Layered protocols : hands-on experience

Int. J. Human–Computer Studies (1996) **44,** 45–72

Layered Protocols: hands-on experience

J. H. Eggen, R. Haakma and J. H. D. M. Westerink†
Institute for Perception Research/IPO, P.O. Box 513, 5600 MB Eindhoven, The Netherlands. email: rxh@philabs.research.philips.com

(Received 26 July 1994, revised manuscript accepted 2 October 1995)

An assessment is presented of the benefits and limitations of the Layered Protocols (LP) model for the analysis and design of user interfaces in the field of consumer electronics. In the assessment a user interface of an existing digital audio recorder which is only partly in line with the LP model is compared with an interface designed according to the model. The observed differences in usability between the two interfaces are mainly caused by deviations from the LP model. It turned out that especially the learnability of an interface is positively influenced by a layered organization of user-system interaction in combination with high-quality E- and I-feedback and optimum similarity between interaction protocols.

1. Introduction

User-system interaction models and user interface design guidelines can be useful tools for designing better conceived and more usable interfaces (Marshall, Nelson & Gardiner, 1987; HUSAT, 1988). They can be of help in structuring the interaction between the user and the system and in explaining the role of the information exchanged. Ways of increasing usability are indicated on the basis of implicitly or explicitly incorporated cognitive principles. In this way, validated interaction models and guidelines can help in reaching a basic level of usability while diminishing the need for extensive user testing.

In this paper the validity of a specific user-system interaction model is assessed: the Layered Protocols (LP) model (Taylor, 1988*a,* 1992. The assessment is carried out in the field of consumer electronics: digital audio recorders. The application area of consumer electronics differs from the professional setting with respect to training. Users operating a digital audio recorder have usually had no formal instruction or coaching. Therefore, more weight is attached to learnability than to efficiency as a usability criterion (Nielsen, 1993).

The LP model is based upon the cognitive principle that humans use superimposed layers of abstraction in perception and performance. From this principle the LP model arrives at an architecture for structuring user-system interaction. The model explicitly addresses the user's and the system's contribution to the interaction and the way in which they relate to each other. In this paper guidelines indicating potential usability consequences are derived from the model. That way the model can be validated.

The LP model was chosen because it seemed somewhat more applicable in consumer electronics than some alternative models. Other models like Cognitive

† The authors are listed in alphabetical order.

1071-5819/96/010045 + 28$12.00/0

Complexity Theory (Kieras & Polson, 1985) and Norman's stages and levels (Norman, 1984) also take a cognitive approach. However, the LP model is more explicit in the implications of the principles for structuring user-system interaction. The GOMS and the keystroke model (Card, Moran & Newell, 1983) are not appropriate because they emphasize efficiency of interaction. Learning is also addressed by Task Action Grammars (Payne & Green, 1986). However, its focus is on structuring the user's part of the interaction. The issue of feedback is disregarded. For a more elaborate discussion of this issue see Engel & Haakma (1993).

The LP model will be validated in the following way. First, the principal concepts of the LP model will be described and some guidelines indicating usability consequences will be derived from the model. After that, these concepts will be made operational in the analysis of an existing interface in terms of the LP model. On the basis of the results of this analysis some interface properties will be indicated as causes of potential usability problems. Next, it will be established whether the LP model is specific enough to guide user interface design by creating a new user interface for a functionally equivalent recorder. The design will be based on the LP model. Subsequently, the two interfaces will be evaluated with users. The results of the experiments will reveal the interface with the highest usability. Finally it will be discussed to what extent the usability differences are accounted for by deviations from the LP model and to what extent the LP model is found to be incomplete in covering user interface aspects influencing usability.

2. The Layered Protocols model

2.1. AN OVERVIEW OF THE MODEL

The layered Protocols model is based on the idea that interaction between communicating partners takes place in a series of layers that represent different levels of abstraction. Taylor (1988*a*) gives persuasive evidence for this claim. As a consequence the LP model organizes user-system interaction as a hierarchically structured framework of communication loops, each operating according to its own interaction protocols.

Figure 1 shows the elementary communication loop that forms the basis of the LP

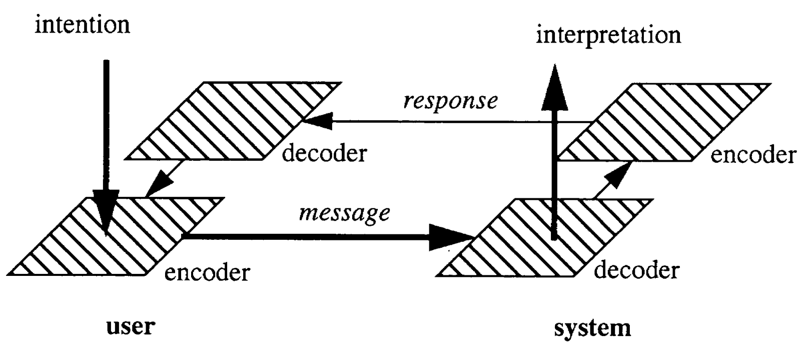

FIGURE 1. Basic communication loop of user–system interaction.

model. The interaction is started by a user with a particular intention in mind. In order to let the system know this intention, the user has to encode it into a message, for example by pressing a sequence of buttons. Upon receiving an action sequence the system decodes this message and arrives at an interpretation. This interpretation will cause the system to take actions aimed at fulfilling the user's intention. The system also provides feedback by encoding either its interpretation of the message received or the results of the actions taken. The communication loop is closed when the user interprets the system's responses and checks to what extent the original intention has been satisfied.

As a consequence of the observation that humans use layers of abstraction in perception and performance, the Layered Protocols model organizes user-system interaction in a hierarchical way (see Figure 2). On the user's side the incoming intention is encoded into sub-intentions at each level. Each sub-intention constitutes an incoming intention for the layer below. In this way the user's overall intention is successively encoded into sub-intentions until it has been transformed into a sequence of physical actions. The system decodes the message it receives in reverse order. At each level it will try to make an interpretation to the next level up on the basis of sub-intentions recognized in the layer below. Feedback is supplied on all (intermediate) interpretations at every level. Those feedback messages may be

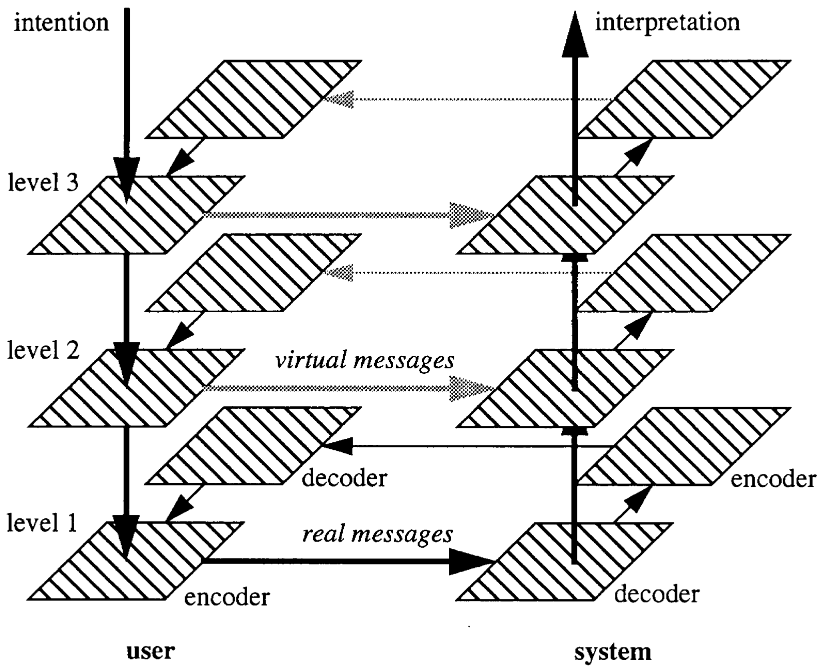

FIGURE 2. The LP model of user–system interaction. The transfer of an intention from user to system is indicated by thick arrows. Via successive stages the main intention is encoded into a message at the lowest level. The message is transmitted to the system via a physical communication channel. The system in turn successively decodes this message into the user's higher-level intention. Feedback at each level (thin arrows) enables the user to verify the system's interpretations at all stages. The grey arrows indicate the exchange of virtual messages between the user and the system in successive layers. A virtual message is considered to pass on the contents of a real message communicated by the underlying layers.

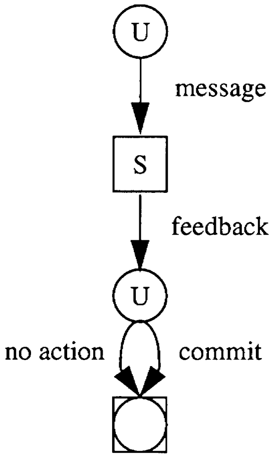

FIGURE 3. The core protocol structure. The User (U) sends a message to the system, the System [S] interprets the message and gives feedback on its interpretation. After that the User can take this interpretation for granted and proceed with the next message or an explicit commitment may be required from the user.

directly transmitted to the user (e.g. they may be fully displayed on a computer screen). Or they may be communicated in bits and pieces requiring the use of another (layered) protocol directed the other way, that is from system to user.

