ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2008. No. 6(86)
ELEKTRONIKA IR ELEKTROTECHNIKA

SYSTEM ENGINEERING, COMPUTER TECHNOLOGY
T 120
SISTEMŲ INŽINERIJA, KOMPIUTERINĖS TECHNOLOGIJOS

Control of Computer and Electric Devices by Voice

A. Rudžionis, K. Ratkevičius
Speech Research Laboratory, Kaunas University of Technology
Studentų str. 65, LT-51369, Kaunas, Lithuania; phone: +370 37 354191; e-mail: alrud@mmlab.ktu.lt
T. Dumbliauskas
UAB "Elintos prietaisai", Terminalo str. 3, Biruliškių k., Karmėlavos sen., Kauno dis., LT-54469, Lithuania; phone: +370 37 351987; e-mail: tomas2mb@gmail.com
V. Rudžionis
Dept. of Informatics, Kaunas Humanities Faculty of Vilnius University
Muitinės str. 8, LT-44280 Kaunas, Lithuania; phone: +370 37 354191; e-mail: vyrud@mmlab.ktu.lt

Introduction

People with physical disabilities take advantage of a variety of methods to gain access to information technology and to electronic assistive technology for communication, mobility and daily living tasks. Many of these access methods are slow and can lead to frustration, a prime example being the use of switch-activated menu scanning. Automatic Speech Recognition (ASR) is potentially of enormous benefit to people with severe physical disabilities. The tremendous richness of human speech communication gives the user many degrees of freedom for control and input. The speed of speech recognition also gives it a potential advantage over other input methods commonly employed by physically disabled people [1]. Speech technology has "always" (at least since the mid 1990'ies) implicitly or explicitly addressed users with visual or mobile disabilities, sometimes "disguised" as a more general goal of enabling eyes-free/eyes-busy or handsfree/hands-busy access to web browsing and other applications [2].

The ability to control the home is an essential aspect of independence and e-inclusion. Environmental Control Systems (ECSs) are available which address many elements of home management for disabled people, such as control of audio-visual equipment, telephones, doors and curtains as well as the ability to summon assistance. More recently, ECS with speech recognition have been introduced and a number of such systems are available on the market [3].

Given the ability for speech recognition to provide a relatively fast and reliable control method, and the availability of high quality speech synthesis, there is the potential to combine these technologies into a speech-controlled communication aid. The intention is to provide the user with a device that performs like a human interpreter, recognizing the disordered speech and synthesizing an intelligible equivalent. This will be of use to people in social situations where they interact with non-familiar communication partners, such as at work, when shopping, in the hospital, on the telephone etc. [1].

A speech input interface based on speech recognition technology has already been applied to many applications: for the Mobility Support Geographic Information System to provide optimal paths to destinations with information on barrier and barrier-free objects related to sidewalks for all pedestrians, especially elderly people and disabled people [4], for pronunciation training to learn the foreign language [5], for people with severe dysarthria as a means of text input to computer [1], to get web information for people with the severe dyslectic and aphasic [6] and so on.

New project „Lithuanian multimodal voice services" funded by Lithuanian State Science and Studies Foundation started in 2007 and will deal with a series of problems arising in the above mentioned applications . This paper presents two prototypes of voice services which are going to be developed for the disabled people during this project:

- voice dialogue models for mobility of handicapped persons (inability to control hands). In this case it is possible: a) to switch on/off electric appliances by voice command; b) to open the necessary web page by voice command;
- a model for blind persons includes a sequence of actions a) a text on the computer screen is synthesized and read by voice; b) the computer control is performed by voice commands.

The program, which reads the text from internet and synthesizes it, was presented in [7]. The universal program for computer and electric devices control by voice commands should be developed during above mentioned project till

2009. The program should satisfy the so called *design for all* (DfA) *or inclusive design* principles [8]. An example of applying the DfA-principles is to equip electronic services and applications with intelligent modality adaptive interfaces that let people choose their preferred interaction style depending on the actual task to be accomplished, the context, and their own preferences and abilities.

