ISSN 1392 – 1215 ELECTRONICS AND ELECTRICAL ENGINEERING 2010. No. 4(100)

ELEKTRONIKA IR ELEKTROTECHNIKA

T 121 SIGNAL TECHNOLOGY

SIGNALŲ TECHNOLOGIJA

Modeling of Call Services for Public Sector

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: algimantas.rudzionis@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

Speech recognition based interfaces could be very useful in many applications. This is particularly true for the applications oriented to the telecommunication users as well as the users of portable and mobile devices. In some situations voice based user interfaces could be single viable solution for the “any-time any-where” information accessibility approach. The tremendous richness of human speech communication gives the user many degrees of freedom for control and input.

The requirements that are more and more often set on the public organizations such as governmental, municipal, healthcare or educational institutions emphasize the necessity to provide the necessary for the user information in a convenient and ubiquitous way. It should be noted that such requirements are often set as judicial documents and public institutions are forced to follow them. It is important to note that such a trend will continue in the upcoming years since it perfectly fits to the general trends of “e-democracy” and openness of the public institutions. Such applications could include the abilities to provide income declarations, to register for the upcoming elections, to get the results of the entrance test to the university and many others. One of the recent examples of such kind in Lithuania is the move to the electronic form of sick-lists that are issued by the healthcare institutions to the state social security foundation. Most and may be even all those electronic services could be realized using traditional graphic user interface, computer and Internet. And such Internet based realizations are the most common way to move the service to the electronic media. But traditional approach not always enable to achieve “anytime anywhere” concept nor always is the most convenient way to provide service. As an example we could provide the electronic forms of the sick-lists: we are very far from the situation when each physician will have computer with Internet connection in each workplace and the abilities to use electronic signatures in each workplace. In such a situation doctor’s need to go for a special places equipped to do this function or healthcare institutions should to hire special staff to do the job of preparation of electronic sick-lists. Additional requirement is the necessity to train the physicians to use electronic system. In any case all those steps increase the expenditures of healthcare institutions. It could be proposed more convenient solution when physician calls to the special data center and dictates necessary data for the sick-list (such as person’s name, duration of illness, code of the disorder, etc.). Then automatic system recognizes commands and automatically generates the sick-list since it is of a predefined form.

Another example of the increased accessibility of a information from public institutions may be the results of the entrance to the university. It is observed that candidates to the future student’s want to know the results of the entrance competition as soon as possible. Despite the fact that all higher educational establishments provide such information on their websites these days for many candidates it could be unavailable immediately since they may be on a trip or in other places. The ability to get this information using mobile phone and speech recognition may provide increased accessibility: after call the user should identify himself/ herself giving name and identification code and system automatically generates and provides to the entrance competition results. It should be noted that voice based user interfaces may increase informational accessibility and provide new level of the user convenience in practically each service provided by the public institutions.

Here one may ask natural question: how complicated speech recognition system should be in such services. Could we need to use continuous or instant speech recognition? Despite the fact that in principle the more natural system will be the higher user convenience level will be achieved in practice we may restrict the vocabulary used in such services with a rather limited set of words.

This restriction may provide essential possibility to increase performance reliability.

Looking at the general requirements to the content of such services we may observe that the essential component of them is the necessity to identify the person (the user, the provider, the patient, etc.). One of the well-known approaches is biometric approach used in [1]. But this approach requires user voice samples and it is inappropriate in many services were we can't have them in advance. Another identification approach is when providing person's name or some assigned code. The identification using names isn't convenient approach due to the unlimited set of the names that may be used for the identification. It should be noted that names themselves doesn't have semantic information and this fact increases the level of complexity: surname should be recognized exactly on a phoneme-to-phoneme basis independently on how similar those phonemes are (e.g., system designers can't give any preference to the name Gražys versus the name Grašys in general).

In this situation recognition using codes formed from the names is better approach. First of all such approach will enable the system designer form vocabulary of limited size. In many cases it is possible to organize pronunciation of the personal identification number using spelling of the digit numbers. In this situation we obtain limited set of ten digit names that could be used to carry out core voice information for the public service. Using this approach we also could implement error – correction codes in some situations to obtain false recognitions and to use necessary measures to avoid misrecognition (e.g., to ask the user to repeat once more one digit instead of asking to repeat whole string of digit names).

