MARPLE Investigates: An 'Adversarial' Approach to Evaluating User Experience

Jane Holt
Lancaster University
Computing Department
Infolab 21
+44 (0) 1524 510311
holtj1@comp.lancs.ac.uk

Simon Lock
Lancaster University
Computing Department
Infolab 21
+44 (0) 1524 510304
lock@comp.lancs.ac.uk

ABSTRACT
User experience of interactive systems has always been difficult to assess due to its very subjective nature. In this paper we present a new approach to evaluating experience based around an adversarial courtroom metaphor.

Categories and Subject Descriptors
H1.2 User/Machine Systems; H5. Information Interfaces and Presentation.

Keywords
User Experience, Interactive, Subjective, Evaluation, Adversarial.

1. INTRODUCTION
The last ten years has seen a growth of interest, within the Human-Computer Interaction (HCI) community, towards User Experience. User experience is concerned with the 'quality of the process' that the user *feels* during interaction. User experience typically includes such aspects as fun [4, 22], creativity [3, 15], playfulness [14] and aesthetics [15]. Efforts are currently being made to design and evaluate systems which take these aspects into account. However, this is a challenging prospect due to the emotional aspects of user experience.

The term 'emotion' is in itself difficult to define. There are many differing opinions as what emotion is and how it functions, from such fields as psychology [6, 7, 9, and 11] and neuroscience [19]. What appears to be a repeating theme though, is that emotion is seen as 'a mental state that does not arise through free will, and that is often accompanied by physiological changes' [9]. These states can be triggered by either an internal or external stimulus, e.g. a pleasant memory or a perceived threat, which can then manifest themselves as an increase in heart rate and breathing, changes in facial expression or tensing of the muscles. There are said to be six basic emotions that we all experience, such as anger, disgust, fear, joy, sadness and surprise [6], though the number and type varies amongst psychologists [6, 7, 9 and 11].

Emotional states are both fleeting and unpredictable, and are open to individual interpretation which can also be subjective. What is interesting to note is that traditionally, emotions were seen in a negative aspect, as self-limiting and obstructive as compared to a logical, rational approach. However, due to research undertaken into the field of affective computing, emotions have come to the forefront with the emphasis now on ensuring systems are enjoyable as well as usable.

Incorporating this into the design process is problematic, and this equally applies to evaluation. Usability has long established methods in order to determine whether a system is efficient, effective and satisfying. Conversely, user experience is a relatively new area which is focused upon developing frameworks and approaches [3, 8, 12], which will include evaluation guidance. Experiments have been conducted into this type of evaluation [1,] which draw upon existing usability techniques, such as questionnaires [5, 18], interviews, [5, 18], heuristics [18] and 'think aloud' protocols [17]. These are well used techniques, in both academia and industry, which appear to provide reliable performance measures.

However, user experience is a wide ranging area of research which has engendered a variety of approaches. Many researchers are adopting a 'holistic' approach by focusing on the subject as a whole [3, 8], whereas others have chosen to concentrate on single aspects such as fun [22, 4], enjoyment [17], play [15] and motivation [8].

2. EXISTING APPROACHES
Currently there appears to be no single standard approach to the issue of evaluating pleasure, as part of user experience. Discussion is centred around whether to rework 'old' or existing methods of evaluation, specifically for user experience, or to develop new techniques. Another option is that of combining techniques from usability with user experience thereby assessing both task performance and emotional impact.

What is recognized by many researchers is the challenge of designing a common methodological approach [3] for this purpose. It has even been suggested that this may be an insurmountable problem due to the imprecise and ambiguous nature of emotions. However, we have already mentioned that there is a growing awareness of the importance of emotions during user interaction and so efforts are being made to find ways of assessing their impact.

We have already mentioned that some researchers have chosen to use existing usability techniques, such as interviews – both structured and semi structured [5, 18] and questionnaires [5, 18]. These and other usability methods are frequently used throughout the usability community and as such can be said to be tried and tested though not entirely conclusive. This may be

more of an issue with qualitative methods which tend to rely upon interpretation of data rather than pure statistics [18].

As well as these, other approaches are being used: one approach uses bio sensing or physiological measures. These are metrics which allow us to obtain physiological data in order to assess user emotion and stress [14]. Commonly used in cognitive psychology, this includes measuring heart rate via an Electrocardiogram (ECG) or the conductivity of the skin via Galvanic Skin Response (GSR). Increases in heart rate or the temperature of the skin can indicate periods of stress and anxiety experienced by the user during interaction.

