Chapter 15

Design and evaluation in the real world: communicators and advisory systems

15.1 Introduction

Textbooks about design and usability testing often make the processes sound straightforward and able to be followed in a step-by-step manner. However, in the real world bringing together all the different aspects of a design is far from straightforward. It is only when you become involved in an actual design project that the challenges and multitude of difficult decisions to be made become apparent. Iterative design often involves carrying out different parts of a project in parallel and under tremendous pressure. The need to deal with different sets of demands and trade-offs (e.g., the need for rigorous testing versus the very limited availability of time and resources) is a major influence on the way a design project is carried out.

The aim of this final chapter is to convey what interaction design is like in the real world by describing how others have dealt with the challenges of an actual design project. As you will have noticed, we have written primarily about design in Chapters 6–9 and evaluation in Chapters 10–14. This was to enable us to explain the different techniques and processes involved during a design project. It is important to realize that in the real world these two central aspects are closely integrated. You do not do one without the other. In particular, the main reason for doing an evaluation is to make progress on a design. Conversely, whenever you develop a design you need to evaluate it. Whether you are designing a small handheld device or a large air-traffic control system, a design that takes months to produce or one that spans years of effort, the two processes must be carried out together.

The chapter provides glimpses into the design and evaluation process for quite different types of interactive systems. The first two case studies discuss the design of mobile communicators for different groups of users, showing how the design issues differ for each group. The third case study examines the redesign of a large interactive voice response system. In the original design, the focus was on developing a system where the programmers used themselves as models of the users. Furthermore, the programmers were more concerned with developing elegant programs than with users' needs for easy interaction. As you will see, this caused a mismatch between their design and how users tried to find information. This is a common predicament and interaction designers are often brought in to fix already badly designed systems.

The main aims of this chapter are to:

- Show how design and evaluation are brought together in the development of interactive products.
- Show how different combinations of design and evaluation methods are used in practice.
- Describe the various design trade-offs and decisions made in the real world.

15.2 Key issues

As we have stressed throughout, user-centered approaches to interaction design involve iterative cycles of design-evaluate-redesign as development progresses from initial ideas through various prototypes to the final product. How many cycles need to take place depends on the constraints of the project (e.g., how many people are working on it, how much time is available, how secure the system has to be). To be good at working through these cycles requires a mix of skills involving multitasking, decision-making, team work and firefighting. Many practical issues and unexpected events also need to be dealt with (e.g., users not turning up at testing sessions, prototypes not working, budgets being cut, time to completion being reduced, designers leaving at crucial stages). A design team, therefore, must be creative, well organized, and knowledgeable about the range of techniques that can be brought into play when needed. Part of the challenge and excitement of interaction design is finding ways to cope with the diverse set of problems confronting a project.

A multitude of questions, concerns and decisions come up throughout a design project. No two projects are ever the same; each will face a different set of constraints, demands, and crises. Throughout the book we have raised what we consider to be general issues that are important in any project. These include how to involve users and take their needs into account, how to understand a problem space, how to design a conceptual model, and how to go about designing and evaluating interfaces. In the following case studies, we focus on some of the more practical problems and dilemmas that can arise when working on an actual project.

We present the case studies through a set of questions that draw out a number of key issues for each project. For example, mapping a large number of functions onto a much smaller number of buttons is key for mobile devices; understanding a child's world is key when designing for children; evaluating the current system is key when redesigning any large system.

15.3 Designing mobile communicators

The first two case studies are about the design of mobile communicators. They focus on some of the design decisions and trade-offs that need to be made. We describe example design practices at two companies, Nokia and Philips, highlighting the differences in requirements and design methods for what is seemingly a similar device.

15.3.1 Background

Mobile communicators often combine the functionality of a mobile telephone, a PDA, and a desktop computer. They allow the user to send and receive email and faxes, to make and receive telephone calls, and to keep contact details, diary entries, and other notes. They are an example of new devices that try to push technological boundaries while at the same time being accessible to a wide range of users. A key design challenge, therefore, is how to make such everyday devices usable and affordable to a heterogeneous set of users. Related to this set of usability goals is the decision about which design approach to use. As you are aware, there are many different approaches to choose from, ranging from ethnographic to more analytic methods. Here, we examine the different approaches of the two companies. To put you in a "design" frame of mind, we begin by asking you to consider the requirements for this kind of device.

ACTIVITY 15.1 In Chapter 7, we introduced a number of different kinds of requirements: functional, data, environmental, user, and usability requirements. Which of these is particularly relevant to the design of a communicator?

Comment All these are relevant in the design of mobile communicators, but one that needs particular attention is environmental requirements. Because the device is aimed at users "on the move" in all kinds of places, the environment in which it should work or its "context of use" is very variable.

Core environmental issues include how to make the device small and light enough to be carried around in a pocket or small handbag. This means the device must be made of light materials and should be physically small, and also the software must be designed to work with a small screen and limited memory. The system must allow for a whole range of situations: noisy or quiet, well lit or poorly lit, hot or cold, wet or dry, vibrating or still, and so on. These constraints have implications for the use of audio, for the levels of display lighting, and for the physical robustness of the device, among other things.

Another consideration in the design of this kind of communication device is what the users are doing when using it. A typical user is likely to be doing something else at the same time as using the communicator. This may be walking around, avoiding obstacles, looking for traffic, etc., or it may be listening for a train announcement or a call from children. So users are trying to combine at least three things: communicating with the device (talking, typing, or whatever), performing the "external" activity (walking, listening, etc.), and operating the device. This creates quite a high cognitive load, so operating the device should occupy as little attention as possible.

Tasks are very likely to be interrupted by external events, so users need to know where in an interaction sequence they are at any time, and be able to restart the sequence after an interruption. For a mobile communicator designed to access the Internet, this raises an interesting design trade-off: how long should a communicator remain connected to the Internet after activity has apparently ceased? A balance is needed between disconnecting so as to minimize connection costs, and remaining connected in a stable state to allow the resumption of an interrupted task. The best option may be to let users set their own time-out period, but this adds to the complexity of operation.

Another implication of the fact that users are likely to be doing other things in parallel with operating the device is that the communicator may need to be operated with one hand, or indeed in a hands-free mode. For example, someone who is walking down the street carrying a bag when the phone rings needs to be able to respond without stopping and putting the bag down, i.e., the operation needs to be one-handed.

For mobile devices in particular, tasks tend to be time-critical, *ad hoc*, triggered by other people or events, relatively brief, low in terms of attention to be applied to the task, and very personal. Because of these characteristics, the flow among tasks must be smooth. It seems that easy transition between contact database, telephone, and calendar is particularly important for mobile devices. The nature of these tasks and the environmental requirements for mobile devices have implications for evaluation, as we discuss in section 15.3.2.

Because this device will be mobile it must be simple to use and not involve much training. It also needs to be robust and reliable, as the user is most likely to be away from any significant technical support.

15.3.2 Nokia's approach to developing a communicator

So how does Nokia deal with these kinds of requirements? And which design and evaluation methods do they use? Here, we look at an example approach of Nokia's, and some of the key decisions in mobile communicator design. A design example of an existing Nokia communicator is illustrated in Figure 15.1. This communicator weighs 244 g, is $158 \times 56 \times 27$ mm, and has a full-color screen. As well

Figure 15.1 The Nokia 9210 communicator.

as email and high-speed WAP connections, it also runs a variety of office applications including word processing, spreadsheets, and presentations.[1]

This case study is based on material from Väänänen-Vainio-Mattila and Ruuska (2000).

