Development overview of TTS-MK speech synthesizer for Macedonian language, and its application

Slavcho Chungurski [1], Sime Arsenovski [1], Dejan Gjorgjevikj [2]

[1] Faculty of informatics, FON University - Skopje
[2] Faculty of Computer Science and Engineering, University of "Sv. Kiril i Metodij" - Skopje
{chungurski@fon.edu.mk, sime.arsenovski@fon.edu.mk,
dejan.gjorgjevikj@finki.ukim.mk}

Abstract. This paper shows the current results of development of TTS-MK – a speech synthesizer for Macedonian language. The basic principles for projecting and building of speech synthesizer for Macedonian language, based on concatenation of speech segments, are shown.

Every language has its respective and specific speech norms and characteristics that should be observed during the speech synthesis. The Macedonian language is phonetic; hence the normative pronunciation does not contain great difficulty, except in some special cases that should be taken into consideration.

The presentation also focuses on the accent in the Macedonian language, which is dynamic and positioned on the third syllable. The rules and regulations for the accent positioning in the Macedonian language can be easily derived, with some deviations that should be resolved.

There are two versions of the system based on different segments corpora. Both of them are presented, as well as their application.

Keywords: Text-To-Speech, Macedonian Language, TTS, TTS-MK, Orthoepy, Speech API – SAPI

1 Introduction

The systems which synthesize speech with connection of previously recorded speech segments take significant place among the systems for text to speech conversion (TTS systems). These TTS systems are called concatenative speech synthesizers. These systems are simple and they do not require deep knowledge of phonetic transitions and co-articulation effects, which is the case with other kinds of speech synthesizers based on rules defined by linguists.

There were some attempts for development of quality concatenative speech synthesizer for Macedonian language, but these developments were based on speech corpora for other Slavic languages, which resulted in unnatural intonation of the synthetic speech in Macedonian language. This paper includes a brief overview of the development of TTS-MK synthesizer for Macedonian language. Concatenative

adfa, p. 1, 2011.

speech synthesizers require setting of serious task for definition and recording of speech and its processing for extracting convenient speech segments. Consequently, this paper presents appropriate definition and development of speech corpus for Macedonian language, used in TTS-MK.

The general functional organization of speech synthesis system for Macedonian language is shown in Fig.1. As in [3], it is made of two modules:

- Natural Language Processing (NLP) module, that gets text as input, makes analysis of the text, create its transcription into phones and recognizes the prosodic elements of the input. The output of this module is symbolic information for the phones and the prosody for the input text.

- Digital Signal Processing (DSP) module, that gets the symbolic information for phones and prosody from the NLP module and after certain processing of the input gives synthetic speech as output.

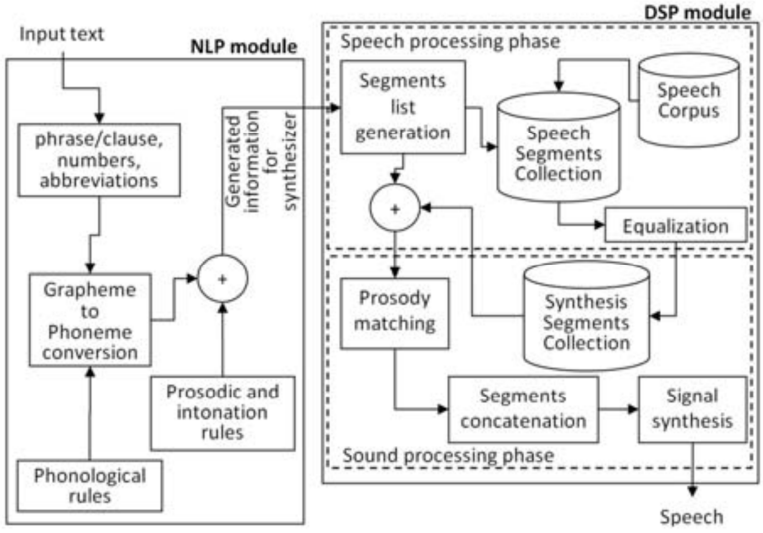

Fig. 1. Structure of general TTS system

The NLP module consists of three main parts: text analyzer, grapheme-to-phoneme unit and prosodic generator. Text analyzer is made of pre-processing unit where the input sentences are transformed into word arrays. This unit also identifies the numbers and abbreviations and transforms them into words. There is a module for morphological analysis that performs morphological analysis of the text for recognition of the affixes (prefixes and suffixes) that are added to the basic forms of the words. This module also confirms the correct accent of the words. This module of TTS systems for Macedonian language seems to be very complicated because in Macedonian language, unlike other languages, a plural of the words as well as words with an articles, adds whole syllables at their end. This can look very difficult to implement and it suggests to some complications in morphological analysis of the text. Thanks to antepenultimate word stress in Macedonian language and the fact that

