

Usability Testing with Cultural Groups in Developing a Cell Phone Navigation System

Young Seok Lee¹, Young Sam Ryu¹, Tonya L. Smith-Jackson¹, Dong Jae Shin¹, Maury A. Nussbaum¹, and Kei Tomioka²

¹Department of Industrial and Systems Engineering, Virginia Tech
Blacksburg, VA, 24060, USA
{youngLee, yryu, smithjack, djshin, nussbaum}@vt.edu

²Design Center, Toshiba Corporation
Tokyo, 105-8001, Japan
kei.tomioka@toshiba.co.jp

Abstract

As the use of cell phones pervades over the world, globalization of user interface (UI) design is emphasized in developing cell phones for successful business. In reality, however, it is rarely possible to design a product that satisfies all international and intercultural customers. UI designers, therefore, need to collect cross-cultural usability information in the product development process in order to: 1) determine universal and to-be-localized components, 2) avoid cultural mistakes that lead to critical miscommunication, and 3) assess the usability of cross-cultural user interfaces. As an initial exploration, this paper reports a case study in which a laboratory-based usability test was conducted with four cultural groups (West African, East European, South American, and North American) to collect multi-cultural usability information in the iterative development process of a cell phone navigation system. Details of the methodology and lessons learned are presented.

1 Introduction

The use of cellular phones is being globalized. According to the International Telecommunication Union (ITU), the number of mobile service subscribers surpassed the number of landline telephone service subscribers in 2002, and grew to 1.37 billion in 2003, resulting in one-fifth of the world population using mobile services (ITU, 2003). The growth rate has been increased rapidly in recent years (2000-2003) along with an explosion of mobile service adoption in Africa and South America, and more rapid adoption is forecasted for future years.

As the use of cell phones pervades over the world, the globalization of user interface (UI) design is becoming more crucial to business success and building a loyal customer base. According to Marcus (2003), globalization of UI design can be achieved by identifying universal components and to-be-customized components of UIs. Universal components of UIs refer to components that can be accepted in many cultural contexts without modification. Use of number and auditory feedback is an example of universal components. On the other hand, to-be-localized components refer to components that should be localized for targeted markets. Language translation is an example of localization.

In reality, it is rarely possible to design a product that satisfies all international and intercultural customers. UI designers, therefore, need to collect cross-cultural usability information in the product development process in order to: 1) determine universal and to-be-localized components, 2) avoid cultural mistakes that lead to critical miscommunication, and 3) assess the usability of cross-cultural user interface.

Collecting cross-cultural usability information is usually time-consuming and expensive since ethnographic methods seem to be the most robust methods to collect this information (Jagne et al., 2004). However, the competitiveness of the global cell phone market requires cell phone manufacturers to develop a cost-efficient process to collect this information. As an initial exploration, this paper reports a case study in which laboratory-

based usability testing was conducted to collect multi-cultural information in the iterative development process of a cell phone navigation system. Details of the methodology and lessons learned are presented.

2 Background of Case Study: Design and testing of the cellular phone navigation system.

The case study described in this paper is based on a research project sponsored by a cell phone manufacturer, aimed at developing a cell phone menu navigation system for three functions: PhoneBook (PB), Voice Activated Dialing (VAD), and Short Messaging Service (SMS). Employing an iterative user-centered design approach, this project consisted of two studies, as shown in Figure 1.

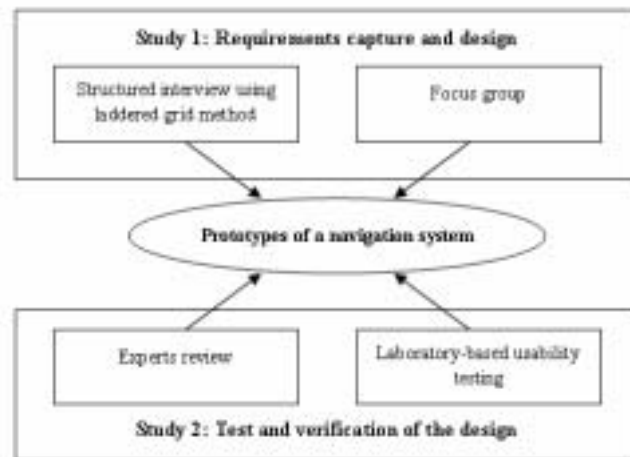


Figure 1: Design process of the cell phone navigation system

In Study 1, structured interviews using the laddered grid method and focus groups were conducted with US citizens to capture requirements such as users' mental models and preferences. Laddered grids, which are graphical representations to express their mental model to complete given tasks, were constructed by participants, and a medium-fidelity cell phone prototype was developed based on the laddered grids as a result of Study 1. In Study 2, the prototype was evaluated in a laboratory-based usability test and verified with usability experts. While conducting the laboratory-based usability test in Study 2, four cultural groups were recruited to examine various cultural issues such as possible cross-cultural differences in navigation patterns. General preferences in cell phone use were also examined to explore globalization of UI design. This paper describes details of the laboratory-based testing in Study 2. It should be noted that the focus of this paper is on presenting the methodology of collecting cross-cultural usability information in a laboratory setting rather than presenting results of the experiment

