The Role of Physical Affordances in Multifunctional Mobile Device Design

Sorin Adam Matei, Purdue University, USA
Anthony Faiola, Indiana University, USA
David J. Wheatley, Motorola Applied Research, USA
Tim Altom, Indiana University, USA

ABSTRACT

As designers of mobile/media-rich devices continue to incorporate more features/functionality, the evolution of interfaces will become more complex. Meanwhile, users who carry with them cognitive models must align them with new device capabilities and corresponding physical affordances. In this paper, the authors argue that based on HCI design theory, users approach objects by building mental models starting with physical appearance. Findings suggest that users who embrace a device’s multifunctionality are prevented from taking full advantage of an array of features due to an apparent cognitive constraint caused by a lack of physical controls. The authors submit that this problem stems from established mental models and past associated behaviors of both mobile and non-mobile interactive devices. In conclusion, users expressed a preference for immediate access and use of certain physical device controls within a multi-tasking environment, suggesting that as mobile computing becomes more prevalent, physical affordances in multifunctional devices may remain or increase in importance.

Keywords: Affordances, Cognitive Models, Human-Computer Interaction, Interaction Design, Media, Mobile Devices, Multifunctionality

INTRODUCTION

The cell phone and other small mobile devices are rapidly becoming the preferred access points to, and storage repositories for, personal messages and media, such as music, photos, and video. Such devices are transforming person-to-person mobile communication into a convergence of voice and media sharing communication, i.e., devices with multifunctional capabilities (Heo, Ham, Park, Song, & Yoon, 2009; Monk, Fella, & Ley, 2004). The functionality of these devices is further enhanced by the possibility of transferring media content to a fixed interface display, such as a personal computer (PC), TV, or stereo system. The emergence of such multimedia-enabled mobile devices creates a
The number of physical and conceptual design challenges that revolve around two issues.

The first issue is related to the fact that commercial devices are relatively shrinking—even the iPad is smaller than a typical laptop—but continue to incorporate more features and functionality. Consequently, controls and interfaces have either become more crowded or have been buried in complex hierarchically structured graphical user interfaces (GUIs) (Vivrou & Kabbasi, 2002). The second issue is that the ever-increasing functionality offered by these novel technologies is limited by the socio-cultural maps and cognitive models that users and designers carry in their minds about mobile and non-mobile device capabilities and their corresponding physical affordances (Faiola, & MacDorman, 2008; Gibson 1977, Hartson, 2003, Norman, 2002).

The concept of affordance designates the capacity of a device to suggest a particular kind of use by virtue of some physical attribute. For example, a cell phone’s most significant affordance is related to voice calling, as an object made to be grasped with one hand and positioned between the ear and mouth. This affordance is reinforced both by the physical design of the device and its controls (softkeys, menu options, hardwired buttons, etc.), as well as the way in which people approach it cognitively.

Physical affordances are extremely effective when they are incorporated into simple/unifunctional devices with limited functions (Norman, 1998). However, problems often emerge when these multimedia devices additionally serve as gateways and transfer devices for video, photos, text, and various other types of information and media content between different types of platforms. For example, how should the device make its non-talking affordances visible and immediately understandable to the user who associates the device with a more basic cell phone or home phone that only makes calls? The process of “unveiling” the functional potential of the multimedia device relies heavily on creating features and interface and interaction design solutions that suggest the idea that the device is not just a cell phone, but also a vehicle for content capture, storage, and transfer between various platforms.

The study described here suggests that users approach such multifunctional devices with cognitive models derived from their prior experiences of using phones, cameras, camcorders, PDAs, and PCs. A directive principle, should be that industrial and interaction designers must consider the users’ prior experience with media devices, in order to avoid conflicts with existing cognitive or mental models associated with the use of these single appliance interactive devices. Forlizzi (2007) specifically recommends thorough examination of prior subjective experiences with mobile products, which can lead to generalizable knowledge for design activities. At the same time, the design process should not be dogmatic. Ecological, participatory design is preferable. Prototypes that reflect a combination of existing and novel features and practices need to be continuously tested and “winning” features sorted out in the process of actual usage. Patterns of use need to be monitored and the conclusions of such monitoring fed back into participatory design activities (Forlizzi, 2007).

This paper, which is theoretically informed by affordance theory (Norman, 2002), addresses the role that particular affordances and cognitive models play in facilitating or hindering the use of the cell phone as a gateway for multimedia content generation, transfer, and visualization. More specifically, we reconsider the role of the touch screen interface and the physical device affordances in such multifunctional mobile devices.

**MULTIFUNCTIONAL DEVICES AND AFFORDANCES**

Discussions about how new audio and video technologies and information management should be embedded in the mobile communication environment and assimilated by users go back to the 1980s and to Xerox’s Palo Alto Research Center (Abowd & Mynatt, 2000; Dourish, 2001). Nevertheless, the practical
integration of multimedia functionality within a mobile device is an evolving concept (Goldstein, Nyberg, & Anneroth, 2003). As third and four generation wireless technologies (3/4 G) become more prevalent, users are finding themselves dealing with the escalating complexities of multifunctional devices. As a result, they will need interface affordances that are capable of bridging the gap between the device’s functional capabilities and the user’s outdated mental models (Agre, 2001; Dourish, 2001).

The importance of physical affordances for mobile devices

True affordance means that each function and its corresponding method of operation are apparent from the device’s visual appearance, haptics, or other direct sensory indicators. Simply put, affordances should provide strong clues about the operation of the device (Norman, 2002) and make its function intuitively obvious, or using a consecrated usability research term, immediately learnable (see below for details about this and other core usability concepts). Norman distinguishes between real and perceived affordances. Real affordances include the physical characteristics of a device or interface that facilitate its operation (Hartson, 2003). In other words, physical affordances help users with their physical actions (Hartson, 2003).

