

Educational Impact of Structured Podcasts on Blind Users

Maria Claudia Buzzi, Marina Buzzi, Barbara Leporini, Giulio Mori

CNR-IIT, via Moruzzi 1, 56124 Pisa, Italy
CNR-ISTI, via Moruzzi 1, 56124 Pisa, Italy
{Claudia.Buzzi, Marina.Buzzi, Giulio.Mori}@iit.cnr.it,
Barbara.Leporini@isti.cnr.it

Abstract. Podcasts are increasingly present in education. However, sequential and continuous listening to didactic material might not be the most appropriate form for learning. For instance, if not well-organized, it can be frustrating to retrieve specific information from educational podcasts. We believe that structured podcasts (i.e., audio files that exploit the document structure) could facilitate navigation and search of content. In this paper we investigate the impact of structured podcasts on blind users, through a user test and satisfaction questionnaire administered to eighteen visually-impaired persons. Results confirm that structured podcasts are simple to use, rapid to explore and are greatly appreciated by blind users.

Keywords: audio podcast, mp3, learning, blind, screen reader

1 Introduction

Audio podcasting is increasingly used in education. Especially appreciated as an ubiquitous tool, podcasting is a flexible technology that adapts to users' needs and habits and facilitates learning anywhere, anytime, at any pace [1]. Numerous studies have shown the effectiveness of podcasting for many students who learn better when listening to educational material instead of accessing the written version [2], [3], [4], [5]. When podcasting is used for educational purposes, easy-to-use content facilitates the learning process. Sequential reading of all content may not be the best way to promote learning, while exploration and the opportunity to skip quickly to a specific part might be more appropriate for memorizing the main concepts and building and consolidating ideas. In our approach we consider features that are useful for providing audio content in educational settings. We propose to offer structured audio materials that permit students to easily obtain an overview of the content as well as find specific parts quickly.

Previous studies have shown that short podcasts (e.g., max 10-15 min) are more effective for learning than a single long unit. Long podcasts may decrease attention, reducing comprehension [6], [4]. Establishing rules and methods for achieving an appropriate content structure is a challenge.

A text-to-speech (TTS) system is software that converts text into speech [7]. Available text-to-audio converters (such as Text2mp3, DSpeech) usually produce a single audio podcast or a set of time-based mp3 files. Adding "break strings" to the

document also makes it possible to split the content into several audio files, but this requires considerable time and effort by the user.

In [8] we proposed a Web-based Text-To-Speech system for automatic generation of a set of structured audio podcasts from a single source document. The system takes a document as input, and generates as output a set of podcasts that reflects the content structure (one mp3 file for each document section), ready to be downloaded and uploaded on an mp3 player.

Listening to structured audio content offers many advantages over sequential listening: the user can easily locate audio parts of interest, and better orient him/herself when searching for specific information within the podcast, aiding the learning process.

In this paper we investigate the educational impact on blind users of listening to structured podcasts, versus a time-based division method. We have defined a set of tasks to perform on a collection of podcasts and a questionnaire for visually impaired users to fill out. We discuss and analyze results collected through this designed and implemented Web user test. After discussing related work, we briefly present the Web-based Text-To-Speech (TTS) system [8] for automatic generation of a set of structured audio podcasts, a description of the user test, analysis of collected data, and some concluding remarks along with indications for future work.

2 Related works

Many studies confirm that podcasting is increasingly useful for exchanging information and making learning approaches more efficient in terms of resource consumption. For many people, listening may be more attractive and less tedious than reading [9, 10]. Podcasts are often used for additional support in teaching and learning. Some common uses in higher education are: taped lectures, guest speakers, group presentations, tutorials, exam reviews, reinforcement of key concepts, and drill or repetition [10].

Recently, to facilitate the preparation of audio materials, several tools have been proposed and developed that transform a text document into a spoken version by using text-to-speech (TTS) technology and a voice synthesizer. A TTS system converts a text into speech. The output generated can be heard immediately while the audio is being produced or can automatically be recorded in audio files. Typically, these audio files are in mp3 format and can be listened to on an mp3 portable player or on smartphones. Tools like Robobrain and vozMe are examples of this converting process. RoboBraille is an email-based service which automates translation of text documents into Braille and speech. Users submit documents (e.g., text files, Word documents, HTML pages) as email attachments. The translated results are returned to the user via email [12]. VozMe (<http://vozme.com/>) is an easy-to-use online service for creating mp3s from text. It only requires typing in or pasting the text and pressing a button. However, audio content created is not structured (is a single file) so it is not very suitable for studying.