3 Case Study: Evaluation of the cell phone navigation system prototype in laboratory-based usability testing

3.1 Participants

A total of 28 participants were recruited for this study, including 16 USA and 12 international cell phone users were recruited for this study. The international users were selected from three non-contiguous geographic regions – West Africa (Cameroon and Ivory Coast), Easter Europe (Bosnia and Serbia), and South America (Argentina and Peru). Four users from each region participated in the test. Each group was equally divided in terms of gender. Mean age of participants was 23 (range = 19 – 44), mean years of cell phone use were 2.83 ($SD = 1.68$), and participants used cell phones on average about six times per day.

3.2 Materials and equipments

The prototype of a cell phone navigation system was developed using Macromedia Flash MX™ as shown in Figure 2. The hardware features of the prototype, such as the key pad layout and function assignment were employed from a cell phone model of the project sponsor. The prototype was presented on a Dell Optiplex desktop computer running Windows XP. Camtasia version 3.0.1 was used to record users' interactions (e.g., mouse movements), the cell phone UI, and users' comments.



Figure 2: A snapshot of the prototype used

An overall experiment script was developed, including an interview script and a verbal protocol script. A demographic questionnaire was used to gather general information and ensure that participants fit the criteria for the study. A post-task questionnaire was developed to acquire subjective usability ratings of the prototype. The usability ratings consisted of six questions regarding ease of use, degree of expectation match, meaningfulness of text labels, meaningfulness of icons, supportiveness, and overall satisfaction. In addition, three post-experiment open-ended questions were developed to explore any differences or similarities related to culture between cultural categorizations. Combining information from Jordan (1998) and Hofstede (1991), we derived three constructs that could be used to discriminate any cultural differences that may occur because of the design of the cell phone. These three constructs were global supportiveness (the degree to which the phone provides information that helps the user understand what to do next), toughness (the user's perspective on the importance of functionality and aesthetics), and cultural customization (users' preferences for how the phone should be designed to appeal to different cultures).

3.3 Procedures

Participants were given general instructions before each usability session began. In this study, five specific tasks were examined for the three functions: saving a phone number using phonebook, calling, setting voice dialing, sending a new message, and deleting a message. Task orders were randomized to minimize any confounding effects due to sequence. Participants performed each task using the prototype on a desktop computer. After each task was completed, the Camtasia recording was presented to participants who were asked to provide the retrospective think aloud (RTA) to describe what they were thinking while completing that specific task. Participants were probed for clarification as needed. After the RTA was completed, participants completed the subjective usability ratings, and additional open-ended questions related to cultural customization. Participants' completion times for each task were also recorded during task performance and observers recorded specific comments made by participants during task performance that reflected critical incidents. In addition, each observer logged critical incidents on the recording sheets.

3.4 Data collection and analysis

We employed a mixed methods approach and therefore collected both quantitative and qualitative data. Quantitative data were obtained from users' performance (task completion time) and subjective ratings acquired by administering a usability questionnaire (ease of use, degree of expectation match, meaningfulness of labels and icons,

supportiveness, and overall satisfaction). These numerical data were analyzed using both descriptive and inferential data analysis. However, since this study was an exploratory study with a small number of participants for each culture group, more emphasis was placed on the qualitative data analysis rather than the quantitative analysis.

Qualitative results were derived from reviewing the task performance recording, verbal protocols in RTAs, responses to the post-task interviews, and critical incident data recorded by observers. Since a variety of mixed data was collected, we developed a systematic data analysis approach to collect cross-cultural usability information. The systematic approach consists of five stages as below.

3.4.1 Development of the menu structure map

The hierarchical menu structure used to develop the prototype was illustrated using MS Excel™. The map was represented using a vertical cascade menu display in which the top-level menu is laid out vertically (Refer to Figure 3).

