MULTIMEDIA SYSTEMS FOR BLIND AND VISUALLY IMPAIRED PERSONS

Andreja Samčović,
Svetlana Čičević

University of Belgrade - Faculty of
Transport and Traffic Engineering,
Serbia

Abstract:

This paper presents the prospective use of multimedia applications for blind and visually impaired persons. Multimedia is gaining a considerable importance and is exploiting the available technologies to allow opportunities for blind people in improving their Quality of Life. Following the description of some multimedia solutions and applications, key elements of systems that help blind persons are described.

Keywords:

multimedia, applications, blind, text, speech, audio.

1. INTRODUCTION

Multimedia unstoppably conquers and changes irreversibly more and more spheres of everyday life. Although numerous multimedia applications only marginally affect the lives of the majority, the full potential of technology has not yet been realized, but has been manifested only by using specialized applications. The blind are certainly part of the population that will benefit with the development of multimedia. Multimedia enriches all areas of life of the blind. Significant improvements have been achieved in their education and employment, with a focus on using computers [1]. Shifts in the areas of mobility and integration into society have also been achieved. The blind have inexorably become more active and independent participants in society. Multimedia technology allows multimedia content to be connected, downloaded, shared and interpreted. The principle of creating content intended for all break down barriers between blind children and their peers, ensuring their integration. A significant share of problems relating to the education of the blind represents access to textbooks and literature, in general. Multimedia applications facilitate the use of literature in electronic form (e-books) in every aspect of the creation, production, distribution, search, usage and presentation [2]. In fact, methods of copyright management are particularly important because they facilitate contracting and access to new multimedia materials, which are often in the form of e-books.

Tiflo-technical multimedia aids increased independence of the blind [3, 4]. Traditional aids, unlike multimedia, are often cumbersome, tedious and poorly acceptable in a social environment. Although the new multimedia

Correspondence:

Andreja Samčović

e-mail:
andrej@sf.bg.ac.rs

technology has advantages, users are still required to make adjustments. With an increasing number of devices, which value is undisputed, it is important to enable their use, comparison, evaluation of the usefulness, and adoption for the blind. Multimedia applications should become part of the everyday life of the blind. Their education, employment, access to information and integration into the information society, depend on the successful application of computers and multimedia technology. The quality of life enabled by the use of multimedia technology, justifies any effort in the popularization of multimedia [5]. Upgrading properties are also an important aspect of multimedia. Audio and visual facilities act synergistically for non-disabled, so it might seem that the blind are handicapped. The disadvantage can be removed by using the sound-descriptive content, wherein the description of the video content is compensated. A description may be auditory or tactile. Since it is difficult to acquire the compensation, especially in old age, it is necessary to enable the blind access to these technologies. Some technologies can be controversial in terms of their association with multimedia. However, their existence could be justified with the development within the complex multimedia solutions. The best example is the mp3 audio compression that was created as part of the MPEG (*Moving Picture Experts Group*) system. This technology now forms the basis for recording audio books, so-called DAISY (*Digital Accessible Information System*). Multimedia solutions for the blind rely on standards and recommendations of MPEG and ITU-T (*International Telecommunication Union*) organization. All developments within multimedia standards, at least, concern indirectly the blind.

There are also specialized hand-held mobile devices with sound recognition and gesture recognition with tactile screens as a viable option on the market of aid devices for the blind. Siqueira et al. explored out the possibility of use Braille text entry on smartphones, because they noticed that, due to the same reason, those devices are less accessible for blind people [6].

Moreover, in the past few years, touch-screen devices have become widely adapted to blind people thanks to screen readers. A blind user can explore a screen by touching it. To select an element, the user performs a second gesture such as a double tap. While this interaction makes an interface generally accessible, the disadvantage is that finding and selecting keys on an on-screen keyboard is slow and error-prone.

The article is organized as follows: some aspects of text-to-speech system are presented in the next section, followed by the section dedicated to multimedia applications designed for the blind. The description of a solution of mobile phone for the blind will be presented. Some advanced multimedia solutions regarding mobile assistive technologies are also described in the following section. Concluding remarks are discussed in the final part of the paper.