Perceived affordances are another class of device features (2002), which offer users not only clues related to the proper operation of a device, but help with their cognitive actions. Both of these fundamental types of affordances should be deployed harmoniously to “show how good design can make the appropriate use of a device clear and obvious to a user” (Dourish, 2001, p. 119).

Developers of multipurpose mobile devices have traditionally assumed that affordances that leverage what users already know about interfaces and social connectivity will result in greater user recognition in the context of new computing experiences (Agre, 2001; Klopfer, Squire, & Jenkins, 2004). Yet, such assumptions are often optimistic, both in terms of appropriateness of such conventions for the hand-held mobile experience and also in terms of the ability of first time users to transfer past experience to the new interface context. In an early study on the interface design challenges that surround multimedia mobile devices, research findings suggested that the efficacy of such devices decreases dramatically with their functional complexity (Goldstein et al., 2003). The study notes that elimination of traditional camera control in a multifunctional device that could take pictures, prevented many users from using this particular functionality. Goldstein et al. drew this broader conclusion:

The porting of the stationary computer metaphor to a mobile multipurpose device may prove ineffective if it is in conflict with previously acquired efficient source metaphors when using information appliances tailored to accomplish a single task. It is mandatory that efficient source metaphors are given proper attention. Omitting well known affordances under the assumption that the intelligent user is capable of accomplishing the task efficiently anyhow is a bold assumption. All applications included in a multipurpose device must be designed with the respective information appliance in mind (Goldstein et al., 2003, p. 373).

To address this problem, Goldstein et al. suggest that physical affordances be generously applied to device design to ensure effective utilization of device capabilities. Weiser’s (1993) paper on ubiquitous computing essentially agrees, emphasizing that the marriage between mobility and multimedia should be marked by a dramatic shift from the desktop paradigm and its ancillary metaphors to devices and interfaces that are already well known (cameras, MP3 players, etc). The reason is that such devices and the conventions they are associated with are easier to incorporate into the expected flow of human behaviors and actions, activity that we might term “natural.” One of Abowd and Mynatt’s (2000) observations in their review of ubiquitous computing is also pertinent here:
We desire natural interfaces that facilitate a richer variety of communications capabilities between humans and computation. It is the goal of these natural interfaces to support common forms of human expression and leverage more of our implicit actions in the world. Previous efforts have focused on speech input and pen input, but these interfaces still do not robustly handle the errors that naturally occur with these systems; also these interfaces are too difficult to build (Abowd & Mynatt, 2000, p. 42).

These requirements are in tune with the emerging research agenda that emphasizes the physical or “embodied” aspect of technological design and use (Dourish, 2001). Brereton and McGarry (2000) have drawn our attention to the fact that people habitually think in terms of objects, even when they design devices that do not require direct physical manipulation to affect the device’s functions. This idea is congruent with what Dourish (2001) describes as the “embodied” nature of technologies. This concept addresses the fact that technologies exist and are interacted with in the physical world in specific contexts and with specific expectations for immediate and facile manipulation. In this process, users prefer tangible (physically manipulable) controls and rely on intuitive use, primed by context and by the task at hand, with “intuitive” here meaning “according to the manner in which we understand affordance theory to suggest that humans use reality and the objects within it.” As already mentioned above and will be detailed below, more intuitive affordances are also easy to learn and degree of “intuitiveness” translates into enhanced “learnability” (Nielsen & Mack, 1994).

From this perspective, the future belongs to small information appliances that maximize, not minimize, external affordances, says Dourish, who envisions adaptive interfaces that rely directly on the physical appearance of the objects. For example, picking up a multimedia mobile device with both hands, and holding it lengthwise makes it ready for taking a picture, while grasping it with one hand and taking it to the ear triggers the cell phone functionality. It is true that users ultimately learn new paradigms of manipulation and control, but it is also true that such learning is difficult and not optimal when existing affordances insufficiently developed.

Although Dourish’s research stretches into the future, his emphasis on the need to turn back to physical controls and physically intuitive devices is a fruitful change of focus for understanding and designing more usable multipurpose mobile devices. In this new light, issues of scale, input/output methods, autonomy, and connectivity as they are relevant to the new generation of portable platforms, represent a universe that is yet to be fully explored. Furthermore, Dourish’s work makes it easier to understand why the new generation of mobile multimedia devices, with their new combinations of on-screen and direct physical affordances can be challenging even for expert users.

Getting beyond calls for simplicity: Finding the right balance between physical and on-screen controls

Typical problems and ideal parameters for designing multipurpose multimedia computing devices have been discussed in the literature by Norman (2002) or Oulasvirta and Saariluoma (2004). Norman argued that the human mind has a restricted working memory, not being able to handle more than four or five items at a time. Hence, external mnemonic aids are critical to support cognition and to prevent human error. Furthermore, features in complex systems should be made as visible as possible to reduce information overload, while functionality and feedback should be clearly visible in the interface. Or, even better, as Raskin (2000) suggested, features should be “detectable,” i.e., the user should be able to pick out the features that he or she needs from the multitude of functionalities available at any particular time. A delicate balancing act is required to achieve this detectability. Emphasizing some features at the expense of others might make certain features unavailable for unsophisticated users. At the same time, an excess of visibility makes the device “gadget-ridden and feature-laden,” which can intimidate even experienced users” (Norman, 2002).
This tendency of devices to appear “busy” is particularly acute in the context of multipurpose devices, which combine the functionality and versatility of many individual technologies, each involving its own complexities. For example, a multipurpose device tasked to capture, enrich, and transfer moving and still pictures would combine some of the more basic features from current digital video and photo cameras, PDAs and desktop computers. Collectively, these devices include over 20 physical affordance points. For example, a typical digital camera has seven contact points located in two areas of the device, a camcorder can have as many as eight contact points located in three areas, a PDA may have six contact points, located in two areas, and the PC may have four major contact points, in three areas. Even if this collection of physical affordances could be reduced or consolidated, it would be difficult to squeeze them within the confines of smaller mobile device. Confronted by this challenge, designers may prefer to hide and then flexibly make functionalities available through a range of GUI solutions. However, for many users, these solutions may prove to be suboptimal, since the main way they recognize a specific functionality is through its corresponding physical affordance (Figure 1).

