Adapting eSpeak for converting text into speech in Albanian

Mentor Hamiti[1], Ramiz Kastrati[2]

**[1] South East European University
Tetova, 1200,Macedonia**

**[2] College Universum
Ferizaj, 70000, Kosovo**

Abstract

Currently there are different types of technology conversion of text to speech (TTS), but the problem is still treated as unsolved. Additional difficulties pose differences in writing and speaking across languages, which limit the use of such universal converter for all languages. Therefore, the research in this field is reasonable and of particular importance, especially for the languages of the different nationalities, following the benefit of the global trend. Such is the case with the Albanian language. In this paper we have explored the possibility of adaptation of the existing text to speech converters, into Albanian language. We have selected eSpeak, as a very popular application with free access, which enjoys the support of several different languages. For generating and improving words in order to adapt them to Albanian is used eSpeakedit, whereas for the analysis and reconstruction of acoustic signals is used Praat software. The results achieved for the benefit of spoken Albanian through eSpeak, are presented within this paper.

Keywords: *TTS, eSpeak, praat, Albanian language.*

1. Introduction

Text-To-Speech (TTS) synthesis means converting written text into understandable and natural speech [6]. The latest development in technology have allowed developing of different software for converting written text into speech. In the last years efforts are being made so the Albanian language can join the group of languages that can be read by technological devices. But although the Albanian language has phonetic rules and belongs to the languages with a well defined structure, these results are in the early stages [7].

The equipment for speech synthesis based on technology in use, are categorized as follows:
- Articulatory synthesis
- Formant synthesis and
- Concatenative synthesis [7]

All three technologies are currently subject to research, however Formant synthesis and Concatenative synthesis are the two main strategies and more hopeful for advancement. The difference between the two technologies is that the Formant synthesis synthesizes the artificial voice by using a specific acoustic model for speech generation, while Concatenative synthesis is based on a series of segments of previously recorded words.

2. eSpeak Software

The software that has a wide use in TTS is eSpeak. This is open source software and under the license agreement allows intervention in order to improve speech in a given language.

eSpeak software is a GUI program that is used to create phonetic records by using a random text. In fact, eSpeak is a program which is used to prepare and compile phoneme data. In this way clear speech is generated at different generation speed and, but the speech is not so natural, as in the case of registration of human speech [5].

In addition eSpeak has basic tools for adjusting the text, supports DDE servers, that allow other applications to send the text and allowing users to be more flexible during usage [5].

The main features of TTS eSpeak are the following [3]:

- Includes different voices, whose characteristics can be altered.
- Can produce speech output or a WAV file.
- SSML (Speech Synthesis Markup Language) is supported (not complete), and Also HTML.
- Compact size. The program and its data, including many languages, Totals about 1.4 Mbytes.
- Can be used as a front-end to MBROLA diphone voices, see mbrola.html. eSpeak converts text to phonemes with pitch and length information.
- Can translate text into phoneme codes, so it could be adapted or a front end for another speech synthesis engine.

- Potential for other languages. Several are included in varying stages of progress. Help from native speakers for these or other languages is welcome.
- Development tools are available for producing and tuning phoneme data.
- Written in C

eSpeak software also provides possibility to edit its content and to analyze a certain number of frame-sequences in editor through imported or generated files [5], as well as allows the selection of the types of sounds.

TTS software is highly advanced for specific languages such as English, German, and French. As for the Albanian language which is part of provisional languages, where "sq Albanian, there is some initial feedback, but needs more work."[3]

However, this system can also analyze files generated from outside, and also enjoys MBROLA support [3]. But since the Albanian language has no registered database in MBROLA, in the context of this paper we analyze only the generation of audio files from eSpeak.

While for editing purposes was used espeakedit software, for acoustic signals analysis is used praat software, and this because both are free software.

Espeakedit software serves to prepare phoneme data for eSpeak which is a TTS. For analyzing the acoustic signals generated by eSpeak, as well as for processing signals which will be part of the eSpeak fund, is used Praat software.

Praat is a software intended to make the analysis and synthesis of words on the computer. The Praat segmentation, labeling and in particular the signal manipulation and re-synthesis facilities are unquestionably superior to any corresponding functionality that currently exists [4].