2. CONVERTING TEXT-TO-SPEECH APPLICATION

Converting text-to-speech (TTS) is a computer application designed to read any text, regardless of its origin. An important difference between devices that reproduce speech and TTS program is the ability to create new words and sentences. It can be said for TTS program that automatically translates graphemes to phonemes in order to pronounce the given sentence [7].

The earliest attempts at synthesizing human speech date back to more than two hundred years. Professor Christian Kratzenstein described the physiological differences between the five vowels (/ a /, / e /, / i /, / o /, / u /) in 1779 in St. Petersburg. He made the device with functionality similar to human speech apparatus. Resonators were activated by vibrating whistle, like that of a musical instrument. Phoneme / i / was obtained by blowing in the lower section without a mouthpiece, in order to obtain a sound of the flute.

Each TTS program is the result of efforts to achieve the appropriate level of speech fidelity within the technology available at the time of program formation. The concept of high-quality speech synthesis appeared in the eighties of the twentieth century as a result of significant progress in the development of speech synthesis techniques and natural language processing.

Potential applications of high-quality TTS programs are diverse:

Telecommunication services – text content can be accessed with ordinary phone by using TTS program. Taking into account that about 70% of all telephone services require a minimum of interactivity, this feature becomes extremely attractive. The message can be simple, such as, for example, cinema program delivery.

Help to handicapped - visual disability is caused by mental or motor/sensory impairments. If the latter is the case, the device is equipped with a custom keyboard and software for rapid synthesis of sentences. Telephone speech services are another example of the successful application. The blind use written contents by the TTS system associated with the system for text recognition (Optical Character Recognition, OCR).

Voice control - the voice message is usually more impressive than the written. It is acceptable, while attention may be kept on other sources of visual information. Hence, we have the use of synthetic speech in measurement and control systems. The best examples of voice control are Global Positioning System (GPS) devices that help driver to navigate easier in traffic by giving verbal instructions.

Multimedia human-to-machine communications - TTS systems are a necessary step (along with systems for speech recognition) in the direction of complete communication of humans with computers. The use of multimedia is a significant starting point in this process.

Fundamental and applied research - TTS synthesizers can be completely controlled, making them as valuable laboratory tools in linguistic laboratories. Full repeatability of results enables intonational, rhythmic and other types of speech.

Although the text-to-speech conversion has become part of everyday life, it is necessary to describe the operation of these programs. Understanding the internal organization of converting gives an insight into real possibilities and limitations of this technology. Text-to-speech occurs successively so that as the output data stream gives PCM (*Pulse Code Modulation*) digital sound. Components of this system include:

- Text normalization;
- Homographic parsing;
- Pronunciation of words;
- Prosody;
- DSP (Digital Signal Processing) component.

Text normalization breaks up the text to a series of words. For example, the sentence: "Lazar drove home with speed of 55.3 km/h" after normalization becomes "Lazar", "drove", "home", "with", "speed", "of", "fifty", "five", "point", "three", "kilometers", "per", "hour", where the occurrence of each punctuation is also recorded. Normalization also works successively as follows:

-The words in the text are isolated firstly. This is mainly trivial, because sequences of letters are isolated with the occasional quote or line.

-Finding the numbers and symbols follows as a next step, for example: the date, time, money. Found instances are replaced by the words consistently to linguistic and contextual rules.

-Abbreviations are translated in the unabbreviated form. This is usually achieved by using a database of abbreviations.

-It is important whether the interpunctuation is pronounced or not. For example, full stop at the end of a sentence is not pronounced, but if it is located within an internet address it is pronounced.

-Normalizer complexity differs between programs. Once the text is normalized and reduced to a series of words, it is forwarded to the next module, homographic parsing.

While *homographic* parsing is usually not a stand-alone module, it logically forms a whole. Therefore, it is described separately. Typically, this module is combined with the text normalization or the pronunciation components. In most languages there are many words that are written alike, but are spelled differently and they are called homographs. The same goes for abbreviations, numbers and symbols. TTS programs use different approaches for solving the problem, but they usually assume the context for which the separate pronunciation is suitable. TTS programs assess the meaning of the text according to the word order in a sentence by parsing into the different types of words.

