KNOWLEDGE – International Journal
Vol. 28.3
December, 2018

VISUALLY IMPAIRED STUDENTS EDUCATION THROUGH INTELLIGENT TECHNOLOGIES

Lindita Ademi
University of Tetovo, Macedonia lindita.ademi@unite.edu.mk
Valbon Ademi
University of Tetovo, Macedonia valbon.ademi@unite.edu.mk

Abstract: The problem for developing a TTS (text-to-speech) is a very active field of research. As the Human-Computer Interfaces (HCI) come of age, the need for a more ergonomic and natural interface than the current one (keyboard, mouse, etc.) is being constantly felt. Talking of natural interfaces, what comes to mind, is sound (speech) and sight (vision). These form the basis of many intelligent systems research like robotics. Moreover, speech can also serve as an excellent interface for visually impaired , or people with motor neuron disorders. In this paper we attempt at developing a TTS system for Albanian Language. A lot of commercial systems are available for many foreign languages (mostly English), but there is yet to be a competitive system available for Albanian language. Although the task of building very high quality, unlimited vocabulary text-to-speech (TTS) system is still a difficult one, with many open research questions, we believe the building of reasonable quality voices for many tasks can serve our needs. Here we have worked with standard Albanian, the most commonly spoken. We hope to easily extend the system to other languages, since there are a lot of underlying similarities between languages. Albanian language being highly phonetic, result in simple letter-to-sound rules. We used the standard concatenative synthesis. The main problem faced by us was to make the synthesized speech sound natural. We investigated the reasons for the mechanical sounding speech and developed different synthesis models to overcome some of those problems. Moreover, we implemented some standard and also novel intonation and duration modification algorithms, which can be incorporated into the TTS at a later stage. Our main achievement was reasonably legible speech with an unlimited vocabulary. The following paper presents a brief overview of the main text-to-speech synthesis problem and its subproblems, and the initial work done in building a TTS for Albanian.
Keywords: Natural Language Processing, Text-to-speech, visually impaired, education.

INTRODUCTION

Intelligent User Interface (IUI), or sometimes an interface agent is a user interface (UI), which includes some aspects of artificial intelligence. The field of intelligent user interfaces covers various topics and deals with the application of artificial intelligence and knowledge-based techniques to the issues of human-computer interaction [11] [14] [16] [20] [21] [28] [32] 34] [35] [37] [38].

Research on new methods of communication focuses on systems of natural language, gestures, images and multimodal interfaces [30] [31].

One of the major benefits of the rapid development of intelligent user interfaces are applications known as screen readers, which greatly facilitate the lives of blind and visually impaired people [2].

According to data from the World Health Organization (WHO), 285 million people in the world are visually impaired. Of them, 39 million people are completely blind and 246 million are impaired. Approximately 90% of all blind or visually impaired people in the world live in developing countries [42]. The number of blind people and persons with visual impairment above 90% in Macedonia is about 2000.

Taking into account that computers and computers, including the Internet, are being accessed everywhere in human life and work, it is more than obvious that they need to be available to disabled people including blind people and visually impaired people. That is why the interest in this paper is to help this group of people with the localization and development of models and opportunities for simpler use of the capabilities of computers and the web.

Assistive technologies include all technologies that help disabled people in the interests of their normal living and functioning, enabling them to participate in society as much as other normal participants. For blind and visually impaired users, such auxiliary technologies include Braille, Braille display, screen readers and other auxiliary tools.

A Text-to-speech synthesizer is an application that converts text into spoken word, by analyzing and processing the text using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here, we developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3.file. The development of a text to speech

synthesizer will be of great help to people with visual impairment and make making through large volume of text easier.

Text-to-speech synthesis -TTS - is the automatic conversion of a text into speech that resembles, as closely as possible, a native speaker of the language reading that text. Text-tospeech synthesizer (TTS) is the technology which lets computer speak to you. The TTS system gets the text as the input and then a computer algorithm which called TTS engine analyses the text, pre-processes the text and synthesizes the speech with some mathematical models. The TTS engine usually generates sound data in an audio format as the output. The text-to-speech (TTS) synthesis procedure consists of two main phases. The first is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is the generation of speech waveforms, where the output is produced from this phonetic and prosodic information. These two phases are usually called high and low-level synthesis [1]. A simplified version of this procedure is presented in figure 1 below. The input text might be for example data from a word processor, standard ASCII from e-mail, a mobile text-message, or scanned text from a newspaper. The character string is then pre-processed and analyzed into phonetic representation which is usually a string of phonemes with some additional information for correct intonation, duration, and stress. Speech sound is finally generated with the low-level synthesizer by the information from high-level one. The artificial production of speech-like sounds has a long history, with documented mechanical attempts dating to the eighteenth century.