In order to facilitate effective interaction between the user and the system, the corresponding coding and decoding processes have to employ the same interaction protocol. Taylor (1988*b*) uses a state-transition diagram to describe a general protocol grammar. The core of this grammar, linking user messages to system feedback optionally followed by an acknowledgment, is shown in Figure 3.

This core grammar suffices when mutual agreement on the protocols is guaranteed. However, in order to set up robust communication in practice, interaction protocols additionally have to indicate how to deal with communication failures. Therefore Taylor's general protocol grammar indicates several expansions of this basic protocol structure to enhance robustness of communiction. The two most important extensions facilitate:

- adjustments—when the system's interpretation deviates from the user's intention in details, the user is offered the opportunity to adjust the interpretation;
- error and abort handling—for dealing with situations in which the interpretation has to be dropped, i.e. when the system's interpretation does not match the user's intention by far, when the system cannot carry out the user's intention or when the user has second thoughts.

In the original LP model the sole purpose of feedback is to provide users with information enabling them to check to what extent their intentions have been carried out. In Engel and Haakma (1993) it was recognized that users not only require information on the system's *interpretation* of their messages (I-feedback). They also need information on how to formulate their messages, i.e. on what information the system *expects* to receive from the user (E-feedback).

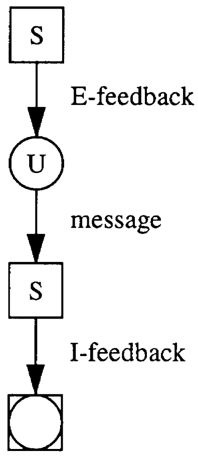

FIGURE 4. A protocol structure incorporating E-feedback. The System [S] provides the user with E-feedback on the options available, the User (U) composes a message on the basis of this information, the System interprets the message and gives I-feedback on its interpretation.

E-feedback assists users in coding their messages. E-feedback has the same layered structure as I-feedback because E-feedback is required for guiding the users in each coding step. E-feedback differs from I-feedback with respect to the moment it is needed: E-feedback is helpful before users code their intentions whereas I-feedback is needed after the message has been communicated. A protocol structure incorporating E-feedback is presented in Figure 4. The difference in timing between E- and I-feedback is clearly indicated.

2.2. GUIDELINES

The mere fact that an interaction model is based on some cognitive principle does not guarantee that it can be successfully used to guide user interface analysis and design. Confidence in the usefulness of the model in practical situations is increased when sensible user interface guidelines can be deduced from the model. Guidelines also make the model more operational. Only when usability consequences are indicated does it become possible to validate the predictions of the interaction model. Therefore this section lists some testable guidelines for structuring user–system interaction that can be derived from the LP model.

• *E-feedback*
E-feedback has to be provided at the various levels of the interaction. In this way the expectations of the system with respect to the message to be received from the user are reflected and users are thus assisted in coding their intentions. E-feedback should give rise to expectations on the part of the users on the system's interpretation of messages.

For example, at the lowest layer the presence of a button may invite users to press it, while the label of the button may give rise to higher-level expectations on the effect that pressing that button may have.

When E-feedback is lacking it is assumed that novice users initially have problems in predicting the effect of a message and in composing the appropriate one.

• *I-feedback*

The LP model stresses the importance of I-feedback. Every user action can in principle have an immediately observable effect. On account of the layered structure of the LP model the system's interpretation is immediately signalled back at the highest level possible: an interpretation is made as soon as the relevant information is available. The provision of immediate I-feedback at the successive levels enables users to detect unintended interpretations at an early stage. Without the layered feedback there would be a good chance of errors remaining unnoticed until later on, when it would be more difficult to correct them. Immediate I-feedback also facilitates learning because users can immediately relate their messages to the corresponding system responses.

Suppose that to a computer with a command-line interface "del ⟨file name⟩ ⟨CR⟩" has the meaning of delivering the file via email to some default address. When users with the intention of deleting some file type "del ⟨file name⟩ ⟨CR⟩", they would only notice the unintended effect after the command had been executed from the system response "Delivered ⟨file name⟩ to ⟨mail address⟩". If the system were to provide intermediate feedback, for example by showing a mailbox on the screen after "del_" has been typed and a trash can after "rm_", users would be able to detect the misunderstanding and initiate error repair at an early stage.

When I-feedback is lacking users are assumed to have problems in establishing the effect of an action. This may make it hard to learn how to operate the interface and to determine to what extent a particular intention has been satisfied.

• *Error and abort handling*

Error an abort handling facilities are required to prevent that situations arise in which users are trapped in an unintended situation and are forced to proceed in an undesired direction. The availability of these facilities at different levels enables users to correct parts of an interpretation selectively. Error and abort handling facilities can be offered explicitly to users, e.g. recording music onto an audio cassette can be interrupted by pressing the stop button. They can also be incorporated more implicitly by offering functions that undo each other's effects: e.g. in the case of a CD player, the effect of pressing the next button can often be neutralized by pressing the previous button.

When error and abort handling are not available at the various levels, it is more difficult to correct interpretations. Exploration of the system is also more difficult because users have to proceed in situations in which they want to backtrack and explore alternatives.

• *Similarity of protocols*

Each layer has its own interaction protocol that interprets the messages coming from the layer beneath. Therefore, the protocols at the different levels are only loosely coupled so as to enable independent protocol analysis: lower-level protocols can be designed and analysed more or less independently of higher-level protocols. Also, a lower-level protocol can support underlying interaction for several higher-level protocols. For example, in the case of audio cassette recorders the same interaction protocol is used for indicating the tape position at which playing should start and for indicating where a new recording should start.

When interaction is structured by employing the same underlying protocols for different tasks the internal consistency is increased, which makes it easier to learn how to operate the system.

In short, it is assumed that user interfaces designed according to the LP model will excel in early detection and repair of communication failures. Learning is facilitated because users can immediately link their actions to observable system responses and because of internal consistency. Guidance in the form of E-feedback assists users in coding their messages.

3. Analysis of an existing user interface

The LP assessment will start with the analysis of an existing digital audio recorder in order to see how well its interface can be described in terms of the LP model. The analysis will result in the identification of potential usability problems.

The analysis was performed using a commercial digital audio tape (DAT) recorder. The DAT recorder offers functionality that is characteristic for digital audio recorders in general. The frequently used features comprise playback and tape transport. The recorder also allows more complex operations, including programming and skipping tracks, recording and tape editing (e.g. merging two tracks).

3.1. RECORDING USING THE DAT RECORDER

In order to give an impression of what the interaction is like, a scenario will be elaborated in which a user has the intention of recording a new track from a CD player immediately after track number 3 and thus overwriting track 4. All the relevant buttons on the front panel and the indicators on the display of the DAT are shown in Figure 5.

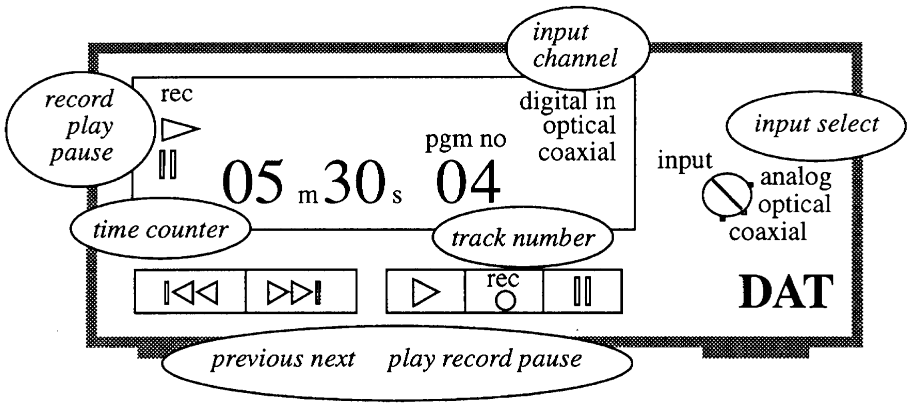

FIGURE 5. Impression of the front panel of the DAT recorder. Only those buttons, knobs and feedback indicators are presented that are relevant to the recording task described in the text. The record task requires the use of five buttons and a knob. The knob is used to select the audio input channel. The selected input channel is also indicated on the display. A time counter indicates the position of the head on the tape. The current track number is also shown. Three other indicators show whether the recorder is recording, playing or pausing. The words in italics are explanations not appearing on the DAT recorder itself.

After the tape has been inserted into the DAT recorder the user has to check on the basis of the indicators in the display whether the intended source has been selected, whether the tape head is at the correct position and whether the recorder is in record-pause mode. If this is not the case, these settings have to be adjusted.

If the tape head is not at the beginning of track 4, the user first has to transport the tape by repeatedly pressing the previous or next button. After one of these buttons has been clicked, the recorder starts making winding sounds, showing a changing counter and most often a new track number. Once the tape head has reached the correct position, the user has to press the record button and will then see the record and pause indicators light up. The DAT recorder is now in record-pause mode.