What kind of lifecycle does Nokia use? Nokia follows a user-centered approach to concept development that includes contextual design techniques. They point out that "one clear strength of the methodology is that it makes ethnographic research manageable in a business environment" (Väänänen-Vainio-Mattila and Ruuska, 2000, p. 197). As discussed in Chapter 9, the "rich" descriptions arising from an ethnographic study are often not in a form that can be readily translated into a design specification. Nokia tries to get around this problem by carrying out ethnographic studies in combination with other methods. This enables them to come up with a set of detailed requirements.

Figure 15.2 shows a top-level model of Nokia's approach. It has four main steps:

1. The cycle begins with data gathering. The data is collected through market research studies, data from previous projects, and contextual techniques.

---

[1]Description summarized from information on the Nokia website *www.nokia.com*, as of February 2001.

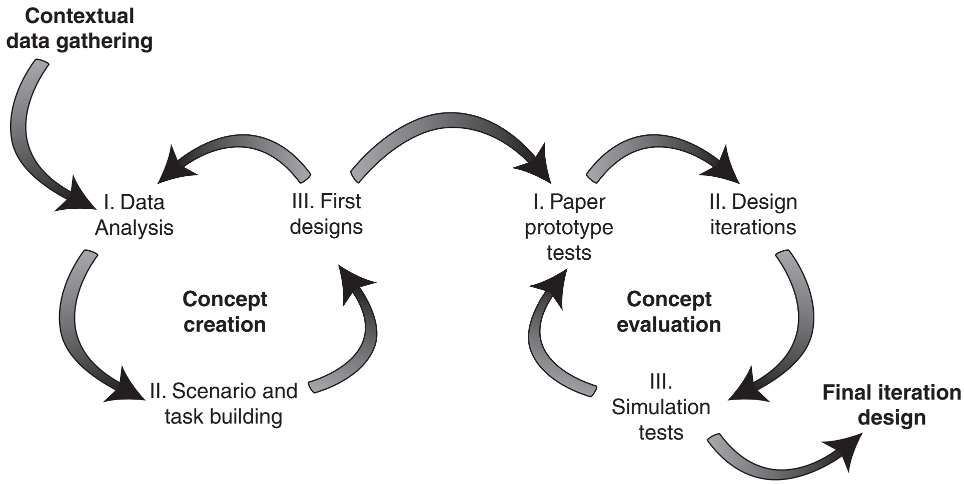

Figure 15.2 The user-centered concept and product development cycle.

2. Scenarios and then task models are built by analyzing the data collected, and initial designs are proposed.
3. Many iterations of design and evaluation are performed before the final design emerges. During this process, it may be found that more data is required, so further data gathering is conducted. The evaluation involves contextual interviews with paper-based prototypes to get feedback on first designs, and usability testing once the design is sufficiently advanced. Evaluation sessions emphasize the most important user tasks, as determined by the data gathering.
   - Once the design is advanced enough, high-fidelity simulations of the design are constructed.
   - Simulation tests are conducted with end users, and expert reviews are performed. Functional prototypes are tested with end users for feedback on long-term acceptability, efficiency, and utility of the concept.
4. During the last iteration phase, the final design is tested with end users and expert usability specialists.

ACTIVITY 15.2 How does this cycle of activities differ from the interaction design model introduced in Figure 6.7?

Comment This cycle also has a focus on iteration through prototyping and evaluation, which is the basis of the model in Chapter 6. However, this cycle distinguishes between concept creation and concept evaluation. Scenarios and task modeling are used at the concept creation phase but simulation tests are used in the concept evaluation phase.

What challenges does this approach raise? Nokia is very conscious of the need for iterative design and evaluation in the development of mobile communicators. They also use participatory design to a degree, but they point out that users will not necessarily have the vision of future possibilities that would allow innovative design in the same way as they might if asked to help design a familiar application like a web browser. Nokia is also well aware of the challenges of evaluating an innovative product like a communicator. These include:

- The difficulty of testing in all possible scenarios.
- The difficulty of testing human communication practices, especially when developing innovative products that will encourage novel behavior.
- The difficulty of testing services that cannot all be known beforehand.

What happens when the product is new and there are no users to test? At Nokia, quick and effortless access to critical tasks is a key design driver, and usability tests are used to evaluate the flow of tasks that have been found critical for mobile devices.

In a competitive and innovative market, other evaluation challenges may also arise. For example, consider the original Nokia communicator (the N9000). This was the first of its kind on the market. This had implications for how it could be evaluated because the device could not be shown to people outside the development team for fear of losing the "first-in-the-market" advantage. Thus the first version on the market did not have the benefit of testing with real users. Although extensive paper-based prototyping and simulations were produced, the evaluations were limited to a small group of people.

What methods does Nokia use? Nokia uses a number of methods in its development cycle, in particular "usage scenarios." Usage scenarios are high-level descriptions of uses of the device, based on data collected from representative stakeholders. They differ from the generic scenarios described in Chapter 7 in that they focus specifically on concept creation and high-level design considerations. An example of a usage scenario developed by Nokia is given in Figure 15.3.

What do design teams do next once they have created a set of scenarios? At Nokia, the design teams use the usage scenarios they have developed to identify critical user tasks and their structure. These task descriptions, which are more detailed than the original descriptions provided in the usage scenarios, are then used to consider lower-level design issues. A sample critical user task is shown in Figure 15.4.

ACTIVITY 15.3 To create scenarios, appropriate tasks and stakeholders will need to be identified. Who would the stakeholders be, and what techniques might be used to investigate their needs?

Comment First, the tasks to be performed and the stakeholders who might be asked about requirements would have to be identified. Stakeholders for a mobile device include users, developers, telephone companies, computer hardware and software vendors, and their shareholders. At least in theory, a user may be almost any member of the population, but in practice, only certain sections of the population are likely to be users. Given the wide functionality of the communicator, the most likely users are professionals.

Example of a Usage Scenario

David works as a legal consultant in an international corporation. He uses a communicator daily for light note taking and communications as well as for his personal organization.

8 A.M. The working day starts with a multiparty conference call to Japan. He uses the communicator as a speakerphone to be able to type notes in it at the same time. At the end of the meeting, he sends everybody a copy of the notes via email directly from the communicator.

1 P.M. At the airport, he downloads all his new email messages to his communicator so that he can start working on them during the flight. On the plane there is always plenty of time to write answers to messages. While downloading, he views the communicator calendar for the day and remembers having promised to send his business card to a potential client. He does this while standing in line for boarding.

At his destination, he switches the communicator phone on, and it automatically starts sending the replies written on the plane. At the same time David can continue reading the rest of the messages.

2:30 P.M. His secretary back in London sends him a calendar reservation for the following week. David checks his calendar in the communicator and accepts the request. His communicator sends the confirmation automatically to the secretary and marks the appointment in David's calendar.

Figure 15.3 An example usage scenario.

If we assume that the user group is professional, then it is necessary to find out more about the tasks they perform. This could be done using questionnaires, interviews and observation, or focus groups, but there would be some other issues to consider. A professional who is constantly on the move will be difficult to track down. However, interviews and questionnaires can be administered in different settings such as at trade fairs where many professionals are all gathered in one place. This would potentially provide a ready audience, reduce travel expenses, and supply immediate responses.

Performing standard observations in an office has its problems, but observing someone on the move, in all the possible locations in which they might use the device, opens up a whole new set of issues. Mobile devices are intended to be used anywhere, so where are observations performed, and how closely can the participants be followed?

