Users and Usage Driven Adaptation of Digital Talking Books

Carlos Duarte, Luís Carriço

LaSIGE/Informatics Department, Faculty of Sciences, University of Lisbon
Campo Grande, Edifício C6, 1749-016 Lisboa, Portugal
{cad,lmc}@di.fc.ul.pt

Abstract

Digital Talking Books (DTBs) offer to visually impaired users an evolution of analogue talking books that mimics the interaction possibilities of print books. They offer the possibility to randomly access book content, present images and tables, annotate books, amongst other features. However, current DTB players with their usability and accessibility problems, limit the enjoyment of the reader. DTB's characteristics demand for a truly multimodal platform to take advantage of all the possibilities offered. Furthermore, varying conditions affect the operation of a DTB player: different physical and cognitive characteristics of the reader, environmental changes and book types. To improve interaction and the reading experience we advocate providing the multimodal player with adaptive features. In this paper we present an adaptive framework, designed to operate in a multimodal environment, which expands on traditional user model only adaptation. We then describe the characteristics of an adaptive multimodal DTB player, built from the ideas presented in the framework.

1 Introduction

Visually impaired people may find standard printed documents inaccessible. For people with such disabilities, access to literary content has been done mainly through talking books. These have been offered on analogue media, such as audiocassettes, providing human speech recordings of a wide array of print material. However, analogue media are limited in several aspects when compared with a printed book: They are linear presentations, which leave much to be desired when reading materials are accessed randomly; Readers cannot interact with analogue books by placing bookmarks or highlighting material; Readers do not have access to the spelling of the words they hear, which is a long time complaint of talking book users; Analogue books offer only one version of the work. If, for example, the book contains footnotes, these are either read when referenced or grouped at a location out of the flow of the text. Neither of these options will please all the readers.

A Digital Talking Book (DTB) is the digital counterpart of a talking book. Digital media allows dealing with the shortcomings mentioned before: DTBs offer the reader the possibility to move around in a book as freely as in a printed book; DTBs allow the creation of bookmarks, highlights and annotations; DTBs store the full text of the printed book, synchronized with the audio presentation, allowing readers to locate specific words and hear them spelled; DTBs allow easily skipping or reading the footnotes when desired.

The DTB goes beyond the limits imposed on analogue books because it can include not just the audio recordings of the book, but full text content and images as well. Because the text is synchronized with the audio, a DTB offers multiple sensory inputs to readers. This can be beneficial to learning-disabled readers and to other target audiences such as blind, visually impaired, physically handicapped and otherwise print-disabled readers (Moreno & Mayer, 2000). For these audiences the DTB offers a significantly enhanced reading experience. For other audiences, balancing DTB modes and media can be explored to overcome the cognitive limitations of human perception and attention (Gazzaniga, Ivry & Mangun, 1998).

DTB developments over the last years lead to the appearance of evolving specifications, with the Daisy Consortium (Daisy, 2005a) being responsible for the major work done in the area, and the publication of several standards (DAISY 2.0 in 1998, DAISY 2.01 in 1999 and DAISY 2.02 in 2001). Finally, in 2002, with cooperation from the Daisy Consortium, the National Information Standards Organization (NISO) published the current DTB standard ANSI/NISO z39.86-2002 (NISO, 2002), also referenced by some as DAISY 3.0. The standard focuses on the files, their structure and content, needed to produce DTBs. However, specifications for playback devices are absent from the standard. An auxiliary document, the Playback Device Features List (NISO, 1999), created during the standard's development, describes the main features that playback devices should possess, but it is not normative and does not present specific implementation solutions. As a result, several DTB players developed either for the earlier DAISY standards, or for the later ANSI/NISO standard, are capable of DTB playback, but adopted different solutions for the presentation and interaction. Some of those solutions suffer from usability flaws that are detrimental to the reading experience. Even from the accessibility point of view, several faults can be observed, preventing the use of the playback devices by some members of the intended audience.

The DTB playback environment, as suggested by its specification, is most notoriously a multimodal environment. The book's textual content is presented synchronized with an audio narration. When supported, speech recognition can be employed as an input mode. Alerting the reader to the presence of a bookmark can be done either visually or audibly. Several other situations can be identified where the use of multimodal interaction will benefit the reader. However, most DTB players don't take advantage of the possibilities presented by the use of multimodalities, and don't go any further than the synchronized presentation of text and narration. We have developed a DTB player that employs multimodalities and thus benefits from improved features and accessibility. Anticipating the different contexts of usage, the diverse playback environments, and the multitude of user characteristics, the DTB player was endowed with adaptation capabilities (Duarte & Carriço, 2004a). By combining the possibilities offered by multimodal interaction and interface adaptability we have developed a DTB player able to overcome the limitations of current DTB players, thus enabling an improved reading experience for both print-disabled and non print-disabled users.

