Comparative Analysis of High- and Low-fidelity Prototypes for More Valid Usability Evaluations of Mobile Devices

Youn-kyung Lim, Apurva Pangam, Subashini Periyasami, Shweta Aneja
School of Informatics, Indiana University
901 E. 10th St., Bloomington, IN 47401
{younlim, apangam, speriyas, saneja}@indiana.edu

ABSTRACT
Validation of low-fidelity prototyping test results is difficult because we cannot claim whether the results are the effect of the prototype itself or the essence of the design concept we try to evaluate. However, it will cost too much if we implement a fully functional prototype for more valid evaluation. In this research, we provide a *qualitative and reflective analysis* of usability evaluations of a text messaging functionality of a mobile phone by comparing three types of prototyping techniques—*paper-based and computer-based* and *fully functional prototype*. This analysis led us to realize how significantly the unique characteristics of each different prototype affect the usability evaluation in different ways. We identify what characteristics of each prototype causes the differences in finding usability problems, and then suggest key considerations for designing more valid low-fidelity prototypes based on this analysis.

Author Keywords
Prototyping, designing mobile devices, usability evaluation.

ACM Classification Keywords
H.5.2 [User Interfaces]: *User-centered design*. D.2.10 [Design]: *Methodologies*.

INTRODUCTION
It is a strong consensus that prototyping is an essential part of evaluating design ideas. However, there is no stabilized consensus on how to validate the evaluation results from prototyping tests. When we design systems that embed newly emerging technologies such as mobile and ubiquitous computing, there are even more variables to consider, and it makes the evaluation process more complicated and unpredictable. Especially if we use low-fidelity prototyping techniques, it becomes harder to claim whether the evaluation findings are originated by the actual concept of the system or by the innate characteristics of the prototype. Some researchers also claim the necessity of high-fidelity prototypes for ubiquitous computing applications because low-fidelity approaches like paper prototyping are insufficient to precisely capture interactivity [12]. However, it is impractical to implement a fully working product for evaluating emerging technologies especially when we cannot decide what architecture or platform will be used for implementation [6].

These contrasting debates are tightly related to the issue of validity of prototyping. Especially for the evaluation of usability aspects of a design, it is difficult to claim the scientific validity of the test results. In this research, we are particularly interested in examining how different types of prototypes affect the evaluation results differently. If we do not carefully design prototypes or utilize them appropriately in the first place, the analysis of the evaluation results will not be meaningful or valid enough, no matter how scientific the tools of analysis are used. Reilly et al. [17] also claim the importance of determining appropriate levels of fidelity and granularity in a prototype before its development and evaluation. We claim that the validation of prototype-based evaluation results is not possible if we use a wrong prototyping technique *apropos* to the evaluation goal. It is well-known that the purpose of prototyping should be defined first as a basis for evaluation [2, 20].

Then how is it possible to select the right prototyping technique from the beginning that helps us evaluate a product in a valid but cost-effective way? In order to help us to answer this question, we devised a study to compare usability test results for an existing mobile phone using three different prototyping techniques so that we can figure out how the evaluation results yielded by each different prototype are different from one another. The three prototypes are namely, a fully-functional prototype which is basically a real working product, a computer-based low-fidelity prototype, and a paper-based low-fidelity prototype. We evaluated the same feature with the same design across the three different prototyping techniques, but retained the most important characteristics of each prototyping technique. For example, the interface elements are

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

represented in different ways because the paper-based prototype is made with hand-drawn interface elements unlike the actual phone with the colorful and digital representations, although the designs of the interface layout and the input elements are kept same. It is critical to keep the original characteristics of the prototyping technique (which in the case of paper prototyping is ‘paper-based’ and ‘easy-to-make’) in order to precisely examine the various roles and effects of different prototyping techniques.

Instead of designing a new product, we evaluated an existing product for this research, because we can use a perfectly working product which is the most reliable high-fidelity prototype. For this purpose, we selected the Samsung VI660 mobile phone, and tested the text-messaging feature. Because we were particularly interested in comparing different prototyping techniques and their effects on usability evaluation, we needed to select a product that is not too complicated but has critical aspects of mobile computing products. Instead of testing pervasively used features like telephoning a person using the phone, we selected the text-messaging feature which is not too new but relatively not very often used in the US culture except teenagers and undergraduate college students who were not selected as participants in our study. In addition, the task for sending a message requires a lot of key-presses compared to other tasks. We believe this task helps us get more valid responses regarding users’ feelings on the handling and operation of the mobile device.

Our research helped us identify different types of usability problems that are revealed by each prototyping technique. We realized that each prototyping technique is not equally useful for finding out all types of usability problems. This investigation is different from other related previous approaches that have mostly focused on quantitative issues of evaluation, such as which technique finds more usability problems or lead more of users’ suggestions about the design across different prototyping techniques. Unlike these approaches, our research emphasizes more on a rigorous *qualitative analysis* of usability evaluation results from each evaluation session. By *qualitative analysis*, we mean which types of usability problems are identified by which prototyping technique, how users’ behavioral patterns are different toward the interaction with each prototype, and how facilitator’s attitudes are different toward each session. Based on this analysis, we identify what characteristics of the prototyping technique resulted in those differences. In addition to this, we identified the problems caused by the characteristics of the low-fidelity prototyping techniques, not by the product design itself.

