The FaceReader: Measuring instant fun of use

Bieke Zaman
Centre for Usability Research, Katholieke Universiteit Leuven
E. Van Evenstraat 2A, 3000 Leuven (BE)
Bieke.Zaman@soc.kuleuven.be

Tara Shrimpton-Smith
Centre for Usability Research, Katholieke Universiteit Leuven
E. Van Evenstraat 2A, 3000 Leuven (BE)
Tara.Shrimptonsmith@soc.kuleuven.be

ABSTRACT
Recently, more and more attention has been paid to emotions in the domain of Human-Computer Interaction. When evaluating a product, one can no longer ignore the emotions a product induces. This paper examines the value of a new instrument to measure emotions: the FaceReader. We will assess the extent to which the FaceReader is useful when conducting usability evaluations. To do this, we will compare the data gained from the FaceReader with two other sources: user questionnaires and researcher's loggings. Preliminary analysis shows that the FaceReader is an effective tool to measure instant emotions and fun of use. However, a combination of the FaceReader with another observation method (e.g. researcher's loggings) is necessary. As regards the user questionnaire, our results indicate that it is rather a reflection of the content of the application or the outcome of a task, than a correct self-reflection of how the user felt when accomplishing the task.

Author Keywords
Emotions, usability, FaceReader, instant fun of use

ACM Classification Keywords
H.5.2. Evaluation/methodology

HCI AND EMOTIONS

Importance of emotions in HCI
The human face is a feature of the human body which deserves special attention when observing users' reactions. Several facets of verbal and non-verbal communication are often defined by the characteristics of the human face. A lot of information is retrieved from facial expressions which can be very diverse, from gestures such as nods and winks and expression of emotions [10].

The sense of the satisfying or dissatisfying quality of our daily lives is determined by the emotions that go hand in hand with the situation [5]. With regards to usability and likeability research, emotions (shown from facial expressions) are a valuable source of information for the evaluation of products. A study has shown that a user-friendly device brings forth more positive emotions than applications filled with usability errors. A user-friendly device will bring forward emotions like satisfaction, enjoyment and excitement, whereas a difficult to use device arouses feelings of frustration from the user [11]. The "fun of use" is a factor that cannot be overlooked, clearly an affective area, and which can be measured through emotions [4]. Because positive or negative emotions induced by a specific product influence the enjoyment of purchasing, possessing and usage, measuring emotions is becoming more and more important every day. In some cases the emotional experience may even be the crucial factor concerning decision-making. For example, if the user likes working with a device, the intention to buy or use that device more often will increase [8].

There are a lot of different emotions a person can experience when working with a specific product, such as angry, happy, sad, relief, amusement [8]. It is not possible to observe and evaluate all the possible emotions a user can experience. Therefore a selection of emotions most relevant to the test situation will be made. It is important that the evaluation of emotions experienced by the test user is measured moment-to-moment. This is because it is difficult for the test user to reproduce exactly the same reactions to events that happened maybe only minutes before and which have been succeeded by other actions [7].

A number of existing instruments to measure emotions
In the field of psychology and sociology a lot of effort has been put into finding a way to measure emotions. In recent years consumer and marketing researchers, as well as computer scientists, have shown a lot of interest in developing their own instruments to assess emotions. A distinction is made between non-verbal (objective) and verbal (subjective) instruments [8].

Non-verbal instruments either measure the expressive or the physiological element of the experienced emotion. The non-verbal instruments include a number of advantages. They are language-independent, they also do not bother the test users during the evaluation and are claimed to be less biased. The expressive component of emotion can be

NordiCHI 2006: Changing Roles, 14-18 October 2006, Oslo, Norway

evaluated either by measuring facial or vocal expressions. When using facial expression instruments, distinct emotions are correlated with specific features. A few examples are the Facial Action Coding Scheme, or the use of electromyogram (EMG) sensor devices. The physiological element is determined by the change in activity in the autonomic nervous system which accompanies emotions. Examples of areas that can be measured are blood pressure, skin conductivity, papillary response, respiration rate and heart responses [8]. For example, skin response can be measured by the Galvanic skin response instrument, which measures both emotional responses as well as cognitive activity. Blood pressure and heart rate can be monitored through cardiovascular measurements or electrocardiograms, where heart rate gives a good indication of the emotional experience. Finally electromyography can be used on the face to plot the distinction between positive and negative emotions [12].