Despite the fact how we will organize public service the essential question remains the same – the quality of speech recognition. For several years we are working with the possibilities to adapt foreign language based recognizers for the Lithuanian spoken commands recognition.

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

Businesses and state institutions, in the countries were such less popular languages are used as a main source of the spoken language communication, face a challenge of the development of own speech recognition tools. The two major ways for the solution are as follows:

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

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

The second approach has the potential to achieve some practically acceptable results faster than developing the entirely new speech recognition engine. Another advantage of this approach is the potential to achieve faster compatibility with the existing technological platforms. Such advantage is often important for the business customers, since they need to follow various technical specifications in order to guarantee the consistent functioning of the enterprise. The idea behind this approach is to transfer the existing source acoustic models from a source language to the target language without using the speech corpora in that language and without full retraining of the speech recognition system [2-4].

There are no Lithuanian speech engines provided by the major speech recognition solutions providers. Same is true for the national companies engaged in the information technologies. So the cross-lingual adaptation of the foreign language based speech recognition engines could be a desirable solution in the case of such languages as the Lithuanian.

In some of our previous papers we presented several prototype systems using Lithuanian speech technologies. In [5] were investigated possibilities to use Microsoft Speech Server'2004 [6] 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. Further investigations were performed using Microsoft Speech Server'2007 [7], but it is obviously, that more comprehensive modeling is needful [8].

Proposed approach

Our choice has fallen to the Microsoft Speech Recognition platforms (Microsoft Speech Application Programming Interface (SAPI) or Microsoft Speech Server based). First of all the Microsoft speech platforms possess the well established technical standards, the requirements and consequently guarantees a high level of compatibility with other software and technical products. Another important factor influencing decision from the technical point of view is the relatively good documentation available for the Microsoft SAPI and Microsoft Speech Server and the potential to exchange experience with other users or developers. As was mentioned earlier, authors in study also used the Microsoft SAPI engine for the Lithuanian commands and proper names recognition.

But the use of the third party commercial recognizer provides at least one significant obstacle trying to adapt the technology. Commercial recognizers are usually closed software platforms and we can not to modify acoustic and language models implemented in the speech engine. It is very clear that acoustic models differ in different languages. So the single parameters that could be modified and adapted are phonetic transcriptions. The phonetic transcriptions are generated using some predefined rules if they are intended to use for the native language. But if they will be used for the recognition of words in another language rules may be not adequate for the task. So the heuristic procedure should be found. The next idea applied in our experiments is the usage of more than one phonemic transcription per word. The phoneme mapping from one language to another inevitably causes some confusions and there is no single mapping for each user and each word: a mapping which can be better for one speaker or for some phonetic contexts may not be so good for another speaker or another phonemic contexts. One of the possible methods to partially avoid such confusions is to try to use several phonemic transcriptions for a single word or command. It means that in some situation one transcription could provide the better recognition while in the other situation another transcription will be used. It is assumed intrinsically that only one transcription will be used in each situation while others will be considered in the same manner as the transcriptions for other words or commands in the vocabulary (same as other words). In principle the number of different transcriptions could be big, but we can expect that some optimal number exists. This number depends on the variety of factors: the number of words in the vocabulary, the similarity of words, the number of speakers, etc. We think that there are no exact algorithms to select the number of transcriptions and this job should be performed experimentally. The possibilities to use the multi-transcriptional approach successfully will depend essentially from the content of a vocabulary but for many practical applications such approach can be appropriate and implemented.

Speech corpora

There were two speech corpora collected for this study. The first one is called initial corpora (denoted ICT further). Here 35 non-professional speakers (17 male and 18 female) pronounced each of the ten digit names 20 times during single recording session. It should be noted that in different experiments we used different number of speakers and their data. The exact number of speakers used in each experiment is specified presenting the experiment. Different speakers data were used in a training session and in a testing session. Training session in our case was the selection of the better suited transcriptions for the recognition. Testing session in our case was the procedure of testing the accuracy of selected transcriptions.

The characteristic property of initial corpora was relatively large number of pronunciation inaccuracies. Since all speakers were inexperienced a lots of pronunciation errors were observed – hesitations, improper stressing, mistakenly pronounced phonemes, lost first and the last phonemes, etc. So the analysis of the speech corpora was done in order to remove the most persistent errors. Analysis showed that about 12.5% of all utterances contained pronunciation errors in the initial corpora.

