Towards an Emotional Validation of Heuristic Approaches for Usability Evaluation.

Yazmín Magallanes*, Ariel Molina-Rueda*, J. Alfredo Sánchez* , Yenny A. Méndez* *

ABSTRACT

This paper presents an initial investigation on how an emotional assessment of user interfaces may extend and validate the heuristic evaluation of the usability of interactive systems. Through the elicitation of responses from two groups of users, we focused on detecting whether a correlation exists between emotional assessment of specific interface issues and the corresponding evaluation by interaction experts. Using a prototypical multi-tactile interface and a heuristic approach, we first obtained feedback from interaction experts regarding its potential problems and issues. We then devised scenarios for observing novel users and eliciting their emotional responses to the issues raised by our experts. Our initial findings indicate a significant correlation exists between the issues raised by expert users and the emotional responses of novel users, but also that some of those issues do not necessarily lead to long lasting negative emotions. We thus posit that emotional evaluation of interactive systems may be helpful for enriching and validating heuristic approaches for usability evaluation

RESUMEN

Este artículo presenta una investigación inicial sobre las maneras en que la evaluación emocional de interfaces de usuario puede extender y validar la evaluación heurística de sistemas interactivos. Mediante la recolección de respuestas de dos grupos de usuarios, este trabajo se enfoca a detectar si existe una correlación entre la evaluación emocional de aspectos específicos de interfaces de usuario y la evaluación correspondiente realizada por expertos en interacción. Primero se obtuvo retroalimentación de expertos en interacción acerca de los problemas potenciales de una interfaz multi-táctil, con base en un enfoque heurístico. Posteriormente, se plantearon escenarios para observar a usuarios inexpertos y obtener respuestas emocionales a los aspectos considerados problemáticos por los expertos. Nuestros hallazgos iniciales indican que hay una correlación importante entre los aspectos problemáticos señalados por los expertos y las respuestas emocionales de los usuarios inexpertos, pero también que algunos de los aspectos problemáticos no necesariamente implican emociones negativas duraderas. Sugerimos entonces que una evaluación emocional de sistemas interactivos puede enriquecer y validar los enfoques heurísticos para evaluación de usabilidad.

INTRODUCTION

Heuristic approaches for evaluating the usability of interactive systems are very popular, mainly because they are inexpensive and relatively rapid to apply. In essence, heuristic usability evaluation relies on the inspection of a user interface by a small group of experts who look for violations to well-established guidelines that would pose problems when systems are released. Its main advantage also represents its greatest drawback: Heuristic evaluation approaches do not require end user participation.

This means that evaluators will not need to recruit users for usability studies, nor design tasks or have users work with the actual interfaces. Whereas it saves time, heuristic usability evaluation lacks key information on user experience that can only be elicited from actual stakeholders. In this paper, we suggest that heuristic usability evaluation may be extended and its findings may be validated by applying an emotional evaluation technique that involves a small number of non-expert users and designing simplified scenarios that result from the observations of expert users.

*Laboratory of Interactive and Cooperative Technologies, Department of Computing, Electronics and Mechatronics, Universidad de las Américas Puebla, Santa Catarina Mártir S/N, Cholula, Puebla, Mexico. E-mail: {yazmin.magallanesvz, ariel.molinara, alfredo.sanchez}@udlap.mx
**Software Engineering Research and Development Group, Universidad del Cauca, Calle 5 No. 4 - 70, Popayán, Colombia.. E-mail: ymendal@unicauca.edu.co

RELATED WORK

There has been a significant amount of research on user experience, which has helped in reaching a commonly accepted definition and developing methods and techniques for evaluation. However, in-depth discussion is still need in academia and industry on the theoretical roots and foundations of these activities [1]. Although the design of interactive products aims to emphasize ease of use via specific product characteristics, there are not clear criteria for enabling the development of enjoyable products and services.

The above statements highlight the importance of understanding how emotions determine the quality of interactions with a product and are related to the evaluation of user experience. User Experience (UX) research focusing on the emotional and experiential aspects of system usage is of highly recognized relevance for the CHI community [1].

Positive emotions will evoke a positive evaluation of the experience. In [2], three important aspects are suggested as important for generating pleasurable and significant experiences with products: understanding, measuring and designing for emotion.

PrEMO [4] is an instrument for measuring emotions experienced by users of products, which has played a key role in our study. PrEMO is a non-verbal, self report instrument that measures 14 emotions that are often elicited in the process of designing products. Instead of relying on the use of words, respondents report their emotions via a set of 'cartoon' animations, each expressing one of the 14 relevant emotions.

