Reports of Enlarged Session of the
Seminar of I. Vekua Institute
of Applied Mathematics
Volume 23, 2009

THE GEORGIAN TEXT READER SYSTEM WITH THE USER'S POSSIBILITY TO BUILD IN AN OWN SYNTHETIC VOICE

Pkhakadze K.,[1,2,4] Chichua G.,[3,4] Vashalomidze A.,[4]
Abzianidze L.,[2,4] Maskharashvili A.,[2,4] Chiqvinidze M.[2,4]

[1]I. Vekua Institute of Applied Mathematics
[2]Iv. Javakhishvili Tbilisi State University
[3]I. Chavchavadze State University
[4]Open Institute of Georgian Language, Logic and Computer
2 University Str., 0143 Tbilisi, Georgia
e-mail: gllc.ge@gmail.com, Web-site: www.gllc.ge

Abstract. In this paper, we shortly and generally describe the Georgian Text Reader System (TRS) with the user's possibility to build in an own synthetic voice (TRSwithUV), by the help of which a computer will be able to read Georgian written texts with the user's own (synthetic) voice. At this moment, in our group, we have already elaborated two different versions of the Georgian TRSs, which are built up independently and are based on the different principles. The experiments, done with the purposes of building these systems, have led us to the idea of a TRSwithUV and brought to light the methods of building it. The main factors, which have influenced us, are simplicity and completeness of the letter-phone correspondences in the Georgian Language. Thus, the paper describes the ideology according to which we are working out the TRSwithUVs for the Georgian Language.

Keywords and phrases: Georgian reader-listener system, Georgian text reader system, building own synthetic voice, method of internal listening, method of transcriptions.

AMS subject classification (2000): 68T10; 68T35.

1. Introduction. A text reader system (TRS[1]) is the system, which reads (converts) natural language text (into speech). The languages have different letter-phone systems, different reading rules and different specifics of word pronunciation. Consequently, for different languages problems of creating TRSs have different complexities. In a certain language, the complexity of the problem is mainly depended on the complexity of reading rules[2] and on the complexity of letter-phone correspondence[3] of the language.

[1] Also known as a speech synthesizer, or a text-to-speech (TTS) system.

[2] We characterize a set of reading rules according to their complexity and completeness. More the reading rules are depended on the letter arrangement more complex are the reading rules (if there is no dependence then the reading rules are simple). A set of reading rules is complete if all words can be read correctly according to these rules; otherwise, the set of reading rules is incomplete. Georgian reading rules are simple and complete.

[3] We characterize a letter-phone system of a language with complexity and completeness: a letter-phone system of a language is called as simple and complete if the letters and phones of the language are in one-to-one correspondence.

TRSs have two major qualities: naturalness[4] and intelligibility[5]. The ideal TRS is both natural and intelligible. There are a lot of high quality (from the point of both qualities) TRSs for major and technologically equipped languages as English, French, German, Japanese, Chinese, Russian, Spanish, Italian etc.

In addition, there is great variety of synthetic voices available for TRSs; after the installation of a certain synthetic voice, your TRS will be able to synthesize a text into this voice. Nevertheless the above-mentioned, there are only few TRSs offering users to build in their synthetic voices in the system. After the long search on the internet, we found out three systems with such kind of the user's possibility.

- *Polluxstar*[6] text-to-speech software for Japanese;
- *MARY*[7] text-to-speech synthesis system for German, English and Tibetan;
- *Festival*[8] Speech Synthesis System with the *Festvox*[9] System mainly for US and UK English.

In Spite of these systems and their functionalities, the users still prefer the systems with already processed synthetic voices because:

- It is hard work to collect data for a new synthetic voice: it is enormous work to collect the speech database for the new synthetic voice. It is necessary to record the voice for a long time in a severe environment. This puts a great load on the users. After all, there is little chance that the created synthetic voice will be high quality synthetic voice like other ones already existing in different voices.
- There is a great variety of high quality synthetic voices available: there are a lot of companies and laboratories offering a great variety of high quality synthetic voices to the users. It is much easier to preview, buy/download and install a new synthetic voice into the TRS, than to build your own synthetic voice.

[4] Naturalness describes how closely does the synthesized speech resemble the human speech.

[5] Intelligibility is the ease with which the synthesized speech is understood.

[6] The software is released by Oki Electric Industry Company (Tokyo, July 24, 2008). "The text-to-speech software that enables reproduction of a user's real voice complete with that individual's unique intonations and tones. By using Polluxstar on a computer, the users can communicate using their own voice, instead of a mechanical non-human voice" "OKI has been researching and developing ways to reproduce voice by collecting a reasonable amount of voice data without the requirement that words necessarily be pronounced accurately. As a result, OKI succeeded in establishing a stable and effective voice database and was able to develop "real voice" technology at a practical level" (see http://www.oki.com/en/press/2008/07/z08050e.html).