Some software such as Natural Reader [13] allows one to jump to the previous or next section. In [9] authors combine sophisticated speech-processing techniques (including audio-based skimming) to create a multimedia player that allows audio-

centered browsing and navigation of recorded presentations. However, these systems require interaction with a PC and thus are not suitable for Mobile Learning.

3 Structured Audio Podcast System

Structured audio podcasts are useful for all kinds of people and are especially appreciated by blind users, who can easily skip from one section to the previous or the next (on an mp3 player by using backward and forward command keys).

Our hypothesis is that for blind users, structured audio content facilitates navigation within the content, reducing frustration when searching for the desired content when learning and memorizing specific parts. Conversely, searching for the desired content in a single long sequential audio file can be quite difficult.

The system we developed is web-based [8]: blind users can interact (via screen reader) with the Web interfaces thanks to the accessible features obtained by applying appropriate design criteria of the Web Content Accessible Guidelines (WCAG 2.0) [14] as well as WAI-ARIA (Accessible Rich Internet Applications) [15] attributes. Our Web system receives a document in input (.doc, .rtf, .txt), and in output provides a set of audio files that reflects the document's internal structure (one file for each document section). Figure 1 shows the architecture of the system: a document is uploaded through a Web page and saved on the server. The system identifies the section titles so that the document can be split and processed by a TTS module to be transformed in an audio streaming, coded and saved as an mp3 file. Each generated audio podcast represents a section of the document, and it contains the title of the section at the beginning in the first few seconds of listening.

In addition, the system transforms each title of sections to audio streaming i.e. talk files (<http://www.rockbox.org/>). With .talk files, an mp3 player equipped with the Rockbox firmware makes files and folders directly accessible through audio.

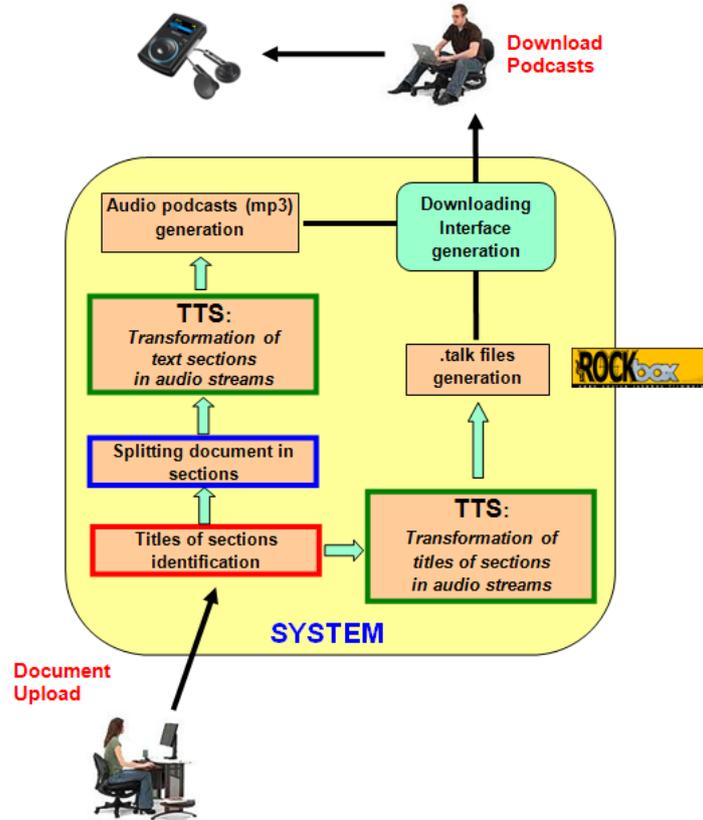


Fig. 1. Architecture of the web-based system for generating structured audio podcasts

Analysis of font properties allows detection of important information highlighted by means of bold or italic font properties by the author, emphasizing them in the generated audio podcasts (for example words in bold are delivered between two short tones in the audio podcast).

