Evaluating Both Usability and Desirability in the Evaluation of Cell Phones: Towards an Integrated Model

Laura Moody and Joan Burtner
Mercer University
Macon, Georgia

The goal of human-centered product design is to create products that are useful, usable, and desirable. For many years, human factors has been about research, modeling, and evaluating the functional usefulness and usability of products and systems. In recent years, however, the role of emotion in product design has become a 'hot topic' for human factors professionals, and a variety of methods have been developed for identifying emotional needs and the emotional response people have to products. In this paper, a preliminary study aimed at developing an integrated understanding of the functional and emotional aspects of user experience with cell phones is discussed. A variety of techniques were employed and the combined data analyzed to develop a holistic picture of the user experience. We anticipate that the results of the study will serve as 'proof of concept' for future development work.

INTRODUCTION

At its roots, traditional human factors is about functionality; in the early years, research and design approaches based on engineering and applied psychology focused on critical aspects of safety, efficiency and effectiveness of complex systems of humans and machines. More recently, the increasing power and ubiquity of computers, and the attendant capability to create a seemingly endless variety of experience, has expanded the role of human factors to include research into larger and more complex groupings of humans and machines. Still, the primary focus has been on functionality - understanding the needs, capabilities, and activities of the various actors; defining design requirements on the basis of robust models of action and interaction; and evaluating the effectiveness of machine and systems design on the basis of functional measures (performance, workload, etc.)

With respect to consumer products, usability methods have developed to address the specific need to understand how ordinary consumers relate to products, software, documentation, web sites, etc. Utilizing methods from design and consumer research, as well as traditional human factors measures of performance and ease of use, usability professionals seek to define and evaluate consumer interaction with products and technology. Although user research and usability studies often include specific attempts to gather consumers' feedback with respect to their emotional response to the product, this feedback tends to take second place to measures of ease of use or performance. Talbot (2000), in evaluating the differences in design process between designers trained in user-centered design techniques and those who focused primarily on aesthetics, noted differences in approach and uniqueness of solutions but could not find specific differences in the relative usability of the resulting designs.

More recently, a variety of techniques have come to the forefront for understanding people's emotional response to products (see, for example, Jordan, 2000; Lavie &Tractinsky,2004; Nagamichi, 2002). Methods such as ethnography, projective techniques, story-telling, product personality profiling, and kansei engineering have all been used to develop and exploit deeper understanding of human needs and emotional response to design. Scenario-based design and the development of user personas can also include specific reference to deeper emotional needs.

An example of a research methodology developed specifically to explore the quality of user interaction with products is the SEnsorial QUality Assessment Method (SEQUAM) (Bonapace, 2002.) It is unique in that it directly links the objective, physical properties of a product (size, configuration, materials, weight, etc.) with users' subjective reaction to the product. SEQUAM has been used to identify specific physical attributes of products that affect users' perceptions and to directly drive product design decisions.

SEQUAM illustrates how an evaluation of user subjective reaction integrated with an examination of specific physical product characteristics can be used to drive design decisions. The purpose of the investigation under discussion here is to explore means of integrating users' subjective evaluation of product characteristics and traditional usability measures to develop deeper understanding of user interaction with products. The product under consideration is the cell phone, both because of its ubiquitous nature in modern society and because a cell phone user's experience with the product is a function both of its usability and deeper emotional aspects (Ling, 2004). The ultimate goal of the research for which this study represents a first step is to develop a comprehensive model of the user experience with cell phones. Such a model would identify the relationships among all the critical aspects of the user experience, including not only the aesthetic and performance, but also such factors as quality and reliability, as well as special features and capabilities. Eventually, such a model could then be used to develop specific design criteria.

PROCEDURE

A variety of tools can be used to investigate user needs, tasks, environments, responses, etc. These tools include 'traditional' human factors techniques (psychology, anthropometry, etc.), as well as tools and techniques borrowed from fields such as anthropology and design, among others. To investigate 'desirability', methods are borrowed from design research, personality studies, etc., and new methods are developed to answer specific questions. To develop an integrated understanding of what makes a product 'useful, usable, and desirable' will require a combination of methods and analytical tools.