Our approach is not to criticize any method, but to figure out optimal parts of each technique, and provide better directions to use low-fidelity prototyping techniques. Our key contribution in this regard is to provide critical considerations in designing a low-fidelity prototype for evaluating mobile devices (e.g. mobile phones) in order to get more reliable results.

BACKGROUND

Prototypes are primarily used for the following purposes [3, 9, 20, 23]: communication, exploration and refinement, and evaluation of design ideas. In this research, we are particularly interested in the purpose of evaluation, especially related to usability and user experience aspects of the design.

Many have discussed about validity of usability studies [4, 5]. There have been discussions on the validity of discount usability methods in terms of the number of users to test, the length of observation, in-situ vs. in-lab tests, and profiles of users.

Another important issue in prototyping is the cost. People have emphasized the benefits of using low-fidelity prototyping techniques in this regard. There are many research examples on low-fidelity prototyping techniques. These techniques include paper prototyping for any type of interactive products which may include computer-based applications, mobile devices, and websites [7, 18, 22]; Switcharoo for physical interactive products [2]; Calder for physical interfaces [8, 11]; Buck prototyping for mobile devices [16]; a rapid prototyping for mobile devices using augmented reality technology [14]; and DART for augmented reality systems [13]. Their work showed benefits of low-fidelity approaches for various types of interactive products. However, no one actually claimed any validity of its use for usability evaluation, nor specified how to validate it. All of the approaches focus more on design exploration and communication. If any approach was used for usability evaluation, it was done in an *ad hoc* way.

Many researchers also have emphasized the advantages and disadvantages of using different prototyping techniques [2, 9, 12, 16, 19, 21, 23, 24]. However, most of the comparisons among the different prototyping techniques are based on anecdotal experiences or assumptions.

Of course, some of the examples were more rigorously conducted, which include [12, 21, 24]. Sefelin et al. [21] examined whether or not the level of users’ willingness on criticizing or making suggestions to the design will be different between paper-based and computer-based low-fidelity prototyping. Virzi et al. [24] claimed that the sensitivity to find usability problems does not differ between low- and high-fidelity prototyping. However, their systems for the evaluation were standardized GUI-based systems which differ from mobile or ubiquitous computing systems, and they did not clarify what types of problems are identified by which type of prototyping technique. Liu and Khooshabeh who studied prototyping techniques for ubiquitous computing environments claimed that it is critical to carefully choose the fidelity and automation level of the prototypes for evaluation [12].

However, none of them rigorously studied about whether each different prototyping technique can reveal the same kinds of usability problems, or if not, what aspects of the design concept and usability problems can be better evaluated by specific prototyping techniques.

It is important because the results from this research help us make decisions on what prototyping techniques we may use when we have limited resources. In addition to this, we emphasize the importance of triangulation of different methods. Using one prototyping technique cannot discover all possible problems [16]. In order to effectively do the triangulation, we must know how they are different from each other in terms of their purposes and roles in design. Our research here addresses this issue as well.

STUDY DESIGN

Our study is to compare the fully functional prototype with two major low-fidelity prototyping techniques, namely, paper-based and computer-based interactive ones. We selected these two low-fidelity prototyping approaches because of the following three reasons: (1) they are most pervasively accepted and used in the real world design practice, (2) they help us examine the effect of representing the physical aspect of a product design with a 2-dimensional form or a roughly represented form, and (3) they are the cheapest approaches to low-fidelity prototyping that are currently available. For the fully functional prototype, we selected a real working mobile phone because it most precisely represents the actual design concept.

The goal is to figure out how the unique characteristics of each prototyping technique affect the usability evaluation differently. The study results led us to discuss how to design the low-fidelity prototypes that help designers do a more valid evaluation of their design.

In this research, the design of the two low-fidelity prototypes was very critical. Each prototype needed to have the same interface layout, the same interface elements, and the same feedback that were already implemented in the actual product so that we can precisely compare the test results across these different prototypes. We also tried to keep the most significant characteristics of each technique that are defined and pervasively accepted by the HCI community [10, 22].

Our study design is very different from the conventional experiment design which normally requires statistical analysis, because we are not trying to select the best prototyping technique out of these three techniques but to identify differences across these three techniques through a rigorous *qualitative analysis* of the results from each session by looking at the details of the types of usability problems identified from the observation and the debriefing interview conducted in each session.

We recruited 5 participants for each type of evaluation, so a total of 15 participants were observed. Each participant was allowed to participate in only one test because a possible learning effect across the different conditions may skew usability problem findings. In our study, the coverage rate of usability problems that can be identified by each technique is the most important factor to consider in terms of the validity of our study. Based on Nielsen's work, "testing five users can find more than 75% of usability problems" [15]. Having more than five users per each evaluation session may still be helpful if we try to cover almost 100% of the problems that can be realized by each technique. However, there is still much debate in this issue about whether we can really cover almost 100% of the problems or not, if we have about 15 participants for usability evaluation [4]. In this regard, it is not really critical for us to struggle with this issue, but more critical to design the study that can be effectively comparable for qualitatively analyzing the results from the evaluation sessions. This discussion is different from the common experiment-based study design issues, which focus more on participant selection and the number of participants to determine its validity.

We had a mixture of participants having different levels of experience in using a mobile phone as well as text-messaging. We included people who do not own a mobile phone, who own a mobile phone, but do not use text-messaging, and who own a mobile phone and use text-messaging for every evaluation type. None of the participants who own a mobile phone own the same phone as the one we selected for the tests. The distribution of the males and females were also almost even for each condition.

