CONTROL OF COMPUTER PROGRAMS BY VOICE COMMANDS

Vytautas SLIVINSKAS*, Gražina PYŽ**, Virginija ŠIMONYTĖ*
*Lithuanian University of Educational Sciences, Lithuania
**Vilnius University Institute of Mathematics and Informatics, Lithuania

Abstract: The paper deals with the computer program control using voice commands. The problem of the voice command recognition is discussed and the results of investigation of the voice command recognition are presented. The use of voice commands is illustrated by an example.

Keywords: Speech recognition, voice commands, computer program control.

1. Introduction

Speech recognition is a process that converts spoken words and phrases to a computer-readable format. Speech recognition is mostly developed for the languages of big countries especially for English language. There exists a number of programs (Dragon Naturally Speaking, Speech Recognition in Windows 7, Dragon Dictate for iPhone/iPad, etc.) that are intended for the use with English language. The languages of small countries cannot boast of such a great attention, nevertheless the local researchers give much effort to developing recognizers of the local languages. Lithuanian speech engineering is a popular field of engineering in Lithuania. Lithuanian speech recognition is one of the Lithuanian speech processing problems taking considerable attention of Lithuanian researchers. One of the best known group works at Vilnius University Institute of Mathematics and Informatics. This group solves such problems as: quality estimation of speech recognition features [1], speaker recognition by voice [2], development of isolated word recognition systems [3], acoustic modelling of Lithuanian speech recognition [4], etc. Scientists from Vilnius University Faculty of Mathematics and Informatics and Forensic Science Centre of Lithuania investigate speaker recognition problems [5]. Researchers from Vilnius Gediminas Technical University deal with control of robots by voice [6], development of biometric systems for person recognition [7], development and implementation of means for word duration signal processing [8], [9]. A group of scientists from Kaunas University of Technology, Vytautas Magnus University and Vilnius University Kaunas Faculty of Humanities carries out investigation of foreign languages models for Lithuanian speech recognition [10], analyse control of electric devices by voice [11], deal with hierarchical phoneme classification approach [12], develop isolated-word Lithuanian HMM speech recognition system [13], investigate HMM for large vocabulary continuous speech recognition [14]. Speech recognition allows one to give input to a computer program using voice. A program that is called the speech recognition engine carries out the speech recognition task.

There exist two cases of the computer program control by voice:

1) the speech recognizer depends on the speaker (speaker-dependent recognition),
2) the speech recognizer does not depend on the speaker (speaker-independent recognition).

Implementation of the second case is much harder although the result is more attractive to the user. The first case only is considered in this work. The results of automatic Lithuanian speech recognition performed by a system that was developed for a construction estimate program are presented in this paper. Section 2 deals with voice command recognition using cepstral approach. The requirements for the recognition system software and hardware are presented in Section 3. Section 4 contains a description of an example of practical application of computer program control by voice commands.

2. Voice command recognition using cepstral approach

The system consists of two parts: the training part and recognition part. The system performs in the following way:

1) A list of commands is compiled;
2) The commands are recorded to a .wav format file (11.025 kHz, 16-bit mono);
3) The extraction of features of each command is carried out;
4) These features are stored in a database;
5) The process of command recognition is performed.

2.1. Voice commands

The voice commands can be divided into a few categories:
1) the commands specified by words (e.g., “atidaryti” (open), “uždaryti” (close), “spausdinti” (print));
2) the commands described by a numerical code – they are pronounced by uttering each digit separately (e.g., the code “999” is uttered as “nine, nine, nine” and means to open);
3) the commands specified by words and a numerical code (e.g., “samata 45” (estimate four five), “laškas 52” (letter five two)); in this case an object (estimate, letter) and the action to be taken with that object are specified.

The number of commands can be large. In this case the command coding has a drawback – the user can encounter difficulty to memorize the command code. In order to avoid this difficulty, the system can return the command name after recognizing it. When the user confirms the command, the system runs it.

If the number of commands is large, then it is useful to divide them into groups. The names of such groups can be as follows: editing, saving, printing, etc.

The system should recognize the commands and their codes without mistakes.

2.2. Feature extraction

The cepstral method is used for speech recognition [15]. The features that specify the spectrum of the uttered word are selected. The main feature selection algorithm steps are shown in Fig. 1.

The result of this algorithm is a feature row-vector for each frame. We then make a matrix whose rows are the obtained vectors. This matrix is called a feature matrix.

