Playback of Rich Digital Books on Mobile Devices

Carlos Duarte, Luís Carriço, and Fernando Morgado

LaSIGE – Faculty of Sciences of the University of Lisbon
Edifício C6, Piso 3, Campo Grande
1749-016 Lisboa, Portugal
{cad, lmc, fmorgado}@di.fc.ul.pt

Abstract. This paper presents the mobile version of Rich Book Player, a Digital Talking Book player for mobile devices. The mobile version is based on a desktop version with multimodal and adaptive features. The development of the mobile version tried to retain the look and feel of the desktop version and as much as possible of the features required for an advanced Digital Talking Book player. We describe how the intrinsic characteristics of mobile devices impacted the performance and interaction aspects of the application.

Keywords: Mobile devices, Digital Talking Books, Speech output, Evaluation.

1 Introduction

This paper presents the development of a Rich Book Player for mobile devices. The mobile player is the follow-up to a PC-based version of the Rich Book Player [1]. That version, in addition to complying with the requirements elicited in the Digital Talking Book standard, supports multimedia outputs (presentation of text, audio and images) and adaptation capabilities. Those features combine to offer a personal platform for Rich Digital Books reading, aiming to promote active reading of digital books, by making available a player that departs from the features that make e-books attractive and extends them: simultaneous presentation of audio and text, powerful searching capabilities, compliance with accessibility recommendations for Digital Talking Books, and the possibilities offered by adaptive interfaces. These combine to produce a player capable of changing its presentation and behavior in accordance to usage conditions.

The mobile version of the Rich Book Player, aims to promote anytime, anywhere reading, while trying to retain as most as possible of the features present in the PC-based version. It was also a goal to have a similar look and feel in both applications, and to keep a high degree of coherence between applications, to try to make the transition from one to the other as easy as possible.

Nevertheless, the porting of the PC-based version to a mobile platform introduces great challenges on both technical and interaction design levels. The intrinsic limitations of mobile devices on performance, storage, and interaction aspects enforce restrictions result in changes to the interface, and the redesign of available tasks,

which have to be dealt with in combination with the specific characteristics of interaction design for mobile devices [2].

Even though mobile platform resources are limited, it was a requirement to maintain the capability to present the book in visual and audio formats simultaneously. This is the main feature distinguishing the Rich Book Player from an e-book player, and is also responsible for introducing one the concept of page, which was absent from the desktop version of the Rich Book Player.

The next section introduces in more detail the concept of Digital Talking Book and briefly describes the desktop version of the application. The following section describes in detail the several aspects of the mobile version of the application, drawing attention to the features that had to be redesigned or introduced in comparison to the desktop version. The subsequent section presents the results of a preliminary evaluation tests conducted with the mobile version of the player. The final section presents conclusions and future work.

2 Digital Talking Books

Digital Talking Books (DTBs) are the digital counterpart of talking books, which have been available for many years to print-disabled readers. Talking books have been offered on analogue media, such as audiocassettes, providing human speech recordings of a wide array of print material. The DTB goes beyond the limits imposed on analogue books, because it can include not just the audio recording of the work, but the 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 [3]. 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 [4].

DTB developments over the last years lead to the appearance of several different specifications, with the Daisy Consortium1 being responsible for the major work done in the area, and the publication of several standards. In 2005, the Daisy Consortium selected by the National Information Standards Organization (NISO) published the current DAISY/NISO standard [5]. 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 Feature List [6], 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.

According to the Playback Device Feature List mentioned earlier, 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.

With this project we aim to develop a player standing in between the second and third types of playback devices: a player that is portable, and, at the same time, able to retain a comprehensive set of features of the most advanced type of players.

2.1 The Desktop Rich Book Player

By combining the possibilities offered by multimodal interaction and interface adaptability we have already developed the Rich Book Player, an adaptive multimodal Digital Talking Book player [1] for desktop PCs. This player can present book content visually and audibly, in an independent or synchronized fashion. The audio presentation can be based on previously recorded narrations or on synthesized speech. The player also supports user annotations, and the presentation of accompanying media, like other sounds and images. In addition to keyboard and mouse inputs, speech recognition is also supported. Due to the adaptive nature of the player, the use of each modality can be enabled or disabled during the reading experience.