In the following section we present related work, starting with a discussion of the characteristics and limitations of current DTB players, resulting of a review of eight DTB players. We then focus our attention on multimodal and adaptive interfaces, their uses, advantages and limitations. Next, we describe a framework for development of adaptive applications designed to operate in multimodal environments. We then present the developed DTB player prototype, based on the adaptive applications framework. Further details on the use of multimodalities and on the implementation of several adaptation components are given. Finally we conclude and present directions for future work.

2 Related Work

2.1 DTB Players

According to the Playback Device Features List (NISO, 1999), the NISO DTB committee recommends that three types of playback devices be developed: First, a basic DTB player, defined as a portable unit capable of playing digital audio recordings, for use mostly by less sophisticated talking book readers who wish to read primarily in a linear fashion. Second, an advanced DTB player, also portable, but designed for use by students, professionals, and others who wish to access documents randomly, set bookmarks, etc. Finally, a computer based DTB player, consisting only of software and being the most complete and sophisticated of the three.

For this final configuration, the Playback Device Features List defines as essential a set of features, including the following: no need to use visual display to operate device, variable playback speed, document accessible at fine level of detail, usable table of contents, easy skips (moving sequentially through the elements), ability to move directly to a specific target, ability to manage notes, reading of notes, setting and labelling bookmarks, automatic bookmark at stop, ability to add information (highlighting and notes), spell words, fast forward and fast reverse, human and electronic speech must be available, presentation of visual elements in alternative formats (speech, for instance), amongst others.

The Daisy Consortium publishes and maintains a list of playback devices (Daisy, 2005b). The majority of these devices support the oldest DAISY standards, while the newest ones support the more recent ANSI/NISO standard. We have conducted a heuristic evaluation of the following players referenced in the list: AMIS 1.3, TAB Player, TPB Reader, Victor Reader Soft, gh Player 2.0 and Book Wizard Reader. Two other players (EaseReader and eClipseReader) were also considered for evaluation, but no demonstration version was available at the time, with only the product information presented in their web pages accessible. The players have been evaluated according to their navigation features, the possibilities of personalization offered, the implementation of bookmarks and annotations, and the use of different modalities.

Navigation wise, a set of features was found in all DTB players: moving forward and backward an element, moving sequentially through the chapters and sections, and the capacity to move to any point of the table of contents. These functionalities meet some of the requirements elicited in the Playback Device Features List. However, some of the implementations present usability problems. Two different implementation's choices for the sequential navigation have been found: the first, in which different sets of commands are used for chapter and section advance, and the second, in which one command selects the navigation unit and another command is responsible for advancing the narration with the previously chosen navigation unit. There may also be different elements (word, sentence or paragraph) for element navigation. All this variability creates situations in which the user may not be aware of the result of his actions (advancing one section instead of a chapter, or a sentence instead of a word). Only TPB Reader deals with this problem, always displaying the navigation element being used. Other problems are found when considering some of the table of contents implementations. Examples include poor presentation of the contents (using only font-size to indicate the heading level of the table of contents' entry, without any kind of alignment or indexing), stopping the narration when consulting the table of contents, not being able to display the table of contents and the book content simultaneously, or display them in overlapping windows. Regarding other features requested, half of the players supported fast forward and reverse, and only two allow the user to navigate to any point of the text by selecting it with the mouse.

Most of the players offer some kind of personalization features. The most common is the possibility to alter the font size. Half the players allow the user to select the text and background colours. These basic personalization capabilities allow visually impaired users some degree of control over the visual presentation, thus enabling the use of this modality as a complement to audio interaction. Nevertheless, there are implementations that almost render inoperative the initial purpose. This happens when, for instance, enlarging the font moves the text outside the display area and the user has no control over the scroll bars to scroll the window contents and move the text back into view. Other players don't support the resizing of the fonts used for the TOC and the annotations. Others yet, have toolbars that are not resizable. On a more positive note, all but one player allow for variable reading speeds. Only one player gives the reader the possibility to select whether or not footnotes and page numbers should be read during the book's narration.