Modified and corrected speech corpora (further denoted MCT) has been obtained in such a way: one male speaker replaced all the corrupted utterances of male speakers while single female speaker replaced all the corrupted utterances of female speakers. Both ICT and MCT corpora were used for the training purposes.

Digit number recognition accuracy

As was described earlier digit number recognition may be implemented in the variety of public services using voice user interface as one of the possible human-computer interaction modes. The efficiency and the quality of the service will be determined by the recognition accuracy. So the main task in these experiments was to try to optimize partially recognition accuracy of the names of ten Lithuanian digit names using Microsoft Speech Server recognition engine: *Microsoft English (U.S. Telephony) v7.0 Server.*

Since Microsoft speech server recognition engine is commercial recognition system so we can't modify nor the acoustic models, nor to retrain them. The single parameter that we could modify (or being more precise to select more appropriate one) is the phonetic transcription of the voice command. In these experiments we used combined selection methodic of multiple transcriptions proposed in [9]. This method showed it's superiority to the method used in [10] when applied for the recognition of the voice commands (Lithuanian proper names). Those results formed the basis for the belief that also in the case of the recognition of digit names could be obtained results that will closed to the recognition accuracy necessary for the commercial and public sector applications.

Using combined transcription generation method we obtained 70 different transcriptions. The number of the transcriptions varied from two transcriptions used for such short digit names as "two", "three" and "no" (in Lithuanian du, trys, ne) up to 12 different transcriptions used for the word "eight" (in Lithuanian "aštuoni") which also contain diphthong. It should be noted that word "yes" (in Lithuanian "taip") also has relatively big number of phonetic transcriptions. This word had 7 different transcriptions since it also contains diphthong.

For the experiments we used speech corpora of Lithuanian names and digit names. In these experiments we used recordings of 10 different speakers. Each of them pronounced each name of the digit and control word 20 times. For the comparison we used all generated transcriptions, two best transcriptions for each word and the best for each word transcription. The transcriptions were selected using the method presented in [9].

It could be seen that average recognition accuracy for ten digit names and two control words obtained was about

81%. Another observation was that recognition accuracy of different digit names varied significantly. While the recognition accuracy for such digit names as „šeši“ (six), „aštuoni (eight) , „septyni“ (seven), „vienas“ (one) was high but average recognition accuracy has been degraded by the recognition of such digit names as „keturi“ (four), „du“ (two), „devyni“ (nine). This could be interpreted in several different ways: while it could be more convenient in some applications to use the names of all digit names but at the same time it is possible to construct reliable person identification using only 7 digit names.

Table 1. Recognition accuracy of Lithuanian digits names and two words

	Recognition accuracy, %		
Command	All transcr.,	Two transcr.	Single transcr.
Nulis	41,5	40,0	67
Vienas	97,5	98,0	97,5
Du	73,5	83,0	83,0
Trys	62,0	67,5	77,5
Keturi	31,5	45,5	39,5
Penki	78,0	77,0	72,0
Šeši	99,5	100,0	100,0
Septyni	96,5	96,5	96,5
Aštuoni	98,5	97,0	97,0
Devyni	64,0	66,5	66,0
Taip	85,5	87,5	87,0
Ne	85,5	90,0	90,0
Average	76,1	79,0	81,1

The next group of experiments were performed trying to investigate the possibilities to improve recognition accuracy of some digit names that were difficult to rceognize with the transcriptions used in experiments above. One of the difficult digit names was the word „Keturi“ (four). In these experiments we used method proposed in [9] to construct phonetic transcription of the word from transcriptions of the syllables.

Initially, the syllable transcriptions were constructed out of 7 consonants (P, T, K, B, D, G, R) and all possible 16 vowel and diphthong combinations. The recognition accuracy of the corpus, containing the syllables, constructed using the consonants in the open, semi open and closed vowel context (word “Keturi” syllables „Ke, Tu, Ri”), was analyzed. The results of this experiment are further noted as “C_7V_{16}”. The recognition result analysis proved, that the syllable “Ke” was recognized as some transcription variation 98,1 %, syllable “Tu” – 94,3 %, syllable „Ri“ – 66,3 %. The most different answers out of 112 possible transcriptions were selected for the syllable “Ri”, although its recognition accuracy was the worst of all three.