In [3], an approach is presented to measure emotions provoked by the visual appeal of websites. Online experiments were conducted to measure the emotional experience associated with twelve screenshots of a university website. In addition to PrEMO [4], de Lera and Garreta [5] propose gestural heuristics to evaluate the emotional state of users through the assessment of their expressive reactions during the process of evaluating an interface. This contribution is mainly focused to acquiring a greater understanding of the user experience. With the use of heuristics, it is intended to complement the analysis obtained through usability testing and subjective responses. Lindgaard et al. [6] conducted a study on first impressions caused by websites, and found that users can assess the visual appeal of a site in less than 50 milliseconds. Patrick [7] states that the visual appeal is a determinant of pleasure during a process of interaction, which focuses on the emotions recalled by the visual appeal as a starting point for the development of a web-based tool intended to uncover the emotional impact of websites, prototypes or concepts. Other efforts aimed at understanding or measuring affective aspects of usability include [8] [9] and [10], which report studies conducted on the relationships between perceptions of user interface aesthetics and usability.

EXTENDING AND ENRICHING HEURISTIC EVALUATION

In order to use the expression of emotions by potential users as a means to enrich and potentially validate heuristic evaluation, we have proceeded in two phases: (i) Heuristic evaluation by experts, and (ii) Evaluation involving emotions.

Heuristic evaluation by experts

Heuristic evaluation is used to identify interface problems with respect to a well known set of heuristics, which are used by experts to produce recommendations. The steps we followed for evaluating a multi-tactile interface we refer to as InnovIMMS involved:

- Free Evaluation. Each evaluator interacted with the application for 1 to 2 hours. Every expert was asked to inspect the application and identify usability problems (See fig 1).

- Directed Evaluation. Experts were invited to numerically evaluate each item in a previously established checklist, as seen in table 1. The assessments were based on the scale presented in table 2.

- Expert Comments. Each expert was asked to fill out a commentary form for each problem in the previous step. Experts used table 2 in addition to the commentary form.

Table 1 .

List of problems to evaluate

Id	Problem to evaluate
1	It is intuitive for users to know what actions are associated with objects in the interface.
2	It is intuitive for users to know what actions are associated with objects in the interface.
3	It is intuitive for users to know what actions are associated with objects in the interface.
4	It is intuitive for users to know what actions are associated with objects in the interface.
5	The actions performed by users are similar to those implemented in the real world.
6	New gestures do not generate additional cognitive load on the user.

Table 2 .

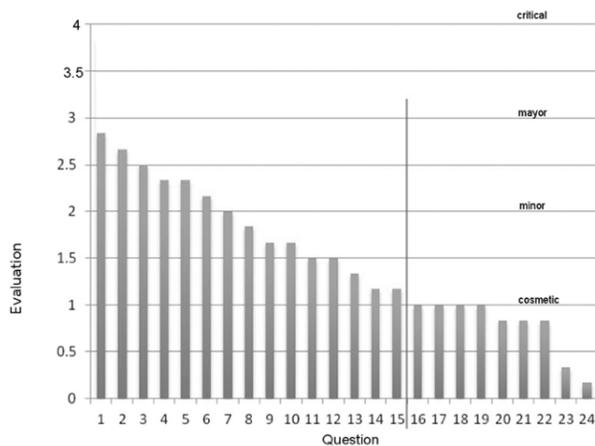| Valor	Degree of severity
Scale for heuristic evaluation	
0	The interface does not present any problem.
1	Cosmetic problem: Does not need to be solved unless time permits.
2	Minor problem: Its prompt solution has low priority.
3	Major problem: it is important to fix it; should be given high priority.
4	Critical problem: It is imperative to fix before releasing

Figure 1 . Experts were presented with 24 issues. Their job was to evaluate the severity of each issue. After gathering the results, the 15 most critical were detected (red line). Cosmetic issues were ignored.

Emotional Evaluation

A group of users with little or no experience with multi-touch interfaces were asked to use InnovIMMS in a semi-directed method. They interacted with the application in some activities, which were selected taking into account the usability problems identified in the previous phase. For each of these activities, users identified the emotions they experienced when the usability problems implied by these.

The objective of this phase is to relate user emotions with interface problems previously detected by experts. We adapted the PrEMO evaluation instrument [4] so emotions were classified as shown in fig. 1 and presented to the users in Spanish, as this was their native language. A brief explanation of each emotion in the figure is presented next, along with the term we used in Spanish (in parentheses).

Negatives:

- **Disgust (repugnancia).** Rejection of the interface.