There was one consideration in selecting participants for the paper-based tests. We made sure that no designers were elected as participants for this type of test because designers tend to know too much about the characteristics of paper prototyping and it is possible to affect their attitudes towards the use of the prototype differently from regular users.

During the test session, we asked all the participants to do the following task:

> *Let's assume that you need to send a text message to your friend using the given mobile phone prototype now.*
>
> *Please send the following message,*
>
> *"Hi, I am sending you this message from a mobile phone usability test. This test evaluates various prototyping methods."*
>
> *to your friend, John, whose contact information can be found in the phonebook.*

The text-messaging feature was selected for the usability evaluation because it is not a new feature but not very often used unlike telephoning feature as aforementioned. Because it requires enough time to manipulate the keypad, users can comment on the operation aspect of the product more precisely.

Each usability evaluation session consisted of two sub sessions like all other common usability evaluation sessions in the real world design practice: (1) the testing session where users conducted the task, and (2) the debriefing session where users answered questions designed to evaluate several key aspects of usability of the product (Figure 1). For design evaluation in general, different aspects of the product should be considered—usability, usefulness, desirability, performance, and style [2]. The evaluation of each aspect requires specific techniques. In our study, we evaluated especially the usability aspect of the product. By usability we mean a broader sense that not only includes the aspect of ease-of-use but also feeling-of-use, appearance of the product, and satisfaction. In order to do this, we incorporated questions for asking these aspects of the product during the debriefing session as shown in Figure 1. After participants fill out this questionnaire, the facilitator further asked them to elaborate the reasons for their answers in detail.

Please circle on the number that answers best on the questions.

1. **Was this prototype easy to use?**
1 2 3 4 5
Very difficult ... Very easy

2. **How do you feel about using this prototype?**
1 2 3 4 5
Not good ... Very good

3. **On the scale below, how would you rate the looks of this prototype?**
1 2 3 4 5
Not good ... Very good

4. **Overall, how would you rate your satisfaction of using this prototype?**
1 2 3 4 5
Not satisfied at all ... Very satisfied

Figure 1. The debriefing questions

In testing sessions, we asked participants to think aloud, and encouraged them to explore the options as many as possible when they had trouble figuring out how to use it. During the observation of the testing sessions, we took videos for detailed analysis.

Fully Functional Prototype

For the fully functional prototype, we used a Samsung VI660 mobile phone (Figure 2). The other low-fidelity prototypes were designed to represent this product.

Figure 2. The fully functional prototype (Samsung VI660)

Because it is an actual product that people use in an actual environment, we could assume that the results of the usability evaluation from this prototype are most realistic and valid. This helped us judge and analyze the results from other low-fidelity prototypes more effectively by comparing them with the results from this actual phone.

It is basically fully functional and interactive, no error exists, and the look-and-feel aspect of the product is real. We tested the usability of the text-messaging feature as mentioned. The layout structure, the interface elements, the feedback, and the general look-and-feel were carefully identified from this product. Their critical ideas were represented in the other low-fidelity prototypes we used for the comparative analysis.

Computer-based Low-fidelity Prototype

The computer-based low-fidelity prototype (Figure 3) was developed by using an emulator, Samsung JaUmi Wireless Toolkit 2.0. We intentionally selected the computer-based low-fidelity prototype that is completely based on the computer screen, among other related approaches such as Buck prototyping [16] or the technique presented by [14]. The reason for this is that we were interested in examining the effect on representing a 3 dimensional product by a 2 dimensional prototype. In order to reduce unexpected effects on using the prototype, we covered the keyboard of the computer with a sheet of white paper (Figure 3).

When a person rolls over the mouse cursor on a key of the mobile phone keypad on the screen, we set the key visually highlighted. For the screens that are not available, we showed an alert message, "This screen is not available" with an alert bell sound. We also tried to accommodate default functionality allowed by the emulator as much as possible, in order to keep the time for building the prototype as short as possible, without altering the actual phone design concept. Through this way, we can keep the benefits of low-fidelity prototyping. For example, we did not spend our effort on implementing the dictionary feature ('T9' feature) in this prototype because no one used the function while using the actual phone for the task we provided. One designer and one programmer spent about 20 hours building this prototype.

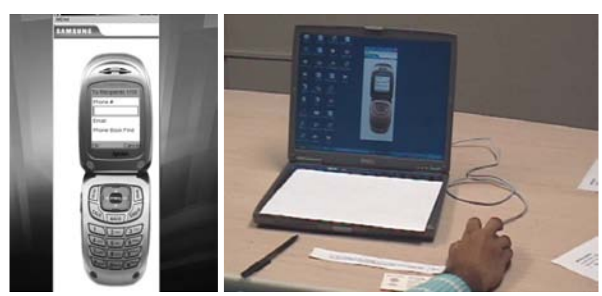


Figure 3. The computer-based prototype and its test setup

Paper-based Low-fidelity Prototype

When building the paper prototype, we also tried to keep the characteristics of paper-based prototyping as much as possible. All paper screens had wooden craft sticks glued to the non-screen side of the paper (Figure 4). This was useful for the facilitator to play the role of the computer. It helped hold the screens (by holding the sticks in the hand) and allowed for quick screen switching. The screens were hand drawn to maintain simplicity and low-fidelity characteristics of the prototype. A small strip of yellow cellophane was used during the test to show when options were highlighted when the user selected them. This provided better visual feedback to the users. The phone prototype, built with foam core, was 1.5 times the original size. This improved user-readability and was easier to work with. This prototype like the computer-based prototype, did not provide the T9 dictionary functionality because the different possibilities on typing the words were too large. Again, most users did not know, or had never used the dictionary feature. Three designers spent about 5 hours building this prototype.