this analysis is strictly for speech synthesis, this procedure is very simple. This is supported with the fact from [4], that in the Macedonian literature language there is no reduction of vocals, which means that the vowels sound very similar in their accented and unaccented form. As a result of this fact, this considerably simplifies the whole speech synthesis process for Macedonian language, because the vowels have only one sound variant, and not multiple, as in French for example (described in [3]). One of the main features of Macedonian language is its very simple rule for grapheme to phoneme transcription. This means that in Macedonian language, like in other Slavic languages, each letter represents one phone. So, the phonetization task for Macedonian language is reduced to trivial checks so solution techniques based on dictionaries and morphophonemic rules, like in English or French for example, are not necessary. The task of the prosodic generator is to supply naturalness to the synthesized speech. Apart from the fact that synthesized speech with natural intonation is more pleasant for listening, it is easier for understanding. Also, during a speech which is practically uninterrupted, the listener can easily recognize the bounds between the words. It is very important for a speech synthesizer to cover prosodic elements and proper intonation, which thanks to accentuation rules is not very complicated in Macedonian language.

The execution in DSP module is divided into two phases: speech processing and sound processing. Speech processing is one of the most important phases in TTS synthesis in general. It means that in this phase it is mandatory to make definition and recording of the speech corpus for Macedonian language, as well as its segmentation. In this phase, the symbolic information for the phones and the prosody for the input text are applied on the recorded speech corpus. The segments from the corpus and their concatenation order are also established in this phase. The collection of segments for speech synthesis obtained from the previous phase is the actual input for the next phase of DSP module - the sound processing. In this phase an adjustment of prosodic elements is performed. Here, as it was described in the part for text analysis, an accentuation is performed. After reconstruction of accentuation aggregate, f0 codebook is implemented in this phase. f0 codebook holds information for the f0 curves for each accentuation aggregate obtained by empirical way and the selection of appropriate curve is made on base on the position of the accentuation aggregate in the sentence (beginning, neutral, before comma, end).

2 Orthoepy of Macedonian language

Orthoepy is normative pronunciation of some standard language and its proper investigation leads to a better speech synthesis for the language. Macedonian language consists of 5 vocals (а /a/, е /ɛ/, и /i/, о /ɔ/, у /u/) and 26 consonants (б /b/, в /v/, г /g/, д /d/, ѓ /ɟ/, ж /ʒ/, з /z/, ѕ /dz/, ј /j/, к /k/, л /l/, љ /ɬ/, м /m/, н /n/, њ /ɲ/, п /p/, р /r/, с /s/, т /t/, ќ /c/, ф /f/, х /x/, ц /ts/, ч /tʃ/, џ /dʒ/, ш /ʃ/).

Every language has its respective and specific speech norms and characteristics that should be observed during the speech synthesis. The Macedonian language is phonetic, hence every voice corresponds to particular grapheme, so normative

pronunciation is not difficult. However, deviations do occur in written and spoken part of the language (discharging, substitution, inserting, voice replacement)[4]. These deviations must be taken into consideration according to the rules of the normative pronunciation during the speech synthesis. The mentioned aspects are considered within the NLP module of TTS-MK. These are some cases which are considered in TTS-MK:

- Double vocals pronunciation
  - Case: when adding a prefix of a word (Example: poodi → po|odi)
    - Solution: dictionary of prefixes
  - Case: when doubling is in the middle or the end of the word, and the accent is not on any of these vocals (Example: vikaat → vikāt; vakuum → vakūm, Exception: E, both vocals are pronounced separately)
    - Solution: processing module
- Vocal R
  - Case: R, where it is the leading grapheme followed by a consonant, is noted with ‘R (Example: ‘rž ; ‘rgja)
    - Solution: phoneme inventory contains @ (schwa)
  - Case: Adding a prefix, which ends with consonant, to words from the previous case leads to discharging of the sign (‘) in the notation, but not in the pronunciation (Example: srska, srža)
    - Solution: R → @R replacement
  - Case: R, where it has a role of a vocal when it is surrounded by consonants (Example: srce → s@rce; brza → b@rza)
    - Solution: R → @R replacement
- Consonants pronunciation
  - Case: Consonant sonority at the end of the word B → P; V → F; D → T; Dz → C; Z → S; Ž → Š; Dž → Č; G → K (Examples: grob →grop; gluv → gluf; led → let; bez →bes; nadež →nadeš; Džordž → Džorč; plug →pluk)
    - Solution: processing module
  - Case: TS → C; DS → C; SS → S; ZZ → Z; ŽD → ŠT (Examples: gradski → gracki; bratski → bracki; bessovesen → besovesen; bezzimen → bezimen; glužd → glušt)
    - Solution: processing module
  - Case: other isolated cases (Examples: ovca → ofca; vkluči →fkluči; gladta → glatta)
    - Solution: None

Macedonian accent is observed in the text analyzer and it can be processed with ease because it is dynamic and positioned on the third syllable (when more than two syllables), but there are some deviations. These cases are observed in TTS-MK:

- Adopted foreign words (Examples: foajè; birò; *ìzam, *ìst etc)
  - Solution: dictionary
- Accent aggregates (Examples: Kisela voda → Kiselàvoda; pri toa → prìtoa)
  - Solution (bad): dictionary, and then third syllable rule

3 Structure of the speech corpus of TTS-MK

Several continuous steps were performed for definition and creation of the speech corpus for TTS-MK[1]:

- Selection of speech segment types needed for the synthesis
- Definition of the set of the phonemes that covers all sound occurrences in Macedonian language
- Selection of the set of speech segments to be used that cover the whole phoneme set from the previous step
- Selection of the texts that cover whole speech segment set from the previous step

As a result of this procedure the following results are achieved:

- Segment types (Diphones, disyllables, whole accented words)
- Phoneme set (TTS-MK phonetic inventory adds ŋ /ŋ/ (allophone of n), ь /ə/ (schwa) and _ (silence))
- Selection of speech segments (34x34 = 1156 diphone units and 150 frequent disyllables)
- Selection of texts for recording of the segments (existing words from a text corpora and logatoms)

The elements required for speech synthesis in the TTS-MK are stored in several files shown in Table 1.

Filename	Protection	Contents
diphone.mk	Encrypted	Diphone set
disyllable.mk	Encrypted	Disyllable set
words.mk	Encrypted	Words set
abbreviations.mk	Open	Abbreviation list
accentAggregates.mk	Open	Accent aggregates list
accentExclusions.mk	Open	Accent exclusions list

Table 1. Storage of TTS-MK corpus

There are two versions of TTS-MK speech corpora.
TTS-MK Emma is the first experimental version. It includes 1156 diphones, 180 disyllables and 392 words from total 1196 recorded segments. It has poor dictionaries for abbreviations, accent exclusions and accent aggregates, contains lot of noise and includes low prosody elements. The whole corpus requires 56 MB of disk storage.

TTS-MK Lence is the second version and it is currently in development. It includes all 1156 diphones, plus 1161 disyllables and 392 words from a total of 1144 recorded segments. Because of its quality, and the quality of the speaker and the tone master, it can be commonly assumed that it is upgradeable. It includes rich dictionaries for abbreviations, accent exclusions and accent aggregates, without noise and with some

prosody elements. It requires 50 MB of disk space. This version is also SAPI 5.3 compatible with rate, pitch and volume adjustments and it is expected to be deployed as a basic tool for blind computer users. It also has some UI settings like Latin texts (spell or phonetic), numbers (digit by digit) and interpunction.

TTS-MK engine is developed completely in C# and .NET framework version 2.0. To interface this engine with SAPI a special engine wrapper was developed in C++. Figure 2 shows the application of TTS-MK for visually impaired persons.

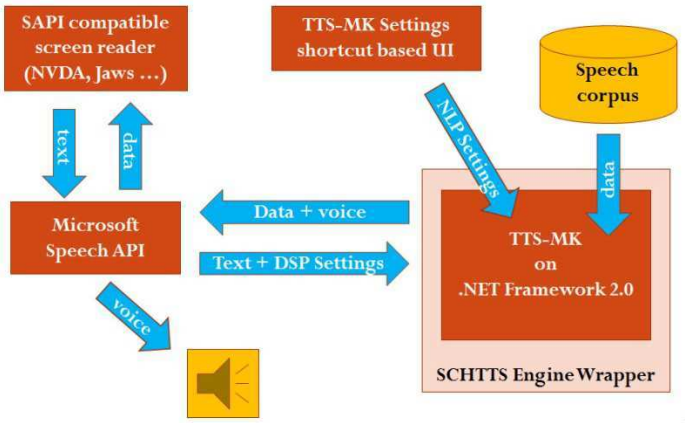

Fig. 2. Application of TTS-MK

4 Future work

Future work on TTS-MK can be separated in two parts. Improvements of NLP module of TTS-MK which will include improvements of orthoepy preprocessor, and improvements of DSP module of TTS-MK which will include unit-selection algorithms and upgrade of the corpus.

5 References

1. Chungurski, S., Kraljevski, I., Mihajlov, D., Arsenovski, S.: Concatenative Speech Synthesizers and Speech Corpus for Macedonian Language. 30th International Conference ITI Cavtat/Dubrovnik, Croatia. 669-674 (June 2008)
2. Chungurski, S., Kraljevski, I., Mihajlov, D., Arsenovski, S.: Evaluation of TTS-MK system for speech synthesis in Macedonian language. ETAI 2009, Ohrid. IE3-3. (September 2009)
3. Dutoit, T.: High Quality TTS Synthesis of the French Language. Ph. D. dissertation, Facultè Politechnique de Mons, France. (1993)
4. Koneski B.: Gramatika na makedonskiot literaturen jazik. Kultura, Skopje. (1987)
5. Chungurski, S., Kraljevski, I., Kakashevski, G., Mihajlov, D.: TTS approach to Macedonian Language. 16th Telecommunication forum TELFOR 2008, Belgrade. (November 2008)