ALBANIAN-SPEAKING VISUALLY IMPAIRED INTERFACING MODEL

Lindita ADEMI, Valbon ADEMI

Faculty of Philology, University of Tetova, Republic of north Macedonia, lindita.ademi@unite.edu.mk
Faculty of Natural Sciences and Mathematics, University of Tetova, Republic of north Macedonia, valbon.ademi@unite.edu.mk

Abstract

Being able to read text, find out information and know about the latest news has always been a challenge for those who cannot access the printed version, such as the blind and visually-impaired. The advent of internet and advancements in User Interfaces (UIs) as well as in assistive technologies make this challenge increasingly feasible. A lot of effort has gone into making software applications and web pages more accessible to the blind and visually impaired students. There are few Windows based stand-alone text-to-speech screen reader applications that provide application-specific speech output through script files written specifically for them. All these screen readers usually cover widely speaking languages. Albanian language is not covered with no one of the screen readers. In this paper a new user-interface model for Albanian speaking blind and visually impaired is presented. The new model is built up on the research results and testing with different screen readers and different groups of blind and visually impaired Albanian speaking students, the screen readers analysis, text-to-speech technologies analysis, specially developed text-to-speech generator from Albanian texts, the mathematical analysis done on general Albanian texts as well as the basic principles of lexicons creations. It has been tested to blind and visually impaired science students. The new user-interface model is a very useful tool for Albanian-speaking blind and visually impaired science students giving them an equal studying opportunity as the other students.

Keywords: Interface, blind, Albanian, text to speech, synthetized speech.

Introduction

Intelligent User Interfaces (IUI), or sometimes called an interface agent is a user interface (UI), that includes some aspects of the artificial intelligence. The area of intelligent user interfaces covers various topics and deals with application of artificial intelligence and knowledge-based techniques on questions of man – computer interaction [1] [2] [3]. One of the major benefits from the rapid development of the intelligent user interfaces are the applications known as (screen readers), which to a greater extent ease the life of the blind and visually impaired people. [4][5].

Having in mind that today the information science and computers, including internet, are entered in every pore of human life and work, the necessity of being at disposal to the disabled people including the blind and visually impaired is more than obvious. The screen readers are text-in-speech systems that transform the text into speech and read the information presented on the computer display. With a help of a combination of keys (shortcuts on the keyboard) the user can move across the user interface and read all the texts available on the screen. With a help of the keyboard, the user can enter a text that is transformed by the screen reader into speech and read loud. One of the screen readers is Web Anywhere [6]. The speech synthesis is an artificial generating of human speech from written texts. The speech synthesis is the central part of the text-in-speech technology, which today has a wide range of applications with a huge significance of their application in auxiliary technologies and tools for blind and visually impaired people and dyslexia, where the screen readers belong. For this purpose, there are appropriate algorithms designed, which afterwards through relevant programs enable to

synthesize the text into speech. At the moment, there are solutions that enable generating natural speech for different world languages. [7][8] [9] [10].
However, they are not universal solutions which can be used for other languages as well, since the speech generated for other languages is not recognizable and natural [8][9]. Because of this, for each language one should look for a solution that refers to its specifics, always aiming towards a voice creation that will respond to the nature of the language. In this study, we describe the basic principles of designing a system for speech synthesis of the Albanian language from written texts. As basic segments that are used for the Albanian language are: the most used words, two letters and separate 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. After the conducted analysis and tests it became clear that the generated speech through letters was very unnatural, as a result of the discontinuity during the 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, we first provide acoustic files of the most commonly used words, which cover a considerable percentage of the written texts, and for other words the generating of the sound 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 withingenerating 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 through 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.

Generating Speech From Written Texts In Albanian Language

At the moment, 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 и 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. Two common assumptions are not true when it comes to the Albanian language: all roots are

not in the sentences and very often lack of syntax information. This unusually increases the importance of the processing of “unknown words”. We have to stress that many verb forms are not clean, but they are “analytical forms” [13].

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: for example, laj (wash) и lahem (be washed). Both forms have a same past participe, 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. This is not written in the vocabulary. 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 Shakespeare, 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 declinations change depending on the position of the uttered adjective or noun in the syntagma in nominative.

From the conducted analysis of the Albanian written texts, it can be concluded 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 36x36 = 1296 possible diphones. 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, we can
conclude that only 45 of them can cover 50,532% 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.

2.1. Application for speech generating from Albanian written texts: The design of the application for speech generating from Albanian written texts consists of three

phases:
1. Normalization of the text,
2. Creation of acoustic database and
3. Conversion of the text
Textual normalization means transformation of the text in the most adequate form for further processing. For the Albanian language, the normalization is realized in three steps [11][12]:
1. First, the entire text is converted in small letters from the alphabet.
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.
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 cipher by cipher. Further segmentation of the text will continue through punctuation symbols. Some of them, such as full stop, question mark, exclamation mark will be equal to the pauses of 1000 ms. Thus, the speech generated later will not respect the intonations, but it will be comprehensive and continuous, which is the goal by itself.
From the different approaches to the speech generating, at the same time respecting the proposed
algorithms in the context of this paper and the perception of the generated sound by the designed
application, we can come to the following conclusions:
1. Generating speech letter by letter is comprehensive but it is not natural and connected to difficult discontinuities in the speech.
2. 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.
3. The use of diphones as basic units is characterized with the fact that it provides homogeneous and stable solution, as any text can be interpreted through a combination of acoustic added files to 892 identified diphones.
4. A possibility which offers compromised solutions, from the aspect of speech quality that is generated and also from the aspect of optimization 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.

Design Of A New Model Of User Interface For Albanian

Speaking Region

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 choose 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 preprocessing the text, it is usually first divided into sentences, then divided into tokens which are separated by characters of blank space. The module for preprocessing 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 analyzed 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.

Conclusion

In the group of existing intelligent interfaces for accessibility, with certain exceptions, dominate those in English. Starting the initiative for effectuating such kind of interface, inevitably comes down to previous analysis of the existing solutions, taking over positive experiences and best practices, or ready-to-go concepts, which inevitably leads to certain modifications and adaptations directed towards efficient use on adequate language region.
The choice of option for intelligent interface for blind people from non-English speaking regions comes down to two options – creating a new product or localization of some existing solution.
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.

References

[1] Galitz O. Wilbert, The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques (CourseSmart), Wiley Publishing 2007.
[2] Mayhew J. Deborah, Principles and Guidelines in Software User Interface Design, Prentice Hall, 2008.
[3] Johnson Jeff, Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules, Elsevier, 2010.
[4] 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.
[5] 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.
[6] Bigham P. Jeffrey, Prince M. Craig, Ladner E. Richard, Web Anywhere: 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.
[7] Feldman R., Sanger J., The text mining handbook - advanced approach in analyzing unstructured data, Cambridge Press 2007.
[8] Gormez Zeliha, Orhan Zeynep, TTTS:Turkish text-to-speech system, 12th WSEAS International Coneference on Computers, Heraklion Greece, pp. 977-981, 2008.
[9] Parssinen Kimmo, Multilingual text-to-speech system for mobile devices: development and applications, Tampere University, Tampere 2007.
[10] Taylor Paul, Text-to-Speech Synthesis, Cambridge University Press, 2009.
[11] 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.
[12] 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.
[13] Ademi V., Напредни Техники за развој на интелигентни кориснички интерфејси, 2019, ISBN 978-608-245-283-8.
[14] Ademi L., Терминологијата од областа на информатичко компјутерската технологија во англискиот и во албанскиот јазик-споредбена анализа, ISBN 978-608-245-412-2, 2019.