Using Affect to Evaluate User Engagement

Jennefer Hart

Manchester Business School

University of Manchester

Booth Street West

Manchester, M15 6PB, UK

jennefer.hart@postgrad.mbs.ac.uk

Alistair G. Sutcliffe

Manchester Business School

University of Manchester

Booth Street West

Manchester, M15 6PB, UK

ags@man.ac.uk

Antonella di Angeli

Department of Information

Engineering & Computer Science

University of Trento

Via Sommarive

14-38123 Provo, Trento, Italy

antonella.deangeli@unitn.it

Abstract

User Experience (UX) emerged beyond the traditional views of usability to account for users' emotional response to the aesthetics of an interactive product. This paper outlines the first of a series of studies on User Engagement (UE), a subset of UX, which focuses upon the quality of the within session interactive experience. The aim of this study is to explore affect through the responses to interactive features and how this impacts upon user judgment. Initial findings indicate that websites with more interactive features generate enhanced positive affect within session, which may predominate over a longer term, thus impacting on the overall user experience.
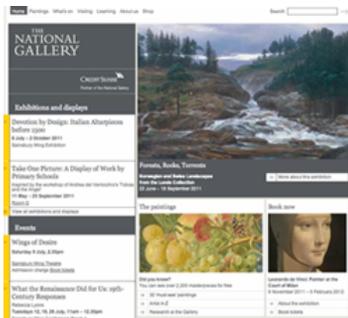
Keywords

User Experience; hedonics; flow; presence; aesthetics.

ACM Classification Keywords

H.5.2 [Information Systems User Interfaces]: Evaluation/Methodology

General Terms

Human Factors, Theory, Measurement

Figure 1: The three interactive websites used within the study. The top is the National Gallery website which acted as a control. The second is the Google Art Project, which offers an immersive 3D experience, where users can navigate through space; the Louvre is the third site, which provides an interactive environment including an audio guide with animation movies.

Introduction

UX research has focused on the relationship between the pragmatic aspects (usability) and the hedonic affects (emotion) generated from the aesthetics or beauty of an interactive product [14]. Findings indicate that the aesthetic appeal of a product can influence perceived usability, although user judgments can vary depending on the task, context, users' prior knowledge and background [4, 13]. Key conceptual frameworks or models of UX have been created as a means of explaining the perceptual constructs of UX [6, 7]. Yet these varying theories have yet to provide a universally agreed definition of UX. This paper focuses upon User Engagement (UE), which refers to a subset of UX, or the within session experience with an interactive product, whereas UX encompasses the whole experience of a product lifecycle from beginning to end [12, 13].

Norman first raised the importance of emotion and how emotions emerge depending on an individual response to an event, agent or object [10, 11]. There have been two approaches to understanding affect within UX: to explore the various constructs by which we judge product quality, or to understand the process of making quality judgments. Lavie and Tractinsky [9] first developed a questionnaire-based instrument to measure users perceptions of pleasure, aesthetics and other design qualities such as usability and service quality. Hassenzahl [7] took a similar questionnaire-based approach to investigate the constructs of pragmatics (usability) and hedonics (emotions). In contrast Hartman et al. [6] proposed a process model of quality judgment that integrated criteria from [9] with influences from user and task characteristics. However both approaches have not accounted for the

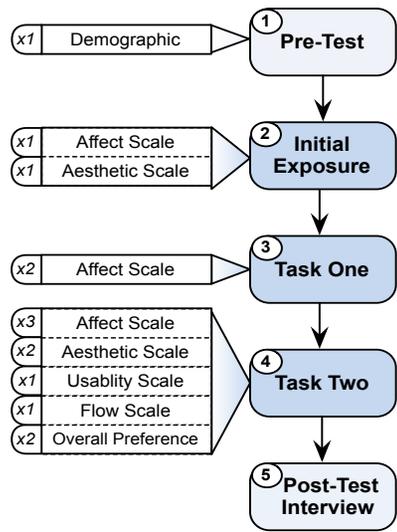effect of interaction on UE the related concepts of flow, immersion and affect.

