ANFORA: Investigating Aural Navigation Flows On Rich Architectures

Romisa Rohani Ghahari
School of Informatics
Indiana University
Indianapolis, IN, USA
rrohanig@iupui.edu

Davide Bolchini
School of Informatics
Indiana University
Indianapolis, IN, USA
dbolchin@iupui.edu

***Abstract*— People use mobile web applications in a variety of contexts, typically on-the-go, while engaged in other tasks, such as walking, jogging or driving. Conventional visual user interfaces are efficient for supporting quick scanning of a page, but they can easily cause distractions and accidents. This problem is intensified when web information services are richer and highly structured in content and navigation architectures. To support a graceful evolution of web systems from a conventional to an aural experience, we introduce ANFORA (Aural Navigation Flows On Rich Architectures), a framework for designing mobile web systems based on automated, semi-controlled *aural* navigation flows that can be listened to by the user while engaged in a secondary activity (e.g., walking). We demonstrate a set of design rules that could govern salient aural interactions with large web architectures. Our approach opens a new paradigm for aural web systems which can complement existing visual interfaces, and has the potential to inform new technologies, navigation models, design tools, and methods in the area of aural web information access. As case study, we are applying ANFORA to the domain of web-based news casting.**

Keywords: aural flows, aural web, acoustic interfaces, information and navigation architecture.

I. Introduction

Accessing the mobile web on-the-go and in a variety of contexts (e.g., walking, standing, jogging, or driving) is becoming more and more pervasive. Mobile users are often engaged in another activity when it is inconvenient, distracting or even dangerous to look at the screen at all times. Although existing visual user interfaces are efficient to support quick scanning of a page, they typically require highly focused attention and may not work well in most situations. Recent studies on the use of mobile devices during secondary tasks indicate that *audio-based* interfaces – although slower to use – are less distracting compared to visual interfaces [5].

Besides the known issue of modality (aural vs. visual), a fundamental concern is the degree of required or desired interactivity with the web application. Let us picture a scenario in which you are walking on a city street and would like to catch up with the weekly local news during your 10-minute walk to work. Continuous interaction with a conventional news site on your smart phone would force you to scan the homepage, get an idea of the latest news, select a category, maybe a subcategory, and then finally choose a news story to read. Once read, you may want to

Figure 1. Typical Mobile Contexts for Eyes-free Aural Experiences.

know more about it or select another news story in the same category, etc. How much of this interactivity is in conflict with the current task of walking to work? Is it a desirable user experience? Well-designed "acoustic" interfaces have the potential to unleash new ways to support appropriate eyes-free browsing of mobile web systems. The design of the aural experience needs to pay particular attention to the viscosity of the interaction and then exploit the aural channel to alleviate continuous, focused attention on the interface mechanics.

Designing for aural mobile consumption is not new. On the one hand, the variety of audio books and audio guides already addresses (in the museum and tourism domain) an important need for aural interaction in specific contexts. On the other hand, in the realm of web accessibility, solutions have been proposed to automate, auralize and "replay" repetitive web tasks (such as paying the monthly bills) [4]. In both cases, however, the nature and structure of these applications follow a very linear and flat pattern (e.g., listening to a book page by page) and do not reflect the growing complexity of information architectures that are pervasive and common in web applications.

As the experience with the aural medium is structurally linear, an unsolved challenge is to properly "auralize" complex web navigation architectures for the mobile user experience. We need therefore to tackle and unpack the tension between the linearity of the aural medium and the non-linearity of the information and navigation architecture of web applications. The non-linearity typical of the web is characterized by the co-presence of different organizational structures (e.g., hierarchical and hypertextual) to support overall user interaction with content. Instead of building from scratch completely separate, "niche" applications for aural interaction, a key solution is to find an approach that could support a *graceful evolution* of current web systems (made of rich content organization and composite structures) toward applications that engender eyes-free browsing experiences.

