A FLEXIBLE INTERFACE ARCHITECTURE FOR DIGITAL TALKING BOOKS

Carlos Duarte, Luís Carriço, Hugo Simões
LaSIGE, Department of Informatics, Faculty of Sciences of the University of Lisbon
Campo Grande, Edifício C6, 1749-016 Lisbon, Portugal

Keywords: Adaptive Interfaces, Multimodal Interaction, Model-based Architecture, Digital Talking Books.

Abstract: Talking books, besides helping the blind and print-disabled communities to have easier access to books, also allow reading in situations when the vision becomes temporarily unavailable, for instance, while driving. In this paper, we present a user interface model-based architecture for digital talking books. The books allow for multimodal interaction in an effort to broaden the usage scope. Considering the execution platform, user and usage characteristics, a great number of configurations for the books are possible. For the situations where the most suited configuration can not be forecast, an adaptable book can be the solution.

1 INTRODUCTION

Bringing together written and spoken words opens up new ways of exploring books. A Digital Talking Book (DTB) accomplishes it by combining the visual reproduction of the book with the audio reproduction of the narration. In comparison to audio tapes with analogue recordings of the books, the digital format brings more than an improvement in the reproduction quality. New interaction possibilities, that before were unavailable or cumbersome, become reality. Searching for a word no longer forces the user to listen to (possibly) the whole book. Annotating and navigating the book is also now made possible.

Besides the visually impaired persons, who are the main target audience for DTBs, several other population segments can benefit from a platform such as this. The multimodal interaction capabilities of the book broaden the reading opportunities to situations where the reader is engaged in visual occupying activities, such as driving or surveillance. Balancing DTB modes and media can be explored to overcome the cognitive limitations of human perception and attention (Gazzaniga et al., 1998).

The expansion of the traditional reading experience, reshaping it into an "immersion" in a multimedia environment, can be envisioned within a multimodal playback environment. The possibility of creative combination of presentation elements, taking advantage of available media resources, offers the support to new ways of telling stories and improving learning (Carriço et al., 2003). The evolution is based on the introduction of new multimedia elements, in a coherent way, during "reading". Possible enrichments are the introduction of background music, environmental sounds related to the "scene of the action", images or videos to complement information presented in the original source, and many more.

The availability of multimodal interaction on the playback platform is desirable to complement this multimedia presentation. In order to improve the coverage, reliability and usability of the interface, it is important to automatically adapt to user, task and environment parameters (Oviatt et al., 2004). A book's presentation and interaction should adapt to the users' characteristics and preferences, to the environmental conditions, and to the playback platform and interaction devices available.

A framework for automatic production of DTBs, that is flexible enough to build books tailored to specific users and usage situations, has already been developed (Carriço et al., 2004). Nevertheless, in situations where the book designer is unable to predict the characteristics of the users, the playback platform and environment, the book must be able to adapt automatically to those changing conditions. That is why we need an adaptive DTB (Duarte and Carriço, 2004).

In the next section we present a brief overview of related work, covering non-visual interfaces, adaptive interfaces and DTB standards and navigation features.

Duarte C., Carriço L. and Simões H. (2004).
A FLEXIBLE INTERFACE ARCHITECTURE FOR DIGITAL TALKING BOOKS.
In *Proceedings of the Sixth International Conference on Enterprise Information Systems*, pages 146-151
DOI: 10.5220/0002636101460151

We then identify some adaptation dimensions specific to DTBs. We follow by describing the model-based architecture of our user interface development framework, relate it to the run-time execution of the interface, and how it can be used to adapt to platform, user and environment changes. We finish by drawing some conclusions.

2 RELATED WORK

2.1 Digital Talking Books

DTBs are intended to give an easier access to books to the blind and print-disabled community. Members of those communities cooperated with several organizations that developed DTB related standards. In Europe, the Daisy Consortium, with collaboration from the European Blind Union developed one of those standards. In the USA a similar work was conducted by the National Information Standards Organization (NISO) in collaboration with The National Library Service for the Blind and Physically Handicapped. From the cooperation between the Daisy Consortium and the NISO resulted the most important DTB specification, the ANSI/NISO z39.86 (ANSI/NISO, 2002).

According to the NISO Document Navigation Features List (NISO, 1999), a DTB should provide basic navigation capabilities (advancing one character, word, line, sentence, paragraph or page at a time, and navigation to specific segments of the DTB), fast forward and reverse, reading at variable speeds, navigation through table of contents or a navigation control file (allowing the user to obtain an overview of the material in the book), reading notes, cross-reference access, bookmarking, searching and others.

However, and wisely, no specific implementation solutions are present in the standard. The solutions must consider aspects related to the proposed specification, but also the non-visual nature of the targeted environment.

2.2 Speech Interfaces

The work on non-visual interfaces can provide us with clues on how to tackle some of the problems faced. Voice browsers are devices that exhibit at least one of the following characteristics: (1) can render web pages in audio format; (2) can interpret speech for navigation. Voice browsers and DTB interfaces share some common problems:

- The audio format is a temporal medium. A visually presented page can render simultaneously images, tables and text, in a spatial format, which is quickly processed by the perceptual human system. Spoken text, however, can present only one word at a time.
- Issuing voice commands, and audio processing, are activities that consume working and short-term memory, conflicting with planning and problem solving tasks. Visual information is processed by separate cognitive systems (Christian et al., 2000).
- The unavoidable recognition errors.

However, the research in the multimodal systems field have made it clear that speech input is advantageous under certain circumstances (Oviatt et al., 2000). Studies (Van Buskirk and LaLomia, 1995; Christian et al., 2000) point out that "the best tasks for speech input were tasks in which the user has to issue brief commands using a small vocabulary".

The interaction characteristics of a DTB are advantageous for the adoption of a speech interface: a relatively small number of commands can be used to implement the needed functionalities. However, some limitations may arise, if, for instance, to follow a table of contents entry, the user is forced to speak the chapter's title.

Research on the efficiency of speech as an input mode is not conclusive, although showing an increase in task completion time (Van Buskirk and LaLomia, 1995; Christian et al., 2000). Some of the recommendations made for constructing voice browsers can be adopted for the design of DTBs: links should be easily spoken text; links should be short (a few words); avoid links with similar sounds; and develop alternatives to numbered links, as these cause cognitive overload.

2.3 Adaptive Interfaces

An adaptive interface has been defined as "a software artefact that improves its ability to interact with a user by constructing a user model based on partial experience with that user" (Langley, 1999). This means that an adaptive interface must have generalization abilities (because the adaptation is based only on partial experience with past user interactions) and that the adaptation is based on a user model. However we may argue that relying only on the user model, as the basis for the adaptation may be insufficient. There are situations where the drive of the adaptation shouldn't originate from the user, but from external events or environmental changes.

Several reviews of user models exist (Kok, 1991; Kobsa, 2001), but we will focus only on two user models here. A number of characteristics could be used to identify users of a certain subgroup, called a stereotype (Rich, 1989). Once the type is known, the interface can be adapted to accommodate the user. This approach can be used when producing a DTB tailored to groups of users. Overlay models (Carr and Goldstein, 1977) can be used to represent the users knowledge (or preferences) on some domain. For

this purpose a set of concept-value pairs is needed. The concepts form elementary pieces of knowledge for the given domain, and the value associated with it represent how well the concept is known to the user. We can envision the use of overlay models whenever producing a DTB that could be tailored to individual users.

Adaptive interface systems cover many areas, with educational applications and on-line information systems being the most popular (Brusilowsky, 2001). We can relate the creation of adaptive DTBs to both of these areas, very clearly when producing a DTB with educational purposes, but also from the information systems area when considering how to adapt based on location and behaviour in physical spaces (Not et al., 1998; Oppermann and Specht, 1999).

3 DIMENSIONS OF ADAPTATION

DTB playback is possible over a broad range of platforms, devices and environments. Also, users with different physical and cognitive characteristics, preferences and knowledge are expected to use DTBs. For this reason an automatic and flexible DTB building platform is a necessity (Carriço et al., 2004). Nevertheless, there are situations where it is not possible to previously identify all the variables governing the book's production. Even when the users and situation share common characteristics, it is impossible to please everyone. DTBs produced by the DiTaBBu platform (Carriço et al., 2004) were evaluated (Duarte et al., 2003b) and some of the subjects' observations are evidence for having a personalized version of the books. Examples are given bellow:

- I would like to have the full text of the annotation shown by default instead of just the subject.
- I would like a sound to signal the start of a new chapter.
- I would like the start of a new chapter to be displayed in the text.
- The annotation frame could be replaced by a link, that when followed would show an annotations window. The space currently occupied by the annotation frame would be used to display more text.
- The annotation frame could be used to show images related to the text.
- I would like to have links to web pages related to the books context.
- I would like for the text to have signals showing where it has been annotated, instead of having just an audio signal.

Each of these observations was made by a different subject, with the exception of the request for a sound signalling the start of a chapter. This is an indication that each person would like a personalized book player.

Considering the blind and print-disabled population, which is the primary target audience for DTBs, we can expect users with very different characteristics from the test participants. These, in addition to not being visually impaired, were familiar with computer interfaces in general. We expect the identification of the aspects to adapt, and the adaptation process, to be a considerably harder task for the target audience, thus being an extra motivation for the development of an adaptive interface for the DTB.

Besides the mentioned "personal preferences", other aspects should be contemplated when considering the creation of an adaptive book. General characteristics include the capabilities of the playback machine (which input/output devices are available), the playback environment (for instance, a noisy background), the users current activities and others that can be found in common adaptive applications. Taking into account all these and other variables specific to DTBs we identified some of the elements of the presentation and interaction that can be adapted, as well as some of the variables responsible for initiating the adaptation procedure. The adaptation initiating variables can be divided into two groups: user related and environment related. Examples of variables are:

User related variables - characteristics, knowledge, preferences, interaction history and current activity.

Environment related variables - interaction devices available, access to media repository, background noise.

The adaptable components of the book may be separated into three dimensions: interaction, content and presentation. Examples of some of the components of each dimension are:

Interaction - The input/output modalities available can be enabled or disabled, and used in cooperation or individually.

Content - Enhancement of the book with the introduction of sound, images, and other available media, translation of the text, hiding or revealing of parts of the book.

Presentation - Size and colour of the font used, rearranging of the elements of the book on screen, type of audio signals used, synchronization units.

Taking into greater account the specific aspects of DTB generation and playback, and of its target audience, we can identify the most important book and user related variables and adaptable components: the visual impairment level of the user (variable) and synchronization units (adaptable component).

We can also examine DTB related aspects of the articulation between some of these variables and components.

- The visual-impairment level has a clear impact on the presentation of the book. From the size of the font used, to not using visual presentation at all for blind users.
- The user preferences, capacities, and past interactions influence the presentation of the book, by setting the level of enhancement used.
- The interaction history can impact the synchronization units. If there is a history of jumping to the same (or near) word in the book, the synchronization units should become gradually finer. In this way, the first jumps would lead to the word's paragraph, then to the word's sentence, and finally to the word itself.
- The output devices used can impact the synchronization units. Absence of a visual display leads to the use of less detailed synchronization units. After searching a word in a sound only output environment, a jump to the searched word would make it more difficult for the reader to understand the "new" context, when compared with a multimodal environment, in which the narration jump can be accompanied with the visual presentation of the surrounding text.

The book characteristics will also play a part in the adaptation process. For example, when processing a novel, it wouldn't make much sense to hide parts of the text, but when processing an educational book it could make sense to hide some of the content already know by the reader.

In order to produce adaptive versions of the DTBs we propose to evolve the DTB production framework. The adaptation dimensions are related to adaptations modules that are responsible for introducing the adaptation capabilities in the generated DTB. The modules and their integration are discussed in the next section.

4 USER INTERFACE ARCHITECTURE

The design of the presentation and interaction aspects of a DTB must consider several issues, including the target user, the execution platform and the reproduction environment. These and other issues impose constraints that the user interface must conform to. If the reader is print-disabled, both presentation and interaction should be done using audio. If the book is to be read in a PDA, then the presentation and the interactors must adapt to the reduced display size. It the book is presented in a noisy environment, then the audio volume must be adjusted, or even discard the audio output in favour of a visual presentation.

When confronted with the diversity of user characteristics, available devices, and all the other issues controlling the presentation and interaction with DTBs we can see the impossibility of having an "one-size fits all" interface. Also, considering the number of possible combinations of all the elements, it would be unreasonable to develop a user interface for each of the possibilities. However, if we are able to establish a set of relations between the interface generation controlling aspects and the usable interactors, then we can create a unique interface whenever the need for it arises. This interface will be tailored to the characteristics of the user and the environment, and will be able to efficiently use the available interaction devices.

A framework with these characteristics has several advantages: flexible creation of user interfaces, allowing an easier interface development process by speeding up the creation-evaluation cycle; development of reusable *presentation and interaction templates* indexed by the characteristics of the platform, the target users and the environment; implementation of multimodal behaviour by integration of templates designed for separate modalities; and adaptation of the interface to changes in the platform, environment or user.

We adopt a model-based approach (Paternò, 2000) to the development of our DTB's user interface. The UI generation architecture is represented in figure 1.

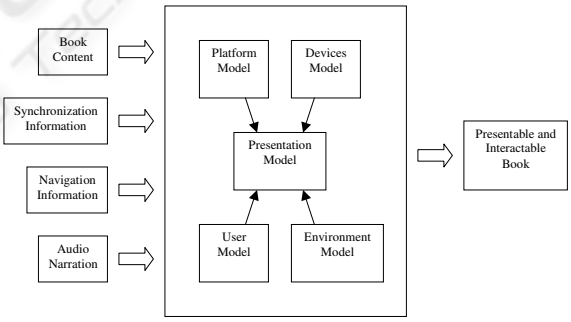

Figure 1: The UI generation architecture.

The interface generation process receives as input the normalized version of the book's content, the book's narration in audio format, the synchronization between text and audio, and the book's navigation information. For details on the creation of these inputs we refer the reader to (Carriço et al., 2004; Duarte et al., 2003a).

Depending on the resources available in the execution platform, the interface can be customized by the user, or even adjust automatically to changes in the environment or in the platform itself (e.g. the connection or disconnection of new interaction devices during the book playback). For this to be possible,

the generation environment needs to be "migrated" to the running environment. A static execution environment is able to present only previously generated books, made for specific combinations of platform and users. In a dynamic execution environment, changes in the platform, user and environment would trigger the adaptation of the interface to its new conditions. When discussing the models used in our architecture we refer to these two situations as *design-time generation* and *run-time generation*. Generation in design-time creates a book tailored for specific execution platforms and user groups. Generation in run-time allows the book to dynamically adapt to changes.

Five models are used to describe the different aspects conditioning the development of the interface:

Platform Model - The platform model describes the various systems capable of presenting a DTB. Examples are a desktop PC or a PDA. The interaction devices that are guaranteed to always accompany the platform are also described in this model. Relevant data that should be described in this model includes screen size and resolution, and access to other media sources (for the possibility of obtaining different materials for the book's presentation, like images or videos).

Devices Model - The devices model describes all the devices that can be added to the execution platform, bringing new modalities into play during the book's presentation. Both hardware and software devices should be included. Examples of hardware devices are a microphone, a braille reader or a braille keyboard. Possible software devices include text to speech converters and voice recognizers. This model can be used in design-time to complement the platform model and create user interfaces for a "fixed platform". However, an increased value can be obtained by using it dynamically in run-time, reacting to changes in the available devices in a way that allows users to exploit the improved interaction possibilities.

User Model - The user model describes relevant user preferences and characteristics. Users can be grouped by their characteristics and abilities, and a design-time user interface generation system creates different interfaces for specific groups. Taking into account the particular domain of application, relevant characteristics for identifying these user groups include the visual ability of the user, the age, etc. A dynamic version of the interface can adapt to perceived changes in the user's characteristics and preferences. Examples include the user's favourite spatial layout and disposition of the on-screen components, or the user's preference for visual or audio interaction.

Environment Model - The environment model describes the environmental characteristics that can have an impact on the book's presentation and interaction aspects. An example is the ambient noise of the reproduction environment.

Presentation Model - The presentation model describes all the components available for presenting and interacting with the book. Examples of components to include in a presentation are the table of contents, the annotations, images, videos or audio clips to enrich the reading experience, the book content itself, and others. Each component has a set of implementation templates available. For each component there exists a template for each platform where it is possible to execute the book's interface, a template for each device that can be added to the execution platform, a template for each identified user group, and a template for each execution environment. From the information in each of the models, a set of rules selects the proper templates, and combines them to form a presentable and interactable book. The content to fill the templates is taken from the inputs to the system, i.e. the book content, the audio narration, and the navigation and synchronization information.

Several of the models can be used to generate the book in run-time, thus offering the possibility of adaptation to changing conditions. To generate a book in design-time, the user interface designer simply has to select the platform, devices, user, environment and presentation models that are most appropriate to the forecast usage situation. This creates a book that is "optimal" (according to the designer's criteria) for the chosen situation, but that has no adaptation possibilities.

To explore an adaptable interface, the execution platform must be augmented with generation possibilities. For example, if a PC is used as the execution platform, with the book being displayed in a web browser, one way to introduce adaptation capabilities in the interface is to install a web server with the capacity to run the generation scripts.

The level of adaptation of the interface can be selected during the design of the interface. By level of adaptation we mean the range of events to which the interface should adapt to. This is achieved by including one or more models in the executable presentation. The only model that is not available for including in the adaptable interface is the platform model. This is justified by the profound changes needed in the interface when migrating to different platforms, and the eventuality of being impossible to execute an adaptable interface in the new platform. All the other models may be included in the execution platform. If the designer is unsure of the availability of some devices in the execution platform, the inclusion of the devices model allows for the adaptation of the interface whenever a device is connected or disconnected

from the platform. If the designer wishes the interface to adapt to the behaviour of the user (for instance, if a user is always requesting to see images enriching the presentation, the interface should start displaying images by default to that user), then the inclusion of a user model may introduce that feature.

An interface without any of the models available in run-time would be a fixed interface. An interface with all the models available would be a fully adaptable interface. The designer should choose the models which will be transferred to the run-time interface, based on her knowledge of the target users, platforms and environment.

5 CONCLUSIONS

In this paper we have presented the motivations for the development of an adaptable interface for DTBs. Variables governing the adaptation and the possible adaptable components have been identified, and several relations between variables and components presented.

A model-based architecture for the creation of flexible multimodal user interfaces for DTBs was proposed. This architecture is the base for further developments, which target the creation of an adaptable user interface. The introduction of some of the interface development models in the run-time version of the interface will allow for the possibility of adaptive behaviour, if supported by the interface presentation platform.

REFERENCES

ANSI/NISO (2002). Specifications for the digital talking book. http://www.niso.org/standards/resources/Z39-86-2002.html.

Brusilowsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2):87–110.

Carr, B. and Goldstein, I. P. (1977). Overlays: A theory of modelling for computer aided instruction. Ai memo, MIT, Cambridge, MA.

Carriço, L., Duarte, C., Lopes, R., Rodrigues, M., and Guimarães, N. (2004). *Computer-Aided Design of User Interfaces IV*, chapter Building Rich User Interfaces for Digital Talking Books. Kluwer Academic Publishers. Accepted for publication.

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

Christian, K., Kules, B., Shneiderman, B., and Youssef, A. (2000). A comparison of voice controlled and mouse controlled web browsing. In *Proceedings of ASSETS'00*, pages 72–79, Arlington, VA.

Duarte, C. and Carriço, L. (2004). Identifying adaptation dimensions in digital talking books. In *Proceedings of IUI'04*, Madeira, Portugal.

Duarte, C., Carriço, L., Chambel, T., and Guimarães, N. (2003a). Producing DTB from audio tapes. In *Proceedings of ICEIS'03*, volume 3, pages 582–585, Angers, France.

Duarte, C., Chambel, T., Carriço, L., Guimarães, N., and Simões, H. (2003b). A multimodal interface for digital talking books. In *Proceedings of WWW/INTERNET 2003*, Algarve, Portugal. Accepted for publication.

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

Kobsa, A. (2001). Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1-2):49–63.

Kok, A. (1991). A review and synthesis of user modeling in intelligent systems. *The Knowledge Engineering Review*, 6:21–47.

Langley, P. (1999). User modeling in adaptive interfaces. In *Proceedings of the 7th International Conference on User Modeling*, pages 357–370, Banff, Alberta.

NISO (1999). Document navigation features list. http://www.loc.gov/nls/z3986/background/navigation.htm.

Not, E., Petrelli, D., Sarini, M., Stock, O., Strapparava, C., and Zancarano, M. (1998). Hypernavigation in the physical space: Adapting presentation to the user and to the situational context. *New Review of Multimedia and Hypermedia*, 4:33–45.

Oppermann, R. and Specht, M. (1999). Adaptive information for nomadic activities: A process oriented approach. In *Proceedings of Software Engineering'99*, pages 255–264, Walldorf, Germany.

Oviatt, S., Darrell, T., and Flickner, M. (2004). Multimodal interfaces that flex, adapt, and persist - introduction. *Commun. ACM*, 47(1):30–33.

Oviatt, S. L., Cohen, P. R., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., and Ferro, D. (2000). Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research directions. *Human Computer Interaction*, 15(4):263–322.

Paternò, F. (2000). *Model-Based Design and Evaluation of Interactive Applications*. Springer Verlag.

Rich, E. (1989). Stereotypes and user modelling. In Kobsa, A. and Wahlster, W., editors, *User Models in Dialog Systems*, pages 35–51. Springer Verlag, Berlin.

Van Buskirk, R. and LaLomia, M. (1995). A comparison of speech and mouse/keyboard gui navigation. In *Proceedings of CHI'95*, Denver, CO.