Software System for Vocal Rendering of Printed Documents

Marian DÂRDALĂ
Academy of Economic Studies, Bucharest
dardala@ase.ro

The objective of this paper is to present a software system architecture developed to render the printed documents in a vocal form. On the other hand, in the paper are described the software solutions that exist as software components and are necessary for documents processing as well as for multimedia device controlling used by the system. The usefulness of this system is for people with visual disabilities that can access the contents of documents without that they be printed in Braille system or to exist in an audio form.
***Keywords**: accessibility, TWAIN, OCR, TTS, SAPI.*

Introduction

With the development of society and informatics technologies, the access to information for people with disabilities exponentially increased. The classic way to print accessible documents, I mean in the Braille form was the first step to make the content of documents accessible for the people with visual disabilities. In this form the information is received using the tactile feeling. This step was then completed with the existence of information on magnetic media (tape or disc), their perception making in an audio form.

The major disadvantages of these forms of accessibility consist in: high costs of transcription, obtaining a large result by printing documents in Braille form because the characters are represented in three-dimensional form (their also have the high) and the access to information stored on magnetic tapes is possible only in a sequential way. All these disadvantages are that the information represented in these alternative forms are mostly general interest, thus the access to a variety of information is limited for people with visual disabilities.

In the new type of society, I mean the information society, that assume an extensive usage of computers, both in institutions and in private life, the access to a wide variety of information can be done through the computer. The development and diversification of peripherals connected to a computing system and usage of multiple media communication have enabled the implementation of the multimodal interfaces. Through these alternative interfaces, computers can be used by people who have different types of disabilities. Development of the computer networks and the emergence of Web sites have allowed the publication of a large and diverse amount of information in the Internet, accessible to the users through computers.

Taking into account the technical progress mentioned above is obviously that the electronic documents do not replace the classic documents, printed on paper. In this form we can find books, magazines, newspapers, bills, notices and so on. Information that exist and that are distributed in the printed form can be accessible to persons with visual disabilities by using the specialized software systems and peripheral that allow to the users to convert the way to represent the information from the visual one to an audio one.

1. System software architecture for documents vocal rendering

To build a software system for documents vocal rendering is appropriate for both hardware devices and software components. In terms of hardware, a system for documents vocal rendering has to have the following peripheral devices:

- Scanner for taking over in the computing system the information printed on paper; and
- Sound card to vocal render the content of documents and possibly to receiving commands in a vocal form from the microphone.

In terms of software are needed components to control peripherals, a voice synthesis en-

gine (TTS - Text to Speech) and a component for document conversion from the bitmap image format, (the result of the scanning process) in text editable format. This process is known as Optical Character Recognition (OCR). The scheme of such a system is shown in figure 1.

