An Approach Based on Multiple Text Input Modes for Interactive Digital TV Applications

Didier Augusto Vega-Oliveros
Diogo de Carvalho Pedrosa
Maria da Graça Campos Pimentel
Renata Pontin de Mattos Fortes
Universidade de São Paulo, São Carlos-SP, Brazil
{davo,diogo,mgp,renata}@icmc.usp.br

ABSTRACT

The development of interactive digital TV applications is hindered by the user-interaction options allowed when traditional remote controls are used. In this work, we describe the model of a software component that allows text entry in interactive TV applications based on an interface with multiple input modes — the component offers a virtual keyboard mode, a cell keypad mode, and a speech mode. We discuss our considerations with respect to the design, development and evaluation of a prototype corresponding to our model, built according to the user-centered design methodology. After conducting a research on existing text input methods in television systems, we interviewed four experts in the interactive TV domain. We also applied 153 questionnaires to TV users, with the aim of gathering a user profile of users who make use of text entry mechanisms. During the development of the prototype, we conducted usability tests using the think aloud protocol, and usability inspections using the heuristic evaluation and cognitive walkthrough techniques. The evaluations allowed the detection of both, a number of problems and of several improvement opportunities; at the same time that they highlighted the importance of using complementary text input modes in order to satisfy the needs of different users. Overall, the evaluation results suggest that the proposed approach provides a satisfactory level of usability by overcoming the limitations of text input in the context of user-interaction with interactive TV applications.

Categories and Subject Descriptors

B.4.2 [**Input/Output and Data Communications**]: Input/Output Devices—voice; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—Input devices and strategies, Voice I/O

General Terms: Design, Human Factors.

Keywords: Text Input, Interactive Digital TV, Multiple Input Modes, Virtual Keyboard, Multi-tap, Predictive Text.

1. INTRODUCTION

The introduction of interactive digital television (iDTV) has the potential of changing the way viewers receive multimedia content by allowing improved quality of video, multiprogramming and user-interaction with applications. Although there are several attempts to develop TV as a main provider of information and entertainment provider, the development of interactive applications is hindered by a small number of user-interaction options allowed by traditional remote controls.

Although the interface of iDTV applications have inherent restrictions to the TV paradigm, it differs from the personal computer (PC) paradigm in key aspects such as viewing and navigation [20], several classes of applications for iDTV can use text input from the user, such as chat, e-mail, search on the electronic program guide (EPG), calendar and forms — in such a scenario, it is important to provide mechanisms to aid the user for entering text.

In this paper we present our model of a software component that allows users to enter text in iDTV applications by means of an interface based on multiple input modes — the component includes a virtual keyboard mode, a cell keypad mode, and a speech mode. We designed the component according with the user-centered design (UCD) methodology. We first carried out a research review with respect to text input methods used in TV systems. Next, we interviewed four experts in the iDTV domain. A third activity involved the application of questionnaires to 153 TV users, aiming at identifying a profile that corresponds to users who make use of text entry mechanisms.

Carrying on with the development of the prototype, which we were able to demonstrate elsewhere [19], we conducted usability tests using the think aloud protocol, and usability inspections using the heuristic evaluation and cognitive walkthrough techniques. The evaluations allowed the detection of a number of problems, which could be dealt with in intermediary versions. The evaluations also pointed to several opportunities of improvement on the design — in particular, they highlighted the importance of using complementary text input modes in order to satisfy the needs of different users. As a result of our design and corresponding evaluation, we suggest that our proposal provides a satisfactory level of usability by overcoming the limitations of text input in the context of the user-interaction with iDTV applications.

The remaining of this paper is organized as follows: systems based on multiple input modes are presented in Section 2; Section 3 visits some text input methods used in TV and discuss some limitations of the TV environment; the design of the model and the proof-of-concept prototype in Section 4; Section 5 presents the evaluations performed; and, the lists of the next steps in Section 6. Finally, Section 7 concludes this paper and points out future research efforts in the final remarks.

2. SYSTEMS BASED ON MULTIPLE INPUT MODES

All natural forms of expression and interaction that could be captured by some kind of technology may be used by systems to allow users to interact naturally and easily. These natural forms extend the concept of interfaces based on mouse and keyboard interaction, known as WIMP (Windows, Icons, Menu and Pointer).

As examples of natural forms of interaction and their corresponding devices we have voice captured by microphone, electronic ink captured by tablets, touch captured by touchpads and touchscreens, gestures captured by accelerometers and cameras, etc.

Technologies that help people to interact with devices involve elements which were "transformed" by taking advantage of the implicit metaphor of people's daily use — such as electronic pens and devices with embedded accelerometers. From a ubiquitous computing perspective, the aim is to allow the user interaction with applications more natural [1].

Several studies in the scientific community have investigated the use of multiple modes of user interaction with computing systems and applications. Many of them use the voice input mode supported by a speech recognition engine. In the work of Patel et al. [18], commands are given to a phone system that provides local news services to Indian farmers.

