SPEECH RECOGNITION FOR EAST SLAVIC LANGUAGES: THE CASE OF RUSSIAN

Alexey Karpov[1], Irina Kipyatkova[2] and Andrey Ronzhin[1]

[1]Saint-Petersburg State University, Department of Phonetics, Russia
[2]SPIIRAS Institute, Speech and Multimodal Interfaces Laboratory, Russia
{karpov, kipyatkova, ronzhin}@iias.spb.su

ABSTRACT

In this paper, we present a survey of state-of-the-art systems for automatic processing of recognition of under-resourced languages of the Eastern Europe, in particular, East Slavic languages (Ukrainian, Belarusian and Russian), which share some common prominent features including Cyrillic alphabet, phonetic classes, morphological structure of word-forms and relatively free grammar. A large vocabulary Russian speech recognizer, developed by SPIIRAS, is described in the paper and especial attention is paid to grapheme-to-phoneme conversion for automatic creation of a pronunciation vocabulary and acoustic modeling at the system training stage. Speech recognition results for a very large vocabulary above 200K word-forms are reported.

Index Terms— Slavic languages, automatic speech recognition (ASR), Russian language, pronunciation vocabulary, grapheme-to-phoneme conversion

1. INTRODUCTION

In the Eastern Europe, over 400 million people speak using Slavic languages (see Table 1). These languages belong to the Balto-Slavic subgroup of the Indo-European family of languages. They are divided into three branches: West, East and South Slavic languages. In the domain of automatic speech analysis and synthesis, the most elaborated systems exist for West Slavic languages: Czech (ASR researches in universities of Plzen [1], Brno [2], Liberec [3]), Slovak [4,5] and Polish [6,7]. South Slavic branch of languages consists of Serbo-Croatian, it includes very similar spoken languages: Serbian, Croatian, Bosnian and Montenegrin [8, 9], plus Slovene [10], Bulgarian [11] and Macedonian [12], correspondingly. There exist also some modern automatic systems intended for multi-lingual speech recognition (e.g., [13]), including a few Slavic languages.

The present paper is mainly aimed to the East Slavic languages: Russian, Ukrainian and Belarusian. It must be noted that Slavic languages does not include other major languages of Eastern Europe: Hungarian [14], Romanian [15] and Moldavian, which can be considered as a dialect of Moldavian. According to the etymology, closest to the Slavic languages are Baltic ones: Lithuanian [16] and Latvian (now there is only a speech corpus [17]), but not Estonian [18], which is the Uralic language.

Table 1. National spoken languages of the Eastern Europe

Language and family	Approximate speakers, Million
Indo-European language family	
West Slavic languages	
Czech	12
Slovak	5
Polish	40
South Slavic languages	
Serbo-Croatian: Serbian, Croatian, Bosnian, Montenegrin	20
Slovene	2.5
Bulgarian	11
Macedonian	2
East Slavic languages	
Russian	165 (up to 300)
Ukrainian	47
Belarusian	8
Balto-Slavic languages	
Lithuanian	4
Latvian	2
Romanic languages	
Romanian	20
Moldavian (dialect of Romanian)	2.5
Paleo-Balkan languages	
Albanian	7.5
Uralic language family	
Hungarian	14
Estonian	1.1

The major similarity between East Slavic languages is Cyrillic alphabet, which is shared by Russian, Ukrainian, Belarusian, some other Slavic languages (Serbian, Bulgarian and Macedonian) and some non-Slavic languages of the former Soviet Union (Kazakh, Kyrgyz, Tajik, etc.). Cyrillic alphabets for Russian and Ukrainian are different, but both

