ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2009. No. 3(91)
ELEKTRONIKA IR ELEKTROTECHNIKA

T170 ELECTRONICS
ELEKTRONIKA

Investigation of Foreign Languages Models for Lithuanian Speech Recognition

R. Maskeliūnas, 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
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

It is well known fact that speech recognition based interfaces could be of great value in many applications. This is particularly true for the applications oriented to the telecommunication users. A speech input interface based on speech recognition technology has already been applied to many applications. Particular benefits could be achieved by disabled people. Automatic speech recognition 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].

Also well known fact is that progress in the field of speech technologies is slow and costly in various senses. Most of the commercial applications are oriented to the users of languages with many native speakers. First of the all this fact shows that successful implementation of speech based systems requires large intellectual and financial investments.

In some of our previous papers we presented several prototype systems using Lithuanian speech technologies. In [2] were investigated possibilities to use Microsoft Speech Server'2004 for Lithuanian speech applications. English speech engine was used in that study to recognize several Lithuanian voice commands, but the recognition accuracy of Lithuanian words was too low for practical applications.

From the advent of speech recognition research and the appearance of first commercial applications the main efforts were devoted to the recognition of widely used languages, particularly English language. The reason of such behavior is very clear – popular widely used languages have bigger market potential for practical applications. So looking at the general trend in the development of commercial speech recognition applications and tools for the development of speech recognition using information systems next sequence could be observed: first version of speech recognition engine oriented to the recognition of English (and particularly US English) is released, then speech recognition system is supplemented with the engines for the recognition of other widely used languages (most often Spanish, French, German and several others) and sometimes but not necessarily with recognition modules of some other relatively widely used languages (in example Dutch, Italian, Turkish, Polish, etc.). Many other less widely used languages remains out of the scope of interest for the major speech recognition solution providers.

In such situation businesses and state institutions in countries were such less popular languages are used as a main source of spoken language communication faces a challenge of development of own speech recognition tools. Two major ways for solution are as follows:

- to develop own speech recognition engine from the scratch;
- to adapt foreign language based engine for the recognition of your native language.

The first approach has potentially higher capabilities to exploit peculiarities of selected language and hence to achieve higher recognition accuracy. But the drawbacks of such approach are the same that major speech technologies providers avoid the implementation of such languages in their products – high costs in the general sense of this word.

The second approach [3] has the potential to achieve some practically acceptable results faster than developing entirely new speech recognition engine. Another advantage of this approach is potential to achieve faster compatibility with the existing technological platforms. Such advantage is often important for business customers since they need to follow various technical specifications in order to guarantee consistent functioning of enterprise. But this approach also requires careful investigation of the ways of adapting and optimizing adaptation algorithms.

This paper presents our activities to adapt and compare several foreign language (English, Spanish, French, German) speech recognition engines for the recognition of limited Lithuanian vocabulary (digits). Recognizing of digits is a common task used by wide variety of applications. Gathering policy numbers, catalog numbers, license plate numbers, etc., are typical tasks that involve digits recognition. Main features of the new version of Microsoft Speech Server - Office Communications Server Speech Server (MSS'2007) are described and the results of experiments carried on this server are presented.

Speech for telephony – Office Communications Server Speech Server

Part of the Microsoft Office Communications Server (MOCS) package, responsible for speech interface control is Microsoft Speech Server (MSS'2007) [4]. Previous version of Microsoft Speech Server MSS'2004 was presented in [5] together with IBM WebSphere Voice Server [6]. Speech Server MSS'2007 is an interactive voice response (IVR) platform that integrates with Visual Studio 2005 or Visual Studio 2008. MSS'2007 provides tools for developing applications that run over a telephone, or *telephony applications*. For example, telephony applications let you check your bank balance via a telephone or get an automated call from your doctor's office reminding you of your next appointment. MSS'2007 is to a telephony application what a web server such as Internet Information Services (IIS) is to a web application. MSS'2007 applications can have the following capabilities:

- speech recognition allows users to respond to application prompts;
- Touch-Tone capabilities, called dual-tone multi-frequency (DTMF), let users respond to application prompts via the telephone keypad;
- text-to-speech (TTS) capabilities allow applications to read and speak written text to users.