There are several reasons to develop and use interfaces based on multiple input modes. One of them is to offer to users with disabilities the possibility to interact with alternative modes that best suit their necessities, in order to democratize the access and increase the number of potential users. Ferati et al. [11] propose a design for acoustic educational applications for blind users. Similarly, Harada et al. [14] are concerned with the development of a form of interaction through voice that brings benefits analogue to the possibility of direct manipulation allowed by the mouse.

In the other hand, there are works that explore different modes of text input. In the work of Cox et al. [5], the main objective was to explore the complementary aspects of the voice and the phone keyboard to overcome the deficiencies of traditional methods of text input in circumstances of user mobility, in which he has busy hands and vision. Finally, Castellucci and MacKenzie [3] present a technique for text input using the motion sensor equipped remote control of the Wii video game to capture gestures that are mapped to characters. An alphabet of gestures is proposed, in which every gesture is composed of only two primitive movements. In common, all these works offer alternative ways to perform a task in order to improve efficiency to reach a wider user population.

3. TEXT INPUT INTERFACES ON IDTV

The use of interfaces based on multiple text input modes in digital television is the focus of this study. For better understanding the problem, we carried out a survey of methods used to input text in iDTV applications. Some of them are even older than the first applications for digital TV, as they are inherited from video game consoles with simple remote controls. In this group, we highlight 2 main forms: virtual keyboard and sequential selection of characters. The first method consists in displaying on the TV screen a set of buttons representing characters. One of them has the focus, that can be moved using the arrow keys. The associated character can be entered using the OK / Enter key of the remote control. Two different layouts are typically used: alphabetic and QUERTY. Devices like iPhone, that do not have a real keyboard, but are equipped with a touch screen, also make use of the virtual keyboard. Recent work [13, 22] explore QWERTY virtual keyboard with two foci controlled by joysticks equipped with two directional controls. The idea is to use the analogy of typing on a computer keyboard, where each hand is responsible for half of the keys.

Sequential selection of characters is a simpler method that consists in presenting to the user a character at the current cursor position: the character can be modified, in alphabetical order, using the arrows on the remote. After selecting a character, the cursor moves to next position and the user should once again go through the list of characters available for selection of the new character. This method avoids using the screen to display the virtual keyboard, but provides a less efficient text entry. Another group of text input methods that can be used in iDTV comes from the ones traditionally found in mobile phones, as traditional iDTV remote controls also have the numeric keys set. In this group, two methods stand out: multi-tap and predictive text. James and Reischel [16] present performance metrics for each of them and show the efficiency of predictive text.

Methods that make use of more recent technologies are also being investigated. In the work conducted by Nakatoh et al. [17], the authors describe the techniques developed for the speech recognition system present in the TV control system. An omnidirectional microphone was used due to the position variation of the speaker in relation to the microphone embedded on the remote control. The captured audio signal was sent to the TV equipment, where it was initially processed by a digital signal processor. Then, the recognized phonemes were passed to be processed by the automatic speech recognition system. The speech recognition system allowed that the same commands were mapped by several items of the phoneme dictionary. The name of a single channel, for example, could be pronounced in 6 different ways. In total, the dictionary had around 400 items, and it was 1300 items when the various ways of pronouncing the same command were considered. To improve the system usability, they developed a technique for reducing ambient noise and a technique of echo cancellation of the TV sound. They also stated that the speaking style varies with the user generation and therefore they developed age-dependent acoustic models.

Another study which used on commands by voice was performed by Wittenburg et al. [23]. In their proposed system, users are free to pronounce any word, without vocabulary or grammar restrictions, and receive as output a list of possibly related programs. To help users understand the results, the authors propose as future work a variable highlight in the words of the result list. That is, words for which the system credits higher probability of been pronounced by the user gain greater prominence.

The literature also reports researches that use the interpretation of user's interaction with specifics devices as a strategy for input text. In this group we have the work of Castellucci and MacKenzie [3], as already mentioned. In it, a technique for text input using the control of the Wii video game is proposed. To achieve it, they capture the motions of the remote control and map them to characters. There is an alphabet in which every gesture is composed of only two primitive movements. The work of Fagá Jr. et al. [10] could be interpreted in the same direction. In this case, the mapping comes from the strokes of a electronic pen. They propose an architecture approach that exploits the automatic capture of the user interaction with personal devices, employs ontology to store context information, and uses the context information to organize users in P2P groups for the collaborative exchange of information. The ontology employed convert the ink-annotation to a annotation that then could be transformed, according to the Watch and Comment paradigm, to input text.

Finally, in the work of de Jesus Lima Gomes et al. [7], an interface using barcode on paper was designed to assist the interaction with a distance learning system using the television. Systems like this, if restricted to the traditional remote control, have the advantage of reaching a wider audience, but offer a more complex interaction. In the system proposed by them, a remote control equipped with a barcode reader is used to request the display of multimedia content related to the topic studied in a printed material.