Bookmarks and annotations are supported by the majority of the players. Only one player does not offer the possibility to create bookmarks, and three don't allow for user created annotations. Two of the players support voice annotations. In most players, bookmarks and annotations can be used as navigational elements, similar to chapters and sections. For the reader, annotation awareness is an important part of the talking book experience. However, only EaseReader alerts for the presence of an annotation with both visual and audio warnings, and TAB Player stops the narration and presents a dialog box alerting the reader to the presence of an annotation. Users of the other players have to consult an annotation's list to be able to read them and be aware of their placement. Only one player allows for user labelling of the bookmarks, and even in this case, the label has to be a number. The other players label the bookmarks with sequential numbers, or with the text of the bookmark creation point. Of the players supporting bookmarks, only one does not create an automatic bookmark on stop.

The Playback Device Features List mentions "no need to use visual display to operate device" as one of the essential features of a computer based DTB player. However, none of the evaluated DTB players would perform adequately in a purely non-visual interaction. The use of voice in a truly multimodal input environment would enable the needed interaction capabilities, but only the AMIS player, with its support for plug-ins, is capable of voice recognition. By relying exclusively on mouse and keyboard inputs, a visual display becomes an indispensable device for taking advantage of all the features offered by current DTB players. Basic navigation and player control (starting, stopping, issuing sequential navigation commands) could still be achieved without a visual display, but for more advanced tasks (annotating for example) the visual display is necessary. This forces users that are blind or suffer from severe visual impairing disabilities to memorize a large number of keyboard shortcuts. In the case of users that, in addition to the visual impairments, also suffer from motor disabilities, the interaction with the DTB player is impossible. Even if the user is able to memorize and use the keyboard shortcuts, she may not be capable of using all the features of the player, because several players do not offer all the available functionalities through keyboard shortcuts.

By contrast, all the players provide visual and audio outputs. Most of the players support speech generation, which is used mainly to provide audio feedback by reading aloud the commands issued by the user, or to vocalize the contents of dialog boxes. In some cases, speech synthesis may also be used for the narration, if the user wishes. Some players also read aloud entries of the table of contents, either using pre-recorded speech, or by synthesis. However, when this is the default unchangeable behaviour, it leads to interruptions in the narration whenever the reader consults the table of contents, even if no navigation action is taken. Another of the essential features, "presentation of visual elements in alternative formats" is disregarded by all the players. The images and tables are presented visually, but there is no alternative way of presentation, thus rendering them imperceptible to blind users. The use of sound to signal the presence of annotations, bookmarks, footnotes and other supporting content is mostly ignored by all the players, with the exception being the aforementioned EaseReader, and its capability to signal the presence of bookmarks and annotations both visually and audibly.

In summary, the players analyzed do not follow several of the recommendations of the ANSI/NISO standard development committee. While for a non visually impaired user those failings aren't critical, for visually impaired users they can detract strongly from their reading experience. A more careful use of multiple modalities can contribute to improve the reading experience by offering more input and output possibilities, and the introduction of a carefully designed adaptive interface could help the readers better explore not only the features offered by the interface, but also the book content itself.

2.2 Adaptive Multimodal Interfaces

Through multimodal interfaces, users can take advantage of their natural communication modes during human-computer interaction, selecting the best mode or combination of modes for each situation or task. Multimodal interfaces may be used with very large or very small devices, and also in sensor-rich environments (Cohen & McGee, 2004). Another important benefit resulting from the ability of using multiple modalities in an interface is mutual disambiguation, where the information provided by one or more sources may be used to resolve ambiguities in information from other sources, thereby reducing errors (Oviatt, 1999).

In order to take advantage of human information processing abilities (which include attention, working memory and decision making) designers of multimodal systems should devise their systems in such a way as to maximize human cognitive and physical abilities. Possible examples of guidelines are:

- Avoid unnecessary information presentation in two different modalities in cases where the user must pay attention simultaneously to both information sources to comprehend the presented material. This redundancy might increase the cognitive load at the cost of the learning material (Kalyuga, Chandler, & Sweller, 1999).
- Maximize the advantages of each modality in order to reduce the memory load in certain tasks and situations, as illustrated by the following modality combinations (Wickens, 1992):
  - System visual presentation combined with manual user input for spatial information and parallel processing. In the case of a DTB player this recommendation can be used to guide the design of the user interaction with the visual components of the interface (windows presenting the main content, table of contents, images, etc.). Their spatial distribution across the screen space can be done in a more efficient and natural fashion using manual input.
  - System auditory presentation combined with speech user input for state information, serial processing, attention alerting, or issuing commands. For a DTB player application, this guideline recommends the use of voice commands for the player state changes, and the use of auditory modes to alert to the presence of images, annotations and other miscellaneous content.