Speech synthesis can be described as artificial production of human speech [3]. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech (TTS) system converts normal language text into speech [4]. Synthesized speech can be created by concatenating pieces of recorded speech that are stored in a database. Systems differ in the size of the stored speech units; a system that stores phones or diphones provides the largest output range, but may lack clarity. For specific usage domains, the storage of entire words or sentences allows for high-quality output. Alternatively, a synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output [5]. The quality of a speech synthesizer is judged by its similarity to the human voice and by its ability to be understood. An intelligible text-to-speech program allows people with visual impairments or reading disabilities to listen to written works on a home computer.

SCREEN READERS

Screen readers are text-to-speech systems that transform text into speech and read the information presented on the computer screen. By means of a combination of keys (keyboard shortcuts) the user can move through the user interface and read all the texts available on the screen. The user can use the keypad to enter text that is transformed into speech by the screen reader and is read aloud. The first screen reader IBM Screen Reader was invented in 1986 by Jim Thatcher. This screen reader was designed to read the text that was written in the DOS operating system command line. In 1988, following the emergence of a graphical interface, IBM Screen Reader / 2 is the first screen reader designed to be used with a graphical interface. It was controlled with numeric keys on the keyboard [1]

The screen reader works on the basis of creating an external screen buffer containing all the information available on the monitor (screen) [18]. This information from the buffer is read to the users. Screen readers use system hooks to grab the user input through the keyboard, allowing the user to control the screen reader using a combination of keys (keyboard shortcuts).

And people with dyslexia can hardly read and write information displayed on the screen. The speech synthesis software helps these two-way people transform the text into speech and speech in text. Text in speech software works just like the screen reader works. Written text is scanned using Optical Character Recognition (OCR) software and the information is then transformed into a voice output. Speech in text, in turn, the software works in reverse, the voice input through the microphone of the computer transforms it into text on the screen [21] [9] [15] [43].

Other screen readers are expensive, costing nearly a thousand dollars for each installation due to their complexity, the relatively small market and high costs. The development of these programs is complex. Because of their cost, screen readers are not installed on most computers, so that blind and visually impaired users are unable to access the Internet from any computer.

To help these users came some new non-commercial high-quality projects that were easily available for use. One of them is the screen reader WebAnywhere available at the Internet address:

This paper presents the basic principles that apply when designing the system to synthesize speech in Albanian language from written texts.

For this purpose, for each language, decisions concerning its specificity should be sought, always for the purpose of creating a voice to respond to the nature of the language. The generation of systems currently used

mainly relies on the use of the concatenation method, during which the acoustic segments of text files are connected, and which have been previously digitized and stored in the database [29] [36] [44] [12].

For the Albanian language, we consider the textual part of the database. The basic segments used are: most commonly used words, two letters and letters separately. However, a certain part of the database includes various shortcuts, that is, textual equivalents and their acoustic files, which should be used during the generation of appropriate speech. The goal is to synthesize different numeric values written in a decimal system. Values are added to the database, corresponding to their sound files, whereby speech is generated for different numbers. Albanian language algorithms explain Albanian language characteristics by writing letters, through words, through syllables and through diphones, and explaining the advantages and problems that occur for each of them [13] [14].

Based on the above, that existing applications do not offer the possibility of generating speech from texts written in Albanian, the only opportunity that remains for us is to explore suitable algorithms that take into account the characteristics and specificities and will provide all the prerequisites for proper application and a design dedicated specifically for this language.

For this purpose, we initially analyzed the statistical texts written in Albanian. Universal, as in any other language, written texts can be considered as compilations of different units, such as sentences, words, syllables, and end letters. After the tests that were made, it became clear that the generation of speech through letters is very unnatural, as a result of discontinuity during the concatenation of acoustic files on special letters with word composing. This is largely conditioned by consonants, usually associated only with vowel ë ə.