3. Adaptation of eSpeak for Albanian language

Albanian alphabet is very similar to IPA (AFN).Digraphs are exception, for which IPA uses special signs [R]. Upon entering a foreign word in Albanian, foreign phonemes suit Albanian phonetic system [9].

Today's Albanian language has its own sound system. These sounds are called phonemes (from Greek: phone - voice). The Albanian language has a total of 36 letters 29 consonants and 7 vowels [1]. 7 vowels: a, e, ë, i, o, u, y

and 29 consonants: b, c, ç, d, dh, f, g, gj, h, j, k, l, ll, m, n, nj, p, q, r, rr, s, sh, t, th, v, x, xh, z, zh.

Table 1: Phonetic transcription of the Albanian alphabet by the International Phonetic Alphabet [8]

Letter:	**A**	B	C	Ç	D	Dh	**E**	**Ë**	F	G	Gj	H	**I**	J	K	L	Ll
IPA value:	A	B	$\widehat{ts}$	$\widehat{t\int}$	d	ð	e	ə	f	g	ɟ	h	I	j	K	L	ł
M	N	Nj	**O**	P	Q	R	Rr	**S**	Sh	T	Th	**U**	V	X	Xh	**Y**	Z	Zh
M	N	ɲ	O	p	c	ɾ	r	s	∫	t	θ	u	v	d͡z	d͡ʒ	Y	Z	ʒ

Vowels are sounds that are articulated in the mouth without encountering any obstacle, while consonants are sounds that are articulated in mouth by overcoming an obstacle.

Translation of Albanian language letters in SAMP based on the data in the sq_list file are given in the table.

Table 2: SAMPA (Speech Assessment Methods Phonetic Alphabet) adapted for Albanian

Letter:	**A**	B	C	Ç	D	Dh	**E**	**Ë**	F	G	Gj	**H** **I**	J	K	L	Ll
SAMP:	A	b	ts	tS	d	D	e	@	f	g	dz;	h I	j	K	L	ll
M	N	Nj	**O**	P	Q	R	Rr	**S**	Sh	T	Th	**U**	**V** **X**	Xh	**Y**	Z	Zh
M	N	N^	O	p	c	R	RR	s	S	t	T	u	v	dz	dZ	Y	z	Z

Syllables are generally formed by the combination of a vowel and a consonant and the other combination is with three letters that contain a vowel and two consonants with a few exceptions.

Albanian language spelling is phonetic, i.e. generally a letter corresponds to a phoneme, or simply said, the words a read as written and the adjustment of the Albanian alphabet with IPA alphabet is realized according to Table 1.

Since we noted that for the Albanian language there is no database for MBROLA, and then the reading of texts is not natural but when using this software, to those who are beginners in the use of the same, it may seem to be a little annoying because it generates a robot like sound that differs from the natural one.

To make interventions aimed at improving eSpeak software we should keep in mind that it contains a list of constraints of phonemes, a file for *_rules and *_list [3]. A phoneme table does not have to specify all phonemes used

by a language but it can inherit a phonemes by previously defined phoneme table.

During the use of espeakedit in a phoneme table of a certain language we can insert rules for a phoneme, but also we can use phonemes from base tables.

Phoneme table defines all phonemes which are used by a language, along with their characteristics and data for their production, how it sounds [3], while for the Albanian language we notice that majority of phonemes are same as those used in English. By having information on the phonetics of the English language we come to the conclusion that the intervention for improvement of these phonemes is a necessary thing to do.

In addition we present some cases of the text analysis from eSpeakedit and Praat software. During the analysis were noticed the following types of problems which are divided into eight groups:

1. Incorrect complete pronouncement of some letters.
2. Sounds that that look like pronounced letters of the Albanian language but not fully accurate.
3. Problems in exact reading when two vowels are next to each other or cases when they are combined with other letters.
4. Incorrect pronunciation of some numbers.
5. Incorrect pronunciation of some symbols.
6. Lack in reading some symbols.
7. Lack in reading the Greek alphabet.
8. Improper placement of the accent in a word

The problems encountered during this analysis and the interventions for improving these defects are shown in groups:

3.1 The first group

In the first group during the analysis were noticed problems in pronunciation of these letters:

Letter "a"

Pronunciation of the letter "a" at the end of the sentence is not acoustic signal that corresponds to this letter, since it has the form of sound "ë". Therefore eSpeak did not read the words properly, where the pronunciation of the vowel does not match the correct pronunciation in Albanian. Acoustic signal analysis was done with the help of Praat software and was noticed that the last letter of the word "shkolla" is read "ë".