What usability and user experience goals are important in designing this kind of device? A mobile communicator would be expected to meet the normal usability goals that we have discussed before. But what about user experience goals? Personalization has been identified as significant in user satisfaction; however, a balance

User Tasks: Classification

(1) Done under pressure: very critical
(2) Done frequently: critical
(3) Medium frequency or medium pressure
(4) Not frequent or not done under pressure

Sample 1: User tasks in person-to-person voice communication

Call-making/in-call

(1) Making a call to an emergency number

(1) Answering a call

(1) Rejecting a call

(2) Making a call to frequently called numbers (usually 4–10 of them)

(2) Making a call by manually entering each digit

(2) Redialing a number/person

(2) Indication of being busy

(3) Making a call to semifrequently called numbers (e.g., a vet, hairdresser)

(4) Making a call to occasionally called numbers (i.e., numbers that are often called only once).

Phone book memory

(1/4) Saving a name and number [1 = very critical during a call]

(2/3) Recalling a name and number and dialing [2 = to a frequently called number]

(4) Editing a name and number

(4) Erasing a name and number

(4) Browsing the contents of a phone book, etc.

Sample 2: User tasks in text messaging

Sending

(4) Sending a text message to a contact in the phone book

(4) Setting a message center number, etc.

Receiving

(2) Reading and replying to a message

(2) Reading and calling back the sender

(3) Reading and erasing a message

(4) Reading and storing a message with a new name, etc.

Figure 15.4 Sample user tasks.

BOX 15.1 Designing an Interface with a Small Number of Keys

What would you do if you had to design a communication device that could accommodate a maximum of only 15 keys? The device has to support numerical and text input. How could you design the mapping of the 15 keys to the various kinds of operations proposed so as to support the range of user tasks identified?

As a minimum, the device will need an on/off switch, a switch for connecting and disconnecting to the network, and a mechanism for entering ten numbers, 26 characters, and the space. You may decide to omit punctuation and capital letters, although this will have implications for the usability of the device. One way in which the functionality can be achieved is described below

One key is a dedicated on/off switch, one key is a dedicated connect/disconnect key, and one function key toggles the keys from numerical to character input. Ten of the keys represent the digits 0–9 and two or three characters each (26 characters plus the space bar means that seven of the keys must represent three characters and three of them must represent two). This uses 13 keys, which means that the keys for on/off, connect/disconnect, and the function button can be made larger to distinguish them from the others.

Alternatively, if you want to include punctuation, then all ten keys could have three characters each (giving room for four punctuation marks), and if you want to include capitals then a 14th key might be used as a "shift" key. Remember, though, that if the device is to be operated with one hand then this must operate more like a "caps lock" key than a shift key.

Some devices let you choose a character using only two keys. One key is repeatedly pressed as it toggles through the character list, and the second is used to accept the choice when the required character is displayed. This design choice requires more time from the user and its suitability depends on the functions the device is to support.

must be struck between allowing flexibility and providing sensible default values so that users don't have to customize settings unless they want to.

Mobile communicators are intended to support users wherever they are, so they must be compatible with the users' lifestyles. Designers must therefore understand the design characteristics that make the communicator attractive to different user groups, and those characteristics that will vary from group to group. If we consider the users as business people, then the important user experience goals are likely to include being helpful, motivating, aesthetically pleasing, and rewarding. If we consider children, then entertainment and fun are likely to be more important, while for teenagers its physical appearance might be more significant.

How does Nokia design a communicator's physical aspects? Deciding how many keys to have and how to map them onto a much larger set of functions is a difficult design challenge in any mobile device (see Box 15.1). For example, in the Nokia 7110 mobile phone, the problem of limited keys and limited space was dealt with by providing softkeys with context-sensitive functions that change depending on where the user is in the interaction sequence. This allows the keys to perform different functions depending on the other contextual issues. The softkeys allow the user to do a variety of things, such as make selections, enter, edit, or delete text. The current label for each softkey is displayed at the bottom of the screen, near the relevant key. There is,

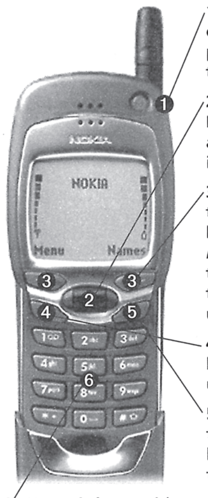

Figure 15.5 The Nokia 7110 mobile phone.

of course, a balance to be struck between having too many softkeys, each with limited functionality, and having only a few keys that can be overloaded with too many functions. In the end, the Nokia 7110 (Figure 15.5) was designed with just two softkeys that performed multiple functions. (Väänänen-Vainio-Mattila and Ruuska, 2000).

Textual input becomes a major problem when the number of input keys is restricted by the design. Having only a small number means the users must constantly "peck" at a few keys, typically using their thumbs. Trying to place too many keys in a heavily constrained space means that the user is likely to press the wrong key or two keys at once. How was this problem handled by Nokia? They opted for a small number of keys but in combination with a way of speeding up the typing of words, through having the communicator guess what the user is writing. In particular, the Nokia 7110 introduced the T9 predictive text method that allows speedy input of words based on a dictionary. The phone proposes a likely word once the user has typed a few characters. The user then either selects the proposed word and moves on to the next word, or rejects it and continues to enter the current word.

Communicators have also been designed to include a function button to let the user customize the interface to a limited degree, for example by allowing a favorite application to be associated with one of the hard keys.

BOX 15.2 Designing Telephones for the Elderly and Disabled

The British Royal National Institute for the Blind (RNIB), together with the British Department of Trade and Industry and British Telecommunications, have compiled a brochure to explain the different impairments affecting many telephone user groups, together with a set of suggested telephone features that could greatly enhance the accessibility of devices for such user groups. They identify 15 impairments and 44 features that could be added to telephones to make their use more pleasant. The impairments include cognitive impairment, weak grip, limited dexterity, speech impairment, hearing impairment, and hand tremor (Gill and Shipley, 1999). Features that could make a difference to these user groups include:

- Guarded or recessed keys to help prevent pressing the wrong key by mistake.
- Sidetone reduction, which reduces the amount of noise picked up from the environment and mixed with incoming speech at the earpiece.
- Allowing the user to adjust the amount of pressure needed to select a key. Apart from the more obvious consequences of too much or too little pressure, unsuitable key pressure may produce muscle spasms in some users.
- Audio and tactile key feedback to indicate when a key has been pressed.

Is it possible to design consistent interfaces, given the physical constraints of a communicator? A particular problem when developing software for a small display with limited input controls is how to make the interface consistent.

The design dilemma of consistency was addressed in Chapter 1. Consistency is often extolled as a virtue, yet it is sometimes appropriate to be inconsistent. In the design of communicators, the problems of consistency arise again. The device needs to have external consistency, i.e., consistency with users' expectations from their use of other similar tools, and also internal consistency, i.e., consistency with other items of software that the device supports. Sometimes these two design goals are in conflict, and it is appropriate to design a new solution for a particular situation.

The N9000 web browser was developed for the Nokia N9000 communicator. Many design decisions had to be dealt with, especially the problem of consistency (Ketola et al., 2000). Nokia has an internal style guide that all its products must follow in order to maintain internal consistency. External consistency with PC-based products is difficult to achieve because of physical constraints, and because the operating system for the N9000 is not commonly used with a PC. Other constraints on the design were:

1. The N9000 does not have a pointing device. Pointing is therefore done by selection using the scrolling bars. Scrolling down causes selection to jump from one hyperlink to the next; scrolling up causes it to jump to the previous link.
2. In cellular devices, connection rate is limited to 9600 bps, which is slower than the fixed-line rate. Connection can also take up to 30 seconds, considerably slower than the fixed-line equivalent. Web users may be accustomed to slow downloading times, but a long connection time is a new

phenomenon. A progress indicator was included in the design so that users would not become frustrated and start pressing other buttons. This leads to a further external consistency issue: should web pages be made to look the same as on faster desktop machines, or should they be designed for faster downloading?