The system is also able to detect tables that are converted into single podcasts, since listening to a large table without seeing it can be extremely difficult and frustrating.

After generating structured podcasts, the system shows their list on a web user interface. A list of downloadable podcasts is a sort of index of document sections. Users can choose whether to download the entire package on their mp3 players, or select only a part: this enables better orientation during a quick search for specific information within a podcast, improving the learning process.

4 User Test Design

This study aims to evaluate the real impact of structured podcasts on learning and on the “going over” and memorizing phase for blind users. The memorizing phase is critical for blind users, especially for quickly retrieving specific information. To evaluate our proposal, we set up a user test involving a group of blind users.

4.1 Target

To involve potential users in our test, a message was sent to a collection of email addresses obtained thanks to collaboration with the Italian Association for the Blind.

Eighteen people answered; 14 are blind and 4 are visually impaired. The sample was composed of 8 women and 10 men, with age ranging from 18 to 74 years. All participants use assistive technology in the Windows environment: the blind use JAWS for Windows¹ screen reader, and the visually impaired use both the screen reader (to rest their eyes) and the magnifier (to enlarge screen content).

Regarding technical skill, 12 users have an intermediate degree of knowledge of the Internet, while 3 users were beginners and 3 users were experts. All of them used an mp3 player device for listening to audio books, music, educational audio files, or downloaded audio podcasts from Internet. In particular, 14 participants used an mp3 player supporting Rockbox firmware for vocal navigation on the device.

4.2 Method

A 12-page document on an e-Commerce topic was created, containing useful information for sightless users. The document entitled “Tips for on-line shopping” was organized into three parts: Psychology of e-commerce web sites, Online Security and Payment methods.

Three audio versions of the document were created to be used in three different phases:

1. A single podcast of duration 28 min and 24 s (first phase).
2. A set of 5-min podcasts (6 mp3 files) where the document broke off after 5 min (when the paragraph ended) A set of structured podcasts (21 mp3 audio files), each one representing a logical section of the document, created using our system [8].

The set of 5-min podcasts was generated using Dspeech (<http://dimio.altervista.org/eng/>), and the set of structured podcasts was created using our system (described in the previous section). Both sets of podcasts were generated using the same SAPI 5 voice, in order to avoid some differences in user perception during listening. Each set of podcasts was grouped in a directory: time-organization (containing 6 files) and content-division (with 21 file audio).

The two sets of podcasts also included a playlist .m3u file, useful for listening to the podcasts using a computer, because the user can skip from one audio track to the

¹ <http://www.freedomscientific.com/products/fs/jaws-product-page.asp>

next or to the previous one as on an mp3 player, in a sequence order as listed in the .m3u file.

Both folders are compressed in a self-extracting file directly downloadable from the user test pages. The folder with the files can be uploaded on an mp3 player device (or listened to on a computer). The audio files are necessary for performing a set of tasks, which mainly consist of answering questions (carrying out seeking tasks).

The user test is organized in three phases. The first phase, for all users, requires choosing any username (e.g. a nickname), to login, and to download the podcast corresponding to the entire document to be listened to.

The second phase requires downloading the **same** document split in different tracks according to the time-based division method (for group A) or the structure-based division method (for group B), which the participant had to use to carry out the first three tasks. The tasks consist in answering three questions, seeking the answer in the folder containing the set of downloaded podcasts (different versions for the two groups).

Analogously, the third phase requires downloading the **same** document split into different tracks according to the structure-based (Group A) or the time-based (Group B) algorithm, in order to make each group listen to both versions, and to execute three other simple seeking tasks in order to answer three questions.

The sequence of time and structure-based tasks (and vice versa) are alternately balanced between participants in order to minimize the learning effect of the structured audio contents. Each group has six questions – i.e. tasks – to complete: three for a set of time-based podcasts, three for the structured ones. The user is alternatively assigned to the group A or B at the first login.

Each of the six questions has the objective of finding the answer within one specific podcast of the set, and the questions do not follow the same sequential order of sections of the original document. This is to force the user to spend time navigating and retrieving the required information.