The user also has to select the correct input channel for the music using the input select button. The selected channel is also shown on the display. The correct input channel can be selected while the other adjustments are being made.

When all the adjustments have been made, the user can start the recording by pressing the play or the pause button. The recorder shows it has started recording by switching the pause indicator off and the play indicator on, while the tape counter starts increasing.

3.2. LP DESCRIPTION OF THE RECORDING FUNCTION

The interaction as described in the scenario can be structured in three layers, each characterized by its own protocol. Figures 6, 7 and 8 show the protocols involved.

The top-level interaction protocol is related to the overall goal of recording. It is a single-step protocol in which users have to pass the message "to record from a particular input channel starting at a particular tape position". This top-layer message is an interpretation of the bottom-layer message "the user has pressed the

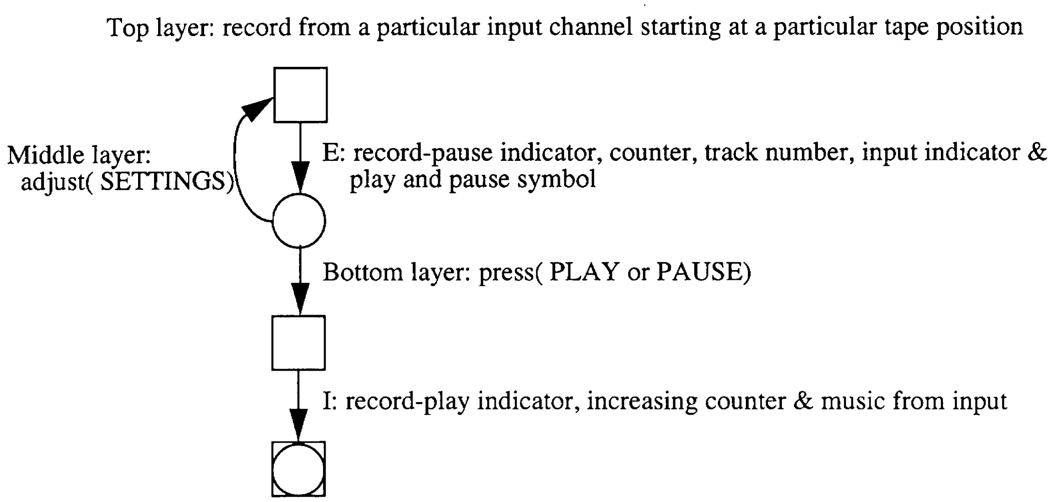

FIGURE 6. Top-layer interaction protocol showing how the message "to start recording from a particular source on a particular track" is constructed from lower-level messages, showing the E-feedback to get users to do it what way and I-feedback to reflect the system's response.

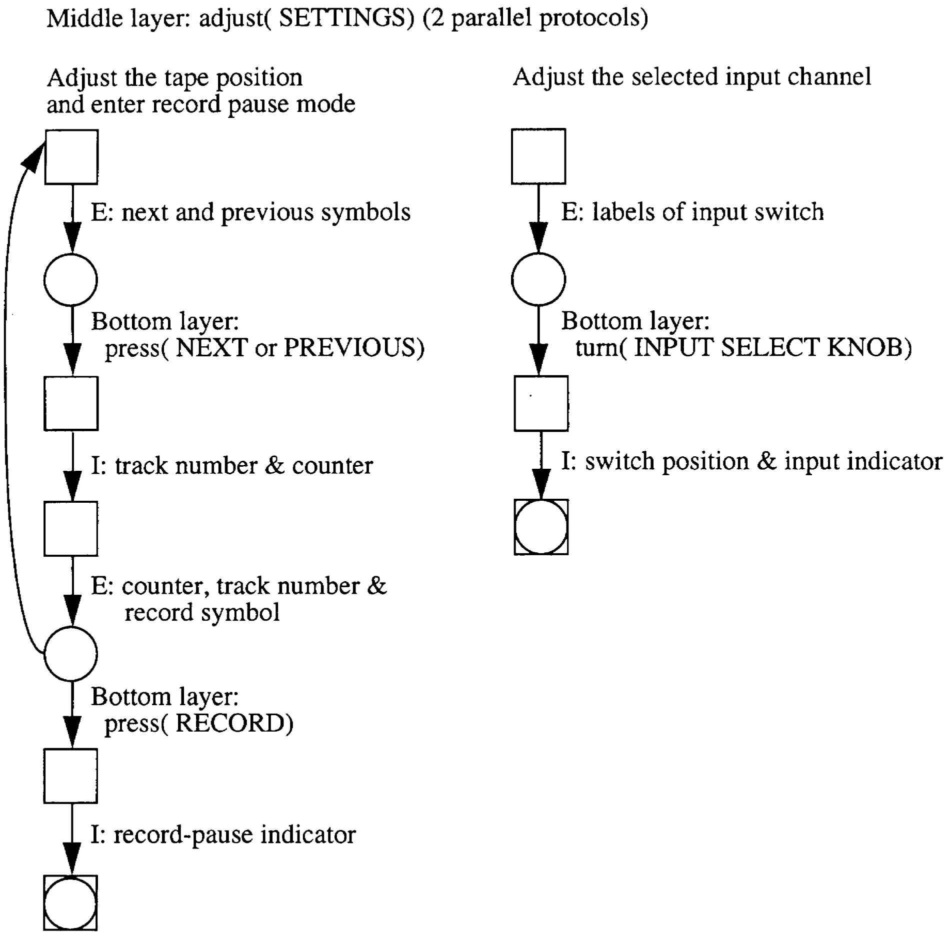

FIGURE 7. Middle-layer protocol for adjusting the recorder's settings.

play or pause button" in combination with some settings that can be adjusted using the middle layer's protocol. That the recording is in progress can be concluded from the increasing counter and the lighting up of the record and play indicators while the music being recorded can be heard. This information therefore functions as I-feedback. The settings that influence the precise effect of pressing the play or pause button are the input channel indicator and the tape head position indicated by the track number and the time counter. The recorder has to be in record–pause mode, indicated by the record and pause indicator, when the recording is started. All the information on the relevant settings in combination with the play and pause symbols are classified as E-feedback since it should make users realize that recording will start after the play or the pause button has been pressed.

The middle layer is concerned with adjusting the settings relevant to the top layer. This layer consists of two parallel protocols: one for changing the selected input channel and another for changing the tape head position and entering the record–pause mode.

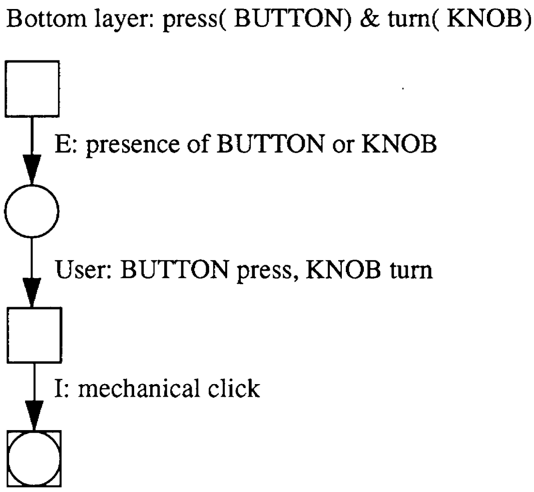

FIGURE 8. Bottom-layer protocol for pressing buttons and turning knobs.

The middle layer's message to change the input channel is an interpretation of the bottom layer's message that the user has turned the rotary input select knob. The selected input channel is shown on the display by the input channel indicator. This information corresponds to the position of the knob. The input channel indicator in combination with the knob position therefore serves as I-feedback. The labels around the knob indicate what input channel will be selected after the knob has been turned. This information therefore serves as E-feedback.

The DAT recorder requires that the tape head position be adjusted before the record–pause mode is entered. The middle-layer message to adjust the tape head position is an interpretation of the bottom-layer message that the user has pressed the next or previous button. The track number in combination with the counter indicates the current tape head position and therefore serves as I-feedback. E-feedback is provided by the previous and next symbols on the buttons. The middle-layer message to enter record-pause mode is an interpretation of the bottom layer message that the user has pressed the record button. The record and pause indicators serve as I-feedback while the symbols on the record buttons serve as E-feedback.

The third and bottom layer involves turning the knobs and pressing the buttons. It generates the bottom-layer message that the user has pressed some button or turned some knob the moment it detects that the user executes such a physical action. The presence of the knobs and buttons should make users realize that they can turn or press them and they therefore function as E-feedback. I-feedback is provided via the tactile and kinesthetic senses, in combination with a clicking sound of the button's mechanism. It should makes users aware that the button has been pressed or the knob been turned.

Apart from on the recording functionality, the LP analysis also focused on other functions, such as programming and tape editing. On the whole, it turned out to be complicated and laborious to describe the interaction between the user and the DAT recorder.

