Albanian-Speaking Blind User Interfacing Model

Vanco E. Cabukovski[1], Valbon Ademi[1,2], and Roman V. Golubovski[1,2]
[1]Faculty of Natural Sciences and Mathematics, Ss Cyril and Methodius University, Skopje, Macedonia
[2]Faculty of Natural Sciences and Mathematics, State University of Tetovo, Tetovo, Macedonia
Email: cabukv@hotmail.com, valbon.usht@gmail.com, roman.golubovski@t.mk

***Abstract*—The full or satisfactory perception of the environmental context is probably the most significant factor in describing "quality of life", among other health and social related issues. For people with impaired sense(s), especially those with unrecoverable loss of sight, being able to somehow compensate for sight dependable activities becomes the very meaning of life. Fortunately, contemporary technologies advancement allows for R&D efforts that result in more or less feasible solutions to most of those challenges. With the Internet being accessible to everybody, news and information is available beyond limits. One of the main focuses of the modern science is providing applications and the Web to blind and visually impaired. It is now possible more than ever by the advancements in User-Interfaces (UIs) and assistive technologies. There are already stand-alone text-to-speech screen reader applications that provide application-specific speech output. All these screen readers usually cover widely speaking languages. Understandable, it is also needed for all other languages. This paper depicts an UI model for Albanian speaking blind and visually impaired based on extensive research embracing screen readers analysis, text-to-speech technologies analysis, specially developed text-to-speech generator from Albanian texts, the analysis done on general Albanian texts as well as the basic principles of lexicons creations.**

***Index Terms*—screen readers, speech synthesis, Albanian-speaking blind user interface, interfacing model**

I. Introduction

Current technologies can greatly improve to high extent the quality of life for all categories of disabled or partially impaired people, especially the blind and visually impaired.

Besides the spatial navigation, main issue related to the loss of sight is also the usual every-day information and news acquisition. With the Internet accessibility today, the programming industry helps these people by providing assistive software known as "screen readers" [1]-[3].

The screen readers are text-to-speech systems that transform the text into speech and read the information presented on the computer display. They provide convenient keyboard combinations and shortcuts allowing the user to navigate through the text interface and retrieve the information. With the use of the keyboard the user can also input text to hear back converted speech. One well known screen reader is WebAnywhere [4].

The core of the text-to-speech implementations are the so called Intelligent User Interfaces (IUI), also known as Interface Agents. As their name suggests, IUIs employ some form of Artificial Intelligence (AI) and knowledge-based techniques covering man-computer interaction aspects [5]-[7].

Speech synthesis is an artificial production of human speech from written texts. It is the central part of the text-to-speech technology, ranging widely in significant application areas, one of which is the implementation of auxiliary technologies and tools for blind and visually impaired people and dyslexia, where the screen readers belong.

The speech synthesis is highly dependent on lingual specifics. It needs set of algorithms tailored to translate written semantics of a particular language into its natural speech. Currently available are solutions for natural speech generation of worldwide mostly used languages [8]-[12]. However, they are more or less unusable for the rest of the languages, since it is impossible to achieve a universal solution that would recognize speech of other languages as "natural" [9], [10], [13].

Therefore, each language should be analysed separately, extracting its unique specifics, especially aiming towards authentic voice creation. This paper presents the basic principles of a system design for speech synthesis of the Albanian language, from written text. Essentially, this approach considers written text as composition of following set of units - sentences, words, syllables and letters.

Originally, we statistically analyzed texts written in Albanian language. Universally, as in every other language, the written texts can be considered as compositions of various units, such as sentences, words, syllables and, at the end, letters. The thorough analysis provided a clear conclusion that generated speech through letters is very unnatural due to the discontinuity during concatenation of the acoustic files of special letters with composing words. This is to a greater extent conditioned by the consonants, which are usually connected only with the vowel ë ə. These discontinuities can be considerably improved if instead of letters, words are used as basic units. However, the basic vocabulary of the Albanian language is present. Since it is not possible to provide acoustic files for all possible words, it is rational to have a compromised solution. So, acoustic files of the most commonly used words were provided first, covering a

Manuscript received February 20, 2016; revised June 2, 2016.

doi: 10.18178/ijeee.5.1.66-70

considerable percentage of the written texts, and for other words the generation of speech should be made by pronouncing single letters.