The multipurpose device studied in this paper is a good illustration of such dilemmas. Capable of taking pictures and videos, group voice calling, and device-to-device media transfer, the device relies on screen interfaces that borrow heavily from the feel and iconic vocabulary of the personal computing and PDA environment. This puts the device squarely in the situation described above by Goldstein et al. (2003). Manipulating a camera phone through an on-screen interface is in latent conflict with the familiar operation and the mental model of clicking a shutter button and with holding the camera with both hands. Goldstein et al suggest that although on-screen controls seem to be more efficient from a usability perspective, they can in fact make devices more confusing and difficult to use. In what follows,

*Figure 1. A multipurpose device that can capture, store, and transfer images would combine about 25 physical affordances currently employed by digital cameras, camcorders, PDAs, and PCs.*
after we will position our paper in the broader field of mobile device usability research, we will discuss the physical-affordance-related design issues that emerged in the prototyping process of a mobile, multifunctional device.

**APPROACH AND PREVIOUS RELATED WORK**

The present paper reports results of a paper-prototyping and early usability study, informed by methodological principles recommended by Nielsen and Mack (1994), Norman (1998; 2002), and in broad theoretical terms by Dourish (2001). The paper builds upon or is related to a class of empirical studies that aim to explain how physical interfaces can impact adoption, use and the evolution of multifunctional devices from stand-alone tools into gateways for a web-based world of multimedia content and services (Brereton & McGarry, 2000; Goldstein et al., 2003; Kangas & Kinnunen, 2005; Jin & Ji, 2010; Heo, Ham, Park, Song, & Yoon, 2009; Marti & Schmandt, 2005; Vivrou & Kabbasi, 2002).

The empirical/mixed methods study reported here aims to determine to what degree multifunctional mobile devices need or not physical affordances and how our findings might change future design decisions. Our study borrows a number of principles and common metrics from usability research. Of the five quality metrics of usability proposed by Nielsen, we focus on three: learnability (How easy is it for users to accomplish basic tasks the first time they encounter the design?), error or failability (How many errors do users make? How severe are these errors?), and satisfaction (How satisfied are the users with the design?). As we will explain below, in our study learnability and failability are operationalized as questions about how “intuitive” the interface is (easy to learn and subsequently operate), while satisfaction is dealt with separately. Because our study evaluates a paper prototype whose functionality was simulated and was limited in time to a single research activity, the other quality attributes of usability research recommended by Nielsen, assessment of efficiency (How quickly can users perform tasks?) and memorability (When users return to the device after a period of not using it, how easily can they reestablish proficiency?), were not included in our study.

Our study extends current research on physical affordances in that it addresses specific usability issues (learnability and satisfaction) utilizing in-depth thematic analysis of data provided through cognitive walk through. Although some research on the right balance between on-screen and physical affordances has been conducted, notably Wheatly (2007) or Faiola and Matei (2010), in depth, mixed methods studies on this topic are not very common. For example, some of the most recent research on physical affordances are still at the stage of determining the broadest outlines of a usability framework for judging physical affordances in mobile devices (Heo et al., 2009; Jin & Ji, 2010). Their goal, to delineate some common metrics and workflow, is mostly orientative and general-heuristic. Compared to them, our study proposes a specific research framework with measurable parameters and actionable recommendations. Our study adds to the research space of physical affordance a specific thematic-analysis methodology, which includes an empirical framework for detecting critical issues. This and the specific methods utilized in our work are described below.

**RESEARCH QUESTIONS**

We use an X1 mobile multimedia device study to investigate the relative value of on-screen and physical affordances. The study was designed and conducted in 2004 in collaboration with the device manufacturer. Its most immediate aim was to support the design process. The broader, more theoretical goal of the study is to better illuminate how physical affordances in multifunctional mobile devices may remain or increase in importance in the design and use of such devices. Specifically, we examined how GUI controls, as opposed to hardwired buttons and input mechanisms, facilitate or hinder user
understanding of the multifunctional nature of the device. The study was driven by a set of research questions, which were designed to tap into the major issues discussed in the literature review. Specifically, the questions are informed by Nielsen and Mack (1994) and aim to explore basic usability dimensions related to device learnability and satisfaction:

1. Is the concept of a multifunctional device positively evaluated by participants?
2. Would participants actually use such a device?
3. Are the affordances primarily embedded within the on-screen GUI interface sufficient for suggesting the functionality and utilization of the device?
4. Do the on-screen affordances facilitate or hinder the use of multimedia functions?
5. What previous conventions and associations did participants use in integrating the device into their use repertoire?
6. Does the level of technological sophistication affect the manner in which participants discovered and understood device functionality?

Method

A user study was carried out to explore reactions to an early prototype of an X cell phone capable of media capture, streaming, and transfer. The primary method of delivering the treatment was through one-on-one interview sessions, in which the seven task scenarios described below were carried out.