3.1 Physical Devices

The new forms of interaction with the television system that are being developed may require physical devices other than traditional remote control. De Miranda et al. [8] conducted a detailed study on the challenges and guidelines that should be considered during the design and integration of new physical artifacts to the Brazilian context. According to the authors, the remote control used with analogue television, still prevalent, can act as a limiting factor for interaction with proposed and developed services for digital television. Among the ten guidelines presented, the ones with number 1, 2 and 4 make explicit reference to the use of speech, recommending that interaction alternatives for people with physical disabilities, visually impaired and illiterate should be provided. However, the authors point out that not all environments can promote this form of interaction and that the system has to be trained to recognize the voice of the user. They mention also the problem of the collective use of TV and the noise and natural atmosphere in which TV is watched. Hence the importance of providing alternative ways to carry out a task, as mentioned in Section 2.

As seen in the previous section, the product made by Nakatoh et al. [17] has a remote control that has an integrated omnidirectional microphone. Other features of the "speech remote control" are: 1) it has only 14 keys instead of more than 70 of the equivalent traditional remote, 2) it allows almost all commands that can be entered with the traditional command, and 3) has an easy to grab format, with a push-to-talk side button.

Currently, the most sold software for speech recognition is the Dragon NaturallSpeaking[1]. It has a large vocabulary, continuous speech recognition and reports a 99% accuracy. It requires that each user create a personal speech model by undergoing a 10-minute reading section of predefined texts. It was the 7th version of the software that was used in the experiments of the work from [5].

3.2 Limitations in iDTV interfaces

Interactive digital TV is an entertainment system that offers a variety of service types using a broadcast reception and a return channel through which it transmits the interaction data. However, the main interaction device with the iDTV is the traditional remote control. The interactivity level that the remote control offers is limited to pressing keys, some of them mapped to a specific command in the television screen. According to Berglund et al. [2] and Piccolo and Baranauskas [20], we can observe tree main problems that iDTV faces:

(1) The user interaction paradigm in iDTV looks like menu navigation, as it could be done in a computer. So, people with little experience on computers are excluded from interacting with iDTV applications. Besides, the TV does not have the same tools as a PC (mouse, keyboard, processing power, etc.).

(2) The interaction design is poor, as the screen resolution is small. This leads people to think that the interfaces are poor in relation to the functionalities, even those who are familiarized with new technologies.

(3) The main interaction device used in the iDTV is the remote control, which is still inadequate and raise difficulties to this task [4]. Most applications are limited to map keys of the remote control in the television screen in order to allow the selection of the commands of the interface [20], providing a hard to understand and monotonous interaction, causing frustration and irritation in users.

There are a variety of proposals to solve this kind of problems. Some of them try to develop better remote controls [22, 2, 6, 13]. Others propose different mechanisms not yet explored, as accelerometers [3], remote controls with barcodes reader [7], interactive applications that reproduce gestures [21], approaches that combine speech and remote control [5], or intelligent prediction mechanisms of text entry from the remote control [12], for instances.

4. OUR DESIGN

The issues discussed here were considered during the development of the prototype of a mechanism for text input in iDTV applications. The project was conducted using a set of techniques that aim to engage the end user during all stages of development, known as User Centered Design [9].

First, we did a requirement elicitation with potential users and experts in the area. In the second step, the features and technical characteristics of the interface model were defined. Finally, we developed a primer prototype in order to make a proof-of-concept usability tests in it.

4.1 Requirements Elicitation

Initially, a study to better understand the future users of the mechanism to be developed was conducted. At this stage, we interviewed four experts in the area, coming from different contexts, and we also applied questionnaires that included potential users from various regions of Brazil. The main contributions are reported below.

[1] http://www.nuance.com/naturallyspeaking/

4.1.1 Questionnaires

The questionnaires were applied on paper (32 responses), in order to reach a diverse audience, and using a on-line surveys system (121 responses), which helped to obtain a greater number of responses. One of the questions asked was what would be the best way to write a message to a friend using the television. The answers to this open question were very diverse and may be clustered into the following categories: QWERTY keyboard; T9 predictive text standard; Speech; Virtual keyboard using a simple remote control; Virtual keyboard using a touchscreen TV; Virtual keyboard using a touchscreen remote control; Thought; Preformulated phrases; Writing on a paper whose text is recognized by the television; Writing with a pen directly on television screen; and Using a mobile phone connected to the TV.