Similarly to the method with words, it is the case with generating a speech from syllables. The idea emerges from the fact that the number of words is too big to include all basic units within the acoustic database, and the number of syllables can be smaller and through them we could cover all the written texts.

The research for the possibility of interpreting texts through a smaller number of basic units, introduced us in the meaning of diphones, as technical solution for this problem regarding this matter. The usage of more common words and diphones could be suggested as optimal solution within generating speech from written texts. This could be justified with the fact that through more common words a considerable percentage of texts can be included and continual speeches will be generated. For the other words the quality of generating thorugh diphones will be lower, but the system will be stable because there will be no words that cannot be composed of diphones. Only for the foreign words or those that cannot be found in the dictionary of the Albanian language, there will be the option of generating speech through specific characters.

According to the characteristics of the Albanian language, the process of conversion the sentences into words is a process of segmentation and classification of the written form of the sentence. Since the classification and segmentation are mutual here, it is not possible to do it one by one. On the contrary, our approach is first to make *temporary* segmentation in potential written forms called tokens, and then the next step will be to examine each token and solve any ambiguity.

In the phase of pre-processing the text, it is usually first divided into sentences, then divided into tokens which are separated by characters of blank space. The module for pre-processing the text is also responsible for management of non-standard words such as the abbreviations and numbers. In the phase of morphological analysis, the input text is analysed in order to find morphemes in each word. Similar to this, in the contextual analysis, each word has a dedicated tag for word class, and in the phase of syntactic analysis, the type of sentence is determined, for example declarative or interrogative. Lastly, in grapheme-in-phoneme module an exact pronunciation of the input words is defined and the prosodic model is responsible for generating the correct prosody for the synthesized speech [14].

II. Genereting Speech from Written Texts in Albanian Language

Currently there are various technologies for converting a written text into speech. Their mutual goal is artificial generating of natural speech that is maximally comprehensive. But, unfortunately, such perfect convertors cannot be found yet. That is why, the researches of this area have an upward tendency because this has a strong influence on the improvement of the efficiency of the existing solutions and defines instant achievements in the field of interest.

The last reform in 1972, brought the standard Albanian literature language. There was a linguistic move to "unite" the two dialects of the Albanian language: Gheg and Tosk. While it seems that the official language is the language of the media and the schools, it is not the language of each speaker of the Albanian language. This language is still a subject of variations [15].

The Albanian language and its alphabet consist of 36 letters, all of which are in Latin script, except for two letters which are written with diacritical script ç and ë The number of vowels is seven: a, e, ë ə, i, o, u, и y. Nine letters are considered as compounds, because they represent combinations of two letters: dh ð, gj, ll l, nj, rr, sh ʃ, th θ, xh dʒ and zh ʒ.

The verbs can have one or two forms: *Z* active form and/or *Z-hem* passive form. Both forms have a same past participle, and the four tenses of the passive form are made from the same tense as the active form by adding the prefix "u": for example, *lava* (I washed -aorist) *u lava (I washed myself*). For some tenses the prefixes *të* and *do* are used, and for other tenses some other prefixes are used. For example, *laj, të laj, do të laj,* (active form) and *lahem, të lahem, do të lahem* (passive form) with 6 different tenses. If we take that *laj* is present tense, the same verb in future tense is *do te laf.*

In the Albanian language, the nouns and pronouns change most. A large number of words of masculine gender in singular change into feminine gender in plural. Hence, the gender is not a constant propriety of a noun. The declination position is at the end of the word, but for three pronouns, it is in the middle, for example, *cilido, ciiitdo, and cilindo.*

Foreign personal names are transcribed according the Albanian phonetics: *Shekspir* Shakespear, *Xhems Xhojs* James Joyce, *Sharl dë Gol* Charles de Gaulle. They change the same as other nouns.

In front of the most adjectives, some nouns and some pronouns there is a particle called article. These articles have declinations: their four forms can be *i, e, të, së.* These declensions vary according to the place of the articulated adjective or articulated noun in nominal syntagm.

From the conducted analysis of the Albanian written texts, it can be noticed that the usage of the letters is not the same. Around 450 most often used words in the Albanian language cover more than 50% of the texts in Albanian. Around 200 words cover about 45%. At generating the speech, first are provided the acoustic files of the most often used words, that cover significant percentage of the written texts, while for other words the generating of the speech should be done through single letters.