MSS'2007 supports a powerful Microsoft .NET Framework-based application programming interface (Voice Response and Windows Workflow) for low-level access to core Speech Server functionalities (latest version of MSS also supports VoiceXML and SALT (Speech Application Language Tags)) [7]. By utilizing Windows Workflow software developer can write managed code and actually see the entire call flow. Microsoft's .NET Framework is a runtime environment and class library that dramatically simplifies the development and deployment of modern, component-based applications.

In Visual Studio, developer may build Voice Response Workflow application framework in the Dialog Workflow Designer (Fig. 1): speech dialog component *answerCallActivity1* answers an incoming call, *questionAnswerActivity1* – asks the question and gets the user's answer, *gotoActivity1* – jumps to another component, *disconnectCallActivity1* – disconnects an existing call. Such framework is suitable for testing of Lithuanian digits recognition by selected speech recognizer. The prompt, grammar and target properties of *questionAnswerActivity1* and *gotoActivity1* speech dialog components should be defined before the testing procedure.

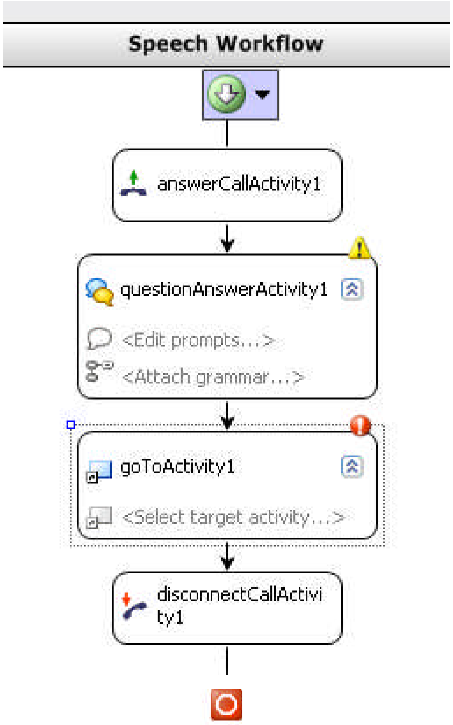
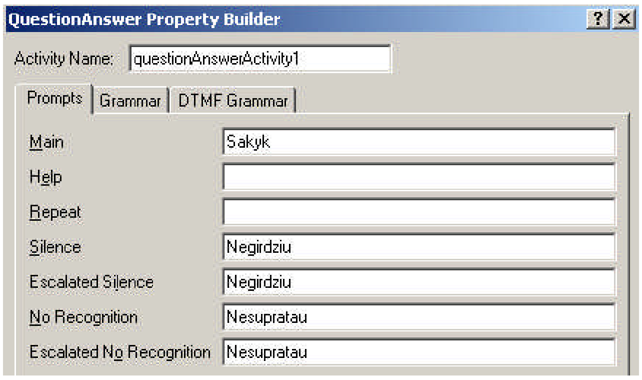


Fig. 1. The view of Dialog workflow designer window

The simplest way to prepare the prompts – to input the desirable text for the synthesizer in the QuestionAnswer Property Builder (Fig.2): the *Main prompt* is the first prompt played after the control is selected. The *Silence prompt* is played when the user remains silent for the duration. If the user gives input that isn't recognized by the specified grammar, the *No Recognition prompt* will play.



Activity Name: questionAnswerActivity1

Main	Sakyk
Help	
Repeat	
Silence	Negirdziu
Escalated Silence	Negirdziu
No Recognition	Nesupratau
Escalated No Recognition	Nesupratau

Fig. 2. The view of QuestionAnswer Property Builder window

For an application to recognize speech, it needs to know which words and phrases to expect. Building grammar involves compiling lists of predicted user responses to specific prompts. Grammar is stored in a W3C-compliant format called Speech Recognition

Grammar Specification (SRGS), specifically in an XML format known as Grammar XML (GRXML). When speech from grammar is recognized, it produces the results in an XML format called Semantic Markup Language (SML). When developing an IVR application, building grammar is the most time-consuming piece. It will take more time than writing the application itself.

Latest MSS version (2007) supports simplified tool for building grammars – Conversational Grammar Builder (CGB). CGB is most appropriate for developing grammars, where the user's answers are single keywords. Conversational Grammar Builder can also be used to develop grammars that use natural language understanding, or Conversational Grammar grammars.