Speech or voice servers, such as *Microsoft Speech Server (MSS)* or *IBM WebSphere Voice Server (WVS)*, offer ASR (Automatic Speech Recognition) and TTS (Text-To-Speech Synthesis) based speech interfaces. Voice servers integrate together telephony, speech and internet [9]. Both servers still don't support Lithuanian voice recognition and Lithuanian text to speech synthesis engines. So far the usage of English transcriptions of Lithuanian words is quite suitable solution of voice servers application for Lithuanian language. Investigation of Lithuanian commands recognition by English speech recognition engine should be useful for voice servers adaptation to Lithuanian speech.

Software for computer control by voice

As was mentioned in the introduction, the universal program for computer and electric devices control by voice commands is under development. The program for computer control by voice commands "Balsas" was developed in the first stage. Since this program will be used only under the Windows operating systems so all requirements for user interface were presented taking into account user environment which is characteristic for Windows operating systems. The program "Balsas" is based on Microsoft SAPI (Speech Application Programming Interface) interface implementation. This interface enables development of speech recognition software without detailed knowledge of recognition algorithms. It allows to separate user interface and logic of software functions from speech recognition and audio signal input/output. It is possible to use the program "Balsas" with any installed in the system speech recognition engine which corresponds with SAPI specification. If new speech engine will be installed or current engine will be updated, the control program will be able to work successfully with new engine.

Program "Balsas" is able to work in the background. Program status is shown in the system tray. Here user is able to see the status of program: "listening mode", "non-listening mode". When clicking on the program icon on desktop with mouse user have the possibilities to change program settings, to change program mode, to close program, to get more information about program status (e.g., which voice commands were accepted). In the "listening mode" or active mode program accepts voice commands and responds to them performing pre-selected actions and in the "non-listening mode" or inactive mode program waits while it will be switched to the „listening mode". The view of the program "Balsas" in the active mode is shown in Fig. 1. The field "Hipotezė" in the program panel represents the hypothesized but not recognized command.

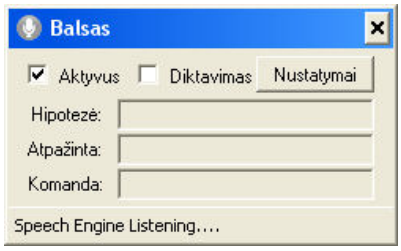

Fig. 1. The view of the program "Balsas" in the active mode

Methods to change program status are as follows: change of mode in the system tray, change of mode using combination of keys, switching to the active mode when predefined key is pressed and switching to the inactive mode when the key is released, change of mode using specified voice command. During research phase it must be determined which program status switching way is the most appropriate for the users and this way will be selected as the default (will be used immediately after installation). All other program control modes user will be able to set using program settings.

Another essential feature of the control program is the possibility to make and to change voice commands and the possibility to change reactions to voice commands using the button "Nustatymai" (Settings) (Fig. 1). The view of the window "Nustatymai" in the dictation mode is shown in Fig. 2.

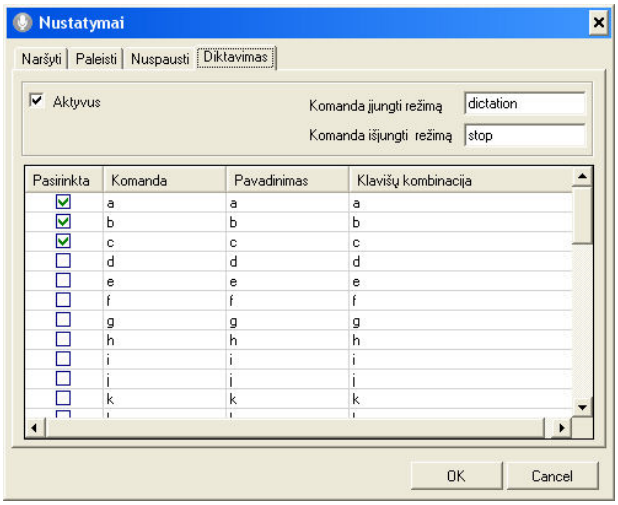

Fig. 2. The view of the window "Nustatymai" (Settings) in the dictation mode

Voice commands and corresponding actions could be grouped into several groups:

Commands for starting of programs (the button „Paleisti" in the Fig. 2). After recognition of some predefined voice commands the actions to start executable files (exe type) are initiated. For example, such command could be used to start internet browser. Also commands of this type could be used to control external devices;

Commands for opening of websites (the button „Naršyti"" in the Fig. 2). After recognition of this type of command website which is attributed to the command is

opened;

Control of key combinations (the button „Nuspausti““ in the Fig. 2). Simulation by control key combinations enables to control with voice commands the operation of program which has corresponding key combinations. The same key combinations are assigned to the commands of the same purpose in Windows operating system independently from program (e.g., ALT+F4 – to close any window, CTRL+S – in all programs means save the data, CTRL+N – in all browsers opens new window). It is important that in program of our interest as many combinations as possible would be implemented. Realization of browser control by voice is simpler after implementation of key combination simulation;

Input of words by spelling (the button „Diktavimas“ in the Fig. 2). System could be switched to the dictation mode using voice command „Dictation“ or pressing the key „Diktavimas“ (Fig.1). In the dictation mode necessary symbols and letters have corresponding voice commands. For example, if we want to write word „four“ we spell word by letters. One of the possible applications is the input of website name into address toolbar of browser: for example, the handicapped person pronounces first voice command „internet“ – internet browser starts, after second voice command „navigate“ - internet browser sets focus to address field, after third voice command „dictation“ – the dictation mode tunes, after that the website name is set by spelling all internet address with the dots, after the next voice command „stop“ - the command mode tunes and after the last voice command „enter“ – the website is opened.

Four fields are presented in the window “Nustatymai“ in the dictation mode (Fig. 2):

- „Pasirinkta“ (Selected) – the new voice command could be turned on or turned off in this field;
- „Komanda“ (Command) – voice command which is used in the recognition process;
- „Pavadinimas“ (Title) – the title of voice command;
- „Klavišų kombinacija“ (Key combination) – the keys which are pressed after the recognition of voice command.

It is obviuosly, that the recognition accuracy of letters could be considerably improved using the appropriate set of words, as in the Morse alphabet (for example, „sun“ is used for „s“, „dog“ is used for „d“ and so on).

The fields in the window „Nustatymai“ in the command mode are similar to the ones presented in Fig. 2, except the field „Adresas“ (address) in the settings of commands for opening of websites and the field „Kelias“ (path) in the settings of commands for starting of programs.

Program „Balsas“ has the possibility to use program plug-ins which could extend functions of program and provide possibilities to control necessary programs or external devices by voice. All installed plug-ins are registered during system startup. It is possible to establish voice commands in the program settings which could be transferred to select plug-in. During operation the program will recognize voice commands and will transfer further operation to plug-in associated with the program. Plug-ins will be realized as DLL libraries (Dynamic Link Library) and will be placed into the appropriate directory. This provides opportunity to write plug-in using any available programming language. Input data for plug-in are recognized phrases and the purpose of plug-in is the realization of expected reaction: this could be control of specialized program with some necessary logic or control of external device via USB or COM ports. It is enough to create new plug-in when adapting the program for new device. When new device will be attached and new plug-in will be installed the program immediately will begin to respond to the implemented commands. User could change voice commands and activate or deactivate them in program settings.

The program “Balsas” employs Microsoft Speech SDK 5.1 version which is compatible with older Windows versions (Windows XP, Windows 2000, and Windows 98). SAPI 5.3 version is provided only with the Windows Vista operating system so it is still unreasonable to use this version. New speech technologies library codenamed Speech FX is intended to use under Microsoft NET environment. Unfortunately this library is included only to the newest version of .NET Framework 3.0 so Speech FX is not used for speech recognition. NET technology - Microsoft Visual Studio‘2005 programming environment with Microsoft .NET Framework 2.0 and Visual C# programming language were selected for realization of program.