Another is that of 'pastiche scenarios': a team at the University of York have been using these as a form of evaluation. These are short narratives which borrow from other texts to show how a user might interact with a device. Similar to a diary yet also closely allied to persona creation, they aim to capture emotionally rich data from everyday occurrences [2].

A more structured method is advocated, based upon an experiential value scale which measures user responses via seven point scale. Based upon marketing research, it uses nine indicators, for example, intrinsic enjoyment, that are then used against a series of statements related to user experience. These statements are measured via this scale according to their levels of agreements [21].

Other new methods have been proposed, for example, 'cultural probes' in which users self report about their interaction experience via a package consisting of a diary, camera and task cards, which records their actions, thoughts and feelings [1, 10]. The use of 'anticipation' interviews in which users are encouraged to keep voice note diaries, followed by a reflection interview [20], and a multiple method which uses in situ observations, 'obstacle cards' and interviews, at various stages in order to assess customer behaviour for an e-commerce website [16].

We have highlighted a few of the various approaches currently being developed, which are drawn from both the qualitative and quantitative domains. Some of these advocate using a single technique whereas others suggest a multiple approach. What could be argued is that either will have advantages and disadvantages which are discussed in the next section.

3. COMPARISON AND CRITIQUE

The use of a single technique, for example, a questionnaire, may be considered perfectly appropriate for user experience evaluation. These have been and are used on a regular basis in both academia and industry: a single technique could be particularly beneficial when both time and money are subject to constraints, for example, conducting an 'expert walkthrough' [18]. On the other hand, if undertaking a protracted series of evaluations, then the utilisation of several techniques may be more profitable. In either case, a systematic approach seems to bring dividends – when determining usability. Usability is often focussed upon the achievement of specific criteria [18] within a context of use, so having what appear to be clearly defined goals and objectives could make things easier.

However, the application of a rigid 'checklist' approach to what can be a fluid, unpredictable and sensual process may not be ideal. It may be more beneficial to use a range of techniques which are based upon interpretation rather than metrics, for example, combining interviews with video observation and/or photography, or cultural probes followed by an interview. As we are dealing with qualitative measures, a single technique might not be strong enough or reliable enough to accurately capture data. Table 1 shows a sample range of evaluation methods which are predominantly qualitative, so it may be more worthwhile to use several of these techniques together in order to provide a solid justification when discussing the results.

4. THE MARPLE METHOD

MARPLE stands for the Multi-lateral Assessment and Review of Pleasure Laden Experiences and is a methodology developed as a response to the issue of evaluating user experience. Due to its interpretative nature, we have devised a largely qualitative approach which is both systematic and integrated. A wide variety of existing approach (such as interviews and video observation) can be used to gather evidence. As illustrated in figure 1, MARPLE provides a framework which drives these individual approaches, integrates their results and supports the building of cases to argue about specific aspects of user experience.

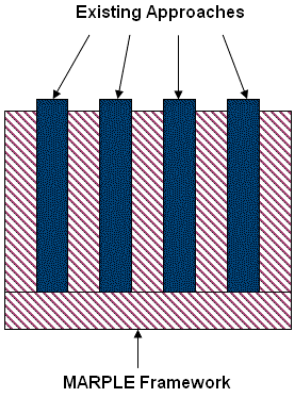

Figure 1. The MARPLE integrated approach

In keeping with the name and the 'adversarial' metaphor, this methodology aims to build a case for the assessment of user experience, from which it will be decided whether that was successful or not. It uses these techniques, in conjunction with each other, to gather evidence, which is then offered as proof at the post session discussion. A single piece of evidence could be seen as too flimsy and not capable of fully capturing the emotional aspects whereas several pieces of evidence, used together could help reinforce an argument.

We also advocate the use of two investigators who will take up positive and negative stances, or in reference to the adversarial metaphor, the 'prosecution' and the 'defence'. These investigators will be looking to determine if the interaction was pleasurable or not and the reasons why. The rationale behind this stance is that they may be able to counter accusations of personal subjectivity and bias. One will argue for the positive determinants of a pleasurable interaction and the other will argue for the negative aspects. Then, the two will swap roles and continue the discussion until a final judgement is reached via a third party. The MARPLE Method is heavily predicated upon the prosecuting/defence investigator stance which we think makes it a novel approach to the issue of user experience evaluation.