- **Indignation (indignación).** Anger due to unfair or unmanageable behavior.

- **Contempt (desprecio).** Disdain, scorn or lack of affect.

- **Dissatisfaction (insatisfacción).** Unsatisfying experience.

- **Disappointment (desilusión).** A negative impression caused by the interface not fulfilling user expectations.

- **Unpleasant surprise (sorpresa desagradable).** A negative unexpected feature.

- **Boredom (aburrimiento).** Uninteresting interface features.

Positives:

- **Desire (deseo).** Impulse to use the interface.

- **Inspiration (inspiración).** Encouragement to do new things.

- **Admiration (admiración).** Appreciation of the interface.

- **Satisfaction (satisfacción).** Contented with available features.

- **Fascination (fascinación).** Intense attraction and good impression caused by the interface.

- **Pleasant surprise (sorpresa placentera).** Positive unexpected feature.

- **Amusement (diversión).** Fun experience.

The evaluation process during this phase consisted of the following steps:

- The user performed a list of predefined activities. At all times an expert was present in order to assist the user in case of difficulty. For each of the selected activities, users were asked to assess 14 emotions. They were presented with a table with images corresponding to emotions (see fig 2) and a scale to rate each of these (see table 3).

Figure 2 . A set of emotions adapted from PrEMO [4].

Table 3 .

Scale for assessment of emotions.	
Value	**State**
0	I do not feel this
1	I feel this somewhat
2	I feel this somewhat

RESULTS AND DISCUSSION

The InnovIMMS application allows users to manipulate three main types of objects: Text, images and a canvas for painting. The application includes a fan menu and other graphical elements that provide easy access to resources and tasks [12]. The evaluation involved six experts during the first phase and eight users for the second.

Results from the expert heuristic evaluation phase

Based on the ratings assigned by experts, we identified those that had a higher score, which meant that problems had been found. Some of these problems are listed below:

- How well users knew what to do with objects in the interface.

- Size of the table to interact with objects

- Similarity of the gestures used with respect to those employed in other applications.

- How comfortable users were when interacting with objects in the application.

- User certainty on what to do with the virtual objects.

- Help for recovering from errors.

- Intuitiveness of gestures to perform functions on the system.

- Accuracy to interact with objects on the surface.

- Functionality to rotate objects on the interface to improve visibility and readability.

- Level of differentiation of new gestures to avoid confusion.

- Appropriate text size.

- Responsiveness of the application.

- Application responds according to user requests and not automatically or randomly.

- Appropriate distance between objects.

- Ability to enlarge and reduce the size of objects.

- Ease of use and access to common or important tasks.

The second phase of the study was based on some of the highest rated problems and the comments made by experts. Table 4 presents some of the problems identified, which are shown in descending order with respect to the number of times the problem has been identified by different evaluators.

Table 4 .

Usability problems identified by experts		
Id	**Information problems**	
	Problem	Quantity
P1	Movement of objects using two fingers.	13
P2	Size of the keyboard.	8
P3	Finding help for the user.	5
P4	Finding the toolbar associated with each object.	4
P5	Navigating using the fan menu.	3
P6	Navigating between nodes.	3
P7	Identifying the icon for linking to new pages	2
P8	Incomplete or inaccessible menus.	1

Table 5 .

Problems associated with activities.		
Id	**Activities to be evaluated**	
	Activity	Problem
1	Insert two images, change their position and size.	P1
2	Enter a text next to the images as a caption.	P2
3	Obtain the toolbar for an image.	P3, P4
5	Pop a menu and discover how it works.	P5
6	Build a hierarchy of linked workspaces.	P6, P7
8	Get the menu in different positions. Cause the menu to go off the screen.	P8

Results from the emotional evaluation phase

In this phase, users evaluated whether or not they felt various emotions. Users experienced a variety of emotions when executing the identified problems. The results are presented graphically in Fig 3.
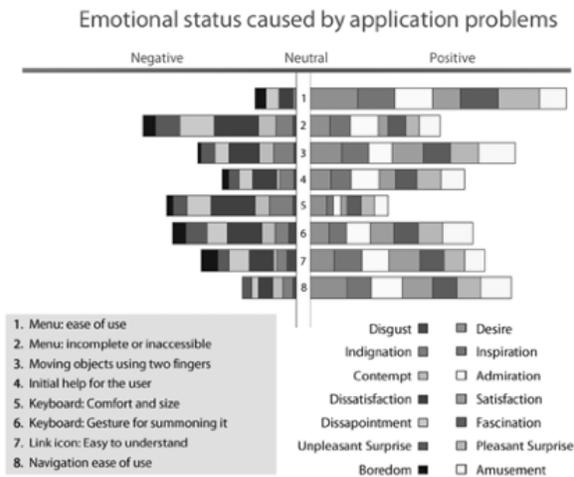
Figure 3 . Evaluation result of emotions.