To address this challenge, we propose a novel, semi-interactive aural paradigm (ANFORA: *Aural Navigation Flows On Rich Architectures*), allowing users to mainly *listen* to the information-rich pages in complex, hypertextual web structures and interact with them infrequently. We introduce the notion of "aural flow" and investigate new ways in which different types of aural flow can be effectively applied to conventional web information architectures. ANFORA explores a number of design alternatives, which have the potential to enhance quick scanning through content-rich pages when the time and contextual or physical constraints are at play. As case study, we are applying ANFORA to the domain of web-based news casting.

II. ANFORA: AURAL NAVIGATION FLOWS ON RICH ARCHITECTURES

A. An "Aural Flow" Scenario

Mary is a college professor who usually goes home around 6 p.m. She wants to know about the most interesting news stories happening in the U.S. this week. A traditional local radio station, however, would likely provide today's news only and include topics not relevant to her, such as economy and public policy in the state. She could go to the National Public Radio (NPR) website and play audio news stories one by one. But this would be inconvenient and dangerous while she is driving. With ANFORA, she uses her mobile phone to listen to the top 10 U.S. news stories of the week during her drive home. She listens to every story, delves deeper into related news stories, as well as skips and pauses the news "flow" at will with a simple swipe.

B. The Application Domain of Newscasting

ANFORA is a generic design framework and is applicable to any non-linear, content-rich architecture that underlies modern web systems. For example, ANFORA is appropriate for any large website that features hierarchical and hypertextual structures, such as commerce, travel planning or tourism sites, just to name a few domains. We chose the news domain as the application example because of its complexity, layered approach to storytelling and appeal with mass audiences. Fig. 2 shows an excerpt of a typical information architecture for a news site.

C. Modelling Semi-interactive Aural Flows

Let us consider the information architecture illustrated in Fig. 2. We use the Interactive Dialogue Model (IDM) [2] primitives and notations to characterize the basic features of the architecture. IDM provides basic concepts to describe and model hypertextual non-linear architectures. IDM is based on the notion that user interaction can be considered a dialogue between the user and the system. In a nutshell, core content entities (e.g., the news) are *multiple topics*. A multiple topic can be structured in *dialogue acts* (news story, commentary on the news story) corresponding to different pages or interaction units composing the topic. Multiple topics are typically organized in *groups of topics* (e.g., U.S. news or world news) at different hierarchical levels. Hypertextual or semantic associations are typed and can be characterized as structural relationships between multiple topics. For example, a news story may have a set of related news stories associated with it on the basis of given semantics (e.g., written by the same field reporter, or about the same "hot" topic). This sample structure can be extended to include a variety of groups of topics, a variety of multiple topics, and a variety of semantic associations. Other existing web design models offer similar primitives and can be used to characterize similar architectural features of web applications.

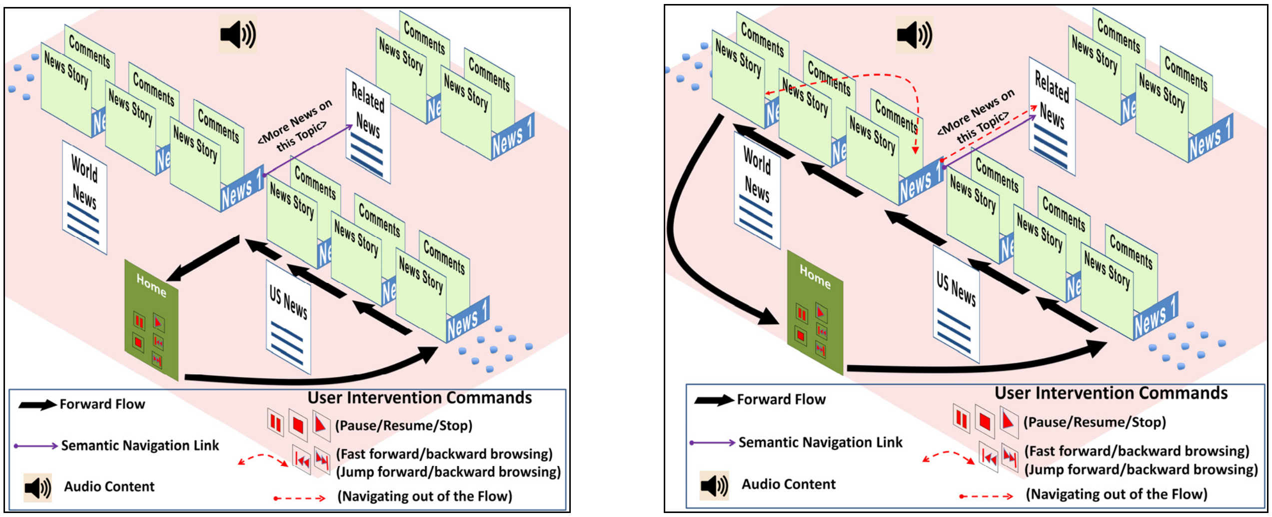
In the context of this traditional architecture, we introduce the notion of "aural flow," defined as follows:

> *An aural flow is a design-driven, concatenated sequence of pages that can be listened to with minimal interaction required. A flow is governed by aural design rules that determine which pages of the information architecture to automatically concatenate and at which point of the flow the user can interact.*

Aural flows are modeled on top of existing web information and navigation architectures and can co-exist with the traditional navigation and interaction paradigm. Existing web applications follow a combination of hierarchical and hypertextual paradigms to organize the structure of the information architecture. An aural flow takes these existing structures and linearizes them appropriately for the aural experience. Therefore no design changes to the existing websites are necessary from the information architecture perspective. Adopting flows can be a powerful way to make web systems evolve as "aural" applications exploiting the current content and structure. For example, users will still be able to "browse" news content by looking at the screen and clicking on links. But as an appropriate alternative, they can also activate and listen to content in "flow mode." Experiencing the flow would allow users to listen to information in a variety of different mobile contexts, with less distraction on their primary task (e.g., walking) and less need to continuously interact with the application to make a selection. As a general strategy, an aural flow can be considered an evolution of the "guided tour" navigation pattern which has been very popular and heavily used in conventional web applications for decades [13]. In guided tour navigation, users let themselves be "led around" by the application (e.g., just selecting "next" or "previous" commands), according to the appropriate sequences of content conceived by the designers.

D. General Interaction Commands

The news casting examples well shows how ANFORA allows users to engage with content on varying levels. For example, after choosing a main group of topic, such as U.S. news, a user could listen to all of the headlines or story



(a) (b)

Figure 2. Aural flows in a typical web architecture: (a) Group Flow through one group of topics at a time; (b) Full Flow through all groups of topics.

summaries in that category. Users would be able to navigate through all the news stories in one category and *continue the flow* with the next category of news or related stories. At any time during the experience, users may pause the flow (in one or more input modalities). The reading shall resume from the beginning of the news story the user was listening to while interrupted (Fig. 2b). To generalize, at any time, users can *interact with the flow* through these basic actions:

- *Pause, resume, replay and stop*: The user can pause and resume the flow. The same dialogue act can be replayed from the beginning. The user can also stop the flow to go back to the home page.
- *Fast forward/backward browsing*: The user can fast forward to go to the next dialogue act of the same topic or fast backward to go back to the previous dialogue act of the same topic.
- *Jump forward/backward browsing*: The user can jump forward to the next topic or jump backward to the previous one at any time.
- *Navigating out of the flow*: The user might want to listen to the related topic by clicking on its link. This action breaks the current flow and moves outside the flow to the desired content (e.g., Related News).

The above interaction semantics could be realized in a variety of input modalities, such as “point and click”, speech, or touch-based gestures. Traditional point and click would require users to focus their attention to small links on the limited interface real estate. As to vocal commands, they would easily increase the cognitive load to remember the vocal vocabulary, and the noise mobile environments would cause non trivial interferences. In our approach, so far, we used gesture commands (e.g. single tap to pause the flow, double tap to stop, right flick to jump forward, etc.,) for their simplicity and availability on the common smart phones platforms (e.g. iPhone). Aural flows, controlled by the basic interaction semantics just illustrated, come in *different types*, to support the variety of requirements of the mobile experience. These types of flows can be selected by the user up front (e.g. from the homepage), or be always available to be activated on demand.