4.3 Remote User Test

The user test, accessible remotely via Web, has been developed using PHP. Web pages of the user test conform the Web Content Accessibility Guidelines (WCAG 2.0) [14] and they have been developed using WAI-ARIA suite [15] criteria to facilitate navigation and interaction for blind and visually impaired users using a screen reader or other assistive technology.

The Web pages of the user test were fully accessible with the most recent version of Firefox while not completely with Internet Explorer. The remote system leads each user to perform the three phases. For each task the starting and ending time was captured and recorded in order to obtain information on how long each task took to find the information. Results evaluate the correctness of the answers to the six questions (three for a set of time-based podcasts, and three for the structured ones), the time for answering, and the answers of the post-test questionnaire.

The time it took to answer any of the six questions was calculated from the time the user uploads the page containing the question until the “Go to the next question” button has been pressed; if the user decided to complete the task at a different time

(s)he pressed the “Pause the test” push button and restarted the test later at the point where (s)he left off. Analysis of the test results is described in Section 5.

4.4 Questionnaire

At the end of the three phases, each user was asked to fill out a questionnaire on personal information (sex, age), computer skills, conditions of test executions (browser, screen reader, etc.), user preferences and a subjective evaluation of both the division methods.

5 Results

As previously mentioned (Section 4), the sequence of time- and structure-based tasks (and vice versa) was alternately balanced between participants. In fact, a pool of users had to look for the answer to a question within the time-based set of audio podcasts, while another pool of users had to look for the same question inside the structure-based set of audio podcasts. We considered task accomplishment and time required for the completed tasks (i.e., for the correct answers).

By analyzing the log files, we observed that not all users completed all the tasks successfully. Probably due to the length of the test, some gave an answer after trying to recall information from memory, resulting in a wrong or approximate answer. If the answer was incorrect we discarded it this time. The number of correct answers (successful seeking tasks) in the two versions was similar (7% and 9%).

Considering the total average time for successfully completing each task, we observed that using the set of structured podcasts led to shorter times for four seeking tasks (T1, T2, T4 and T6), compared to the use of the time-based set, and vice versa for the other two tasks. However, the factors that favored time-based organization are: 1) the position of the information sought is at the end of the track in the structured podcast while it is at the beginning in the time-based file and 2) the title did not indicate clearly which track to explore. In these cases the exploration tasks were nearly equivalent. Instead, where the title clearly suggested where to look for the answer, the structure-based organization proved more useful than the time-based one.

The total average time spent completing the six tasks with the structured organization was 24% less than the total average time for the time-based organization.

However, a possible factor contributing to the time differences (and some users confirm this hypothesis) involves the opportunity for users to answer a question from memory, after reading the single mp3 version -- thus executing the task without searching for the answer in the podcasts and so taking less time. To summarize, after analyzing all times spent carrying out the assigned tasks, we can conclude that there is a time savings of about one quarter (24%) by using the structured-based version. This result is important, considering that when the user reads again (i.e. listens over and over to) a podcast he/she is able to identify easily and immediately where to find information, thanks to structured organization using the logical structure of the document, and it would be easy to remember.

Concerning user satisfaction and preferences, the feedback of participants is clear, concordant and unambiguous: all participants prefer listening to small audio podcasts instead of a single long audio file, and they prefer structured podcasts to the set composed of 5-min podcasts, for many reasons:

- a) it takes less time to retrieve specific information
- b) titles of sections simplify seeking specific information
- c) shorter average duration of each podcast
- d) more intuitive and practical search
- e) focus and perception of more details
- f) simplifies memories and information access
- g) a structure-based division method maintains the entire audio flow of a section logically, while the time-based division method can interrupt it abruptly.

All this positive feedback encourages the adoption of the structure-based method to produce more usable podcasts that might better satisfy user needs.

6 Conclusions

In this paper we present a study to verify the impact of structured and time-based podcasts on learning for blind users. The study was conducted with 18 users (14 blind and 4 visually impaired) via a user test to compare the two division methods: structured (according to document structure) and time-based (fixed duration) versions. Analyzing the collected data, all users greatly appreciated having a set of podcasts instead of a single long file, and they expressed a preference for structure-based podcasts over the time-based (5-min) division method.