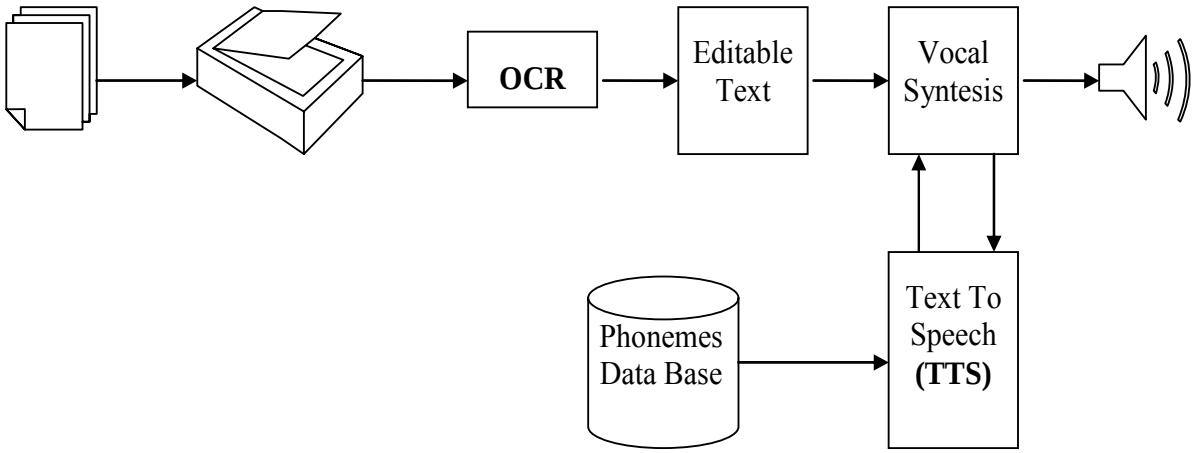

Fig.1. The software system architecture

Such a system should be built to have a great flexibility concerning the connection to specialized software components to perform the characters recognition and to make the vocal synthesis. Needed of adapting the software system to the software components appear because of this system is depending of the cultural elements as: symbols of the alphabet of a language, and on the other hand the language phonemes set. The efforts to develop such a software system have to be focused on building interfaces. The interfaces have to be able to connect the various specialized software components that are made by different software companies.

2. Scanning device control

Developers of specialized devices to purchase of fixed images such as scanners, digital cameras as well as software developers who use such devices have felt the need to standardize the communication between devices and applications.

TWAIN was developed as a standard protocol for communication and programming interface between devices and application programs. The main entities included in an application that uses TWAIN are presented in Figure 2. From the figure 2 we can notice that TWAIN realizes the communication between application programs and peripherals for fixed images achieved.

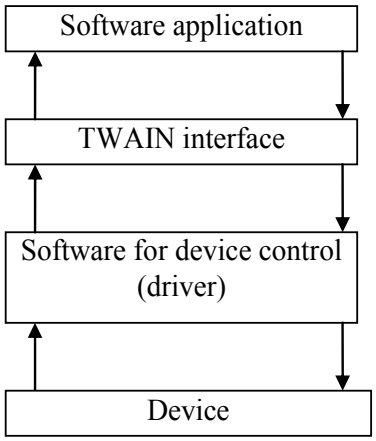

Fig.2. The role of TWAIN interface in communication between application and device

Using the TWAIN interface has multiple advantages. From the viewpoint of the software developers it ensures the independence between application programs and various types of peripherals specializing in the acquisition of fixed images. It is well known that there are many producers firms of such devices (for instance: Canon, Epson, HP, so on). In these circumstances applications are written independently of the particular devic-

es just because of this interface. TWAIN allows to the owners of devices to build their own interface to control peripheral no longer required building a custom interface in the user applications.
An application that uses such devices should have in its interface at least two options:
• Selecting the source that allows to the user to choose the device through which to acquire the images;
• Image acquisition process that determines starting the image transfer from the device to the application.
In Windows operating system the TWAIN interface application programming is available through dynamic link libraries called twain32.dll.

3. Optical character recognition process

The object oriented model developed under the MODI name (Microsoft Office Document Imaging) allows the software development that is capable to process documents that exist in an image representation to recognition within the meaning of texts contained in them. The MODI programming interface is available since Office 2003 software package and the references that have to be included in the project are in the Microsoft Office Document Imaging 11.0 Type Library.
The module contains an object hierarchy specialized in processing and displaying images, the most important ones in terms of the functionality are:
- *Document* object used to manage an orderly collection of pages (images);
- *Image* object used to process a single page from the document;
- *Layout* object used to attaching the result obtained by characters recognizing process to a single page from the document;
- *MIDocSearch* object used to provide searching functionality to the document;
- *MIDocViewer* object is an ActiveX control used to display pages from the document.
The simplest sequence to obtain the text from an image by optical characters recognition process using the MODI model is:

```
MODI.Document doc = new Document();
doc.Create(@"exemplu.tif");
doc.OCR(MODI.MiLANGUAGES.miLANG_ENGLISH,false,false);
doc.Save();
MODI.Image img = (MODI.Image)doc.Images[0];
obtb.Text = img.Layout.Text;
```

where, *obtb* is an object of the *TextBox* type that has the multiple lines property activated. From the code sequence, we can notice that the OCR method realize the characters recognition in correlation with a parameter that denote language information.