A similar experiment was made using the all 24 consonants. The 384 syllable transcription variations were created in this case.

The detailed result analysis proved that the recognizer is capable of correctly recognizing the syllables, based on the open / semi open vowel: “Ke” and “Tu”. Syllable “Ke” was recognized ~ 98 % in both types of the experiments. The number of omissions was larger for the syllable “Tu”. The syllable “Ri” was the most poorly recognized syllable: ~ 66,3 % and 59,1 %. The recognition accuracy was a bit higher in the case of “C_7V_{16}” possibly due to a more limited set of the transcriptions used.

Most of the frequently recognized syllable transcriptions were quite phonetically different from their Lithuanian counterparts. The phonetically similar alternatives were selected very rarely, proving a disadvantage on the use of such linguistic knowledge based methods as [11].

The value of the combinative transcription selection method may be illustrated in the next experiment using the digit name “Keturi” (four). At the first stage the transcriptions were created, using the syllable recognition results and the best of those were selected using the iterative selection method. Trying to further improve the recognition accuracy, the most dissimilar syllable transcriptions were replaced with a more phonetically close, though less frequently recognized transcriptions. For example, the second syllable’s transcription from the word “Keturi” - D OW (pronounced *dau*) was replaced with a more phonetically similar alternative - T UH (pronounced *tu*). The best transcriptions from that modified set were iteratively selected. The overall recognition accuracy achieved (91 %) was almost 3 times higher than for the LBT based transcriptions. The results are presented in table 2.

Table 2. The best iteratively selected transcriptions of word „Keturi“

Transcription	Transcription frequency, %
G EH T UH D IH	40,5
G EH T UW D IY	20,5
G EH T UW D IH	19,0
G EH T OW G EY	3,5
K EH T UH D IY	7,5
All:	**91,0**

Similar transcription development method has been applied to get transcriptions of all ten digit names and the words “Taip” (yes) and “Ne” (no). The results are presented in the Fig. 1 together with the recognition accuracy obtained using the earlier method:

It could be seen that combinative method allowed significant increase in overall recognition accuracy. It could be seen that recognition accuracy of 9 words from 12 exceeded 90% and the recognition accuracy of 6 words exceeded even 95% (this accuracy level is important since many research on customer satisfaction emphasizes 95% recognition level as the level from which customer satisfaction is enough to use the service and to be ready to pay money). We may to assume that some further optimization is possible but since the number of the digit names recognized worse isn’t big we performed another group of experiments using restricted set of digit names.

The knowledge obtained in previous experiments allowed us to organize next group of experiments. The words “Taip“ and “Ne” were removed from the recognition vocabulary, as these voice commands are usually modeled using separate semantic rules and recognition grammars.

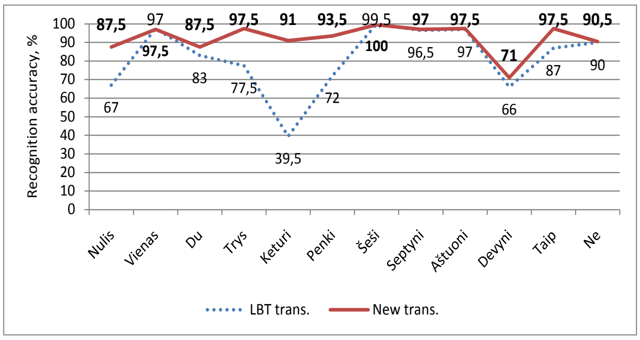

Fig. 1. Recognition accuracy of ten Lithuanian digit names and the words "taip" and "ne" using transcriptions obtained with combinative method (solid line) and most similar transcriptions method (dashed line)

Table 3 shows recognition accuracy of seven digit names: there were used 3 different digit names sets. Data sets were obtained removing three digit names from the total set of 10 digit names. The number in the table means the Lithuanian pronunciation of the denoted digit name.