E. Short Aural Explorations with “Group Flow”

There are situations in which a user may have only a few minutes to listen to interesting content while on-the-go. “Group Flow” provides users with aural access to a selected group of topics (e.g. world news) and plays all the instances of news stories belonging to the selected group. The flow stops as all the news stories in the group have been read and the user is led back to the homepage (Fig. 2a). One advantage of “Group Flow” is that users can decide from the outset which category of content they would like to listen to, and they have this choice every time a group is ended. A favorite group flow (e.g., news in politics) can also be “bookmarked” and readily proposed to the user to play at the start of every usage.

F. “Full Flow” for Prolonged Aural Experiences

There are situations in which users have longer periods of time (e.g., 30 minutes or so) to devote to listening to interesting and engaging content, for example during long drives or extensive walks. Audio books, for example, exploit exactly this type of situation while in the car. In a typical web information architecture (such as the one exemplified in Fig. 2), an aural “Full Flow” is the concatenation of all instances of multiple topics across some or all of the designed groups of topics. This allows the user to experience the full set of main content pieces that are available. The length of the flow is of course determined by the number of topic instances in each group and by the number of groups. For example, a typical full flow may involve two main groups of topics (U.S. news and world news), each composed of three to four “hot” news stories, for a total flow experience of approximately 20 minutes. A

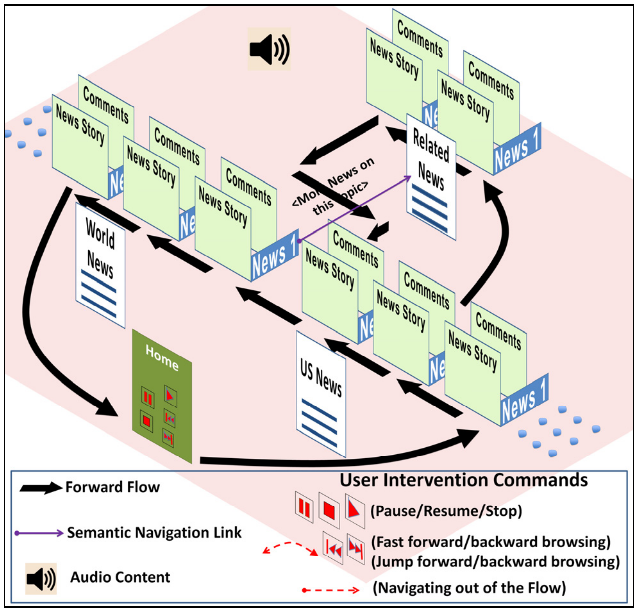


Figure 3. Automated deep flow through all groups of topics and semantic associations.

full flow would work as follows (Fig. 2b). On the user's request, full flow would start from the homepage, read the name of the first group of topics (U.S. news) and play the content of the first U.S. news story (e.g., "Gas Prices: Lessons From The Carter Years"). After the first news story has been read, the following story is read, and so on. Eventually, as a new group of stories are encountered (world news) a simple aural cue (text-to-speech or sound snippet) indicates the change of category. At any time, full flow can be paused and resumed, stopped, started over, or fast browsed back and forth (Fig. 2b). Navigating out of the flow (e.g., following a semantic association, or hitting any main landmark link such as "home") will stop the flow. Some of the obvious disadvantages are that users might not perceive changes from one group of topics to another and may have difficulty building a mental model of the content structure being played. Therefore, we plan to provide visual orientation information for the user by displaying on the screen the full text of the story, as well as any labels necessary to illustrate to the user which type of flow and category of news is currently being read.

G. Advanced Aural Browsing with "Deep Flow"

A user may have even longer periods of time available, such as one hour, to listen to broad range of content. "Deep Flow" provides the user with the opportunity to *also* automatically follow the semantic associations from a given topic. For example, as shown in Fig. 3, the user selects "U.S. news" to start the aural flow. The recent headline is "Gas Prices: Lessons From The Carter Years." There are also two related stories associated with it: "The Cycle Of Fear Over Rising Gas Prices" and "Prices Going Down Next Week?" which will also be played. After these related stories are played, the flow will continue to read the second most recent news story in the U.S. category.

Table 1. Different characteristics of aural flow types.