The fact that the DAT gives no I-feedback of a more abstract nature at the top level (e.g. no information is provided on the tape contents) made the analysis *complicated*. It would have been impossible to distinguish between the top layer and the middle layer by solely looking at the feedback supplied. In this analysis the way in which the user's part of the interaction is organized is used as a guide in distinguishing between layers. It then becomes clear that the top layer codes its feedback solely in terms used at the level beneath.

The work was also *laborious because* of the great number of different protocols used in the DAT interface. Almost each function has its own protocols, which differ from the protocols applied for other functions. The only major exceptions are the transport controls for tape-head positioning, which are used in both the recording and the editing functionality.

3.3. USABILITY CONSIDERATIONS

It will now be discussed to what extent the DAT recorder was found to follow the guidelines derived in Section 2.2. On the basis of these guidelines some potential usability problems will be indicated.

- *E-feedback*

E-feedback is offered in the form of button labels on the front panel. The system settings relevant for predicting the interpretation of a user message are shown on the display.

- *I-feedback*

The immediate I-feedback provided by the DAT recorder often consists of changes in the status information on the display. At the highest level no I-feedback of a more abstract nature is given about the effect of the current settings on the recording being made. Instead, the users have to infer these effects from the I-feedback given at the middle level. Therefore users may have difficulty in establishing whether the tape has been affected and if so, how. This can have the consequence that users are slow in learning how to use the recorder and in detecting communication failures.

- *Error and abort handling*

The protocols generally contain various possibilities of repairing errors and aborting the interaction. Functions that undo each other's effects are provided, e.g. "next" and "previous". Some explicit abort facilities are also provided such as a "stop" button.

- *Similarity of protocols*

There is not much similarity between the interaction protocols used for selecting different functions. The interaction is structured according to a great number of different protocols. This again may slow down users in learning how to control the recorder.

In short, it is assumed that learning how to operate the DAT recorder is slowed down by the lack of I-feedback at the top level and by insufficient similarity between the applied interaction protocols.

4. Design of a new user interface

A new user interface offering functionality equivalent to that of the DAT recorder was designed. The interface was a prototype for a Digital Compact Cassette (DCC) recorder. Like the DAT recorder, a DCC recorder is a type of digital audio recorder.

The DCC prototype was explicitly designed to be in line with the LP model. The interaction is structured in three layers. At the top level functions operate on the tape contents. To visualize this, the tape is graphically displayed in the lower part of the screen. By filling in a form, presented in the upper part of the screen, users can select the function that matches their intentions (see Figure 9). The middle layer provides the means for changing the contents of the form. At the bottom level the physical interaction, consisting in pressing buttons, takes place.

Three forms were designed, each addressing a different type of user intention: one form for *programming,* one for *recording* and one for *editing* functions. These different functions are selected by using the same underlying form-filling protocol.

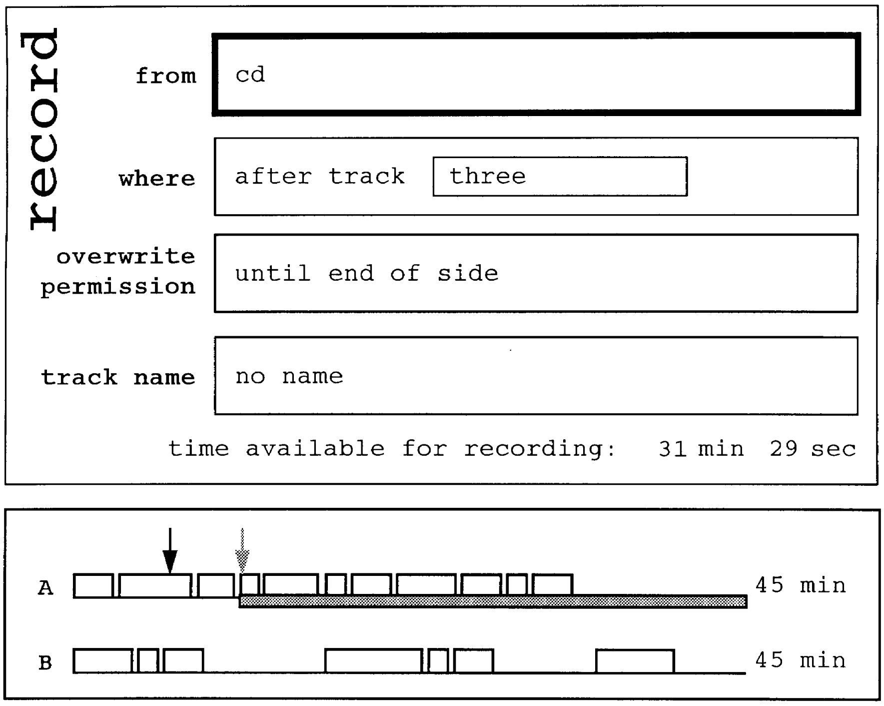

FIGURE 9. The screen of the DCC prototype. The upper part of the screen shows the form used for recording. One of the fields in the form is surrounded by a thick black frame. A graphical representation of the tape contents is presented in the lower part of the screen. The tracks on the two sides of the tape are presented as white boxes on two lines. The black arrow indicates the tape head position. Additional information on the interpretation of the form is supplied by the dark grey bar indicating what part of the tape is reserved for the recording, the grey arrow indicating where the recording will start and the field indicating the time available for recording.

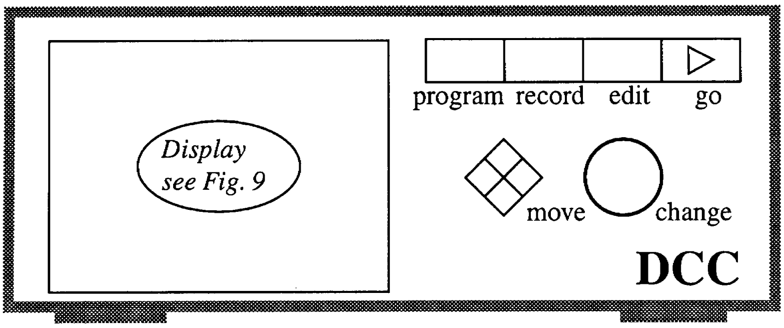

FIGURE 10. Impression of the front panel of the DCC prototype. Indicated are only those buttons and knobs that are relevant to the recording task described in the text.

4.1. RECORDING USING THE DCC PROTOTYPE

In order to give an impression of the interaction, a scenario similar to that of the DAT recorder will be elaborated (see Section 3.1): a user has the intention of recording a new track from a CD player immediately after track number 3 and thus overwriting track 4. All the relevant buttons on the front panel of the DCC prototype are shown in Figure 10.

First, the user has to select the record form by pressing the "record" button on the front panel of the digital recorder. The record form containing four fields appears on a display (see Figure 9). Next, the user may wish to change the contents of these fields in such a way that they correspond to the recording intention. With the aid of four cursor keys labelled "move" a thick black frame can be moved over the record form to indicate which field must be adjusted, e.g. the "from" field. The contents of the field indicated by the frame can be changed by turning the knob labelled "change". The user can feel every turn of the rotary knob and can see that the next or previous option is displayed in the selected field, e.g. from "radio" to "cd". In this way the user can scan all the options available for the field. The recorder gives indications of the interpretation of the form: a grey bar indicates what part of the tape is reserved for the recording and the time available for recording is shown to allow the user to check whether the reserved amount of space will be sufficient. After the record form has been adjusted, the user has to press the "go" button to start the recording. The system now carries out the instructions on the record form, starting with the necessary tape-transport actions. During the recording the graphical representation is constantly updated. A new track emerges on the tape while the grey record bar gradually disappears.

4.2. LP DESCRIPTION OF THE RECORDING FUNCTION

The interaction as described in the scenario is structured in three layers, each characterized by its own protocol. The protocols at the top and middle levels are shown in Figures 11 and 12. The protocol used at the bottom level is identical to the bottom-layer protocol of the DAT recorder (see Figure 8).