The initial investigation discussed in this paper focuses on using a combination of methods and analytical tools to develop an integrated understanding of what makes a product both 'usable' and 'desirable'. The study involved specific physical characteristics of cellular phones and involves a combination of product personality and traditional usability measures. The study was conducted using students in two undergraduate industrial engineering courses. All participants owned and/or use a cell phone frequently or regularly. Each participated in an experimental session lasting approximately 1 - 1.5 hours. In addition, several of the students observed the study and each of the classes analyzed a portion of the data as a class assignment. The purpose of this is to provide undergraduate industrial engineering students with an in-depth exposure to the conduct of human factors research.

The cell phones chosen for this study are shown in figure 1 below. The phones were similar in configuration, and differed in size, number and configuration of keys, and the feel and feedback from the keys.

Figure 1. Cell phones used in the study

Nineteen students participated in the study. Each experimental session consisted of three phases, as described below. Specific measures taken at each session are also listed.

Phase 1: "Storytelling"

The purpose of this phase was to understand the participant's personal experience with cell phones and to make the participant comfortable with the process. The conversation began with a discussion of participant's experience with cell phones. The participant was then asked to "tell a story" about cell phone use, provide three words or short phrases to describe his or her current phone, and provide three words to describe his or her "ideal" cell phone.

Phase 2: Semantic Difference

In Phase 2, participants were shown the three cell phones and given time to interact freely with them. Following this, the participants completed a semantic differential questionnaire designed to elicit first impressions of each phone's 'personality'.

Phase 3: Ease of Use

In phase three, the effect of the phone design and configuration on performance was evaluated. Each participant was asked to dial a series of numbers, first while looking at the phone and then while not looking. The numbers presented to the participants were classified as either 'simple' or 'difficult' to dial, depending on whether subsequent numbers were close together on the keypad or far apart. Typical performance measures, i.e., number of errors and time to dial the number, were recorded. The order of phones used was balanced across subjects.

Prior to each phone's *test*, each participant was allowed a practice session with the phone. During the practice session, the participant was encouraged to think aloud and comment on the feel of the phone, the layout of the keypad, and any other aspect of the experience. These comments were recorded for further analysis.

Final Review and Wrap-Up

In the wrap-up phase, participants were asked to provide a three word description of each of the phones in the study. They were also asked to select which phone they would be most likely and least likely to purchase. Finally, any remaining questions and comments were addressed before the participants left.

RESULTS

The evaluation of the results of the experimental session was conducted in two layers. The first layer involved an analysis of individual measures to identify key responses and usability issues. This analysis was followed by a 'meta-analysis' designed to integrate the results.

Note that, due to changes in the data collection procedure, the results from the first participant could not be used and the following analysis is based on 18 participants.

Analysis of Individual Measures

Phase 1. The stories relayed by participants in phase 1 centered around the "usefulness" aspect of the product. That is, participants tended to recall incidents in which they found themselves in a dangerous or uncomfortable situation and their cell phone was either helpful to them or failed when they needed it.

The descriptors provided by the participants for their current phone and their ideal phone indicated several striking similarities that may provide insight into the priorities of the study participants. The majority of responses with respect to

both current and ideal cell phones dealt with the aesthetic aspects of the phone (24 of 54 and 18 of 56 responses, respectively.) The most often used aesthetic descriptors centered around the size or style of the phone, with the ideal phone described as "compact" and "stylish". Other descriptors focused on ease of use, as well as versatility, reliability, quality of service, and (to a lesser extent) features such as internet capability and additional functions (pictures, voice dial, etc.)

