

Digital Talking Books - a Report from a Practical, Ongoing Project

Kjell Hansson, Lars Sönnebo* and Jan Lindholm*

The Swedish library for Braille and talking books (TPB)

* Labyrinten Data AB

Summary: This project aims to develop a new system for storing and retrieving talking book material using digital technology. The project shows how radically new means of information access can be put to work using a digital talking book system based on standard personal computer technology.

1 Introduction

In 1992, TPB was granted funds from the Swedish Ministry of Education to investigate how new technology could be used to give students new and better means to access talking book information.

A three-year project was initiated to evaluate the possibilities of digital technology for talking books. The first year was to be spent on getting general and unbiased information about suitable technology. During the second year, the project was to enter a more pragmatic stage, intending to actually create practical results, that could serve as a basis for demonstration and real-world testing.

This report describes some of the results from the project.

The project focused on human voice as the main information carrier for the talking book, rather than speech synthesised from machine-readable text by a text-to-speech synthesiser. Other projects at TPB have been investigating the latter technology from the student's point of view.

1.1 The Preliminary Survey

The first year of the project aimed at taking a close look at which technology was available, and what could be achieved by using it. The result was a survey of both methods for processing and storing voice data as well as suitable media for digital talking books.

The report is available to all interested. A full version is available in Swedish, and an abridged version is available in English.

1.2 Some Starting Points

After the preliminary study, it was decided that the project in its second year should concentrate upon some of the most interesting results from the preliminary survey, and from these build a prototype digital talking book system, suitable for demonstration of access principles, voice coding techniques etc. The system could also be used as a basis for practical testing at a later stage.

Before starting the second year of the project, some basic goals and standpoints were stated:

- When choosing a storage/distribution media for a digital talking book system, it is desirable to use one that is widely spread and used for other purposes, well standardised and one that will be available for a long time. CD-ROM was chosen as media for the project
- The system should be based on a standard PC, to allow the student to use the computer for other tasks, e.g. word processing, information retrieval and reading with synthetic speech etc. The PC would also be a good platform for the software development and for testing of necessary peripheral equipment.
- Standard computer technology was to be used as much as possible. We therefore decided to use the platform offered by the Multimedia Extension under Microsoft Windows, running on an IBM-compatible PC.
- The base for the system should be a general data structure that allowed for easy future expansion with new data types, new data formats etc. The format specification should be made open for other talking book producers and system developers to make their own implementation. This was seen as important, since it would be highly beneficial if a commonly adopted standard could be established amongst libraries, even in different countries. Such a common standard would make interlending services possible.
- The system should be capable of storing vast amounts of material, so that long recordings should be possible to store on one media unit. The output quality of the voice recording should be good, comparable with the quality offered by standard audio cassette systems.
- The system should be expandable in the future in terms of storage media, distribution methods, data format for voice recording etc.
- The system should give the user fast, random access to the book's material. The material should be structured in much the same way as a normal printed book, to allow the user to read the talking book in a similar way to a normal, printed book. The response of the system should be fast, without long delays before they desired material was played back, no matter which part of the material was requested.

2 Our PC based Talking Book System

The following is a brief description of the PC-based talking book system that has been developed during the second year of TPB's talking book project.

2.1 Some unique aspects of the system

The uniqueness of the system being developed in the project can be summarised by the following properties:

- Using digitised human speech as the information carrier gives much better voice quality than in systems using synthetic speech.
- Voice data is treated in comparatively small units, allowing common database-type techniques to be used for storage and retrieval of data. Furthermore, the data can be organised in a structural and hierarchical manner, resembling the original printed material. This renders random access to any part of the material, from the section (chapter, part etc.) down to the sentence (phrase) level.
- Since the system is not restricted to a particular kind of storage media or device, technology advancements can easily be integrated in future versions.
- The system uses standard computer technology as its hardware base, which will help keeping costs for playback equipment down, as well as simplifying system development. The use of a standard personal computer as the core of the system will allow for integration with other computer applications.
- The system is open, constructed upon a foundation of common computer standards and methods, which will be published, so that other developers may implement their own solutions.

