Web-based Application of the Text-to-Speech Synthesis System “Speak Macedonian”

Branislav Gerazov, Toni Janevski and Zoran Ivanovski[1]

Abstract – **This paper describes an online presentation environment for the text-to-speech (TTS) system “Speak Macedonian”. The project included the generation of a standalone MATLAB executable of the TTS system and a website environment that would allow people to use it. The website environment was realized with a combination of an HTML website and a PHP script. The PHP script realizes the interface between the user and the system and also provides data logging so that every activation of the application can be logged for analysis.**

Keywords – **online presentation, MATLAB, PHP, HTML, Internet**

I. Introduction

The artificial synthesis of the human voice has inspired people’s work for centuries. Speech synthesis has truly come a long way, from the XVIII century mechanical contraptions of von Kempelen, the 1930’s mechanic VODER, the analog electronic synthesizers of the 1950’s, to today’s digital computer based systems. Development of speech synthesis in the digital era has been marked by three distinct paradigms: formant, articulatory and concatenative synthesis. Formant and articulatory synthesis are based on parametric modeling of the speech waveform and the speech production process, respectively. They served as the basis of TTS systems of the 1970’s up until the early 1990’s, [1]. Because of the difficult and expensive process of developing the high quality models used in these TTS systems, they were developed only for a handful of languages.

With the increase of computing power and memory, TTS systems of the 1990’s moved towards the relatively simpler paradigm of concatenative speech synthesis. Concatenative synthesis is based on straight-forward generation of the required output speech through the concatenation of small prerecorded speech segments, also called units. The simplicity behind this approach, spurred the creation of new TTS systems for various languages around the world. In the Slavic speaking part of the Balkans various systems have appeared, most prominent of which is the “AlphaNum” unit-selection TTS system for Serbian and later Croatian, [2]. Other systems include “SpeechLab” for Bulgarian and “Govorec” and “Proteus” for Slovenian.

Several attempts have been made to synthesize speech in Macedonian. First attempts of building such a system date back to 1997, [3]. Later trials emulated Macedonian using the Croatian diphone inventory under the MBROLA framework, [4], with a similar approach by the AlfaNum team, that used the system’s Serbian diphone inventory for the job, [5]. Today, two high quality TTS systems for Macedonian have been developed: TTS-MK, which has been finalized in 2009 [6], [7], and “Speak Macedonian”, our system, which is in its final stages of development, [8].

This article presents an online presentation environment for our TTS system. It is designed to bring our system closer to the public in Macedonia and the world. This is to serve both for promotional purposes but also for gaining user insight on various aspects of the synthesized speech quality. The environment is not designed for testing purposes. Final tests of the system will be carried out in a controlled environment in the facilities of our faculty, as work on the system reaches completion.

The core of the addressed problem is enabling a TTS system realized completely in the MATLAB programming language to execute on a server environment without a running MATLAB installation, and to build a web script able to pass arguments to and from the MATLAB application and the user. The presented solution was realized using MATLAB’s built-in compiler and a combination of HTML and PHP. In the following sections we outline the designed solutions.

II. The TTS sytem “Speak Macedonian”

The TTS system “Speak Macedonian” as presented in [8], is based on the paradigm of concatenative speech synthesis. It uses a mixed unit inventory with segment lengths of one and two voices. This offers quality speech output with an optimized size inventory. The system includes automatic pitch, duration and amplitude modeling of the segments providing for artificial prosody generation. It also includes a voice transformation module presented in detail in [9], that allows transformation of the default male voice to female, childlike and elderly, as well as various in-between modifications. The TTS system’s graphical user interface (GUI) is given in Fig. 1. It features a text field for text input and two pushbuttons – one for speech generation and one for repetition. It also includes controls for the voice transformation module.