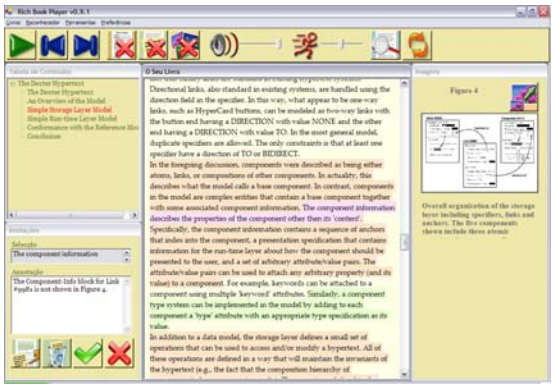

Fig. 1. The Rich Book Player's interface. The center window presents the book's main content. On the top left is the table of contents. On the bottom left is the annotations panel. On the right is the figures panel. The content being presented in the player is the article "The Dexter Hypermedia Reference Model", by Halasz and Schwartz.

Figure 1 shows the visual interface of Rich Book Player. All the main presentation components are visible in the figure: the book's main content, the table of contents, the figures panel and the annotations panel. Their arrangement (size and position) can be changed by the reader, or as a result of the player's adaptation. The other visual component, not present in figure 1, is the search panel. Highlights are used in the main content to indicate the presence of annotated text and of text referencing images. The table of contents, figures and the annotations panels can be shown or hidden. This decision can be taken by the user and by the system, with the system behavior adapting to the user behavior through its adaptation mechanisms. Whenever there is a figure or an annotation to present and the corresponding panel is hidden, the system

may choose to present it immediately or may choose to warn the user to its presence. The warnings are done in both visual and audio modalities.

All the visual interaction components have a corresponding audio interaction element, with one exception. Since the speech recognizer currently used in the player does not support free speech recognition, annotations have to be entered by means of a keyboard. All the other commands can be given using either the visual elements or vocal commands.

3 The Mobile Rich Book Player

The mobile version of the Rich Book Player was developed with three main goals in mind: 1) Allow for an anytime, anywhere entertaining and pleasant reading experience; 2) Retain as most as possible of the features available in the desktop version; 3) Support a similar look and feel and foster coherence between both applications.

To achieve these goals, architectural and interaction changes had to be made. The two major limitations of the mobile platform are the limited screen size and processing power.

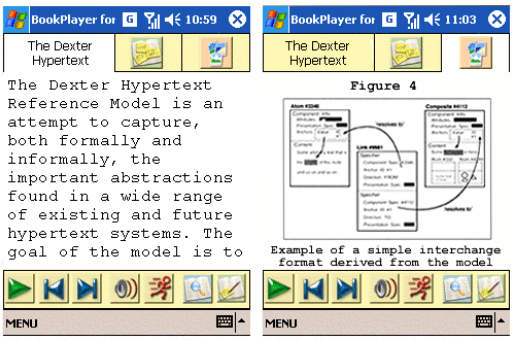

Fig. 2. The mobile Rich Book Player. Main content view on the left and images view on the right.

3.1 Main Components Display

Figure 1 displays the main components of the Rich Book Player: main content, table of contents, annotations and images windows. On the desktop version it is possible to display all the components simultaneously and users can find the arrangement that best suits them. Due to the much smaller screen size of the mobile device it is impossible to follow the same approach.

Figure 2 presents the main content view of the mobile version of the Rich Book Player. The four main areas of interaction can be seen in the figure. On the top, three tabs allow for the selection of the current view. The left tab opens the content view

(figure 2 left). The tab is used to display the current chapter number, which, in this way, is quickly available to the user. The middle tab displays the annotations view. This is the view used to read previously entered annotations and write new ones. The right tab is the images view (figure 2 right). In this tab users can see the images that are part of the book, together with their title and caption. All these contents, book text, annotations and images are displayed in the biggest of the interaction areas. Bellow this area is a toolbar which displays the commands to control book playback, navigation, and other features. The final interaction area is the menu bar located at the bottom of the screen. Besides being used to present the menu, the menu bar is also used to display command buttons whenever necessary.

Of the four main components of the desktop version, three of them were already mentioned and although they cannot be displayed simultaneously as in the desktop version, they all can be displayed on their own view. The component not yet mentioned is the table of contents. This component has been downgraded to a menu entry and retained just one of its functions. In the desktop version, the table of contents was used to display the current chapter being read, by highlighting its entry, and as a navigation mechanism, allowing users to jump to a particular chapter by selecting its entry. In the mobile version, only the navigation function was retained. The current chapter feedback is now provided has the header of the main content tab.