The top layer enables the selection of functions operating on the tape contents. This layer constitutes the highest level at which the system supports the users'

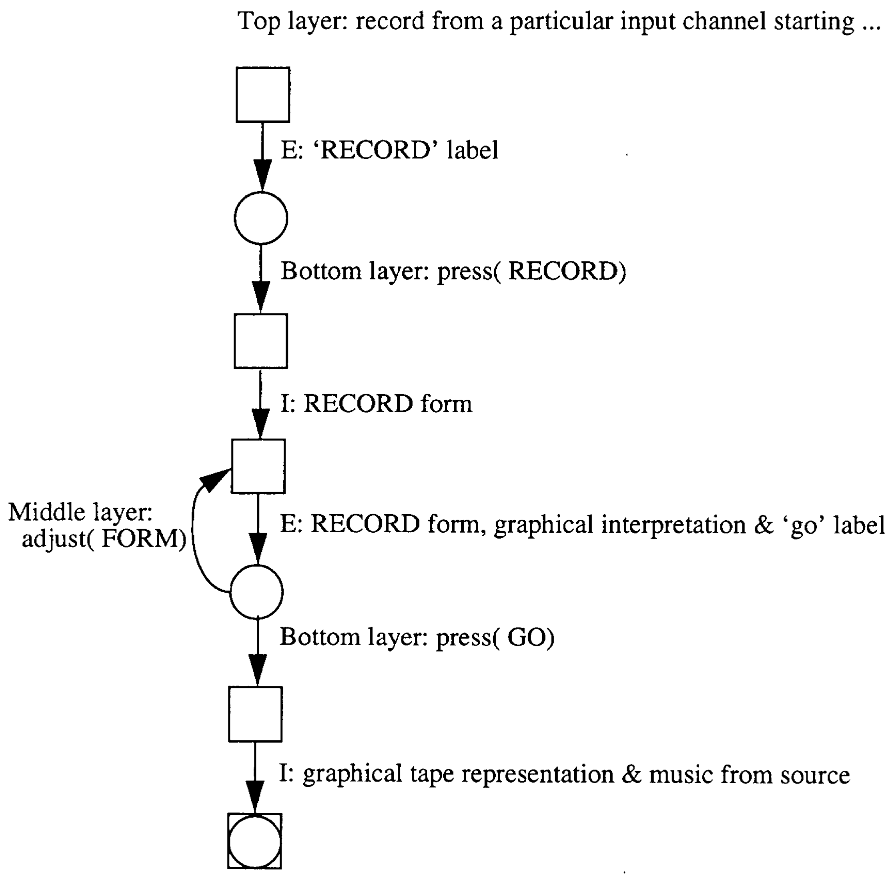

FIGURE 11. Top-layer interaction protocol for the recording function.

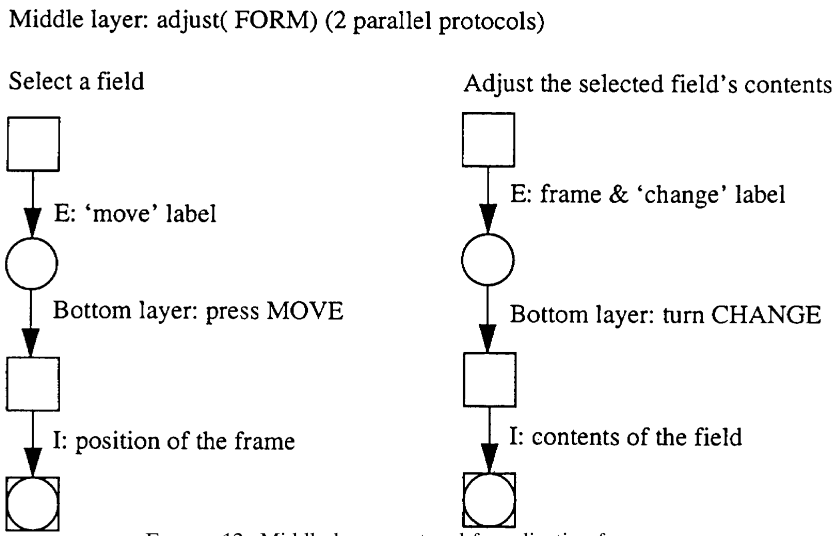

FIGURE 12. Middle-layer protocol for adjusting forms.

intentions. Its two-step protocol includes the selection of the appropriate form, enables adjustment of the form and commitment to execution.

The primary top-level message that the user wants to make a recording is an interpretation of the bottom level message that the user has pressed the record button. The label "record" should make users press the corresponding button when they have this intention. Therefore the label functions as E-feedback. After the button has been pressed, the record form is displayed on the screen. At this stage the record form functions as I-feedback, making users realize the system has interpreted they want to make a recording.

At this point users can adjust the form to their wishes. The middle layer supports the functionality for adjusting the contents of a form.

In the second step of the top-layer interaction the record form functions as E-feedback, since on the basis of this form users form expectations on what will happen after they commit to the system's interpretation by executing the physical action of pressing the go button. These expectations are strengthened by the additional E-feedback supplied in the graphical representation on the interpretation of the form. The label "go" should make users realize that they have to press that button to start the recording. The graphical representation, continuously updated according to the changing state of the tape in combination with the music, functions as I-feedback, making users realize the recording is in progress.

The middle layer is concerned with adjusting forms. This layer comprises two parallel protocols: one for selecting a field in the form and another for changing that field's contents. A field is selected by positioning a frame around it. Since the position of the frame should make users realize whether the correct field has been selected the frame serves as I-feedback. The label "move" near the cursor keys should make users realize that these keys enable repositioning of the frame. This label therefore functions as E-feedback.

The middle-layer message to change a field's contents is an interpretation of the bottom-layer message that the user has turned the change knob. The contents of a field serve as I-feedback, since they enable users to judge whether the field has been filled in correctly. In this protocol the frame functions as E-feedback, since its position influences the exact effect of turning the knob. The label "change" should make users realize that the knob can be used for changing the contents of a field.

The bottom layer of the DCC prototype is organized in the same way as that of the DAT recorder (see Section 3.2). The three layers provide independently encoded I-feedback of a different level of abstraction. The same information that serves as I-feedback at a particular level may also turn up as E-feedback in other interaction protocols when it indicates system settings relevant at a higher level.

4.3. USABILITY CONSIDERATIONS

It will now be discussed to what extent the DCC prototype succeeded in following the guidelines derived in Section 2.2, which should lead to certain improved usability aspects.

- *E-Feedback*

  E-feedback is offered in several forms: the labels of the buttons on the front panel refer to the user intentions which are supported by the new interface. The forms show what system settings influence its interpretation. Finally, additional information in the tape representation reflects how the form will be interpreted. This E-feedback is expected to give rise to correct expectations on the part of the user as to how the system may behave. It helps users in selecting the correct buttons.

- *I-feedback*
  The I-feedback supplied is organized according to the layered interaction structure. By looking at the changes in the selected form and in the tape representation, the effects of user messages can be immediately observed. This should make it easier for users to establish to what extent their intentions have been carried out. It is therefore expected that erroneous actions and system settings will be detected and adjusted at an early stage. Users will learn how to operate the interface faster.

- *Error and abort handling*
  Error and abort facilities are provided. The use of the rotary knob and the cursor keys for changing forms makes it intuitively clear to users how to obtain the inverse effect of a preceding modification: by turning the knob in the opposite direction or pressing the opposite cursor key. It is constantly possible to select another form.

- *Similarity of protocols*
  Thanks to the form-filling interaction protocol the more complex functions of the recorder are accessible in a similar way. This leads to increased internal consistency of the interface, which makes it easier to learn how to operate the interface.

In short, it is assumed that it will be easier to learn how to operate the DCC prototype because of its well structured E- and I-feedback and its high internal consistency.

5. Experimental comparison of the two user interfaces

The benefits and limitations of the LP model have to be assessed in a usability test: does the design based on the LP model yield an interface with improved usability? Therefore the usability of the newly designed DCC prototype was compared with that of the existing DAT recorder.

The usability of the recorders was measured according to ISO 9241 part 11 (1993, draft version). This standard uses the following definitions:

- usability is the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments;
- effectiveness is the accuracy and completeness with which users achieve specified goals;
- efficiency is the accuracy and completeness of goals achieved in relation to resources expended;
- satisfaction is the comfort and acceptability of using a system.

The ISO standard gives guidance in specifying and measuring usability in the context of office work involving visual display terminals. The application area of consumer electronics differs from professional settings with respect to training, as users operating a digital audio recorder are not usually given formal instruction or coaching. Therefore, the recorder should aim at being self-explanatory: the functionality and the organization of the interaction have to become clear while users explore the system. That is why a fourth usability criterion is added in this paper: learnability, which is defined as the speed at which the user's effectiveness increases over time.

5.1. EXPERIMENT

The ISO standard indicates that when measuring the usability of an overall system a description is needed of the relevant characteristics of the context of use, which includes equipment, environment, users and users' task goals. These aspects will be described below.

5.1.1. Equipment and environment

The recorders were evaluated in a laboratory setting. The subjects were seated in front of a table with either the DAT recorder or the DCC prototype positioned at the centre. In addition, a simple CD player was available as a recording source. The DAT recorder is an existing product and was used unaltered. The newly designed DCC prototype was simulated on a SUN workstation. It was fully functional and very realistic. The prototype included auditory cues as winding sounds and a front panel with real push-buttons positioned in front of the computer screen.

Normally environmental factors may hamper information exchange between the user and the system at the physical level. For example, poor illumination may make the labels on the front panel hard to read or the position of the product may influence the accessibility of the buttons. In the experiment, care was taken to ensure that no environmental factors impeded information exchange at the physical level of communication.

5.1.2. Subjects

Both recorders were evaluated by subjects of the same user group. The subjects were between 20 and 40 years old and had Dutch as their native language. Their educational background was non-technical and of a vocational or university level. Subjects with such a profile fall within the target user group of digital audio recorders.

In total 20 subjects, 12 men and eight women, participated in the experiment. They were divided into two groups of 10 subjects each, one for testing the DAT recorder and one for testing the DCC prototype. It was verified that the subjects working with the DAT recorder had, on average, the same experience with consumer appliances like VCRs, CD players and analogue cassette recorders as those working with the DCC prototype. The subjects had no experience with digital recorders.

