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

THE GEORGIAN READER-LISTENER 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: E-mail: gllc.ge@gmail.com, Web-site: www.gllc.ge

Abstract. In the paper very shortly and generally we talk about the Georgian Reader-Listener System with the user's possibility to build in an own synthetic voice, by the help of which a computer will be able to read Georgian written texts and to typewrite Georgian spoken texts. At this moment, our group[1] has already created the first versions of the Georgian Text Reader System and the Georgian Speech Listener System. Below we will shortly describe them: it must be mentioned that we consider the Georgian Reader-Listener System with the user's possibility to build in an own synthetic voice, on which we are currently working, as the system, which is obtained through some type of integration of some type Georgian Text Reader System (TRS) and the Georgian Speech Listener System (SLS).

Keywords and phrases: Georgian Reader-Listener System, Georgian Text Reader System, building synthetic voice, method of internal listening, method of transcriptions.

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

1. Introduction. In the paper, we talk very shortly and generally about the Georgian Reader-Listener System (RLS) with user's possibility to build in an own synthetic voice (RLSwithUV), with the help of which computer will be able to read Georgian written texts and to listen (to typewrite) Georgian speech.

At this moment, our group has already created the first versions of the Georgian Text Reader System (TRS) and the Georgian Speech Listener System (SLS). Below we will shortly describe them: it must be mentioned that we consider the Georgian RLSwithUV, on which we are currently working, as the system, which is obtained through some type of integration of some type of TRS and SLS.

2. The Georgian Reader-Listener System with User's Possibility to Build in an Own Synthetic Voice. At this moment, we are building the Georgian TRSwithUV [5], which is considered as one of the main parts of the Georgian RLSwithUV. The main idea, which is elaborated in this project, is to construct such type of SLS,

[1]The authors - members of the Open Institute of Georgian Language, Logic and Computer (www.gllc.ge)

which will be able to recognize the Georgian speech without teaching of spoken forms of the words and, also, which will be able to recognize freely almost any spoken text of almost any Georgian speaker[2]. As it was mentioned, we have already created the first versions of the Georgian TRS and SLS. Below, we shortly describe them:

• There are two Georgian TRSs collaborated in our group. We call them the Georgian TRS-1 (by George Chichua) and the Georgian TRS-2 (by Aleksandre Vashalomidze). These Georgian TRSs are created independently from each other and therefore the Georgian TRS-2 is not bug fixed and updated version of the Georgian TRS-1[3]. Moreover, the evolution, algorithms, and speech databases of these reader systems are different from each other. At this moment, on the base of that various experience, which we have already received, we aim to construct such type of Georgian TRSwithUV, which will require reasonable work from a user to train the system in order to build his/her synthetic voice in the system.

• In our group, researches in speech recognition are directed by George Chichua. For 2004, he had elaborated the system, which recognized Georgian words from a fixed database of words. The System worked based on the principle of teaching written and spoken forms of the words. The listening accuracy (recognition percentage) was 80% in the database of about 100 words. In 2007, the system was able to recognize sentences, but a user had to speak discretely (to make slight pauses between the words of the sentence). The recognition algorithm was improved, but the teaching[4] principle was the same. In the sentences, the listening accuracy was 95% in database of about 350 words. From 2008, we are working out on the Georgian SLS, which does not need spoken forms of words in the database (we will shortly describe the algorithm of the system). The system uses only textual database of words. The system is user dependent[5].

In general, the monolingual systems like SLS (speech recognition software, voice recognition software, a speech-to-text system) have listening/recognizing abilities only in one natural language and as a rule, they need some training: this means that, in order to recognize a user's speech, the system offers to the user some texts to read and records the speech. After the training, the system recognizes the user's speech with the high accuracy (in some systems, accuracy is up to 97%) in a large database of words. That is why these SLSs are called user-dependent systems.

There exist user-independent systems but they have high recognition only in very restricted databases (often they need some additional information about the users like: the user is a male or a female, a man, a woman, a boy, a girl or a child). The Georgian RLSwithUV is planned to be a user-dependent system in the above mentioned sense. However, because of the users' possibility to build in the RLSwithUV their synthetic voice, this dependence does not make any practical difficulties for a user. We are going to improve the method used in our current Georgian SLS. In order to recognize the

[2]This means that our aim is to construct such type of RLSwithUV, recognizing abilities of which will not be limited by a user's voice.

[3]Though the Georgian TRS-1 is created earlier than the Georgian TRS-2.

[4]We have attempted to make the SLS without teaching the spoken form of the words, but recognition percentage of the system is about 30-40% in database of 800 words.