The above discontinuities can be significantly improved if instead of letters, words are used as basic units. However, the basic vocabulary of the Albanian language is present. By not being able to provide acoustic files for all possible words, a compromise solution is reasonable. So, first, acoustic files of the most commonly used words are provided, covering a significant percentage of written texts, and in other words generating the sound should be done in special letters. In a similar way, as is the case with words, the case of generating speech from syllables is considered. The idea stems from the fact that the number of words is too large to be able to include all the basic units within the acoustic base, and the number of syllables may be smaller and could cover all written texts through them.

The research into the possibility of interpreting texts through a smaller number of basic units introduced us to the significance of the diphones as a technical solution to the problem in this issue.

The use of more frequent words and diphones can be suggested as an optimal solution within the context of generating speech from written texts. This can be justified by the fact that through more frequent words a significant percentage of texts will be included and continuous conversations will be generated. In other words, the quality of the diffusion generation will be lower, but the system will be stable because there will be no words that cannot be composed of diphones. Only for foreign words or for those who are not present in the Albanian dictionary will there be the option to generate speech through specific characters.

Although they had great incentives to improve access, largely blind web users were left out.

This paper explores how intelligent interfaces can enable blind people to independently improve their web access. Solutions created by users of these tools can be shared so that blind users need to help them together to create awareness of the Internet.

Disabled people should not be seen only as access clients, but as effective partners in achieving better access for everyone. With intelligent interfaces that support them, blind and normal users can jointly and effectively contribute to improving the accessibility, usability, and availability of a web-based approach [22] [26].

LANGUAGE SPECIFICS

In the group of existing intelligent accessibility interfaces, with certain exceptions, those in English are dominant. Taking an initiative to effect this kind of interface inevitably leads to a preliminary analysis of the existing solutions, and taking positive experiences or ready-made concepts inevitably leads to certain modifications and adaptations aimed at the effective use of the appropriate language area.

A particular aspect is the dominant use of English in the use of ready-made applications, as well as the prevalence of English in existing web content.

In such a situation, the issue of exclusive use of a particular language or the possibility of choosing a language - English or local - arises.

In either case, both cases involve extensive preparatory work in creating a database in a format appropriate to the selected application type. Emphasis is placed on a database that will represent a vocabulary or glossary with a voice interpretation. Of course, the basic abstract interpretation of the local language is also needed.

In many respects, the problems in creating an intelligent interface for blind people would be reduced to

localization problems. Of course this also includes the special dictionary or glossary of information technology and the preparation for its own voice interpretation.

In spite of the developed and widely used technology for the use of local language development code development pages, specific challenges are associated with the use of different code pages in the same application, at the user's choice, especially in web applications.

CONCLUSIONS AND RECOMMENDATIONS

Choosing an Intelligent Interface for blind people from non-English speaking areas comes down to a choice of two options - creating a new product or localizing an existing solution.

In both cases, prior language preparation is required. This preparation covers the contents of the basic databases on a handwritten or glossary, including special information terminology as well as their vocalization.

In the case of a choice of a cash solution, it is advisable to select an application developed in free software. The advantages are multiple, both in the possibility of involving the free software community, as well as in the participation of the very blind and visually impaired in the preparation, maintenance and updating of the application. It's not about neglecting the price, copyright, and the ability to act independently in subsequent software changes.

The characteristics of the solution are choices that can be made according to the best world experiences, which are previously listed in the section for desirable characteristics. The basic principle is, of course, the maximum respect for the principle of intuition, as well as the possibility of easy adoption of the technique of use, preferably without the necessary help from third parties.

Future research in the field of the development of intelligent user interfaces in the world aspire to new and highly developed technologies that extend the capabilities of all users, especially disabled people such as blind and visually impaired people, as categories for which they are and research in this paper.

As for the Republic of Macedonia, the aforementioned technologies in the future are expected to develop text-to-speech modules in specific software ready-made solutions, for applications that are very necessary for the blind and visually impaired, as well as other categories handicapped people from our linguistic area, who are taken care of in some of the responsible institutions, and their homes so that they can facilitate access and usability on the computers, the Internet, etc.

We hope that this paper with the research and reviews we make will contribute to future research and software solutions in the field of development, implementation and use of intelligent user interfaces in Macedonia, in this way, blind people will gain access to open content, instructions, books, instructional software-s, and the like. This would be a significant step for their inclusion and will enable them to realize their generation rights for modern education.