However, as shown in the previous section, currently available DTB players do not follow these recommendations, resulting in a possibly overly complex reading experience for the user, with usability and accessibility problems. To improve their coverage, reliability and usability, multimodal interfaces should be designed to automatically learn and adapt to important user, task and environment parameters (Oviatt, Darrell & Flickner, 2004). Multimodal interfaces should adapt to the needs and abilities of different users, as well as different usage contexts. Dynamic adaptivity allows the interface to degrade gracefully by introducing complementary and supplementary modalities according to changes in task and context. Individual differences (for example, age, preferences, skills, sensory and motor impairments) can be captured in a user profile and used to determine interface settings such as: allowing gestures to augment or replace speech input in noisy environments, or for speech impaired users; adapting the quantity and method of information presentation to the user and visual display.

3 An Architecture for Adaptation in Multimodal Platforms

DTB playback should be possible over a broad range of platforms, devices and environments. Also, users with different physical and cognitive characteristics are expected to read DTBs. To cope with all these varying conditions two solutions can be envisaged. First, create books that are tailored to expected playback conditions and to probable user characteristics. To enable this design time adaptation, a flexible DTB building framework, called DiTaBBu, has been developed (Carriço, Duarte, Lopes, Rodrigues & Guimarães, 2005). Nevertheless, there will be situations where it is impossible to previously identify or predict all the variables governing the book production. Even in situations where users and platforms share their characteristics, evaluation shows the desire for personalized DTB players (Duarte, Chambel, Carriço, Guimarães & Simões, 2003).

When considering books for a more heterogeneous audience, the identification of characteristics concerning the book's production is more difficult. This leads to the proposed second solution: moving the adaptation from design time to run time, and create an adaptive DTB player. The adaptive player should consider different aspects, from personal reader preferences, to capacities of the playback machine and the playback environment.

To take into account all the changing conditions we propose an adaptive framework. Adopting a model-based approach (Paternò, 1999), the framework expands on the traditional user model driven adaptation, and includes several other models as can be seen in figure 1.

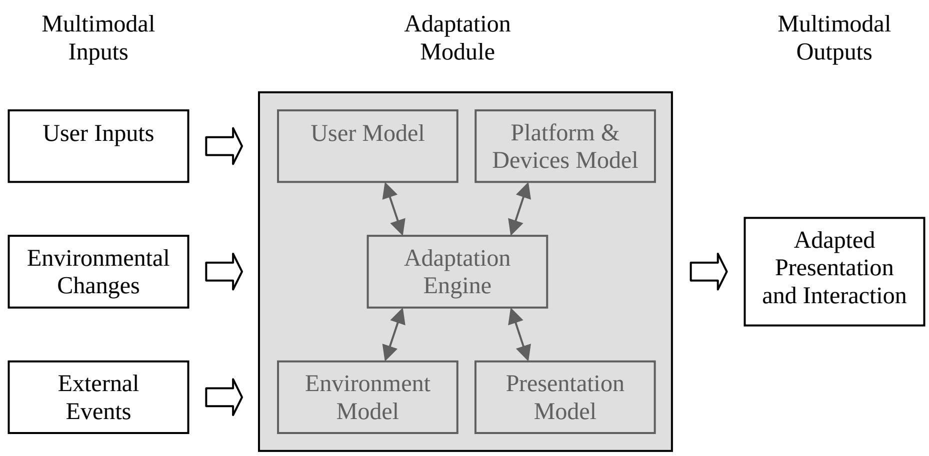

Figure 1: The adaptation framework

This framework for adaptation in a multimodal platform derives its inputs from user interaction, from perceived changes in the environment (a raise in the ambient noise might trigger the disabling of voice inputs, for instance) and from events external to the adaptation module, but that can be raised by the application (displaying an image might prompt the redistribution of the other visual components). The adaptation engine, based on the information contained in the models and the inputs received, adapts the presentation and the interaction possibilities of the interface, and performs the necessary updates to the different models. Next, the four models are described in greater detail.

- *User Model:* This is the traditional model used in adaptive applications, where relevant user preferences and characteristics are described. For a DTB player application, the model may include physical attributes such as user's visual impairment level, and preferences describing the preferred interface behaviour for dealing with annotations, or favoured text size, for example.
- *Platform & Devices Model:* The Platform & Devices Model describes the characteristics of the execution platform and of the devices attached to it. Platform attributes relate to invariant characteristics of the

platform, and include, for instance, screen size. Devices represent the software and hardware artefacts that can be present or absent in the platform. The application should be able to adapt in run-time to changes in this model, meaning that if the user plugs a microphone to the system, the application should be able to use speech recognition (if available) as an input modality.