Figure 4. The paper prototyping setup and its use situation

Table 1 shows our usability evaluation setups for each prototyping technique. Interface elements and layouts are very carefully designed to keep the characteristics of each low-fidelity prototyping technique as well as to maintain the testing conditions consistent throughout.

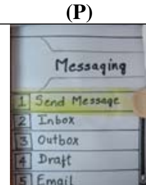
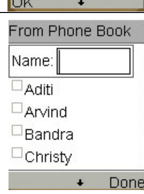
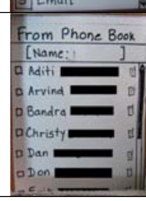
	Fully functional (F)	Computer-based (C)	Paper-based (P)
Main screen for messaging menus			
Screen for phonebook			
Number of participants	5	5	5
Number of facilitators	1	1	1
Number of observers	2	2	2
Given task	Send a text message to John whose number is in the phonebook.		
Tested feature	Text messaging feature		

Table 1. Screens and testing setups for the three techniques

RESULTS AND DISCUSSION

The analysis of the results is done in three steps.

1) We examined how many same usability problems and positives are identified by the different prototyping techniques, and how many different problems and positives are identified by those techniques (Figure 5). This step of the analysis was necessary in order to figure out how much difference in overall exists among the different prototyping techniques.
2) We then identified what types of usability issues and how many instances in each issue were found by each prototyping technique (Figure 6 and Table 2). This step of the analysis was for a deeper look of the difference among the techniques which helps us to figure out which techniques are good for evaluating which types of usability issues.
3) We also identified how different unique characteristics of each prototyping technique affected the attitudes and behaviors of the users, the facilitator, and the observers, who were the key members of a usability testing environment (Table 3). This step of the analysis was helpful for providing a guide for preparation and execution of the different techniques for more effective evaluation.

Based on this analysis, we will explain as a conclusion a set of guidelines to design more effective and valid low-fidelity prototypes for usability evaluation. This analysis will also help us decide how the prototype should be set up, what facilitator's roles are, and what observers should keep in mind when using a specific prototyping technique for evaluation.

For the analysis of the results, our focus is not on determining which technique is better than others, but more on figuring out what the pros and cons of each technique are and how we can accommodate them for better and more valid prototype design.

First, we identified usability *problems* and *positive aspects* of the product that users experienced. The *problems* were identified from both the observation of the testing sessions and the users' comments recorded in the debriefing sessions. The *positive aspects* were identified from the users' comments in the debriefing sessions.

We recorded the following actions as usability problems during the testing sessions:

1) Users push a wrong key (button);
2) Users explicitly express problems in their think-aloud protocols and gestures; and
3) Users repeat the action they already completed in order to check if they did right or not.

We did not measure time information for each action because the technique like paper prototyping is not appropriate for considering the time information. A major part of time consumption in paper prototyping sessions was caused by facilitator's selecting and displaying appropriate screens according to users' inputs.

All the problems were identified by four researchers' careful reviews of the videos taken during the testing sessions. The factors we captured from the testing sessions include usability problems encountered, users' general attitudes towards the prototype, facilitator's behavior, observer's behavior, and any interesting characteristics of the prototype encountered. The factors from the debriefing sessions include participant's answers on the questions, user's general attitudes towards the prototype while commenting on the design, and facilitator's behavior during debriefing.

We did not count the observed frequencies of each finding. A finding that is observed more frequently than others may imply that its significance-level is higher than others. However, in our research, this information is not necessary because our goal is not to redesign the phone for better usability but to figure out types of problems or positive aspects realized by specific prototyping techniques. In this regard, we identified only each distinctive instance instead of total number of encounters of each instance.

For the simpler denotation of the different prototyping techniques, we used the following convention: 'F' stands for *fully-functional prototype*, 'C' stands for *computer-based low-fidelity prototyping*, and 'P' stands for *paper-based low-fidelity prototyping*.

Overall Patterns of Problems and Positive Aspects of Usability Identified by the Three Prototyping Techniques

If many of the problems identified by one technique do not overlap with the ones identified by other techniques, we need to make sure how many of them were realized distinctively by each different prototyping technique, and what problems were identified by one technique but not the other.

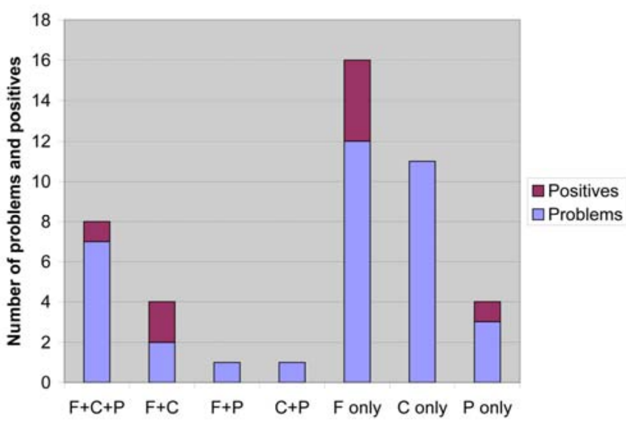


Figure 5. The distribution of number of problems and positive aspects identified by different sets of prototyping techniques

Figure 5 shows the distribution of the number of problems and positive aspects identified by all of the three different techniques (F+C+P), two of the techniques (F+C, F+P, or C+P), or each of them (F, C, or P). We can easily see that a large number of the problems are realized by the fully functional prototype and the computer-based one. And also there are some of the problems identified only by the paper-based one, but far fewer than the case for the other ones.