Pronunciation of words block receives text and sends a series of phonemes for further processing, like the spoken vocabulary. In order to obtain the correct word pronunciation, the word is firstly searched in an internal pronunciation lexicon. If it could not be found, the program determines a pronunciation according to the rules. The complexity of this module is also very different in various TTS programs.

Prosody refers to the pronunciation melody, speed and emphasis. The pronunciation is "mechanical" or "robotic" without the prosody. If the prosody is poor, pronunciation is weird. Although the TTS programs differ, there are some general rules that determine prosody. First, the start and the end of a sentence are determined, showing the tone pitch of the pronunciation, in accordance with the type of language. Breaks are inserted between sentences. Phrases are pointed out like the sentences, if successfully identified. Breaks are also inserted at the comas. Some programs seek to find the most important words in the sentence, and they emphasize them particularly. Such words are pronounced as louder, longer and with more melody variation. Words that are less important are pronounced unstressed. Output data of prosody are series of phonemes with tone pitch, duration and loudness.

DSP component for speech synthesis behaves in accordance with the properties of natural human speech. It was shown that transitional properties between phonemes were more important than stationary for better speech understanding.

3. MULTIMEDIA APPLICATIONS IN TIFLO-TECHNIQUE

Tiflo-technique is an umbrella term for all the aids for the blind. Objects of prestige in tiflo-technique could be distinguished from the useful and necessary objects, as in all other spheres of life. Having in mind that objects of prestige enhance the Quality of Life (QoL), they should not be dismissed as unnecessary. Professional tiflo-experts or blind users should assess the importance of that objects.

Tiflo-technical multimedia aids can be considered as all devices that are compatible with appropriate interface in the multimedia field. These aids facilitate access and make available multimedia content such as internet content, texts, and moving pictures.

Braille display

Braille display is a specific peripheral device for the blind in which all the data can be seen on the monitor. This device is shown in Fig. 1. Data appear in character letters for the blind (Braille letters).

Fig. 1. Braille device and electronic notebook [8]

There are Braille displays with 20, 40, 65 and 80 characters horizontally and up to 13 characters vertically. The number of characters depends on the price of aids. These devices use programs WinDots, Hal/Supernova and Jaws under Windows operating systems.

Plextalk PTP1

Plextalk PTP1 can be used as a pocket recorder and player. This small and powerful device is ideal at work, home and school. The player is designed for blind people and is suitable for all ages and needs, due to large and easily recognizable buttons. All keys are wired as all other commands. The device supports all known audio formats such as *mp3, wav, wma, audio cd* and many others. It has a rechargeable battery, built-in speaker and microphone, as well as input jack for an external microphone. *Plextalk* PTP1 is supplied with its own memory card and supports charger, USB cable, carrying case and headphones. *Plextalk* PTP1 is illustrated in Fig. 2.

Fig. 2. Plextalk PTP 1 device [9]

Braille+ device

Braille+ is one of the most flexible, affordable and user-friendly mobile devices available today. This device coordinates the multimedia collections, can read and correct e-mail and documents, plays and records voice and music, can browse the web, read digital audio books. Also, Braille+ follows the headings and subheadings, manages contacts and calendar, makes notes, blogs, plays music from the internet radio stations, and many more. This device is small enough to fit in a purse or a pocket. It offers non-visual means of communication, learning, information, entertainment and managing lifestyle through the built-in speech output.

Ergonomic tasters help to identify control keys, stereo speakers enhance audio quality, and multiple ways of typing mean that *Braille+* is ready for just about every job, on a desk or on the move. *Braille+* device is represented in Fig. 3.

Book port plus

Book port plus (Fig. 4) is also, as well as Braille+, a relatively new instrument designed for listening to digital audio sound, internet radio, reading computer folders

and excellent sound recording. Hardware and software make this device much easier, more reliable and more flexible than mobile phones, which support some of the functions listed above.

Fig. 3. Braille+ device [10]

Fig. 4. Book port plus device [10]

DAISY audio books format

DAISY (*Digital Accessible Information System*) consortium is founded by a group of audio book libraries in 1996 with the aim of transition from analogue to digital sound books. Vision of DAISY consortium was the availability of information to people with reading difficulties, in a format that is accessible, well equipped, enables navigation, with a price equivalent to an adjusted book [11].