A concern regarding the need to offer more than one alternative to text input could be noticed, as in the following examples (our translation):

- "*Using speech when I'm alone and using T9 when I'm in an ambient with other people.*"

- "*Speech would be ideal, but I think we could correct some possible errors or [make some] modifications using a keyboard, for example*"

- "*I would like to be able to choose:* `i)` *If I'm in the living room with my mother in law: conventional keyboard;* `ii)` *if I'm with other people: speech to text.*"

- "*Using speech and phonemes recognition or, to a reality closer to ours, a remote control with a LCD touchscreen...*"

- "*... but there must be other ways, that are accessible from those who are dumb or are currently without voice, for example.*"

Another recurrent concern was regarding to the interference of the environment and TV sound, in cases in which the text input is performed using speech, as in the following examples (our translation):

- "*... I know there are limitations and difficulties, such as external interference from other people, background noise, etc.*"

- "*... but there may be noise problems*"

- "*... without interfering with the program audio that the person is watching.*"

4.1.2 Interviews

In order to better understand the problem to be solved, we interviewed four experts in the area, each one with different specializations. They were a college professor researcher in the areas of multimedia, hypermedia, middleware and interactive applications for digital TV, a usability engineer of a research and development Software Company, and an interaction designer and a products consultant of a cable TV Company. The main contributions are reported below.

The college Professor:

- He thinks that the media use various auxiliary information that help in understanding the message. There are a number of ambiguities that are only broken because of the context.

- He finds the possibility of sending messages through TV interesting. He imagines that the user makes simple operations using the remote control, and would be interesting to write the message using speech. But the interference could disrupt the environment. "*It should occur in a restricted environment.*"

- "*There are several ways in which communication can be done: voice, body movement, interaction with objects, wind, blowing stronger or weaker can generate an alphabet, for instances.*"

- He considers necessary to define reserved voice commands in order to let the machine knows when to take certain actions instead of writing what is being spoken.

- "*Would be interesting to search for what techniques should be used and in what situations.*"

- He notes that young people prefer to communicate with text, even though the communication by voice are faster and more direct,"*... youths today are so good writing with the numeric keypad that perhaps other ways are not best suited to them ... Maybe it's because texts are more reserved*" he said.

The usability engineer:

- He's seen only one application that used text entry. It used a virtual keyboard. "*It was an alphabetic keyboard, not QWERTY. One of the letters had focus and the enter key was used to insert the letters focused.*"

- He considered feasible text entry by voice. "*It would require less effort and it is interesting because it is more natural.*" he said.

- " *The advantage would be to insert long texts. The disadvantages would be when people were watching TV together with someone, because of privacy, or when there were someone sleeping.*"

- He said he did not know any project for text entry by voice on TV.

The interaction designer and the products consultant:

- *"Once we have analyzed several kinds of infrared remote controls. Some were of normal shape and others more complex-shaped, like a mouse, keyboard or joystick. Cost-benefit has led us to choose the simple remote control."*

- *"The major difficulty is that there are no character keys in the remote control and a simulation has to be done. One way is by showing a virtual keyboard on the screen and allowing the user to navigate through the arrow keys and press OK. We tested the QWERTY and alphabetic keyboards. According to the performance tests, typing in the alphabetic keyboard had more success than QWERTY when the user was at are slightly larger distance from TV."*

- *"The tendency is to use text input format similar to the cell phone, where the letters are also printed on the keys."*
- *"Another difficulty in TV is that the user has to look at the screen to see the result and to look at the remote control to seek for the letters, what does not occur in cell, since the keyboard and screen are very close. Even when the remote controls have the letters printed, we will continue showing the keyboard map on screen."*
- *"What we have today is that the remote control is easily available to watch television."*
- *"Another problem is the environment where the TV is. It may be that i am dictating a text and my son goes on the side screaming and disturbs."*
- *"The TV has a special characteristic of being a familiar device and not individual."*

This step was crucial to help us to understand the importance of offering alternative ways of entering text, in order to meet the needs of different user profiles and to be flexible enough to be used in different environments, as discussed in Section 2. Still aiming to better understand the problem to be solved, we conducted a survey on the ways of text input currently used in television systems, which was presented in Section 3.

Requirements elicitation allowed us to define the functional requirements and key criteria of usability to be considered. Nine functional requirements for the system were elicited, addressing basic issues related to inserting and deleting characters and words, and the substitution of letters both using speech and remote control. The following usability criteria were considered the most important in the context of the project: (1) *Familiarity*: User should be able to use some of the knowledge it already has in the context of writing during his first interactions with the proposed mechanism, (2) *Substitutivity*: Complementary forms of text input must be provided for the user, and (3) *Responsiveness*: A iDTV decoder usually has little processing power compared to a personal computer. The system must be fast enough to let the user notice changes in its state.

4.2 Text entry model based on multiple input modes on iDTV

Our proposal aims to explore and to develop a mechanism for text entry in iDTV. During the project we realized that no single mechanism is able to meet the needs and characteristics of all system users. The approach based on multiple input modes serves a larger number of users in a satisfactory manner.

The importance of the project lies in how our users will interact with the component, if the first interactions between the user and the system are clear, if the user finds the various mechanisms proposed interesting enough so that he can perform a specific task, and if the system can make the user to perceive readily the state changes in response to actions.

We designed the model considering three main methods of entering text:

(1) Speech recognition with a microphone located on the user's remote control, which is activated via a push-to-talk key;

(2) Cell keypad mode using the remote control, with the mapping between buttons and letters shown on the GUI. That method can be used with text prediction (T9) or without a dictionary aid (multitap);

(3) Virtual keyboard mode in alphabetical order, in which the user navigates through the letters using the arrow keys and inserts the selected letter in the text using the OK key on the remote.