As seen in the figure, there is a clear tendency towards the positive side even though all the activities were designed to recreate usability problems. It is interesting to note that all positive feelings appeared in all activities and this is what makes the general trend to move towards the positive side. Users had mixed feelings when performing problematic tasks.

Discussion

Activities leading to highly negative emotions in users were related to these tasks:

- Interaction with the menu in various locations over the table to evaluate emotions when the menu is incomplete or partly outside the display area.

- Interaction with the keyboard to evaluate comfort and size.

- Gesture to display and hide the keyboard.

In these three activities, users were highly dissatisfied and disappointed mainly because they could not interact naturally with the application. Keyboard handling is one of the activities that caused highly negative emotions. The least negativity was related to:

- Interaction with the menu to evaluate ease of use and navigation (see menu in fig. 4).

- Simplicity of navigation between workspaces.

- Initial help provided to the user.

Interestingly, the activity *interaction with the menu to evaluate ease of use and navigation*, resulted in a high level of positive emotion experienced by end users. Menu navigation was evaluated as problematic by experts during phase 1, but users did not generally rated it negatively and even enjoyed navigating and found it a pleasant surprise to use the fan menu. Users were highly inspired and wanted to continue using this menu.
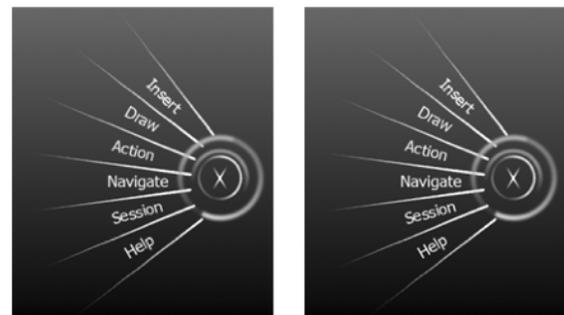

Figure 4 . InnovIMMS fan menu. Left. Main menu with several submenus available Right: Submenu Insert. The option to close is located in the center.

On the other hand, a high level of emotional negativity occurred because users were not able to use the keyboard easily (fig. 5). This was evaluated in the activity *interaction with the keyboard to evaluate comfort and size*, and was rated with a low positive emotion. This case confirms the experts' opinion about the keyboard. It should be improved significantly and with a high priority.

The fan menu and its navigation mechanism produce the greatest amount of positive emotions in the user, followed by navigation between pages and the gesture for moving objects with two fingers. The fan menu already had received positive ratings from users in previous studies [11].

The mechanism for link navigation received good reviews from the experts but also suggestions to improve its response time. Users had few negative comments about navigation, but there were comments that confirm expert opinions about the need to make it faster and more straightforward. Currently, navigation to parent nodes requires at least three touches on the surface (popping up the fan menu, selecting navigation, and then navigating back). Experts and users confirmed the need to reduce this to a mechanism that requires a single touch. There were also suggestions to use a special gesture to perform the actions inspired by existing commercial devices.
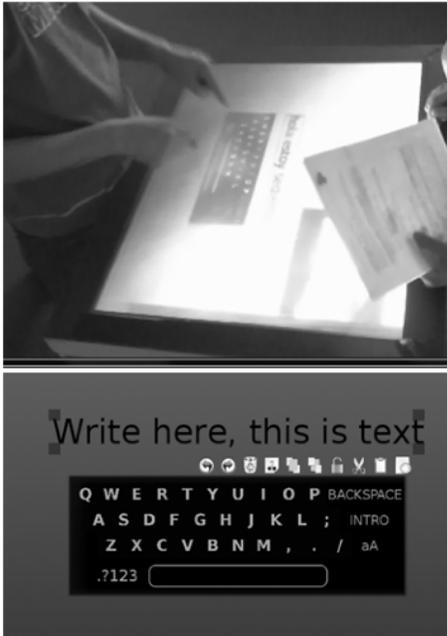
Figure 5 . Above: user manipulating a keyboard for entering text. Below: screenshot of the keyboard and editable text.

The gesture to move images with two fingers (see fig 6) had a low acceptance in studies that do not involve emotions. Experts and users of previous studies [12] and the current study report that it is preferable to move objects using only one finger.