Table 3. Recognition accuracy of seven digit names in 3 sets

Recognition accuracy of voice commands, %							
1	2	3	5	6	7	8	All
99.0	98.5	97.0	95.0	99.0	96.0	100	97.8
1	2	3	4	6	7	8	All
99.0	98.0	98.0	91.5	100	97	100	97.6
1	2	4	5	6	7	8	All
99	96.5	94.5	95.5	96.5	98.5	100	97.2

At the first stage of the experiment the two most poorly recognized words „devyni“ and „nulis“ were removed from the recognition vocabulary and the experiment was repeated with the remaining 8 digit names. This time the acceptable 97,8 % recognition accuracy was achieved.

Trying to achieve even higher recognition accuracy for the 7 digit set recognition application the most poorly recognized word of the 8 digit name set was removed, achieving the 97,6 % recognition accuracy (1 set of digit names).

Another experiment was made this time removing the second most poorly recognized word. The recognition accuracy achieved was a bit lower -97,2 % (2 set of digit names).

Trying to determine, how the removal of a short voice command influences the recognition accuracy, the experiment was repeated once again, this time removing one of the most accurately recognized short word "Trys". The recognition accuracy was lower than in the previous cases - 96,6 % (3 set of digit names).

The results of the experiments were verified using a new 10 speaker corpus (not used in any of the previous experiments, every speaker uttered each digit name 20 times) and 10 live speakers (uttered each digit name 20 times using the headset microphone). The results were confirmed, as the overall recognition accuracy was higher than the desired 95 % (overall 97,5%) for all three 7 digit sets (97,8 %, 97,6 % and 97,2 %)

The results of the experiments in the graphical form are presented in Fig. 2.

It should be noted that in this case all digit names used in this experiment were recognized with the accuracy exceeding 90%. It is important since it enabled us to have vocabulary comprised from the commands which were recognized with only very high accuracy level.

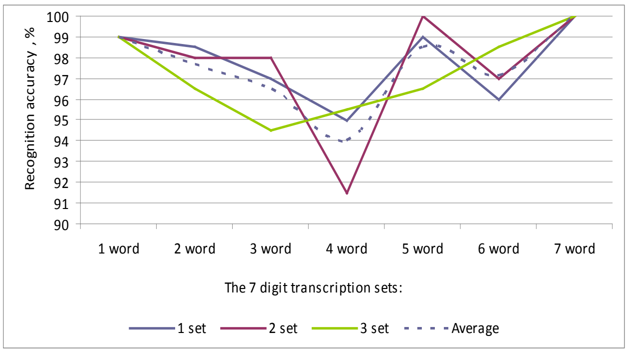

Fig. 2. Recognition accuracy of seven Lithuanian digit names

Conclusions

Successful implementation of voice user interfaces forms the possibilities to create new informative services in public sector. Such interfaces allows to access information more effectively and rapidly. In this way the realization of paradigm for the development of human-computer interaction called "anytime – anywhere" comes closer.

Many services in public sector (as well as in commercial) require the identification of user. It could be done giving users name/surname or some nickname as well as giving some code (personal number). The second solution may be more appropriate since it enables the development of services using limited vocabulary in their automatic speech recognition engines.

Combined transcription selection method allowed increase significantly recognition accuracy of ten Lithuanian digit names and two control words "Taip" and "Ne". In this case recognition accuracies for most the words in the vocabulary were above 90% and many words were recognized with accuracy level exceeding 95%.

Experiments with the recognition of seven digit names out of the ten possible showed that in this case it is possible to achieve overall recognition accuracy of more than 95%. This accuracy level is the requirement for the commercial applications as the necessary precondition to achieve customer satisfaction. The experiments with three different digit name sets showed that in all cases overall recognition accuracy exceeding 95% was achieved. In the best case overall recognition accuracy close to the 98% was achieved.

Those experiments show that implementation of Lithuanian voice command recognition into some services provided by the public sector enterprises may be feasible. The success will depend on the proper selection of the automated functions. Those services will extend access to the public sector information and will improve the relations and interaction between public enterprises and the society in general.

