

Cross-Cultural Design and Evaluation of the Apple iPhone

Michael A. Oren, Utkarsh Seth, Fei Huang, and Sunghyun Kang

Springer-Verlag, Computer Science Editorial, Tiergartenstr. 17,
69121 Heidelberg, Germany
{moren, useth, feihuang, shrkang}@iastate.edu

Abstract. In this paper, we report the design and results of a study to improve the usability of the iPhone for a global audience, particularly in India and China. With extensive research in cultures of three countries China, India, and the United States, the iPhone interface was redesigned with an eye to culturally universal (for the three cultures of interest to this study) and ease of access of functions most used by mobile phone users in these cultures. Both the iPhone and the new prototype interfaces were tested to measure their usability and results are reported here.

Keywords: Global Design, Usability, iPhone, Mobile Interfaces.

1 Introduction

In an ever shrinking world and increasing globalization, cross cultural studies while designing products is proving to be a valuable tool for its effectiveness. From Switzerland to Singapore, from Boston to Bombay, the latest electronic designs are familiar icons of modern technology & culture. Despite this globalization, it is detrimental to the selling value of the product if it is not designed keeping the culture in mind where the product is going to be used in. Previous work on mobile device redesign has been conducted by Lee, et al who looked into cultural issues with menu hierarchies of mobile devices [1].

The product in question, the iPhone, was launched in June 2007 by Apple Inc. The phone, hailed for its revolutionary technology, indeed had lots of design problems. There were functions that were present in the phone but the user had no clue how to access them. The visual cues, which could tell the user what a function or icon does, provided to the user were also not very intuitive or logical. Some of the most common features that a user wished to access in a minimum number of steps were embedded down in the hierarchy of menus. The phone also lacked consistency among the screens where it performed different actions on the same inputs (such as tapping a name found in recent calls resulting in calling the person, whereas tapping a name in contact would take the user to more details about the contact). And most important of all, the phone was not designed to cater to the needs of various cultures where the mobile phone market is surging.

The current iPhone has several flaws which makes it unsuitable for different cultures. For example the icon of the “Maps” is not very intuitive when we talk about

various cultures. The icon displays an interstate Logo which is indeed very common to the people living in the USA. But when we talk about people from India and China there is no concept of an “interstate” there so if this version of the iPhone is launched in these countries people there might be clueless about what the icon represents.

To explore similar type of problems and solve them eventually we need to interview people from different countries and ask them about their inhibitions with the iPhone. We will have them do specific tasks and will ask for their input on the product. This way we will get a clear idea as to what people from different cultures like India and China expect from the product.

There were also general usability problems that were found in the iPhone. Those were inconsistency problems with the interface that are discussed in detail in the design goals. There were also opportunities to maximize the use of the multi-touch functionality in the iPhone. We came up with several efficient designs that would provide more uses for the touch functions beyond simply just tapping the screen.

Our study of the iPhone took place from January through May 2008 using version 1.14 of the iPhone OS.

2 Early Evaluation

Before redesigning the iPhone, we conducted exploratory research to determine the most used features of mobile phones among the three cultures. We also conducted a study of the current iPhone icons to determine how easily identifiable users found the icons to determine which icons required a redesign.

2.1 General Cell Phone Study

One of the early evaluation tools we used was a short survey to poll potential users regarding what features they most commonly use on their cell phones and the steps they have to take to use those features. One of the most striking things we learned from this survey was that although the majority of respondents indicated that the alarm clock was one of the most common features they use, only one person (out of 5 who listed alarm as a critical feature) listed fewer than seven steps in order to set an alarm. One person even had to take twenty steps on his LG cell phone. The iPhone also suffered from superfluous steps in order to set the alarm, so we decided that this would be one of the key tasks for users to complete. As part of this survey, we asked non-iPhone users what features they most desired to see from the iPhone and the response was overwhelmingly data based features such as the web browser and Google Maps. Current iPhone users often stated that they feature they wanted improved the most were missing features, such as copy & paste rather than requesting changes to the current design. However, this might be early adopter bias not willing to admit Apple makes design mistakes.