Participants

There were 23 participants between the ages of 18 and 28 (12 female and 11 male, average age 20.7 yrs). They were recruited through announcements distributed through the student class registration and grade notification accounts. Students saw the announcements when they checked their grades or registered the classes. Subjects were recruited from a pool of 71 individuals, of which 23 made and came to a research appointment. They represented a variety of majors (Engineering, Liberal Arts, Management, Hospitality). One was a graduate student, four were juniors, eight sophomore, and ten seniors. Each student received financial compensation for her/his time. Although not a fully random sample, the respondents reflected the group of young, relatively technically literate individuals, who could become one of the target user bases for this device.

Treatment

To illustrate the functions of this future application in a way that participants could understand and relate to, a Microsoft PowerPoint storyboard scenario was used for the study, which hypothesized the context of use and included specific functional tasks. The scenario was animated and presented using a laptop computer. It contained seven discrete sequential user tasks, which were carried out by the participants using three paper prototypes of device interfaces. It should be noted that to reduce the possibility of “priming” the subject, the scenarios were not device-specific, but mentioned only a generic type of task that is generally performed with the type of device being investigated here. The scenarios referred to core device capabilities, especially its multifunctional features. They aimed to capture basic user reactions to core features, such as capturing and moving media across platforms and sharing between users. These were considered core and novel capabilities, which involved a new approach to control and interface design.

The prototypes consisted of a basic foam-core block, representing the approximate size and format of the device, onto which interchangeable screens could be placed in response to the participants’ operations and menu selections on previous screens (Figure 2). The paper prototypes were based on an X cellular handset, which used touch screen and stylus operation.

Procedure

Participants were asked to perform the tasks by tapping on various screens or buttons. The tasks
were not device specific, but activity specific. People were shown life scenarios, such as the need to call two other people simultaneously. For each action they were presented with different versions of paper screen interfaces. Paper screen prototypes were specifically chosen so that participants would perceive them as being very early in the development process and would therefore be more willing to provide both positive and negative feedback to influence the development. They were also hand drawn, rather than computer generated, to help reinforce this early stage, developmental feel. If the prototype and screens were perceived as being more finished, participants could feel that their feedback would be less influential in the development of the application and consequently they might be more reluctant to be critical.

The task scenarios specifically probed user values associated with the following functional capabilities, for which physical device affordances were not necessarily obvious:

1. Group voice calling, providing the capability to simultaneously initiate a multi-way voice call between three or more people.
2. Group presence information, providing visual information relating to the availability and geographic location of call recipients prior to the initiation of a voice call.
3. Wirelessly sending and receiving media files concurrent with a group voice call, providing the capability to share still images and short video clips within a multi-way call without interrupting the voice call.
5. Transfer of media from the cell phone to a nearby device, providing the capability of transferring a picture or streaming video from the cell phone to a fixed display device such as a PC or TV.

In addition to performing the tasks, participants were asked to comment on their actions during execution, using a “think-aloud” technique, and also to answer a number of open-ended questions about their understanding of the tasks, the prototype features, and additional contexts or situations in which these functions might be used. Furthermore, if a screen was not understood, or if one was felt to be missing, the subject was asked to sketch what they felt would work better for them, or to sketch out the missing screen. Participants were also asked to answer a series of open-ended questions related to their subjective evaluation of the group voice calling and multimedia capabilities of the device. The questions were open-ended to reduce bias. In order to determine that the participants were a representative sample, including highly technical and less technical users and both early and later adopters of technology, they completed short questionnaires that assessed...
their level of technological sophistication and speed of adoption.

The questionnaire for technical ability included two sets of questions. The first set referred to frequency of using personal computers, mobile devices (cell phone, iPod, digital camera, or PDA), and mobile/digital communication services (instant messaging, text messaging, or email). The actual question was: “How often have you used the following devices or services in the last month?” The response alternatives were: more than once a week, once a week, once or twice a month, and never in the last month. The answers allowed quantitative separation into groups on the basis of interest in using common technology. An implicit assumption we make is that increased use frequency leads to greater use facility. Frequency of use is often associated with digital literacy or technological sophistication; the more technologically capable users are also frequent users. Van Braak (2003) has found strong and significant effects for time and frequency of computer use and computer use competence. (For more on the relationship between frequency of use and technological ability, see Ballantine, McCourt Lares and Oyelere, 2007).

Answers to this set of questions were weighted on different scales, according to degree of complexity and relevance of the particular device to this study. Thus, frequency of use of most common devices and services (personal computer, cell phone, or email) was evaluated on a 0-1 scale (0 never, 0.25 once-twice a month, .5 every week, 1 more than once a week), while more sophisticated devices or services (cameras, instant messaging, or text messaging) were rated on a 1-3 scale (0 never, 1 once-twice a month, 2 every week, 3 more than once a week). Use of iPods and PDA was rated on a 0-5 scale (0 never, 3 once-twice a month, 4 every week, 5 more than once a week). This higher rating reflected a lower frequency of occurrence of these devices among participants, though for different reasons. The iPod was relatively new and expensive and not pervasive among the student population at the time of the study, while the PDA was becoming somewhat redundant due to the prevalence of PDA-like functions on other mobile devices, including cell phones.

Frequency of use offers only a very crude measure of technical ability. A second set of questions more directly captured technological ability, adding a qualitative dimension to the quantitative dimension of frequency of use. This second set of questions collected information on participants’ proficiency (speed) at adopting and using new devices. The literature on user adoption is extensive, and we do not propose to summarize all current models here, but it is well accepted that users fall into distinct groups of adoption patterns.