Our model allows the user to use the text input mode that he chooses when he wants, as illustrated in Figure 1. We choose the virtual keyboard mode as the default state of the component to the detriment of the cell keypad mode, because this mode can be more intuitive and comprehensive. Users not used to write text on cell phones feel difficulties if they tried to write using this mechanism. As we noticed during the analysis of questionnaires, they were answered mostly by people with high levels of education (80% at the beginning of undergraduate college) and age not too young nor too old (80% of people were between 18-44 years old). 30% of people said that seldom or almost never send messages via cell phone.

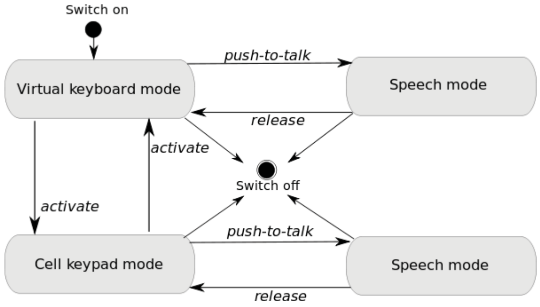

Figure 1: State diagram of the model

The user can use the speech mode concurrently with other text entry modes in a natural way by pressing the push-to-talk key on remote. Moreover, the switch between modes "Virtual keyboard" and "Cell keypad" is made through the activation of a button on the interface.

4.3 The prototype in use

Based on the requirements and principles defined on the model, two interfaces were independently designed, and the strengths of each were combined to create a third interface. Next, a functional prototype was implemented to allow the evaluation of the interface designed. We made a horizontal prototype that simulates the features that were described in the model. It was created in Java to run on a PC, aiming a short development time. Figure 2 shows the four major states of the text input component prototype.

The speech mode would allow the user to write in the selected text box dictating the words he or she wants, but no speech recognition engine was implemented. This functionality was tested using the Wizard of Oz technique. In addition to the Dictate state, shown in the upper left of Figure 2(1), the speech mode of the component has also a state where only editing commands are recognized by the speech recognition engine (although it is not available in the current version). The cell keypad mode (Figure2(2)) allows the user to write in the selected text box using the numeric keys of the remote, just as it is traditionally done in telephones and

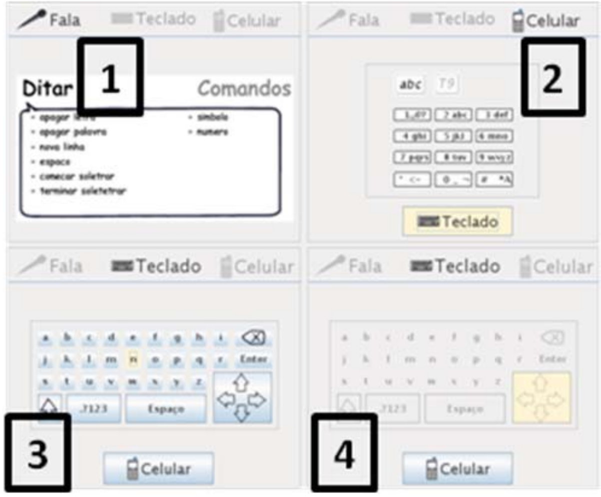

Figure 2: Four main states of the text input component prototype (in Portuguese): (1) Speech mode in Dictate state, (2) Cell keypad mode, (3) virtual keyboard mode and (4) virtual keyboard mode in cursor movement state

cell phones. It should allow both multi-tap and predictive text (although only the first mode is currently available). Finally, the virtual keyboard mode (Figure 2(3))allows writing text by selecting the desired characters using the arrows and OK key. By selecting the button on the bottom right of the keyboard, the function of the arrow keys on the remote control changes to let the user to move the cursor in the text box and, therefore, the buttons on the virtual keyboard become disabled, as shown in Figure 2(4).

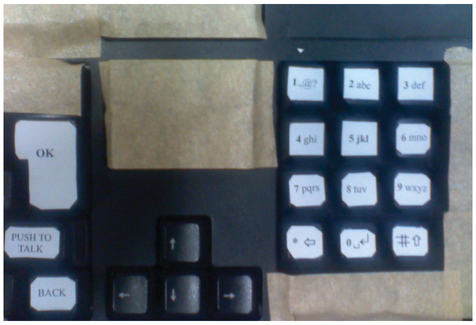

Figure 3: Adapted keyboard to interact with the prototype

The switch between the modes “virtual keyboard” and “cell keypad mode” is easily performed. To go from the virtual keyboard to the cell keypad mode, users should press the directional keys to give focus to the “celular” button, and press OK on the remote control (Figure 2(3)). The switch in the opposite direction is even easier, because it is necessary only to press the OK key (Figure2(2)). In case the user wants to activate the speech recognition system, he only needs to press and hold the push-to-talk key on the remote control. The interface of the component changes to show the commands that can be intercalated while the user dictates the desired text. As soon as the push-to-talk key is released, the interface shows a mode according to the previous state.