Specific design decisions and solutions taken under these constraints were as follows:

1. The default page for a desktop web browser is a home page, but because of the connection time and the speed of downloading, the N9000 browser defaults to a list of favorite pages (called the Hotlist) instead. Thus, the default state is offline. This violates external consistency, but proved to be acceptable to users.
2. The functionality of the N9000 browser had to be carefully examined. Because of the Nokia style guide, only three buttons were available for navigating through the function hierarchy, so navigation became a major issue. To cope with the limited availability of command buttons, the N9000 employs the idea of *views*, within which only certain functions are possible. For the web browser, three views were provided: Hotlist view, Document view, and Navigation view. Users can select a document in the Hotlist view and enter the Document view. From here they are able to save, read, disconnect from the network, and close the document. However, they cannot navigate through the document. For this they need to go to the Navigation view. This conceptual shift was difficult for users to come to terms with.
3. The style guide dictated that the fourth command button be used to move upwards in the view hierarchy. It is also a part of the style guide that this button should be called "Back." In other applications this may not be a problem, but in the context of a web browser, a button labeled "Back" is interpreted differently. Internal consistency had to be obeyed here, and so the command that moved back to the previous page in the history list was called "Previous." This caused considerable confusion for users.
4. Optimizing web pages for display on mobile communicators involves the following three issues: content, because it's important to optimize download times; page layout, because of the small size of the screen; and navigation, because it's important to minimize the number of file downloads. User trials showed that, in the mobile context, users are more interested in getting the text information quickly than in downloading the graphics. Downloading unwanted pages also proved to be considered a key aspect of usability. Good link naming and clear, predictable behavior were important because of the long downloading times; locating the wrong page expends much time and cost.

ACTIVITY 15.4 If you are sitting near a desktop computer, study the interface of the piece of software that is running. If you are not near one, then think of the application you run most regularly on a desktop machine. Imagine what this interface would look like if you were to reduce the screen size to a mere 158 mm $\times$ 56 mm (the size of the Nokia 9210 communicator). What difficulties can you see? What implications do you think this has for software design, and also for the user who is swapping between desktop systems and mobile systems on a regular basis?

Comment If the same screen design is carried over to the mobile device then either everything will have to be miniaturized, so that the tool bars, icons and menus will become unreadable, or left at the same size, so that they will take up too much space on the screen. The interface therefore must be designed differently. This has implications for consistency for users who might be using the same application in a desktop environment and on the mobile device.

What kind of user testing does Nokia use? As mentioned earlier, there were confidentiality problems in testing the first generation of communicators on the intended user population. Hence, user testing could be done only after the product was released on the market. One kind of summative testing Nokia did was to find out what questions people have when first using the communicator. Users were given the device to use for some weeks and were then asked to report on positive and negative features. The results from this study confirmed the developers' concerns about the effects of consistency with other similar applications designed to run on desktop machines. Another study involved sending questionnaires to more critical communicator users whose experience ranged from 0 to 12 months, to find out if their reactions were similar.

As can be seen from this case study, Nokia uses a number of methods to develop their communicators for the general public. Furthermore, many design decisions and problems have to be dealt with, ranging from the lack of real users for testing, to how to let users send text messages with only a few keys and a very confined space.

15.3.3 Philips' approach to designing a communicator for children

We now consider how another company went about designing a mobile communicator aimed at a specific user group, children (mostly girls) aged between 7 and 12. Developing a tool for this user group is quite different from developing a tool for use by the general public, where there is likely to be a huge range of different users. An advantage of designing a device for a smaller set of users is that they are likely to have similar needs and preferences, meaning that the device can be customized much more to their requirements. This case study draws on material reported in Oosterholt et al. (1996).

Which approach did Philips use? The Philips process of development for this particular communicator made extensive use of prototyping techniques and participatory design. Children were involved from the initial concepts stage right through to final product testing. Each time a prototype was produced, it was shown to children for comment and feedback. A central part of the design process involved developing interface metaphors. Again, when ideas for metaphors were

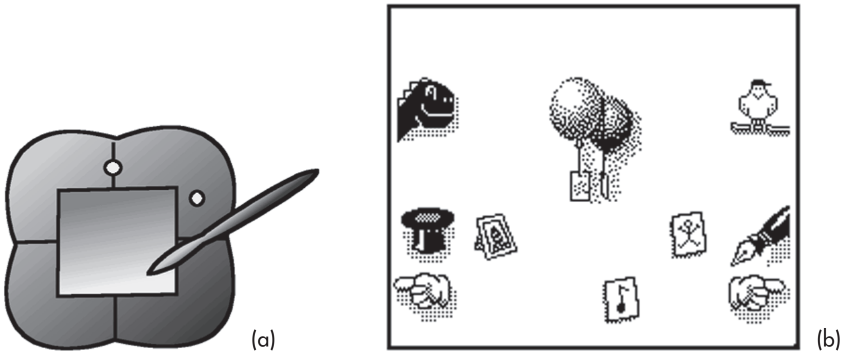

Figure 15.6 (a) The communicator with pen. (b) Product display showing 'the world'.

proposed, the designers turned to the girls in a spirit of participatory design in order to elicit their responses.

What usability and user experience goals were considered important? In the Nokia communicator example we saw the importance of usability goals focusing on effectiveness and efficiency, especially the need to move smoothly among critical tasks. In contrast, Philips focused more on the user experience goals of being enjoyable, entertaining, and fun. Other goals were that it should encourage creativity and provide personal and magical applications. The girls had expressed a specific desire for these.

What functionality did the communicator provide? The communicator was designed to have a touch-sensitive screen, pen input, infrared communications, and audio output (see Figure 15.6(a)). The interface was built on the metaphor of a world in which the users can move around freely, picking things up and starting applications (see Figure 15.6(b)). Available applications include a calendar, alarm clock, photo album, fortune teller, and communicator. The user can also perform tasks such as writing letters, composing tunes, drawing pictures, and sending them to other similar devices (see Figure 15.7).

What methods were used? Development of the product was divided into four phases: initiation, concept creation, specification, and finalization. Whereas Nokia adopted techniques from contextual design, Philips used mainly low-fidelity prototyping techniques for this particular project. Different prototypes were used throughout the development and for different purposes.

During the initiation phase, foam models were used to elicit feedback on the color, shape, size, styles, and robustness of the device, among other things. Using group discussions to encourage the youngsters to express their opinions a lot of feedback was gained from the foam models, even though the models contained no functionality. For example, children liked the idea of protecting the screen when carrying it, so they wanted different bags and cases to be provided for it; privacy was an important aspect, so they did not want it easily accessible by others; the pen should be stored safely within the device rather than underneath it for fear of it