5.1.3. Tasks

A set of eight tasks was devised comprising five relatively simple tape-transport and playing tasks and three more complex tasks: a programming, a recording and a tape-editing task. The order in which the tasks had to be executed was fixed. The subjects started with three tape-transport and playing tasks since users confronted with a new recorder are expected to explore this functionality first. After that, the subjects were confronted with the more complex tasks. In order to prevent the risk of subjects becoming discouraged when failing in these complex tasks, simple playing tasks were inserted between the programming and the recording task and between the recording and the tape-editing task. Since both recorders offer largely the same functionality, the same tasks were used for both recorders. These tasks were assumed to be representative of the entire functionality offered by the recorders: if users were able to complete these tasks they had mastered all major aspects of the recorder's operation. The tasks were expressed in every-day wording. Technical terms were avoided as much as possible.

5.1.4. Experimental procedure

The experiment consisted of three sessions. In each session the same set of eight tasks had to be completed. Before they started the first session, the subjects were given a written introduction of about 10 lines in which the most important aspects of the operation of the particular recorder being tested were indicated. They were allowed to study the introduction and explore the recorder for about 8 min. When they started the first session, the subjects were given the set of tasks and they were told a fully-fledged manual was available on request. The subjects were instructed to think aloud. If they had fruitlessly attempted to carry out a task for 20 min they were given the opportunity to proceed to the next task. At the end of the first session there was a 20-min break before the second session. At the end of the second session the subjects were requested to fill in a questionnaire and they were explained the procedures for correctly executing the tasks. One week later the third session took place.

5.1.5. Measurements

For each task in each session several measurements were made. A task-completion score was objectively assessed by the leader of the experiment on the basis of explicit criteria. If subjects successfully carried out a task they scored 1. That score was reduced for all aspects of the task erroneously executed, e.g. in the case of the recording task the completion score was decreased by 0.5 when users made the recording at the wrong position on the tape. The minimum completion score was 0. The time spent on each task was also measured. Furthermore, it was recorded whether the manual was requested or not. All sessions were video-taped to enable closer investigation of the subjects' behaviour and identification of the usability problems encountered. These measurements made it possible to compare the interfaces in terms of effectiveness by comparing the task completion scores, in terms of efficiency by comparing the execution times and in terms of learnability by comparing how the completion scores and execution times evolved over the sessions. The degree of satisfaction was not directly assessed but was derived from the answers of the questionnaire.

5.2. QUANTITATIVE RESULTS

5.2.1. Effectiveness

An analysis of variance with repeated measures (MANOVA) was made of the task completion scores for the recorders, sessions and tasks. The main effects were found to be significant: recorder [$F(1, 18) = 6.3$, $p < 0.05$], session [$F(2, 17) = 18$, $p < 0.0001$] and task [$F(7, 12) = 5.8$, $p < 0.005$]. The task completion scores for the DCC prototype were on average higher than those for the DAT recorder. On average the scores increased over sessions. The significance of the main effect "task" demonstrates that the subjects generally performed better in some tasks than in others. None of the interactions were found to be significant on a 5% level.

The average task completion scores for the two recorders per session are presented in Figure 13.

A more detailed study of the task completion scores revealed that for sessions 1 and 2 the differences between the DAT recorder and the DCC prototype as shown in Figure 13 were significant [$F(1, 18) = 6.4$, $p < 0.05$ and $F(1, 18) = 7.1$, $p < 0.05$]. For session 3 this was no longer the case on a 5% level. For the DAT recorder the increases between sessions 1 and 2 and between sessions 2 and 3 were significant [$F(1, 36) = 4.0$, $p = 0.05$ and $F(1, 36) = 5.9$, $p < 0.05$]. For the DCC prototype the increase between sessions 1 and 2 was significant [$F(1, 36) = 13$, $p < 0.001$]. There was no significant difference between the scores of sessions 2 and 3.

5.2.2. Efficiency

An analysis of variance with repeated measures (MANOVA) was made of the task completion times for the recorders, sessions and tasks. All main effects were found to be significant: recorder [$F(1, 18) = 15$, $p < 0.001$], session [$F(2, 17) = 57$, $p < 0.0001$] and task [$F(7, 12) = 63$, $p < 0.0001$]. The task completion times for the DCC prototype were on average lower than those for the DAT recorder. On average the

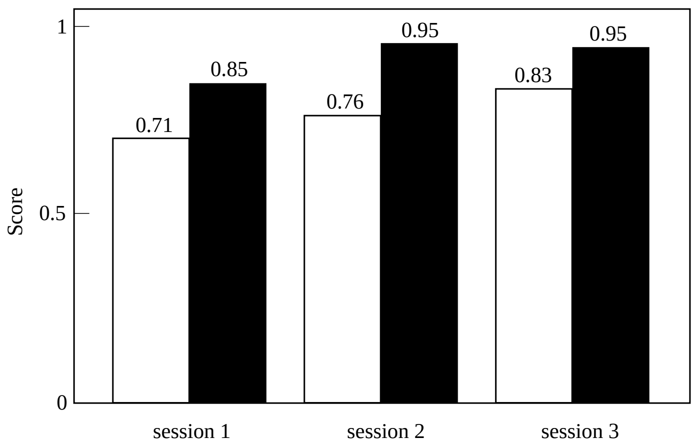

FIGURE 13. The average task completion scores per session for the two recorders. □: DAT recorder; ■ DCC prototype.

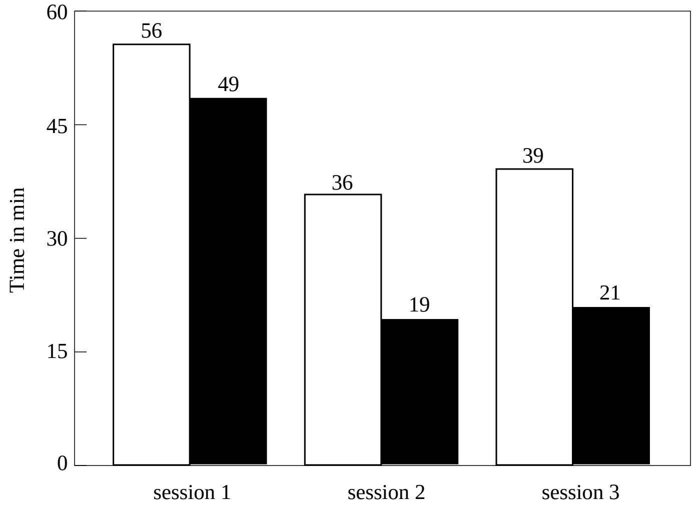

FIGURE 14. The average amount of time needed to complete the total set of tasks per session for the two recorders. Key as Figure 13.

first session took longer than the other two sessions. The significance of the main effect "task" demonstrates that the subjects generally performed better in some tasks than in others. The only significant interaction was that between the tasks and recorders [$F(7, 12) = 7.9$, $p < 0.001$]. The other interactions were not significant on a 5% level.

The average amounts of time required by the subjects to complete the total set of tasks are presented in Figure 14 for the two recorders.

A more detailed study of the task completion times revealed that for sessions 2 and 3 the differences between the DAT recorder and the DCC prototype as shown in Figure 14 were significant [$F(1, 18) = 15$, $p < 0.001$ and $F(1, 18) = 11$, $p < 0.005$]. However, for session 1 this was not the case on a 5% level. For both the DAT recorder and the DCC prototype the decrease between sessions 1 and 2 is significant [$F(1, 36) = 28$, $p < 0.0001$ and $F(1, 36) = 62$, $p < 0.0001$]. The difference between sessions 2 and 3 is not significant.

5.2.3. Satisfaction

In general, the subjects who had used the DCC prototype expressed a more positive attitude in the questionnaire than the subjects who had used the DAT recorder. This was most apparent when the subjects were asked what they thought of the experiment. In the case of the DAT recorder eight subjects said that they found the experiment difficult whereas two subjects said they liked the experiment. In the case of the DCC prototype only two subjects had found the experiment difficult whereas four subjects said they had liked the experiment and another four subjects indicated that they had ultimately appreciated the experiment although they had experienced difficulties at first. Though this information does not indicate satisfaction in a direct way, it does reflect a difference between the two recorders in terms of some subjective aspect of usability.

5.2.4. Learnability

In the case of both recorders a learning effect was apparent because both the scores and the task completion times yielded a significant effect over the sessions. A significant effect over the recorders was also observed: the task completion scores and the task completion times were generally better for the DCC prototype than the DAT recorder. Since the subjects were unfamiliar with the particular recorder tested and digital recorders in general this indicates that subjects learned how to operate the DCC prototype faster than the DAT recorder. The following more detailed evidence supports this.