LITERATURE

[1] Asakawa C., What's the web like if you can't see it? In Proceedings of the 2005 International Cross-Disciplinary Workshop on Web Accessibility (W4A). pp. 1–8. 2005. http://dl.acm.org/citation.cfm?id=1061813

[2] Bigham J. Jeffrey P., Lau Tessa, Nichols Jeffrey, TrailBlazer: Enabling blind users to blaze trails through the web. In Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI 2009), Sanibel Island, FL, USA, 2009.\

[3] Bigham P. Jeffrey, Prince M. Craig, Ladner E. Richard, WebAnywhere: Enabling a Screen Reading Interface for the Web on Any Computer, In Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A), pp. 132-133, ACM New York, 2008.

[4] Bigham P. Jeffrey, PrinceM. Craig, Ladner E. Richard E., WebAnywhere: A Screen Reader On-the-Go, In Proceedings of the 2nd Cross-Disciplinary Conference on Web Accessibility. Beijing, China, 2008.

[5] Blenkhorn P., Evans, G., Architecture and requirements for a Windows screen reader. In Speech and Language Processing for Disabled and Elderly People (Ref. No. 2000/025), IEE Seminar on. pp. 1–1. 2000. http://ieeexplore.ieee.org/xpls/abs_all.jsp? arnumber=846939.

[6] Borodin Yevgen, Bigham P. Jeffrey, Dausch Glenn, Ramakrishnan I.V., More than Meets the Eye: A Survey of Screen-Reader Browsing Strategies, In Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A), ACM New York, 2010.

[7] Clark A.J. Robert, Richmond Korin, King Simon, Festival 2 - build your own general purpose unit selection speech synthesiser. In Proc. 5th ISCA workshop on speech synthesis, 2004.

[8] Designing the Moment: Web Interface Design Concepts in Action by Robert Hoekman, New Riders, 2008.