[1] Branislav Gerazov, Toni Janevski and Zoran Ivanovski all are with the Faculty of Electrical Engineering and Information Technologies, Karpos II B.B., 1000 Bitola, Macedonia, E-mail: {gerazov , tonij , mars} @feit.ukim.edu.mk

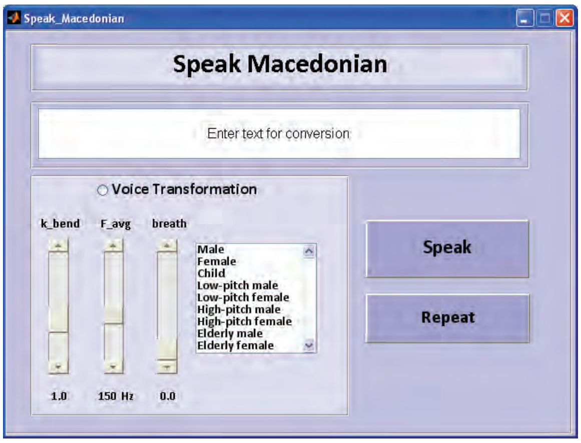


Fig. 1. The graphical user interface of the TTS system "Speak Macedonian"

An example of synthesized speech output is given in Fig. 2. The figure shows the time-domain waveform, spectrogram and the intonation contour of the synthesized word "makedonski" (eng. "Macedonian").

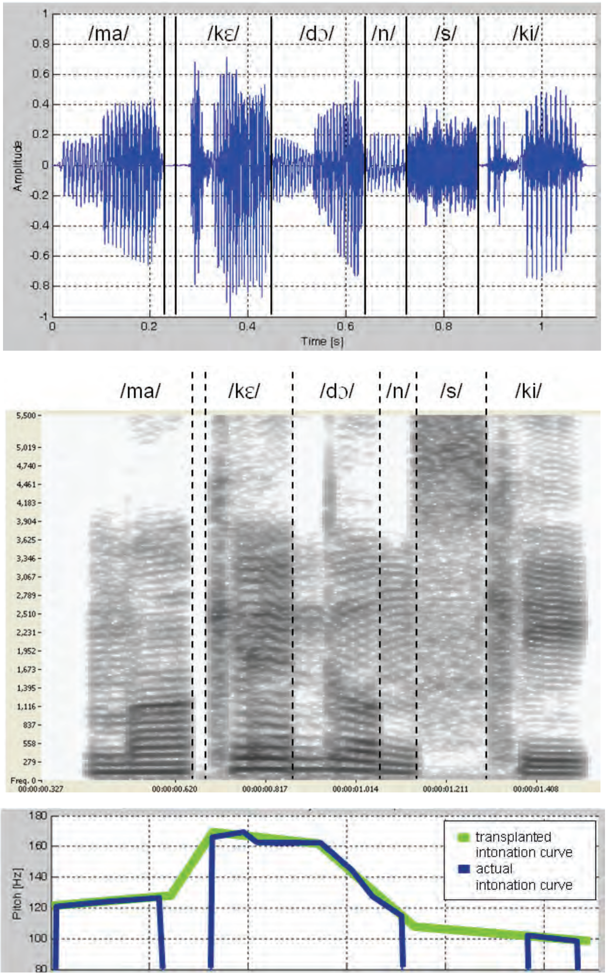


Fig. 2. Time-domain waveform (top), spectrogram (middle) and the intonation contour (bellow) of the synthesized word "makedonski"

III. Standalone MATLAB executable

For the purposes of relocating the TTS system to a server platform it was necessary to make it work without the need of a running MATLAB installation. This was carried out by generating a standalone executable file from the TTS system's MATLAB code. The executable file was generated using the built-in MATLAB Compiler, [10]. The principal purpose of the compiler is to offer distribution of MATLAB applications and libraries to users that lack a MATLAB installation. It can be used to compile m-files and other MATLAB code to standalone C/C++ applications for UNIX, Windows and Macintosh platforms as well as C/C++ libraries such as Windows' Dynamically Linked Libraries (DLL's).

Applications generated by the MATLAB Compiler consist of two parts: a binary executable file and an archive file. The executable file contains the project's main function, and is created using wrapper files that nest the m-code and provide an interface to it. The archive file or the Component Technology File (CTF), contains the user created MATLAB functions and the data files used by the application.