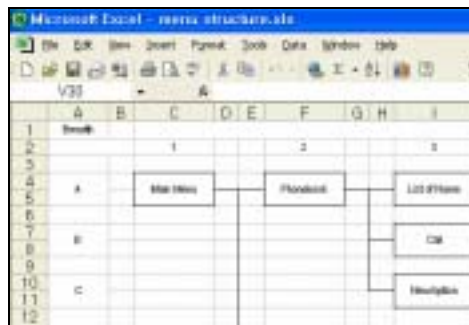


Figure 3: An example of menu structure map

3.4.2 Review of the recorded task performance

Four usability engineers reviewed the recorded task performance and post-task interviews. They were given the menu structure map (Figure 3) and a data logging template that allowed them to record their observations of several important events. Figure 4 illustrates the example of the data logging template and descriptions for each column are given below.

Number	Location in Menu	Description	Time duration	Incorrect Keys/pressed during task performance	Critical incidents / Verbal protocol	Problems faced	Design recommendation
A2	Home	Homebook	3				
A3	List of Menu	List of Menu	8 s		"He pressed 'ok' key twice but nothing happened. I assumed that it would bring some menu such as delete, add number, different things to do with her entry." "Finally pressed Newspaper" "I think here just options will be better than newspaper because the two are different tasks."	"Limited function of the prototype" "Unclear labels"	"OK should bring detailed information of the selected name; however, it should not show nothing directly." "New and options are different tasks. It should be separated. Maybe 'New' goes to the left and place instructions for users to press 'ok' at the screen."
B4	Newspaper	ADD voice	1		Automatic process		
F5	voice prompt	voice prompt	2		"I don't need to listen to my recorded voice again"	Different user preferences.	Hearing his/her own recorded voice might not necessary even though users' preference conflict. However, his/her voice did should allow it.
F5	confirmation	confirmation	1		"I don't need to know it is saving. I want to know if it has saved."	Not descriptive confirmation message.	Confirmation word feedback help users understand what's been done. Replace the progress bar with word confirmation saying voice did for records has been saved.
T1	complete		27				


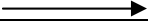
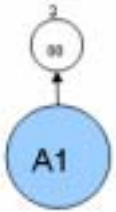
Figure 4: The data logging template

- *Number: Ordinal number of interaction*
- *Location of menu: corresponding code in the menu structure map where users' interaction was observed.*
- *Description: title of the node in the menu structure map.*
- *Time duration: time taken to move to next menu item (in second).*
- *Incorrect key pressed: key pressed that is unnecessary to perform a given task.*
- *Critical incidents/verbal protocol: detailed description of critical users' interaction and corresponding users' explanation during the RTA (retrospective think aloud).*
- *Problem faced: observers' interpretation of the critical incidents.*
- *Design recommendation: design recommendations were generated based on the verbal protocol and users' interaction.*

3.4.3 Representation of navigation paths

Referring to the data logging template, navigation paths taken by participants during the experiments were represented using a navigation path diagram (NPD). The NPD is a network diagram that consists of nodes and arcs. The notation used in the NPD is described in Table 1.

Table 1: Notations of the navigation path diagram

Title	Description	Graphical representation
Navigation node in blue	Menu items that users must select to complete tasks. The code in the circle refers to the one in the menu structure map. For example, A1 is the main menu.	
Arc	Movement in the menu structure	
Deviation node in four colors	Deviations from the correct paths. For example to the right, two deviations were observed at A1 menu item since participant(s) moved back from A1(main menu) to 00 (standby mode). Four colors were used to represent users' nationality. - White: US citizen - Yellow: Eastern European - Red: South American - Purple: African	

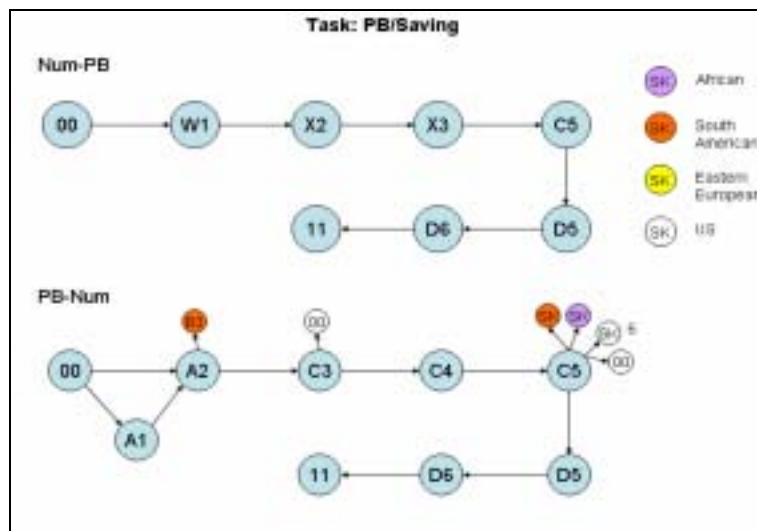


Figure 5: The NPD for phonebook/saving task

The use of NPDs made it easy to identify where in the hierarchical menu usability problems occurred with the prototype. Also, navigation patterns between cultural groups could be compared and contrasted more easily. The NPD was drawn for each task scenario, and an example for Phonebook-Saving task is illustrated in Figure 5. As shown in the figure, participants used two paths to complete this task. The first path was to type a phone number at the standby mode and then save the number using menu options (Num-PB in the figure). The second path was to find a phonebook menu option, type all necessary information, and save it (PB-Num in the figure). The NPD revealed that all deviations occurred in the second way. Furthermore, most of deviations were found at the menu item C5, and all cultural groups made similar errors at C5 except the Eastern European group. In this manner, navigation patterns were examined for each task, and a list of problematic menu designs was generated.