- The average completion score of session 1 is significantly higher for the DCC prototype than for the DAT recorder (see Figure 13).
- In session 2 the average completion score was 0.95 for the DCC prototype. This high score was again obtained in session 3. This indicates that the subjects had already mastered the operation of the interface in session 2 and only made minor mistakes after session 1. On the other hand, the completion scores of the subjects who operated the DAT recorder significantly increased by 0.07 from session 2 to session 3.
- In session 1 the average task completion time was not significantly shorter for the DCC prototype than for the DAT recorder. However, the average task completion times for sessions 2 and 3 differ significantly: the average task completion time for the DCC prototype is about half that of the DAT recorder (see Figure 14).
- The data regarding manual use constitute further evidence (see Figure 15). An analysis of variance with repeated measures (MANOVA) of these data revealed a significant difference between the two recorders [$F(1, 18) = 29$, $p < 0.0001$]. This indicates that the subjects who operated the DAT recorder generally consulted the manual more often.

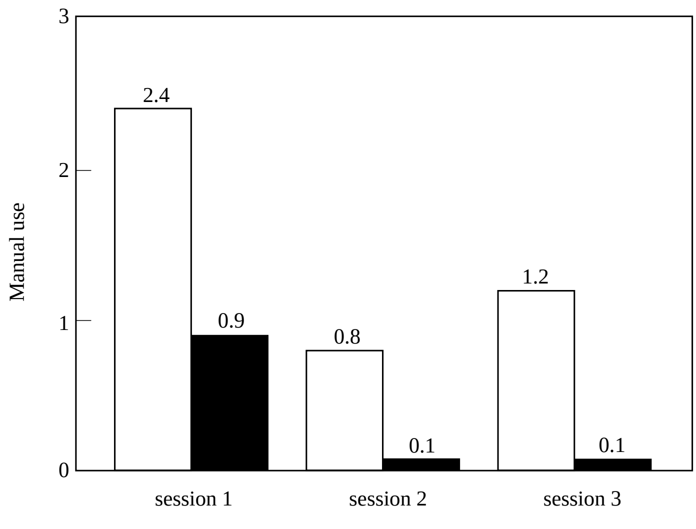

FIGURE 15. The average total number of times that the subjects had to consult the manual to be able to complete all the tasks per session for the two recorders. Key as Figure 13.

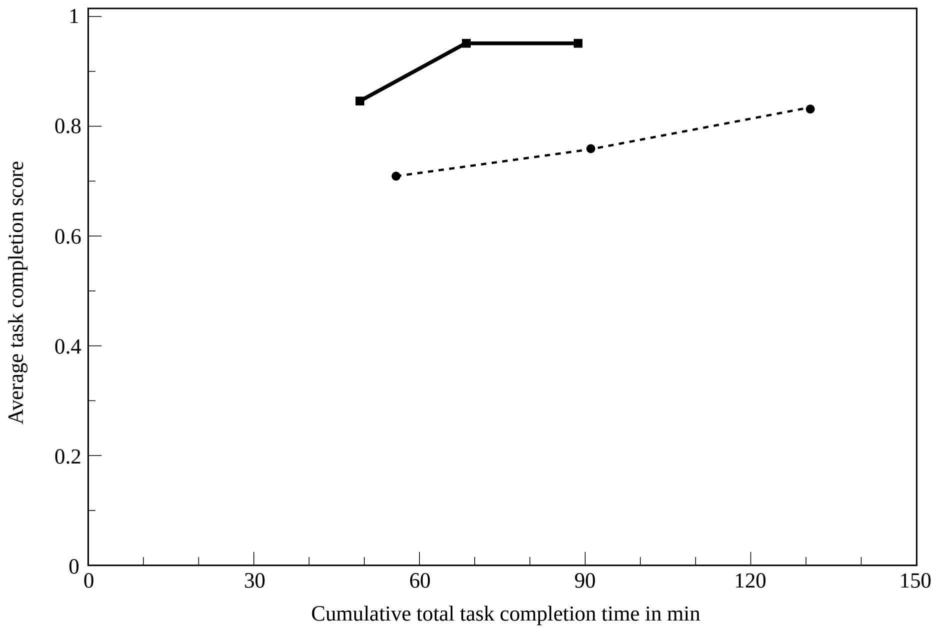

FIGURE 16. The "learning curves" for the two recorders. For each session a data point is plotted, representing the average task completion score for that session on the vertical axis, and the total time spent in interaction with the recorder (cumulative total task completion times) on the horizontal axis. The graphs thus give an impression of how much was learned in what amount of time. ■—■: DCC prototype; ●- - -●: DAT recorder.

Figure 16 summarizes the data presented in Figures 13 and 14 and shows the average task completion scores as a function of the cumulative total task completion times. This yields a comprehensive picture indicating how quickly effectiveness increases over time. The difference in learnability is apparent.

5.3. OBSERVATIONAL RESULTS

5.3.1. Usability

The data presented above show that the DCC prototype is more usable than the DAT recorder in terms of effectiveness, efficiency, satisfaction and learnability. Since the environment, the functionality and the set of tasks were kept constant in both evaluations, and because the subjects were comparable in terms of age, educational background and experience with consumer appliances, the measured usability differences between the two recorders must be mainly attributable to the fact that they have different user interfaces.

These two user interfaces differ with respect to the extent that they are in line with the LP model: the design of the user interface of the DCC prototype was based on the LP model whereas the DAT user interface deviates from the LP model in various respects. It remains to be discussed to what extent the user interface aspects indicated by the LP model account for the usability differences between the two interfaces. By reviewing the video tapes it was investigated whether the expected usability consequences had actually occurred and whether any unexpected usability problems had turned up.

5.3.2. Observed usability aspects addressed by LP

The discussion of what elements of the LP model account for the usability differences between the DAT recorder and the DCC prototype will be based on the derived guidelines: how they were expected to work and how they actually worked (see Sections 2.2, 3.3 and 4.3).

- *E-feedback*

  The subjects who used the DAT recorder often had trouble deciding which buttons had to be pressed for the various tasks because many labels were rather cryptic and did not give rise to expectations as to the effect of pressing these buttons. The buttons involved in tape editing presented the most problems. The subjects were also uncertain about the relevant settings and their influence on the ultimate effect. The DAT recorder provides virtually no E-feedback of this kind.

  The DCC prototype also had labels that were not immediately understood, viz. the labels of the buttons involved in changing the contents of a form. The subjects understood the relevance of the fields on a form quite well and had no trouble in predicting the effect that the execution of a form would have on the tape. The additional information in the graphical tape representation was of help in interpreting the forms and making the predictions.

  Thus, E-feedback proved to have a strong influence on usability. However, the mere availability of E-feedback at each level is not sufficient. It must also be of high quality: it must provide users with sufficient information for predicting the effect of a message.

- *I-feedback*

  Although the subjects who used the DAT recorder could often infer from lower-level I-feedback that the recorder was doing something with the tape, they had problems determining the effects of their actions because I-feedback was lacking at the highest level. For example, when some buttons were pressed the corresponding label was repeated in the display. Although these subjects realized that the system had reacted, they were often unable to infer the effects from what they knew about the previous state of the recorder in combination with their knowledge of the function of the button. So the subjects had to establish the effects of their actions by transporting and playing the tape. As a result, they found it difficult to determine what action had caused the effect.

  The subjects who used the DCC prototype were quite capable of immediately assessing the effects of their messages by observing the changes on the display.

  This leads to the same conclusion as for E-feedback, namely that the mere availability of immediate I-feedback at each level is not sufficient. It must also be of high quality: the users must be provided with sufficient information for determining to what extent the intentions have been carried out, for detecting communication failures and for learning the operation of the interface by linking their actions to system responses.

- *Error and abort handling*

  No usability differences due to lacking error and abort facilities were observed.

The DAT recorder and the DCC prototype did not differ much in this respect. All the subjects frequently used these facilities in exploring the interface. This indicates that these facilities form an indispensable part of interaction protocols.

- *Similarity of protocols*
  When confronted with the form-filling strategy for the first time, the DCC subjects required some time to grasp this protocol. When using it for the second time, for a different task, they made extensive use of the experience gained during the first encounter. The subjects were quite capable of transferring their experience to another task and generalizing the interaction protocols. During the first session they often needed to consult the manual only once: namely for the programming task when they were confronted with the form-filling strategy for the first time. The DAT recorder shows virtually no internal consistency of this kind. During the first session the subjects often needed to consult the manual up to three times: namely for the programming, the recording and the tape-editing tasks (see Figure 15). With the DCC prototype these tasks could all be executed using the same form-filling strategy whereas the DAT recorder employed three different interaction protocols. It is therefore concluded that similarity of protocols is indeed an important asset predicted by the LP model.

To summarize, it can be observed that the guidelines derived from the LP model have worked as assumed but that the influence of the quality of the E- and I-feedback has been underestimated.

5.3.3. Observed usability aspects not addressed by LP

Although the LP model provides a unified and consistent framework for structuring user-system interaction, it may very well be incomplete and may miss some aspects that also affect usability. A usability issue that is not addressed in the LP model but whose importance was observed in the video tapes is the familiarity of subjects with the user interface concepts employed. Three observations lead to this conclusion.

