices, unexpected or dead end situations. Positive reactions like surprise or interest may be linked to task success or to a positive orientation reaction toward a stimulus. All these contextual factors can be inferred using appropriate and effective task tracking techniques in order to support affect recognition. In addition to supporting recognition, contextual factors can also be used to identify the cause of affects. For example, the precise icons or interface elements, which are inefficient and cause, in the context of a given task, impatience and task abandon. This information is highly important for HCI diagnostic and quality evaluation.

Within human-computer interactions, task recognition usually aims to analyze the interaction flow coming from the user, using a filter corresponding to the model of the task being executed [7]. The goal is to identify which *part of the task* the user was executing when he generated a particular *sequence of interactions*. In our work, we developed a task recognition approach based on machine learning algorithms – Layered Hidden Markov Models [22], where data (*observed interactions*) are associated to targets (the *task model components*) in a hierarchical structure.

The ever growing complexity of today's interfaces makes difficult the use of task recognition techniques based only on standard user's interactions (mouse clicks and keyboard). Our approach uses eye tracking to overcome this problem. In most interfaces, gaze position can be separated and viewed in terms of areas of interest (AOI). Therefore, the *data* used in our machine learning technique are sequences of interactions, including mouse clicks, keyboard events and gaze position events (AOI). The *targets* of our machine learning technique are each sub-tasks defined in hierarchical task model (Card, Moran and Newell, 1983). A task model links the way a user mentally structure a task (what to) to the different possible interactions to achieve the task (how to) [3]. It is represented as a hierarchical structure of goals, sub goals.

The task recognition system is developed in two phases. First, the *training stage*, the Layered Hidden Markov Models (LHMM) task recognition algorithm is trained using data collected while subjects execute the proposed task. In order to train the system, the tasks given for training are more specific or the recordings may be tagged to categorize what the subject is doing at a given time in the interaction. During the second phase, the *experimental phase*, djfferent subjects are given the same tasks but they are left free to find a way to do them. The trained LHMM model is then used to analyze their interactions with the system.

The Assesment of Affect in e-commerce interaction

With e-commerce and the multiplication of retail Web sites, taking emotions into consideration becomes crucial as it has been clearly established that consumers are not always rational in their choices, nor utilitarian in their motivations [10]. Whereas there is much more than cognitive information processing going on while consumers are online [12] measuring behavior for interactive systems evaluation has almost exclusively limited its focus to cognitive activities. Models combining emotions with human information processing variables are scarce and when available, they rely on reported data [9]. Research has shown though, that emotions are experienced both at the conscious and preconscious levels [3].
So far, studies on consumers' online participative behavior have mainly focused on the "why" and "how" of participation, identifying:
1) *drivers of contribution* such as motivations to create and maintain a blog [5], or to articulate themselves on consumer-opinion platforms [2] as well as
2) factors encouraging *participation* in virtual communities, such as offline interactions and perceived usefulness [17] or
3) the modes of *inter-personal influence* which can take place in these online communities for the adoption and use of products and services usefulness [5].
Reactions to interface design are very often measured in passive situations [26], but in the context of e-commerce applications, it is important to develop researches when there is more interaction. For example in a web site offering online transactions, it is necessary to follow the specific steps according to a task model, where actions and relevant areas of visualization are defined. It is therefore necessary to follow the user activities and to study changes in his physiological and eye-tracking measures accordingly.

EXPERIMENTATION

We were interested to assess how users are influenced by "user generated content" in the context of e-commerce. We experimented using the Amazon web site, where books have comments and assessment (number of stars) from other users. Using a classical experimental design, we tested the differences in consumers' attitude toward contribution, online intention to contribute and actual contribution with and without social functionalities, within Amazon Web site.

To assess the usability and the UGC social perspective we also used think-aloud protocols [9] and qualitative questionnaires. We complemented both sources of data with eye-tracking and physiological measures, so as to obtain a richer perspective on the phenomenon.

User testing was conducted in the Bell Solutions Web Laboratory. We used the Internet Explorer® Web browser with a TOBII (X120 eye-tracker). We recorded comments from the user and the output of the control monitor, with an overlay of the gaze movement. Physiological measures (GSR, BVP) were monitored using a Biograph system. Each system was running on a different computer, for which the clocks had been synchronized. The Tobii was calibrated and then started with the Biograph both having a measure at each 1/60 sec. Data were synchronized using the time stamp in each data file. Though raw data were collected, they were reduced to means for each fixation in a zone. Screen interactions and navigation activity are also recorded in order to integrate the task analysis dimensions. The task model was described as the structure of possible actions in relation to exploring information on a series of books in the web site. For example they could read descriptions, look at assessment and contributions from other users, add an evaluation, etc. In the training phase, five users were asked to accomplish the tasks. For each user data were collected from physiological and eye-tracking measures (pupil size, AOI), along with navigation information. These observations were used to train the recognition of the structure of actions, so it was possible to recognize what a user was doing, when given more general instructions.