[7] The project is maintained by DFKI's (German Research Institute for Artificial Intelligence) language technology lab. In MARY TTS system (see http://mary.opendfki. de/wiki/VoiceImportToolsTutorial), with the help of the voice import tools a user can build new German and English voices (even own one) from wave (a type of audio files) files.

[8] The Center for Speech Technology Research (The University of Edinburgh) collaborates with Carnegie Mellon University's speech group on Festival (see http://www.cstr.ed. ac.uk/projects/festival). The last version of Festival was released in July 2004.

[9] This project is part of the work at Carnegie Mellon University's speech group aimed at advancing the state of Speech Synthesis. The Festvox project aims to make the building of new synthetic voices more systemic and better documented, making it possible for anyone to build a new voice (see http://festvox.org/index.html). This project do not guarantee that a user will end up with a high quality acceptable voice, but with a little care the user can likely build a new synthesis voice in the supported languages (US and UK English) in a few days, or in a new language in a few weeks (more or less depending on the complexity of the language, and the desired quality). The last version of Festvox was released on January 21, 2007.

In the paper, we will describe a Georgian TRS, on which our group[10] is working. This Georgian TRS will be extended by the user's possibility to build in an own synthetic voice. In the construction of the new TRS, we will use components of our two different Georgian TRSs.

At the same time, we will describe these components and their functions in the paper, in order to, the readers feel evident perspectives for the construction of the new Georgian TRS with the user's possibility to build in an own synthetic voice (shortly, the Georgian TRSwithUV). We are developing the user's voice building tool in the Georgian TRS, in order, to use the user's voice building technology in speech recognition: attaching the Georgian TRSwithUV to the Georgian Speech Listener System as a result gives the Georgian Reader-Listener System with the user's possibility to build in an own synthetic voice [5].

2. Some solutions of the Georgian TRS with user's possibility to build in an own synthetic voice. At the beginning, we want to declare that the Georgian reading rules (GRR) as Georgian letter-phone correspondence are simple and complete ones.

The Georgian TRSwithUV divides a Georgian word into syllables before its synthesizing. In this procedure, we use the method of natural non-semantic division of a word into syllables. The method is based on below described principles:

• In a word, a vowel[11] makes only one syllable. Therefore, in any word, the number of syllables is equal to the number of vowels.

• In a word, a vowel is characterized by its operating scope, which is given by the pair (m, n); $m, n \in \{0, 1, 2 \ldots\}$.

• If in a W word, the scope of some V^1 vowel is (0,0), then this V^1 is a syllable of this W word. This type syllable, generally, syllable of $Vform$ is called as $Vtype$, or trivial syllable.

• If in a W word, some V^1 vowel has (m, n) scope, then $C_{Lm} \ldots C_{L1} V^1 C_{R1} \ldots C_{Rn}$ is the syllable made by the V^1. C_{Lk} (resp. C_{Rk}) is a consonant, placed in left-k^{th} (resp.right-k^{th}) position from V^1 in the W. Generally, a syllable of $C \ldots CVC \ldots C$ form is called as complex syllable. At the same time, a syllable of $VC \ldots C$ (resp. $C \ldots CV$) form is called as right (resp. left) complex syllable.

• If in a W word, some V^1 vowel has (1,0) (resp. (0,1)) scope, then the syllable, which is made by V^1, is called as simple left (resp. right) syllable. Generally, a syllable of CV (resp. VC) form is called as simple left (resp. right) syllable.

Short review of our researches and results in the text reading technology: there are two Georgian TRSs collaborated in our group. In order to distinguish these TRSs from each other, we will call them the TRS-1 (by G. Chichua) and the TRS-2[12] (by A. Vashalomidze).

It must be mentioned, as some of the realizations of the Georgian TRSwithUV

[10] Authors, which are members of the Open Institute of Georgian Language, Logic and Computer (www.gllc.ge)

[11] There are 5 vowels and 28 consonants in the Georgian language

[12] The TRS-2 is not updated version of the TRS-1, though the TRS-1 is created earlier than the TRS-2.

may be considered the TRS-1[13] and TRS-2, but in these systems, the voice building procedures require many efforts and is too hard for a user. So, in constructing the Georgian TRSwithUV, we consider following key factors having crucial importance:

- The quality of synthesizing (including both naturalness and intelligibility);
- Reasonable work required from user to build in an own Synthetic voice;

According to our approaches any Georgian TRSwithUV must consists of three different principal procedures:

- Speech Recording Procedure (SRP);
- Database building Procedure (DBP);
- Text Reading Procedure (TRP);

Below we will describe some different realizations of each procedure; a certain combination of them gives a certain realization of the Georgian TRSwithUV.

Speech Recording Procedure: any SRP is built and functions in the following way: the system displays vowels/ consonants/ syllables/ morphemes/ words/ phrases/ sentences/ texts after each other. After the beep, a user has to pronounce the displayed vowel/ consonant/ syllable/ morpheme/ word/ phrase/ sentence/ text according to the instruction attached to it. The system records the vowels/ consonants/ syllables/ morphemes/ words/ phrases/ sentences/ texts. The system may ask a user to repeat the vowel/ consonant/ syllable/ morpheme/ word/ phrase/ sentence/ text. Also, a user has possibility to delete or/and rerecord them.