More advanced Speech Grammar Editor (SGE) is provided for developing grammars where the user's answers are more complex than a single phrase, for example, a grammar used to recognize the Lithuanian digit “du” (Fig. 3): two transcriptions of Lithuanian digit “du” are involved in the SGE providing two alternatives for the recognition of this digit. Another application for Speech Grammar Editor is speech recognition scenarios involving mixed-initiative user responses, where the user can provide information that the system has not yet asked for.

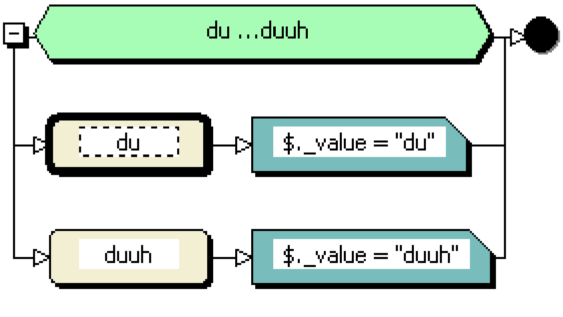


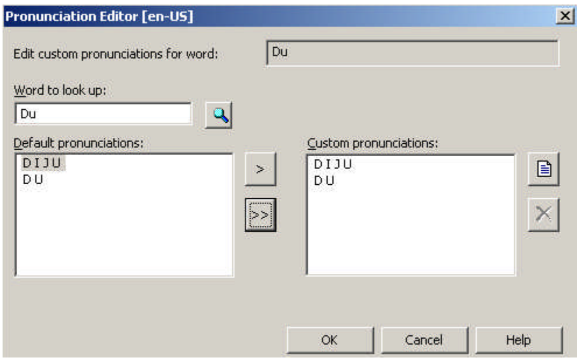
Fig. 3. The view of Speech grammar editor window

Also it is possible to customize recognized pronunciations in speech applications running on Speech Server. Lexicons are supported by Speech Grammar Editor and Conversational Grammar Builder. Any word could be placed in the field “Word to look up” in the Pronunciation editor (Fig. 4) and the proposed transcriptions of this word would be generated according the UPS (Universal Phone Set) alphabet [8] and placed in the field “Default pronunciations”. The user can add any transcription to the recognition grammar by copying it to the field “Custom pronunciations”.

The made-up transcriptions could be added in the field “Custom pronunciations” using UPS phoneme and prosodic symbol labels which are specific for each language. The primary stress S1, the secondary stress S2, the syllable break and other prosody UPS labels could be used in the transcriptions.

Currently, the following language packs are available for speech and DTMF recognition:

- English (United States);
- English (United Kingdom);
- German;
- French (Canada);
- Spanish (United States).



Edit custom pronunciations for word: Du

Word to look up: Du

Default pronunciations:
DIJU
DU

Custom pronunciations:
DIJU
DU

Fig. 4. The view of Pronunciation Editor window

Chinese, Italian, Japanese, Korean and Portuguese language packs supports only DTMF recognition and text-to-speech synthesis.

Investigation of Lithuanian digits recognition

Investigation of Lithuanian digits recognition by English recognizers was carried out for many years. In order to improve the accuracy of Lithuanian digits recognition by recognition engines of other languages, it is possible to use foreign transcriptions of Lithuanian words. This way speech recognition engine of foreign language interprets spoken word as native one and sometimes the improvement is quite noticeable.

Ten Lithuanian digits from 0 to 9 were chosen to investigate recognition accuracy. Very high averaged accuracies of Lithuanian digits recognition by *Microsoft English (U.S.) v6.1* recognizer were achieved using trained speaker profile and experimentally chosen recognizer settings in 2008 for one speaker (MR, woman):

- 99.8% - using many transcriptions for each digit;
- 99.0% - using one transcription for each digit.

SAPI-based (*SAPI – Speech Application Programming Interface*) transcriptions were used in above mentioned experiment, for example, the transcription of digit “0” was *nuhlihs.*

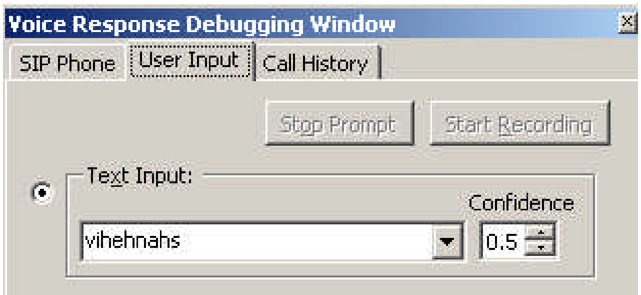
In order to check the suitability of selected transcriptions for MSS’2007, the test for the measuring of the accuracy of the digits recognition was prepared on MSS’2007. One transcription for each digit from the previous experiment was used for the measuring of the accuracy of digits by English recognition engine (*Microsoft Speech Recognizer 9.0 for MSS (English – US)*). Four speakers took part in the experiment: each digit was spoken 100 times through the mobile telephone. The averaged results of recognition accuracy of Lithuanian digits are presented in table 1.