Flow	Characteristics	Time	Advantages	Disadvantages
Group	A selected group of topics	5 min.	Decide the category from the outset	Interact every time to select a different category
Full	All groups of topics	Longer period of time - 30 min.	Less interaction	Difficulty building mental model
Deep	All groups of topics + semantic associations	Longer period of time – 1 hr.	In-depth coverage of content	Difficulty building mental model
Light	Agile overview of each topic (default dialogue act)	Shorter period of time	More stories in less time (agile overview)	Details of each topic will not be played
Rich	Extensive coverage of each topic (all dialogue acts)	Longer period of time	Extensive coverage	Time-consuming and constraining

One advantage of this design, which differentiates it from the previous two designs, is that it provides in-depth coverage of the content available by exploiting associations among content pieces suggested by designers. A disadvantage, however, is that users might lose the track of where they started and which group with which they are currently engaged.

H. Light and Rich Flows

To enhance the flexibility of aural navigation, there are also two sub-modalities through which to experience the aural flow of content (the multiple topics), which we call "light" and "rich." Note that these are applicable to all types of aural flows.

In non-trivial web architectures, a news story with multiple topics is typically structured in different pages. For example, every news story has a full story and the comments from other users. Regardless of the specific type of flow selected, the user can decide to aurally browse the news only by the full story page (before moving to the next news in the flow), or by also listening to the user's comments. We call "Light Flow" the former aural strategy and "Rich Flow" the latter. Light flow allows quick exploration of the main content by providing an overview of the material (the first page has typically the most representative content of the story package). Rich flows would allow a far more extensive exploration of the details (as spread out in multiple pages) of each new story.

One advantage of light flow is that user can listen to more stories in the less time. This modality may work well to provide agile overviews of the content available. The details of each topic, however, will not be played (without the user necessarily being aware of it). Rich flow provides extensive, ample coverage of the full content of each topic. If users have limited time, however, rich flow may be too time-consuming and constraining. A user can switch between the different flow types at any time by going back to the homepage and selecting a different interaction pattern.

Table 1 shows all the different types of flow and their respective characteristics.

III. Related Work

In the following sections, we review relevant previous works on auralizing navigation architectures in different contexts and research communities.

A. Automating Browsing Tasks

WebVCR allows users to record and replay their browsing steps (filling out a series of forms in order to access data in travel websites) in smart bookmarks as a shortcut to web content [1]. Similarly, Chickenfoot is a Mozilla Firefox extension, which allows users to automate and customize their web experiences without changing the source code [3]. Along the same line, Koala – which is a collaborative system – allows users to record and replay their personalized business processes, such as purchasing or making travel arrangements [11]. Another collaborative scripting environment which is an enhancement on Koala is called CoScripter [9]. It has been recognized that automating repetitive browsing tasks (such as checking email and paying bills) can reduce the user interaction level with the application. Borodin [4] also developed an approach by recording personalized macros and replaying them later. Hence users do not need to actively repeat the same tasks over and over again. These applications automate the linear tasks in the architecture to reduce the continuous user interaction and repetition. Moreover, these applications allow users to record and replay the browsing tasks later. However, automating the non-linear tasks in the complex and non-linear architecture is not addressed. Complex architectures are very frequent, and they are among the majority of existing web applications. ANFORA automates complex information architecture and allows users to select how they prefer to listen to information to reduce interaction and distraction.

B. Audio Interfaces compared with Visual Interfaces

The following studies emphasize using audio interfaces over visual interfaces, as well as some of the reasons audio interfaces may be preferred. Recent study shows that audio interfaces in the car are less distracting compared to traditional visual interfaces, though the former might be slower to use [5]. This study addresses the distraction involved with visual interfaces and discusses how visual interfaces might still be preferred to audio interfaces because they enable users to scan quickly through pages. Li also introduces the blindSight prototype, which replaces the traditional visual interface and provides audio feedback. This could help users access their calendars or contact lists without looking at the mobile screen [10]. Li's study shows how audio interfaces could play important roles in quickly accessing and interacting with the system while engaged in another primary task. ANFORA research introduces an automated aural interface to reduce interaction levels and enable quick scanning through information, which was not addressed as a possible solution in previous studies.