Verbal instruments assess the subjective feeling component of emotions, evaluated through self-report by means of rating scales or verbal protocols. One example is the Self-Assessment Manikin [8]. A major disadvantage is the language barrier between different cultures, whereas an advantage is that rating scales can be used to evaluate all emotions, as well as mixed emotions.

One of the measurement instruments that was developed to assess emotional response to products is the Product Emotion Measurement instrument (PrEmo). A combination of the advantages of non-verbal and verbal instruments; no verbalisation is needed, but it does evaluate separate emotions [8].

In our project, we will test a new non-verbal instrument: the FaceReader (see Figure 1). The FaceReader, recently developed by VicarVision and Noldus Information Technology bv., recognizes facial expressions by distinguishing six basic emotions (plus neutral) with an accuracy of 89% [3]. More particularly, the FaceReader classifies happy, angry, sad, surprised, scared, disgusted and neutral. The system is based on Ekman and Friesen's theory of the Facial Action Coding System (FACS) that states that basic emotions correspond with facial models [6]. A study has shown that different cultures respond with comparable facial expressions to specific experiences [5]. We will estimate the value of the FaceReader when consulted in a usability test. It is important to take into account that the best method to conduct a usability test, which is also focused on the experienced emotions of the test user, is in a controlled test environment. The facial expression of emotions made by the test user are a result of the reaction to the device or software that is tested. In an uncontrolled test environment it is uncertain which emotion expressed is linked to using the device tested [1].

METHODOLOGY

The primary focus of our experiment was to estimate the value of the FaceReader. We explored whether it would be interesting to work with the FaceReader when conducting usability research.

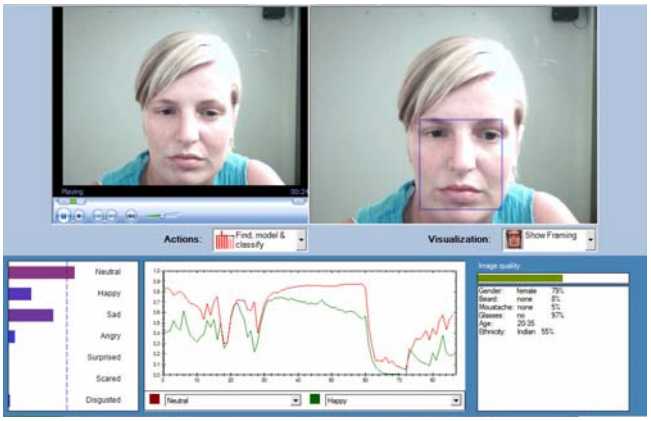

Figure 1. The FaceReader measures 6 emotions.

Participants

Seven male and five female test users participated in the usability test session. Their age varied between 20 and 60 years old. We tried to include test users representing different levels of computer/Internet knowledge and skills. Nevertheless, there was a slight overrepresentation of more experienced test users.

Test material

Each test session comprised 9 tasks to be performed on a personal computer. The test users were given the same tasks and the same amount of time to complete them. After each task, the test users were asked to fill out a post-task questionnaire. One part of the questionnaire concerned the test user's emotions during the task. The questionnaire collected the self- reported 'subjective' emotions once the task had been finished. In addition to the emotional part, the questionnaire also dealt with the usability of the application tested during the task. The usability questions were based on the QUIS questionnaire, which is an appropriate instrument to measure user interaction satisfaction [2].

Test setting

The test users were invited to undergo the usability test in a living room that is fully equipped to serve as an observation room of a usability lab. The living room was separated by a one-way mirror from the control room. The researcher logged from the control room and recorded the images on video without disturbing the test session. The collection, analysis and presentation of the observational data was structured and facilitated by using 'The Observer 6.1', a Noldus software programme.

RESULTS

A preliminary analysis was undertaken. We will discuss some of the highlights of our experiment. One has to take into account that it concerns work in progress. The goal of our experiment was to estimate the value of the FaceReader. To do this, we analyzed the data of the

FaceReader and compared it to the loggings of the researcher and the data gained from questionnaires. Table 1 presents the results of this comparison.