have 33 letters, and there are 32 letters in Belarusian. The Ukrainian alphabet does not have graphemes Э, Ъ, Ы, Ё, but has some other letters such as Є, І, Ї, Ґ plus apostrophe (‘), there are some corresponding changes in phonetic alphabet as well (e.g., letter Г corresponding to Russian consonant /g/ is pronounced as velar /ɣ/, there is such a phoneme in South Russian dialects only). Belarusian is a bit closer to Russian, it uses grapheme І instead of И, has one more letter Ў (short У, like English /w/) plus apostrophe instead of Ъ and letter Щ does not exist. It is more or less easy to read and understand Belarusian texts for the Russians, though less than one third of the words are orthographically similar in both languages (including letter changes), but much more word pairs have phonetic similarity. Morphological and stressing rules for word-formation, orthographic-to-phonetic transcription, grammar for all three languages have very similar principles as well, that makes simple texts readable and slow speech understandable for non-speakers. There are some common features for all East Slavic languages. Stress is moveable and can occur on any syllable, no strict rules to identify stressed syllable in a word-form (people keep it in mind and use analogies, proper stressing is the major problem for learners of East Slavic languages). These are synthetic and highly inflected languages with lots of roots and affixes including prefixes, interfixes, suffixes, postfixes and endings, moreover, these sets of affixes are overlapping, in particular one-letter grammatical morphemes У, А, О. Nouns and adjectives are divided into feminine, masculine, and neuter gender (like in German) and have some cases (7 or 6) and number. Verbs are marked for first, second, and third persons, singular and plural, perfective and imperfective. Some verbs can generate above two hundred grammatically correct word-forms. In syntax, main verbs agree in person and number with their subjects. One of the Slavic language features is a high redundancy: the same gender and number information can be repeated some times in one sentence. Word order is grammatically free with no fixed order for words indicating subject, object, possessor, etc. However, the inflectional system takes care of keeping the syntax clear, semantic and pragmatic information is crucial for determining word order.

At present, there are no speech and language databases for Ukrainian and Belarusian in the ELRA catalogue or in other multilingual corpora like GlobalPhone, SpeechDat, Speecon, Speech Ocean, etc. Research on Ukrainian speech recognition are carried out only in Ukraine, for instance, by ICITS Center in Kiev [19] headed by T. Vintsyk, who was a pioneer in dynamic programming methods in the USSR, as well as by IAI Institute in Donetsk [20]. A corpus of continuous and spontaneous Ukrainian speech has been collected and processed for the tasks of ASR and TTS [21]. Also there exists a text corpus available on-line (www.mova.info/corpus.aspx?l1=209) that was created by National University of Kyiv. There are also some studies on automatic processing of the so called Surzhyk (a mixture of written and spoken Ukrainian and Russian) [22], which is widely used in regions of Ukraine near borders between both countries and in Crimea, and it is considered as a dialect.

Text-to-speech systems for Belarusian and some components of Belarusian ASR including some language resources and analyzers have been developed by UIIP Institute of NAS [23], as well as by BSUIR [24] and BSU [25] Universities in Minsk. However at present time, no model for large vocabulary continuous Belarusian speech recognition as well as no core component of an ASR system, i.e. a segmented multi-speaker Belarusian speech database acceptable for acoustic models training.

The situation with the most other languages of the former Soviet Union is even worse, there were no any research on automatic processing or recognition for Armenian, Tajik or Uzbek languages with millions of native speakers and only initial steps are known towards Kazakh (over 10M speakers, [26] reports only preliminary results), Azerbaijan, Georgian or Moldavian ASR.

Russian is not only language of multi-national Russian Federation, there are up to 150 other languages spoken by different peoples [27]. These spoken and written languages have from five millions to hundreds of native speakers, the major of them are (above half million speakers): Tatar, Bashkir, Chechen, Chuvash, Avar, Kabardian, Dargin, Ossetian, Yakut, Udmurt. Many of these languages are national ones in the corresponding Republics of the Russian Federation (for instance, the Republic of Dagestan officially has 14 national languages). Some speech corpora for several of these languages have been collected recently and studied with forensic, safety or other purposes [27].

The Russian language also has many dialects and accents because of the multi-national culture of Russia. There exist essential phonetic and lexical differences in Russian spoken by Caucasian, Tatar or Buryat people caused by influence of other national languages. Major inner dialects of the standard Russian are North, Central and South Russian in the European part. The North Russian dialect (it is used in the cities of Arkhangelsk, Murmansk, Vologda, Kostroma, nearby Ladoga and Onega, etc.) is characterized, for example, by clear pronunciation of unstressed syllables with vowel /o/ (without typical reduction to /ɐ /), the so called “okanye”, a lot of words from Old-Russian are used as well. On the contrary, the South Russian dialect (it is spoken in Belgorod, Kursk, Ryazan, Smolensk, Tambov, Tula, etc.) has more distinctions including so called “akanye” (no difference between unstressed vowels /o/ and /a/) and “yakanye” (unstressed /o/, /e/, /a/ after a palatalized consonant are pronounced as /æ/ instead of /i/ as usually) [28], velar fricative /ɣ/ is used instead of the standard /g/ (like in Belarusian and Ukrainian), semivowel /w~u̯/ is often used in the place of /v/ or final /l/, etc. Central Russian (including the Moscow region) is a mixture of the North and the South dialects. It is usually considered the Standard