In a similar way as it was the case with words, it is the case of generating the speech from syllables. The idea is based on the fact that the number of syllables can be smaller and through them to cover all written texts. However, the results of the statistical analysis of texts in Albanian confirmed the total number of 10.619 syllables

used in the Albanian language. Theoretically, in the Albanian language there are 36×36=1296 possible diphones [16]. But, on the basis of the analysis of the written Albanian texts it was detected that only 892 of them are in fact used in the semantic meaning of the language. On the basis of the calculation of the frequency of the diphones in the Albanian language only 45 of them can cover around 50% of the written texts. Since their total number is small, it is reasonable that all diphones should be included in the database of the acoustic files.

The process for speech generating from Albanian written texts consists of following three phases:

- Normalization of the text,
- Creation of acoustic database and
- Conversion of the text.

Textual normalization means transformation of the text in the most adequate form for further processing. As per [16] and [17] for Albanian language, the normalization is realized in following three steps:

- *Step 1*: First, the entire text is converted in small letters from the alphabet.
- *Step 2*: The specific letters dh, gj, ll, nj, rr, sh, th, xh, zh, ë и ç are replaced with equivalent symbols: D, G, L, N, R, S, T, X, Z, E и C.
- *Step 3*: At the end, this is used for creation of a table which is used for converting the textual contents into numbers, special symbols, acronyms and abbreviations.

The equivalent mass consists of 200 units which are considered to be necessary for the Albanian language, but also, additional units can be added to it. All words that are not in the equivalent mass, will be interpreted letter by letter. Numbers that are out of this pattern will be interpreted digit by digit. Further segmentation of the text will continue through punctuation symbols.

From all these approaches to the speech generating, at the same time respecting some proposed algorithms and the perception of the generated sound by the developed application the following main conclusions are drawn:

- Generating speech letter by letter is comprehensive but it is not natural and connected to difficult discontinuities in the speech.
- Generating word with quality speech results – we get comprehensive and natural speech. However, the only disadvantage is that not all the words can be included in the acoustic database.
- The use of diphones as basic units is characterized with the fact that it provides homogenous and stable solution, as any text can be interpreted through a combination of acoustic added files to 892 identified diphones.
- A possibility which offers compromised solutions, from the aspect of speech quality that is generated and also from the aspect of optimisation of the acoustic database. Considerable number of the most used words are uttered according to the record of the original, while others are comprised of diphones, and several are uttered through special letters.

III. Description of a User-Interface Model for Albanian Speaking Language

In the group of existing intelligent interfaces for accessibility, with certain exceptions, the English language ones dominate. The choice of the option for intelligent interface for blind and visually impaired persons from non-English speaking regions comes to choice of two options – creation of a new product or localization of some existing solution.

In case of choosing a ready-to-go solution, it is recommendable to choose an application that is made in free software. By following the studies for other languages, we try to treat the idea of speech synthesis by merging the acoustic files of textual segments stored in the database.

The main problem that occurs during the synthetizing process is connecting the acoustic segments (acoustic segments with certain number of words), in the process of generating words.

If the word from the written text is not saved as a separate segment, then it is divided in two letters, then a request is made of these two letters in the database of certain acoustic segments, and thus by merging these acoustic segments we get a certain word.

The Albanian language is not included in the set of languages for which there is a final solution for this subject. We are dedicated to find a model in which texts written in Albanian language can be transformed into speech.

In this case we need a preliminary language preparation. This preparation covers the contents of the basic databases of a vocabulary or terminology book, including special IT terminology, as well as their vocalization. The speech is a vocalization form of human communication. Every uttered word is created by phonetic combination of limited set of vocal, sound units–vowels and consonants.

The quality of the text conversion into speech is measured by two main parameters: comprehensiveness and naturalness. The comprehensiveness should be done with clarity of the heard speech, and naturalness refers to the similarity of the speech with the artificially generated common speech.

According to the characteristics of the Albanian language, the process of conversion the sentences into words is a process of segmentation and classification of the written form of the sentence. Since the classification and segmentation are mutual here, it is not possible to do it one by one. On the contrary, our approach is first to make temporary segmentation in potential written forms called tokens, and then the next step will be to examine each token and solve any ambiguity. This first process is called tokenization; the step that generates words from the tokens is called text analysis. With tokenization (segmentation) of the text, we open an opportunity for the algorithms that are used for the text analysis to focus on one token at a time.