Below we list some important features of the prototype:

(1) The prototype was developed to be simple and to allow the tester to focus its attention only in the text entry component. In the screen, only a text box and the designed text entry component are shown.

(2) The font sizes used in the prototype led to a component visible area of approximately 400 x 330 pixels, which fits even on a standard-definition television without causing visualization difficulties.

(3) The names chosen to the three text entry modes presented in the interfaces needed to be small because of the low screen resolution adopted. One of the modes is identified by just one word. The icons were also chosen in order to be easily identified and associated with the mode.

(4) No help screen was developed because the text entry component would be described in the help screen of the application that uses it.

Finally, interaction with the real system should be performed with a special remote control containing a built-in microphone and a push-to-talk key that must be pressed to activate the speech recognition engine. In user tests with our prototype, the input data was performed using an adapted computer keyboard where some keys were relabeled and unused keys were covered with adhesive paper, in order to not confuse the user (Figure 3).

5. USABILILITY EVALUATION

So far, the latest stage in our project was the usability evaluation of the prototype. We used two usability inspection methods: Heuristic Evaluation and Cognitive Walkthrough. We decide to used them given the strong acceptance and recognition they have [15]. We also conducted user testing through the thinking aloud with Wizard of Oz technique.

The think aloud tests were performed by 4 pairs of users, following one of the recommendations of Flores et al. (2008), who note that to test pairs allows individuals to express more naturally their actions and opinions. Figure 4 show some frames taken from the recorded videos during the tests. The Heuristic and Cognitive Walkthrough evaluations were applied each one to three experts and the Heuristic Evaluations were performed using the general heuristics proposed by Nielsen and Mölich [2].

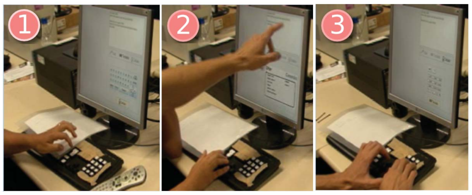

Figure 4: Frames taken from the recorded videos during the tests, in which the 3 different modes were used: (1) Virtual keyboard mode, (2) speech mode and (3) cell keypad mode.

All the information of evaluations and tests were summa-

[2] The original list of 9 heuristics from Nielsen and Mölich (1990) has been refined by Nielsen (*http : //www.useit.com/papers/heuristic/heuristic$_l$ist.html*)

rized in order to make a general compendium of a number of problems in the proposed mechanism. Each of these problems was awarded its corresponding level of severity. Seven of them, due to greater severity, were used to create a list of recommendation changes. The main problems were:

(1) The speech mode has a different activation mechanism from the others, but this is not indicated in the interface. There was not a consensus among the evaluators if this mode should or not be activated only by clicking on the push-to-talk button, as it is now. However, if this happens, that difference must be clearly indicated in the interface.

(2) The activation mechanism of the mobile phone style mode and the virtual keyboard mode should be better crafted. The buttons that allow the activation are distant from the tabs that indicate which mode is active. The most intuitive seems to be the use of own tabs as activation mechanism.

(3) The mobile phone style mode does not give a clear indication that the buttons shown are merely illustrative. The buttons of the interface are shown only to serve as a guide for users who use remote controls with no letters printed on keys. This causes a large amount of error situations, in which users try to use the arrow keys to focus one of the illustrative buttons.

(4) The term "Celular" (cell phone), used to indicate the text input style often used in mobile devices, may not be easily understood by users, who may think it refers to their own mobile phone or to a cell phone call.

(5) The "BACK" button caused a lot of frustration because it could almost never be used. This may have occurred because the component has been evaluated outside the context of an application, where it would certainly have clearer functionalities. Another problem caused by this key was the association made between the key and the backspace key of a traditional computer keyboard.

(6) The key used to enter and to exit the cursor movement state, accessible from the keyboard mode, was not the same. The key to enter was the "OK" key, but the key to exit was the "BACK" key. This was extremely counterintuitive. The possibility of inserting new lines using the "OK" key, offered when this state is activated, is not worth the amount of errors generated.

(7) The interface of the speech mode is not clear enough and causes a lot of problems. The use of two columns of commands in the Dictation state makes the user associate the right column with the Commands state. The interface is not clear neither on how to access the Commands state, nor indicate clearly the difference between the Dictate state and Commands state. Not even the list of available commands in the Dictate state is satisfactory.

In addition to the problems identified in the design of the text input component, some serious problems on the specific prototype implementation ended up gaining more prominence in the evaluations than it was expected:

(1)The cursor that indicates where the next character will be inserted is not shown in any of the input text modes.

(2) The adjustment made on the numeric keypad resulted in the "Num Lock" key being used to insert symbols. This made the rest of the numeric keypad stop working whenever the symbols key was pressed an odd number of times.