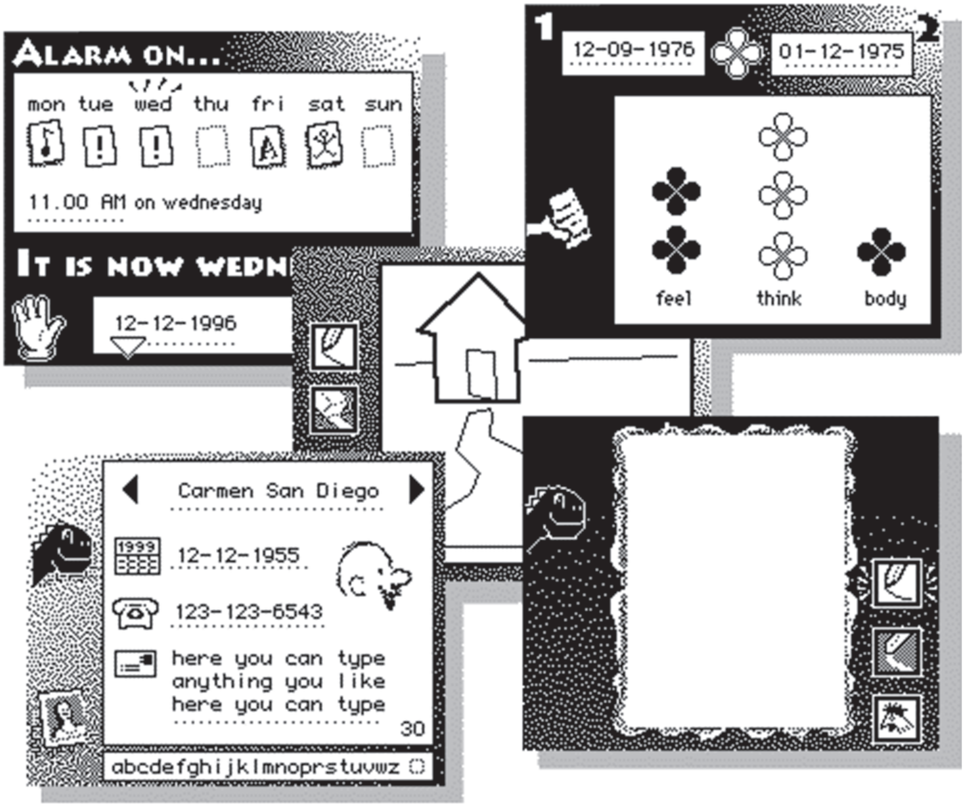

Figure 15.7 Some of the built-in applications.

being lost. One surprising result was that the children did not like the colors. The initial colors were bright (See Figure 15.8 on Color Plate 8), but they wanted dark colors more akin to their parents' hi-fi equipment at home.

The session with the models also provided input for the first user interface design, which was animated using a computer-based tool. This was used to explore navigation, pen-based dialog, types of application, and visual style.

During the concept creation phase, dynamic visualizations, which are like the storyboards described in Chapter 8 but are computer-based, were used to capture the initial ideas about interface and functionality (see Figure 15.9).

During the specification phase, foam models were again used to decide the size of the screen appropriate for writing on while standing up. As well as the size, different display formats were simulated (see Figure 15.10). These prototypes proved to be effective, again eliciting a lot of useful feedback. For example, left-handed users used the upper left part of the product to lean on while writing and the right-handed children used the lower right portion, yielding the design implication that the product should have hand resting places at these two points.

Also during specification, ideas for the interface design were evaluated by youngsters at a fair. There were two main contenders for the interface design.

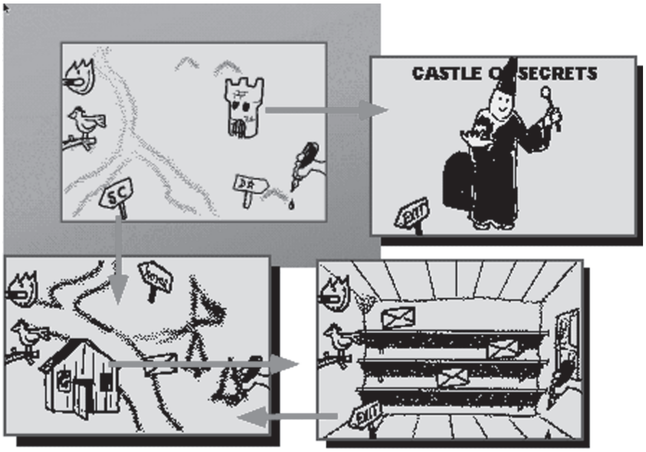

Figure 15.9 The first dynamic visualizations.

One provided direct access to each of the applications in the device, represented as a static matrix of options. This meant that the visual presentation and size of the applications was limited by the size of the screen. The other interface worked by indirect access, through a navigation model based on the idea of a window moving over a linked list of options.

Prototyping was also used in the finalization phase for market evaluations.

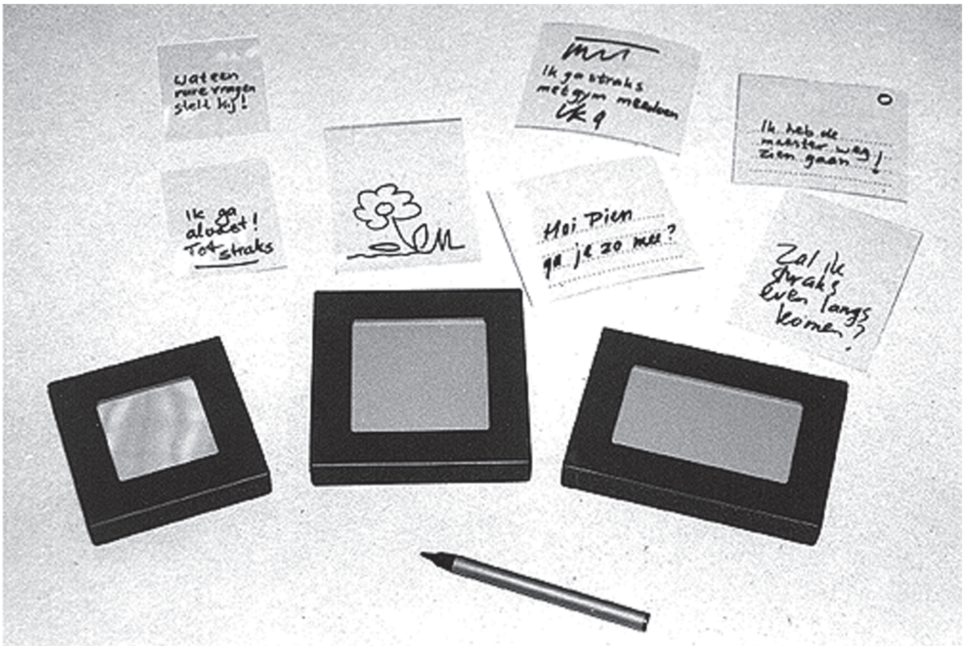

Figure 15.10 Foam models for investigating display size and screen format.

ACTIVITY 15.5 Prototypes are often used to answer specific questions. In this development, what questions were answered by producing and evaluating the foam models?

Comment Foam models were used at two specific points in the development to answer clear questions. The first set was used to consider the physical design such as size and color. They also elicited comments about storing the pen, covering the display, and having a carrying bag. The second set was used to design the display size and format. This also had the side effect of finding out useful information about where children would rest their hands on the device.

How much did the children participate in the design? One of the problems with participatory design is knowing how much to involve the users. Trying to involve children too much can be counterproductive, boring them and sometimes making them feel out of their depth. Asking children to participate too little can end up making them feel as if their views and ideas are not being sufficiently taken into account.

The Philips design team involved the children in design and evaluation from the very beginning. The first participatory design session was held during the initiation phase at a local international primary school. The session investigated the social and personal lives of 7 to 12 year-olds. Groups of 8 to 10 children were engaged in discussions and were asked to draw sketches of their ideal product. They were also asked to write stories about the use of the product, so that designers could get some contextual information about how it might be used. From this first session, it was clear that the concept was well received by the children. They particularly liked the communication, the pen-based interface, and its multifunctionality.

There were clear differences between boys, who wanted a broader range of functionality, and girls, who focused on communication. The ability to personalize was important to both groups. For example, one girl wanted the device to cough when a message arrived so that the teacher wouldn't know she was using it during class.