Detailed Patterns of Problems and Positive Aspects of Usability Identified by the Three Prototyping Techniques

This overall pattern helped us to understand that each different technique has a unique capability to find different problems and positive aspects which may not be easily found by other techniques. In order to figure out then which types of the problems and the positive aspects are found by which specific technique, we examined the individual instances of the problems and the positive aspects one by one, and categorized them into the key *types of usability issues*.

Ten major usability issues in designing a mobile phone application were identified as follows:

a. Unclear meanings of labels
b. Icon/symbol/graphical representation issues
c. Locating appropriate interface elements
d. Mental model mismatch—unexpected feedback of the product
e. Appearance/look of the product
f. Physical comfort in handling
g. Physical comfort in operation
h. Concept itself
i. Comparison with other similar products
j. Performance issues

Some examples of the actual instances for each type of issue are shown in Table 2.

Figure 6 shows how many instances of each type of usability issues were identified by each prototyping technique. The types *a* to *e* were identified by all the three types. However, the types *f* to *j* were not identified by the paper-based prototyping approach. Especially the types *f*, *h*, and *j*, namely, *physical comfort in handling*, *comments on the concept*—text messaging—*itself*, and *performance issues of the product*, were identified only by the fully functional prototype.

In this research, our goal is to look into the causes of these differences residing in the different prototyping techniques. The analysis shown in Figure 6 provided a starting point for us to determine where to examine in detail. Through the detailed examination, we could extract two important considerations for selecting and designing an appropriate prototyping technique for more valid evaluation:

1) understanding of the *erroneous characteristics of the*

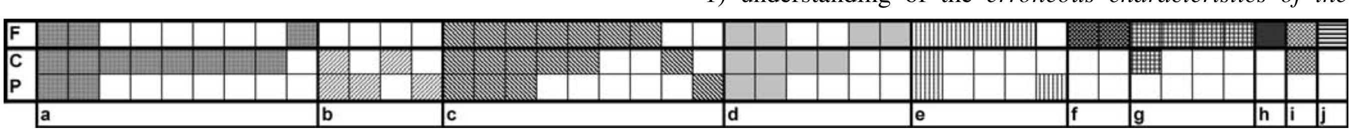


Figure 6. The distribution of the problem and positive instances related to different types of usability issues identified by each prototyping technique (F, C, and P). Each column represents each instance.

prototype, which conflict with the nature of the actual design idea; and,

2) identification of *valid aspects of the prototype*, which are analogous to the actual design idea.

Erroneous Characteristics of Prototypes

The erroneous characteristics of prototypes can be identified if we look at which problem or positive instances found by the low-fidelity prototypes (paper-based and computer-based) are not identified by the fully functional prototype (Figure 6).

We can see that the instances related to the issue of *unclear meanings of labels* (type *a*) were identified mostly by the computer-based one. From the analysis of actual instances of the problems related to this issue identified by the computer-based prototyping technique, we realized that people's preconception about GUI-based interfaces for computer-based applications influenced their misinterpretation of the interface elements such as labels represented by the prototype. For example, people were confused whether the labels are clickable or just labels. The ways of identifying the executable links on the mobile phone which does not allow people to touch the screen or move a mouse-cursor around, are different from the ways of identifying them on the computer-based applications. As you see the examples in Table 1, we tried to represent the labels and highlighting indicators as similar as possible across the different prototypes. However, many users who participated in the computer-based condition tried to click the interface elements on the screen part directly with the mouse cursor instead of clicking the keys on the keypad part. Some people were not sure whether or not the links with the labels on the computer-based one are clickable, although they are represented in the same way with the ones on the actual mobile phone. While typing in the text message using the computer-based prototype, one user tried to use the computer keyboard although it was covered by a blank paper. The facilitator asked the user to use the keys on the keypad part of the prototype instead of the actual keyboard when s/he tried to use it.

These instances tell us that it is important to carefully consider how to represent interface elements in different types of prototyping approaches, in order to minimize possible differences that may be caused by certain interface conventions of the prototyping approach. Some initial pilot studies to figure out which way is appropriate may be necessary. When we interpret the evaluation results from these tests, we should take into account these possible differences.

Another interestingly noticeable finding related to the *erroneous characteristics* was that the users mentioned about the physical operation experience also in the 2D-based computer-based prototyping session, as we see in the graph (Figure 6)—type g, *physical comfort in operation*. A user mentioned that the size of buttons is large enough for typing the given text message. The critical issue we must think about here is that the form of the physical operation for the computer-based one is completely different from the one for the fully functional prototype. When there is less difference in physical manipulation between an actual hardware product and a software-based prototype such as the cases for an ATM machine or a telephone installed in a fixed location, the difference in evaluation results may not be significant [1]. However, in the case of mobile devices, we should not overlook this issue. For example, the physical experience of mouse-clicking on the keypad shown on a large computer display for operating the text messaging is completely different from the physical experience of grabbing a small mobile phone for typing in the message by finger-inputs. In this regard, we cannot conclude whether or not the size of the buttons is actually large enough, because we cannot claim if it is the represented form on the large display that is manipulated by a mouse cursor, or it is the button size that is actually large enough even in a real situation. For example, a user who used the fully functional prototype mentioned that typing was difficult (Table 2). When we use a different platform from the actual product for cheaper and easier prototyping, we must consider this aspect of influence in evaluation.