C. Navigation Strategies in Audio Book

The first digital talking book (DTB) was called DAISY (Digital Audio-Based Information System). "The DAISY recording system allows digital audio to be recorded and coded at several levels of detail such as phrases, groups of phrases and section headings." Moreover, the DTB gives access to full text, and users can interact with it using keyboard [12]. Similarly, the Mobile Rich Book Player prototype is DTB and uses the Windows Mobile platform [6]. On the same line, Jain and Gupta [7] present a system called VoxBoox, which generates automatic interactive talking books. This system converts digital books to audio books and makes them accessible to blind users. Blind users can interact with VoxBoox using voice commands (skip, back, repeat, start, end, and pause) and call a toll free number to connect with a voice browser [7].

Audio books follow linear architectures (table of contents, chapters, sections, etc.). Existing web applications, however, tend to be non-linear (news category, summary, full story, related news, etc). These studies did not address auralizing the non-linear structures. In designing ANFORA, however, we reused generic commands such as play, stop, pause, and resume used in audio books.

D. Audio News Navigation Strategies

Sound News is an audio and tactile interface which lets users to browse the web version of the newspaper eyes-free. Sound News, however, uses three keys and three tracks for touch pad [15]. This application provides eyes-free navigation; users however, still need to interact with the touch pad continuously.

E. Navigation Strategies in Audio Wiki

Wang et al. use audio as the interaction medium for a wiki called Mobile Audio Wiki [14]. Similarly, Kolias et al. created a different type of audio wiki application which is available for mobile phones and web browsers. Blind users can call a number, which is fixed with the application, and speak their selections (category and term). Finally, the audio article content is read for the user [8]. Both of these systems allow users to navigate the linear structure of wiki using the audio version, which plays content automatically. To date, there have been no solutions that automate and auralize content-rich, non-linear information architectures to offer users a less distracting non-visual way to browse the web. We address the non-linear structure because the majority of existing web applications are non-linear and users often use them while doing other primary tasks. ANFORA represents a viable solution in which automation and auralization can significantly improve non-visual web browsing.

IV. ANFORA: Potential and Limitations

The overall potential of ANFORA is that it offers simple aural design strategies that can support analysts and designers in the graceful and disciplined evolution of current web systems from a visually-driven perspective to an aurally-driven paradigm. Within this vision, there are number of advantages and drawbacks. The first advantage of ANFORA is that it enables users to get an eyes-free, aural experience of architected web content when they are engaged with other primary tasks. This goal is achieved by reducing user interaction through automated aural flows. A second advantage is that users can select among different types of aural flow according to the contingent, contextual requirements, information needed, and time available. A third advantage is that users can still intervene and control the flow with simple interaction commands.

One of the limitations of the current version of ANFORA is the fact that a specific topic may belong to different groups and could be therefore repeated several times in automated flow. This issue could be overcome by marking those topic instances and play them only once according to user's navigation history in the current session. A second disadvantage of ANFORA is that users may become bored from listening to TTS content originally written for the reading experience rather than the listening experience. However, more and more, news organizations offer stories in multiple formats – i.e., text-based stories, as well as audio stories that are "written for the ear." Thus, we may overcome the issue of boredom by providing stories in multiple formats, which are more suitable for audio information consumption. A third limitation is related to the complexity of creating a mental model for users so they do not become disoriented. For example, users might not know after few minutes which news category they are listening to. Moreover, they might not know how many news stories are in a particular category and how many more stories are left in that category. To overcome this problem, we could add short numeric cues (aurally and visually) to clearly communicate navigation status and orientation in the flow to the users. As a corollary to this, we may also provide users with cues conveying the approximate time length of the flow to be started, or – on demand – indicate the time left for the flow to finish.