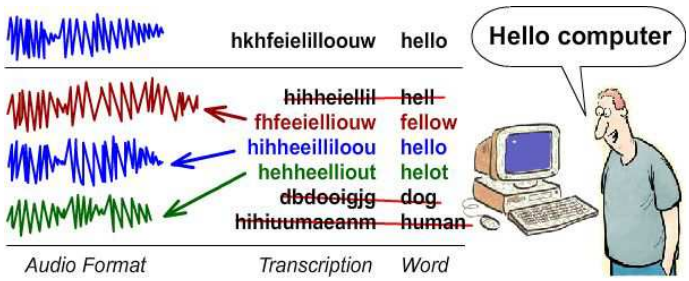
[5]This means that a user has to train the system with his/her voice.

speech, the Georgian SLS uses the method of splitting a sentence into words and the method of transcriptions of heard words. The transcription is a long string and any phone (i.e. consonants and vowels) has some corresponding substring in the string. The system converts a heard sentence into the array of heard transcriptions (each heard transcription corresponds to the heard word in the sentence).

The system has two textual databases: a database of words and a database of transcriptions of these words, derived from spoken form of the words. The problem of the recognition of the word is reduced to the problem of identifying which transcription from the database is similar to the heard transcription. The idea of this method is that the system converts words of the database in the very format in which listened speech is input in the system. This method is a modified (distorted) version of the main method, in which we build a database of synthetic voice for solving the speech recognition problem.

The main method, which we call a method of internal listening, is that the system is able to synthesize any word of the textual database into its spoken form and calculate its audio data. After this, the system can compare input spoken word and its audio data to its generating spoken word and audio datum. We call this method the method of internal listening.

Our group has theoretically elaborated this general approach for solving speech recognition problem in 2005, and from 2005 we began to work on the Georgian TRS. According to this theoretical approach, the Georgian TRS with high quality will give us, in perspective, the Georgian SLS with high quality[6]. But, it is clear, that using only primitive application of the method of internal listening (synthesizing and comparing spoken forms and datum of the listened words) in the speech recognition, may require much time for finding in the database the word similar to heard one. Because of this complexity, which we are presuming, we plan to use some different type additional methods[7] to overcome mentioned complexity. For example, we use, in constructing of improved Georgian SLS, the method of transcriptions and the method of internal listening in combination (see the image):



• The method of transcriptions - The method will help the system to accelerate search process in the database. In this method, the system has already the database of transcriptions; therefore, there is no need for word synthesis, in order to compare the

[6]We believe that this approach is a natural approach to solve the speech recognition problem.

[7]Here is shortly named some different type of additional methods: time oriented additional methods, Memory oriented additional methods, Grammatical additional methods, Semantic additional methods, and different combined additional methods. But, by now, we do not concentrate on them.

word of database to the heard word. However, using this method will give the system a set of the words (not a word as it does in the current Georgian SLS) more similar[8] to the heard word.

• The method of internal listening - In the system, the method will serve as a powerful tool, with the help of which, in the given set[9], the system will detect the most similar words to the heard word. There will be few words in the set, so that the system will need little time to synthesize these words and compare the synthesized words to the heard word (they will be in the same audio format).

It is clear, that in order to get the high listening accuracy, it will be better, if the system will compare the input word to the internally synthesized words in the same voice. The system needs to synthesize words in the user's voice for the comparing to the heard words pronounced by the same user. Therefore, the Georgian RLSwithUV will include a user dependent SLS with the high listening accuracy[10].

A user has to train the system with his/her voice. In the training procedure, we aim to integrate the speech recording procedure of the Georgian TRSwithUV and the training procedure of the current Georgian SLS. Therefore, a user will be able to pass both (training and speech recording) procedures in one integrated procedure. This will ease up the work for the user.

After the procedure, the user will get as the Georgian SLS trained with his/her voice, as the Georgian TRS with his/her own synthetic voice. One of the benefits of the construction of the Georgian RLSwithUV is that a user with the same effort (used in speech recording or training procedures) will get the Georgian RLSwithUV (the SLS with high listening accuracy and the TRS with the user's synthetic voice).

Conclusion. The construction of the Georgian RLSwithUV is supported by several reasons:

• A new method is used in the speech listening procedure, which results the high listening accuracy.

• Without increasing the load on a user, the user simultaneously can train the SLS and build in an own synthetic voice in the TRS.

• The most of users use both systems: the Georgian TRS and the Georgian SLS. From this time, a user can easily have integrated version of them - the Georgian RLSwithUV.

R E F E R E N C E S

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

[8]In the method of transcriptions, each pair of words (a word of database and heard word) has a coefficient of similarity; more the coefficient is, more similar are these words.

[9]The set of words, defined by the method of transcriptions.

[10]We can control the listening accuracy with the help of size of the database of words. We can achieve high recognition accuracy if we decrease the size of the database or build the database of orthographically very different words.

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

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

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

5. K. Pkhakadze, G. Chichua, A. Vashalomidze, L.Abzianidze, A. Maskharashvili, M. Chiqvinidze, The Georgian Text Reader 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