In order to improve the phoneme table we selected a signal that generates voice that responds to the letter "a".

The adjustments realized for this group have to do with setting the conditions for the selection of the type of acoustic signal that corresponds to letter "a" when the vowel is short. The appearance of the generated signal by eSpeak for the word "shkolla" is presented in Figure 1 and Figure 2.From the graphical presentation we can see the difference between the two analyzed cases.
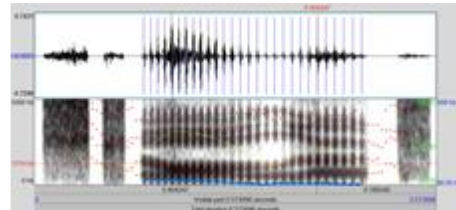
Fig. 1 The word "shkolla" before the intervention
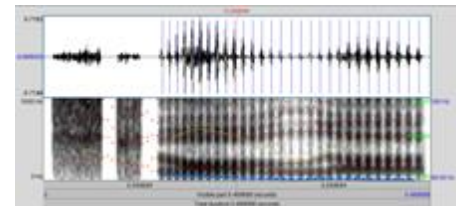
Fig. 2 The word "shkolla" after intervention

In addition is presented the part of the code for proper reading of the letter "a" for generating the sound signal

```
phoneme a  // shkronja a
  vowel  starttype #a  endtype #a
  length 170
  IF thisPh(isDiminished) THEN
    FMT(vowel/a)
  ENDIF
  FMT(vowel/a_3)
endphoneme
```

The letter "c"

The letter "c" at the beginning of words is very similar to the letter "s" and while reading the text, the words that begin with this letter are pronounced incorrectly. For example: "centimetri". To eliminate this, the signal that is selected for generating the sound for the letter "c" is WAV (ustop/ts, 50), while in the phoneme table is set according to the rules.

```
phoneme ts  // shkronja c
  vls pla afr sibilant
```

```
lengthmod 2
voicingswitch s
Vowelin f1=0  f2=1900 100 200  f3=-100 80
WAV(ustop/ts, 50)
endphoneme
```

The letter "r"

It is noted that many times when hearing the reading of this letter it seems like the acoustic signal generated by reading the letter "rr".

The selection of an acoustic signal that would be the same with the letter "r" in the eSpeak software folder that contains phonemes, is not the sound that will meet our requirements.Therefore, for creating the sound that is identical to the acoustic signal, we registered syllable "ra", from which we separated the sound by using Praat software and through the given code we attached it to the group of phonemes for Albanian language in the eSpeak phoneme table.

Fig. 3 Recording the syllable"ra"
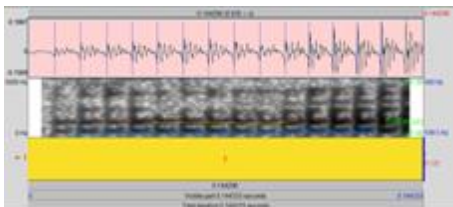
Fig. 4 Separating the letter "a" from syllable"ra"

The separated sound from this syllable based on the rules set by espeakedit is inserted in phonemes table.

The letter "g"

While reading this letter from eSpeak software, an identical sound with "k" is generated.Same as in the cases of problems with the sounds of other letters, the file that corresponds most to the sound of the letter "g" has been selected.For example. "guaca" "gu'atsa".

FMT (g/xg)

3.2 The second group

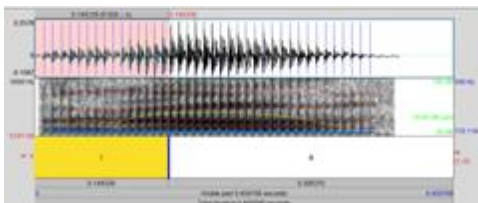In the second group there are number of letters that through reading generate sounds that are like many sounds generated by the pronunciation of the letters of the Albanian language and the native Albanian speakers notice these errors.

