

NEW WAYS TO READ DIGITAL TALKING BOOKS

Miguel Rodrigues, Luís Carriço, Carlos Duarte, Nuno Guimarães
LasiGE, Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa
Campo Grande, Edifício C6, 3º Piso, 1749-016 Lisboa Portugal
{mrod, lmc, cad, nmg}@di.fc.ul.pt

António Serralheiro, Isabel Trancoso
L²F Spoken Language Systems Lab., INESC-ID/IST
Rua Alves Redol 9, 1000-029 Lisbon, Portugal
{antonio.serralheiro, isabel.trancoso}@inesc-id.pt

ABSTRACT

In this paper we present new ways to read Digital Talking Books (DTBs). DTBs aim to provide better access to literary contents for the visually impaired population. However, these books offer limited capabilities when reading the book freely. Only continuous reading or jumping through sections of the text is allowed. This lack of flexibility leaves out many of the types of reading users usually do, such as those that provide the user with an overview of the book content or readings based on a specific set of key points that have more importance to the user. The work we present here overcomes these limitations by providing the user with a whole new range of interaction possibilities with DTBs.

KEYWORDS

Digital Talking Books, Multimedia Readers, Media integration, Reading.

1. INTRODUCTION

With the increasing need to make written knowledge available to every person, new efforts have emerged to bring together new technologies and the traditional ways of content presentation. In particular, Digital Talking Books (DTBs) target a print-disabled or visually impaired community, using visual reproduction of the book together with spoken reproduction (ANSI/NISO, 2002). This combination of different media manages to overcome some of the cognitive limitations of human perception (Gazzaniga et al., 1998). On the other hand, the digital nature of DTBs, not only allows to overcome the limitations of analogue audio books, but also enables the exploration of different and improved ways of interaction, beyond those permitted by printed material. Cross-referencing and word searching becomes easier, as well as mark and annotate. But, for instance, multimedia and multimodal selective and structured marking and annotations can be easily envisaged, without destroying the medium, or even conceived as summaries or new reading patterns.

The combination of visual and audio components and the weight each has in the interaction greatly depends on each person using the DTB. As such, the possibility to personalize the interface by changing font size, color, disposition of elements on screen or even changing the audio reproduction speed becomes increasingly important. This last feature assumes crucial importance specially when reading technical or reference books. Most of these aren't usually read in a continuous way and some parts are read several times over. For this type of reading, special features such as different reading speeds, jumping over unimportant sections or marking portions of the text for future reference, assume bigger relevance and deserve an attention that they haven't yet received from current DTB players.

This paper presents RiPlay (Rich Digital Talking Book Player). It is a DTB player that fulfils all the requirements for DTB reading, emphasizing the usability and accessibility issues. The interaction and presentation alternatives it offers enable its usage in pure audio mode (speech reproduction and interaction), still maintaining the navigation and annotation capabilities. Beyond DTB playing, RiPlay offers a couple of mechanisms supporting the study of documental material (e.g. technical and scientific articles and books): fast reading/listening overview, through a set of simple text selection criteria, highly customizable;

mark/annotate text excerpts, with different importance degrees, that later, on a document revision, can be read/listened to selectively, using different reading speeds or skipping uninteresting parts. In this sense, RiPlay provides the ability to listen to different perspectives of a text.

2. RELATED WORK

Digital Talking Books (DTBs) are the digital counterpart of analogue talking books, overcoming several limitations of the analogue media, particularly its linear presentation nature. Digital media, on the other hand, offers readers the possibility to move around in a book as freely as in a printed book, place bookmarks or highlight materials. These features naturally go beyond the limits imposed by printed books (e.g. speed of search), and the inclusion of the audio narration provides reading approaches not possible in paper versions.

DTB developments over the last years led to the appearance of several different specifications, with the Daisy Consortium being responsible for the publication of several standards. Finally, in 2002, with cooperation from the Daisy Consortium, the National Information Standards Organization (NISO) published the current standard (ANSI/NISO, 2002). The standard focuses on the structure and content of the files needed to produce DTBs. An auxiliary document, the Playback Device Feature List (NISO, 1999), describes the main features that playback devices should possess, including: 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; reading of notes; setting and labeling bookmarks; automatic bookmark at stop; ability to add information (highlighting and notes); fast forward and fast reverse; availability of human and electronic speech; presentation of visual elements in alternative formats (speech).