- *Environment Model:* This model describes the environmental characteristics that can have an impact on the presentation and interaction aspects of the application. An example is the ambient noise of the reproduction environment influencing both speech input and output.
- *Presentation Model:* This model describes the components available for presentation and interaction. Each component has a set of templates available, covering the broadest range possible of devices, user groups and environments. For instance, for the Table of Contents component, the visual presentation should be complemented with the possibility to present it audibly, either using pre-recorded content, or by synthesizing it. Another set of higher level templates organize the available components into a presentable version of the interface, and select the input modalities available at any given instant.

Translating the framework into a practical application begins with the process of determining the adaptation variables to be used, the adaptable components of the interface, and the attributes of the different models. Some of the models' attributes will be equivalent to input variables, while others have to be derived from those, and others yet will have to be acquired by other means (Kobsa, Koenemann & Pohl, 2001). In the case of a DTB player, input or adaptation variables are, for instance, the commands issued by the reader (user input), measurements made on the noise level (environmental changes) and events raised by the book's playback, like displaying an image or alerting for the existence of an annotation (external events). The adaptable components can be categorized according to three dimensions: Interaction components reference the input and output modalities available, their status (enabled or disabled) and their use (cooperatively or individually); Content components are related to the enhancement of the book with available sounds, images or other media (Carriço, Guimarães, Duarte, Chambel & Simões, 2003), to text translation and to hiding or revealing parts of the book; Presentation components deal with size and colour of used fonts, placement of visual components, type of audio signals used for alerts, synchronization units, etc. The models will store relevant information for the presentation and interaction with the DTB player, and will be updated whenever necessary with information extracted from the interaction.

4 The Digital Talking Book Player

The elements of a DTB can be separated into several components. Four component's categories can be identified: the book content, the table of contents, the annotations, and other miscellaneous content, including tables, images and side notes. For a DTB to reach all the proposed audiences all the components must be available in several modalities, thus justifying the need for a truly multimodal environment. Regarding output modalities, all the elements should be presented visually, as well has having a recorded version of the text, or, if not available, the playback platform should be able to use speech synthesis to reproduce them. Special care needs to be devoted to the reproduction of images and tables. While visual reproduction is trivial, how are images and tables transmitted using only words and sounds? These are problems that have already been studied by developers of audio browsers (Goose & Möller, 1999) (James, 1997), and some solutions are presented by the W3C Web Accessibility Initiative (World Wide Web Consortium, 2004). To present either images or tables to visually impaired users the solution is to prepare a description of the object and present that description using speech. A description is essential as an alternative means of presenting images but also particularly important for tables. An apparently good solution for tables would be to read them row by row or column by column. However, using that solution it is harder to transmit the contents of the table. A solution where a description is provided conveys the meaning of the table to the reader in a more satisfying way.

The benefits of an adaptive multimodal platform extend to other situations. During book reading several elements may take the reader's attention away from the main content, or from the established linear presentation. These elements include the images and tables mentioned before, but also cross-references, footnotes, side notes, floating text boxes, and even user created annotations. Whenever one of these elements is encountered during the narration one of the following behaviours must be selected: present the element, alert to its existence without presenting it immediately, or simply ignore it. The selection of the behaviour to adopt can be left to the user, or can be an initiative of the adaptive interface based on past user selections. Whoever decides, if the current option is to alert the user to the presence of one such element, the signalling should be done with more than one modality. This enables reaching users that may not perceive signals in one of the modalities, and even for other users it may be a more effective way to alert them. If the users consider such alerts to be too distracting, the option to select the modality to use should be available.

Regarding the input, multimodalities should also be present. The traditional input modalities for a PC-based application, the mouse and keyboard, could be combined or replaced in some situations by speech input. Evaluation of a previously developed HTML based DTB player (Duarte & Carriço, 2004b) proved the merit of offering speech input as an alternative interaction modality. This was shown by user's adoption of verbal commands for some tasks, most notably playback control tasks, which are the most used. Nevertheless, special care must be taken to avoid some of the problems of using speech as an input. The major difficulty with the use of speech is providing a noise free signal as input to the speech recognizer. Care should also be taken to avoid saturation of the signal. When the playback environment is not controlled possible noise generating factors are the surrounding environmental noise, and the audio narration of the book itself. This last factor can be controlled by the use of headphones whenever possible. The first factor is harder or maybe even impossible to control when considering mobile DTB players used in outdoor settings. A possible workaround is the adoption of a push-and-talk interface, where the speech recognizer only accepts inputs after being ordered to by the user. A simple implementation consists in having a button in the interface that is pressed by the user whenever a command is to be issued. An evolution to this approach consists of having the speech recognizer always on, waiting to recognize a command word. When the user wishes to issue commands to the book, the command word must be spoken first, changing the speech recognizer to interaction commands recognition mode, and then the user can proceed as usually. In a completely controlled environment, where the user has headphones available or opts not to listen to audio, and with a low noise level, the speech recognizer can be left to operate in continuous speech recognition mode.