2.3. Speech command interpreting and running

After the voice command is given, the program controlled by voice starts to compute the command cepstrum parameters if the microphone sound signal amplitude is higher than a certain level. The command utterance is considered to be finished if the pause following it is longer than a certain prespecified value, e.g., 0.2 s.

A feature vector is obtained for each command that is compared with the feature vectors of the etalon commands. The program selects that etalon whose parameter correlation with the uttered word parameters is the highest.

How to pass the result to the controlled program? The simplest way is to use a text file as an inbox that is regularly checked by the controlled program with the help of a small agent procedure. For example, if a command identificator is stored in this text file then this procedure cleans the inbox and runs the command. Such a connection is sufficiently flexible and universal.

3. Requirements for the voice command software and hardware

In order to control successfully a computer by voice, the computer software and hardware has to meet certain requirements. The main requirement for the software is the following: the processing of the uttered word has to be done in a sufficiently small time less than 0.5 s.

A proper selection of the computer microphone and sound card is very important.

All current sound cards support signal audio playback at 44.1, 22.05, 11.025, and 8 kHz. The sampling frequency is selected to be equal to 11.025 kHz, the dynamic range is of 16 bit, and quality is MONO. This is a good compromise between the processing rate and quality. The sound card should not have an inner noise.

It is important to have a good microphone for speech recognition. The microphone should be resistant to environment noise. The microphones used in speech recognition usually do not pass frequencies lower than 100 Hz.

4. The experiment results

In this section, an example of application of the computer program control by voice is presented. In order to control the program by voice, it is necessary to run the recognition program at the same time. The window of the recognition program is given in Fig. 2.
The recognition program stores the uttered command in a text file. The program controlled by voice uses this text file as an inbox.

Before running the recognition program it has to be trained, i.e. the etalon parameter database has to be created. In order to create the etalons and track the recognition process, a specialized user-program interface has been developed. This window is presented in Fig. 3.

![Fig. 3. The recognition process tracking window](image)

The program user records the program control voice commands into a database. He/she records these commands uttered in his/her own voice. These commands are called the etalons. Each command can be uttered several times therefore a single command can have many etalons. In the experiment, a list of commands described by integer numbers from one to nine has been made. Each command is pronounced two times and the obtained speech signals are stored into the etalon database. The feature matrix is determined for each etalon using the algorithm described in Section 2.2.

The user gives a voice command to the program. Then the features for this command are computed analogously as in the etalon case. An example of the feature matrix of the word “septyni” (seven) is shown in Fig. 4 (each row of the matrix is depicted by a single sequence).

![Fig. 4. The feature matrix of the word “septyni” (seven)](image)

These features are compared with the etalon features and the correlation between them is determined. The etalon is ascribed to the command whose correlation with that command is the largest. The largest ten values of the correlation coefficients between the feature matrix of the speech signal of the word “septyni” and the feature matrices of etalons are shown in Fig. 5.
As can be seen from Fig. 5, the highest correlation coefficient corresponds to the etalon 7_2 (the second speech signal of the word “septyni”). Therefore one can state that the system recognized the command “septyni”.

The result of investigation of the recognition accuracy is presented in Table 1.

Table 1. The recognition accuracy

<table>
<thead>
<tr>
<th>Command</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>vienas (one)</td>
<td>100 %</td>
</tr>
<tr>
<td>du (two)</td>
<td>95 %</td>
</tr>
<tr>
<td>trys (three)</td>
<td>95 %</td>
</tr>
<tr>
<td>keturi (four)</td>
<td>80 %</td>
</tr>
<tr>
<td>penki (five)</td>
<td>90 %</td>
</tr>
<tr>
<td>šeši (six)</td>
<td>85 %</td>
</tr>
<tr>
<td>septyni (seven)</td>
<td>95 %</td>
</tr>
<tr>
<td>aštuoni (eight)</td>
<td>90 %</td>
</tr>
<tr>
<td>devyni (nine)</td>
<td>95 %</td>
</tr>
</tbody>
</table>

5. Conclusions

1. In this paper a speaker-dependent Lithuanian speech recognizer is used.
2. The speech recognition process comprises of two parts: the training part and recognition part.
3. The cepstral approach is used for speech parameter estimation.
4. If several etalons are created for the same command then the recognition result is better.
5. The experiment results show that the recognition process is sufficiently accurate what is important for successful control by voice.
6. In order the recognition process is successful, the computer sound card and microphone have to be of good quality.

6. References