The MATLAB compiler can be easily operated through the Deployment Tool GUI shown in Fig. 3. The compilation process generates C/C++ source code and the EXE and CTF files that are needed for the distribution of the project. The two files can additionally be packaged by the Deployment Tool in a self-extracting archive file.

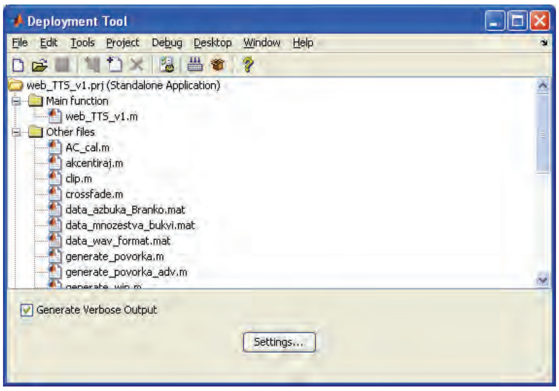
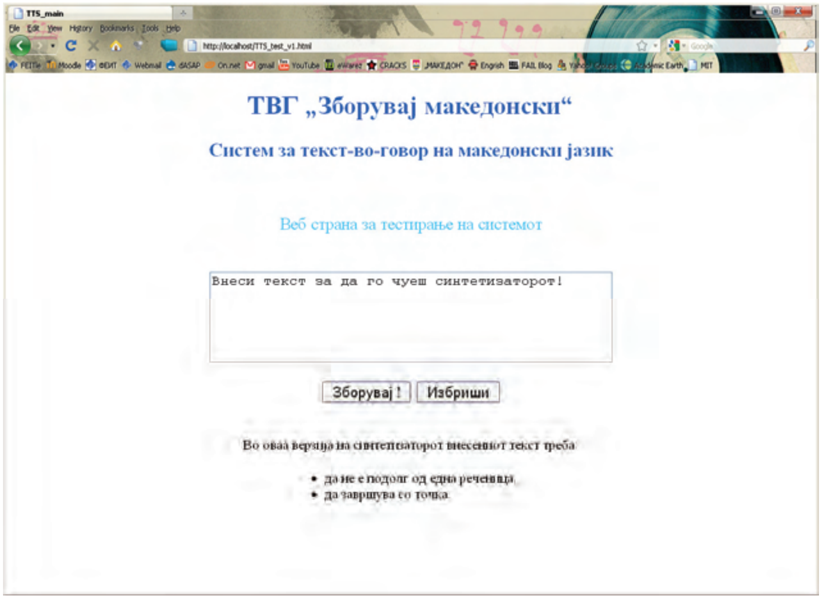


Fig. 3. MATLAB Compiler's Deployment Tool GUI compiling the TTS system

To use the compiled application the executable and the CTF archive (or the package) need only to be copied (or extracted) to the destination platform. If the destination platform lacks MATLAB, then it is also necessary to install the MATLAB Component Runtime (MCR) on it. The MCR contains a data base of all the MATLAB built-in libraries that allow m-code execution.

When the EXE file is first started it will extract the CTF archive into a new folder. Although the filenames of the user-written m-code are left intact, the code within these files is protected using the Advanced Encryption Standard (AES) with keys encrypted with 1024-bit RSA keys. This denies access to the project's code by third-party users of the distribution package. The binary executable and the mat-data structures are not encrypted.


TTS_main
File Edit View History Bookmarks Tools Help
http://localhost/TTS_test_v1.html
FEITle Moodle ФЕИТ Webmail dASAP On.net gmail YouTube eWarez CRACKS „МАКЕДОН" English FAIL Blog Yahoo! Groups Academic Earth MIT
Зборувај ! Избриши

ТВГ „Зборувај македонски“

Систем за текст-во-говор на македонски јазик

Веб страна за тестирање на системот

Внеси текст за да го чуеш синтетизаторот!