DAISY format is an international standard for production, exchange and use of digital audio books. The format is designed to be integrated into existing technological developments, to ensure the availability of information for people with reading disabilities.

Objectives of DAISY consortium are the following:

- Create and promote a global standard for navigation and structuring of digital audio books;
- Encourage the opening and development of the digital library of audio books both in developed and developing countries;
- Increase the accessibility and use of electronic books and multimedia documents for people with reading disabilities;
- Provide recognition and acceptance of the DAISY standard for multimedia documents with the possibility of navigating among the leading publishers;
- Encourage the creation and development of a global digital library of audio books that exceed the political limits and linguistic differences, but at the same time accept cultural diversity.

DAISY format is based on HTML/XML and SMIL (*Synchronized Multimedia Integration Language*) program language. Although the format of the media content is not obligatory, audio contents are typically encoded in mp3 format. SMIL allows easy creation of interactive audio-visual presentations and is typically used for "rich media"/multimedia presentations that include audio and video streaming, text or any other form of content.

Using a SMIL can specify sequence of media at a time, synchronized with each other and designed according to user demands. Multimedia presentation is easily created and corrected, as the text makes SMIL format. SMIL is developed by a group under the coordination of the *World Wide Web Consortium* (W3C) with the cooperation of representatives from various multimedia products (interactive TV, Web and audio/video streaming).

4. MOBILE TELEPHONE FOR THE BLIND

There are possible solutions on the market for a mobile phone designed for blind and visually impaired persons. One available solution is shown in Fig. 5. The phone has a speech device which transmits information to a user by sound, instead of a screen. Each key on the keypad on the phone is associated with the different sounds, for easier handling. The text message or name and number of a caller are transmitted via the sound to the phone.

Fig. 5. Mobile telephone for the blind [12]

This mobile phone has other features:

- dual band mobile phone adapted for the blind;
- serial port and headphone jack;
- synthetic speech and a clock;
- loudspeaker for "hands-free" use;
- address book;
- advanced calling features;
- SMS messages.

There are some other phones specially developed for the blind with some additions, such as reading an e-mail, listening to internet radio and other contents over the internet network. However, the blind mostly rely on speech synthesizers for mobile phones and wireless applications, especially in our country. Mobile phones are no longer used just for talking and text messaging, with the development of Android operating system. Smart phones have already begun to install the applications with various features and options.

5. MULTIMEDIA ASSISTIVE TECHNOLOGIES

Nowadays smartphones are everywhere in our society. Smartphones are used not only for communication, but also for entertainment and organization of our daily lives. However, the transition from hardware keys to virtual keys on touch-screen made their use more difficult for blind people. Price of smartphones is another discouraging factor for blind users, both smartphone itself and expensive accessories. For example, an external Braille keyboard for a smartphone is expensive and cumbersome to handle. Moreover, despite awareness of more accessible devices, some blind users prefer to buy cheaper options. Some of the advanced mobile devices and applications for assistance will be described in the following [13].

AudioBrowser is one modern assistive multimedia technology for the blind. It is similar to the information access tool for touch-sensitive screens. AudioBrowser allows users to browse stored information and commands by a combination of both speech- and non-speech audio feedback. Users are guided by speech and non-speech audio while the screen of smartphone is split in two to allow users to view different information from the display panel. As users' fingers move across the screen, non-speech audio is used to inform them when crossing a boundary. In a given segment of the screen, speech audio informs the user what is contained. A key advantage of *AudioBrowser* is support a structure which allows users to access information while in motion.

V-Braille is an assistive Braille aid that transfers information through vibration on touch screen [14]. It enables users who are familiar with Braille to interact with interfaces for mobile smartphones. The traditional structure of Braille approximately shares a mobile screen into six parts, which is shown in Fig. 6. When the screen is touched within a dedicated part, vibration of different strengths forces a character that enables users to make difference among characters.

Fig. 6. V-Braille device [14]

MobileEye aims to help the visually impaired to see and understand their environment during travel and other activities by using a mobile smartphone's camera and TTS technology. The system consists of four subsystems adapted for different types of visual disabilities: (a) a color channel to help the user distinguish colors; (b) software to provide image magnification, reading and understanding of objects; (c) pattern recognizer for objects recognition; and (d) document retriever to allow access to printed materials by using only a snapshot of a page and download the document from a large database. Each operation of the software is driven by a voice message.