In summary, this adaptive multimodal DTB player employs text, audio narration and images, as output modalities responsible for transmitting the book content to the reader. Sound is also used as a means of alerting the user to the presence of content and interaction events that occur during playback. For input, the player supports interaction via keyboard, mouse and speech. The adaptation is responsible for selecting available modalities for input and output, reacting to changing interaction conditions.

Another responsibility of the adaptation module is the arrangement of the various elements of the player visually. The player’s visual interface can exhibit simultaneously up to four windows: the main content window, the table of contents window, the annotations window and the images window. Figure 2a shows an instance of the player with all the windows visible. The main content is shown in the centre of the window, the table of contents to the left, with annotations in the top right and images in the bottom right, a possible visual configuration.

If the reader is not satisfied with the configuration she can move any window to a general position, and the player will rearrange all the windows' positions automatically. Figure 2b shows the interface after the user moved the images window to the bottom left. As can be seen, the height of the table of contents and annotations windows was changed in response to the user order. The user may also choose to hide any of the visible windows. Figure 2c shows the interface after the user hid the annotations window. The main content window automatically reclaimed the space left unused by the vanishing window.

The adaptation engine reacts to these user originated events by selecting the presentation template to use, which determines, first, the position of each window and then their width and height. The template selection is determined by the number of visible windows and their general placement. This means that there is a different template for each combination of number of windows and their placement. The screen space is divided in general areas determined by the width of visible windows and screen dimensions. The user does not need to place the windows precisely. Moving the window to the desired general area is enough, with the system being responsible for the precise placement. Another event that fires the adaptive window positioning is the resizing of the main application window. If the resizing decreases the width of the application window to an amount bellow the sum of the minimum width of the visible windows, one or more windows will be hidden by the player. If the decrease does not drop the width bellow that limit, or instead there is a width increase, only the widths of the visible windows are affected.

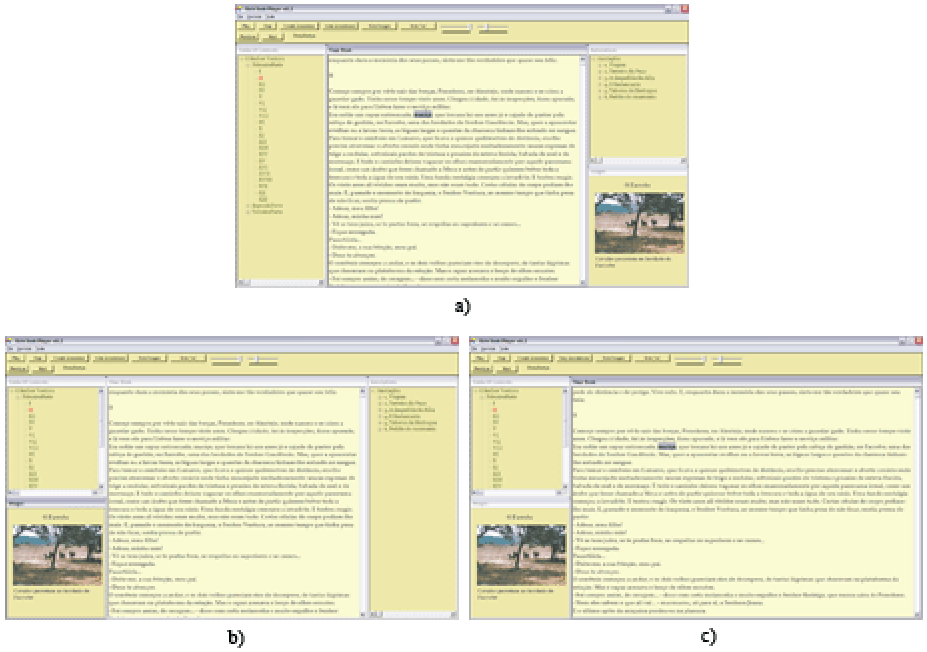

Figure 2: a) The player showing main content, table of contents, annotations and images. b) The interface adapts after the user moved the images window. c) The interface adapts again after the user hid the annotations window