Work is currently being undertaken on a toolkit which can be used to conduct evaluation of interaction experiences. This toolkit will comprise of a series of heuristics – for an expert review, a set of guidelines as to what to consider during evaluation, a tutorial which outlines the pleasure evaluation process, from the initial setting up of the equipment through to

data analysis and post session discussion. Its systematic approach includes obtaining participants, the setting up of the test situation, the evaluation session using video and interviews, data analysis and then the post session discussion with the prosecuting/defence stance. The post session discussion will be followed by the production of an evaluation report.

The MARPLE process model provides an overall framework for structured evidence gathering, analysis and interpretation. This model starts with the setting up of the system, the interaction itself and evidence gathering, through to data analysis, the post session discussion and the creation of a case for a positive or negative experience.

MARPLE can be used at various stages throughout the design cycle, where evaluation needs to be undertaken, e.g. low fidelity stage, and, after implementation, before the system released into the public domain. It is also designed to complement current usability methods.

5. EVALUATION: HCI 2006

MARPLE applied to assess user experience of an interactive system during HCI 2006. In order to demonstrate and evaluate MARPLE we set up a 3D interactive installation (illustrated in figure 2) which used anthropomorphic interface agents, a games console and two unusual input devices. The two unusual input devices - a couple of children's dolls were chosen as we thought they epitomised playfulness and hopefully, would engender a pleasurable experience.

As this was to be a 'quick and dirty' evaluation, we set up the interaction in a location where workshop delegates tended to congregate. Due to time constraints we were only to conduct short evaluations with a maximum of five users. These users were mainly from the performing arts/digital arts domains and were evenly balanced in terms of gender, age and expertise. Each user would sit in front of the installation and would emotionally interact with two 3D interface agents via the doll input devices. These dolls were attached via retractable wires to a games console which the user could manipulate in any way they saw fit – see Figure 2. Each session involved an informal interview with video observation and photography and lasted, on average, around 5 to 10 minutes.

This was followed by data analysis of the interview transcripts and a post session discussion. The data analysis involved grouping the stills into a linear sequence or 'storyboards', to ascertain a pattern of interaction and any distinguishing features. The video footage was viewed in much the same way, with the emphasis on looking at facial expression, gesture and body language in order to highlight positive/negative aspects.

Similarly with the interview transcripts: these were analysed in order to look for repeating words or emerging patterns of behaviour. As the interviews themselves were short, this meant a limited amount of user feedback, though we still managed to elicit some useful comments. The results showed that overall, the users enjoyed the interactive installation, with such words as 'playful', 'funny' and 'good' which suggest that this was a positive experience. There were a few suggestions for improving the installation, such as the 'time lag' between the user's action and the agent's reaction.

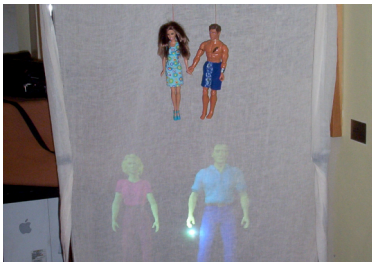

Figure 2. Interactive Installation (HCI 2006)

The resulting data was then open to discussion by both investigators who adopted the prosecuting and defence lawyer stance. One investigator argued in favour of a positive experience by describing various features which they felt supported this; such as expressive body language, smiling face etc, whereas the other investigator opposed this by pointing out indicators of a negative experience. Eventually, agreement was reached on a variety of issues though it was suggested that a third 'neutral' party might be needed.

The evaluation of the MARPLE approach proved very successful and yielded some useful and positive feedback. It also raised a number of important issues relating to the evaluation of user experience. One issue is the problem of interpreting photographic evidence, and as such, can not be taken as conclusive. Also, one aspect of MARPLE still open to debate is the friendly, 'chatty' and yet semi structured interview or 'stealth interviewing' which is designed to be unobtrusive and still obtain useful feedback. We have designed a series of suggested interview questions, which are concerned with conceptual issues rather than issues of 'how do you do this', 'what did you expect to happen' and so on. There is the tendency to slip into usability interviewing, particularly if the investigator has previous experience in this area and so there needs to be an awareness of this issue.

The MARPLE Method is still in its infancy and so further work needs to be undertaken before advocating its adoption by both academia and industry. This means a look at other multi-faceted approaches, an appraisal of the techniques we have chosen and why, and, its suitability and appropriateness as a means of evaluating user experience.