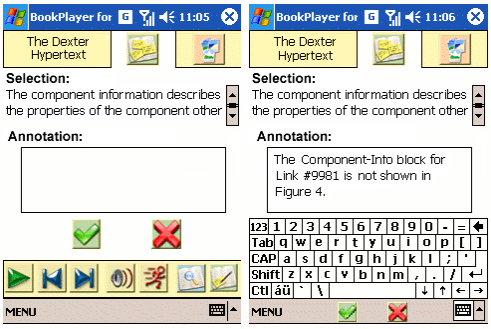

Fig. 3. Annotations view during an annotation creation process. On the right, buttons are reallocated to the menu bar when using the virtual keyboard.

3.2 Annotation Creation and Display

Creating annotations is a two stage process. The user must first select the text to be annotated and only after that enter the annotation. The text selection process is done by selecting the text in the main content view with the stylus and then pressing the create annotation button (the rightmost button in the toolbar). This takes the user to the annotations view in create annotation mode (figure 3, left).

In the annotations view, the user is still able to see the selected text while entering the annotation. When the text box is selected, the virtual keyboard is displayed, and the *confirm* and *cancel* buttons are reallocated to the menu bar (figure 3, right).

Figure 4 presents the annotations view in annotation display mode. When the user changes to the annotations tab, the annotations menu (figure 4, left) displays all the existing annotations in a tree view. By selecting one of the annotations the user is taken to the annotations detail view (figure 4, right). In this view the user can read the current annotation, edit the annotation (which means going to annotation creation mode), delete the annotation, and navigate to the text that has been annotated. The navigation buttons in the toolbar, which in the content view navigate to the next and previous pages, in this view navigate to the next and previous annotations.

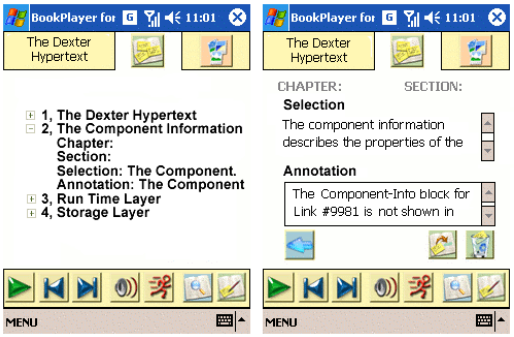

Fig. 4. Annotations view in annotation creation mode. Annotations menu on the left, and details of an annotation on the right.

3.3 Speech Recordings and Synchronization

The main feature distinguishing a Digital Book Player from an e-book player is the possibility to present the book's content using speech, either recorded or synthesized. The desktop version of the Rich Book Player supports both modes of speech presentation. The mobile version currently supports the presentation of recorded speech, using the Windows Media Player for Pocket PC for playback.

Speech presentation opens up interaction possibilities that are not available with a visual only interface. With speech, users can change tabs and view images or read annotations while listening to the narration, thus avoiding the forced pause in reading if speech had not been available. Speech also allows users to access the book content without having to look at the device, allowing usage scenarios that, up until now, were available only with portable music devices, but without the limitations of those. For example, performing a search in such a device is extremely cumbersome. In comparison, with the mobile Rich Book Player, search is extremely simple due to the presence of a digital version of the book.

With the benefits of incorporating speech into the interface new challenges are uncovered. With speech comes the need for synchronization mechanisms. The application needs these to be able to know when and what images and annotations to present. We were able to port the synchronization mechanism of the desktop version to the mobile version without loosing synchronization granularity, meaning the mobile version also supports word synchronization. The synchronization mechanism

only had to be adapted to the page concept, introduced in the mobile version, which was absent from the desktop version. The synchronization mechanism allows the application to turn to the next page when the narration reaches the end of the current one. It is also used to visually highlight the word currently being spoken, in order to make the narration easier to follow when the user chooses to both read and listen to the book.