Database Building Procedure: any DBP is built and functions in the following way: the system for each recorded vowel/ consonant/ syllable/ morpheme/ word/ phrase/ sentence/ text has a sample (the sample may be simple or complex. A complex sample is constructed through the different integrations of different simple samples). With the help of these samples, system builds database for synthetic voices[14].

Text Reading Procedure: Now the Georgian TRSwithUV has the database of synthetic voice and with the help of it, the system reads Georgian texts. According to the type of database, the text reading procedures differ from each other. Generally, we differ from each other semantic and non-semantic reading. By now, we are realizing a non-semantic reading system, in which we use the method of natural non-semantic division of the words into syllables and the methods of reading complex syllables by the help of simple ones. Below we describe some techniques of constructing of non-semantic reading system:

- If database of synthetic voice is built only by the consonants, vowels and CV-syllables, the system uniquely divides the words into constituents by giving priority to the CV-syllables. For example: *gaumjobeseba [ga-u-m-jo-be-se-ba]* (this methods of

[13] The Georgian TRS-1 has an additional program Start-Length Importer, with the help of which a user can import parameters from one speech database into another one; the program's function is to ease up a new voice building procedure for the user, but the import of parameters does not always give a desirable result.

[14] Some example of simple samples: the system separates words from each other in a phrase/sentence/text. Also, it can extract and trim the silence from recorded data. The system with the help of some different methods elaborated in our group separates syllables in a word.

reading is used in TSR-1).

• If database of synthetic voice is built only by the consonants, vowels and VC-syllables, the system uniquely divides the words into these constituents by giving priority to the VC-syllables. For example: *gaumjobeseba [g-a-um-j-ob-es-eb-a]*.

• If database of synthetic voice is built only by the consonants, vowels and simple right and left syllables, the system divides the words into these constituents with giving different priority to different type syllables. For example: *gaumjobeseba [ga-um-jo-be-se-ba]*.

• If database of synthetic voice is built only by the vowels and simple right and left syllables, the system naturally divides the word into syllables. After that, the system begins to synthesize each syllable; then joins two neighbor synthesized syllables and gets the synthesized word. The complex syllable synthesis is performed according to either of rules:

$$\begin{array}{ll}
C_k \ldots C_2C_1V\ : & C_kV \ldots C_2VC_1V \longrightarrow C_kv \ldots C_2vC_1V \\
VC_1C_2 \ldots C_k\ : & VC_1VC_2 \ldots VC_k \longrightarrow VC_1vC_2 \ldots vCk \\
C_m \ldots C_1VC_1 \ldots C_n : & C_mV \ldots C_mVVC_1 \ldots VC_n \longrightarrow C_mv \ldots C_mvVvC_1 \ldots vC_n
\end{array}$$

where, in right side, C_nV (resp. VC_n) is a sound of C_nV (resp. VC_n) syllable in the database; C_nv (resp. vC_n) is C_n consonant extracted from C_nV (resp. VC_n) sound.

3. Conclusion. Different realization of SRP, DBP and TRP give different Georgian TRSwithUV. The Georgian TRSwithUV require reasonable work from users, and as a result, users will get the synthesis in his/her voice with the quality not worse than our current Georgian TRSs have. Finally, it must be mentioned that the methods, used in the Georgian TRSwithUV, will be used in the Georgian Reader-Listener System with the user's possibility to build in an own synthetic voice [5].

REFERENCES

1. Chicua G., Division of a speech into words and identification of the discourse, *II Conference in Natural Language Processing, Ar.Chiqobava Institute of Linguistics, Tbilisi*, 2004.

2. Chicua G., Computer Speech Recognition, *III Conference in Natural Language Processing, Ar.Chiqobava Institute of Linguistics, Tbilisi*, 2005.

3. Chicua G. Abzianidze L., The speech recognition and synthesizing, *V Conference in Natural Language Processing, Ar.Chiqobava Institute of Linguistics, Tbilisi*, 2007.

4. Chicua G., Trial program of speech recognition, *VI Conference in Natural Language Processing, Ar.Chiqobava Institute of Linguistics, Tbilisi*, 2008.

5. Pkhakadze K., Chichua G., Vashalomidze A., Abzianidze L., Maskharashvili A., Chiqvinidze M., *The Georgian Reader-Listener System with the User's Possibility to Build in an Own Synthetic Voice, I. Vekua Institute of Applied Mathematics, Tbilisi*, 2009.

Received *.**.2009; revised **.**.2009; accepted **.**.2009.

Authors' addresses:

K. Pkhakadze
I. Vekua Institute of Applied Mathematics of
Iv. Javakhishvili Tbilisi State University

2, University St., Tbilisi 0186
Georgia
E-mail: gllc.ge@gmail.com