6. REFERENCES

[1] Axelrod, L and Hone, K.S. *Affectemes and Allafects*: *a novel approach to coding user emotional expression during interactive experiences*. Behaviour and Information Technology, Vol. 25, No 2, March-April 2006, pp 159-173.

[2] Blythe, M. *Pastiche Scenarios*. Interactions, September + October 2004.

[3] Blythe, M; Wright, P; McCarthy, J and Bertelsen, O.W. *Theory and Method for Experience Centred Design*. CHI 2006, April 22-27, 2006, Montreal, Quebec, Canada.

[4] Carroll, J.M. *Beyond Fun*. Interactions, September 2004.

[5] Dix, A; Finlay, J; Abowd, G.D; and Beale, R. *Human-Computer Interaction: Third Edition.* Pearson Prentice Hall, Essex, England, 2004.

[6] Ekman, P. *Facial Expression and Emotion*. 1992 Award

Address. American Psychologist, April 1993, Vol 48, No 4, pp 384-392.

[7] Ekman, P. *Basic Emotions*. Chapter 3, Handbook of Cognition and Emotion. Sussex, UK. John Wiley & Sons Ltd, 1999.

[8] Forlizzi, J and Battarbee, K. *Understanding Experience in Interactive Systems*. DIS2004, August 1-4, 2004, Cambridge, Massachusetts,

[9] Goertzel, B. *A General Theory of Emotion in Humans* and Other Intelligences. DynaPsych Table of Contents. http://www.goertzel.org/dynapsyc/2004/Emotions.htm

[10] Jaasko, V and Mattelmaki, T. *Observing and Probing*. DPPI '03, June 23-26, 2003, Pittsburgh, Pennsylvania, USA.

[11] James, W. *What is an Emotion?* Classics in the History of Psychology. An internet resource developed by Christopher D. Green, York University, Toronto, Ontario. http://psychclassics.yorku.ca/James/emotion.htm

[12] Knight, J and Jefsioutine, M. *The Experience Design Framework: from Pleasure to Engagability*. Conference HCI, the Arts and Humanities, 2003, University of York, UK. http://www-users.cs.york.ac./ ~pcw/KM_subs/Knight_Jefsioutine.pdf.

[13] Mandryk, R.L; Atkins, M.S and Inkpen, K.M. *A Continuous Objective Evaluation of Emotional Experience with Interactive Play Environments*. Proceedings of the SIGCHI conference on Human Factors in computing systems, Novel Methods: Emotions, Gestures, Events (CHI 2006), (Montreal, Quebec, April 22-27, 2006).

[14] Mandryk, R.L; Inkpen, K.M and Calvert, T.W. *Using psychophysiological techniques to measure user experience with entertainment technologies*. Behaviour and Information Technology, Vol 25, No. 2, March-April 2006, pp 141-158.

[15] Myers, D. *What's good about bad play?* Proceedings of the Second Australasian Conference on Interactive Entertainment, 2005, Sydney, Australia 2004.

[16] Petre, M; Minocha, S; and Roberts, D. *Usability beyond the website: an empirically-grounded e-commerce evaluation instrument for the total customer experience*. Behaviour & Information Technology, Vol 25, No 2, March-April 2006, pp 189 – 203.

[17] Picard, R.W. *Affective Computing for HCI*. Ergonomics and User Interfaces 1999.

[18] Preece, J; Rogers, Y; and Sharp, H. *Interaction Design: Beyond Human-Computer Interaction*. Wiley, USA, 2002.

[19] Simon, V.M. *Emotional Participation in Decision-Making*. Psychology in Spain, Vol 2, No. 1, pp 100-107.

[20] Swallow, D; Blythe, M and Wright, P. *Grounding Experience: Relating Theory and Method to Evaluate the User Experience of Smartphones*. Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics, Chania, Greece. Session: On theories, methods and techniques. Vol. 132, 2005, pp 91-98.

[21] Toms, E.G; Dufour, C and Hesemeier, S. *Measuring the User's Experience with Digital Libraries*. JCDL '04, June 7-11, 2004, Tucson, Arizona, USA.

[22] Wiberg, C. *Usability and Fun: An overview of relevant research in the HCI community*. http://www.sics.se/~kia/evaluating_affective_interfaces_/ Wiberg_2.doc.