The concepts of affect (mood and emotion), flow (immersion) and presence (involvement) have been researched in other fields [2, 3, 16], yet how these principles influence UE is poorly understood. Affect is defined here as a combination of emotion, which is typically of shorter duration, and mood that is often more longer lasting [1, 5]. To maintain engagement during interaction positive affect and high arousal need to be promoted by interesting, stimulating and exciting interfaces. Interaction design may promote positive affect and arousal through serendipitous effects, variable pace, use of avatars and virtual environments, thus generating UE. The influence of task, context and user characteristics on perceptions of engagement are also important.

The Study

This study tested the hypothesis that highly interactive websites provide positive UE. Three different real-world interactive art gallery websites were selected for their variation of interactive features, e.g., interactive guides, animations and 3D effects (see Figure 1). The application context of all three websites was in the same domain (art galleries), although they varied in their use of metaphors, interactive features and the constructs tested, as shown in Table 1.

	Metaphor	**Interactive Features**	**Constructs Tested**
National	Menu Based	Traditional Grid Layout	Control
Google	Interactive Fly-Through	3D Environment	Flow/Immersion Presence (Individual)
Louvre	Interactive Graphical Object	Animated Avatar Guide	Flow/Immersion Presence (Social)

Table 1: Overview of three websites selected for study

Figure 2: Diagrammatic overview of the study method and scales used.

Affect	Aesthetics	Flow/Pres.
Overall Mood	Attention	Pace
Fun	Confusing	Control
Attractive	Cheap	Challenged
Curious	Complicated	Time
Interesting	Tacky	Awareness
Pleasurable	Unpredictable	Natural
Absorbing	Amateurish	Involved
Exciting	Unimaginative	Compelling
Engaging	Dull	
	Conservative	
	Cluttered	
	Colour	**Overall**
Usability	Unbalanced	**Preference**
Convenient	Images	Visit Again
Functionality	Discord	Recommend
Learnable	Bad	Overall
Navigation	Ugly	Ranking

Table 2: Overview of the six scales used: (affect, aesthetics, flow/presence, usability and overall preference).

Two sites Louvre and Google were chosen for their interactive features, an avatar guide that enabled users to interact with graphical animated objects (Louvre), and a 3D environment that displayed an interactive fly-through metaphor (Google). The National Gallery site acted as control as it offered limited interactivity with simple menu link navigation.

Measures

To investigate the emotional impact on user judgment during interaction a mixed methods approach was used. This paper focuses upon questionnaire scales that form part of a wider study that includes interviews, observation and video capture. Six questionnaire scales were administered and repeated after various tasks, as shown in Figure 2. Apart from the demographics, all scales employed 7-point Likert scales, as follows:

▪ Demographics: General information about the participants was collected, e.g., age, gender, etc.

▪ Affect: A 9-item scale based on PANAS, which focused on one emotive word in order to capture value-charged affect straight after each task [15, 8]. The emotive words were drawn from a mix of sources on arousal, hedonics and emotion [2, 3, 7, 9, 13], as shown in Table 2. This scale was applied three times after each task.

▪ Aesthetics: A 17-item scale was used to measure aesthetic ratings, 11 of these items were taken from the AttrackDiff scale [7], while the remaining 6 items were used to evaluate different media-design characteristics associated with aesthetics. The pragmatic and identification items were excluded to focus on the hedonic experiences. To compare first

impressions with post interaction this scale was applied twice.

▪ Flow/Presence: An 8-item scale derived from a variety of scales was used to capture flow, presence and immersion [2, 3, 16, 13]. This (and the remaining scales) was applied once after completion of all three tasks.

▪ Usability: A 4-item scale was used to capture users perception of the functionality, navigation and utility of the websites after interaction [9, 7].

▪ Overall Preference: A 3-item scale asked if participants would visit the site again, recommend the site and to rate their overall experience of using the site. During the interview participants were asked to rank the sites in order of preference.

Experimental Design

The three websites were manipulated within subjects and counterbalanced in a repeated measures (3 exposures x 3 websites) design. The procedure was as follows (illustrated in Figure 2).

1. Pre-test: After completing the consent form and demographic questionnaire, participants were given prior training by showing them a dummy home page then asking them to complete the affect questionnaire. This prepared them for the initial exposure task and allowed them to become familiar with the affect scale.