2.2 Icon Evaluation

Sixteen participants (six Chinese, five Indian, and five American) participants completed a survey of the thirty-two icons used by the iPhone in version 1.14 of the OS. With the exception of icons appearing within menus, such as the clock under phone,

no contextual information was given about the icons on the survey. Participants were given the survey with blank lines beneath the icons and told to write down what mobile phone functions they thought the icon represented.

Table 1 presents the number of participants (out of 16) that correctly determined what each icon in the original interface represented.

Table 1. The number of users (of 16) properly identifying the icon

Map	Calendar	Notepad	Weather	Stocks	Clock	YouTube	iTunes Download
13	16	4	16	7	16	3	4
Calculator	Settings	Photos	Camera	SMS/Text	Phone	Mail	iPod/Music
15	7	12	8	9	14	8	14
Mute	Keypad	Speaker Phone	Add to Conversation	Hold	Contacts	Recent Calls	Browser
11	6	2	3	13	3	0	2
Contacts	Keypad	Voicemail	Favorites	World Clock	Alarm	Stopwatch	Timer
4	3	4	9	7	14	11	4

Sixteen of thirty-two icons performed very poorly, with fewer than eight people properly recognizing the icons. In order from “least recognized” to “most recognized”, the icons receiving fewer than three correct identifications were: recent calls (0); speaker phone (2); Internet browser (2); YouTube (3—the response “video” was accepted); add to conversation (3); contacts—under Phone (3); keypad—in a call (3); contacts—while in a call (4); timer (4); voicemail (4); iTunes Download (4—the response “download” was accepted); notepad (4); keypad—under phone (6); stocks (7—both American and Chinese participants had over 50% recognition, but none of the Indian participants recognized this icon); settings (7); world clock (7). In addition, the mail icon had 50% recognition rate (8 participants recognized it), but it was skewed because all five American participants recognized it while only two of six Chinese participants and one of five Indian participants properly identified it.

3 Methods

All studies took place in an isolated office or room. The original iPhone interface was tested using a 4 GB iPhone running the latest version of the iPhone OS (1.14) and participants hands were recorded via a mini-DV camcorder mounted on a tripod placed behind the participant’s shoulder. A Windows XP laptop was used to run the prototype of the new interface, which used Camtasia to record the screen and the participant’s voice.

Participants were randomly assigned to start with either the original iPhone interface or the prototype interface and after completing all tasks they would switch to the other version. There were seven tasks users were asked to complete, which were evaluated based on completion time and the number of clicks it took them to complete the tasks. The seven tasks users were asked to complete were

1. Turn phone on
2. Set the alarm to recur every weekday at 4:45 AM (time varied)

3. Call the contact Google411 and place the call on speaker phone
4. Add a recent call as a new contact named “Joe Smith”
5. Rearrange the home-screen
6. Silence the phone
7. Turn phone off

These tasks were randomly assigned to participants, with the notable exception of turning the phone on and turning the phone off, which relied on the current state of the phone to determine whether the participant would be able to complete the task.

Participant performance was measured by the amount of time taken (measured in seconds) to complete each task as well as by the number of clicks the user took to complete tasks. Correct clicks when typing in a name were not counted, nor were clicks (or the time taken) that led to portions of the interface that had not been implemented in the prototype but would have normally worked (clicks that led to unimplemented areas that would not have led to successful completion of the task were counted). Significance level was calculated with 95% confidence using a Student's t-test.

3.1 Participants

We recruited five (3 males, 2 females) American participants, five (1 male, 4 female) Indian participants, and six (3 male, 3 female) Chinese participants, primarily through a word-of-mouth recruitment strategy. One American participant owned an iPhone, while the remaining American participants owned other phones (two Motorola—one was a smart phone, two Sony Ericsson). Three Indian participants owned a Samsung phone, one a Sony phone, and the last a Motorola—not of the phones were smart phones. Likewise, none of the Chinese participants had a smart phone as their primary phone—four had an LG brand phone and two had Motorola brand phones. Three American participants fit into the primary target age range of 20-30; however, one participant was over the age of 50 and one was under 20. The participant under 20 was not yet a college graduate. All other participants had completed high school or higher. Other than the participant over the age of 50 rating the intuitiveness of many of the original interface functions lower than the other participants, there was no notable difference between the users that fit into our target demographic and the outliers. All five participants in the Indian group were between the ages of 20-30. Five of the six participants in the Chinese group were between the ages of 20-30, the sixth was between 30-40 years of age. Unlike the American outlier, however, the Chinese participant in the age range of 30-40 did show significant performance differences when compared to the other Chinese participants—taking a significantly longer amount of time spent on several tasks (they were removed from the analysis of those tasks noted below).