The Daisy Consortium publishes and maintains a list of DTB players that adopt different solutions for the presentation and interaction. According to a heuristic evaluation presented in (Duarte & Carriço, 2005), most of them adhere to the navigation related recommendations, with some support for bookmarks and annotations. Nevertheless, several usability and accessibility flaws are reported that hinder the reading experience: lack of mechanisms to make the reader aware of the presence of annotations in a given part of the book; narration's interruption caused by the examination of auxiliary structures, like the table of contents; font sizes that render the text in such a way that it exceeds the available space in a non-scrollable window; impossibility to customize the presentation (size and color) of some elements; and an overall impossibility to use the players in a non-visual environment for all but the simpler operations, due mainly to the absence of speech recognition and the consequent over-reliance on keyboard shortcuts. The possibilities opened by annotating and marking and its integration with the other playing features are left unexplored by all the players. Different playback speeds, for instance, are available in some players, but only as a global setting thus with small usage impact (read the whole book slower/faster). Furthermore, if reading of technical or reference material is considered, even more limitations become visible. This kind of material is usually not read in a linear fashion, with the user typically skipping over the segments considered less important. Fast and selective reading (Fairbairn & Winch, 1996; Leeuw & Leeuw, 1990) allow users to decide about the text interest before engaging in a deeper study. None of the players provide support for these reading patterns.

3. RIPLAY

RiPlay provides to DTB readers/listeners most of the functionalities proposed in the Playback Device Feature List document (NISO, 1999). Some of those features are, for example: synchronization between text and audio, highlighting the text unit (word, sentence, ...) being heard at that moment; variable reading speed; visualization and insertion of textual/voice annotations; table of content navigation; text unit jumping, through search and direct choice (e.g. clicking); or even personalizing the user interface by changing font sizes and menu positions. Special care was taken in usability and accessibility issues, stressing, for instance, the coherence of size in the presentation font of DTB content, indexes and interaction elements; the duration and articulation of audio warnings with narration; and in general the feasibility of using the player functionality in an audio only (or centered) perspective.

In this audio perspective, RiPlay provides the ability to work with speech interaction and introduces audio context-spaces that enable the listener to acquire the notion of context when navigating through the book. For example, when a word search is issued, the reader in a visual perspective often reads a set of words (context) around the marked word. In an audio standpoint that is impossible unless the player provides that audio context. In this case RiPlay reproduces the words from the beginning of the sentence or paragraph, and indicates the section or chapter as needed (if changed). The player is also highly customizable, enabling, for example, the definition of size of the highlighted text unit and the way annotations and footnotes are notified. When added, annotations are associated with the text being reproduced at that moment (narration usually stops, depending on the users' customization). When the player reads that text at a later time, the user is notified that there is an associated annotation. Again, content narration may continue, stop or pause temporarily. Notification can be done audibly (a sound or through synthesized speech), visually (e.g. highlighting the text) or both. The annotation itself can be opened and reproduced or maintained closed. All these options are configurable.

Also beyond the guidelines of DTB and the current implementation of DTB players, a couple of features, described next, were introduced to address new ways to "read" DTBs

3.1 Overview Reading/Listening

Overview reading pertains to the ability to make fast readings of whole texts by picking only specific excerpts according to a selection criterion. Those criteria range from purely syntactic ones (e.g reading the beginning of the first sentence of each section), to more semantic ones, where the reader picks keywords, scans the text for them, and reads the surrounding text.

RiPlay provides a choice over several selection criteria that enable different book reproduction. In the simplest one, only the beginning of each paragraph is reproduced. The player plays a few words and then jumps to the beginning of the next paragraph. All parameters can be customized, specifying how many basic units (e.g. words or sentences) will be played in the beginning of each larger unit (e.g. sentence, paragraph, section or chapter). Also, the user can select specific sections (or chapters) to be reproduced, for example, abstracts, introductions or conclusions. Finally, instead of jumping (not playing) the text between the selected excerpts, the user may also opt to play it faster, in order to still have an overall idea of the text.