The whole design team was present at participatory design sessions. Spending time to get the children's opinions and to enter their world to understand how they perceive things was important for the success of the product.

One lesson that the designers drew from this exercise echoes a comment by Gillian Crampton Smith in the interview at the end of Chapter 6: users are not designers. In this instance, the children were limited in what they could design by what they knew and what they were used to. Another stakeholder group, parents, expected keyboard input, as they believed this to be more sophisticated than pen input, which was seen as old fashioned.

On the other hand, children are often more imaginative than adults, so involving the children was useful when discussing innovative ideas, or when only partial ideas were available. Working with children like this rather than adults requires a different approach, yet both adults and children need to appreciate each others' strengths and weaknesses. Box 15.3 describes the intergenerational design teams that Druin works with in projects at the University of Maryland.

BOX 15.3 Children and Adults Bring Participant Observation Close to Design

Allison Druin designs innovative technology with intergenerational design teams in which children and adults work together (Druin, 2000). In her teams children and adults observe children interacting with low-fidelity prototyping materials—crayons, pens, paper, glue, scissors, felt, furry cloth, Lego, animal parts, etc. (Figure 15.11), to explore ideas. By keeping more complex technology, such as computers, out of the picture during early observation and brainstorming sessions, adults do not dominate the scene.

Both adult and child members of Druin's teams observe and take notes while other children interact with the prototypes. This enables the team to capture impressions from both child and adult perspectives. Originally observations were recorded on a data-capture form like that in Figure 15.12 but many children prefer to draw and write simple notes like those in Figure 15.13. Adults, on the other hand, generally prefer to write, so now the team uses both techniques.

A typical observation session includes a pair of observers, an interactor and a child. The interactor's role is to ask questions that initiate discussion about the activities. Without this essential role, the children tend to feel that they are on stage being observed. But when engaged in discussion they are more likely to relax and reveal their real behavior and opinions. It is also important that the interactor does not take notes because this can make the children feel that they are being tested. Therefore, video is used to record observations.

Figure 15.11 Early design ideas for "fluffy-fuzzy" robots using low-tech prototyping materials.

RAW DATA:			DATA ANALYSIS:		
Time	**Quotes**	**Activities**	**Activity Patterns**	**Roles**	**Design Ideas**
10.05	E: Can you draw whatever you want? [Gustav: Yes.] E: (To K) A Christmas Tree? K: Yes!	K. takes the mouse rapidly, draws a red tree, takes the yellow crayon	Drawing	Artist	
		draws something in the corner, rubs out, continues	Drawing Erasing	Artist	
		E. tries to take the mouse	Struggling for control of input device	Leader	Multiple input devices
	E: But I want the long one! E: Noo! [Difficult to erase.] E: There. E: But what's this? [Windows Start menu appears.]	E gets the mouse, tries to get the blue crayon, looks irritated when she cannot get the blue one, gets it	Difficulty selecting tools	Frustrated User	Easier way to select tools

Figure 15.12 Excerpt from a data capture form.

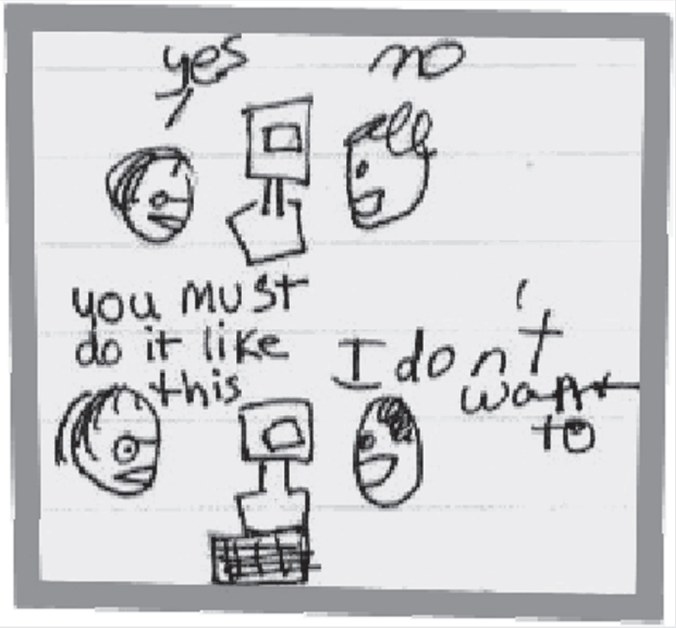

Figure 15.13 Sample notes illustrating a child's observation.

ACTIVITY 15.6 Suggest ways of helping adults and children feel comfortable together and gain mutual acceptance.

Comment Allison Druin asks everyone to dress casually in jeans, sneakers and T-shirts. The group works together at shared tables or on the floor. Snacks are important in creating a relaxed environment, and everyone uses first names. The goal is to create a group in which everyone respects each other's contributions and accepts and welcomes different contributions. Children are used to being controlled by adults and adults are used to being in control, and it takes time to break down these ingrained stereotypes.

What conceptual models did they design? By the concept creation phase, the importance of four goals for the product and its interface had emerged:

1. to support communication by stimulating social interaction among children
2. to evoke creativity and fantasy
3. to be "alive"—unexpected fun things should happen, surprising and pleasurable to the user, that give the product more character
4. to enhance intimacy—the product is a personal asset containing personal information

Five metaphors were developed by designers based on these values. Each metaphor was represented by a story. Figure 15.14 shows an illustration of one metaphor: the wizard. Specific metaphor workshops were conducted to find out how the girls reacted to the metaphors. They were asked to create a collage to visualize the metaphors, showing what they understood by them. The collages were a combination of drawings, essays, and existing pictures. The metaphor workshop showed that the girls were interested in being able to create, communicate, and organize personal things.

Figure 15.14 One of the metaphors: the wizard.

How did they evaluate the conceptual model? During the finalization stage, usability evaluations with children were performed to investigate the user interface itself and also to answer specific questions concerned with ideas for games, and writing performance. In most sessions, users were asked to play with the device for a certain period of time before giving feedback.

What lessons were learned from this case study? Many lessons were learned from developing an innovative product using a combination of participatory design and user testing. Some practical advice offered by Oosterholt and colleagues that can be generalized to the design of other interactive products is:

> *Specify Your User Requirements And Define Milestones* The rationale behind specifying user requirements is not just to develop them, but to make sure that the team agrees on the assumptions and realizes how and when they have been and can be changed.
>
> *A Product Is Not Designed in a Vacuum* Start thinking about additional and follow-up products at an early stage, so one does not have to change suddenly or add extra functionality in a later phase.
>
> *Users Are Not Designers* Not all answers can be generated by user or market tests. Users will generally relate any new product concept to existing products.
>
> *Act Quick And Dirty If Necessary* Often, the purpose of user testing is not to decide whether one interface concept is more usable than an alternative concept, but to discover issues that are important to the children. Small qualitative sessions of user involvement are therefore often appropriate. Furthermore, such sessions provide an opportunity for designers to "enter" the children's world.

15.4 Redesigning part of a large interactive phone-based response system

In this case study, we focus on quite a different kind of system, one being redesigned for a specific application intended to provide the general public with advice about filling out a tax return—and those of you who have to do this know only too well how complex it is. The original product was developed not as a commercial product but as an advisory system to be interacted with via the phone. We report here on the work carried out by usability consultant Bill Killam and his colleagues, who worked with the US Internal Revenue Services (IRS) to evaluate and redesigned the telephone response information system (TRIS).