Low recognition accuracy achieved by English recognition engine indicates that used transcriptions are not suitable for Lithuanian digits recognition on MSS’2007. Universal Phone Set (UPS) [7] should be used on MSS’2007 instead of SAPI-based alphabet for Lithuanian digits transcription.

Table 1. Accuracy of Lithuanian digits recognition by English recognizer through the mobile telephone

Speaker	Recognition accuracy, %
RM, man	64.2
VR, man	52,1
KR, man	61,0
MR, woman	41,0
Average	54.6

Before the investigation of UPS transcriptions the attempt to find the most suitable recognition engine for Lithuanian digits recognition was done. German (*Microsoft Speech Recognizer 9.0 for MSS (German-Germany)*), English (*Microsoft Speech Recognizer 9.0 for MSS (English-US)*), French (*Microsoft Speech Recognizer 9.0 for MSS (French-Canada))* and Spanish *(Microsoft Speech Recognizer 9.0 for MSS (Spanish-US)*) recognizers were used in next experiment. Foreign transcriptions of Lithuanian digits were chosen using speech synthesizers of foreign languages [9]: each Lithuanian digit was synthesized and the most similar to Lithuanian pronunciation foreign transcriptions of digit were selected. The number of transcriptions varied from five transcriptions for short digits (2) to fifteen transcriptions for long digits (7, 8). Separate recognition grammars were prepared for each digit and for each language using SGE: overall forty recognition grammars. Forty Lithuanian digits recognition tests were carried out with each recognition grammar in order to find the most suitable transcriptions of each digit. Experiment was conducted in noisy computer room environment, using headset microphone. MSS'2007 debugger was used in this experiment (Fig. 5): it provides the recognized text and the confidence score. By default, the speech recognizer engine returns a single result with the highest confidence score. The recognition rejection threshold, by default, is equal 0.20: the digit is not recognized if the confidence score is less than 0.21.



Text Input:

Confidence

vihehnahs

0.5

Fig. 5. The view of Voice Response Debugging window

Each digit was spoken 100 times by one speaker-man and one speaker-woman and the recognized transcriptions were calculated. The transcriptions which were recognized mostly times were selected and used in the next experiments. As the example of the given results, the best transcriptions of Lithuanian digit "nulis" for German, English, French and Spanish recognizers are presented in the table 2. In that case if two or three transcriptions of one digit received similar evaluations in the previous experiment, all of them were used in the next experiment.

Four recognition grammars were prepared for each language using SGE. Each digit was spoken 100 times by one speaker-man and one speaker-woman and recognized digits together with the confidence scores were recorded.

The averaged accuracy and the averaged confidence score of Lithuanian digits recognition by German, English, French and Spanish recognizers are shown in the tables 3 and 4.

Table 2. The best transcriptions of Lithuanian digit "nulis" for German, English, French and Spanish recognizers

	German	English	French	Spanish
0	*nuhlihs*	*nulis*	*nouluece*	*nuhlihs*

Table 3. The averaged accuracy of Lithuanian digits recognition by German, English, French and Spanish recognizers

Speaker	German	English	French	Spanish
KR, man	58.4	76.4	52.8	88.2
GB, woman	51.8	59.0	76.8	98.8
Average	55.1	67.7	64.8	93.5

Table 4. The averaged confidence score of Lithuanian digits recognition by German, English, French and Spanish recognizers

Speaker	German	English	French	Spanish
KR, man	0.44	0.48	0.48	0.60
GB, woman	0.42	0.37	0.57	0.77
Average	0.43	0.43	0.53	0.68

The best averaged accuracy and the best averaged confidence score of Lithuanian digits recognition was achieved by Spanish recognizer. The big difference between the results of speaker-man and speaker-woman is noticeable: the averaged accuracy of Lithuanian digits recognition by Spanish recognizer for speaker-man is 88.2% when the averaged accuracy of Lithuanian digits recognition by Spanish recognizer for speaker-woman – 99.8%.