2. Initial exposure: Screen shots of each of the three websites homepage was shown for 0.5 seconds using PowerPoint that controlled the exposure time by replacing the image with a mask (grey screen). After each website exposure the affect and aesthetic scales were completed.

Affect Scale Item	Principle Components (X3 Websites) Factor 1		
	Initial Exposure	Task One	Task Two
Current Mood	.904	.932	.933
Fun	.900	.928	.932
Attractive	.900	.927	.916
Curious	.882	.913	.915
Interesting	.855	.909	.907
Pleasurable	.843	.904	.907
Absorbing	.829	.837	.879
Exciting	.764	.807	.801
Engaging	.469	.620	.651
Engenvalue	6.15	6.8	6.9
Variance	68.3%	75.6%	76.7%

Table 3: Factor validity of Affect Scale across three activities.
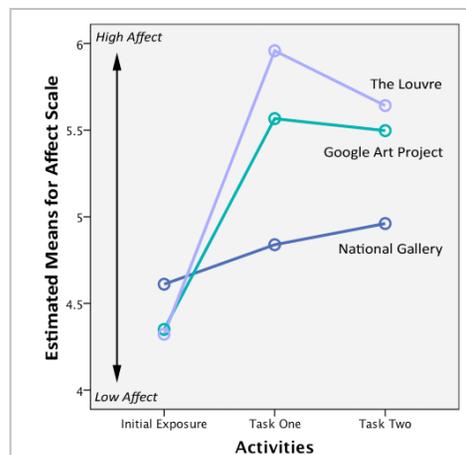
Figure 4: Aggregated Affect Scale for Site and Task, showing significance differences.

3. Task One and Two: Participants carried out two interactive tasks, *navigation* (task one), a general orientation task, and *directed* (task two), that focused on key interactive features. After completing each task, participants were given up to 3 minutes to *explore* anywhere within the site. Repeated measures were collected for affect (3), aesthetics (2), flow (1), usability (1) and overall preference (1).

4. A short interview using semi-structured questions gained more insight into participants' preferences and experiences while interacting with the websites.

The experiments took place within a university audio-visual lab so participants' facial expression and task navigation could be recorded. They were carried out over the summer of 2011 and all participants received a £10 Amazon gift voucher for their participation.

Forty participants (21 female) participated in the study, with age ranging from 18-25(27.5%), 26-35 (52.5%), 36-45 (17.5%) and 46-55 (2.5%). The majority of participants were students (70%), while the remaining were university staff. All participants were educated to degree level and came from a variety of subject areas including business (37.5%), business computing (7.5%), computing (7.5%), humanities (10%), and science (9%).

Results

Questionnaire scores (affect, aesthetics, flow, usability and overall preference) were aggregated (averaging individual items scores) since the scales all produced high levels of internal reliability with Cronbach α ranging from .89 to .96.

Affect Scale

A principal component analysis was run on the affect scores across all three websites with oblimin rotation. The analysis revealed the affect scale was composed of one factor for all three activities (see Table 3). A two-way ANOVA was carried out on the aggregated affect scales using websites (3) and tasks (3) as within subject factors. Mauchly's test indicated that the assumption of sphericity had been violated therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity. The results revealed a significant main effect of task; $F (1.4, 56.2) = 51.2$, $p < .001$, $\eta = .57$, website $F (2, 74.8) = 4.5$, $p < .05$, $\eta = .1$, and the interaction of task and site $F (3.9, 111.1) = 7.7$, $p < .001$, $\eta = .16$.

Post hoc tests using Bonferroni correction indicated that participants' affective responses showed a significant difference for both the Google Art Project and the Louvre site from initial exposure to task one, whereas no difference emerged for the National Gallery (Figure 4). The individual affect items (Figure 5) indicated that for initial exposure participants were more curious and interested than absorbed, excited or engaged etc., although this difference decreased after interaction.