In phase two, the system was tested with 27 subjects. The users were given general instructions to explore the books: descriptions, comments, and evaluations. These instructions were given at the beginning and subjects were left free to explore as long as they wished. After the exploration, they were asked to comment on the interaction and the components of the interface. They were also ask questions about social and credibility dimensions which may have influenced their appreciation of the user generated content. For the experiment, we asked subjects to explore a set of books of general interest, who were saved locally to insure that each subject would have access to the same content. For those pages, different areas of interest (AOI) were defined corresponding to the structure of presentation: description, detailed description, keywords, price, statistics of evaluation (number of stars), comments positive or negative, recent comments, ordering zones, etc.

The data collected gives for each fixation: what was the page and the zone being explored, physiological measures (GSR, BVP) and Pupil size (mean for both eyes), the task structure was also recognized (goal, subgoal) . Results were analyzed in order to see how patterns of interactions could be extracted and how they could be related to emotional reactions. GSR and BVP measures were normalized using the baseline for each subject, using data collected at the beginning of the session. The baseline refers to the average measures of GSR during a rest period before a session. We used Lisetti &al.(2004) equation to correct observations :

$GSR_{normalized}=(GSR-GSR_{relaxation})/GSR_{relaxation}$

Pupil size and BVP were accordingly corrected for each subject. Even though they don't vary as much between individuals, we wanted to insure that observed differences were only due to the context of navigation and not to individual differences or individual mood in general.

ANALYSIS AND PRELIMINARY RESULTS

Eye fixations on areas of interest defined around the social functionalities elements of interface, as well as GSR and BVP measures are analyzed along with reported measures of contribution intent and buying intent. As far as eye-tracking data is concerned: the more eye fixations on a social functionality element of interface design related to one of the book's content, the more useful is the social functionality for decision-making. Also the physiological reactions are compared to eye-tracking measures and impact on attitudes toward the book.

The hypothesis is that the longer the eye fixations on a social functionality element of the interface related to one of the book's content and the stronger the reaction to it, the stronger should the participant's intent to contribute online and to buy the book.

Pretests have shown that comments were attracting a lot of fixation time, so user were spending time trying to assess the comments. However there was more GSR reactions to description than to comments.

When looking at the types of activity, it appears clear that reading the table of statistics of evaluation was an important part of the interaction. For judging the book, reading the description appears important, and looking at the keywords also. This may look surprising considering the keyword zone is very small. But keywords content is dense and it takes long for the user to read and reflect on implications. We also compared how the task analysis could be compared to results on AOI statistics. In general both measures agree, but in some cases, especially when the zones are small task recognition appears to give better results (for example: reading keywords vs reading description, reading comments vs looking at the statistics of evaluation (which is a small zone at the beginning of description). In the experimentation with more observations for training the task recognition algorithm and also with more subjects, the agreement will certainly be improved. But still first results suggest that the structures of activity might be more precise than AOI, since in some cases AOI associated might be small and eye-tracking only might be imprecise. Another interesting result is the analysis of the scan paths and how they are linked to the types of zone.

REFERENCES

1. Baccino, T., & Manunta, Y. (2005). *Eye-Fixation-Related Potentials: Insight into Parafoveal Processing.* Paper presented at the Journal of Psychophysiology.
2. Backs, R.W., Boucsein, W. (Eds.), 2000. Engineering Psychophysiology. Issues and Applications. Erlbaum, Mahwah, NJ.
3. Bailey, B. P., Adamczyk, P. D., Chang, T. Y., & Chilson, N. A. (2006). A framework for specifying and

monitoring user tasks. Computers in Human Behavior, 22(4), 709-732.