Во оваа верзија на синтетизаторот внесениот текст треба

- да не е подолг од една реченица
- да завршува со точка

Fig. 4. Layout of the web page for online presentation of the TTS system “Speak Macedonian“

For the purposes of interfacing the TTS system through a website it was necessary to redesign its interaction method. Namely, the role of the GUI was to be taken entirely by the website. This was carried out with the use of two auxiliary files – one text file and one audio file. When started, the executable loads the text TXT file that contains the text input by the user and starts the TTS synthesis process. After the synthesized speech is output the executable writes the audio data in a WAV audio file. The creation of the text file with user input text and the transfer of the audio file back to the user is to be done by the website interface.

IV. Website interface

The website interface that users can use to access our TTS system was built using a combination of PHP and HTML. The layout of the finished website is shown in Fig. 4. It consists of a text area for user input and two pushbuttons. The submit button “Зборувај!” (eng. “talk”), calls the PHP service script that generates a text file with the user input, starts the MATLAB executable and sends the generated audio file to the user. The pushbutton “Избриши” (eng. “clear”) resets the text area to null.

The project has so far been tested on a local-host machine using a WampServer installation. Wampserver (Windows Apache MySQL PHP server) is a free internet development environment for Windows, [11]. It can be used to locally develop web sites and applications based on Apache, MySQL and PHP, before they are put online.

A. HTML code

The HTML code was written in accordance to the XHTML 1.0 Transitional standard. All styling and formatting was done using the Cascading Style Sheets (CSS) formatting language. A key part of the webpage is the form declaration. The action attribute in the form declaration points to the PHP service script that executes the TTS system. The form is declared to accept user input in UTF-8 i.e. Unicode formatting allowing for the use of Cyrillic’s. The method of sending user input is chosen to be POST, as it does not limit the length of the input text. The textarea element which creates the text input field, sets the onfocus attribute so that the default text (“Внеси текст за ...”) is cleared when the user clicks on it. Further clicks on the text field won’t clear it allowing the user easy modification of the input in successive trials of the TTS system.

B. PHP script

In this section we will present the key points of the PHP script that uses the user’s input from the HTML form to call the MATLAB application and send its output to the user. The PHP script executes in the steps shown in Fig. 5.

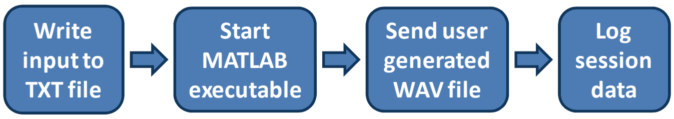
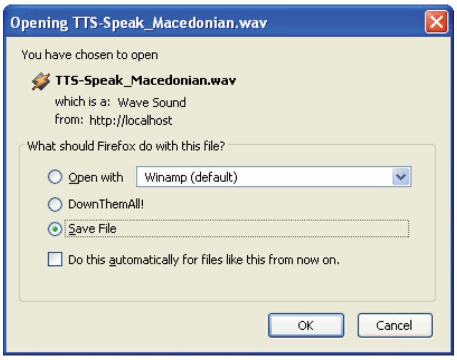


Fig. 5. Block diagram of PHP service script

First, the script takes the input text from the HTML form and writes it to an auxiliary TXT file. Then, it activates the MATLAB executable, which processes the text and generates an audio WAV file containing the synthesized speech. Next the PHP script sends this WAV file to the user in the form of an attachment, through the use of HTTP headers. At the end, the script logs the data including the current time, user’s IP address, input text and success of transfer, in a log file.

The transfer of the WAV file was realized through the sending of HTTP headers for Cache-Control that disabled cashing of the sent file along the links of the server-host connection, and HTTP headers that declared the length of the file, MIME file type (WAV audio), and the method that the file is to be sent with (as an attachment). The sending of these headers opens a file download window in the user's browser window, Fig. 6. Next the PHP script reads the WAV file to the server's output buffer and flushes the buffer to start the file transfer. At the end, the script checks whether the transfer was successful by checking the status of the connection.