Measuring usability

The International Organization for Standardization (ISO) defines usability by three aspects: effectiveness, efficiency and satisfaction [9]. When evaluating the usability of a system, the input of the expert usability researcher is crucial to rate the effectiveness and efficiency. A researcher will observe usability problems which users cannot describe. Users are indeed not used to focusing on the small steps involved in every task. However, the problems observed cannot be interpreted without understanding what the user experiences, feels, thinks or expects. Without entering the world of their inner feelings, one cannot assess the satisfaction of a product. That is why test users were asked to think out loud while performing tasks [13]. Another way of investigating what the user thinks and feels concerning the product is by user satisfaction questionnaires. Our post-task questionnaire was based on QUIS to measure the three usability aspects. In contrast to the researcher's loggings, a questionnaire allows easy comparison over different applications and test users. Another advantage is that the questionnaire exclusively provides feedback from the user's point of view. The FaceReader only measures one of the three ISO usability aspects: satisfaction. User-friendly devices indeed elicit satisfaction, enjoyment and excitement whereas a difficult to use device causes feelings of frustration [11]. The data of the FaceReader cannot be interpreted without context. A researcher has to verify which factor caused a certain emotion (e.g. content, environment, usability).

Researcher: loggings	FaceReader	User: questionnaire
Measuring: usability, emotions, fun of use	Measuring: satisfaction, emotions, fun of use	Measuring: usability, emotions
Expert view	'Objective' instrument	'subjective' end user's view
Instant data (distraction: danger of missing gaps)	Instant data (complete)	Self-reflection after the events happened (recall: danger of missing gaps)
Observed behaviour/ emotions	Predefined/ ascertained emotions	Self-evaluated (biased) emotions
Subtle differences in emotions	6 basic emotions + neutral	Recalled emotions
Interpretation of context	No context	Self-interpretation

Table 1. Comparing data from researcher's loggings, FaceReader & the test user's questionnaire

Measuring emotions

Although a researcher can observe emotional cues on a monitor in the control room, he/she often misses some crucial information. The researcher has to combine different tasks: logging, observing, (if necessary) probing to think out loud, adjusting video and audio systems, etc. Hence, it is very normal that he/she cannot observe every single emotional behaviour. By contrast, the FaceReader registers every small change in emotion. The FaceReader is limited though in the number of emotions it discerns. Only 6 predefined emotions are registered whereas a researcher can distinguish a larger number and more subtle distinctions in emotions. Moreover, a researcher can also interpret essential verbal cues such as comments, voice intonation, stop words and sighs. As regards the user questionnaire, emotional responses were measured using a five-point scale for each of the six basic emotions ranging from 'strongly agree' to 'strongly disagree'. The advantage of the questionnaire is the comparability and the feedback from the user's point of view and not an outsider's.

When comparing the data from the three sources (researcher, FaceReader & user), we found that the data of the researcher and the FaceReader were very similar. A preliminary analysis revealed that the user questionnaire is in fact measuring the wrong thing. Although the user had to state how he/she felt during the task, the ultimate answers were rather a reflection of either the content of the application or the reaction dependent on the outcome of the task (successful or not). We further found a small distortion in the data of the FaceReader. A comparison of the researcher's loggings and the FaceReader's results showed that the FaceReader registers angry behaviour when the test user seems concentrated and serious.

Measuring fun of use

Desmet (2003) defines 'fun of use' as: "the fun one experiences from owning or using a product". Because of the affective character of 'fun of use' [4], the FaceReader was used to measure it. However, fun of use is more than an emotion as such [3]. Therefore, a combination of the FaceReader and another observation method was necessary.

The FaceReader proved to be an efficient aid to register the fun of use of an application. In contrast to the FaceReader that registers the emotions more than twice a second, a researcher cannot observe every small change in emotions. As regards the user questionnaire, we already stated that it does not correctly measure what the user thinks and feels during the task-solving process.

The first results show two different scenarios depending on the outcome of the task. We will illustrate this with an

example of a task that must be accomplished using a news website that was rated bad with regards to usability by a preceding expert evaluation. The first scenario concerns the test users who could not accomplish the task. During the task, the FaceReader mostly registered a sad or angry behaviour pattern. But when filling out the questionnaire, the test users evaluated themselves as if they had been rather happy and 'neutral'. As regards the second scenario in which the test users successfully finished the task, the FaceReader firstly registered an angry behaviour response (confirmed by loggings). Similar to the first scenario, the test users were somewhat frustrated by the bad usability of the website. As they managed to accomplish the task, the FaceReader no longer registered angry behaviour but a more neutral one. When filling out the questionnaire however, the test users evaluated themselves as 'happy'. The self-evaluation seemed to be biased by the successful outcome of the difficult task.