For example, the letter "L" while reading this letter by eSpeak we notice problems when after the letter "l" comes the letter "j".A same problem also occurs to the letters e, i, u, t, s, q, z, sh, dh, th.

For improving this TTS software we analyzed the pronunciation of the letters that are inside the software and with the help of espeakedit we selected the certain intonation for these letters, for example, for the vowels "e", "i", "u", appears the problem of the length of the sound, in the cases when the accent falls on these letters and for this reason the letter "e", when the accent falls on this letter the registration "vowel/e" is taken, length 200 while when the accent does not fall on this letter then is taken the registration "vowel/e_mid" length 140.

The same substitutions were made and for the letters. The letter "u", length 140.

"FMT (vowel/u_bck)" and when the accent falls on this vowel length is taken 2000 "FMT (vowel/u_2)".For the letter "i", length 160 "FMT (vowel/i_6)", when the emphasis falls on the vowel length is 200 Marree "FMT (vowel/ii_6)" so that the duration of the acoustic signal to sound and speech language fits the Albanian.

3.3 The third group

In this group are encountered the problems in reading the words where in this sentences are located. For example: "ua" that is read as "wa", "ie" that read as "je".But 1 the ërehen v errors while reading the syllable "ay" to read as "he".

"munguan"   "mungu'an" , ""vlerësoj ""vl'ER@sOj" , " ajo" "aj'O"

Solving this problem is achieved by removing certain groups of letters such as these diphthongs since these diphthongs are used by the English language, and it is known that the Albanian language is not similar to the one from the phonetic aspect.

After the intervention on the groups of letters we can notice that the result meets the requirements. For example,

for the word "munguan" the vowels "ua" now are read as "ua" and not as one above.It is the same for both taken cases for demonstration.

"munguan" "mungu'an" , ""vlerësoj ""vl'ER@sOj" , " ajo" "aj'O"

3.4 The fourth group

In this group are placed the reading of numbers, where are noticed irregular spellings of numbers 1,9,11,19 and numbers that contain these figures. "1" "n^'I", "11" and "n ^'I".

Intervention for improving the pronunciation of these defects on these numbers and their combinations is implemented in sq_list. In addition we present improvements in reading of these numbers and their combination.

"1" "n^'@ "11" "n^,@mb@Dj'et@ "

3.5 The Fifth Group

Although in sq_list file are located a considerable number of usable symbols while reading textbooks, it is noticeable that the translation of these in Albanian language has many problems.

Table 3: Symbols found in the file sq_list

	*	+	.	=	!	%	@	&	€	#	~	^	,	-	:
;	>	<	?	"	(	)	]	[	{	}	\	#	tab	$	"

For example. "5>2" "p'Es@ _:_: d'y", "5%" "p'Es@ p@tS'indja", "2#" "d'y h'aS".

After the intervention, there have been improvements in the pronunciation of these symbols which were with problems. With the intervention for eliminating these problems we got these results. "5>2" "p'Es@ m'@Em'aDE se d'y ", "5%" "p'Es@ p@RtS'ind@ "2#" "d'y tar'ab@"

3.6 The sixth group

Besides symbols that are in sq_list, there is also a considerable number of symbols used and are unable to read and pronounce in Albanian. These symbols are shown in the table below:

Table 4. Symbols that are added to the sq_list file

Symbol	©	¤	§		°	'	—	"	•	"	‚	‰		&	«
				µ	¶	'	→	"		"	»		‹		»
∀	®	™	`	'	∴	⇒		≥	≤	±	÷	≠	≈	∩	∪	⊇	⊄	
⊆		⊂	∈	∉	∠												⊒

In this set of symbols, in the appropriate file were inserted the adequate symbols and the reading that fits to the Albanian language.For example. "6 ‰ "dZ;'aSt@ pROm'ilAd'y pROm'ilA"

3.7 The seventh group

By reading a large number of textbooks, the need for pronouncing the Greek alphabet letters appears. The list with Greek letters is taken from the "en_rule" file and it is presented in the following table.