Four self-reported items were formulated as affirmative statements with which respondents were asked to agree, disagree, or be neutral. The statements were:

1. I am one of the first people among my friends or relatives to buy a new electronic device.
2. I like to customize the settings on my computer or electronic devices.
3. I rarely use the more advanced features on my cell phone, computer or electronic devices.
4. I learn faster than most of my peers how to use a new technological gadget.

The three answer categories were weighted using a score of 5 for agreement, 0 for neutral, and -5 (negative value) for disagreement.

A final index was constructed through item summation. $M=20.9$, $SD=11.5$, Median=22, $Range=1$ to 42. To facilitate further analysis, the sample was split at the median. Two groups, almost equal in size, were created, labeled as “high” ($N=11$) and “low” ($N=12$) technological sophistication respondents.

ANALYSIS

Video recordings of the “think-aloud” narrative and responses to open-ended questions were transcribed from all subject sessions. Transcriptions for each subject were unitized
into discrete items, one for each task (T), question (Q), or subject. EZAnswer, a qualitative analysis software developed by CDC (Centers for Disease Control) was used for managing and analyzing the data.

In addition to unitizing the data by task, question and subject a series of codes was generated for identifying discrete user reactions to the tasks and questions raised by the scenarios. The coding procedure relied on the Applied Hermeneutic Methodology (AHM) developed by Ross and Wallace, which was successfully applied in analysis of large-scale qualitative databases (Wallace, Ross, & Davies, 2003). The procedure emphasizes, in line with the hermeneutic tradition of qualitative analysis, the fact that inter-rater reliability is a social consensus process. AHM requires that the coding taxonomy be agreed upon by all the coders. When a suitable taxonomy has been agreed upon, a process of reading based on the ‘hermeneutic circle’ is formalized. The text is broken down into units and re-read a number of times by all the coders who then decide on the proper unitization and coding as a group. When a consensus has emerged and a common interpretation of the codes is achieved, the coding can continue either in a group setting or individually.

Eighty-four unique codes were generated through this iterative procedure. They reflected the most important reactions to the prototype identified in the data. Codes included generic and specific issues. They focused on actual use, immediate or intended, and on positive and negative reactions. For example, when evaluating user actions and reactions during the task that involved transferring media content to a nearby display, we used 4 codes, 2 for evaluation: “Respondent positively (negatively) evaluates content transfer to nearby device” and 2 for intention to use: “Respondent would (not) transfer content to nearby device.”

We applied the codes to a text unit, which comprised the outcomes of specific tasks performed and the answers to questions asked during each users’ interaction with the prototype. To each segment we allocated one or more codes by a panel of three coders, based on consensus voting and iterative reading of the session transcriptions as described above.

The present paper analyzes only the reactions to the tasks and the answers to the questions that directly referred to the multimedia/multipurpose capabilities of the device. These were as follows:

T3 Subject sends live video from a remote location.
Q6 Would the subject point camera at self or at surroundings while capturing video at a remote location?
T4 Subject transfers live video received on a cell phone to a nearby TV.
Q9 Would subject transfer media content to a nearby device if they had a device like the one evaluated?
T5 Subject sends photo to a group concurrent with a voice call.
Q13 What current multimedia sharing practices are used by the subject?
T6 Subject sends photo to a group listed in address book.
T7 Subject transfers received photo to a PC followed by a call back.
Q19 Is the device a good idea?
Q20 What do you think about the device?
Q21 What did you like about it?
Q22 What did you dislike about it?
Q23 What did you find most difficult?
Q24 What would you change in the prototype?
Q25 Open ended comments.

Lists of code frequencies for each question/group of questions or tasks were generated during the analysis and each list presented in descending order the number of participants that offered answers or comments that fit a specific code. The position of each code (issue) on the list was assumed to indicate its relevance and importance.

For example, Table 1 presents one of the “importance” lists generated during the analysis. It lists the codes applied while analyzing the question: “What did you like about the device?” (Q21). The figure indicates that the preference
for the “all in one” device is the most important issue found in the answers to this question, with 43% of participants indicating this preference, followed by preference for media sharing, media transfer, etc. Each task was similarly analyzed and relevant code rankings were generated.

Furthermore, 96% of the respondents positively evaluated the idea of transferring media content to a PC (2nd most salient issue).

**RESULTS**

**Subjective Evaluation of the Multifunctional Approach**

The first research question was aimed at determining whether the idea of a multifunctional device was evaluated positively by participants. The results indicate that the theoretical concept of a multifunctional device capable of media capture, storage, transfer, and voice calling was generally well received by participants. Analysis of all the answers to the questions and of the comments made while the tasks were performed indicates that 70% of participants positively evaluated the ability to exchange media content during an ongoing voice call (7th most salient issue). In addition, when asked to make open-ended comments about the device (Q19, Q20, and Q25), 66% positively appreciated its ability to share content during an ongoing voice call.