Phase 2. While the number of participants was too small for a statistical factor analysis, visual inspection of the semantic differential questionnaire results indicate clear differences among the phones, with the most notable difference being between phone C and the other two phones. Phone A, for example, was most often found to be "heavy", "sturdy", "plain", and "uncomfortable". Phone C, on the other hand, was most often found to be "light", "compact", "stylish", and "comfortable". Similar results were found for other concept pairs.

Figure 2 provides a visual representation of the results of the semantic differential. In the graph below, the location of each bubble indicates the mean (centered) score for each phone on each concept pair on the semantic differential scale, while the size of the bubble reflects the standard deviation of the score. Visual inspection indicates that the cell phones in this study appear to be differentiated by the following concept pairs: light/heavy, compact/expansive, affordable/expensive, easy to use/difficult, simple/complex, powerful/weak, secure/vulnerable, innovative/traditional, and classic/trendy.
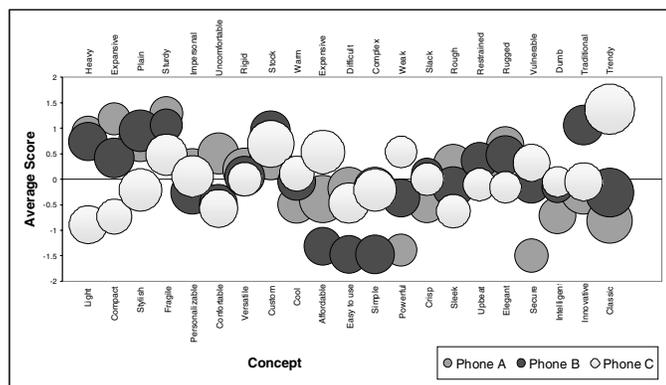
Figure 2. Average Centered Semantic Differential Responses

Phase 3. Analysis of the performance measures also indicates a difference among the phones when respondents were not looking at the phone while dialing. Analysis of the number of errors show a significant difference in phones (F = 3.17, p < 0.05) and dialing difficulty of the numbers (F = 14.36, p < 0.01). Interaction effects were not significant (see Table 1.) As shown in Figure 3, phone C resulted in significantly more errors than either A or B.

Table 1. GLM Results of Errors vs Phone, Number (not Looking)

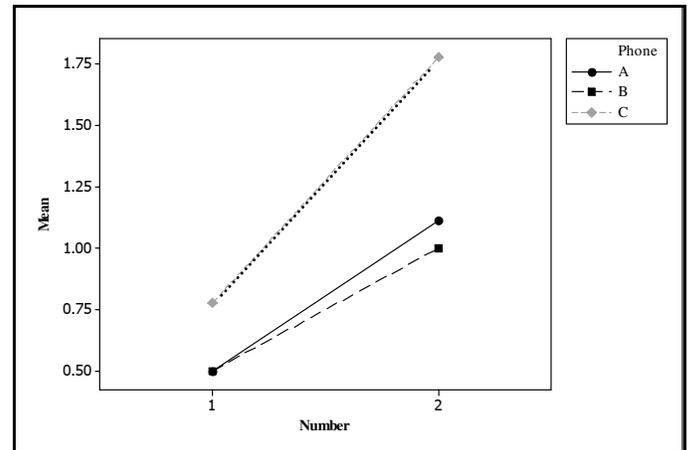| Source	DF	Seq SS	Adj SS	Adj MS	F	P
Phone	2	6.0556	6.0556	3.0278	3.25	0.043
Number	1	13.3704	13.3704	13.3704	14.36	0
Phone*Number	2	1.2407	1.2407	0.6204	0.67	0.516
Error	102	95	95	0.9314		
Total	107	115.667				

Figure 3. Effect of Phone and Number (1 = "easy to dial", 2 = "difficult") on Errors (Not Looking)

Due to the variability in individual dialing times, analysis of the time required to dial each number showed no clear indication of an effect of phone type on time to dial numbers in either the "looking" or "not looking" condition. This individual variation may be explained by the differences in experience and comfort level among the participants when it came to dialing cell phones.