The accuracy and the confidence score of separate Lithuanian digits recognition by German, English, French and Spanish recognizers are presented in the figures 6 and 7. Lithuanian digits "vienas", "du", "penki" and "devyni" were recognized without errors by Spanish recognizer and the digit "du" – by French recognizer. The confidence score of these digits recognition validates the previous conclusion.

Lithuanian digit "keturi" was not recognized by German and English recognizers. Lithuanian digits which were not recognized or were recognized poorly:

"2" – German recognizer (speaker-woman);

"3" – German and Spanish recognizer (speaker-man);

"4" – all recognizers except Spanish recognizer (speaker-woman);

"5" – German and French recognizers, English recognizer (speaker-man);

"6" – German recognizer (speaker-woman);

"7" – English recognizer (speaker-woman), French recognizer (speaker-man);

"8" – French recognizer (speaker-man).

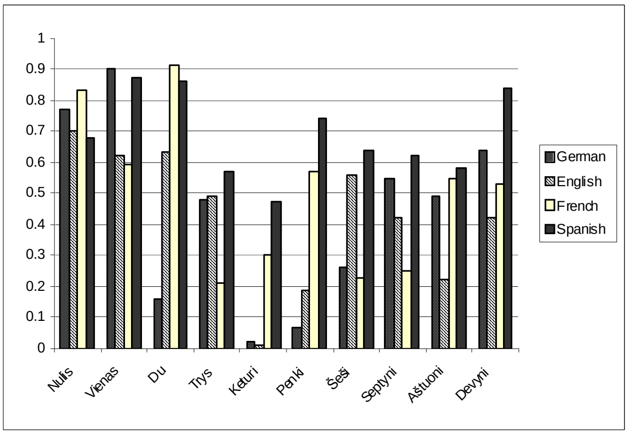


Fig. 6. Confidence score of Lithuanian digits recognition by German, English, French and Spanish recognizers

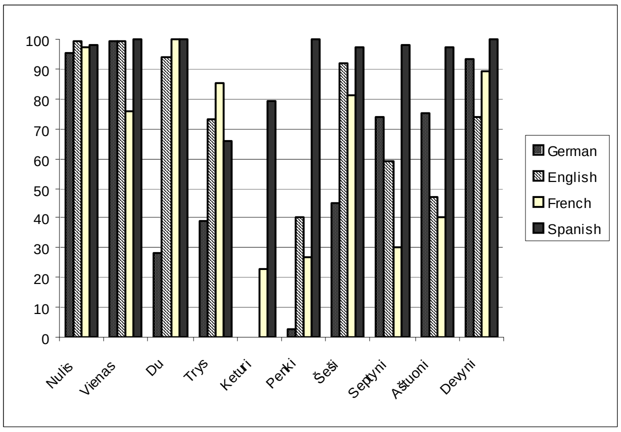


Fig. 7. Accuracy of Lithuanian digits recognition by German, English, French and Spanish recognizers

It would be difficult to analyze all cases of bad recognition, so only Spanish recognizer was examined further. The accuracy of Lithuanian digits recognition by Spanish recognizer for speaker-man and speaker-woman are presented in the Fig. 8.

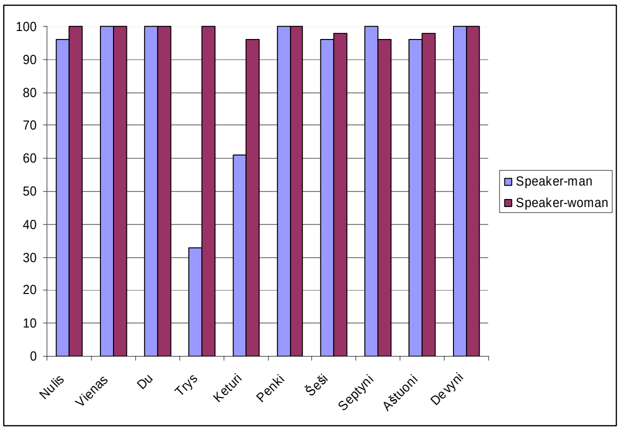


Fig. 8. Accuracy of Lithuanian digits recognition by Spanish recognizer

There are only two cases of bad recognition of Lithuanian digits by Spanish recognizer (Fig. 8):