3.4.4 Verbal protocol/critical incidents examination

When a list of problematic menu items was generated for all tasks, we sorted data logging templates by cultural groups and reviewed verbal protocols to determine causes of the deviation and possible cultural differences. For example, reviewing verbal protocols with respect to the C5 menu option in the PB-Saving task revealed that use of unclear icons caused deviations at this menu, and perceptions for the same icon is different across cultural groups. Also, the need for icon localization to differing cultural groups was found. For instance, some cultural groups showed difficulty recognizing the office building icon representing ‘work phone number’ in the phone book, and they expected different icons such as desks or hammers to represent their work place, which indicates a strong need of localization for this icon.

3.4.5 Qualitative analysis for open-ended questions related to cultural customization

Content analysis was conducted to identify important themes, objects, and referents contained in their responses for the three constructs (supportiveness, toughness, and cultural customization). One of the results revealed that supportiveness seemed to be a function of the extent to which the phone’s text labels, icons, and menus helped users to figure out how to do a task and this was consistent across cultures. Users seemed to rely on these attributes of the UI to infer meaning during task performance, and when these attributes were difficult to understand, overall supportiveness was negatively affected. According to Jordan (1998) and Hofstede (1991), individuals in certain cultures will expect more supportiveness from a technology. For example, users from such areas as Sweden, Peru, Iran, France, and West Africa will expect more supportive technologies than Japan, Austria, Italy, Colombia, and the USA. This information is based upon where each culture scores on Hofstede’s Masculinity Index. However, no real differences could be seen in the qualitative data regarding supportiveness.

4 Summary and discussion

This paper reports a case study where a laboratory-based usability test was conducted to test a cell phone menu navigation system prototype and collect multi-cultural usability information regarding the cell phone UI design and cultural preferences. Although a small number of participants representing each cultural group were involved, various cross-cultural information, such as navigation pattern, mental model, metaphors, and appearance, was obtained using a systematic approach to data analysis. However, it should be noted that this approach is a quick approach that can be used in the product development process.

UI designers or usability engineers using this quick solution need to be cautious for the following reasons. First, data from only four participants can not represent their cultural group. When a potential difference between cultural groups is found in the usability testing, further studies must be performed to verify if the difference is a cultural factor or individual characteristics. Use of existing cultural models such as Hofstede (1991), Tromprenaars (1993) and Hannerz (1992) can help verify the difference.

Secondly, the cultural questions that we used for the three constructs need to be refined further to elicit more specific information about cultural preferences and patterns. Although we refined the questions several times to make it easy for participants to answer, we found that some participants experienced difficulty comprehending terms, such as supportiveness, functions of the phone, culture, and customization. Further research is necessary to identify how to more directly and clearly elicit cultural information in interviews.

Acknowledgements

This study was a collaboration between Virginia Tech and the Toshiba Corporation of Japan. We are most grateful for the research funding and resources provided by the Toshiba Corporation.

References

- Hannerz, U. (1992). *Cultural complexity; studies in the social organisation of meaning*. New York: Columbia University Press.
- Hofstede, G. (1991). *Cultures and organisations: Software of the mind*. New York: McGraw-Hill.
- ITU. (2003). *Mobile cellular, subscribers per 100 people*. International Telecommunication Union. Retrieved February 10, 2005, from the World Wide Web: http://www.itu.int/ITU-D/ict/statistics/at_glance/cellular03.pdf
- Jagne, J., Smith, S. G., Duncker, E., & Curzon, P. (2004). *Cross-cultural interface design strategy* (IDC-TR-2004-006): Middlesex University.
- Jordan, P. W. (1998). *Designing pleasurable products*. London and New York: Taylor and Francis.
- Marcus, A. (2003). Global and intercultural user-interface design. In Jacko, A and Sears, A. (Ed.), *The human-computer interaction handbook* (pp. 441-463). Mahwah, New Jersey: LEA.
- Trompenaars, F. (1993). *Riding the waves of culture: Understanding cultural diversity in business*. London: Nicholas Brealey.