Software and hardware for control of electric devices by voice

The Universal Serial Bus (USB) is a fast and flexible interface for connecting devices to computers. Every new PC has at least a couple of USB ports that you can connect to a keyboard, mouse, scanners, external disk drives, printers, and standard and custom hardware of all kinds. Inexpensive hubs enable you to add more ports and peripherals as needed.

The human interface device (HID) class was one of the first USB classes to be supported under Windows. On PCs running Windows 98 or later, applications can communicate with HIDs using the drivers built into the operating system. For this reason, USB devices that fit into the HID class are some of the easiest to get up and running. The designation *human interface* suggests that the device interacts directly with people. A device may detect when someone presses a key or moves a mouse or joystick, or the host may send a message that translates to a joystick effect that the user experiences. The classic examples of HIDs are keyboards, mice, and joysticks. Other HIDs include front panels with knobs, switches, buttons, and sliders; remote controls; telephone keypads; and game controls such as steering wheels.

Windows provides two ways for applications to communicate directly with HIDs: Windows API functions and the APIs supported by DirectX. Communicating with a HID isn't as simple as opening a port, setting a few parameters, and then reading and writing data, as you can do with RS-232 and parallel ports. Before an application

can exchange data with a HID, it has to identify the device and get information about its reports. To do this, the application has to jump through a few hoops by calling a series of API functions. The application first finds out what HIDs are attached to the system. It then examines information about each HID until it finds one with the desired attributes. For a custom device, the application can search for specific Vendor and Product IDs or the application can search for a device of a particular type, such as a mouse or joystick. After finding a device, the application can exchange information with it by sending and receiving reports [10].

While the universal program for computer and electric devices control by voice commands is under development, the simple recognition program "Simple Dictation" was adapted for control of electric devices by voice. This program is supplied together with Microsoft Speech SDK. HID API functions for HID device searching, sending and receiving reports were added to this program. USB interface board VM110 and relay board VM129 were used as hardware for control of electric devices by voice [11]. The developed system can turn on or turn off eight electric devices by voice commands, but it can be easily supplemented with other functions, because USB interface board has digital inputs and analog inputs/outputs too.

The projection based speech recognition algorithm

Lithuanian speech recognition engine "ARVRKRPK Lithuanian Recognizer" was created using speaker dependent projection based recognition algorithm [12]. This algorithm was developed as a highly phonetically motivated alternative to DTW (Dynamic Time Warping) or HMM (Hidden Markov Models) speech recognition models [13]. The projection method is based on phonetically segmented speech features. The segmentation function is smoothed using linear and nonlinear filtering. Each utterance then is divided to stationary and transient phoneme like segments. The average number of the stationary - transient segment pairs is approximately equal to number of phonemes per utterance. The stationary and transient markers of the three different utterances of the same word, which has the following phonetic content: fricative-1st diphthong-sonant-2nd diphthong, are presented in Fig.3. The only single phonetic segmentation error in the right column between 1st diphthong and sonant is observed. The internal segmentation of diphthong was correct.

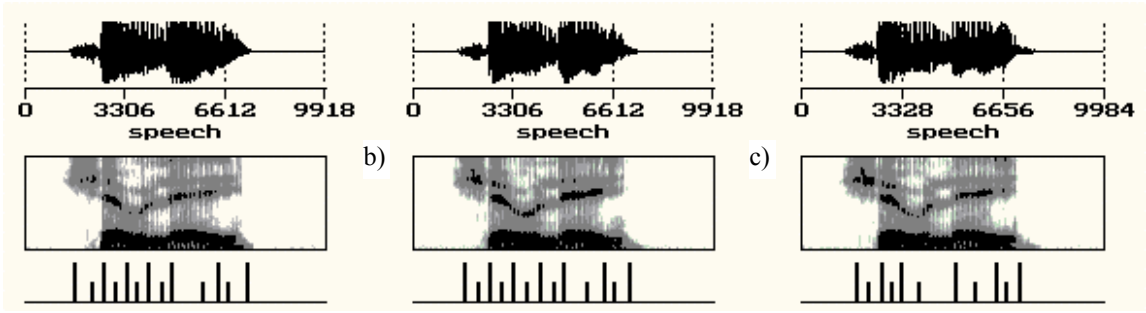

Fig. 3. Three phonetically segmented speech examples: from top: a – speech, b – sonogram, c – stationary (lower) and transient (higher) segment markers