Besides user originated events, other events can initiate the visual adaptation procedure. These playback initiated events can be raised whenever the narration reaches a point in the book where an image or annotation is supposed to be presented. The player's behaviour depends on previous user's actions. The desired player's behaviour can be selected on an n-dimensional behaviour space, where each dimension relates to a behaviour component. A different behaviour space should be built for each of the visual components. We next refer to the annotations window behaviour, but the image window behaviour is similar. Six dimensions have been identified for selecting the player's behaviour when an annotation related event is raised, either by the system or the user. The first dimension is *action* and can take the following values: show, alert and ignore. With show the annotation is presented to the reader, alert warns the reader to the presence of an annotation, but does not take any action to present it, and ignore does not take any action at all. The second dimension is *visibility* and can have one of the following values: always and hide. With always the annotations window is ever present, and with hide the window is closed after the annotation has been presented. The third dimension, *modality*, can take the values visual, audio and all. These values decide what modality (or combination of modalities) to use when presenting the annotations. The fourth dimension is *reading* and takes the values pause and continue. With pause the narration stops when the annotation is presented, and with continue the narration goes on. The fifth dimension is *reaction* and may take the values advance and remain. With advance the book's narration jumps to the annotation's creation point whenever the user selects an annotation. With remain the annotation is presented but the narration does not advance to the annotation's creation point. The sixth dimension is *content* and can have the values list and annotation. With list the annotations' window preferentially presents the list of created annotations, replacing the list with the annotation's text only when the annotation is presented. With annotation the annotations window preferentially exhibits the annotations' text, waiting for a user request to show the created annotations list. These dimensions can be considered independent, but relations between some of them can be enforced, leading to better usability. An example follows: when the modality is audio and the action is show then the reading should be pause, in order to prevent two audio tracks (the book and annotation narrations) to play over each other.

The adaptation engine is responsible for updating the behaviour of the player in response to user's actions. For that to be possible a set of rules had to be formulated. These rules' preconditions are the user inputs relating to annotation interaction. The user can act upon the annotation in three ways: by selecting the view/hide annotations command, by pausing/playing the narration whenever an annotation is being presented, and by selecting to hear the audio presentation of an annotation when she is aware of its presence. From the user behaviour, the adaptation engine updates the player behaviour. If the user ignores alerts to annotation presence a certain number of times, then the engine will change the action dimension to ignore. If, on the other hand, the user always responds to alerts, the engine updates the action dimension to show. If the user pauses the narration whenever there is an annotation present, or if she goes back to an annotation creation point after it has been hidden, the engine will update the reading dimension to pause. If the user resumes narration immediately after the player paused it, the engine updates the reading dimension to continue. If the user repeatedly chooses to hear an audio narration of the annotations the engine updates the modality dimension to all. If the user selects to view the annotations window even when there is no annotation present, the engine updates the visibility dimension to always. These are some examples of the updating rules used to define the player behaviour when dealing with annotation presentation. A set of similar dimensions and rules are also defined for image presentation.

5 Conclusions

DTBs are important to improve the reading experience of visually impaired readers by offering features similar to those of a print book, and that were previously unavailable to users of analogue talking books. However, it is essential that the implementation of those features in DTB players is done considering their usability and accessibility. Our study of eight DTB players referenced by the Daisy Consortium revealed several usability and accessibility problems that hinder their usage. The most serious accessibility problem encountered is the lack of alternative inputs to the mouse and keyboard. A visually impaired user is forced to memorize a large amount of keyboard shortcuts to be able to take advantage of the features of the player, if they are made available in this fashion. Another problem is the lack of alternative presentations to images and tables. These and other problems can be solved by developing a DTB player that is a truly multimodal platform.

To allow for playback in varying environmental conditions, with multimodal inputs and outputs, of different types of books, for users with particular physical and cognitive abilities, we propose the use of adaptation to harness all the changing conditions, and improve the reading experience. We developed an adaptation framework, reacting to user input, environmental changes and application events external to the adaptation module. The adaptation module expands on the traditional user model only adaptation, to include three other models: the platform and devices model, the environmental model and the presentation model. The adaptation engine outputs define a multimodal presentation and interaction platform.

In this paper we also presented details of our first prototype of an adaptive multimodal DTB player, with greater emphasis on the adaptation of the visual components, in response to user generated and playback generated events. We defined presentation behaviour dimensions, and shown rules to update the player behaviour in response to user commands.