Other reproduction criteria capitalize on the assumption that the relevant text is around the relevant words. As a basic, yet effective approach (Gernsbacher, 1990), RiPlay provides an automatic reproducing mode where the played words are in the neighborhood of the most used ones: the number of occurrences of significant words is computed; the list of those above a specified frequency is selected; and the sentences where those words occur are reproduced. The frequency threshold can be specified. Still on the same neighborhood rationale the user has the possibility to define the keywords he wants to search (a relevant words' list). In both cases, the neighborhood size is customizable and, again, the user may decide that only those words are played or that they are narrated at a smoother pace, whereas the rest of the text is read faster.

3.2 Re-reading/listening

Re-reading pertains to the revisiting of a document in order to remember or get a deeper comprehension of specific excerpts considered important on previous readings. The reader usually underlines relevant passages of the text, sometimes, using different colors or marking styles, in order to convey different relevance levels or categories. In a later re-reading, the reader attention is usually focused on those passages.

The RiPlay tool also provides mechanisms for re-reading, again capitalizing on the ability to narrate the content. The user is able to mark excerpts of the text, indicating, currently, three types of excerpts. From a menu the user can choose: **Highlight personal markings**, to mark the text as relevant; **Highlight for fast speed reading**, meaning lesser relevancy; and **Highlight for slow speed reading**, indicating strong importance. Text selection is done as in common text editors/viewers in a visual interaction approach or through voice commands when a speech-based strand is in use. In the latter case the user issues the "begin and end selection" voice commands while listening to the text. As the commands are usually given after the actual words are played, RiPlay marks the text a couple of words before (the number of words is adjustable).

When the text selection is performed, RiPlay visually highlights the text using different colors and different patterns that depend on the relevancy choice that was set (see figure 1). Naturally, it also associates

different reading speeds. Color, patterns and reading speeds are customizable. Once the "highlights" are defined, if the whole document is played, then only the marked excerpts are read. Personal ones are read at the normal pace, whereas the others are read at the specified slow and faster speeds.

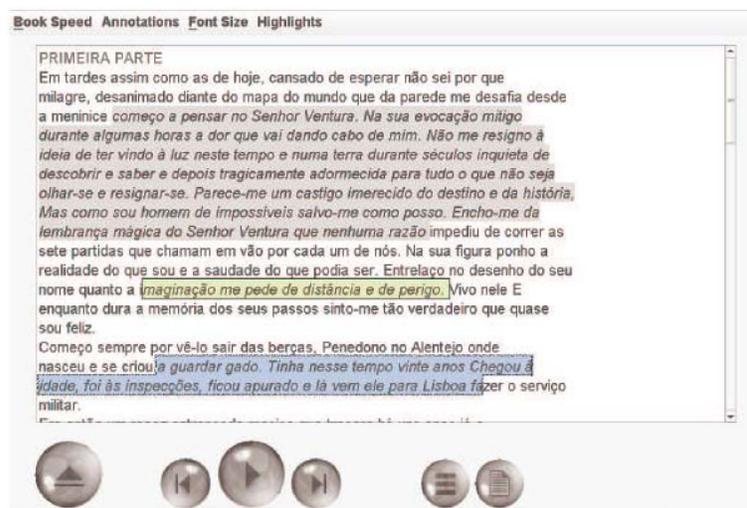


Figure 1 - Highlighting for different speed readings in RiPlay.