The comparison of test and reference patterns is performed using local similarity matrix where only segmented feature vectors are implemented. The projection is the max value of similarity, bound with the given row or column of similarity matrix and satisfying the restrictions of the global similarity measure. The essence of the restrictions: the global similarity measure is evaluated only from 2 or 3 diagonals, which show the greatest sum of projections.

The projection algorithm outperforms DTW (error decreases about twice) for the same phonetically segmented sequences [12]. The projection algorithm allowed to implement the "averaging" (the repetitions of a given phonetically segmented phrase). This feature can be applied for the averaging of many pronunciations of the same word and, in this way, it could help to create the more reliable word references or to detect and to collect the same phonetic units automatically [14]. The original method of detecting boundaries of words was used in projection based recognition algorithm [15].

Investigation of Lithuanian commands recognition

The program for computer control by voice commands "Balsas" is in the stage of developing: the set of voice commands with the best recognition accuracy should be determined. Another set of words should be defined for the reliable input of words by spelling. Then the investigation of selected Lithuanian voice commands recognition accuracy should be performed.

The program for control of electric devices by voice was tested with Lithuanian and English speech recognition engines: ten Lithuanian commands of electric devices switching were chosen for recognition accuracy investigation (Table 1). All commands differ by acoustic content: instead of using the similar commands "turn-on the lamp", "turn-off the lamp", unlike commands "fire the lamp", "out the light" are used.

Lithuanian speech recognition engine "ARVRKRPK Lithuanian Recognizer" and *Microsoft English US 6.1* recognition engine were used in the investigation of recognition accuracy. In order to improve the recognition

quality of Lithuanian commands recognition by English recognition engine, it is possible to use English transcriptions of Lithuanian words. This way English speech recognition engine interprets spoken word as English one and sometimes the improvement is quite noticeable. English transcriptions of Lithuanian commands were chosen using Microsoft English speech synthesizer: each Lithuanian command was synthesized and the most similar to Lithuanian pronunciation English transcription of command was selected.

Table 1. Lithuanian commands and their english transcriptions

Electric device	Command	English transcription
Lamp	*Uždek lempą (turn on)*	*uzhdeck lemmpah*
	Gesink šviesą (turn off)	*geh sink shwieessaah*
Radio	*Paleisk muziką (turn on)*	*pah leighsk muzzikaah*
	Stabdyk radiją (turn off)	*stabdeek muzzikaah*
Voltmeter	*Pamatuok įtampą (turn on)*	*paamaattuoook eetaampaah*
	Baik matavimą (turn off)	*bike matawymaah*
Oscillo-graph	*Parodyk grafiką (turn on)*	*paarrohdeek graaffickaah*
	Išjunk oscilografą (turn off)	*ishjuuunk vostsellogrrhaaa ffaa*
Recorder	*Įrašyk signalą (turn on)*	*eerrasheeck signaaalaah*
	Nutrauk įrašymą (turn off)	*nuhtrraauk eerraaashimaah*

The accuracy of Lithuanian commands recognition by Lithuanian and English speech recognition engines is shown in the Fig. 4.