As the aggregated value returned significant effects, a one-way repeated measure ANOVA was carried out on each of the individual items within the affect scale. The National Gallery showed no significant effect on any of the individual items for task. A significant difference was found on all items for Google ($p < .05$ to $p < .001$), apart from curiosity. Louvre also showed a significant effect ($p < .001$) for all individual items. Google showed no significant difference during interaction (between task 2 to task 3), whereas Louvre showed a
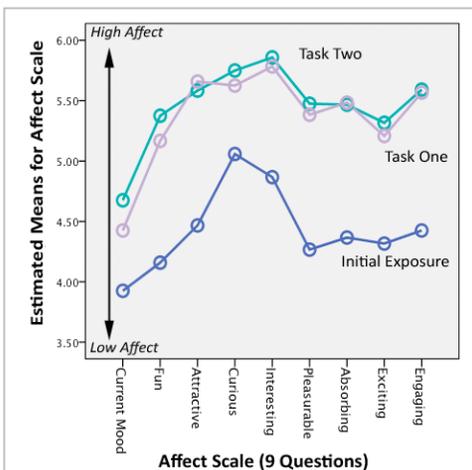
Figure 5: Aggregated Mean for the Individual Affect Scales, showing that mood significantly increases after interaction.

Figure 6: The Aggregated Mean for the Aesthetic scale, showing a significant increase on task (after interaction), where both Google and Louvre ratings increase far greater than National after the interactive experience (also showing the interaction effect).

significant increase for current mood (p< .05), fun (p< .01) and exciting (p< .05).
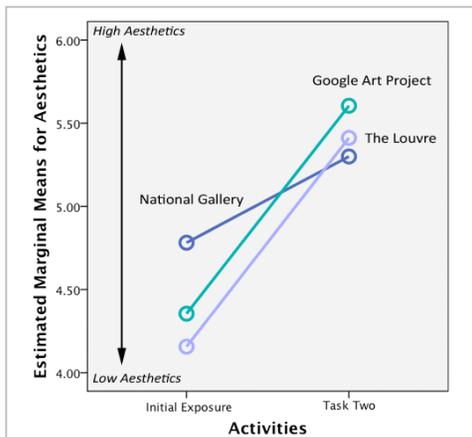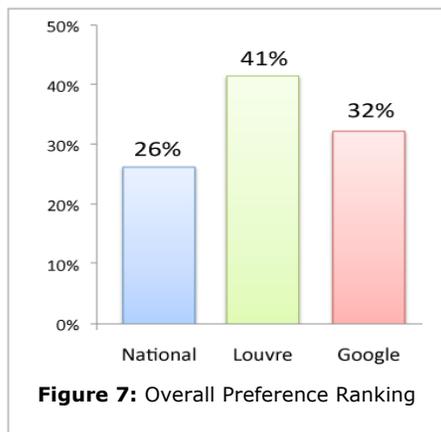
Aesthetic Scale

A repeated measures ANOVA was carried out on the 17-item aggregated aesthetic scale with website (3) and task (2) as factors. Results returned a significant main effect of task; $F(1, 39) = 77.3$, $p < .001$, η = .67, and interaction (task * site), $F(2, 78) = 4.9$, $p< .05$, η = .11. No significant main effect for website was found. Participants rated aesthetics higher for all sites after interaction, however both the Louvre and Google improved more than the National (Figure 6).

Flow/Presence and Usability Scales

A one-way repeated measure ANOVA showed no significant effect across the 3 sites for either scale. Analysis of individual questions within the scales also showed no significant effect.

Overall Preference

A one-way repeated measure ANOVA was carried out on the aggregated 3-item preference scale and no significant effect was found across the three sites. Participants ranking of each website preference were weighted to give an overall percentage rating per site, where 100% resulted if all participants rated the site as first preference. The Louvre site was rated the highest at 41%, followed by Google at 32%, then the National at 26% (as shown in Figure 7).

Discussion

The preliminary findings of this study showed that participant's affective responses changed after users' initial exposure to highly interactive websites (such as Google and Louvre) but not to more traditional ones (e.g., National Gallery). This suggests that interactive

features have an important effect on UE as measured by the affect scale questions. The individual affect items (mood, fun and exciting) continued to improve during interaction only in the Louvre, which suggest this site was the most engaging, which is reflected within the overall ranking, where Louvre was favored the highest. As the Louvre offered a mix of interactive features (animated objects, avatar guide and audio) compared to Google (3D interaction), and National Gallery (menu only), this supports the hypothesis that highly interactive websites provide positive UE. Change in the affective responses was partially associated with the aesthetic ratings, which increased significantly for task, interaction, but not site. Both the Google and Louvre ratings improved after interaction (for affect and aesthetics), although there were no significant differences between them and the control (National Gallery).