3.4 Awareness Raising Mechanisms

Images and annotations require a notification system to alert the user to their presence. Users just listening to the narration need to be alerted through sound signals, while users reading the text need to be alerted through visual signals. Both mechanisms coexist also in the mobile version of the Rich Book Player. Auditory icons [7] are played when the narration reaches a page with annotations or associated images. Visually, when the user reaches such a page, the annotations or images tabs flash to indicate the presence of an annotation or image.

3.5 Pagination

One of the main presentation and interaction differences between the desktop and mobile versions of the Rich Book Player is the introduction of the page concept in the mobile version. The main motivation behind this decision was the desire to avoid the use of scroll bars to read the book. With moderate to large books even small scroll bar movements might give origin to large displacements of the text being displayed, which would quickly turn into a usability problem. Loading a text control with the books full content could also raise performance issues.

To support changing font and font sizes while reading the book, the pagination runs in real-time. Whenever a book is loaded, or the font settings are altered, a new pagination is started. This implies storing the current reading point (when speech playback is active) or current page (when speech playback is inactive), repaginate, and present the page of the current reading point. Several choices are possible for the pagination starting point. Starting the pagination algorithm from the book's first page would mean the user would have to wait a variable period for the display to refresh. This period would vary depending on the current reading point. Reading points near the end of book would mean substantially longer waiting periods, due to the lack of processing power of most mobile devices. Starting the pagination from the current reading point, would mean pages with different contents would result from runs of the algorithm with different starting points, even with the same font settings. This might confuse users used to pages holding the same contents on print books. Both approaches raise usability issues. To overcome these issues we employed the notion of forced page break. A forced page break is a location in the text that is guaranteed to be at the start of page. Employing this notion, the pagination algorithm always produces the same results for the same font settings. The pagination starts from the first forced page break prior to the current reading point and runs to the first forced page break after the current reading point. Since we can control the frequency of

forced page breaks, we can guarantee an upper limit on the time that is necessary to paginate until the current reading point, thus assuring the user will not have to wait unacceptably long periods and the pages stay coherent with every run. After this stretch of the book is paginated, the algorithm can run in the background, paginating the rest of the book. Possible choices of forced page breaks, which will be adequate in most situations, are all the entries in the table of contents. This has the added benefit of page breaks being associated with structural book elements, which is something a reader would expect, or, at least, not find confusing.

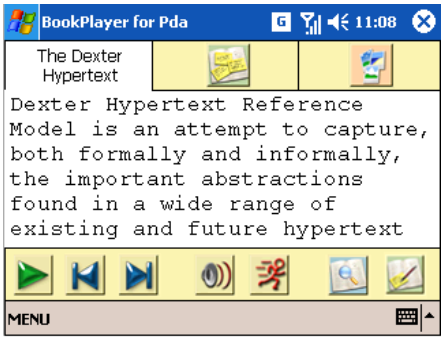

Fig. 5. The Rich Book Player layout is landscape mode.

3.6 Adapting the Layout to the Device

Nowadays, mobile devices exist with a multitude of screen resolutions, and even with the possibility of altering the screen orientation. The Rich Book Player visual layout is able to adapt itself to changes in orientation and to different screen resolutions. Figure 5 presents the layout of the Rich Book Player in landscape mode.

Besides changing the layout of the different interface components, a changing in screen orientation also requires a new run of the pagination algorithm, as described in the previous section.

3.7 Performance and Storage

The book's presentation involves parallel processing of three threads: the main interaction thread with audio playback, the synchronization thread and the pagination thread. This might impose some performance constraints on the reproduction platform. We have tested the Rich Book Player on two devices: an HP iPAQ h5500 with a 400 MHz XScale processor, 40 MB ROM and 128 MB RAM running Microsoft's Pocket PC 2003, and a QTEK 9100, with a 200 MHz TI OMAP 850 processor, 128 MB ROM and 64 MB RAM running Microsoft's Windows Mobile 5. We have successfully been able to use the application on both devices to read a book of circa 25000 words (a 279 Kb text file), corresponding to a narration with 2 hours and 15 minutes, recorded in a 158 Mb mp3 file.

Due to the size of the audio files, books will have to be made available in storage cards, since it cannot be expected that internal storage of the mobile devices be able to hold such large amounts of data. The Digital Talking Books can be shared between the two platforms, as well as annotations that have been made in one platform can also be read in the other.