Figure 6 . User manipulating an object

Given this, the high level of positive emotion that users expressed is interesting and unexpected. This could be explained if we consider that users had not previously used a horizontal multi-touch surface of the size used during the study. More than one user clearly showed fascination and amusement when discovering that the application allowed resizing images. After handling and moving images, users were prompted to evaluate the action. There were negative

comments but in general the activity produced positive emotions in the user. There were more negative emotions when the user moved the text, but this activity was not evaluated.

The data show a clear correlation in terms of negative emotions on the problems noted by experts. Somewhat surprisingly, a high range of positive emotions and negative emotions occurred simultaneously. A possible explanation of this is the enthusiasm of users to continue working with a new system and tolerance to frustration in the early stages of learning to use a new toy.

There is a unique case is which users had a high positive score when moving with two fingers. It's special because they are several reasons to believe that positive emotions were not expected: (i) experts clearly detailed problems with this gesture, (ii) videos and user observation indicate it was a difficult task to perform, (iii) comments from users indicate that this gesture is easily confused with the gesture to increase and decrease the size and (iv) users suggested that movement of objects should be made with the usual gesture: dragging with a single finger, as most existing commercial devices do. Despite this, the gesture earned high scores on positive emotions. One way to explain it is because the technology was new and is known for past studies that the ability to scale and move objects with your fingers is attractive to new users. Further information is required in this case to give a final conclusion.

CONCLUSIONS

One important implication from our work at this point is the potential benefit of coupling emotional evaluation techniques with the more conventional heuristic approach. Even though more work is needed in order to obtain more conclusive evidence, our initial results show a significant correlation between the issues raised by interaction experts and the emotions expressed by end users. However, some of the issues raised by experts have not necessarily led to negative emotions from end users. Even after discounting the effect of technological novelty in our experimental setting, it is clear that users are willing to accept or adapt to some of the interface features deemed problematic by expert users and rapidly overcome any negative emotions towards those features. By systematically following heuristic evaluations by an abridged emotion-oriented study, results can be validated and more accurately reported, thus addressing the lack of user involvement in heuristic approaches to usability evaluation. From a methodological standpoint, more applications of coupled heuristic-emotional eval-

uations are needed in order to determine the strengths and opportunities for improving this extended approach.

ACKNOWLEDGMENT

REFERENCES

[1] Obrist, M., Law, E., Väänänen-Vainio-Mattila, K., Roto, V., Vermeeren, A., Kuutti, K. (2011). Ux research: what theoretical roots do we build on - if any?. *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems,* pp.165-168.

[2] Hout, M. V. (2008). Comprendiendo, midiendo, diseñando (para la) emoción, *Faz,* pp. 88-97.

[3] Capota, K., van Hout, M., van der Geest, T. (2007). Measuring the emotional impact of websites: a study on combining a dimensional and discrete emotion approach in measuring visual appeal of university websites, *Proceedings of the 2007 conference on Designing pleasurable products and interfaces, ser. DPPI '07,* pp.135-147.

[4] Desmet, P. (2002). *Measuring emotion: development and application of an instrument to measure emotional responses to products.* Kluwer Academic Publishers, pp. 111-123.

[5] de Lera, E. and Garreta, M. (2008). *10 heurísticos emocionales: pautas para evaluar la dimensión afectiva de los usuarios de forma fácil y económica,* pp. 60-67.

[6] Lindgaard, G., Fernandes, G., Dudek, C., Brown, J. (1998). Attention web designers: You have 50 milliseconds to make a good first impression!. *Behaviour & Information Technology,* 25(2):115-126.

[7] Jordan, P. W. (1998). Human factors for pleasure in product use. *Applied Ergonomics,* 29(1):25-33.

[8] Tractinsky, N. (1997). Aesthetics and apparent usability: empirically assessing cultural and methodological issues. *Proceedings of the SIGCHI conference on Human factors in computing systems, ser. CHI '97,* pp. 115-122.

[9] Tractinsky, N. (2000). What is beautiful is usable. *Interacting with Computers,* 13(2):127-145.

[10] Spillers, F. (2003). Emotion as a cognitive artifact and the design implications for products that are perceived as pleasurable. *Cognition,* 7(6):1-14.

[11] Magallanes, Y. (2011). *InovIMM:Modelo de Interfaces para Innovación Usando Superficies Multitáctiles.* Master's thesis, Universidad de Las Américas Puebla.

[12] Magallanes, Y., Molina, A., Sánchez, J. A. (2012). Towards an optimal combination of gestures and graphical elements for collaboration using multitouch surfaces. Accepted in *Proc. CONIELECOMP 2012.*