2.2 General System Description

The system uses digitised human voice recordings, stored in a compressed (coded) data format on CD-ROM media. CD-R technology is used as the main production system, since a library such as TPB mainly makes a very small number of copies of each talking book. Nothing hinders mass-production of talking books from a master, though.

The data representing the voice recordings are divided into small data blocks called "phrases", using an automatic pause-detection system in the recording software. The phrases are organised into "section", which in turn are organised into a hierarchical structure, resembling the printed book's table of contents. The material can be accessed from software as a database-like structure.

The voice data is stored in compressed format on the CD, allowing a standard CD-ROM disc to hold 15-50 hours of continuous speech, depending on voice coding

technique and required output voice quality. The system allows for new compression/decompression technologies to be added as they become available, allowing higher voice quality and/or more hours of recording on each CD-ROM. Our system currently uses ADPCM-type coding and decoding.

The user reads the book with a specialised software, currently running under Microsoft Windows. The reading software will have a fully talking interface, using digitised voice.

The reading system consists of a standard IBM-compatible PC (a 486 SX or better), equipped with a standard CD-ROM XA and photo CD-compatible drive, a 16 bit audio card and the playback software. The PC can be complemented with a speech synthesiser and screen reading software, to give the user access to other Windows applications.

Beside the audio data, the data structure also allows for the book's original text (and, if desired, its illustrations) to be placed on the media in machine-readable form, sharing the same structure as the voice recording. Thus, the text could be viewed on screen or accessed via a speech synthesiser if desired.

2.3 User Interface Functionality

While designing the first version of the user interface for the reader, the goal has been to give the user access to the recorded material in a way similar to a sighted user, reading a normal printed book. The underlying data structure was created with this in mind, and the user interface is therefore mainly a continuation of the book's structure.

The reading is controlled from a hierarchical structure, resembling the original book's table of contents (ToC). Each item in the talking book's ToC is a "section", made up of a collection of "phrases" (voice data blocks). A section has a certain level in the ToC's hierarchy, which determines when and in what order its phrases are read back.

To select which part of the book to read, the user simply selects a section from the Book's ToC. As section headings are selected, the section's title is read back, using the appropriate phrases from the recording. The titles of the sections in the ToC are also shown on screen for sighted or partially sighted users.

After selecting the starting point, playback can be started at the desired position by a play command. The user can also jump forward or backward in a section by phrase, by phrase group (as defined by the producer) or by page. The latter relies on page breaks which have been placed in the section by the producer, based upon the page layout of the printed book.

The book can be read in continuous mode, section by section or it may be read phrase by phrase, page by page etc. Skipping between phrases, groups etc. within a section is a quick and easy process, giving the reader convenient "skimming" abilities.

Pauses made in the recorded material are interpreted in the recording process and stored as tokens rather than voice data of zero amplitude. This saves storage space, but more importantly, it gives the reader /user control over pause length, thereby controlling an important aspect of the reading. Together with the fast random access to the material even on low level (phrases), the system gives the user a very quick retrieval system, especially compared with sequential access as on a standard magnetic tape recording. To allow for even faster reading, the system offers "intelligent time compression" (ITC), which gives higher playback speech rates compared to the original recording.

In addition, the data structure makes it easy to implement special features in a PC-based system such as user notes, marking of phrases for later retrieval, bookmark placement etc. Such data is stored as a "user book" on the PC's local hard disc.

The playback voice quality is good, comparable with results from ordinary compact cassette recordings. Such a good quality can be retained with the current data compression technique, even for books of 20-25 hours in length. Longer recordings can be stored on a single CD-ROM with reduced voice quality, using the current compression technology. As an alternative, extremely large talking books can be split up onto two or more CD's, where each CD-ROM will be treated as a separate book.