"3" – Spanish recognizer (speaker-man), recognition accuracy - 31%, the confidence score - 0.31;

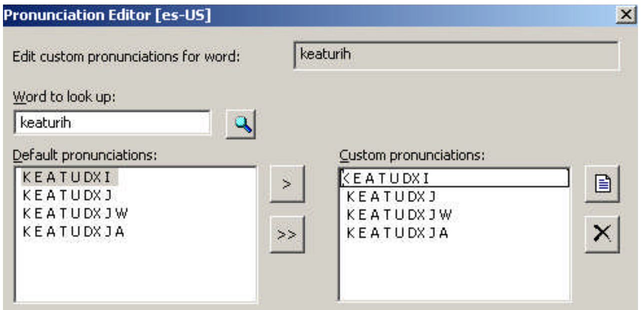
"4" – Spanish recognizer (speaker-man), recognition accuracy - 61%, the confidence score - 0.24.

The accuracies of recognition of these digits for speaker-woman were very high:

"3" – Spanish recognizer (speaker-woman), recognition accuracy - 100%, the confidence score - 0.83;

"4" – Spanish recognizer (speaker-woman), recognition accuracy - 96%, the confidence score - 0.70.

In order to improve the recognition accuracy of Lithuanian digits "trys" and "keturi" the recognition grammars of these digits were changed by copying the SGE generated default transcriptions of these digits from the field "Default pronunciations" to the field "Custom pronunciations" (Fig. 4). The example of such operation for the transcription "keaturih" of the digit "keturi" is shown in the Fig. 9.


Pronunciation Editor [es-US]

Edit custom pronunciations for word: keaturih

Word to look up: keaturih

Default pronunciations:
KEATUDXI
KEATUDXJ
KEATUDXJW
KEATUDXJA

Custom pronunciations:
KEATUDXI
KEATUDXJ
KEATUDXJW
KEATUDXJA

Fig. 9. Custom pronunciations of the transcription „keaturih" of the digit „keturi" of Spanish recognizer

The experiment of recognition accuracy measuring was repeated: Lithuanian digits "trys" and "keturi" were spoken 100 times by the same speakers and recognized digits together with the confidence scores were recorded.

The accuracy of recognition of Lithuanian digit "trys" by Spanish recognizer using the custom pronunciations of this digit increased from 31% to 70% and the accuracy of recognition of Lithuanian digit "keturi" increased from 61% to 88% for speaker-man. The same high results as in previous experiment were received for speaker-woman. These results confirm that the accuracy of Lithuanian digits recognition by Spanish recognizer could be increased using the custom pronunciations of Lithuanian digits prepared using Universal Phone Set (UPS) labels.

Conclusions

Voice servers integrate together telephony, speech and internet and provide tools for developing applications that run over a telephone. Using of transcriptions of Lithuanian words so far is the only solution of voice server application for Lithuanian language. Main features of the new version of *Microsoft Speech Server - Office Communications Server Speech Server (MSS'2007)* are described and the results of experiments carried on this server are presented.

The accuracy of recognition of ten Lithuanian digits by German, English, French and Spanish recognizers was checked on voice server MSS'2007. The best accuracy of Lithuanian digits recognition (93.5%) was achieved by Spanish recognizer. It could be increased using the custom pronunciations of Lithuanian digits prepared using Universal Phone Set (UPS) labels.

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. **Rudžionis A., Ratkevičius K., Maskeliūnas R., Rudžionis V.** Investigation of Voice Server Applications for Lithuanian Language // Electronics and Electrical Engineering. – Kaunas: Technologija, 2007. – No. 6(78). – P. 46–49.
3. **Zgank A., et al.** The COST278 MASPER initiative – crooslingual speech recognition with large telephone databases // Proc. of 4th International Conference on Language Resources and Evaluation LREC'04. – 2004. – P. 2107–2110.
4. **Dunn M.** Pro Microsoft Speech Server 2007: Developing Speech Enabled Applications with .NET. – Jun 2007. – 275 p.
5. **Rudžionis A., Ratkevičius K., Rudžionis V.** Speech in Call and Web centers // Electronics and Electrical Engineering. – Kaunas: Technologija, 2005. – No. 3(59). – P. 58–63.
6. **Xiaole Song.** Comparing Microsoft Speech Server 2004 and IBM WebSphere Voice Server V4.2 [interactive] [February 12, 2009] – Accessed at: http://www.developer.com/voice/ /article.php/3381851.
7. Microsoft .NET Speech Technologies [interactive] [February 12, 2009]. – Accessed at: http://www.microsoft.com/speech.
8. Phoneme Table for English (United States) [interactive] [February 12, 2009]. Accessed at: http://msdn.microsoft.com/ /en-us/library/bb813894.aspx.
9. Free Text to Speech software online for English, Spanish, French, German, Italian, Portuguese, Korean, Japanese, Chinese and Russian languages [interactive] [February 12, 2009]. Accessed at: http://text-to-speech.imtranslator.net.