Usability issues	Example instances
(a) Unclear meanings of labels	- The 'Phone #:' label was not clear whether is it for entering the telephone number or for entering person's name to access to the phonebook. (F+C+P) - The function of the key, '0/Next', was not clear to the users. (C)
(b) Icon/symbol/ graphical representation issues	- No one noticed that the OK button with the messaging icon allows accessing to the messaging function directly. (F+C+P) - Arrow graphics for browsing direction on a menu screen was not clear in terms of what they mean. (P)
(c) Locating appropriate interface elements	- Associating the right key for the right command execution was not clear. (F+C+P) - It was unclear how to enter the comma while typing the given text message. (F+C+P) - It was unclear where the switching button for lower- or upper-cases for messaging. (F+C)
(d) Mental model mismatch	- Users tried to access the phonebook first to send the message to 'John'. However, the prototype does not allow users to send messages directly from the phonebook list. They must access the messaging page first and then access the phonebook. (F+C+P)
(e) Appearance/ look of the product	- Users liked how it looks in general. Some users specifically liked the look of keypad. (F+C+P) - One user mentioned that the look of the prototype is not appealing to her/him. (F)
(f) Physical comfort in handling	- The phone itself was comfortable to grab. (F)
(g) Physical comfort in operation	- Typing was easy. (F) - Typing was difficult. (F) - The size of the buttons was large enough. (F+C)
(h) Concept itself	- Some users specifically mentioned that they do not like the idea of text messaging. (F)
(i) Comparison with other similar products	- Some users mentioned that the interface is similar to the ones for the standard phones. (F+C)
(j) Performance issues	- A user mentioned that the LCD display of the phone is bright enough so the letters were easy to read. (F)

Table 2. Example instances for each type of usability issues

We also found an example of the erroneous characteristics of prototypes from the paper-based prototyping technique. The number of findings in type *b*, *the icon/symbol/graphical representation issues,* identified by the paper prototyping technique was the largest one unlike the cases of other types. When we carefully examined the instances of the problems of this type identified from the paper prototype, the key cause was the abstractness and unfinished look of the paper prototype. Because the key characteristics of paper prototyping include quick-to-make and early-on use in design, it is not possible to have detailed and complete look of visual design when we are in the stage of using a paper prototype for evaluation.

The analysis of these causes guide us to decide how to design a specific prototype to capture more valid data which are caused not by the erroneous characteristics of the prototype but by the characteristics of the product design idea itself which is represented by the prototype. In addition, if such erroneous nature cannot be easily corrected in a design of the prototype, recognizing this factor will help designers being cautious and thoughtful on the interpretation of the data they can gather from the evaluation using the prototype.

Valid Aspects of Prototypes

The valid aspects of prototypes can be identified if we look at which instances found by the low-fidelity prototypes (paper-based and computer-based) are also found by the fully functional prototype (Figure 6).

Some of the findings from the paper prototyping were, however, still valid enough especially for types *c*, and *d*, namely, *locating appropriate interface inputs* and *mental model mismatch*, because many of the instances from these types were also realized by the fully functional prototype.

In the case of the computer-based prototyping, several problems which are related to detailed interface elements design in type *c*, *locating appropriate interface elements*, were also identified in the fully functional one. For example, from both of those techniques, we could identify that some of users could not figure out how to switch the modes of lower- or upper-cases for typing text messages (Table 2). The instance like this which can be realized only when we simulate the real text messaging typing experience, cannot be easily demonstrated through the paper prototyping technique.

This analysis tells us that some of the key usability issues can be evaluated using a certain prototype in a same valid way as if we evaluate the actual product. We can effectively prevent these types of usability problems early enough in the design process by applying these low-fidelity prototyping techniques.

Other Findings and Considerations

One interesting fact we found was that no one testing in the paper-based condition suggested new design ideas. However, the participants in other conditions—computer-based and fully functional ones—suggested some new functional ideas such as voice typing and button label changes. This was different from the pervasively accepted claim about paper prototyping approaches. Many have claimed that these approaches provide better environments for users to suggest more of new ideas than high-fidelity prototypes [19, 22]. We realized that we cannot generalize this situation to all cases. Based on how we design the prototype, what type of product we evaluate, what aspect of the product we evaluate, and how we facilitate the user in using the prototype, we will have different results on the level of user involvement in suggestions.

In this analysis, we can realize the importance of triangulation of different techniques for more valid and enriched evaluation process, as well as the importance of carefully planned application of appropriate techniques at the right phases in the design process.

Different Attitudes in Conducting Tests with the Three Different Prototyping Techniques

In addition to this analysis, we also observed carefully how each member in the testing sessions behaved differently in different prototyping situations. The members include users, facilitators, and observers. Table 3 shows the summary of our findings from this observation. When we think about a usability testing situation using a prototype, we can readily imagine a chain reaction of the prototype towards the attitudes of a user, a facilitator, and observers. What we mean by the chain reaction is that the unique characteristics of the prototype cause not only users' different styles of interactions with the prototype, but also facilitators' different behaviors in facilitation and observers' different difficulties or readiness in interpreting users' interactions and behaviors with the prototype. All of these differences directly or indirectly affect the evaluation results. In this regard, we must be clearly aware of the characteristics of the attitudes by users, facilitators, and observers, caused by different prototyping techniques, in order to interpret the evaluation results more precisely.