<table>
<thead>
<tr>
<th>Code</th>
<th>Issue</th>
<th>Number of respondents who mention the issue at least once</th>
<th>Percent of respondents who mention issue at least once*</th>
</tr>
</thead>
<tbody>
<tr>
<td>WFMULAY</td>
<td>Likes all in one</td>
<td>10</td>
<td>43.48%</td>
</tr>
<tr>
<td>WFMSAY</td>
<td>Likes media sharing</td>
<td>6</td>
<td>26.09%</td>
</tr>
<tr>
<td>WFMXAY</td>
<td>Likes media transfer</td>
<td>4</td>
<td>17.39%</td>
</tr>
<tr>
<td>MQUALAY</td>
<td>Video quality concerns</td>
<td>2</td>
<td>8.70%</td>
</tr>
<tr>
<td>PHYBOAY</td>
<td>Keyboard/hard-button related</td>
<td>2</td>
<td>8.70%</td>
</tr>
<tr>
<td>ICONHAY</td>
<td>Understands icon meaning</td>
<td>2</td>
<td>8.70%</td>
</tr>
<tr>
<td>PHYHAAY</td>
<td>Physical deployment related</td>
<td>1</td>
<td>4.35%</td>
</tr>
<tr>
<td>CORDSAY</td>
<td>Expects cords and cables</td>
<td>1</td>
<td>4.35%</td>
</tr>
<tr>
<td>SIMPLAN</td>
<td>Screens, menus sufficient</td>
<td>1</td>
<td>4.35%</td>
</tr>
</tbody>
</table>

Note=* percentages within each category; figures in parentheses represent absolute values.

The data indicated that participants would use such a multifunctional device, but that they would not use its entire array of functionality equally. Detailed analysis indicated that participants were much more likely to transfer a media file to a nearby device than to stream media in real-time. While all participants would use the file transfer feature, only 21% would send live video to other people or stream media. These reactions indicate that, although multimedia capabilities seemed like a good idea to the participants, these functions might not be immediately utilized when participants interacted with a cell phone. Moreover, from some of the verbal comments, it appeared that the participants’ inclination toward the most commonly used multimedia feature, i.e., media transfer to a nearby display device, was driven, at least in part, by obstacles which existed in transferring camera-phone media content to other devices. Many of the participants repeatedly asked, before performing the transfer.
procedure, if the transfer was to be done using wired connections. When told that this was not necessary, participants generally expressed relief and amazement that the procedure could be so easy to perform, implying that the current procedures were not as facile as those presented in the study scenario.

Intuitiveness of GUI-Based Affordances

The GUI-based affordances related to image capture and content transfer/sharing that were common at the time of this study (and were integrated into the experimental prototype) were not adequate to suggest the multifunctional possibilities (voice + media sharing/transfer) of the device. Although the GUI symbols used to indicate the various multimedia features of the device were quite conventional and straightforward—mostly icons and metaphors borrowed from desktop and web interfaces (e.g., camera, address book, handset, home, etc.)—40% of participants affirmed that they did not understand their meaning in the current context (while 40% indicated that they recognized the icons and 20% did not express an opinion). Thus they were not easy to learn and the interface partially failed the test of “learnability” (Nielsen and Mack, 1994).

This finding becomes more understandable in light of the fact that many participants expected to control the multimedia features of the device through hardwired buttons and controls instead of by selecting icons on a touch screen. Sixty-five percent of the participants raised various issues about the physical deployment of the device, especially its lack of clear function affordances. Their specific verbal comments frequently indicated that they expected a button for taking a picture or video, just like still image and video cameras. Other comments suggested a preference for a clamshell design, which would permit a larger usable area to accommodate both a larger LCD screen and a greater number of hardwired controls. Suggestions also included having hardwired buttons not only for video recording and image capture, but also for sending video, and using a physical cradle for synching/sending media with/to PCs or TVs.

An interesting comment, which indicated that current use patterns might impede the full and advantageous adoption of device multifunctionality, was that cell phones are often used in multitasking situations (e.g., shopping, walking, doing house chores, etc.). In such contexts, embedded multimedia or other controls in a touch screen display, which require the user to look at the screen, would make utilization of those functions very difficult. For many respondents, a multimedia, multifunctional device made sense only insofar as it allowed hardwired controls that facilitated multitasking. They indicated a preference for controls that can be found and operated by touch, using one hand, and without having to look at the screen.

These comments are best summarized by what is probably the most important finding of our study. Sixty-one percent of participants indicated that they wanted a hardwired button for controlling the multimedia capabilities of the device. Moreover, when specifically asked to indicate which feature they disliked the most, 52% indicated physical deployment issues, such as those related to buttons/control layout, number of screens or menus that users had to go through to accomplish a specific task, the size of the prototype, and the lack of hardwired buttons.

Intuitiveness of Multifunctional Affordances

As mentioned in the literature review, devices with too few affordances are insufficiently informative, while devices with too many are confusing. Therefore, it was important in this study to determine whether the multifunctional affordances led to confusion when it came to conceptualizing and using the multimedia functions. Or, using a term consecrated in usability research (Nielsen & Mack, 1994), the study aimed to find out if the multifunctional affordances are “learnable.”
Streaming live video wirelessly from a remote location was one of the central and most novel purposes of the device. Although this functionality was explicitly described to the participants, and the “live” nature of the sharing was also indicated in the on-screen menus, when asked to send live video, a majority of participants (87%) behaved as if they were sending a pre-recorded file (i.e., recording a video clip, then saving it, and finally sending it as a digital file). The fact that the task required a live feed, or that the GUI included functions that allowed live broadcasting, did not seem to be sufficient to define the live streaming capability. Not surprisingly, while carrying out the task a third of the respondents indicated that they were confused by the GUI.

The failure of many users to engage this feature clearly highlighted how a function can be ignored or misunderstood if lacking an obvious physical affordance. Many respondents indicated during debriefing that what was missing was a sense of “YOU ARE BROADCASTING LIVE NOW.” To make this functionality apparent, they asked for a prominent and clearly labeled “GO LIVE” on-screen button or, better still, to include an actual physical button for triggering the video capture and/or video transmission, just like the ones on video-cameras. The fact that the entire operation took place in a desktop-like GUI environment might have interfered with the mental model of live broadcasting. Users rarely capture streaming video on a computer, generally preferring to download pre-recorded video clips, and the “download first” stereotype appears to have been transferred to this mobile device.