Russian originates from this group. A survey of state-of-the-art ASR methods and systems for Russian can be found in our recent publications [29, 30]. The present paper deals with the lexical level of Russian speech recognition and presents details of the module for automatic orthographic-to-phonemic transcription in order to prepare a pronunciation vocabulary for ASR. After a few modifications, this module can be applied to other East Slavic languages too.

2. LEXICAL LEVEL OF ASR

One of the important challenges at development of ASR systems for spoken Russian is grapheme-to-phoneme conversion or orthographic-to-phonemic transcription of a recognition lexicon that is not a trivial task in comparison with German or agglutinative Turkish and Finnish. There exist several troubles: grapheme-to-phoneme mapping is not one-to-one, stress position(s) in word-forms is floating, substitution of grapheme Ё (always stressed) with E in the most of printed and electronic text data, phoneme reductions and assimilations in continuous and spontaneous speech, many homographs, etc.

There are two main approaches to grapheme-to-phoneme conversion and pronunciation modeling [31]: knowledge-based and data-driven methods. And for both methods direct and indirect modeling can be distinguished. In the knowledge-based methods, transcriptions are generated using existing phonetic and linguistic knowledge (rules) formulated by the experimental phonetics at analysis of speech data, acoustic and articulation characteristics of phonemes. In the data-driven methods, existing speech data of many speakers are analyzed in order to produce pronunciation vocabulary based on real data. The data-driven methods are more adequate for a small and medium-sized ASR lexicon [32], whereas knowledge-based and hybrid methods are more appropriate for a (very) large pronunciation vocabulary.

The aim of grapheme-to-phoneme conversion consists in an automatic generation of pronunciation vocabulary from a lexicon represented in the orthographic form: $W = \{w_1,...,w_N\}$, where N is a number of words in the ASR lexicon. The pronunciation vocabulary can be represented as a set $V = \{v_1,...,v_i,...,v_N\}$, each element of which is a pair $v_i = <O_i, \{T_i\}>$, where O_i is an orthographic representation of a word and T_i is a set a phonemic transcriptions of this word. At that, the orthographic representation of i-th word is a vector of graphemes $O_i = (o_i^1,...,o_i^K)$, and $o_i \in G$, where $G = \{g_0, g_1,...,g_{33}\}$ is the set of Cyrillic letters and the sign of stress g_0 ("!"). The set of i-th word phonemic transcriptions $T_i = \{t_i^0, t_i^1,..t_i^L\}$ contains one canonical transcription and can have a few alternative transcriptions, where $t_i^j = (s_{i_1}^j,...,s_{i_J}^j)$, $s_i^j \in P$, and $P = \{p_1,...,p_{46}\}$ is a set of phonemes in the language. So, essence of the grapheme-to-phoneme conversion consists in projection of the orthographic lexicon into the pronunciation vocabulary ($W \xrightarrow{R} V$) by means of the set of phonemic transcription rules $R = \{r_1,...,r_X\}$.

According to the SAMPA phonetic alphabet, there are 42 phonemes in the Russian language (for 33 Cyrillic letters): 6 vowels and 36 consonants including plain and palatalized versions of some consonants. Russian consonants are: voiced-unvoiced pairs /p/ (Cyrillic grapheme П) and /b/ (Б), /t/ (Т) and /d/ (Д), /k/ (К) and /g/ (Г), /f/ (Ф) and /v/ (В), /s/ (С) and /z/ (З) (they have palatalized versions as well), /S/ (Ш) and /Z/ (Ж); sonorants /l/ (Л), /r/ (Р), /m/ (М), /n/ (Н) (these consonants are not paired, but have palatalized versions) and /j/ (Й), plus velar /x/ (and a soft version /x'/, grapheme Х), /ts/ (Ц), /tS'/ (Ч), /S':/ (Щ). However, according to the International Phonetic Alphabet (IPA), there are 17 vowels in Russian with different levels of reduction between stressed and unstressed vowels up to complete disappearance. Recent experiments showed [33], that distinction between models for stressed and unstressed vowels allows decreasing WER at ASR. Thus, six stressed (/a!/, /e!/, /o!/, /u!/, /i!/ and /1!/ in SAMPA format) and four unstressed vowels are used (/o!/ and /e!/ may have only stressed versions in the standard Russian with a few exceptions).