4. Beatty, J., Lucero-Wagoner, B. (2000). Pupillary System. Chapter 6 in CACIOPPO, J.T., TASSINARY, L.G., AND BERNTSON, G. Handbook of Psychophysiology. Cambridge University Press.
5. Bonie Nardi, Diane J. Schiano, Michelle Gumbrecht, Luke Swartz (2004), "Why we blog?", *Communications of the ACM*, vol 47, no12, pp 41-46.
6. Boucsein, W. 2000. The Use of Psychophysiology for Evaluating Stress-Strain Processes in Human-Computer Interaction. In Backs and Boucsein 2000,289–309.
7. Carberry, S. (2001). Techniques for Plan Recognition. User Modeling and User-Adapted Interaction, 11(1-2), 31-48.
8. Card, S., Moran, T., & Newell, A. (1983). The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale: NJ.
9. Chamberlain, L. and Broderick, A.J. (2007). *Qualitative Market Research: An International Journal,* **10 (2)**: 199-216.
10. Childers, T. L., Carr, C.L., Peckc, J., and Carson, S. (2001). Hedonic and utilitarian motivations for online retail shopping behavior. Journal of Retailing, 77(4): 511-535.
11. Conati C. and Merten C. (2007). Eye-Tracking for User Modeling in Exploratory Learning Environments: an Empirical Evaluation. Knowledge Based Systems, Volume 20, Issue 6, Elsevier Elsevier Science Publishers B. V. Amsterdam, The Netherlands
12. Ethier, J., Hadaya, P., Talbot, J., Cadieux, J. (2006). B2C web site quality and emotions during online episodes: An empirical study, *Information & Management,* **43**: 627–639.
13. Hirschman, E. C., & Holbrook, M. B. (1982). Hedonic Consumption: Emerging Concepts, Methods and Propositions. *Journal of Marketing,* **46**: 92-101.
14. Holbrook, M.B., Batra, R. (1987). Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising, *Journal of Consumer Research,* **14(3):** 404-419.
15. UGC: More Popular than Profitable*, eMarketer*, 2009
16. Ji, Q. and X. Yang, Real-time eye, gaze, and face pose tracking for monitoring driver vigilance, *Real-Time Imaging 3,* (2002) (8), pp. 1077–2014.
17. Joon Koh,Young-Gul Kim, Brian Butler, Gee-Woo Bock (2007). Encouraging participation in virtual communities, *Communications of the ACM*, vol 50, no2, pp 68-73.
18. Kirwan, B., & Ainsworth, L. K. (1992). A Guide to Task Analysis. Taylor & Francis, Ltd.
19. Lee S., McQuiggan S. W., Lester J. C. (2007), Inducing User Affect Recognition Models for Task-Oriented Environments. User Modeling 380-384.
20. Lisetti C. L., Nasoz F. (2004), Using Non-invasive Wearable Computers to Recognize Human Emotions from Physiological Signals, EURASIP Journal on Applied Signal Processing - Special Issue on Multimedia Human-Computer Interface, vol. 2004, n 11.
21. Nasoz F, Lisetti CL, Alvarez K, Finkelstein N (2003). Emotion Recognition from Physiological Signals for User Modeling of Affect. International Journal of Cognition, Technology and Work - Special Issue on Presence.
22. Oliver, N., Garg, A., & Horvitz, E. (2004). Layered representations for learning and inferring office activity from multiple sensory channels. Computer Vision and Image Understanding, 96(2), 163-180.
23. Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge [England] ; New York: Cambridge University Press.
24. Partala T., Surakka V. (2003). Pupil size variation as an indication of affective processing.International journal of human-computer studies. vol. 59, no1-2, pp. 185-198.
25. Picard R. W. (1997) Affective computing. MIT Press.
26. Prom Tep, S., Dufresne, A., Saulnier, J., & Archambault, M. (2008). Web site quality evaluation combining eyetracking and physiologicial measures to self-reported emotions: an exploratory research. In Measuring Behavior (pp. 224-225). Maestricht, Netherland.
27. Rowe D.W, Sibert S., Irwin D. (1998). Heart Rate Variability: Indicator of User State as an Aid to Human-Computer Interaction, Proceedings of ACM CHI 98.
28. Salvucci, D. D. and J. R. Anderson (2001). Automated eye-movement protocol analysis. *Hum.-Comput. Interact.*, 16(1): 39-86.
29. Siegle, G. J., N. Ichikawa, et al. (2008). "Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses." *Psychophysiology* 45: 679-687.
30. Subramani and Rajagopalan (2003), "Knowledge-Sharing and Influence in Online Social Networks via Viral Marketing" *Communications of the ACM*, vol 46, no12, pp 300-307.
31. Tao L., Akinobu M., Shigeo M. (2008). Using subjective and physiological measures to evaluate audience-participating movie experience. ACM Press
32. Thorsten Hennig-Thurau,. Kevin P. Gwinner; Gianfranco Walsh; Dwayne D. Gremler (2004), "eWOM via consumer-opinion platforms: what motivates consumers to articulate themselves on the Internet?", *Journal of Interactive Marketing*, vol 18, no1, p.38.
33. Ward, R. D., P. H. Marsden, B. Cahill, and C. Johnson. 2001. Using skin conductivity to detect emotionally significant events in human-computer interaction. In Proceedings of BCS IHM-HCI 2001, 10-14 September, 2001, Lille, France, 2: 20-25.

ACKNOWLEDGEMENTS
This research was made possible by a NSERC grant for reserach and development in collaboration with Bell Canada.