5.1 Future Work

Our next step is to perform usability and accessibility evaluation of the develop prototypes. Two evaluation studies are planned. The player will be used by students of an Advanced Interaction Topics course during the spring semester for article reading and annotating. Although the user population is not visually impaired, the test will allow us to find usability problems and to perform a comparison with printed articles reading as not all of the articles of the course will be presented in DTB format. It will also give us a greater understanding of the problems associated with the presentation of technical and reference works, whose characteristics greatly differ from other literary styles. The second planned evaluation setting will occur with blind readers of the Portuguese National Library. This study will allow a thorough evaluation of our implementation of the accessibility features, and show us how a visually impaired population, familiarized with analogue talking books, is capable of evolving to a digital based player.

Simultaneously with the evaluations, we plan to continue evolving the prototype. The greatest challenge expected will come with the migration to platforms with different characteristics. A PDA based player will offer the mobility lacking with PC based players (even laptops), but has reduced processing capabilities and reduced screen size. It will be interesting to see how the adaptation rules developed for one platform translate via the platform and devices model to other platforms.

Acknowledgements

The research discussed here has been funded by the Fundação para a Ciência e a Tecnologia through grant POSI/EIA/61042/2004.

References

Carriço, L., Guimarães, N., Duarte, C., Chambel, T., & Simões, H. (2003). Spoken books: Multimodal interaction and information repurposing. *Proceedings of HCII'2003, International Conference on Human-Computer Interaction*, 680-684.

Carriço, L., Duarte, C., Lopes, R., Rodrigues, M., & Guimarães, N. (2005). Building Rich User Interfaces for Digital Talking Books. In R. Jacob, Q. Limbourg, J. Vanderdonckt (Eds.), *Computer-Aided Design of User Interfaces*. Springer

Cohen, P., & McGee, D. (2004). Tangible multimodal interfaces for safety-critical applications. *Communications of the ACM*, 47(1), 41-46.

Daisy Consortium (2005a). Welcome to the DAISY Consortium Web site. Retrieved February 21, 2005, from http://www.daisy.org

Daisy Consortium (2005b). Playback Tools. Retrieved February 21, 2005 from http://www.daisy.org/tools/playback.asp

Duarte, C., Chambel, T., Carriço, L., Guimarães, N., & Simões, H. (2003). A multimodal interface for digital talking books. *Proceedings of WWW/INTERNET 2003*, 153-160.

Duarte, C., & Carriço, L. (2004a). Identifying adaptation dimensions in digital talking books. *IUI'04: Proceedings of the 9th international conference on Intelligent User Interfaces*, 241-243.

Duarte, C., & Carriço, L. (2004b). Usability evaluation of digital talking books. *Interacção 2004: Actas da 1ª Conferência Nacional em Interacção Pessoa-Máquina*, 9-17.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (1998). Cognitive Neuroscience – the Biology of the Mind. W. W. Norton & Company.

Goose, S., & Möller, C. (1999). A 3d audio only interactive web browser: using spatialization to convey hypermedia document structure. *MULTIMEDIA'99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, 363-371.

James, F. (1997). Presenting HTML structure in audio: User satisfaction with audio hypertext. *Proceedings of ICAD'97*, 97-103.

Kalyuga, S, Chandler, P, & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351-371.

Kobsa, A., Koenemann, J., & Pohl, W. (2001). Personalized hypermedia presentation techniques for improving online costumer relationships. *Knowledge Engineering Review*, 16, 111-155.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, 97, 117-125.

NISO (1999). Digital Talking Book Player Features List. Retrieved February 21, 2005, from http://www.loc.gov/nls/z3986/background/features.htm

NISO (2002). Specifications for the Digital Talking Book. Retrieved February 21, 2005, from http://www.niso.org/standards/resources/Z39-86-2002.html

Oviatt, S. (1999). Mutual disambiguation of recognition errors in a multimodal architecture. *Proceedings of Conference on Human Factors in Computing Systems*, 576-583.

Oviatt, S., Darrel, T., & Flickner, M. (2004). Multimodal interfaces that flex, adapt, and persist. *Communications of the ACM*, 47(1).

Paternò, F. (1999). Model-based Design and Evaluation of Interactive Applications. Springer-Verlag.

Wickens, C. (1992). Engineering Psychology and Human Performance. Harper Collins.

World Wide Web Consortium (2004). Web Accessibility Initiative (WAI) Homepage. Retrieved February 21, 2005, from http//www.w3.org/WAI/