At grapheme-to-phoneme conversion the following positional changes of sounds are made: (1) changes of vowels in pre-stressed syllables, which are presented in Table 2; (2) changes of vowels in post-stressed syllables, which are shown in Table 3; (3) positional changes of consonants can happen in the following cases [34]:

- At the end of a word or before an unvoiced fricative consonant, voiced fricatives are devoiced.
- Before voiced fricatives (excluding /v/ and /v'/) unvoiced fricatives become voiced.
- Before palatalized dentals /t'/ and /d'/ phonemes /s/, /z/ become palatalized, as well as before /s'/ and /z'/, consonants /s/, /z/ are disappeared (merged into one phoneme).
- Before palatalized dentals /t'/, /d'/, /s'/ /z'/ or /tS'/, /S':/, hard consonant /n/ becomes palatalized.
- Before / tS'/, consonant /t/ (both for graphemes Т and Д) is disappeared.
- Before /S/ or /Z/, dental consonants /s/, /z/ are disappeared (merged).
- Two identical consonants following each other are merged into one.
- Some frequent combinations of consonants are changed: /l n ts/ → /n ts/, /s t n/ → /s n/, /z d n/ → /z n/, /v s t v/ → /s t v/, /f s t v/ → /s t v/, /n t g/ → /n g/, /n d g/ → /n g/, /d s t/ → /ts t/, /t s/ → /ts/, /h g/ → /g/, /s S':/→ /S':/, etc.

The algorithm for automatic grapheme-to-phoneme conversion of word-forms operates in two cycles, consisting of the following steps:

Table 2. Positional vowel changes in pre-stressed syllables

Original vowel (for grapheme)	Resulting phoneme depending on position				
	At the beginning of a word	After velar consonants	After paired hard consonants	After paired palatalized consonants	After fricatives /S/, /Z/, /ts/
/e/ (Э,Е)	/1/	/i/	/1/	/i/	/1/
/i/ (И)	/i/	/i/	-	/i/	-
/1/ (Ы)	-	-	/1/	-	/1/
/a/ (А,Я)	/a/	/a/	/a/	/i/	/a/
/o/ (О,Ё)	/a/	/a/	/a/	/i/	/a/
/u/ (У,Ю)	/u/	/u/	/u/	/u/	/u/

1) Stress positions are identified using a morphological database.
2) Hard consonants before graphemes И, Е, Ё, Ю, Я become palatalized (if possible) and these graphemes are converted into phonemes /i/, /e/, /jo!/, /ju/, /ja/ in the case if they are located in the beginning of the word or after any vowel, otherwise they are transformed into /i/, /e/, /o!/, /u/, /a/, correspondingly.
3) A consonant before grapheme Ь gets palatalized and the grapheme is deleted (it has no corresponding phoneme).
4) Transcription rules for positional changes of consonants (presented above) are applied.
5) Transcription rules for positional changes of vowels in pre-stressed and post-stressed syllables (presented above) are applied.
6) Steps (4)-(6) are repeated one again, some changes may result in some other changes in preceding phonemes.
7) Grapheme Ъ is deleted (it has no corresponding phoneme), this letter just shows that the preceding consonant is hard.

Table 3. Positional vowel changes in post-stressed syllables

Original vowel (for grapheme)	Resulting phoneme depending on position		
	After velar consonants	After paired hard consonants and /S/, /Z/, /ts/	After paired palatalized consonants and /tS'/, /S':/
/e/ (Э,Е)	/i/	/1/	/i/
/i/ (И)	/i/	/1/	/i/
/1/ (Ы)	-	/1/	-
/a/ (А,Я)	/a/	/a/	/a/
/o/ (О,Ё)	/a/	/a/	/a/
/u/ (У,Ю)	/u/	/u/	/u/

For the grapheme-to-phoneme conversion, we employ an extended morphological database of more than 2.3M word-forms with a mark ("!") for stressed vowels/syllables. This database is a fusion of two different morphological databases: AOT (www.aot.ru) and Starling (starling.rinet.ru/morpho.php). The former one is larger and has above 2M items, but the latter one contains information about the secondary stress for many compound words as well as words with grapheme Ё, which is always stressed at pronunciation, but it is usually replaced to E in official texts that results in loosing information on stress.