Considering the time it took to complete the proposed tasks, structured podcasts seem to support the search for information, decreasing the time needed to search (at least in 4 questions out of 6). Results highlighted a time savings of about 24% obtained using the structured-based sequence.

Results regarding time were not totally in accord with our expectations for greater time-saving, possibly due to several factors: external factors may have influenced the test, increasing the overall time (due to remote testing that did not allow precise monitoring of user activities); difficulty and position of the answer inside the content; and the test was long so some users may have preferred to answer by trying to remember what they heard before, without actually searching for the answers in the podcasts.

All participants expressed their preference for structured podcasts since they required less time to retrieve specific information, titles of document sections provided a clue for seeking specific information, there was shorter average duration of each podcast, and overall they offered a more intuitive and practical search. In addition, reflecting the logical structure of the document, a structured podcast allows the focus and perception of more details, simplifies memories and information access, and maintains the entire audio flow of a section logically, while the time-based division method can interrupt it abruptly.

In conclusion, the positive feedback of all participants encourages the adoption of a structure-based method to produce more usable podcasts that better satisfy user needs by decreasing the time required to explore and seek information.

In future work we aim to refine the user test and repeat it with more blind people, but also with sighted users in order to compare results of the two samples. Moreover, we are planning a new user test with blind users, aiming to evaluate the web system we have developed to create structured podcasts, to verify its usability, the impact of structured podcasts on learning, and the delivery of special information (such as bold, italic, tables, etc.) contained in the original document.

References

1. Mermelstein, B., and Tal E.: Using Cellular Phones in Higher Education, in *Wireless and Mobile Technologies in Education (WMTE 2005)*, 2005.
2. Aldrich, D., Bell, B., and Batzel, T. Automated Podcasting Solution Expands the Boundaries of the Classroom. in *Proceedings of the 34th annual ACM SIGUCCS conference on User services*, 1-4, 2006.
3. Lauer, T., and Wolfgang, H. Audio-based Methods for Navigating and Browsing Educational Multimedia Documents. in *Proceedings of the inter. workshop on Ed. Mult. and mult. ed. Poster and demo session*, 123-124, 2007. DOI=<http://portal.acm.org/citation.cfm?id=1290166>.
4. Pitt, I., and Edwards, A. *Design of Speech Based Devices. A Practical Guide*. Springer, London, UK, 2003.
5. Wolff T. B.: Podcasting Made Simple, in *Proceedings of the 34th annual ACM SIGUCCS conference on User services*, pages: 413 – 418, 2006.
6. Deibel, K. Course experiences of computing students with disabilities: four case studies. in *Proceedings of the 39th SIGCSE technical symposium on Computer science education*, 454-458, 2008.
7. Allen, J., Sharon Hunnicutt, M., and Klatt, D. *From Text to Speech: The MITalk system*. Cambridge University Press: 1987. ISBN 0521306418.
8. Leporini B., Buzzi M. C., Buzzi M., and Mori G. Automatically Structuring Text for Audio Learning. in the *Proc. of HCI International 2009*. in Springer LNCS. Vol. 5616/2009 *Universal Access in Human-Computer Interaction. Applications and Services*, 73-82, 2009.
9. Cebeci, Z., and Tekdal, M. Using Podcast as Audio Learning Objects. *Interd. Journal of Knowledge and Learning Objects*. 2, 2006.
10. Ormond, P. R. Podcasting enhances learning. *Journal of Computing Science in Learning* 24, 1 (2008), pp. 232-238.
11. Campbell, G.: There's Something in the Air: Podcasting in Education. *EDUCAUSE Review*, vol. 40, no. 6 (November/December 2005); <http://connect.educause.edu/Library/EDUCAUSE+Review/TheresSomethingintheAirPo/40587>.
12. The RoboBraille Consortium. The RoboBraille email service, <http://www.robobraille.org>
13. NaturalSoft Text-to-Speech, <http://www.naturalreaders.com/index.htm>
14. W3C. Web Content Accessibility Guidelines 2.0. <http://www.w3.org/TR/WCAG20/>, 11 Dec 2008.
15. W3C. WAI-ARIA Technical Specification, <http://www.w3.org/TR/wai-aria/>