Users' attitudes

F	During testing - *Interactions with the prototype are real time*. - *The physical experience with the prototype is exactly same with the ones for the actual product*. - *Users can explore all possible options* to complete the task without worrying about the limitation of its functionality. During debriefing - *Users discuss the look in detail*. - *Users grab and use the prototype as a reference* to explain their thoughts and comments.
C	During testing - *Users are confused about non-working buttons* which made them feel doubtful about whether the situation that is not working is due to their incorrect actions or due to the prototype's errors (although there were no errors in the prototype). - *The posture to use the prototype* is not exactly same with the one for the actual phone. Some users need to bend over to see the screen better. - *Users have a strong preconception on how the conventional computer works.* Users try to click the interface elements directly on the display part of the prototype rather than using its keypad

	part. When entering text for messaging, some users even try to use the computer keyboard instead of the keypad graphics on the prototype. - *Users' preconception on the GUI conventions in computer-based applications* affected their interpretation of the interface shown on the prototype. Some label-based clickable links were not clear for users to understand them as clickable links. During debriefing - *Users discuss the look in detail*. - *Users point and use the prototype as a reference* to explain their thoughts and comments.
P	During testing - *Users trust a person more than a computer*. Incomplete sets of screens and limited functionality help users easily figure out whether they did right or wrong at the first time unlike the cases for the computer-based one. - *Users need to think aloud all the time* in order to clearly express what button they have pushed so that the facilitator knows what feedback needs to be given to the user. - *The long lag of the response time on each input* makes users sometimes feel afraid of pushing a wrong button or not wait till getting the feedbacks while typing texts. Due to this reason, we shortened the text message that they needed to type because it took a tremendous amount of the time to type all the texts we originally assigned to the users. During debriefing - *Users cannot judge the look precisely*. - *Users grab the prototype and use it as a reference* to explain their thoughts *without screens*.
Facilitator's attitudes	
F	- *A facilitator does not have to worry about technical errors* so that s/he can completely focus on users' behaviors.
C	- *A facilitator should be aware of the limited functionality of the prototype* in order to facilitate users correctly. S/he needs to know whether the result of the action made by the user is due to user's error or prototype's error.
P	- *A facilitator should be more careful to encourage a user to think aloud all the time*. - *A facilitator should fully understand and memorize how the actual product (or design) should work*.
Observer's attitudes	
F	- *It is relatively easy for an observer to read user's thinking process* because s/he can observe user's instant and real-time responses on interactions. - *It is necessary to adjust either user's position or camera's position* to capture the right scenes. - *Time information is valid*.
C	- *It is relatively easy for an observer to read user's thinking process* because s/he can observe user's instant and real-time responses on interactions. - *We can use screen-capture software for recording user's interactions*. - *Time information is valid*.
P	- *Without user's think-aloud verbalization, it is hard to understand what they think*. - *It is necessary to adjust either user's position or camera's position* to capture the right scenes.

Table 3. Users, facilitators, and observers' attitudes in the testing session when using each prototyping technique

CONCLUSION AND FUTURE STUDIES

In this research, we figured out that major usability issues such as *unclear meanings of labels*, *icon/symbol/graphical representation issues*, *locating appropriate interface elements*, *mental model mismatch*, and *appearance/look of the product*, were identified by all the three types of prototypes, namely, *the fully-functional prototype, the computer-based low-fidelity prototype,* and *the paper-based low-fidelity prototype*. However, some of the issues like *physical handling and operation*, *comments on the concept itself*, *comparison with other similar products*, and *performance-related issues*, were identified by either the fully-functional one or the computer-based one, which are both real-time interactive.

We also examined the details of the usability testing result instances in terms of *erroneous characteristics of prototypes* and *valid aspects of prototypes*. This helped us figuring out key limitations of each low-fidelity prototyping techniques for the evaluation as well as the useful parts of the prototyping techniques that help extracting valid evaluation results.

We realized several key limitations of the two low-fidelity prototyping techniques we studied. First for the computer-based one, the most significant limitation that affected the evaluation results was users' preconception of the conventional computer such as the GUI-based interface and the inputs like the computer keyboard. Another limitation related to this technique was the possible difference in physical experience. Because it is a representation of the 3-dimensional product onto the 2-dimensional space, the physical manipulation and operation experience was significantly different. However, an interesting finding was that many users mentioned during the debriefing session that they could imagine how this would be felt when it is implemented as a 3-dimensional form, just by looking at the image on the screen. There is no valid proof for correctness of the answers on this aspect of the product because the answers are based on their assumptions, not actual experiences. Even for these limitations, this prototyping technique helped us find critical and valid problems in most types of usability issues as shown in Figure 6. In this regard, if we consider these limitations carefully when designing this prototype, it can be used effectively for usability evaluation for mobile device design.

For the paper prototyping technique, the major limitation was on the abstractness and unclearness of the images that are represented in general. Another issue was the long lag of response time for each input as well as the critical need of think-aloud protocols. However, when we think of what it takes to make the paper prototype compared to other prototypes, it is still a valuable approach to evaluate usability of the product in terms of identifying possible problems in key types of usability issues that include the types *a* to *d*.