4 Results

Only three tasks resulted in a statistically significant improvement for the American group with a 95% confidence level between the original design and the new design: calling a contact on speakerphone, adding a recent contact, and turning the phone off. This improvement was both in the time needed to complete the task as well as the

number of times participants clicked on elements of the interface (both results were significant). Marginal significance ($p \leq 0.10$) was also found for the tasks of silencing the phone and rearranging the home-screen, but this marginal significance was only an indication of reduced number of clicks in the new interface as compared to the old interface not in the amount of time used to complete tasks.

All but one task (setting the alarm) resulted in statistically significant improvement in terms of number of clicks for participants in the Indian group with a 95% confidence level. The task of setting an alarm did result in marginally significant improvement for Indian participants. Likewise, five of seven tasks resulted in statistically significant improvements in the time spent to complete the tasks on the new prototype for Indian participants, only the times for adding a recent call and turning the phone off did not show significant improvement.

Chinese participants saw four of seven tasks show a significant improvement in the time to complete the task with a 95% confidence level, with the task of placing a call on speakerphone showing marginally significant improvement. The tasks of turning the phone on and setting the alarm did not show significant improvement for completion time, and neither of these tasks showed a statistically significant improvement in terms of number of clicks required to complete tasks. Four tasks showed statistically significant improvement for the number of clicks required to complete tasks, and silencing the phone showed marginally significant improvement for the number of clicks required.

While on most tasks the American participants still performed better than the Indian and Chinese participants, the difference between the number of extra clicks and time needed for Indian and Chinese participants as compared to the American participants was reduced (this was not a statistically significant improvement for the majority of tasks).

4.1 Survey Results

For all tasks, nine out of the sixteen participants indicated that the new interface made the task more intuitive. For the task of silencing the phone, three of the five participants preferred the new interface. Only one American participant preferred the new interface for the task of setting the alarm (all six Chinese participants preferred the new interface and three Indian participants preferred it for the alarm task), but one of the American participant indicated this response had more to do with the style of the sliders, which we were unable to duplicate for this study. While five participants each in both the American and Chinese groups preferred the new interface for silencing the phone, only two of five participants in the Indian group preferred the new interface. It was unclear why the Indian participants felt this way, quantitatively the participants did significantly better using the new silence function ($p < .003$) for both clicks and time required to silence the phone.

All but four (2 Chinese and 2 Indian—all 4 of whom had heard of it) participants indicated that they knew about the iPhone beyond its name, three (2 American and 1 Chinese) had used one before but did not own one, and one American participant owned an iPhone. Two (1 American and 1 Chinese—the Chinese participant also used Windows) participants used OS X on their primary computer; one American participant used OSX, Linux, and Windows on a daily basis; one American participant used

Windows XP and Linux on a daily basis; all remaining participants (5 Indian, 3 American, and 5 Chinese) participants used Windows XP exclusively.

Participants were asked to rate the intuitiveness of each task on each interface based on a 5-point scale (1-Not at all intuitive; 5-Very intuitive).