In the phase of pre-processing the text, it is usually first divided into sentences, then divided into tokens which are separated by characters of blank space. The module for pre-processing the text is also responsible for

management of non-standard words such as the abbreviations and numbers. In the phase of morphological analysis, the input text is analysed in order to find morphemes in each word. Similar to this, in the contextual analysis, each word has a dedicated tag for word class, and in the phase of syntactic analysis, the type of sentence is determined, for example declarative or interrogative. Lastly, in grapheme-in-phoneme module an exact pronunciation of the input words is defined and the prosodic model is responsible for generating the correct prosody for the synthesized speech.

The next phase is prosodic modelling, which in the linguistic science includes intonation, rhythm and lexic accentuation in the speech. The prosodic characteristics of the speech unit can be grouped in the characteristics of syllable, word, phrase or sentence.

The perceived quality of the synthesized speech is largely determined by the naturalness of the prosody generated during the synthesis.

After finding the optimal solutions for adaptation of the specifics and characteristics of the Albanian language, on the basis of the above said for the way of work on multilingual text into speech system, we continue with the following stages: text normalization, creation of acoustic database and at last, text conversion with the help of applications for speech generating from text written in Albanian language. The whole process is given in Fig. 1.

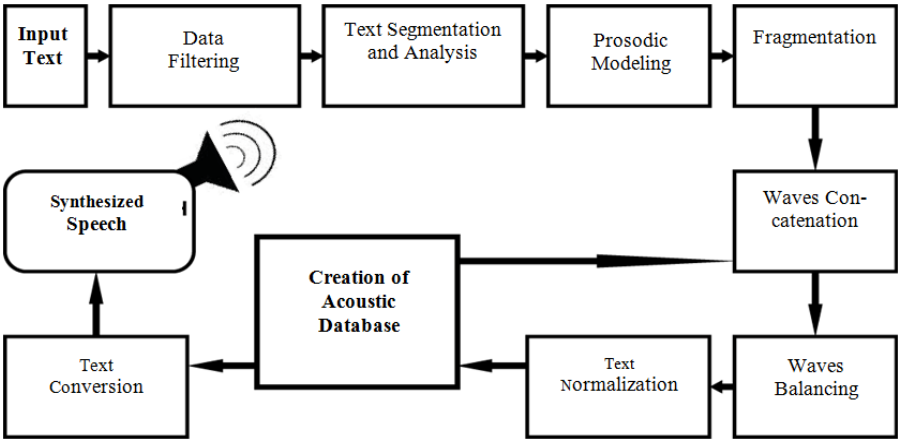

Figure 1. The user interfacing model.

IV. Conclusion

The existing accessibility tools based on intelligent interfaces are largely intended for the English speaking region. The need of this kind of assistance for blind and visually impaired in other regions logically urges for localization, which however cannot be performed in a straightforward translation manner due to the deep conceptual and vocal differences among languages. Initiating such a job inevitably addresses issues like specific language analysis, investigating existing solutions, careful selection of successful experience and best practices or ready-to-go concepts, all leading to appropriate modifications and adaptations directed towards efficient implementation in the particular language.

So the choice of building a particular non-English text-to-speech tool comes down to either create a new product or localization (through adaptation) of an existing solution. The latter requires language preparatory operations such as basic vocabulary or terminology databases with appropriate vocalization.

The proposed user interfacing model for Albanian speaking blind and visually impaired is based on an optimal solutions for adaptation of the specifics and characteristics of the Albanian language for work on multilingual text into speech systems. It has been successfully tested on Albanian speaking blind and visually impaired with WebAnywhere screen reader. The speech generated provides homogenous and stable speech quality fully understandable.

This user interfacing model is a very useful tool for Albanian-speaking blind and visually impaired users giving them an equal Internet access opportunities as the other users.

The research for the possibility of interpreting texts through a smaller number of basic units, introduced us in the meaning of diphones, as technical solution for this problem regarding this matter. The usage of more common words and diphones could be suggested as optimal solution within generating speech from written texts. This could be justified with the fact that through more common words a considerable percentage of texts can be included and continual speeches will be generated. For the other words the quality of generating thorugh diphones will be lower, but the system will be stable because there will be no words that cannot be composed of diphones. Only for the foreign words or those that cannot be found in the dictionary of the Albanian language, there will be the option of generating speech through specific characters.

References

[1] J. J. Bigham, T. Lau, and J. Nichols, "TrailBlazer: Enabling blind users to blaze trails through the web," in *Proc. 12th International Conference on Intelligent User Interfaces*, New York, 2009, pp. 177-186.