Additionally, some alternative phonemic transcriptions are generated for word-forms in order to model the effects of phonemes' reduction and assimilation in spontaneous speech using a set of cross- and intra-word phonetic rules:

- Phoneme /j/ located at the word end is completely reduced if it follows an unstressed vowel.
- An unstressed vowel is reduced up to complete disappearance, in the case of its position between identical consonants.
- Words-homographs, like ВСЕ ("everybody" in Eng.) and ВСЁ(Е) ("everything" in Eng.), get alternative transcriptions, if there are orthographic words differentiating in graphemes Ё and E in two morphological databases, etc.

The quantitative evaluation showed that 98% canonical phonemic transcriptions, made automatically for some evaluation texts, were correct (comparison with transcriptions made by an expert). Some errors were caused by incorrect stress position for multi-syllable word-forms absent in the morphological database or for homographs.

With minor changes in the phonemic alphabet and the knowledge-based transcription rules, it is possible to adapt the conversion module for Belarusian and Ukrainian languages too.

3. EXPERIMENTAL RESULTS

We have implemented the SIRIUS automatic system [30] for very large vocabulary ASR of spoken Russian. Continuous density Hidden Markov Models (HMMs) with 3 emitting states and mixtures of 16 Gaussians per state are applied to model context-dependent phones. At feature extraction, 12-dimentional Mel-Frequency Cepstral Coefficients (MFCC) plus segment energy with their deltas and double deltas are calculated from the 26-channel filter bank analysis of 20 ms frames with 10 ms overlapping.

We have proposed an integral language model (LM) that takes advantages of statistical and syntactic text analysis, [30] gives details of text processing for LM creation. We have collected and processed a text corpus consisting of archives of four on-line newspapers for the last five years: "Новая газета" (www.ng.ru), "СМИ" (www.smi.ru), "Лента.ру" (www.lenta.ru), "Газета.ру" (www.gazeta.ru). This text corpus was processed in parallel by two analyzers identifying N-grams and syntactic dependencies in sentences and then the results of both analyzers are merged in the

integral stochastic model that takes into account frequencies of the detected word pairs. These text analyzers complement each other: syntactic parser (AOT VisualSynan) is used to find long-distance dependencies between words, i.e. some potential bigrams absent in the training data, whereas relations between adjacent words are covered by the statistical analyzer (CMU SLM Toolkit).

The volume of the text corpus after its normalization and deletion of doubling or short sentences is over 110M words, and it has about 940K unique word-forms. As the result of the statistical analysis, we have obtained above 6M unique bigrams (cutoff = 1) and the syntactic analysis has allowed us to extend the integral LM up to 6.9M bigrams. This language model includes almost 210K unique words, which form the pronunciation vocabulary created with the help of the grapheme-to-phoneme convertor presented above.

For ASR system training and evaluation, we used a corpus of Russian speech created during 2008-2009 in the framework of the Euronounce project. It includes approximately 22 hours of continuous speech of 52 native male and female speakers from the St. Petersburg region, while pronouncing a special set of 330 phonetically rich sentences and texts, plus 30 minutes of spoken Russian for ASR evaluation. Speech data were recorded with 44.1 KHz sampling rate (for ASR downsampled to 16 KHz), 16 bits per sample, SNR was 35dB at least, by a stereo pair of Oktava MK-012 stationary microphones (close talking ≈20 cm and far-field ≈100 cm microphone setup) connected to PC via Presonus Firepod sound board.

Table 4 summarizes model parameters (LM informational entropy and perplexity, amount of out-of-vocabulary OOV words, bigram hit) and recognition results for the given corpora in terms of the word error rate (WER) and grapheme (or letter that is the same) error rate (LER). The pronunciation vocabulary contains almost 210.1K word-forms and the integral syntactic-statistical bigram LM is used. According to our previous results [30], this LM outperforms pure statistical bigram and trigram LMs created with the same training text data.