V. CONCLUSIONS AND FUTURE WORK

ANFORA is a conceptual framework, based on the notion of "aural flow", to design innovative audio interfaces on top of existing web information architectures. We presented three main types of aural flows to cater for different contexts of use and time constraints of the eyes-free, mobile experience. We are currently developing prototypes for each type of aural flow. ANFORA provides a generic framework that can be applied in any content-rich web information architecture to make it accessible for sighted users while they are engaged in another primary task to reduce continuous user interaction and distraction. Our current and future work is mainly focused on improving the conceptual design of the aural flows, finalizing a series of system prototypes and carrying out a comprehensive user-centered evaluation. The evaluation of ANFORA will be accomplished in two main steps. First, we will conduct an exploratory, qualitative study to collect initial user's feedback on the utility, usability (e.g. interaction mechanics, orientation and listening time) and desirability of the aural flows in mobile contexts. On this basis we will improve the conceptual approach and the aural flow rules and then run a controlled experiment to compare the effect on the user experience of the developed aural prototypes with conventional, mobile web systems.

ACKNOWLEDGMENTS

This research material is based on work supported by the National Science Foundation under Grant #1018054. We also thank Tao Yang, Mexhid Ferati and Jennifer George-Palilonis for comments to improve the manuscript.

REFERENCES

[1] Anupam, V., Freire, J., Kumar, B., & Lieuwen, D. (2000). Automating Web navigation with the WebVCR. *Computer Networks, 33*(1-6), 503-517.

[2] Bolchini, D., & Paolini, P. (2006). Interactive dialogue model: a design technique for multichannel applications. *Multimedia, IEEE Transactions on, 8*(3), 529-541

[3] Bolin, M., Webber, M., Rha, P., Wilson, T., & Miller, R. C. (2005). Automation and customization of rendered web pages. *Proceedings of the 18th annual ACM symposium on User interface software and technology*, 163-172.

[4] Borodin, Y. (2008). Automation of repetitive web browsing tasks with voice-enabled macros. *Proceedings of the 10th international ACM SIGACCESS conf. on Computers and Accessibility*, 307-308.

[5] Brumby, D. P., Davies, S. C. E., Janssen, C. P., & Grace, J. J. (2011). Fast or safe?: how performance objectives determine modality output choices while interacting on the move. *Proceedings of the 2011 annual conference on Human factors in computing systems*, 473-482.

[6] Duarte, C., & Carriço, L. (2009). When You Can't Read It, Listen to It! An Audio-Visual Interface for Book Reading. *Universal Access in Human-Computer Interaction. Applications and Services*, 24-33.

[7] Jain, A., & Gupta, G. (2007). VoxBoox: a system for automatic generation of interactive talking books. *Universal Access in Human-Computer Interaction. Applications and Services, 4556/2007*, 329-338.

[8] Kolias, C., Kolias, V., Anagnostopoulos, I., Kambourakis, G., & Kayafas, E. (2010). Design and implementation of a VoiceXML-driven wiki application for assistive environments on the web. *Personal and Ubiquitous Computing, 14*(6), 527-539.

[9] Leshed, G., Haber, E. M., Matthews, T., & Lau, T. (2008). CoScripter: automating & sharing how-to knowledge in the enterprise. *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 1719-1728.

[10] Li, K. A., Baudisch, P., & Hinckley, K. (2008). Blindsight: eyes-free access to mobile phones. *Proceeding of the 26th annual SIGCHI conference on human factors in computing systems*, 1389-1398.

[11] Little, G., Lau, T. A., Cypher, A., Lin, J., Haber, E. M., & Kandogan, E. (2007). Koala: capture, share, automate, personalize business processes on the web. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 943-946.

[12] Morley, S. (1998). Digital talking books on a PC: a usability evaluation of the prototype DAISY playback software. *Proceedings of the third international ACM conference on assistive technologies*, 157-164.

[13] Paolini, P., Garzotto, F., Bolchini, D., Valenti, S. (1999). 'Modelling by Pattern' of Web Applications, in Peter P.S. Chen (Eds). *Proceedings of Advances in Conceptual Modeling (ER'99) Workshops on Evolution and Change in Data Management, Reverse Engineering in Information Systems, and the World Wide Web and Conceptual Modeling*, 293-306.

[14] Wang, L., Roe, P., Pham, B., & Tjondronegoro, D. (2008). An audio wiki supporting mobile collaboration. *Proceedings of the 2008 ACM symposium on applied computing*, 1889-1896.

[15] Zancanaro, M., Stock, O., & Alfaro, I. (2003). Using cinematic techniques in a multimedia museum guide. *Proceedings of Museums and the Web 2003, 12*.