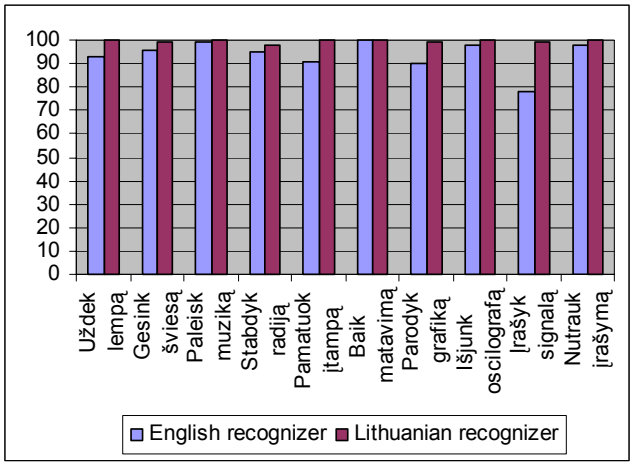

Fig. 4. Recognition accuracy of Lithuanian commands

The average accuracy of Lithuanian commands recognition by English recognizer was 93.8%, the average accuracy of Lithuanian commands recognition by Lithuanian recognizer – 99.5%. The training procedure is needful for Lithuanian recognizer while English recognizer is speaker-independent.

Conclusions

Two programs were developed during the first stage of the project „Lithuanian multimodal voice services“: one – for computer control by voice (Internet Explorer or any other computer program, which supports special keyboard combinations could be controlled by voice commands, input of words by spelling could be performed by voice), second – for switching of electric devices by voice (electric devices could be turned on or off by voice with additional USB-based hardware). Both programs are SAPI-based and can work with any Windows operation system‘s recognition engine.

The accuracy of recognition of ten Lithuanian commands by Lithuanian and English speech recognition engines was checked: the average accuracy of Lithuanian commands recognition by English recognizer was 93.8%, the average accuracy of Lithuanian commands recognition by Lithuanian recognizer – 99.5%. The training procedure is necessary for Lithuanian recognizer while English recognizer is speaker-independent.

The universal program for computer and electric devices control by voice commands should be developed till 2009 and used for disabled people.

References

1. **Hawley M. S., Green P., Enderby P., Cunningham S., Moore R. K.** Speech Technology for e-Inclusion of People with Physical Disabilities and Disordered Speech // Proc. Interspeech. – Lisbon. – 2005. – P. 445–448.
2. **Brøndsted T., Aaskoven E.** Voice-Controlled Internet Browsing for Motor-Handicapped Users. Design and Implementation Issues // Proc. Interspeech. – Lisbon. – 2005. – P. 185–188.
3. **Vovos A., Kladis B., Fakotakis N.** Speech Operated Smart-Home Control System for Users with Special Needs // Proc. Interspeech. – Lisbon. – 2005. – P. 193–196.
4. **Jitsuhiro T., Matsuda S., Ashikari Y., Nakamura S., Yairi I. E., Igi S.**. Spoken Dialog System and its Evaluation of Geographic Information System for Elderly Persons' Mobility Support // Proc. Interspeech. – Lisbon. – 2005. – P. 197–200.
5. **Granström B.** Speech Technology for Language Training and e-Inclusion // Proc. Interspeech.– Lisbon. – 2005. – P. 449–452.
6. **Kvale K., Warakagoda N.** A Speech Centric Mobile Multimodal Service Useful for Dyslectics and Aphasics // Proc. Interspeech.– Lisbon. – 2005. – P. 461–464.
7. **Rudžionis A., Ratkevičius K., Rudžionis V.** Voice based internet services // Electronics and Electrical Engineering. – Kaunas: Technologija, 2004. – No. 3(52). – P. 5–9.
8. **Juozėnas A., Kasparaitis P., Ratkevičius K., Rudinskas D., Rudžionis A., Rudžionis V., Sidaras S.** DfA Implementations for People with Vision and Hearing Disabilities: Application and Development for Information Society // 4th International Conference on Universal Access in Human-Computer Interaction. – UAHCI 2007. – Beijing, China, July 22-27, 2007. – P. 686–695.