[9] Feldman R., Sanger J., The text mining handbook - advanced approach in analyzing unstructured data, Cambridge Press 2007.
[10] Freedom Scientific. http://www.freedomscientific.com/jaws
[11] Galitz O. Wilbert, The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques (CourseSmart) , Wiley Publishing 2007.
[12] Gormez Zeliha, Orhan Zeynep, TTTS:Turkish text-to-speech system, 12th WSEAS International Coneference on Computers, Heraklion Greece, pp. 977-981, 2008.
[13] Hamiti M., Dika A. Analyses of Same Content Texts Written in Different Languages. 31st International Conference on Information Technology Interfaces ITI 2009; June 22-25, Cavtat, Croatia, 2009.
[14] Hamiti Mentor, Dika Agni, Speech Generation for Albanian Written Texts. In *Proceedings of the ITI 2010, 32nd International Conference on INFORMATION TECHNOLOGY INTERFACES, pp. 85-90.* Cavtat/Dubrovnik, Croatia, 2010.
[15] Jeffrey P. Bigham- Intelligent Interfaces Enabling Blind Web Users to Build Accessibility Into the Web. http://www.cs.rochester.edu/u/jbigham/papers
[16] Johnson Jeff , Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules, Elsevier, 2010.
[17] Laurel Brenda, The Art of Human-Computer Interface Design, Morgan Kaufmann Series, 1990.
[18] Mayhew J. Deborah, Principles and Guidelines in Software User Interface Design, Prentice Hall, 2008.
[19] MBROLA. Diphone Studio: Tool for developping and maintaining a set of diphones, 2004. http://tcts.fpms.ac.be/synthesis/mbro la.html [09/20/2009]
[20] McKay N. Everett, Developing User Interfaces for Microsoft Windows , Microsoft Press, 1999.
[21] Mittal O. Vibhu O, Yanco A. Holly, Aronis John, Simpson C. Richard, Assistive Technology and Artificial Intelligence: Applications in Robotics, User Interfaces and Natural Language Processing (Lecture Notes in Computer ... / Lecture Notes in Artificial Intelligence), Springer, 1998.
[22] Nichols Jeffrey, Lau and Tessa, Mobilization by demonstration: using traces to reauthor existing web sites. In Proceedings od the 13th International Conference on Intelligent User Interfaces (IUI 2008), pages 149-158, New York, NY, USA, 2008.
[23] Obendorf Hartmut, Minimalism: Designing Simplicity (Human-Computer Interaction Series), Springer, 2009.
[24] Parssinen Kimmo, Multilingual text-to-speech system for mobile devices: development and applications, Tampere University, Tampere 2007.
[25] Piton Odile, Lagji Klara, Përnaska Remzi, Electronic Dictionaries and Transducers for Automatic Processing of the Albanian Language, Natural Language Processing and Information Systems, Lecture Notes in Computer Science Volume 4592, pp 407-413, 2007.
[26] Richard Atterer, Monika Wnuk, and Albrecht Schmidt. Knowing the user's every move - user activity tracking for website usability evaluation and implicit interaction. In Proceedings of the 15th International Conference on World Wide Web (WWW '06),pages 203-212, New York, NY, 2006.
[27] Ryu Hokyoung , Mobile User Interface Analysis and Design: A Practitioner's Guide to Designing User Interfaces for Mobile Devices, Nova SciencePublishers, 2009.
[28] Saffer Dan, Designing Gestural Interfaces: Touchscreens and Interactive Devices, Oreilly, 2008.
[29] Schroeter Juergen, Basic Principles of Speech Synthesis, Springer Handbook of Speech Processing, pp 413-428, Springer, 2008.
[30] Scott Bill, Neil Theresa, Designing Web Interfaces: Principles and Patterns for Rich Interactions, O'Reilly, 2009.
[31] Shao Ling, Shan Caifeng, Luo Jiebo, Etoh Minoru, Multimedia Interaction and Intelligent User Interfaces: Principles, Methods and Applications (Advances in Pattern Recognition Springer, 2010.
[32] Shneiderman Ben, Plaisant Catherine, Cohen Maxine , Jacobs Steven, Designing the User Interface: Strategies for Effective Human-Computer Interaction, 2009.
[33] Speech Acts and Prosodic Modeling in Service-Oriented Dialog Systems (Computer Science, Technology and Applications) by Christina Alexandris, Novinka, 2010.
[34] Stephanidis Constantine, User Interfaces for All: Concepts, Methods, and Tools (Human Factors and Ergonomics), 2000.
[35] Stone Debbie, Jarrett Caroline, Woodroffe Mark , Minocha Shailey, User Interface Design and Evaluation (Interactive Technologies), Elsevier, 2005.
[36] Taylor Paul, Text-to-Speech Synthesis, Cambridge University Press, 2009.

[37] Teodorescu Horia-Nicolai, Mlynek Daniel, Kandel Abraham , Zimmermann Hans-Jürgen, Intelligent Systems and Interfaces, Kluwer Boston, 2000.
[38] Tidwell Jenifer, Designing Interfaces, O'Reilly, 2011.
[39] V. Ademi, Advanced techniques for developing intelligent user interfaces, INTERNATIONAL CONFERENCE ON - RESEARCH AND EDUCATION - CHALLENGES TOWARD FUTURE, University of Shkodra "Luigj Gurakuqi", Albania, 2013.
[40] V. Ademi, A. Madevska Bogdanova, Development of Intelligent User Interfaces, International Journal of Arts & Sciences(IJAS) - Conference for academic disciplines, Malta, 2013.
[41] V. Ademi, Development of Intelligent User Interfaces, 12th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Las Vegas USA,2013.
[42] WHO | Visual impairment and blindness. WHO. http://www.who.int/mediacentre/factsheets/fs282/en/
[43] Witten I. H., Frank E., Data Mining – Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufman Publishers, 2005.
[44] WIMMER, K., AGUERO, P., BONAFONTE, A. Joint extraction and prediction of Fujisaki's intonation model parameters. Proceedings of International Conference on Spoken Language Processing, 2004.
[45] BLACK, A. W., TAYLOR, P., CALEY, R. The Festival Speech Synthesis System. Manual and source code avaliable at http://www.cstr.ed.ac.uk/projects/festival.html, 1996-2006.
[46] AYLETT, M. P. Synthesising hyperarticulation in unit selection TTS. In Proceedings of Eurospeech 2005, 2005.
[47] ALVVAREZ, Y. V., HUCKVALE, M. The reliability of the itu-t p.85 standard for the evaluation of text-to-speech systems. In Proceedings of the International Conference on Speech and Language Processing 2002, 2002.