On the original interface, three tasks had mean scores that were below three for the American group: turning the phone on (2.6); turning the phone off (2.8); and rearranging the home-screen (2.8). The Chinese group only scored one task below 3 (rearranging the home-screen with a mean score of 2.3) and silencing the phone had a mean score of exactly 3 points for the Chinese group. The Indian group was by far the most favorable of the original interface with no task receiving a mean score of less than 3.4. When all 16 participants were combined, only the task of rearranging the home-screen had a mean score less than three (only 6 of 16 participants scored it 4 or higher—all other tasks had at least 50% of participants scoring it 4 or higher). The fact that turning the phone on scored lowest for the American participants, despite taking most participants fewer clicks and less time than either of the other two tasks seems strange, but may indicate that participants feel this is an important task to keep as simple as possible. Furthermore, only two tasks scored above a mean score of 4 on the original interface (when all participants were combined): setting the alarm (4.3 and 14 of 16 participants scoring it 4 or higher) and making a call on speaker phone (4.3—with 13 of 16 participants giving it a score of 4 or higher).

In contrast, all but one task (turning the phone off—3.8 for the American group but over 4 for Indian and Chinese groups) had a mean score above 4 in each respective group when participants rated the intuitiveness of completing tasks on the new interface. When combining all participants, all tasks scored at least a 4.2 for the new interface and had at least 13 of 16 participants rating it a 4 or higher on intuitiveness. Furthermore, while 8 of 16 users ranked rearranging the home-screen as the most difficult task on the original interface, there was no clear “worst” task on the new interface.

It should be noted that the original interface scored an overall intuitive mean score of 3.6 (Chinese 3.83; American 4; Indian 3.6) whereas the new interface received a just barely better score of 3.8 (Chinese 3.7—worse than original; American 3.6—worse than original; Indian 4.2—only better than original). However, one American user rated the overall intuitiveness of the new interface a ‘5’ on all but one task (rearranging the home-screen—4) and indicated that they preferred the new interface for all tasks (and overall), but scored the overall intuitiveness of the new interface a ‘1’. This seems to be a mistake by the participant that has skewed this particular result significantly due to the small sample size. A similar mistake appeared to have occurred with one Chinese participant (who also rated the new interface a ‘1’ and the old interface a ‘4’ despite very high marks for the new interface as compared to the old and listing preferences for the new interface for all tasks other than silencing the phone)—this participant also noted that the original interface definitely needs to be improved.

At least 9 of 16 participants preferred the new interface on all tasks (turning the phone off—9 participants; and setting the alarm—10 participants had the fewest people prefer them). Overall, 14 of 16 (all but 1 American and all but 1 Indian) users indicated that they preferred the new interface to the old interface. For the alarm task, it is our belief, based on feedback from participants that the old interface was preferred only due to the fact that we were unable to mock up the iPhone-style sliders on

the computer. In general though, this indicates a remarkable success for our new design within the American, Chinese, and Indian marketplaces.

5 Design

Most of the tasks showed considerable improvements in terms of fixing general usability issues and bringing all three cultural groups to a roughly equal level of performance on tasks they were completing for the first time. However, a large performance gap was still very apparent in adding a recent call as a new contact between the Chinese group and the other two cultural groups. Furthermore, 16 of the 32 icons on the iPhone were identified correctly by fewer than half of the participants. Using these findings we undertook efforts to redesign the remaining elements of the iPhone.

While all tasks showed statistically significant improvement overall from the original design to the first prototype and each individual group always showed improvement (although not always significant) from the first prototype to the new prototype on all tasks, adding a recent call as a new contact appeared to have a cultural gap between the Chinese group and the other two cultural groups (see figure 1).

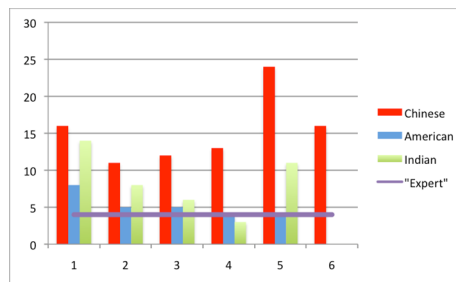


Fig. 1. A graph showing the number of clicks each participant took to complete the task

As this chart clearly shows, Chinese participants used 10 or more clicks than the average American participant. In reviewing the recording of the participants, these excessive clicks appeared to be due to participants in this group not noticing either the green plus or blue arrow that they needed to click in order to complete this task. Whether this was due to the buttons not being salient enough or whether it had to do with the button functionality not being obvious enough was unclear. As such, we chose to go about a redesign that would attempt to address both of these potential problems (see figure 2).