References

1. **Šalna B., Kamarauskas J.** Evaluation of Effectiveness of Different Methods of Speaker Recognition // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P.67–70.
2. **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.
3. **Cohen P., et al.** Towards a Universal Speech Recognizer for Multiple Languages // In: proceedings of Automatic Speech Recognition and Understanding, 1997. – P. 591–598.
4. **Fugen, C., et al.** Efficient Handling of Multilingual Language Models // In: proceedings of ASRU–2003. – ASRU, St. Thomas, 2003. – P.441–446.
5. **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.
6. **Xiaole Song.** Comparing Microsoft Speech Server 2004 and IBM WebSphere Voice Server V4.2.
7. **Dunn M.** Pro Microsoft Speech Server 2007: Developing Speech Enabled Applications with .NET. – 275 p.
8. **Maskeliunas R., Rudzionis A., Ratkevicius K., Rudzionis V.** Investigation of Foreign Languages Models for Lithuanian Speech Recognition // Electronics and Electrical Engineering. – Kaunas: Technologija, 2009. – No. 3(91). – P.15–21.
9. **Maskeliunas R., Rudžionis A., Rudzionis V.** Analysis of the Possibilities to Adapt the Foreign Language Speech Recognition Engines for the Lithuanian Spoken Commands Recognition. // COST Action 2102 International Conference on Cross–Modal Analysis of Speech, Gestures, Gaze and Facial Expressions, Lecture Notes in Artificial Intelligence, LNAI5641. – Springer, 2009. – P. 409–422.
10. **Kasparaitis P.** Lithuanian Speech Recognition Using the English recognizer // Informatica, 2008. – No. 19(4). – P. 505–516.
11. **Zgank A., et al.** The COST278 MASPER initiative – croslingual speech recognition with large telephone databases // Proc. of 4th International Conference on Language Resources and Evaluation (LREC'04), 2004. – P. 2107–2110.

Received 2010.02.25

R. Maskeliūnas, A. Rudžionis, K. Ratkevičius, V. Rudžionis. Modeling of Call Services for Public Sector // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 4(100). – P. 81–86.

Paper deals with application of English speech recognizer for multi-speaker Lithuanian digit names recognition. Many services in public sector require the identification of user or recognition of Lithuanian digit names with 95% or more accuracy. Combined multiple transcription selection method allowed to achieve multi-speaker recognition accuracy of ten Lithuanian digit names and two auxiliary words above 90%. It is possible to construct reliable user identification using only 7 digit names, so the experiments with three different sets of 7 Lithuanian digit names were performed and in all cases overall recognition accuracy exceeded 95%. Ill. 2, bibl. 11, tabl. 3 (in English; abstracts in English, Russian and Lithuanian).

Р. Маскелюнас, А. Руджёнис, К. Раткявичюс, В. Руджёнис. Моделлирование телефонных услуг для государственного сектора // Электроника и электротехника. – Каунас: Технология, 2010. – № 4(100). – С. 81–86.

В статье анализируется возможности использования английского распознавателя для многодикторного распознавания литовских названий цифр. Большинство телефонных услуг для публичного сектора требует идентификации пользователя, а для этого необходимо распознавать литовские названия цифр с точностью не менее 95%. Комбинированный метод определения множественных транскрипций для литовских названий цифр позволил получить точность многодикторного распознавания литовских названий цифр и двух слов более 90%. Точность многодикторного распознавания семи литовских названий цифр, достаточных для идентификации пользователя, превышает 95%. Ил. 2, библ. 11, табл. 3 (на английском языке; рефераты на английском, русском и литовском яз.).

R. Maskeliūnas, A. Rudžionis, K. Ratkevičius, V. Rudžionis. Telefono paslaugų valstybiniam sektoriui modeliavimas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2010. – Nr. 4(100). – P. 81–86.

Nagrinėjamos galimybės anglų kalbos atpažinimo įrenginį panaudoti daugiadiktoriam lietuviškų skaičių pavadinimų atpažinimui. Dauguma balso sąsaja pagrįstų telefono paslaugų remiasi vartotojo identifikavimu, o tam reikia atpažinti lietuviškų skaičių pavadinimus bent 95 % tikslumu. Aprašomi kombinuoto skaičių pavadinimų daugybinių transkripcijų paruošimo metodo testavimo rezultatai. Dešimties lietuviškų skaičių pavadinimų ir dviejų pagalbinių žodžių daugiadiktorio atpažinimo su anglų kalbos atpažintuvu tikslumas viršija 90 %, o septynių skaičių rinkinio, pakankamo vartotojui identifikuoti, atpažinimo tikslumas yra didesnis kaip 95 %. Il. 2, bibl. 11, lent. 3 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).

DOI: 10.5755/j02.eie.9882