These major findings led us to claim that it is important to *determine what aspects of usability we need to evaluate before we build a low-fidelity prototype for the evaluation*. Our research results regarding different distributions of usability issues identified by different prototyping techniques (Figure 6 and Table 2) provide a useful reference for this decision.

In this research, we primarily focused on the usability—ease of use—aspect of the design. For the future research, we also like to further examine how people *feel* differently about the product if they use different prototypes. Although we asked this question during the debriefing session (Figure 1), we plan to do more rigorous research on this by dissecting the issues of the determinants of feeling-related experience, as we did for the usability issues in this research.

Another research question that has not been answered in our research is how to determine the effect on evaluation by limited functionality of low-fidelity prototypes. Because we usually do not implement full functionality for low-fidelity prototypes, it is hard to claim if the evaluation results from a partially functional prototype will be the same as the results from a fully functional one. In future research, this issue should be treated as another consideration for designing valid low-fidelity prototypes.

ACKNOWLEDGMENTS

We appreciate all the participants to our study, and also we like to thank Hyewon Ko for her valuable ideas on making the paper-based prototype for a mobile phone, and Keval Mehta for his help on building the computer-based prototype. We also specially thank Erik Stolterman for his invaluable comments.

REFERENCES

1. Archer, N. P., and Yuan, Y. Comparing telephone-computer interface designs: Are software simulations as good as hardware prototypes? *International Journal of Human-Computer Studies*, 42, 2 (1995), 169-184.
2. Avrahami, D., and Hudson, S. Forming interactivity: A tool for rapid prototyping of physical interactive products. *Proc. DIS2002*, 141-146.
3. Buchenau, M. and Suri, J.F. Experience prototyping. In *Proc. DIS2000*. ACM Press, NY, 2000, 424-433.
4. Cockton, G. and Woolrych, A. Sale must end: should discount methods be cleared off hci's shelves? *Interactions*, 9, 5 (2002), 13-18.
5. Convertino, G., Neale, D.C., Hobby, L., Carroll, J.M., and Rosson, M.B. A laboratory method for studying activity awareness. *Proc. NordiCHI2004*, 313-322.
6. Davies, N., Landay, J., Hudson, S., and Schmidt, A. Rapid prototyping for ubiquitous computing. *IEEE Pervasive Computing*, 4, 4 (2005), 15-17.
7. Grady, H. M. Web site design: a case study in usability testing using paper prototypes. *Proc. SIGDOC'00*, 39-45.
8. Greenberg, S., and Boyle, M. Customizable physical interfaces for interacting with conventional applications. *Proc. UIST '02*, 31-40.
9. Gutierrez, O. Prototyping techniques for different problem contexts. *Proc. CHI '89*, 259-264.
10. John, B.E. and Salvucci, D.D. Multipurpose prototypes for assessing user interface in pervasive computing systems. *IEEE Pervasive Computing*, 4, 4 (2005), 27-34.
11. Lee, J. C., Avrahami, D., Hudson, S. E., Forlizzi, J., Dietz, P.H., and Leigh, D. The calder toolkit: wired and wireless components for rapidly prototyping interactive devices. *Proc. DIS2004*, 167-175.
12. Liu, L., and Khooshabeh, P. Paper or interactive? A study of prototyping techniques for ubiquitous computing environments. *Proc. CHI '03*, 1030-1031.
13. MacIntyre, B., Gandy, M., Dow, S., and Bolter, J. D. DART: a toolkit for rapid design exploration of augmented reality experiences. *Proc. UIST '04*, 197-206.
14. Nam, T., and Lee, W. Integrating hardware and software: augmented reality based prototyping method for digital products. *Proc. CHI '03*, 956-957.
15. Nielsen, J., and Landauer, T.K. A mathematical model of the finding of usability problems. *Proc. ACM INTERCHI'93*, pp. 206-213.
16. Pering, C. Interaction design prototyping of communicator devices: Towards meeting the hardware-software challenge. *Interactions*, 9, 6 (2002), 36-46.
17. Reilly, D., Dearman, D., Welsman-Dinelle, M., and Inkpen, K. Evaluating early prototypes in context: trade-offs, challenges, and successes. *IEEE Pervasive Computing*, 4, 4 (2005), 42-50.
18. Rettig, M. Prototyping for tiny fingers. *Communications of the ACM*, 37, 4 (1994), 21-27.
19. Rudd, J., Stern, K., and Isensee, S. Low vs. high-fidelity prototyping debate. *Interactions*, 3, 1 (1996), 76-85.
20. Schneider, K. Prototypes as assets, not toys: why and how to extract knowledge from prototypes. *Proc. ICSE-18* '96, 522-531.
21. Sefelin, R., Tscheligi, M., and Giller, V. Paper prototyping - what is it good for? A comparison of paper- and computer-based low-fidelity prototyping. *Proc. CHI '03*, 778-779.
22. Snyder, C. *Paper Prototyping: The Fast and Easy Way to Define and Refine User Interfaces*. Morgan Kaufmann Publishers, San Francisco, CA, 2003.
23. Thompson, M., and Wishbow, N. Prototyping: tools and techniques: improving software and documentation quality through rapid prototyping. *Proc. SIGDOC '92*, 191-199.
24. Virzi, R. A., Sokolov, J. L., and Karis, D. Usability Problem Identification Using Both Low- and High-Fidelity Prototype. *Proc. CHI '96*, 236-243.