(3) The predictive text functionality (T9), despite not being implemented, is indicated by a label on the interface.

(4) The tests of the speech mode, using the technique of the Wizard of Oz, did not allow the text to be inserted into the text area of the prototype. A parallel screen was used.

Observing the tests with users, we noticed that the think aloud test was not adequate when evaluating the Speech mode, since users were asked to say all they were thinking and this action did not fit very well in the case of voice interfaces.

6. IMPROVEMENTS

In order to solve the main problems of the designed component listed in the previous section, the following changes should be made for the next version:

(1) To allow the three text input modes to be activated by selecting the corresponding tab. Thus, the "Celular" ("Mobile Phone") button is removed from the virtual keyboard mode and the "Teclado" ("Keyboard") button is removed from the mobile phone style mode. When the virtual keyboard mode is activated, the focus can move freely between the character buttons and the two other tabs. When the speech or mobile phone style mode are activated, the focus can move only between the two other tabs. The button push-to-talk should be preserved and the user should be able to press it regardless of the active mode, which makes it a shortcut to the speech mode. However, if it is pressed while the speech mode is not active, when released, the previous mode should be activated again. In addition, to increase the clarity and to consider users who do not speak English, it should be relabeled to "Segure para falar" ("Hold to talk"). These changes aim to solve the problems 1 and 2, and also have some impact on problem 3, as it does not let the focus fixed on a single button while the mobile phone mode is active.

(2) The graphical interface of the mobile phone style mode should be improved so that no doubt remains that the buttons shown are merely illustrative. This change also aims to solve the problem 3.

(3) The tabs that identify the text input modes should be bigger to allow each mode be identified by more than one word. The mobile phone mode would be called "Estilo celular" ("Mobile phone style"), the virtual keyboard mode would be called "Teclado virtual" ("Virtual keyboard"). The name of the speech mode could remain the same. This aims to eliminate possible confusion explained by problem 4.

(4) The "BACK" key should be relabeled to "VOLTAR" ("back") to consider users who do not speak English and also solve the problem 5. If time is available, the context of the application should be used in the prototype, to make it clear what is the function assigned to the BACK key.

(5)The OK button should also be used to exit the cursor movement state. This would solve problem 6.

(6) The organization of the items shown in the speech mode should be improved, to make it clear to the user that "Dictate" is just one of the possible states of this mode. The command "Comandos" ("Commands") should be listed along with other possible commands of the Dictate state. Also, the command "Ditar" ("Dictate") should be listed along with other possible commands of the Commands state. The "Apagar linha" ("delete line") command should be added to the

list of commands of the Dictate state. The commands that require some extra word, as in the case of "Insert symbol" and "Insert number", should indicate that in the interface. All these changes address the problems related to the speech mode, and grouped in item 7.

In future work, we intend to implement a new prototype that includes the suggested changes and targeting a set-top box (STB), so that new evaluations may be carried out, taking into account previously ignored factors, such as performance and use of a real remote control.

7. FINAL REMARKS

Given the differences between the iDTV and PC platforms with respect to user tasks associated with viewing, navigating, and interacting, in this paper we have proposed a interface model based on multiple input modes, and its corresponding software component, to deal with the problem of text entry in iDTV applications, and presented a prototype making use of the component.

During the development of the prototype, we conducted usability tests using the think aloud protocol, and usability inspections using the heuristic evaluation and cognitive walkthrough techniques. The evaluations allowed the detection of both a number of problems and of several improvement opportunities; at the same time they highlighted the importance of using complementary text input modes in order to satisfy the needs of different users. Overall, the evaluation results suggest that the proposed approach provides a satisfactory level of usability by overcoming the limitations of text input in the context of the user-interaction with iDTV applications.

With respect to future work, we plan to tackle both the problems and new requirements identified by the specialists in the usability evaluations. We also intend to implement a new prototype that, including the suggested changes, runs on a set-top box, so that new evaluations may be carried out — this is important since it allows to take into account factor which have been ignored in the current evaluations, in particular the user performance when a real remote control is used.

Acknowledgments. We thank the financial support received from FAPESP, CNPq, CAPES, FINEP and the MCT.

References

[1] G. D. Abowd, E. D. Mynatt, and T. Rodden. The human experience. *IEEE Pervasive Computing*, 1(1):48–57, 2002.

[2] A. Berglund, E. Berglund, A. Larsson, and M. Bång. Paper remote: an augmented television guide and remote control. *Universal Access in the Information Society*, 4(4):300–327, 2006.

[3] S. J. Castellucci and I. S. MacKenzie. Unigest: text entry using three degrees of motion. In *CHI '08 Extended Abstracts of the ACM Conference on Human Factors in Computing Systems*, pages 3549–3554, 2008.

[4] P. Cesar, K. Chorianopoulos, and J. F. Jensen. Social Television and User Interaction. *Comput. Entertain.*, 6(1):1–10, 2008.