Post-test usability evaluation showed no significance difference suggesting all three sites were considered equally usable. The failure to capture any flow effect between task and site could be due to the experimental procedure employed. Retrospective questionnaires rely on participants' memory of the event, which often can be inaccurate. In addition the flow scale was completed after the affect, aesthetic and usability scales, possibly masking any flow experiences that may have occurred.

This study explored users perceptions of UE through measures of affect, flow and presence. Initial findings showed that affective responses change over time and increases within session, especially when using websites with more interactive features. However there were no overall differences post test between the sites for usability, flow/presence or overall preference.

Figure 7: Overall Preference Ranking

Our conjecture is that the usability of all three sites was good, and that equitable content produced no difference in overall preference. However, the preference order, and post-test aesthetics did agree with the affect ratings. It appears that interaction does increase user engagement as measured by affect and, while this effect may be mainly transient within a session it may also have a longer lasting effect on user experience. Further analysis of the data collected during the interviews and video observations will be used to see if users' reports and behavior corroborate with their perceptions of affect and user engagement.

Acknowledgements

References

[1]   Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999) The Role of Emotions in Marketing. *Journal of the Academy of Marketing Science*, 27(2), 184-206.

[2]   Berlyne, D. E. (1960) Conflict, Arousal and Curiosity, *New York: McGraw Hill.*

[3]   Csíkszentmihályi, M. (1990) Flow: The Psychology of Optimal Experience, *New York: Harper and Row.*

[4]   Diefenback, S. & Hassenzahl, M. (2009) The "Beauty Dilemma": Beauty is valued but discounted in product choice, *Proceedings CHI 2009*, pp.1419-1426, ACM Press.

[5]   Frija, N. H. (1993) Moods, Emotion Episodes, and Emotions, In *Handbook of Emotions*, Eds. Lewis. M., & Haviland. J. M. New York; Guilford, 381-403.

[6]   Hartman, J., Sutcliffe, A., & De Angeli, A. (2008) Towards a Theory of User Judgment of Aesthetics and User Interface Quality, *TOCHI,* Vol.15 No.4, pp.1-30.

[7]   Hassenzahl, M. (2004) The Interplay of Beauty, Goodness, and Usability in Interactive Product, *Human-Computer-Interaction,* 19(4), 319-349.

[8]   Karapanos, E., Martens, J., & Hassenzahl M. (2010) On the retrospective assessment of users 'experiences over time, in *Proceedings of the 28th International Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '10*, ACM Press.

[9]   Lavie, T., and Tractinsky, N (2004) Assessing dimensions of perceived visual aesthetics of web sites, *International Journal of Human Computer Interaction*, 60, 269-298.

[10] Norman, D. A. (2004) Emotional Design: Why We Love (or Hate) Everyday Things, *Basic Books, New York.*

[11] Ortony, A., Clore, G. L., & Collins, A (1988) The cognitive structure of emotions, *Cambridge: Cambridge University Press*.

[12] Roto, V., Law, E., Vermeeren, A., & Hoonhout, J. (eds.) (2011) User Experience White Paper, *Results from Dagstuhl Seminar on Demarcating User Experience* Schwartz, M. *Guidelines for Bias-Free Writing*. Indiana University Press, Bloomington, IN, USA, 1995.

[13] Sutcliffe, A. (2010) Designing for User Engagement: Aesthetic & Attractive User Interfaces, (Eds). Carroll, *J.M, Morgan & Claypool Publishers*, Chapter 1.

[14] Tractinsky, N., Katz, A. S., & Ikar, D. (2000) What is beautiful is usable, *Interacting with Computers, 13,* 127-145.

[15] Watson, D., Clark, L. A., & Tellegen, A. (1988) Development & Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales, *Journal of Personality and Social Psychology*, 54, 1063-1070.

[16] Witmer, B. G., & Singer, M. J. (1999), Measuring Presence in Virtual Environments: A Presence Questionnaire, *Presence*, 7(3), 225-224.