2.4 The Producer's Perspective

The data structure allows for recording of data in arbitrary order. Recording is done phrase-by-phrase, where each phrase represents a sentence or paragraph in the original, printed material. Phrases can be corrected, inserted, deleted or moved. The producer can also group related phrases together and insert page breaks between phrases, following the printed book's layout, thus simplifying the user's access even to highly complex material.

The silence-detection system relieves the recorder from having to keep in time with a moving tape. There is no need to manually enter record mode - the reader just starts talking. This recording process should mean less stress and also make editing much easier than on tape.

The recorded material is stored on a large hard disk, and can be transferred to MO (Magneto-Optical) or tape media for backup or for transfer to another system, e.g. a system for proof-reading.

After recording all material for a section, a section file is created. All the section files making up a talking book are organised on their highest level by a "book file". The book file and section files can be transferred to a CD-ROM image using CD-R equipment or to a master for production in larger scale.

The multi-session capability of today's CD-R devices allows for more than one book on a CD-ROM, which can be recorded one at a time.

The intention behind the recording tools that have been, or will be developed by our project, is that the person who is the recorder should also be able to edit the material as well as produce the distribution media.

A streamlined production system can give very cost-effective talking book production, compared to the methods commonly used at present.

3 Project Status Autumn 1994

A prototype PC-based talking book system was finished in early May 1994. The system was capable of demonstrating the general principles for data structure and user access. Basic recording tools were created to create the talking book material for the demonstration. A CD-ROM disc containing demonstration material, taken from several publications was accompanying the system.

The original intention was that the project should be reported and finished in its third year. Since the project has been very successful and have come up with remarkable and practical results, we are now trying to get founding to prolong the project. The third year will then be spent on turning the prototype system into a fully working system.

More work is needed to achieve a stable and fully working system. For example, on the recording side, streamlined tools are needed and the user interface for the reader needs to be proven and developed according to user's needs and demands.

The prototype system set-up will be copied and distributed to several students for field testing, probably during the autumn of 1994. A set of CD-ROM titles will be created, containing appropriate literature for the testers.

The continued project will aim to create a fully working complete talking book system, ready for mass production. To do this, we will work together with the test users as well as some of today's producers of normal talking books. Their demands and opinions will be collected and the result will be implemented in the system. It is estimated that by the end of 1995, a version 1.0 of the system will be presented.

New voice coding/decoding techniques will be tried for the system in the future, giving better voice quality and/or higher compression rate than the algorithms currently used.

During the next stage of the project, we will also try to interest other talking book producers in Sweden and other countries in the digital talking book format that has been developed in the project. Hopefully, this format can form a common base for future production systems, so that talking books can be interchanged between different libraries.

4 Future Plans

4.1 Standardisation

It is highly desirable that digital talking books can be shared amongst different libraries, services and production systems in various countries. The project will therefore make serious efforts to promote the general data structure to all who are interested in creating a digital talking book system. The data format will be open, so that different producers can implement their own tools. Alternatively, our software tools could be made available to other libraries.

The data structure of our system is advanced enough to cope with the most complex publications, but may equally well be used also for material of a simpler structure, for example fiction literature. In those cases, the material could simply be stored in a small number of sections, representing the parts or the chapters in the book. Thus, our data structure will be usable as a standard for all kinds of digital talking book material.

4.2 Building Purpose-built Equipment

The current system is based around a personal computer, since the computer gives many advantages to the student. The core of the system may however be implemented in a specialised computer-system, resembling a stand-alone unit, much like a normal compact disc player. Such a machine could still give the user many of the features found in a PC-based system, of course without features like making notes in relation to the recording etc. With such a device, the digital talking book standard could be used by today's average talking book reader, and so the positive aspects of digital talking book technology could be spread to a wider range of readers.

Since production of talking books with the aid of digital technology seems likely to be rather cost-effective, there is even a chance that talking books may be a competitive distribution media for information, even aimed at sighted users. If a suitable talking book playing device can be made available at a competitive price, today's producers of commercial talking books might be interested in the new technology created by this project.