4. Vocal synthesis

Transforming text to speech is very useful for people with visual disabilities. Vocal synthesis is realized by the components known as TTS (Text To Speech) Engine. In the world there are many software products of TTS type. Their main disadvantage is that these systems are developed in correlation with a particular language that is used to playing the text. Because the Romanian language has a small widespread area, the interest for the development of the Romanian voice synthesis engine adapted for it is quite small.
Phonemes of the Romanian language were included in a few TTS engines as: MBROLA, IVO Software that created the Carmen voice and Baum Engineering that created *Ancutza* voice.
Speech can be done by using several technologies, the most important ones are:
- By concatenating the pre-recorded phonemes; and
- Acoustic models.

The words in visual form are reflected by characters while in audio form are reflected by sounds. The atomic sounds that form a word are identified by phonemes. It's easy to understand this thing because some languages, for instance English language uses two ways to present a word: by characters and by phonemes. In this case a phoneme is also represented by a graphical symbol but expressed an atomic sound, useful to correctly pronounce the word.
Speech by phonemes concatenation means that the word is built by linking the elementary sounds that form it then it is played. The elementary sounds (phonemes) are stored in

a phonemes database. It also stores for each phoneme a graphic symbol (one or more characters) to visual identification of the atomic sound. This technique to transform the text to speech is advantageous because of there are relatively few phonemes in a Latin language. But to get a synthesized sound as close to natural language must be taken into account and sounds of transition called diphonemes. A diphoneme consists of two phonemes and these sounds are in a very large number in the language.
Speech based on acoustic models does not use samples of human speech to generate the voice signal, but the voice synthesis using acoustic models. Thus, this technique generates sounds similar to those produced by vocal cords and by applying filters it is possible to obtain different intonation.
Vocal signal produced in this way is artificially in comparison with speech obtained by linking the prerecorded phonemes. Among the advantages of this method, most important are:
- The vocal synthesis module occupies less space because it does not use a database to store the phonemes;
- Voice obtained by signal noise models is more artificial but it is more comprehensible to the high speed playback.
Speed playback voice is an important parameter of a voice synthesizer especially when it is used by people with visually impairment because the screen readers provide much redundant information to the user. This parameter can determine the speed of the interaction between the user who has visual disabilities and the computer.
Under the Windows operation system, the programming interface that enables the development of applications that perform the conversion between text and speech is called SAPI (Speech Application Programming Interface).
The SAPI model is divided into two distinct levels:
- The SAPI high level - this level of services offers voice basic commands in the form of voice recognition and simple text-to-speech outputs;
- The SAPI of low level - that provides access to detailed voice services, including direct interfaces to control human behavior for handling voice recognition and text conversion to the voice signal.
Each level of service SAPI has its own set of objects and methods.

Conclusions

The software products developed to achieve accessibility by converting the contents of the documents thus it to be render by vocal signal are particularly useful for people with disabilities who can not use the classical computer interfaces. These systems must be designed in an open system to allow the location interfaces based on specific regional language used in communication.

References

1. Dârdală, M., Vetrici, M., Ioniţă, C., *Creşterea gradului de accesibilitate al dispozitivelor mobile prin dezvoltarea de componente software*, Volumul de lucrări al celei de-a patra Conferinţe Naţionale de Interacţiune Om-Calculator, Rochi 2007, Universitatea Ovidius Constanţa, 20-21 Septembrie 2007, Editura MatrixRom, Bucureşti;
2. Moulton, G., *Accessible Technology in Today's Business: Case Studies for Success*, Microsoft Press, 2002;
3. Reveiu, A., Dârdală, M., *Designing of accessible software for economic and social inclusion*, The Proceedings of the 6th Biennial International Symosium, SIMPEC 2006, Braşov;
4. Shami, M., *Automatic Clasification of Emotions in Speech Using Multi-corpora Approaches*, The Second annual IEEE BENELUX / DSP Valley Signal Processing Symposum, Antwerp, Belgium, 2006;
5. ***, Accessibility in Microsoft Products http://www.microsoft.com/enable/products/;
6. ***, Microsoft Developer Network Library, Microsoft Press, 2007;
7. *http://www.twain.org/*