PictureSensation is another mobile application for hapto-acoustic exploration of images [15]. It is designed to enable the visually impaired to obtain direct perceptual access to images via an acoustic signal. *PictureSensation* introduces a gesture-based and speech-guided user interface that guarantees the autonomous use by the blind. It implements recently proposed exploration and audification principle, which helps visually impaired in orientation in everyday situations. A user actively explores on a touch screen and receives auditory feedback on the content at current position of their finger. *PictureSensation* provides an extensive tutorial to allow the blind to become familiar with the use of the application, as well as the principles of image content to sound transformation, without any assistance. This application has potential to help visually impaired to explore and understand the whole scenes, even on screens of smartphones.

PictureSensation presents a valuable mobile application to grant the blind access to the visual world through exploration.

Fig. 7. PictureSensation mobile application [15]

6. CONCLUSION

Multimedia irreversibly changes the lives of the blind for the better. Measure of changes can be described in terms of improving their Quality of Life (QoL). The prestige that is associated with the use of various devices in daily life is fully justified in terms of improving QoL.

However, multimedia develops independently of the needs of the blind. Suitable solutions for blind people are not necessarily in the mainstream of development, but they are never contradictory. Most aids for the blind use speech which can be synthetic, or recorded. Fewer aids use sound, whether in primitive or in an abstract form. The development of screen reader started in economically and technologically more developed part of the world.

Blind people have started to use computers, which help them be involved in all contemporary streams of life. Screen readers are becoming better and more efficient with constant improvements. On the other hand, their price decreases rapidly, so that they are becoming available and widely used among people they were primarily intended for. However, the need for such a new technology progress occurred also in less developed communities, which imposed the problem of linguistic diversity.

ACKNOWLEDGMENT

REFERENCES

[1] L. Bian, "Research on multimedia technology in higher special education", 2008 International Conference on MultiMedia and Information Technology, Three Gorges, 2008, pp. 727-730. doi: 10.1109/MMIT.2008.152

[2] D. Virkes, "Multimedija za slijepe", Magistarski rad, Fakultet za Elektrotehniku i Računarstvo, Zagreb, Hrvatska, 2004.

[3] S. Dikić, "Tiflologija", Univerzitet u Beogradu, Defektološki fakultet, I izdanje, Beograd, 1997.

[4] J. S. Garcia, "Tiflotecnologia", Acción social, Revista de Política social y Servici os socials, Vol. I, No. 5, pp. 97-107, 2017.

[5] V. Delić, N. Vujnović, "Stereo presentation and binaural localization in a memory game for the visually impaired", Lecture Notes in Artificial Intelligence, A. Esposito et al. (Eds.), COST 2102 Int. Training School, pp. 354-363. Springer, Heidelberg, 2009.

[6] J. Siqueira, et al., "Braille text entry on smartphones: a systematic review of the literature", 40th Annual IEEE Computer Software and Applications Conference (COMPSAC), Vol. 2, 2016.

[7] M. Sečujski, "Accentuation dictionary of Serbian language intended for text-to-speech synthesis", Proceedings of Conference Digital Speech and Image Processing (DOGS), Bečej, pp. 17-20, 2002.

[8] http://www.abledata.com

[9] http://www.plextalk.com

[10] http://tech.aph.org

[11] http://www.daisy.org/

[12] http://www.tuvie.com

[13] L. Hakobyan, J. Lumsden, D. O'Sullivan, H. Bartlett, "Mobile assistive technologies for the visually impaired", Survey of Ophthalmology, Vol. 58, No. 6, 2013. doi: 10.1016/j.survophthal.2012.10.004

[14] C. Jayant et al, "VBraille: haptic Braille perception using a touch-screen and vibration on mobile phones", 12th International ACM Conference on Computers and Accessibility ASSETS, Orlando, Florida, USA, 25-27 October 2010.

[15] M. Banf, R. Mikalay, B. Watzke, V. Blanz, "PictureSensation – a mobile application to help the blind explore the visual world through touch and sound", Journal of Rehabilitation and Assistive Technologies Engineering, Vol. 3, No. 1-10, 2016. doi: 10.1177/2055668316674582