4 Preliminary Evaluation

We have conducted some preliminary evaluation on the mobile player. That assessment was carried out in the sequence of a deeper evaluation of the PC version, reported elsewhere [8]. In that experiment, twenty four students used the Rich Book Player in a laptop, with a mouse, headphones, microphone and a fixed large monitor. They had to read/listen-to an article ("The Dexter Hypermedia Reference Model"), prepare a summary and answer a set of questions, as part of a Hypermedia Systems' course evaluation. To attain these objectives, students browsed through the article, highlighting specific parts of the text, reread/listen-to some of those parts, took notes, and later searched for answers to the posed questions.

Seven of those students, randomly selected, a month later were requested to perform a reduced version of the first experiment, now using the Mobile Rich Book Player. They were handed a PDA (the aforementioned HP iPAQ h5500), with the application and the article. They were asked to reread it, listen to some parts, take notes on the parts of the text they considered important, and look at the images. During the half-hour of the trial, they were requested to stand up for about 20 minutes. During 10 minutes they could use the stylus and for the other 10 minutes they could use only one hand to hold and operate the device. A researcher accompanied the process, observing, annotation behaviors, and posing opportune pertinent questions (e.g. "what are you doing?", "do you remember what you have done on the PC version?") as part of a contextual inquiring approach. This trial was no longer a part of the course evaluation. At the end the students were encouraged to criticize the mobile application.

Overall, we had a very positive feedback and no major complaints were made relatively to the PC version. Nevertheless, some issues have been identified that require some changes to the interface of the mobile version. The first, concerns the visual appearance of the tabs. The symbols used in the annotations and images tabs resemble a button, thus initially misleading users. Other issue that has been identified becomes relevant in the usage scenario where the user had only one hand available. When operating the device with just one hand, and using the fingers instead of the stylus, reaching for the tabs covers the display, thus causing a slight annoyance. Another related issue concerns the placement of the buttons in the toolbar. The most frequently used buttons, from our experience with the application, are the page forward and backward buttons (when the user reads the book without listening to the narration). These buttons are placed on the left side of the toolbar, which makes one-hand operation more awkward for right-handed people, particularly in landscape mode. A version of the toolbar with the navigation and playback control buttons on the right side of the toolbar will solve this issue.

5 Conclusions

This paper presented the mobile version of Rich Book Player, an application for reading Digital Talking Books in mobile devices. With the Rich Book Player it is possible to read, listen to a book, or have both modalities working in a synchronized fashion.

The introduction of books with speech narration in mobile devices represents a step forward from the traditional text based e-book reader. Besides the narration, the Rich Book Player also supports the presentation of images within the book, and the creation and displaying of annotations. All the different contents are synchronized between themselves, allowing the player to alert the reader to the presence of an annotation or image. This is achieved in both reading scenarios, with or without narration.

The work on the mobile Rich Book Player will carry on with a detailed usability evaluation of the application. More books, with different focus, will be used for evaluation purposes. More devices and more usage scenarios will allow us to have a more precise idea of the performance and the usability problems of the player.

References

1. Duarte, C., Carriço, L.: A conceptual framework for developing adaptive multimodal applications. Proceedings of the 11th International Conference on Intelligent User Interfaces, Sydney, Australia. ACM Press, New York (2006) 132–139
2. Jones, M., Marsden, G.: Mobile Interaction Design. John Wiley and Sons Ltd (2005)
3. Moreno, R., Mayer, R.E.: A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. Journal of Educational Psychology 97 (2000) 117–125
4. Gazzaniga, M.S., Ivry, R.B., Mangun, G.R.: Cognitive Neuroscience - the Biology of the Mind. W. W. Norton & Company (1998)
5. National Information Standards Organization: Specifications for the Digital Talking Book. NISO Press (2005)
6. National Information Standards Organization: Playback Device Guidelines. Available at http://www.loc.gov/nls/niso/features.htm (1999)
7. Gaver, W.W.: Auditory Icons: Using Sound in Computer Interfaces. Human-Computer Interaction, Vol. 2, N. 2. Lawrence Erlbaum Associates, Inc., Mahwah, NJ (1986) 167–177
8. Duarte, C., Carriço, L., Guimarães, N.: Evaluating Usability Improvements by Combining Visual and Audio Modalities in the Interface. Proceedings of the 12th International Conference on Human-Computer Interaction, Beijing, P.R. China, Springer-Verlag, Berlin Heidelberg New York (2007)