9. **Cox R. V., Kamm C. A., Rabiner L. R., Schroeter J., Wilpon J. G.** Speech and Language Processing for Next-Millennium Communications Services // Proceedings of the IEEEE. – August, 2000. – Vol. 88, No. 8. – P. 1314–1337.
10. **Anderson D.** USB System Architecture (USB 2.0) // Addison-Wesley Developer's Press, 2001. – 506 p.
11. **Velleman** Components N.V. Accessed at: http://www.velleman.be. – 2008.
12. **Rudzionis A.**. Isolated word recognition by fully phonetical word template // Contribution to the COST232 final report. – 1994. – P. 11–13.
13. **Rudzionis A.** Recognition by averaged templates // COST249: "Continuous Speech Recognition Over the Telephone", draft minutes of the 1st Management Committee Meeting. – Brussel, Belgium, 1994. – P. 41–47.
14. **Rudzionis A., Rudzionis V.** Noisy speech detection and endpointing // Voice operated telecom services. Do they have a bright future? Workshop Proceedings. – May 11-12, 2000. – Ghent, Belgium. – P. 79–82.

Submitted for publication 2008 02 15

A. Rudžionis, K. Ratkevičius, T. Dumbliauskas, V. Rudžionis. Control of Computer and Electric Devices by Voice // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008. – No. 6(86). – P. 11–16.

Paper deals with the speech recognition software for controlling computer programs and electric devices by voice. Internet Explorer or any other computer program, which supports special keyboard combinations could be controlled by voice commands. The essential features of the control program are: the possibility to work with any installed in the system speech recognition engine, the possibility to make and to change voice commands and to change reactions to voice commands, the possibility to input the text by spelling. Electric devices could be turned on or turned off by voice with additional USB-based hardware. Speaker dependent projection based recognition algorithm, which was used in Lithuanian speech recognition engine, is described. The accuracy of recognition of ten Lithuanian commands by Lithuanian and English speech recognition engines was checked. Demonstrations of computer and electric devices control by voice commands are prepared. Ill. 4, bibl. 14 (in English; summaries in English, Russian and Lithuanian).

А. Руджёнис, К. Раткявичюс, Т. Думбляускас, В. Руджёнис. Управление компьютера и электрических приборов голосом // Электроника и электротехника. – Каунас: Технология, 2008. – № 6(86). – С. 11–16.

Анализируются программы для управления компьютером и электрическими приборами голосом. Любая компьютерная программа, поддерживающая специальные управляющие комбинации клавишов, может быть управляема с помощью голосовых команд. Особенности программы управления: возможность работы с любым инсталлированным распознавателем, возможность вводить новые команды и менять реакцию на них, возможность вводить текст по одной букве. Электрические приборы включаются или выключаются голосом при помощи дополнительного устройства, основанного на USB интерфейсе. Описан зависимый от диктора алгоритм распознавания речи, на основе которого реализован первый литовский распознаватель речи. Представлены результаты исследования точности распознавания десяти литовских голосовых команд с распознавателями английского и литовского языков. Подготовлены демонстрации программ управления компьютером и электрическими приборами. Ил. 4, библ. 14 (на английском языке; рефераты на английском, русском и литовском яз.).

A. Rudžionis, K. Ratkevičius, T. Dumbliauskas, V. Rudžionis. Kompiuterio ir elektros prietaisų valdymas balsu // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2008. – No. 6(86). – P. 11–16.

Nagrinėjama programinė įranga kompiuteriui valdyti ir elektros prietaisams įjungti ir išjungti balsu. Interneto naršyklė ar bet kuri kita kompiuterio programa, palaikanti specialias klaviatūros klavišų valdymo kombinacijas, gali būti valdoma balso komandomis. Kompiuterio valdymo balsu programos esminės ypatybės: galimybė dirbti su bet kuriuo kompiuteryje įdiegtu atpažinikliu, galimybė įvesti naujas balso komandas bei keisti reakciją į balso komandas, galimybė paraidžiui įvesti tekstą. Elektros prietaisai gali būti įjungiami arba išjungiami balso komandomis per papildomą USB įtaisą. Trumpai supažindinama su atpažinimo algoritmu, panaudotu ruošiant lietuvių kalbos atpažiniklį. Pristatomi dešimties lietuviškų balso komandų atpažinimo su anglų kalbos ir lietuvių kalbos atpažininkliais tikslumo tyrimo rezultatai. Paruoštos kompiuterio valdymo balsu ir elektros prietaisų įjungimo ir išjungimo balsu demonstracijos. Il. 4, bibl. 14 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).