Table 4. Summary of the Russian speech recognition results

Entropy, bit/word	Perplex.	OOV words, %	N-gram hit, %	WER, %	LER, %
9.6	772	0.75	84.1	36.4	12.6

Relatively high speech error rates can be explained by the inflective nature of the given Slavic language, where each stem corresponds with tens/hundreds of endings, which are usually pronounced in continuous speech not so clearly as the beginning parts of the words and often different orthographic word-forms have identical phonemic representations. We have also applied inflectional word error rate (IWER) measure [35, 30], which assigns a weight k_{inf_1} to all "hard" substitutions S_1, where lemma of the word-form is wrong, and a weight k_{inf_2} to all "weak" substitutions S_2, when lemma of the recognized word-form is right, but ending of the word-form is wrong:

$$IWER = \frac{I + D + k_{\text{inf\_1}} \cdot S_1 + k_{\text{inf\_2}} \cdot S_2}{N} \times 100\,\%$$

In our experiments, the IWER measure with $k_{inf\_1}=1.0$ and $k_{inf\_2}=0.5$ was equal 31.8%, so in total about 10% of the errors were caused by misrecognized word endings.

4. CONCLUSION

The paper proposed a review of ASR systems and lexical/speech resources for three East Slavic languages (Russian, Ukrainian and Belarusian) and other languages of Eastern Europe, a particular attention was paid to very large vocabulary speech recognition for standard Russian with a focus on automatic production of the pronunciation vocabulary and knowledge-based grapheme-to-phoneme conversion. After a few modifications, it can be adapted to other East Slavic languages too. Experimental setup and ASR results for a very verge lexicon of 210K word-forms and syntactic-statistic language model were reported.

5. ACKNOWLEDGEMENTS

This research is supported by the Saint-Petersburg State University (project No. 31.37.103.2011), by the Ministry of Education and Science of Russia (contract No.11.519.11.4020), by the Russian Foundation for Basic Research (project No. 12-08-01265), by the Russian Humanitarian Scientific Foundation (project No. 12-04-12062), as well as by the grant of the President of Russia (project No. MK-1880.2012.8).

6. REFERENCES

[1] P. Ircing, J. Psutka, J. Psutka, Using Morphological Information for Robust Language Modeling in Czech ASR System. IEEE Transactions on Audio Speech and Language Processing, Vol. 17, 2009, p. 840-847.

[2] I. Oparin, O. Glembek, L. Burget, J. Černocký, Morphological random forests for language modeling of inflectional languages, In Proc. IEEE Workshop on Spoken Language Technology SLT'08, Goa, India, 2008.

[3] J. Nouza, J. Silovský, Adapting Lexical and Language Models for Transcription of Highly Spontaneous Spoken Czech. In Proc. "Text, Speech and Dialog" International Conference TSD'10, Brno, Czech Republic, 2010, pp. 377-384.

[4] M. Lojka, J. Juhar, Finite-state transducers and speech recognition in Slovak language. In Proc. Signal Processing Algorithms, Architectures, Arrangements, and Applications Conference SPA'09, Poznan, Poland, 2009, pp. 149-153.

[5] S. Darjaa, M. Cernak, M. Trnka, M. Rusko, R. Sabo, Effective Triphone Mapping for Acoustic Modeling in Speech Recognition. In Proc. Interspeech'11, Florence, Italy, 2001, pp. 1717-1720.

[6] M. Ziolko, J. Galka, B. Ziolko, T. Jadczyk, D. Skurzok, M. Masior, Automatic Speech Recognition System Dedicated for Polish. In Proc. Interspeech'11, Florence, Italy, 2011, pp. 3315-3316.

[7] J. Loof, C. Gollan, H. Ney, Cross-language Bootstrapping for Unsupervised Acoustic Model Training: Rapid Development of a Polish Speech Recognition System. In Proc. Interspeech'09, Brighton, UK, 2009, pp. 88-91.

[8] P. Scheytt, P. Geutner, A. Waibel, Serbo-Croatian LVCSR on the dictation and broadcast news domain. In Proc. ICASSP'98, Seattle, USA, 1998, pp. 897-900.

[9] V. