Chapter 27

BUILDING RICH USER INTERFACES FOR DIGITAL TALKING BOOKS

Luís Carriço, Carlos Duarte, Rui Lopes, Miguel Rodrigues and Nuno Guimarães
LaSIGE, Department of Informatics, Faculty of Sciences of the University of Lisbon

Abstract: This paper presents a framework for the automatic production of Digital Talking Books (DTB). The production process converts existing audio tapes and OCR-based digitalization of text books into full-featured, multi-synchronized, multimodal digital books. The framework deals with the standardization processes, media enrichment and User Interface definition. The latter is based on abstract, yet DTB specific, pattern-based UI specifications. This allows the definition of various forms of interaction and presentation, required by the diversity and constraints of targets users (e.g. visually impaired persons) and situations of use (e.g. learning). Balancing the focus of production between personalized, situation-based UI and adaptive ones is also considered. The article also summarizes some usability tests on generated DTBs that contributed to the refinement of the framework.

Key words: UI generation, model-based tools, accessibility, user diversity

1. INTRODUCTION

Audiotapes have served as an important medium, and sometimes the only alternative, for print-disabled reader's access to books. In several public libraries, in particular in the Portuguese National Library, a long time effort was made in speech recording of a large amount of printed material. However, the limitations of this analogue approach, even when compared

R. Jacob et al. (eds.) Computer-Aided Design of User Interfaces IV, 335-348.

with their printed counterparts are noteworthy (e.g. difficulties in indexing, annotation and cross-referencing). Moreover, for particular disabilities and situations of use the visual complement is also required.

Digital Talking Books (DTBs) are a logical answer. Involving several groups related with visually impaired people, DTB work identifies requirements, directions and recently a standard based on emerging Web technology (ANSI/NISO, 2002; Daisy, 2002). Nevertheless, the standard, wisely, does not intentionally propose specific solutions for interaction. In fact, the required combination of synchronization, structural navigation and annotations management, using visual, audio, speech and standard interaction devices, poses ambiguity and cognitive problems that must be dealt with at the UI design level (Carriço et al., 2003a; Duarte et al., 2003; Morley, 1998). These issues are further stressed by the diversity of targeted users, their particular disabilities and perspectives. It is essential to explore and evaluate distinct UIs for the same book, with different multimodal combinations, eventually enriched with new media contents not present in the original book. Balancing DTBs modes and media, for example, can be explored to overcome the cognitive limitations of human perception and attention (Gazzaniga et al., 1998).

This paper describes DiTaBBu (Digital Talking Books Builder), a framework for the production of DTBs based on media indexing, speech alignment and multimodal interaction elements. The work has been carried out in the context of the IPSOM project, joining the Portuguese National Library (owner and publisher of analogue talking books), speech processing technology experts (providing tools for speech recognition and speech alignment) and multimedia interaction designers and engineers. The framework balances its requirements between: (1) the existence of large amounts of recorded material; (2) the flexibility and simplicity needed for the generation of UIs; (3) the DTB recommendations and standards; and (4) the ability to integrate, explore and adjust multimedia units in the production process. At its current status, the framework already builds on a set of results from usability evaluation studies over produced DTDs, which consolidated and refined several decisions on the execution platform and particularly on the UI specification.

In the following section this article presents the requirements imposed on the production framework by: the particular project needs; the related standards and recommendations; and the results of evaluation tests. Design decisions are also referred. Next, the architecture of the DTBs generated by DiTaBBu framework is presented. The following section describes the frameworks itself, covering the book's content organization and the modular UI generation. Afterwards, some related work is discussed. The paper ends drawing conclusions and delineating future work.

2. REQUIREMENTS AND DESIGN OPTIONS

The construction of the DiTaBBu framework is the result of a set of requirements and design decisions that evolved throughout the project.

2.1 Project needs

The Portuguese National Library (BN) provides services for visually impaired persons. It possesses a large amount of analogue spoken books, recorded by volunteers and stored in analogue audio tapes. At the same time the BN is also committed to build a digital version of books - actually scanned books within a XML/HTML envelope. Although both results are available, a need for its integration, along with the introduction of DTB general functionalities, was clearly felt, particularly by the visually impaired community. Two basic problems were raised at this level:

- the huge amount of existing books, audio tapes and digital copies - requiring an automated form to produce the integrated multimodal books;
- the poor quality of the audio tapes - making it very difficult to automatically generate computable digital audio versions.

The first issue should, of course, consider mechanisms for easy and reusable specifications of books' UIs. Those should adapt to different book contents and cope with various types of users and use settings.

The latter problem was solved by recording a clean audio version of some books, or using existing digital forms. The initial pilot corpus was the "O Senhor Ventura" (a novel by Miguel Torga), read by a professional reader in a sound proof booth. Other digital versions were later used (Serralheiro et al., 2003). For the moment, then, the automation process departs from a digital (and computable) version of the audio and from the existing scanned text. Refinement of the speech alignment component is currently under work, in order to depart from the original analogue tapes.

2.2 DTB Recommendations and Standards

The work around DTBs has recently resulted in a standard specification (ANSI/NISO, 2002). Throughout the process a list of features and functions was identified (NISO, 1999a):

- Support basic navigation (advancing one character, word, line, sentence, paragraph or page at a time, and jumping to specific segments);
- Fast forward and reverse, reading at variable speeds;

- Navigation through tables or control files (allowing the user to obtain an overview of the material in the book);
- Reading notes, cross-reference access, index navigation, bookmarks, highlighting, taking excerpts, searching, and other capabilities.

All these requirements are fully considered in the generated DTBs, except for the variable speed reading and the thinner (character and sometimes word based) basic navigation support. For the first one, a complete speech model must be available in order to maintain low voice distortion. An alternative, currently under evaluation, is the reduction or extension of sentence separation (silence, or breathing times), combined with small speed changes. The implementation of the second feature strongly depends on the ability to isolate character and word sounds from the continuous speech recording. An alternative is the introduction of speech synthesis. Here, cognition issues are raised (Gong and Lai, 2001), which should still be a focus of further evaluation studies. Currently the production framework allows the definition of DTBs minimum synchronization unit, from word up, that determines the basic navigation and playback granularity.

Another result from the DTBs specification process is a categorization of DTBs (Daisy, 2002), according to the functionalities that could be made available to the user under different scenarios:

- full audio with title element only - allows sequential playback and is particularly useful for small devices and mobile settings;
- full audio with navigation control - adds direct access through structural items (e.g. table of contents);
- full audio with navigation control and partial text - adds textual search on specific components;
- full audio and full text - all features available and usually requires fix desktop settings with sophisticated resources;
- full text and some audio - allows listening to some textual components (e.g. pronunciation aids);
- text and no audio - structured text, allowing Braille production.

On this multifaceted perspective of a DTB, the same book “edition”, and to some extent the same book (structure and content), could be presented and interacted in different ways, using different devices (NISO, 1999b) and different media and mode combinations. As a direct consequence, the DTB production mechanism or the DTB execution platform or both should build on an architecture that promotes a clear separation between the books' contents, including the logical and semantic structure (e.g. media correspondence) and the books' user interface (UI). This will reinforce coherence between the several usage settings of the same book, facilitating the maintenance and the specification of UI and navigation.

The DiTaBBu framework enforces this architectural separation and allows the generation of all these categories including the most complex DTB format - full audio and full text.

Finally, the proposed standard (ANSI/NISO, 2002) defines a model of DTBs around a set XML-based Document Type Definitions. A basic architecture is also proposed identifying the modules (files) that should be present (Navigation files, Media files, Synchronization files …). Presentation specifics are handled with style sheets (CSS or XSL) and for synchronization purposes, SMIL 2.0 is recommended. The proposed DTB architecture enables different presentation and interaction designs, and the choice of web-based technology ensures the required wide dissemination.

However, in the final representation, content and presentation are dispersed and intermixed in several modules. For example, for the book's content, the media correspondence is defined in the Synchronization file, where the presentation sequence and timings are also established. On the other hand, for navigation elements, the media correspondence (e.g. table of contents text and speech) is specified in the Navigation file, whereas the time-related presentation is in the Synchronization module. This DTB proposed architecture, although coping with several configurations for the same book (a DTB for each configuration), hardly embraces the intrinsic correspondence among them. It can (as a standard) be used as a final format for DTBs, but a clearer separation of content and UI is required, either on DTB production frameworks or on DTBs architectures that provide an enhanced run-time flexibility or even adaptability (Duarte & Carriço, 2004). Furthermore, other final DTB formats should also be made available, that, for example run on off-the-shelf browsers and common devices. The DITaBBu framework can produce different arrangements and different formats for final DTBs.

2.3 Impact of DTB's Evaluation Results

A set of usability studies on several UIs variants were done. The variants, generated by the first versions of the DiTaBBu framework, departed from the same book ("O Senhor Ventura"). Different synchronization units and different visual and audio marks for the synchronization of navigation anchors, playback and annotation were used, as well as different forms of interaction (pure voice-based, mouse, keyboard and combined). Wizard of Oz tests were conducted to solve the language related problems of speech recognition software.

In terms of the DiTABBu framework the impact of the usability tests was essentially felt in the identification and characterization of the UI

specification language that enables DiTaBBu to generate UI variants for the DTDs. In accordance with the results, the components controlling the interface generation were classified into several main classes, each with a different focus, covering presentation and interaction aspects, for the main book content, as well as for annotations, navigation structures, enriching media, etc. Particular relevant results where obtained for the synchronization facet. For example, on DTBs with multiple media presentations, users require contextual information (such as containing sentence, paragraph or section) when navigation or continuous presentation occurs. Furthermore, evaluation results point to the need for different temporal and spatial based contextual units (e.g. the further the navigation ``jump'' the bigger the required context). The detailed test results can be seen elsewhere (Carriço et al., 2003a, Duarte et al., 2003).

3. DTB ARCHITECTURE AND PLATFORM

In view of the above-mentioned recommendations, the generated DTBs must cope with a diversity of devices and modality combinations, which address the specific characteristics of the books' content, the situation of use and the users. Thus a DTB architecture design that handles a flexible execution is a major prerequisite. Additionally, an easy form of dissemination and integration, with emerging digital publication technologies, is not only a requirement imposed by the source material provider (BN), but for the main target users in general.

A Web-based technology approach was adopted, based on DTB and XML related recommendations, but several final DTB formats and organizations are possible. The general architecture for a generated DTB includes:

- An XML-based content specification, embracing text and other media (in specific formats), media anchoring points, media correspondence (to text or between media) and structure. No UI presentation or synchronization issues are considered at this level.
- A set of XSLT-based specifications enabling the creation of UIs for the content.
- The UI, including presentation and specific interaction objects when required. Presentation could follow several formats and organizations, from plain SMIL (plus CSS), to versions compliant with the DTB standard architecture.

The introduction of the XSLT level permits to build the several UIs, using alternative DTB formats and still maintaining the coherence with the books' content. On the other hand, it also allows balancing the generation of

the UI, between the production framework and the execution platform itself (Figure 27-1 shows an example with a DTB using an HTML+TIME format). If the execution platform is able to process XSLT, a book following the above three-layer organization could be directly used. If performance is an issue or the execution platform does not support it, the DiTaBBu framework could generate the final DTB configurations (e.g. a DTB fully compliant with the standard or a simpler SMIL version, in any DTB category).

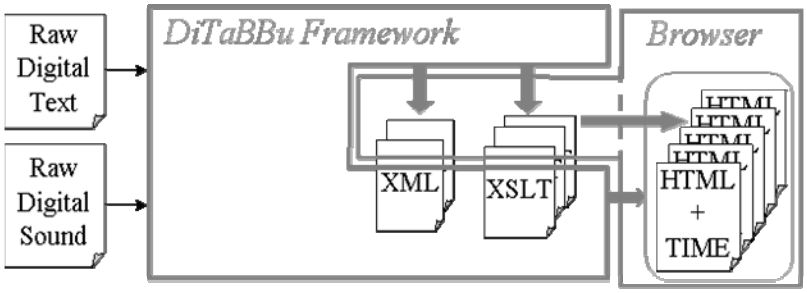

Figure 27-1. Balancing performance and run-time flexibility

As a basic, yet powerful, execution platform, Internet Explorer 6, was adopted. This choice enabled the use of HTML+TIME (and CSS), as a representative of a SMIL 2.0 profile, and Microsoft's variant of VoiceML, for voice interaction. Both architectural organizations are supported, since the browser processes XSLT. In the simplest form, the digital book is a (set of) HTML+TIME, CSS and media specific files. An initial version of the generated DTBs used HTIMEL (Chambel et al., 2001), instead of HTML+TIME. Both languages are still available as a result of the DTB production process. However, only HTML+TIME is currently maintained. For voice interaction, off-the-shelf products, recognizing Portuguese language, were initially used with very bad performance results. The Microsoft's implementation of VoiceML, provided better results, but using English as interaction language. Currently, Portuguese speech recognition software, developed within the IPSOM project's teams, is being integrated.

4. THE FRAMEWORK

The DiTaBBu framework generates DTBs through an automatic production process, configured by a set of specification files that allow the required flexibility. Figure 27-1 presents the framework’s inputs and outputs.

Internally the framework can be decomposed into two main phases: content organization and UI-generation.

4.1 The Content Organization Phase

The content organization phase is represented in figure 27-2. The first step is to determine the time of the written words on the recorded speech (alignment). The book's original text is initially expanded (e.g. abbreviations and numerals are replaced by their complete textual representation) and stripped from punctuation signs. This is feed in the alignment module that generates a table with the audio stream timings for each of the spoken words. Besides the words, this process also identifies the silences (reader's pauses) present in the narration - details on the alignment process can be found elsewhere (Serralheiro et al., 2002).

From the alignment table, the expanded text (text as spoken) and taking again the digital copies of the source text (raw digital text), two XML tagged descriptions are generated. In the first one, derived from the source text, an ID is assigned to every word and the correlation between written and "as spoken text" is maintained whenever is needed (e.g. <word id="10" sounds="one">I</word>). The second file contains the words' timing (e.g. <anchor id="10" unit="word" begin="13"/>) and silences and represents the anchors into the media file (e.g. an "mp3" file with the book's narration).

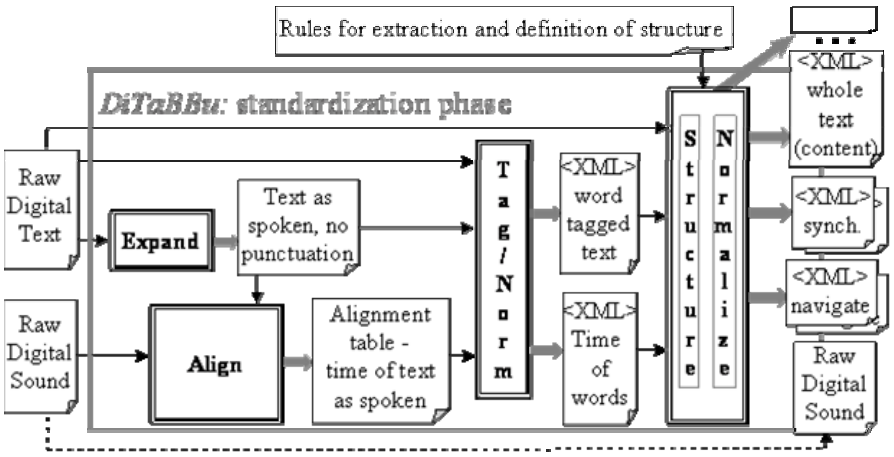

Figure 27-2. DiTaBBu: the content organization phase

The final step on the content organization phase is actually twofold: the book structure is included in the XML-based content file and the remainder DTBs standard files are generated. In the first process a set of rules in terms of regular expressions are used to extract the structure (e.g. paragraphs, sections …) from the digital source text. Alternatively and additionally specific structure definitions can be introduced. The second process provides the main DTB file, the navigation files (extracting table of contents, etc.) and the connection between different media (referred as synchronization). Synchronization units for syntactic constructs and for spoken divisions (breathing and pauses) are added, enabling an easier (tagged based) production of multi unit DTBs (Duarte et al. 2003).

This phase result is mostly compliant with the ANSI/NISO recommendations. Exceptions are the inexistence of UI elements, including the SMIL-based synchronization specification defined on the standard. Instead, the synchronization file generated in this phase of DiTaBBu describes simple media correspondence (through common id tags or meta-information).

4.2 The UI-generation Phase

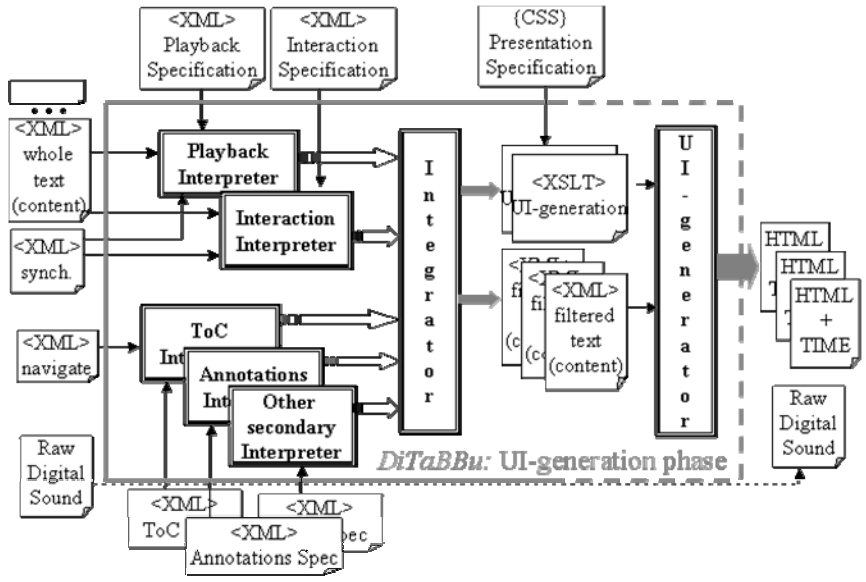

Figure 27-3. DiTaBBu: UI-generation phase

The UI-generation phase is represented in figure 27-3. In the initial steps of this phase the framework presents a set of interpreter modules. Each module receives as input: a set of XML-based files with content; and a specification file, describing the patterns and rules to be applied to that content. Those specification files follow XML-based dialects dependent on the module. Internally the module also uses XSLT code and XSLT templates that are selected and adjusted according to the specification, in order to generate the module's output (XML + XSLT files). Cocoon (Langham & Ziegler, 2002) is used in the transformation process. Two groups of interpreters can be identified relating to primary and secondary material. The primary material modules are the playback and interaction interpreters. Basically they deal with the main book content, not considering footnotes, margin notes and navigation auxiliaries (indexes, tables of content …). The respective dialects handle the visual and audio logical markup and their synchronization. For example:

- <showsync delay="2s" sunit="silence"/> means that playback will show visual synchronization marks (the visual effect is specified later in the CSS) delayed by 2 seconds and using the words between reading silences as a unit - the whole unit is marked (e.g. underlined) as narration evolves;
- <onsearch sunit="word, paragraph" basedon="paragraph, section" /> means that in result of a search, the narration (sound) should start on the word found or on the beginning of the paragraph containing that word depending on how distant (paragraph of section) from the current reading position the searched text is (see impact of evaluation results).

The secondary modules handle auxiliary navigation structures, user annotations, side margin notes, etc. Apart from the specificity of their dialects (e.g. <show summary> on annotations), the synchronization rules with the primary content are also specifiable.

The remainder steps of the UI-generation phase provide the integration and filtering of the results of the interpretation modules and the (optional) generation of the final presentation. The former generates a set of XSLT and XML content files that can be also interpreted by the execution platform.

5. RELATED WORK

Multimodal systems have a high degree of complexity. Even if considering only the integration of speech into a point and click interface, there are a great number of problems to be considered (Oviatt et al., 2000). Nevertheless, research has indicates that speech input is advantageous in several circumstances (Oviatt et al., 2000), and identified the task

characteristics that better suit speech input (Van Buskirk & LaLomia, 1995, Christian et al., 2000). Those tasks are the ones where the user has to issue brief commands using a small vocabulary, which are approximate to the interaction characteristics of a DTB. However, research on the effectiveness of speech as an input mode has not been conclusive (Martin, 1989, Visick et al. 1984). As such, there is still need to experiment systems with different configurations of voice and other mode commands. DiTaBBu platform offers the advantage of flexibility in the creation of DTBs with different presentation and interaction characteristics. That will allow us to try out different ways to convey document structure and assist navigation, such as the use of 3D audio (Goose & Moller, 1999), auditory icons (Gaver, 1993, Blattner et al., 1990), multiple speakers and sound effects (James, 1997), etc.

Referring to DTB formats and architectures, the use of the DTB standard has recently gained momentum. Several software (Dolphin Audio Publishing, 2003; Innovative Rehabilitation Technology inc., 2003; VisuAide, 2003b) and hardware (VisuAide, 2003a) players were made available. However, other web-based solutions should be envisaged, if a wider dissemination and ease of evolution is pursued. For example, formats fully compatible with common Web browsers, like the one proposed in this work, potentially executable in general purpose mobile devices and adopting the mentioned flexible architecture should definitely be available.

Our platform of DTBs and its UI generation shares some of the characteristics of model-based UI development environments, namely, the infrastructures needed for the automation of tasks related with the design and implementation of UI processes (Szekely et al., 1996), and the higher level of abstraction in the description of the interface (Wiecha & Boies, 1990; Puerta & Maulsby, 1997). For example, model-based approaches were adopted to handle flexible generation of UIs for different users and devices (Paternò, 2000). There are several model-based projects addressing the issue of creating UIs for multiple devices (Eisenstein et al., 2001; Ali and Pérez-Quiñones, 2002; Lin and Landay, 2002) or to adapt to different devices (Calvary et al., 2001). This, in fact, is a field where the transition to the commercial software world has not yet occurred, in part because of the abstraction level used in the specification of the models, which contradicts adopted user interface design techniques. However, in the case of DTB production, with the particularities of the domain, there is not such a great emphasis on abstraction. The generation process can, thus, be more easily adopted. Besides, the common notion of “book collection” can definitely compel to reuse, a characteristic reinforced by the automation process.

Still related to model-based UI development is the use of several specifications that are conceptually similar to the models employed in those environments: application model, task model, dialogue model and

presentation model. Future developments, which target the construction of an intelligent interface for the DTBs, will see the inclusion of new specifications to allow an adaptation to the reader and the reading environment. This is similar to the introduction of user and environment models presented by some model-based frameworks (Markopoulos et al., 1992; Puerta, 1996; Elwert & Schlungbaum, 1995).

6. CONCLUSIONS AND FUTURE WORK

This paper presented DiTaBBu, a framework for the production of DTBs. We have described the requirements and design options taken in view of those requirements and of the usability tests already performed on generated DTBs. Currently, the produced DTBs provide most of the functionalities intended in the standards literature, including audio and text synchronization, annotations, navigation through mouse and keyboard interaction and through voice commands.

The platform itself was described. Its architecture based on modules that derived from DTB specific concepts, enables the required flexibility for the creation of multiple UI for DTBs, maintaining the automatic generation premise. The fact that those modules are rule based and template-supported stresses that flexibility.

As ongoing work, we are integrating tools for an higher level of specification for the modules specification files. In line of hypermedia related works (Carriço et al., 2003b, Kraus and Koch, 2002) its being defined an UML description of those specification dialects, that in turn will generate the XML specifications. Some work has also started in the integration of images as secondary book material, including speech based description of such images. The enrichment book process already present in the framework will handle explicit and semi-automatic inclusion of other multimedia related contents in the produced DTBs.

REFERENCES

Ali, M. F. and P´erez-Qui˜nones, M. A. (2002). Using task models to generate multi-platform user interfaces while ensuring usability. In *Proceedings of Human Factors in Computing Systems: CHI 2002 Extended Abstracts*, pages 670–671, Minneapolis, MN.

ANSI/NISO (2002). Specifications for the Digital Talking Book (ANSI/NISO Z39.86-2002), Use and Maintenance. http://www.loc.gov/nls/z3986/v100/index.html.

Blattner, M., Sumikawa, D., and Greenberg, R. (1990). Earcons and icons: Their structure and common design principles. Gilbert, E. (ed.) Visual Programming Environments: Applications and Issues, Los Alamitos, CA, IEEE Computer Society Press.

Calvary, G., Coutaz, J., and Thevenin, D. (2001). A unifying reference framework for the development of plastic user interfaces. In *Proceedings of Engineering for Human-Computer Interaction: EHCI 2001*, p. 173–192, Toronto, ON, Canada. Springer Verlag.

Carriço, L., Guimarães N., Duarte, C., Chambel, T. and Simões, H. (2003a). Spoken Books: Multimodal Interaction and Information Repurposing. HCII'2003, International Conference on Human-Computer Interaction, Crete, Greece.

Carriço, L., Lopes, R., Rodrigues, M., Dias, A. and Antunes, P. (2003b). Making XML from Hypermedia Models. In: Proceedings of the IADIS International WWW/Internet Conference. Algarve, Portugal, November 2003

Chambel, T., Correia, N. & Guimarães N. (2001). Hypervideo on the Web: Models and Techniques for Video Integration. Int. Journal of Computers & Applications, 23 (2).

Christian, K., Kules, B., Shneiderman, B., and Youssef, A. (2000). A Comparison of voice Controlled and Mouse Controlled Web Browsing. Proceedings of ASSETS'00, Arlington, VA.

DAISY (2002). Daisy Structure Guidelines. http://www.daisy.org/publications/ guidelines/sg-daisy3/structguide.htm.

DAISY (1999). Statement of Principles for the Creation and Production of Accessible Books and Materials. http://www.daisy.org/dtbook/guidelines/draft/ principles.htm.

Dolphin Audio Publishing (2003). EaseReader - the next generation DAISY audio eBook software player. http://www.dolphinse.com/products/easereader.htm.

Duarte, C., Chambel, T. Carriço, L., Guimarães N., & Simões, H. (2003). A Multimodal Interface for Digital Talking Books, WWW/Internet International Conference, Algarve, Portugal.

Duarte, C. and Carriço, L. (2004). Identifying Adaptation Dimensions in Digital Talking Books. Submitted for publication in IUI'2004, Intelligent User Interfaces, ACM Press. Madeira, Portugal.

Eisenstein, J., Vanderdonckt, J., and Puerta, A. (2001). Applying model-based techniques to the development of UIs for mobile computers. In *Proceedings of the International Conference on Intelligent User Interfaces: IUI 2001*, p 69–76, Santa Fe, NM. ACM Press.

Elwert, T. and Schlungbaum, E. (1995). Modelling and Generation of Graphical User Interfaces in the TADEUS Approach. Designing, Specification and Verification of Interactive Systems, Vienna, Springer.

Gaver, W. (1993). Synthesizing Auditory Icons. Proceedings of INTERCHI'93, Amsterdam.

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

Gong, L. and Lai, J. (2001). Shall we mix synthetic speech and human speech? impact on user's performace, perception, and attitude. In *Proceedings of the ACM CHI 2001*, pages 158–165, Seattle, Washington.

Goose, S., and Moller, C. (1999). A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. Proceedings of the 7th ACM Conference on Multimedia, Orlando, FL.

Innovative Rehabilitation Technology Inc. (2003). eClipseReader. http://www.eclipsereader.com/.

James, F. (1997). Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext. Proceedings of ICAD'97, Palo Alto, CA.

Kraus, A. and Koch, N., 2002. Generation of Web Applications from UML Models using an XML Publishing Framework. *In 6th World Conference on Integrated Design and Process Technology (IDPT)*

Langham, M. and Ziegler, C. (2002). Cocoon: Building XML Applications, New Riders.

Lin, J. and Landay, L. (2002). Damask: A tool for earlystage design and prototyping of multi-device user interfaces. In *Proceedings of the 8th International Conference on Distributed Multimedia Systems*, p. 573–580, San Francisco, CA.

Markopoulos, P., Pycock, J., Wilson, S. and Johnson, P. (1992). Adept - A task based design environment. Proceedings of the 25th Hawaii International Conference on System Sciences, IEEE Computer Society Press.

Martin, G. (1989). The Utility of Speech Input in User-Computer Interfaces. International Journal of Man-Machine Studies 30, 4.

Morley, S. (1998). Digital Talking Books on a PC: A Usability Evaluation of the Prototype DAISY Playback Software. ASSETS '98, ACM Conference on Assistive Technologies.

NISO (1999a). Document Navigation Features List. Available at http://www.loc.gov/nls/z3986/background/navigation.htm.

NISO (1999b). Playback device guideline. http://www.loc.gov/nls/z3986/background/features.htm.

Oviatt, S., Cohen, P., 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).

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

Puerta, P. (1996). The Mecano Project: Comprehensive and Integrated Support for Model-Based Interface Development. Computer-Aided Design of User Interfaces, Namur, Belgium, Namur University Press.

Puerta, A. and Maulsby, D. (1997). Management of Interface Design Knowledge with MODI-D. Proceedings of IUI'97, Orlando, FL.

Serralheiro, A., Caseiro, D., Meinedo, H., Trancoso, I., and Neto, J. (2002). Spoken Book Alignment Using WFSTS, in *Proceedings of HLT'02*. San Diego CA.

Serralheiro, A., Trancoso, Caseiro, D., Chambel, T. Carriço, L. & Guimarães, N. (2003). Towards a repository of Digital Talking Books. Eurospeech 2003.

Szekely, P., Sukaviriya, P., Castells, P., Muthukumarasamy, J. and Salcher, E. (1996). Declarative Interface Models for User Interface Construction Tools: the MASTERMIND Approach. Engineering for Human-Computer Interaction, London, UK, Chapman & Hall.

Van Buskirk, R. and LaLomia, M (1995). A Comparison of Speech and Mouse/Keyboard GUI Navigation. Proceedings of CHI'95, Denver, CO.

Visick, D., Johnson, P., and Long, J. (1984). The Use of Simple Speech Recognizers in Industrial Applications, in Proceedings of INTERACT'84, London, UK.

VisuAide (2003a). Victor reader classic. http://www.visuaide.com/victorclassic.html.

VisuAide (2003b). Victor reader soft. http://www.visuaide.com/victorsoft.html.

W3C (2001). Synchronized Multimedia Integration Language (SMIL 2.0). http://www.w3.org/TR/smil20/.

Wiecha, C. and Boies, S. (1990). Generating user interfaces: principles and use of ITS style rules. Proceedings of UIST'90.