CONCLUSION

The present study examined the value of a new tool for measuring emotions, the FaceReader. A comparison of the results obtained by the researcher, user questionnaires and the FaceReader revealed that the latter is an effective tool to measure instant emotions and fun of use. The advantage of the FaceReader over the loggings of a researcher comprises the precision whereby every small change in emotion is registered. As regards the user questionnaire, we found that it does not measure instant emotions or fun of use. The users' self-reflection is more a reflection of the content of the application or the outcome of a task (successful or not). The data of the researcher and those of the FaceReader complement each other well. The observation loggings of a researcher are needed to notice subtle differences in the basic emotions discerned by the FaceReader. In addition to this, a researcher can also interpret context (e.g. environment, content) and verbal cues such as sighs, voice intonation and what is said.

FURTHER WORK

Further research has to confirm whether the preliminary results found in this initial research with the FaceReader can be confirmed and the assumptions can be substantiated. Another focal point in a following research set up will be to evaluate if the emotions the test users communicate, correspond with the emotions they express. The communication of emotions in this test set up occurs through thinking aloud (verbal), through questionnaires (verbal) and through facial expressions (non verbal).

ACKNOWLEDGMENTS

This paper is the result of research carried out as part of the CIcK project, funded by the Interdisciplinary institute for Broad Band Technology (IBBT). The project is being carried out by a consortium of companies. The article would not have been possible without the co-operation of Noldus and VicarVision. In particular, we are grateful to Hans Theuws (Noldus) and our colleague Kenny Verbeke for their support and assistance with the use of the FaceReader. We would also like to thank David Geerts for the contributing suggestions to this paper. Finally, special thanks are due to Simon Shrimpton-Smith, for the correction of language errors.

REFERENCES

1. Branco, P., Firth, P., Encarnaçao, L.M., and Bonato, P. Faces of emotion in Human-Computer Interaction. In *Late breaking results CHI 2005*, ACM Press (2005), 1236-1239.
2. Chin, J.P., Diehl, V.A. and Norman, K.L. Development of an Instrument Measuring user satisfaction of the Human-Computer Interface. In *Proc. CHI 88*; ACM Press (1988), 213-218.
3. Den Uyl, M. J. and van Kuilenburg, H. The FaceReader: Online Facial Expression Recognition. In *Proc. Measuring Behaviour* (2005), 589-590.
4. Desmet, P. M. A. Measuring emotions. In M. Blythe, C. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology: From Usability to Enjoyment.* Kluwer, Dordrecht, the Netherlands, 2003.
5. Ekman, P. Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. *Psychological Bulletin 115*, 2 (1994), 268-287.
6. Ekman, P. and Friesen, W. V. *Manual for the Facial Action Coding System.* Consulting Psychologists Press, Palo Alto, CA, 1977.
7. Hazlett, R. Measuring emotional valence during interactive experience: boys at video game play. In *Proc CHI 2006*, ACM Press (2006), 1023-1026.
8. Hirschman, E.C., and Holbrook, M.B. Hedonic consumption: emerging concepts, methods and propositions. *Journal of marketing*, 46 (1982), 92-101.
9. ISO/IEC. 9241 Ergonomic requirements for office work with visual display terminals (VDTS) Part 11. Guidance on usability, ISO/IEC, Switzerland, 1998.
10. Lyons, M.J., and Bartneck, C. HCI and the Face. In *Workshop CHI 2006*, ACM Press (2006), 1671-1672.
11. Mahike, S., Minge, M. and Thüring, M. Measuring multiple components of emotions in interactive contexts. In *Work in progress CHI 2006*, ACM Press (2006), 1061-1066.
12. Mandryk, R.L., Atkins, M.S., and Inkpen, K.M. A continuous and objective evaluation of emotional experience with interactive play environments. In *Proc. CHI 2006*, ACM Press (2006), 1027-1036.
13. Nielsen, J., Clemmensen, T. and Yssing, C. Getting access to what goes on in people's heads?: reflections on the think-aloud technique. In *Proc. of the second Nordic conference on Human-computer interaction,* ACM Press (2002), 101-110.