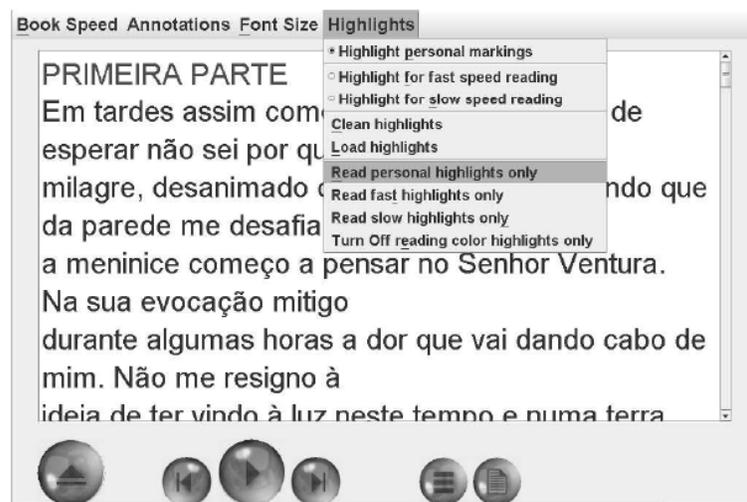


Figure 2 - Types of reading in RiPlay.

Apart from the whole text reading, the user can, for example, read only the text he marked for each of the relevancy choices previously done. Reading options can be changed from the menu shown in figure 2.

3.3 Topic Re-reading/listening

RiPlay further permits the combination of the above functionalities in order to provide relevant reading/listening over specific topics of document. A simple one, for example, combines the list of relevant words criteria and its neighborhoods, with the relevancy choices. This way when the document is played, only the neighborhoods of the words that lay inside selected relevant text are narrated. Moreover they are narrated at the speed defined by the particular selections.

On the other hand, since relevancy selections can be saved and loaded from the player (see figure 2), the user can associate "highlighting" sets to specific document topics or argumentations.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have presented RiPlay. It is a DTB player that fully addresses the DTB standard and recommended requirements, providing usable and coherent interaction approaches, for different usage settings, that are not available in existing DTB players. Riplay also goes beyond DTB requirements, in some of the "standard" features, but most of all, addresses reading patterns integrating visual and audio counterparts, thus providing new ways to explore DTBs. Automatic and keyword-based overview readings, user selected excerpts, and the combination of both, integrated with annotation, audio reproduction and variable reproduction speed are features that provide richer possibilities to study and review documents.

As future work we are considering other approaches to find the most important sentences in the book, combining syntactic and semantic analysis of the text. User's annotations can be further developed, with the corresponding enhancements over the control of re-reading/listening. One possibility is allowing the reader to define an annotation hierarchy, and use it to categorize the annotations. By associating playback settings with each category, the re-reading or listening of the book would be further tailored to the user's interests. This will be extended to the use of a metadata language to classify annotations and control future book's reproductions.

To evaluate all the features implemented in the RiPlay player, we are performing a set of usability tests, focusing not only the interface and interaction options available, but also the efficiency of the approach for studying and reviewing documents.

Research is also being made on the interaction the player allows as we are trying to make this player available to the greatest number of people possible. Future solutions might include for example the connection between the RiPlay player and a Braille interface machine, allowing other output options.

ACKNOWLEDGEMENT

This work is being carried within the context of project RiCoBA (Rich Content Books for All) POSI/EIA/61042/2004, funded by Fundação para a Ciência e Tecnologia.

REFERENCES

- ANSI/NISO, 2002. *Specifications for the digital talking book*. <http://www.niso.org/standards/resources/Z39-86-2002.html>,
- Carriço, L., Duarte, C., Guimarães, N., Serralheiro, A. and Trancoso, I., 2004. Modular production of flexible digital talking books. In Proceedings of the 6th International Conference on Enterprise Information Systems, ICEIS 2004, Porto, Portugal.
- Duarte, C. and Carriço, L., 2005. *Users and usage driven adaptation of digital talking books*. In Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas.
- Fairbairn, G. J. and Winch, C., 1996. *Reading, Writing and Reasoning - A guide for Students*. Open University Press.
- Gazzaniga, M. S., Ivry, R. B. and Mangun, G. R. 1998. *Cognitive Neuroscience - the Biology of the Mind*. W. W. Norton & Company.
- Gernsbacher, M. A. 1990. *Language Comprehension as Structure Building*. Lawrence Erlbaum, Hillsdale NJ.
- Leeuw E. D. and Leeuw, M. D, 1990. *Read Better, Read Faster*. Penguin Books.
- NISO, 1999. *Digital talking book player features list*. <http://www.loc.gov/nls/niso/features.htm>.