Although this case study is situated in the US, such phone-based information systems are widespread across the world. Typically, they are very frustrating to use. Have you been annoyed by the long menus of options such systems provide when you are trying to buy a train ticket or when making an appointment for a technician to fix your phone line? What happens is that you work your way through several different menu systems, selecting an option from the first list of, say, seven choices, only to find that now you must choose from another list of five alternatives. Then, having spent several minutes doing this, you discover that you made the wrong choice back in the first menu, so you have to start again. Does this sound familiar? Other problems are that often there are too many options to remember, and that none of the options seems to be the right one for you. In such situations, most users long for human contact, for a real live operator, but of course there usually isn't one.

TRIS provided information via such a myriad of menus, so it was not surprising that users reported many of these problems. Consequently a thorough evaluation and redesign was planned. To do this, the usability specialists drew on many techniques to get different perspectives of the problems and to find potential solutions. Their choice of techniques was influenced by a combination of constraints: schedules, budgets, their level of expertise, and not least that they were working on redesigning part of an already existing system. Unlike new product development, the design space for making decisions was extremely limited by existing design decisions and the expectations of a large existing user population.

15.4.1 Background

Everyone over age 18 living in the US must submit a tax return each year either individually or included in a household. The age varies from country to country but the process is fairly similar in many countries. In the US this amounts to over 100 million tax returns each year. Completing the actual tax return is complex, so the IRS provides information in various forms to help people. One of the most used information services is TRIS, which provides voice-recorded information through an automated system. TRIS also allows simple automated transactions. Over 50 million calls are made to the IRS each year, but of these only 14% are handled by TRIS. This suggested to the designers that something was wrong.

15.4.2 The redesign

How do users interact with the current version of TRIS? The users of TRIS are the public, who get information by calling a toll-free telephone number. This takes them to the main IRS help desk, which is in fact the TRIS. The interface with TRIS is recorded voice information, so output is auditory. Users navigate through this system by selecting choices from the auditory menu that they enter by typing on the telephone keypad. First, the users have to interact with the Auto Attendant portion of the system—a sort of simulated operator that must figure out what the call is about and direct it to the proper part of the system. This sounds simple but there is a problem. Some paths have many subpaths and the way information is classified under the four main paths is often not intuitive to users. Furthermore, some of the functionality available through TRIS is provided by two other independent systems, so users can become confused about which system they are dealing with and may not even know they are dealing with a different system. Users get very few clues that these other systems exist or how they relate to each other, yet suddenly things may be quite different—even the voice they are listening to may change. Navigating through the system, with its lack of visual feedback and few auditory clues, is difficult. Imagine being in a maze with your eyes blindfolded and your hands tied so you can't feel anything, and where the only information you get is auditory. How can you possibly remember all the instructions and construct an accurate mental model in your head to help you?

Once in TRIS, users can take various paths that:

- Provide answers to questions about tax law (provided by one of the two other computer systems accessible through TRIS).
- Allow people to order all the forms and other materials they need to complete their tax return (provided by the two other systems accessible through TRIS).
- Perform simple transactions, such as changing a mailing address, ordering a copy of a tax return, or obtaining answers to specific questions about a person's taxation.
- Reach a live operator if none of the above options are applicable or the user cannot figure out how to use the system.

ACTIVITY 15.7 Why is developing an accurate mental model of TRIS difficult for users?

Comment Much of TRIS is hidden to the users. Their interaction with it is indirect, through listening to responses from the system and pressing various keys (whose meaning is always context dependent). There is no visual interface and users have only speech output to support their mental model development. Because speech is transient, unlike visual feedback, users must work out the conceptual model without visual cues. The user interface to this system is a series of menus in a tree structure and, since human short-term memory is limited, the structure of the system must also be limited to only a few branches at each point in the tree. Another problem is that TRIS accepts input only from the telephone number keypad, so it's not possible to associate unique or meaningful options with user choices.

What are the main problems identified with the existing version of TRIS? Because one of the main problems users have when using TRIS is developing a mental model of the system it is hard for users to find the information they need. In addition, TRIS was not designed to reveal the mapping of the underlying systems and often did things that made sense from a processing point of view but not from the user's. This is probably because the programmers took a data-oriented view of the system rather than a user-oriented one. For example, TRIS used the same software routine to gather both a social security number and an employee identification number for certain interactions. This may be efficient from a code-development standpoint, since only one code module needs to be designed and tested, but from the user's perspective it presented several problems. The system always had to ask the user which type of number was expected, even though only one of these numbers made sense for many questions being asked. Consequently, many users unfamiliar with employee identification numbers were not sure what to answer, those who knew the difference wondered why the system was even asking, and all users had yet another chance to make an entry error.

What methods did the usability experts use to identify the problems with the current version of TRIS? To begin with the usability specialists did a general review of the literature and industry standards and identified the latest design guidelines and current industry best practices for interactive voice response (IVR) systems. These guidelines formed the basis for a heuristic evaluation of the existing TRIS user interface and helped identify specific areas that needed improvement. They also used the GOMS keystroke-level modeling technique to predict how well the interface supported users' tasks. Menu selection from a hierarchy of options is quite well suited to a GOMS evaluation, although certain modifications were necessary to estimate values for average performance times.

What did they do with the findings of the evaluation? Once the analysis of the existing interface and user tasks was complete, the team then followed a set of design guidelines and standards, to develop three alternative interfaces for the Auto Attendant part of TRIS. An expert peer panel then reviewed the three alternatives and jointly selected the one that they considered to have the highest usability. The usability specialists also performed a further GOMS analysis for comparison with the existing system. The analysis predicted that it would only take 216.2 seconds to make a call with the new system, compared with 278.7 seconds with the original system. While this kind of prediction can highlight possible savings, it says little about which aspects of the redesign are more effective and why. The usability specialists, therefore, needed to carry out other kinds of user testing.

ACTIVITY 15.8 Why is it that the results from a GOMS analysis do not necessarily predict the best design?

Comment The keystroke-level analysis predicts performance time for experts doing a task from beginning to end. Not all of the users of TRIS will be experts, so performance time is not the only predictor of good usability.

The usability specialists did *three iterations* of user testing in which they simulated how the new system would work. When they were confident the new Auto Attendant interface had sufficient usability, they redesigned a subset of the underlying functionality. A new simulation of the entire Auto Attendant portion of TRIS was then developed. It was designed to support two typical tasks that had been identified earlier as problematic, to:

- find out the status of a tax refund
- order a transcript of a tax return for a particular year

These tasks also provide examples of nearly all of the user–system interactions with TRIS (e.g., caller identification, numeric data entry, database lookup, data playback, verbal instructions, etc.). A separate simulation of the existing system was also developed so that the new and existing designs could be compared. The user interaction was automatically logged to make data collection easier and unobtrusive.

What conflicts can arise when suggesting changes for improvement? When carrying out an evaluation of an existing product, often "jewels in the mud" stick out—glaring usability problems with a system that, if changed, could result in significant improvements. However, conflicts can arise when suggesting such changes, especially if they may decrease the efficient running of the system. The usability specialists quickly became aware that the TRIS system was making too many cognitive demands on users. In particular, the system expected users to select from too many menu choices too quickly. They also realized that immediate usability improvements could be gained by just a few minor changes: breaking menu choices into groups of 3–5 items; making the choices easier to understand; and separating general navigation commands (e.g., repeat the menu or return to the top menu) from other choices with pauses. However, to make these changes would require adding additional menus and building in pauses in the software. This conflicts with the way engineers write their code: they are extremely reluctant to purposely add additional levels to a menu structure and resist purposely slowing down a system with pauses.