Finally, analysis of the descriptors provided by participants for each phone indicate clear differences among the phones. In general, there was a clear difference in the nature of the descriptors, with more negative descriptors of all types attributed to one phone (A), more positive to a second (B), and a mixed result for the third. The majority of descriptors for phone A focused on its size and bulk: "big," "bulky," and "awkward" were commonly used negative descriptors, while positive descriptors included "rugged" and "easy to feel the buttons." Positive descriptors given for phone B focused on its simplicity and relative ease of use, while the negative comments included "out of date" and "dull." Phone C results were more polarized, with some participants referring to the phone as "stylish" and "cool", while others described it as "cheap." Phone C also resulted in the most mixed results, with an almost equal number calling it "hard to use" as "easy to use," and with some calling it "comfortable" while others described it as "cramped."

Despite the result of these descriptors, however, a clear majority (10 out of 18) selected phone C as their first choice out of the three phones, with six respondents voting for B and only two for A. Conversely, 12 out of 18 participants chose phone A as their last choice, which seems in line with the results of the descriptor analysis.

Meta-Analysis

Inspection of the data collected for the individual measures appear to be consistent with respect to participant's initial perceptions, performance, and final evaluation of each cell phone. For example, participants scored cell phones A and C as relatively less easy to use than B on the semantic differential questionnaire. The score for cell phone C is consistent with the higher error rates using that phone to dial without looking, as well as with some participants' description of the cell phone as "cramped" and "difficult to use."

The data also seem to indicate (to a certain extent) the relative impact of aesthetics and performance issues on users' preferences for a particular phone. In particular, although cell phone C resulted in poorer dialing performance and was described by some participants as "difficult to use" and "cramped," it was preferred by a majority of participants, including two whose description included "hard to use." This apparent contradiction is probably best explained by one participant who, when selecting cell phone C as the first choice, said, "you don't have to be fast when dialing the phone."

Formal analyses of correlations among the data yielded no statistically significant relationships among the various measures used in this study.

CONCLUSIONS

Initial analysis of the individual measures taken in this study indicate that the results support expectations, especially with respect to the effect of cell phone size and keypad configuration on dialing errors, as well as the patterns found in the semantic differential data. However, the size and scope of this initial investigation were insufficient to determine specific relationships among measures of emotional response and performance.

In addition, this study was designed to ignore several critical factors that are known to have significant impact on user satisfaction and performance using cell phones, including service quality ("number of bars"), interaction design (menus, features and how they are accessed, etc.), additional features (picture phone, PDA capability, etc.), and other factors. Inclusion of those factors is critical to developing a complete model of user experience with cell phones.

The study described in this paper is intended to be a preliminary 'starting point' for a larger investigation into the integration of functional and emotional aspects of human-product interaction. To develop a truly integrated understanding of what makes a product 'useful, usable, and desirable' will require a combination of methods and analytical tools. The researchers anticipate that the study will represent a start in that direction and provide some guidance for future research. Future work will include refining and modifying the measures used here, exploring additional measures, expanding the participant pool to include a wider population, and expanding the model to include all aspects of the user experience.

REFERENCES

Bonapace, L. (2002). Linking product properties to pleasure: The sensorial quality assessment method - SEQUAM. In Green, W.S. & Jordan, P.W. (Eds.) *Pleasure with products: Beyond usability* (pp. 189-217). London: Taylor & Francis.

Jordan, P.W. (2000). *Designing pleasurable products: An introduction to the new human factors.* London: Taylor & Francis.

Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies, 60,* 269-298.

Ling, R. (2004). *Mobile connections: The cell phone's impact on society.* San Francisco: Morgan Kaufman.

Nagamichi, M. (2002). Kansei engineering as a powerful consumer-oriented technology for product development. *Applied Ergonomics, 33,* 289-294.

Talbot, J. (2000). Style, function, interaction: Do designers disregard user-product interaction when proposing product design concepts?