In order to address the issue of what the button does, we added the contact icon (address book) to the icon to make it clear that the “+” meant “add to contacts”. While undergoing this redesign, we also decided to make it more obvious what the blue arrow on the right-hand side of each contact represented, which we achieved by grouping the time/day of the call with the blue arrow, which suggests that clicking that button will lead the user to additional information about the call. As an added bonus, this redesign also increases the surface area where the user can press in order

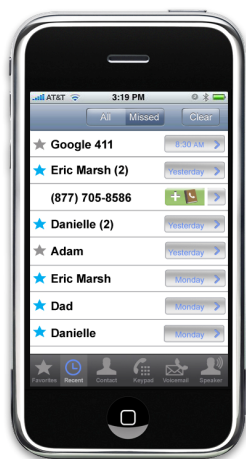


Fig. 2. The redesigned recent calls screen



























to add a call as a new contact or get additional information since one of the problems we found with this task was that the blue arrow was simply too small for most users to easily click.

In order to address the problem of saliency (e.g. the button going unnoticed), we increased the size of the button (partially as a natural function of adding the icon). We also chose a shade of green that stood out more from the blue of the rest of the interface. In addition, we made the contact icon brown as an additional method of catching the users eye as the combination of green, white, and brown on an otherwise consistent color palette draws the eye's attention to the button and ensures that users are able to notice its presence.

As we were unable to conduct a second study, we are unable to verify this claim. In addition, it is possible that the redesign creates an interface that distracts users from their primary goal on the recent calls page in order to promote a (possibly) rarely used feature. These are both considerations that would need to be addressed in a future study but are beyond the scope of this current project.

With half of the icons currently on the iPhone failing our low expectations of at least 50% user recognition, we knew that serious problems existed. To be fair, many of the icons that users failed to identify with 50% accuracy are ones that are fairly common on phones in the US (such as voicemail) so it is understandable why Apple may have chosen them; however, in order for Apple to succeed in the international marketplace they must consider icon designs that are easily recognizable to a wider audience. Out of these icons, we chose to redesign 15. We chose not to redesign the 'recent calls' icon (despite 0 of 16 users recognizing it) due to the fact that none of the users had trouble finding/going to recent calls when asked to complete that task so we felt that was simply an issue of the survey not providing enough context for that icon. In addition, we chose not to redesign the settings ('gears') icon as our group was unable to decide on a universal design for it. Many of our redesigns used the concept

Table 2. Original icons and the redesigned versions

Function	Original	New
Keypad		
YouTube		
Maps		
Music Store		
E-mail		
Contacts		
Notepad		
Internet Browser		
World Clock		
Timer		
Voicemail		
Speakerphone		
Add Caller		

of combining two common, easily recognizable icons in order to create one icon that was easy for all three cultures to recognize and interpret. The redesigned icons can be seen in Table 2.

6 Conclusion

We managed to identify key cultural design flaws and general usability problems present in the iPhone and successfully redesign it to reduce (although not completely eliminate) these problems. With the information presented here, Apple and other cell phone manufacturers should be able to guide their design in order to ensure that the widest possible world population can use it. The three cultures that were the focus of this study represent roughly half of the world's population as well as some of the largest, fastest growing mobile markets.

As noted earlier, these design solutions do not completely solve the cultural bias of the original iPhone and further iterative testing and redesign would help to further reduce the level of culturally based performance differences. Given more time, it might also be possible to infer more general cultural usability design guidelines based on these findings; however, such a goal was beyond the focus of this study. If our team had more time, it would have been ideal to run another study to confirm that the

final redesign successfully addressed the most glaring remaining issues. However, overall, we feel that this was an extremely successful redesign that will help prevent calls to Apple's help desk when the iPhone is officially released in India, China, and other world markets.

Reference

1. Lee, Y., Ryu, Y., Smith-Jackson, T., Shin, D., Nussbaum, M., Tomioka, K.: Usability Testing with Cultural Groups in Developing a Cell Phone Navigation System. In: The Proceedings of Human-Computer Interaction International (2005)