ACTIVITY 15.9 The gap between programmers' goals and usability goals is often seen in large systems like TRIS that have existed for some time. How might such problems be avoided when designing new systems?

Comment It can be hard to get changes made when a system has been in operation for some time, but it is important for interaction designers to be persistent and convince the programmers of the benefits of doing so. Involving users early in design and frequent cycles of 'design-test-redesign' helps to avoid such problems in the design of new systems.

How were the usability tests devised and carried out? In order to do usability tests, the usability specialists had to identify goals for testing, plan tasks that would satisfy those goals, recruit participants, schedule the tests, collect and analyze data, and report their findings. Their main goals were to:

- evaluate the navigation system of the redesigned TRIS Auto Attendant
- compare the usability of the redesign with the original TRIS for sample tasks

Twenty-eight participants were recruited from a database of individuals who had expressed interest in participating in a usability test. There was an attempt to recruit an equal number of males and females and people from a mixture of education and income levels. The participants were screened by a telephone interview and were paid for their participation. The tests were conducted in a usability lab that provided access to the two simulated TRIS systems (the original design and the redesign). The lab had all the usual features (e.g., video cameras) and a telephone. Timestamps were included in the videotape and the participants' comments were recorded.

The order of the tasks and the order in which the systems were used was counter-balanced. This was done so that participants' experience on one system or task would not distort the results. So, half the participants first experienced the original TRIS design and the other half first experienced the redesigned TRIS system. That way, if a user learned something from one or other system the effects would be balanced. Similarly, the usability specialists wanted to avoid ordering effects from all the participants doing the same task first. Half the participants were therefore randomly allocated to do task A first and the other half to do task B. Taking both these ordering effects into account produced a 4×4 experimental design with eight participants for each condition.

ACTIVITY 15.10 Compare the description of this testing procedure with that for HutchWorld in Chapter 10. What differences do you notice and how can they be explained?

Comment The testing for HutchWorld is more typical. There were fewer participants and only one version of the system was tested at any time. In the TRIS test a larger number of participants were involved and the tests were more like an experiment. TRIS is complex, particularly the mapping between TRIS and the underlying functionality, although the system's purpose is clearly defined. By the time the usability specialists started the tests, they believed that they had fixed the major usability problems because they had responded first to the expert reviewers' feedback and then to the GOMS analysis. They were therefore confident that the new design would be better than the original one, but they had to demonstrate this to the IRS. This style of testing was also possible because there were thousands of potential users and the cost savings over 50 million calls justified the cost of this elaborate testing procedure.

How did they ensure that the participants tested were a representative set of users? In order to get demographic information to make sure the participants were representative, a questionnaire was given to all of them. It revealed a broad range of ethnicity, educational accomplishment, and income among the 18 women and 14 men who took part in the tests. Most had submitted tax returns during the last five years and most were experienced with interactive voice response systems. Eight participants indicated strong negative feelings about IVR systems, saying they were frustrating, time-consuming, and user-unfriendly.

What data was collected during the user testing? A total of 185 subnavigation steps made up the two tasks for the current TRIS. Participants successfully completed 91 steps on their first attempt (49% of the total). This was compared with a similar number of steps for the redesigned system: 187 subnavigation steps made up the same tasks for the redesigned TRIS. Participants were able to complete 117 of the steps on the first attempt (62% of the total), indicating an improvement of over 10%.

The average time to perform tasks was also analyzed. The summary data for the two tasks is shown in Table 15.1. As you can see, performance time on the redesigned system was much better for both tasks.

How was the user's satisfaction with the system assessed? At the end of each task, participants were asked to evaluate how well they thought the system enabled

Table 15.1 Average total task completion time by systems in seconds (s)

Task	Original system (s)	Redesigned system (s)
A	264.3	186.9
B	348.7	218.1

them to accomplish their tasks by completing a user satisfaction questionnaire. The responses again indicated that participants thought the redesign was easier to use and they preferred it. Regardless of the order in which participants used the two systems, the scores on the *redesigned* system were consistently much better than for the *original* system. The questionnaire provided statements that the participants had to rate on a 7-point scale. The difference between the two systems was highly significant, averaging over 3 rating-scale points higher on each statement.

ACTIVITY 15.10 User satisfaction questionnaires like the ones just described enable usability specialists to get answers to questions they regard as important. How can you make sure you collect opinions on all the topics that are most important to users?

Comment Asking users' opinions informally after pilot testing the questionnaire helps to make sure that you cover everything, but it is not foolproof. Furthermore, you may not want to increase the length of the questionnaire. Two other approaches that could be used separately are to ask users to think aloud and to use open-ended interviews. However, the think aloud method can distort the performance measures, so that is not such a good idea. Open-ended interviews are better, and this was done by the usability specialists in this case.

Participants were also invited to make any additional comments they wanted about the two systems. These were then categorized in terms of how easy the new system was considered to navigate, whether it was less confusing, faster, etc. Specific complaints included that some wording was still unclear and that not being able to return to previous menus easily was annoying. No matter how much usability testing and redesign you do, there is always room for improvement.

Would it have been better to redesign the entire system? It would have been far too expensive and time-consuming to redesign and test the whole system. A skill that usability specialists need when dealing with this much complexity is how to limit the scope of what they do and still produce useful results.

What other design features could be considered besides improving efficiency? Given that the system is aimed at a diverse set of users, many whose native language is not English, a system that uses different languages would be useful (the Olympic Messaging System used in the Los Angeles games did this very successfully). A range of voices could also be tested to compare the acceptability of different kinds of voices.

This case study has illustrated how to use different techniques in the evaluation and redesign of a system. Expert critiques and GOMS analyses are both useful tools for analyzing current systems and for predicting improvements with a proposed new design. But until the systems are actually tested with users, there is no way of knowing whether the predictions are accurate. What if users can theoretically carry out their tasks faster but in practice the interface is so poor that they cannot use it? In many cases, testing with real users is needed to ensure that the new design really does offer an improvement in usability. In this case study, results from usability testing were able to indicate that not only was the new design faster but users also liked it much better.

Summary

The three case studies illustrate how different combinations of design and evaluation techniques can be used effectively together to arrive at a design for a new product or redesign of an existing system. Quite different demands are placed on the design team when redesigning an existing product compared with designing a new product. Many practical problems and constraints will be encountered in both situations and experience of designing different systems will help you learn how to deal with them.

Key points

- Design involves trade-offs that can limit choices but can also result in exciting design challenges.
- Prototypes can be used for a variety of purposes throughout development, including for marketing presentations and evaluations.
- The design space for making changes when upgrading a product is limited by previous decisions.
- The design space is much greater when building new products.
- Rapid prototyping and evaluation cycles help designers to choose among alternatives in a very short time.
- Simulations are useful for evaluating large systems intended for millions of users when it is not feasible to work on the system directly.
- Piecing together evidence from data from different sources can provide a rich picture of usability problems, why they occur, and possible ways of fixing them.

Further Reading

BREWSTER, S., AND DUNLOP, M. (2000) (eds.) *Personal Technologies*. Special issue on Human Computer Interaction and Mobile Devices, 4, 2&3. This collection of articles discusses many issues in the design of mobile devices and would be a good starting point for anyone interested in pursuing this area.

BERGMAN, ERIC. (2000) (ed.) *Information Appliances and Beyond*. San Francisco, CA: Morgan Kaufmann. This book contains an excellent collection of practical articles describing how different information appliances have been developed, from interactive toys and games to a vehicle navigation system.

KILLAM, H. W. AND AUTRY, M. (2000) IVR interface design standards: A practical analysis. In Proceedings of HFES/IEA 44th Annual Meeting. This paper describes aspects of the TRIS study in more detail.