Embedded conventions and mental models

The data also provided some insights into how previous conventions and mental models were used by participants for integrating the multifunctional device into their repertoire. Throughout the study, participants seemed to rely heavily on their previous mobile experience for making sense of the new device. Many indicated that they approached the device as a cell phone, the multimedia functions being seen as optional/added services. They understood the functionality of the device in terms of the cell phone sets they already possessed. Some of the participants even used their personal cell phones to show the interviewer what the prototype functionalities should look like. In addition, when confronted by the new capabilities, the participants often referenced other mental models: “this is like a camera-phone, right?” or “like an instant messaging program,” or “like a palm-top.” This strongly indicates that participants called upon, or referred to, multiple pre-existing mental models when trying to understand and integrate the device functionality into their use-repertoire.

Effect of Technological Sophistication

Data was collected to assess the level of technological sophistication of the study participants so as to determine whether it had any impact on the manner in which they discovered and understood the device functionality and took advantage of its advanced utilities. Utilization and understanding differences arising from level of technological sophistication were assessed by cross-tabulations. We used the dichotomized level of technological sophistication (low vs. high) as the independent variable, and presence/absence of mentions of key study issues by each subject (coded according to the system described in the data analysis section) as the dependent variables.

This analysis produced two relevant findings. Highlighting the importance of physical affordances in the context of multipurpose devices, more technologically aware respondents indicated in higher proportion than those with a lower level of technological awareness that they would prefer a physical button over an on-screen menu item for accessing certain device functions. More specifically (Table 2), one third (4) of the highly sophisticated, versus none of the low sophisticated, respondents expressed a preference for physical buttons and controls ($\text{Chi-squared}=5.3, p<.05$). In addition, 90% (10) of the highly sophisticated, versus half (6) of the low sophistication respondents, indicated
that they would prefer to have seen a clearly labeled “media” place or icon incorporated in the controls (Chi-squared = 4.5, p < .05).

These findings suggest that physical affordances and clear (on-screen or otherwise) indicators of multimedia device functionality become more, not less relevant, to users who are more technologically sophisticated. One possible explanation for this is that this category of users has a greater breadth of experience with a wide range of media and communications devices. Consequently, they may have developed, and called upon, a wider range of mental models or functional expectations, which, rather than being advantageous, could lead to greater confusion when learning a new device, particularly one that does not fit into one of these pre-existing mental models or device categories.

DISCUSSION

HCI researchers are increasingly in need of methodologies that can assist them in creating mobile devices that take into consideration the limitations of human cognition and the context of use. As Raskin (2000) holds, to truly be “responsive to human needs and considerate of human frailties” (p.7) design must conform to the natural limitations of human physical and cognitive abilities. The role human-computer interaction design plays in the design of complex multifunctional media devices will depend heavily on our understanding of the challenges associated with human-centered design and the tangible issues surrounding physical affordances (Faiola, 2006). Consequently, affordances and direct device controls are essential for fitting new capabilities and experiences into users’ old mental maps.

Findings from this study indicate that users who theoretically embrace a device’s multifunctionality are prevented from actually using its capabilities by an apparent cognitive constraint. Participants in the study preferred the multifunctional device, but could not always take advantage of its full range of features. Moreover, participants seemed to fall back on known metaphors especially related to file sharing, saving, and sending and mental models, sometimes from a desktop computing environment, which did not serve their purposes well in the context of the study. A recurring theme throughout the study was the need for more physical controls and hardwired buttons. This response emerged not only because the users had formed particular mental models and associated behaviors from their past use of other devices. They also indicated their positive preference for immediate access and use of the controls within a multi-tasking environment.

Table 2. Affordance related issues: Overall and specific (by technological sophistication) analysis.

<table>
<thead>
<tr>
<th>User Issue</th>
<th>Low sophistication N=12</th>
<th>High sophistication N=11</th>
<th>All subjects N=23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondents would prefer a physical vs. a screen control</td>
<td>0%* (0)</td>
<td>36%* (4)</td>
<td>14% (4)</td>
</tr>
<tr>
<td>Respondents would prefer a “Media Home” button or place among the device controls</td>
<td>50%* (6)</td>
<td>91%* (10)</td>
<td>70% (16)</td>
</tr>
<tr>
<td>Concurrent media sharing – positive evaluation</td>
<td>70% (16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media transfer – positive evaluation</td>
<td></td>
<td>96% (22)</td>
<td></td>
</tr>
<tr>
<td>Icons not fully understood</td>
<td></td>
<td></td>
<td>40% (9)</td>
</tr>
<tr>
<td>Usability issues; physical deployment</td>
<td>65% (15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respondents want hardwired controls</td>
<td></td>
<td></td>
<td>61% (14)</td>
</tr>
</tbody>
</table>

Note=* percentages within each category; figures in parentheses represent absolute values.
environment, i.e., one-handed operation with little or no visual demand. Ultimately, the participants clearly preferred a device that had more direct entry points to some basic functionality, which would make it easier to handle.

In part, our findings are validated by the development of devices that use controls, which involve physical gestures that mimic or rely on hardwired controls. In this respect, devices such as the Apple iPhone do not contradict, but rather validate our conclusions. The iPhone is considered by many HCI researchers a “convergent device,” because its multi-touch on-screen interaction style converges with affordances associated with physical controls (Anderson, 2008). The multi-touch screen, and the gestures that are associated with it, require users to physically interact with the device by pressing buttons on the display. These actions build on and involve gesture repertoires and conventions derived from physical affordances.