- In general the subjects had little trouble executing the tasks involving playing and transporting the tape, despite the low quality of the E- and I-feedback. This can be explained by their familiarity with CD players and analog cassette recorders. This experience could be easily transferred to the DAT recorder and the DCC prototype because the same interaction concepts were applied.
- The facts that the subjects were not familiar with the user interface concepts of the DAT recorder and that they differed from their intuitive notions caused confusion. The subjects' intuitive notion was that a track is the central concept that the various functions work on. This is probably due to their familiarity with CDs. However, the DAT user interface uses the ID concept: e.g. a "start ID" marking the tape position at which a track begins can be "written" on the tape and can be "erased". This led to discrepancies between the subjects' expectations and the actual behaviour of the system. Sometimes the subjects confused their intuitive notions with the recorder's concepts, e.g. they feared losing the music of a track when erasing its start ID.
- The subjects who used the DCC prototype had problems learning to fill in a form. Although all the subjects succeeded in mastering the form-filling protocols, these

protocols caused serious delays in the task completion times of the first session. Because the subjects were unfamiliar with this kind of form-filling, it took them a long time to learn these new interaction concepts.

To summarize, it can be observed that the users' familiarity with the user interface concepts employed also has an effect on usability. Since this factor was found to be influential in the case of both interfaces, it is unlikely that it fully accounts for the measured usability difference between the DAT recorder and the DCC prototype.

6. Discussion

The goal of this paper was to assess the benefits and limitations of the Layered Protocols model for the analysis and design of user interfaces in the domain of consumer electronics. A first step towards the achievement of this goal was the derivation of a set of guidelines from the LP model. These guidelines were used to analyse an existing interface and to design a new one. Next, the hypothesis was formulated that the user interface based on the LP guidelines is more usable, which was investigated by measuring the usability of the two interfaces. Finally it was discussed to what extent the differences in usability can be explained by deviations of the interfaces from these guidelines. This assessment of the LP model will be discussed in this section.

Describing the user interfaces in LP terminology

Analysing and designing user interfaces with the aid of the LP model involves describing the interface in terms of the model. Although the LP model is so general that it can describe virtually any interaction, the amount of work involved proved to be different for the two interfaces. Describing the DCC prototype was easy since it was explicitly designed according to the LP model. But the analysis of the DAT recorder was rather tedious because each function had to be described separately. The interaction protocols for different functions showed little similarity. Since the feedback supplied does not reflect a layered structure, it was questionable whether an LP analysis would be appropriate. However, the user's part of the interaction shows a layered structure: the execution of a function like record means that some settings, such as the tap head position, have to be adjusted. The lack of layered feedback made the analysis of the DAT recorder more difficult. It is also assumed that this lack makes it harder for users to become aware of the layered interaction structure.

Guidelines

The LP model as presented by Taylor (1988*a*) is a very general interaction model. For this study, a selection was made of the interaction concepts introduced in the LP model. The most important concepts were elaborated into guidelines indicating potential usability consequences. By making the model operational in this way it becomes possible to test whether the model and the guidelines correctly indicate factors influencing usability.

On the basis of the interface descriptions in combination with the guidelines, potential usability problems were pointed out. Although the usability issues indicated by the guidelines actually occurred, it was felt that it should be possible to pinpoint usability consequences more precisely on the basis of the detailed interface descriptions. The current guidelines are not specific enough in relation to the detailed LP descriptions. More hands-on experience is needed to extend and refine the guidelines.

Evaluating usability

The ISO 2941 standard was used to compare the usability of the two interfaces. Because the assessment was carried out in the domain of consumer electronics, learnability was added as a usability criterion. It was defined as the speed at which the users' effectiveness increases with time. Though it can be argued that learnability is simply the way in which usability develops with time, it is considered important enough to be treated as a separate usability criterion. Learnability is more likely to influence what functionality will actually be used on the long term that efficiency. Nielsen (1993) even considers it the most fundamental usability attribute. Unfortunately the ISO 2941 standard does not address this issue.

Because the experiment was conducted in a laboratory setting, it has to be considered whether the measured usability relates to the usability experienced in real life. A major difference between the laboratory setting and a real-life situation concerns the goals: the subjects were given explicit, well-defined tasks whereas intentions of users may be less articulate. It is also felt that the subjects were more tenacious in their attempts to fulfil a task than users would be in real life. Nevertheless, the experimental conditons in this study make it plausible that the measured usability differences reflect real-life usability differences.

While usability focuses on making the system's functionality available to users, the usefulness of a system indicates whether its functionality serves a purpose for its users. In real life, a good match between the system's functionality and the users' intentions is likely to yield a more valuable system for its users. That is why in this study the two recorders were kept functionally equivalent.

The LP model

A general interaction model should be a useful tool for designing better conceived and more usable interfaces. In principle, it should assist in structuring the interaction and explaining the role of each piece of information exchanged while indicating usability consequences. A general interaction model cannot be expected to indicate the usefulness of functionality, nor can it be expected to describe the actual contents of interaction messages. However, it could indicate criteria for high-quality feedback and provide simple usability tests for measuring quality differences between alternative messages and interaction structures. In this way an interaction model can decrease the need for more extensive user tests in interface design and help in reaching a basic level of usability.

This study has shown that the LP model is a step in the right direction. Its distinction between E- and I-feedback indicates the different roles feedback can have. It advocates structuring interaction in a layered way while the guidelines indicate potential usability consequences. The conclusion of this paper is that these interaction aspects actually influence the usability of a system. The main drawback of the LP model is its lack of precision. More explicit criteria for distinguishing between layers and classifying information as E- or I-feedback should make the LP model more clear-cut. The guidelines must also be refined to predict usability consequences more precisely. The LP model would be far more useful in user-interface design if simple standardized user tests were available for evaluating interaction per layer, while the overall interaction architecture would guarantee a basic level of usability.

7. Conclusion

On the whole, it can be concluded that the LP model indicates interaction aspects influencing the usability of user interfaces. Guidelines can be derived from the LP model that can be of help in reaching a basic level of usability in user interface design.

Some of the most important aspects of the LP model include the layered organization of the interaction in combination with immediate high-level I-feedback, which facilitates early detection of communication failures. Furthermore, E-feedback guides users in coding their intentions. The quality of E- and I-feedback turned out to be crucial: E-feedback must provide users with sufficient information for predicting the effect of a message, and I-feedback must provide users with sufficient information for determining to what extent their intentions have been carried out and for detecting communication failures.

The LP model also indicates that the organization of I-feedback at the various levels strongly influences the learnability of an interface. It helps users to establish the effects of their messages immediately. A second factor contributing to learnability is an optimum similarity between interaction protocols, which leads to increased internal consistency within user interfaces. In consumer electronics in particular high learnability is essential, because most consumers will have had no formal instruction or coaching.

One influential usability factor for which the LP model does not account was observed: the users' familiarity with the interaction concepts applied in the user interface. With this limitation in mind, the overall conclusion of this assessment is that the LP model indeed addresses user interface aspects influencing usability and can thus be of help in reaching a basic level of usability in user interface design.

We gratefully acknowledge the layered feedback received from Don Bouwhuis, Frits Engel, Floris van Nes, Martin Taylor and John de Vet on earlier versions of this paper. We thank Rudy van Hoe and René Wijnen for their help with the data analysis.

8. References

Card, S. K., Moran, T. P. & Newell, A. (1983). *The Psychology of Human–Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Engel, F. L. & Haakma, R. (1993). Expectations and feedback in user-system communication. *International Journal of Man–Machine Studies,* **39,** 427–452.

HUSAT Research Centre (1988). *Human Factors Guidelines for the design of Computer-Based Systems,* Technical Report. Loughborough.

ISO (1993). Guidance on specifying and measuring usability, ISO CD 9241 part 11 (draft).

KIERAS, D. & POLSON, P. G. (1985). An approach to the formal analysis of user complexity. *International Journal of Man–Machine Studies,* **22,** 365–394.

MARSHALL, C., NELSON, C. & GARDINER, M. M. (1987). Design Guidelines. In M. M. GARDINER & B. CHRISTIE, Eds. *Applying Cognitive Psychology to User-Interface Design,* pp. 221–278. Chichester: Wiley.

NIELSEN, J. (1993). *Usability Engineering.* Boston, MA: Academic Press.

NORMAN, D. A. (1984). Stages and levels in human–machine communication. *International Journal of Man–Machine Studies,* **21,** 365–375.

PAYNE, S. J. & GREEN, T. R. G. (1986). Task-action grammars: a model of the mental representation of task languages. *Human–Computer Interaction,* **2,** 93–133.

TAYLOR, M. M. (1988*a*). Layered protocols for computer-human dialogue I: principles. *International Journal of Man–Machine Studies,* **28,** 175–218.

TAYLOR, M. M. (1988*b*). Layered protocols for computer–human dialogue II: some practical issues. *International Journal on Man–Machine studies,* **28,** 219–257.

TAYLOR, M. M. (1992). *Principles for intelligent human–computer interaction, a tutorial on Layered Protocol theory.* DCIEM report no. 93–32, North York, Ontario, Canada.