[5] A. L. Cox, P. A. Cairns, A. Walton, and S. Lee. Tlk or Txt? Using Voice Input for SMS Composition. *Personal Ubiquitous Comput.*, 12(8):567–588, 2008.

[6] M. J. Darnell. Making Digital TV Easier for Less-Technically-Inclined People. In *UXTV '08: Proc. 1st Int. Conf. Designing Interactive User Experiences for TV and Video*, pages 27–30, 2008.

[7] F. de Jesus Lima Gomes, J. V. de Lima, and R. A. de Nevado. O papel Comum como Interface para TV Digital. In *IHC '06: Proceedings of the VI Brazilian Symposium on Human Factors in Computing Systems*, pages 29–32, 2006.

[8] L. C. De Miranda, L. S. G. Piccolo, and M. C. C. Baranauskas. Artefatos físicos de Interaç ao com a TVDI: desafios e diretrizes para o cenário brasileiro. In *IHC '08: Proc. VIII Brazilian Symposium on Human Factors in Computing Systems*, pages 60–69, 2008.

[9] A. Dix, J. Finley, G. Abowd, , and Beale. *Human-Computer Interaction.* Prentice Hall, 3rd edition, 2004.

[10] R. Fagá Jr., B. C. Furtado, F. Maximino, R. G. Cattelan, and M. d. G. C. Pimentel. Context Information Exchange and Sharing in a Peer-to-Peer Community: A Video Annotation Scenario. In *SIGDOC '09: Proceedings of the 27th ACM International Conf. Design of Communication*, pages 265–272, 2009.

[11] M. Ferati, S. Mannheimer, and D. Bolchini. Acoustic Iteraction Design Through "audemes": Experiences With the Blind. In *SIGDOC '09: Proc. 27th ACM International Conf. Design of Communication*, pages 23–28, 2009.

[12] G. Geleijnse, D. Aliakseyeu, and E. Sarroukh. Comparing Text Entry Methods for Interactive Television Applications. In *EuroITV '09: Proc. Seventh European Conf. European Interactive Television Conf.*, pages 145–148, 2009.

[13] K. Go, H. Konishi, and Y. Matsuura. Itone: A Japanese Text Input Method for a Dual Joystick Game Controller. In *CHI '08: CHI '08extended abstracts on ACM Conference on Human Factors in Computing Systems*, pages 3141–3146, 2008.

[14] S. Harada, J. O. Wobbrock, J. Malkin, J. A. Bilmes, and J. A. Landay. Longitudinal Study of People Learning to Use Continuous Voice-Based cursor Control. In *CHI '09: Proc. 27th Int. Conf. Human Factors in Computing Systems*, pages 347–356, 2009.

[15] T. Hollingsed and D. G. Novick. Usability Inspection Methods After 15 Years of Research and Practice. In *SIGDOC '07: Proc. 25th ACM International Conf. Design of Communication*, pages 249–255, 2007.

[16] C. L. James and K. M. Reischel. Text Input for Mobile Devices: Comparing Model Prediction to Actual Performance. In *CHI '01: Proceedings of the SIGCHI Conf. Human Factors in Computing Systems*, pages 365–371, 2001.

[17] Y. Nakatoh, H. Kuwano, T. Kanamori, and M. Hoshimi. Speech Recognition Interface System for Digital TV Control — Special Issue Applied Systems. *Acoustical science and technology*, 28(3):165–171, 2007-05.

[18] N. Patel, S. Agarwal, N. Rajput, A. Nanavati, P. Dave, and T. S. Parikh. A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India. In *CHI '09: Proceedings of the 27th Int. Conf. Human Factors in Computing Systems*, pages 51–54, 2009.

[19] D. d. C. Pedrosa, D. A. Vega Oliveros, M. d. G. C. Pimentel, and R. P. d. M. Fortes. Text Input in Digital Television: a Component Prototype. In *Adjunct Proc. of EuroITV '10: Proc. 8th Int. Interactive Conf. Interactive TV and Video*, pages 1–4, 2010.

[20] L. S. G. Piccolo and M. C. C. Baranauskas. Desafios de Design para a TV Digital Interativa. In *IHC '06: Proc. VI Brazilian Symposium on Human Factors in Computing Systems*, pages 1–10, 2006.

[21] G. Verhulsdonck. Issues of Designing Gestures into Online Interactions: Implications for Communicating in Virtual Environments. In *SIGDOC '07: Proc. 25th ACM International Conf. Design of Communication*, pages 26–33, 2007.

[22] A. D. Wilson and M. Agrawala. Text Entry Using a Dual Joystick Game Controller. In *CHI '06: Proc. SIGCHI Conf. Human Factors in Computing Systems*, pages 475–478, 2006.

[23] K. Wittenburg, T. Lanning, D. Schwenke, H. Shubin, and A. Vetro. The Prospects for Unrestricted Speech Input for TV Content Search. In *AVI '06: Proc. Working Conf. Advanced Visual Interfaces*, pages 352–359, 2006.