It is also important to note that while the iPhone has eliminated physical buttons (except for the one primary and second volume control), it has limited usage. The lack of a physical keyboard is a clear disadvantage when it comes to business applications. The Blackberry mobile device, with its keyboard and plethora of physical buttons, is preferred for sending and receiving email (Visiongain, 2009). Moreover, the iPhone has limited multimedia capabilities. At least presently, it does not support live video streaming. Hence, based on these existing limitations to current mobile devices, our findings need to be qualified. That is to say, when designing a new generation of multifunctional devices it is imperative to utilize gestures and tools that leverage to our embodied experience and that employ real or simulated physical affordances. That is, the specific choice for physical or onscreen controls should be decided based upon the nature of the task.

CONCLUSION, LIMITATIONS, AND FUTURE WORK

In summary, our findings suggest that the increasing portability of devices with greater computing power and multifunctionality will force interaction designers to rethink the current emphasis on abstract, screen based controls, derived from desktop computing. Although on-screen controls are necessary, due to the increasing capabilities and shrinking physical size of mobile devices, they can become cumbersome and obscure in a media rich device. Natural interfaces, which take advantage of existing behaviors, physical capabilities, and human-centered abilities, are more appropriate in this context. At the same time, on-screen controls should emphasize simplicity and should utilize intuitive icons.

These recommendations need, however, to be considered in the context of the limitations implicit in design of our study. Our study reports 6 years old data. A new generation of devices that use only on screen controls, especially the iPhone, seem to have put to rest the issue of physical affordances. Yet, as it is immediately obvious for any iPhone or iPad user, these devices do not eliminate physical affordances. They in fact turn the entire touch screen into a powerful touch, sweep, or pinch interface that acts like a physical affordance. The iPhone repertoire of gestures and icons associated with multimedia (e.g., camera controls) are far closer associated with physical devices than with computer screen metaphors, as it was the case for Windows Mobile devices of the pre-Windows Mobile 7 generation. Moreover, the staying power of physical affordances is suggested by the resilience of Blackberry and of a variety of Android phones that use various physical controls, including physical keyboards. In fact, the fastest growing smartphone brands are those that utilize multiple physical affordances, such as the Motorola Droid (Wireless and Mobile News, 2010). Our findings and interpretations...
are thus still relevant and can add a possible explanation for these developments.

Another potential limitation of our study is the small sample. In size and range of user experiences it offers a rather small segment of possible reactions and experiences. The size of the sample was dictated, however, by the in-depth research protocol, which involved about 10 hours of data collection and processing per individual. Paper prototype cognitive walk-through activities are powerful heuristic tools, yet they have inherent limitations, notably their inability to correctly represent system status and action feedback. These limits might have triggered exaggerated negative reactions to on screen features that could not be fully represented in a paper prototype and artificially stimulated our respondents’ declared need for physical affordances. However, the findings are too consistent and emerge in even straightforward use situations to completely warrant this alternative explanation. Another possible limitation of our study is the fact that the sample did not cover a very wide range of technological abilities. The finer differences we have proposed for groups of respondents that vary in terms of technological ability need to be interpreted as orientative indicators. The limited nature of this data also prevented us from deeper analyses in this direction. This and other issues need to be more and better tested on actual physical devices in a study that utilizes a larger and more diverse sample.

In the future, designers of multi-functional devices should investigate easy-to-use physical affordances, while optimizing their size, location, and interrelationship with groups of functions. Another key issue to be explored is users’ ability to configure new or revised mental maps of their functions, starting with their physical and outward appearance. Hence, next generation designers of mobile multifunctional media-rich devices should be cognizant of the emerging findings related to physical affordances in HCI and human behavioral studies. In this way, the complex issues surrounding content creation and delivery, context of use, accessibility of information, and universal access can be adequately addressed in the design and implementation of future mobile devices.

REFERENCES


**ENDNOTES**

1. X stands for a cell phone brand, which was anonymized for the review process.

2. Though the examination of device affordances was not one of the primary research motivations for this study, the acquired data suggested an inquiry into these issues. By utilizing basic paper prototyping rather than a tangible physical handset platform, participants were nonetheless able to conceptualize and predict some of the physical usage issues which would arise with a real physical device. At the same time, we agree that not all physical usage issues may be replicated in a paper prototype. Hence, we submit that it has sufficient physicality to test the applicability of affordances, more so than doing so virtually would offer.

Paper prototyping has long since proved its worth in both design and research applications. Prototypes have four distinct dimensions: breadth, depth, interaction, and look (Snyder, 2003). Breadth is the span of functionality presented, depth is the degree to which the real world is replicated in detail, interaction is the faithfulness with which interactivity is replicated, and look is the appearance. Of these, we agree with Snyder that breadth is the least important. We also agree that paper prototypes offer the best balance of depth and look when simply constructed and versatile. In addition, paper prototypes can reproduce affordances with reasonable fidelity, a key benefit for the present study.

3. Paper prototyping has long since proved its worth in both design and research applications. Prototypes have four distinct dimensions: breadth, depth, interaction, and look (Snyder, 2003). Breadth is the span of functionality presented, depth is the degree to which the real world is replicated in detail, interaction is the faithfulness with which interactivity is replicated, and look is the appearance. Of these, we agree with Snyder that breadth is the least important. We also agree that paper prototypes offer the best balance of depth and look when simply constructed and versatile. In addition, paper prototypes can reproduce affordances with reasonable fidelity, a key benefit for the present study.

4. Salience in this context signifies the relative ranking of the code among all the 84 codes generated through the analysis. Specifically, the code was the 7th most mentioned issue among the questions studied.