While reading these letters, errors are noticed in the majority of the letters. For example: "α " "_:(en)'alf@_(sq) ", "β " "_:(en)b'i:t@_(sq)", "π" "_:(en)p'aI_(sq)"

The Greek letters	α	β	γ	δ	ε	έ	ζ	η	ή	θ	ξ	ι	ί	κ	λ
ά	μ	ν	o	π	ρ	σ	ς	τ	υ	ύ	φ	χ	ψ	ω	ώ

Same like toward the special symbols and for Greek letters is undertaken the same procedure for the introduction of these letters, naming in Albanian.For example. "α " " 'alfA", "β " "b'EtA", "π" "p'i"

3.8 The eighth group

The problem of placing the accent while reading the text from eSpeak is quite significant. Wrong placement of the accent in question is quite problematic not only for Albanian but also for the majority of the languages that support eSpeak.

Since the syllable is a group of phonemes centered on a vowel, problems were noted in placing the accent on the word, although letters are pronounced right, placing the accent in wrong place does word deformation.

But the original purpose of this study was the possibility of intervention in the pronunciation of the letters in the Albanian language and the establishment of rules that characterizes this language.

Since syllables in Albanian language are created by a vowel and a consonant and word emphasis falls on the vowel, depending on where the accent of the word falls, the Albanian language distinguishes these cases. Words with end accent, penultimate, and third syllable from the

IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 4, No 2, July 2014
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org

26

end. But even a single voice may compose a syllable. In the spoken language, because of not pronouncing the unstressed "ë", often open syllables are becoming closed.

Considering that the Albanian language has a total of four diphthongs (ie, ye, ua, ue), where these diphthongs and vowels form a syllable. Since majority of the words in Albanian language are with penultimate accent then appears the need for creating a database where are located all the words of the Albanian language and the place where is located the accent in question. But there is no electronic dictionary for Albanian language, for this reason, the intervention for adjustment of these vowels is done for the case when they are aligned and unstressed.

4. CONCLUSIONS

TTS every day more and more is taking place at a large number of users who have problems with visual impairments, children with special needs, reading messaging application, but also and from companies that deal with the production of highly intelligence technology. Given that up to now there is no commercial software for TTS in Albanian, eSpeak software with the interventions and improvements that have made, will be able to help these people in need.

After adjusting the pronunciation of the letters and arranging emphasis on examples above shows that sound generated by the eSpeak software meets the requirement to be understandable in Albanian.

The improvements are quite significant since the managers of the software will be able to incorporate these improvements within the software with a new version, which has been one of the goals of this paper.

Introduced

Reading the words of eSpeak software with these improvements sounds pretty good for the Albanian language, while for further work it remains to work in introducing a dictionary of Albanian language in order to establish the emphasis on each word, as well as the introduction of rules and improving the phoneme table for Albanian language. Also the creation a database in Albanian language for mbrola will enable the Albanian language to be part of the group with digitized languages.

References

[1] IPA, The international Phonetic Association, Available: http://www.langsci. ucl.ac.uk/ipa/fullchart.html Accessed: 19.12.2012

[2] Shaban Demiraj. "The Origin of the Albanians", Academy of Sciences of Albania, Tirana, 1988.

[3] eSpeak, Available: http://espeak.sourceforge.net/index.html, Accessed: 05.05.2013

[4] Jonathan H, Steve C., Tina J. and Michel S." Building an interface between EMU and Praat: a modular approach to speech database analysis", 15th ICPhS Barcelona, ISBN 1-876346-48-5 © 2003 UAB

[5] Ramiz Kastrati, Mentor Hamiti. "The opportunity of using eSpeak as Text-to-Speech synthesizer for Albanian Language", , Tirana,15.06.2013

[6] Bickley C, Syrdal A, and Schroeter J. (1998). Speech Synthesis: in the Acoustics of Speech Communication, Picket J.M, Ed., Boston, NY:Allyn and Bacon.

[7] Lemmetty S. "Review of Speech Synthesis Technology", MSc. Thesis, Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland, 1999.

Mentor Hamiti, holds a PhD diploma in Faculty of Contemporary Sciences and Technologies at South East European University (2010). Currently he is Assistant Professor at the same University, R. Macedonia (www.seeu.edu.mk).

Ramiz Kastrati, holds a MSc diploma in Faculty of Contemporary Sciences and Technologies at South East European University (2012). Currently he is Assistant at the University College Universum, R. Kosovo.