[2] C. Asakawa, "What's the web like if you can't see it?" in *Proc. International Cross-Disciplinary Workshop on Web Accessibility*, New York, 2005, pp. 1-8.
[3] A. Sharma, A. Srivastava, and A. Vashishth, "An assistive reading system for visually impaired using OCR and TTS," *International Journal of Computer Applications*, vol. 95, no. 2, pp. 13-18, June 2014.
[4] J. P. Bigham, C. M. Prince, and R. E. Ladner, "WebAnywhere: A screen reader on-the-go," in *Proc. International Cross-Disciplinary Conference on WEB Accessibility*, New York, 2008, pp. 73-82.
[5] W. O. Galitz, *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques (CourseSmart)*, Indianapolis: Wiley Publishing Inc., 2007.
[6] D. J. Mayhew, *Principles and Guidelines in Software User Interface Design*, NJ: Prentice Hall Inc., 2008.
[7] J. Johnson, *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules*, Burlington, MA: Elsevier Inc., 2010.
[8] S. R. Feldman, *The Text Mining Handbook - Advanced Approach in Analyzing Unstructured Data*, Cambridge Press, 2007.
[9] Z. Gormez and O. Zeynep, "TTTS: Turkish text-to-speech system," in *Proc. 12th WSEAS International Conference on Computers*, Heraklion, Greece, 2008, pp. 977-981.
[10] K. Parssinen, "Multilingual text-to-speech system for mobile devices: Development and applications," Doctoral dissertation, Tampere University, Tampere, 2007.
[11] P. Taylor, *Text-to-Speech Synthesis*, Cambridge University Press, 2009.
[12] N. Obin, C. Veaux, and P. Lanchantin, "Exploiting alternatives for text-to-speech synthesis: From machine to human," in *Speech Prosody in Speech Synthesis: Modelling and Generation of Prosody for High Quality and Flexible Speech Synthesis*, K. Hirose and J. Tao, Eds., Berlin, Heidelberg: Springer-Verlag, 2015, ch. 13, pp. 189-202.
[13] T. Boros, D. Stefanescu, and R. Ion, "Handling two difficult challenges for text-to-speech synthesis systems: Out-of-Vocabulary words and prosody: A case study in Romanian," in *Where Humans Meet Machines: Innovative Solutions for Knotty Natural-Language Problems*, A. Neustein and J. A. Markowitz, Eds., New York: Springer Science + Business Media, 2013, pp. 137-162.
[14] K. S. Rao, *Predicting Prosody from Texts for Text-to-Speech Synthesis*, New York: Springer Science + Business Media, 2012.
[15] O. Piton, K. Lagji, and R. Përnaska, "Electronic dictionaries and transducers for automatic processing of the Albanian language," *Lecture Notes in Computer Science, Natural Language Processing and Information Systems*, vol. 4592, pp. 407-413, 2007.
[16] M. Hamiti and A. Dika, "Speech generation for albanian written texts," in *Proc. 32nd International Conference on Information Technology Interfaces*, Cavtat, Dubrovnik, Croatia, 2010, pp. 85-90.
[17] M. Hamiti and A. Dika, "Analyses of same content texts written in different languages," in *Proc. 31st International Conference on Information Technology Interfaces*, Cavtat, Dubrovnik, Croatia, 2009, pp. 527-532.

Vanco E. Cabukovski is Full Professor of Software Engineering at Faculty of Natural Sciences and Mathematics, Sts. Cyril and Methodius University in Skopje, Republic of Macedonia. His main research interests include intelligent systems, e-learning systems and information systems. He has published over 70 scientific papers and over 25 books in informatics and ICT. He is author of 30 software applications and has participated in over 20 domestic and international projects.

Valbon Ademi is an Assistant Professor of Computer Sciences with the Faculty of Natural Sciences and Mathematics, State University of Tetovo, Tetovo, Republic of Macedonia. His main research interests are focused on intelligent systems and intelligent user interfaces.

Roman V. Golubovski is an Assistant Professor of Software Engineering with the Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia. His main research interests are focused on signal acquisition and processing, computerized automation and information systems, and intelligent systems. Assist. Professor Golubovski is author and co-author of 2 books and more than 33 papers. He is an author of 7 software and hardware development project applications in various areas, and participant in more than 50 application projects as well as in 4 domestic and international research projects.