Opening TTS-Speak_Macedonian.wav

You have chosen to open

TTS-Speak_Macedonian.wav

which is a: Wave Sound

from: http://localhost

What should Firefox do with this file?

○ Open with Winamp (default)

○ DownThemAll!

● Save File

□ Do this automatically for files like this from now on.

Fig. 6. Sending the generated synthesized speech output to the user by the PHP service script as seen by the user

Two more practical key points of the PHP script are also worth mentioning:

- set_time_limit( ) is used to prolong the maximum time execution limit of the PHP script because the MATLAB executable can run longer than the default 30 sec, especially when started for the first time (due to the extraction of the CTF archive) or when synthesizing long text inputs.
- When writing the text to the TXT file it was necessary to first write a header in the file that declared it as a Unicode TXT file "\xEF\xBB\xBF".

V. Results and Future Work

The project has so far been tested on a local server installation. It has proven to work for various user inputs, and it is the authors' opinion that it is ready to be put online. Before this can be done, it is necessary to finalize the TTS system in respect to finishing the intonation generation module and rounding up the unit inventory.

Our plan is to upload and integrate the project on the Digital Image Processing Team (DIPteam) website, [12], which is already active at our Department. This will be followed with promotion of the system in the Macedonian Internet community, which will hopefully lead to presentation of the system to the masses.

VI. Conclusion

The main purpose of the project was the online presentation of the TTS system "Speak Macedonian" developed at our department. The project included the generation of a standalone MATLAB executable of the TTS system and a website environment that would allow people to use it. The website environment was realized with a combination of an HTML website and a PHP script. The PHP script realizes the interface between the user and the system and also provides data logging so that every activation of the application can be logged for analysis. The project was tested on a local server installation and is expected to be put online after finalization of our TTS system. New projects based on other digital signal processing systems built in MATLAB will follow.

References

[1] D. O'Shaughnessy, "Modern Methods of Speech Synthesis", IEEE Circuits and Systems Magazine, vol. 7, No 3, 3rd Quarter 2007, pp. 6 – 23.

[2] M. Sečujski, R. Obradović, D. Pekar, Lj. Jovanov and V. Delić, "AlfaNum System for Speech Synthesis in Serbian Language", In Proc. of the 5th Conf. Text, Speech and Dialogue, Brno, 2002.

[3] L. Josifovski, D. Mihajlov and D. Gorgevik, "Speech Synthesizer Based on Time Domain Syllable Concatenation", SPECOM '97 Cluj-Napoca, Oct. 27-30, 1997.

[4] M. Zrmanovska, *Sinteza na makedonskiot govor vrz baza na tekst so prozodija*, Master Thesis, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 2005.

[5] V. Delić, M. Sečujski, D. Pekar, N. Jakovljević and D. Mishković, "A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian", IS-LTC 06, Ljubljana, Slovenia, 9. - 10 october, 2006.

[6] S. Chungurski, I. Kraljevski, D. Mihajlov and S. Arsenovski, "Concatenative Speech Synthesizers and Speech Corpus for Macedonian Language", 30th International Conference ITI Cavtat/Dubrovnik, Croatia, Jun. 23-26, 2008.

[7] S. Čungurski, *Govorni sintetizatori na makedonski jazik*, PhD Dissertation, Faculty of Electrical Engineering and Information Technologies, Skopje, Macedonia, 2009.

[8] B. Gerazov, G. Shutinoski and G. Arsov, "A Novel Quasi-Diphone Inventory Approach to Text-To-Speech Synthesis", MELECON '08, Ajaccio, France, May 5-7, 2008.

[9] B. Gerazov, S. Bogdanova, and Z. Ivanovski, "Linear predictive based voice transformation module for Macedonian TTS", TELFOR 2009, Belgrade, Serbia, Nov 24 – 26, 2009.

[10] *MATLAB Compiler User's Guide,* The MathWorks, 1995–2010.

[11] http://www.wampserver.com/en/

[12] http://dipteam.feit.ukim.edu.mk