Received 2009 02 14

R. Maskeliūnas, A. Rudžionis, K. Ratkevičius, V. Rudžionis. Investigation of Foreign Languages Models for Lithuanian Speech Recognition // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 3(91). – P. 15–20.

Paper deals with application of *Microsoft Office Communications Server Speech Server* or *MSS'2007* for Lithuanian digits recognition. Voice servers integrate together telephony, speech and internet and provides tools for developing applications that run over a telephone. Using of transcriptions of Lithuanian words so far is the only solution of voice servers application for Lithuanian language. The results of investigation of Lithuanian digits recognition by German, English, French and Spanish speech recognition engines implemented on MSS'2007 are presented. The best accuracy of Lithuanian digits recognition was achieved by Spanish recognizer. It could be increased using the custom pronunciations of Lithuanian digits prepared using Universal Phone Set (UPS) labels. Demonstration of user identification by telephone using Spanish recognizer is prepared. Ill. 9, bibl. 9 (in English; summaries in English, Russian and Lithuanian).

Р. Маскелюнас, А. Руджёнис, К. Раткявичюс, В. Руджёнис. Исследование по применению моделей зарубежных языков для распознавания литовского языка // Электроника и электротехника. – Каунас: Технология, 2009. – № 3(91). – С. 15–20.

Анализируется возможности использования *Microsoft Office Communications Server Speech Server - MSS'2007* для распознования литовских цифр. Речевые серверы интегрируют вместе телефонию, речь и интернет и предоставляет инструменты для разработки приложений, которые используют телефоны канал. Использование транскрипций литовских слов до сих пор является единственной возможностью создать приложения для речевых серверов на литовском языке. Представлены результаты исследования распознавания литовских цифр с исползованием немецкого, английского, французского и испанского распознавателей для MSS'2007. Лучшие результаты распознования литовских цифр были достигнуты используя испанский распознаватель. Она может быть увеличена с помощью транскрипций, созданых с помощью меток Universal Phone Set (UPS). Демонстрация идентификации пользователя по телефону подготовлена. Ил. 9, библ. 9 (на английском языке; рефераты на английском, русском и литовском яз.).

R. Maskeliūnas, A. Rudžionis, K. Ratkevičius, V. Rudžionis. Užsienio kalbų modelių panaudojimo lietuvių kalbos atpažinimui galimybių tyrimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2009. – No. 3(91). – P. 15–20.

Nagrinėjamos *Microsoft Office Communications Server Speech Server* arba *MSS'2007* balso serverio panaudojimo lietuvių kalbos skaičiams atpažinti galimybės. Balso serveriai apjungia telefoniją, kalbos technologijas ir internetą ir leidžia ruošti taikomąsias telefonines programas. Lietuviškų žodžių užsienietiškų transkripcijų panaudojimas kol kas yra vienintelis balso serverių taikymo lietuvių kalbai būdas. Pateikiami lietuviškų skaičių pavadinimų atpažinimo su vokiečių, anglų, prancūzų ir ispanų kalbų atpažintuvais rezultatai. Nustatyta, kad ispanų kalbos atpažintuvas geriausiai atpažįsta lietuviškų skaičių pavadinimus. Skaičių atpažinimo tikslumas gali būti pagerintas panaudojant transkripcijas, paruoštas su *Universal Phone Set (UPS)* alfabetu. Paruošta vartotojo identifikavimo demonstracija, kurioje lietuviškų skaičių pavadinimai atpažįstami su ispanų kalboas atpažintuvu. Il. 9, bibl. 9 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).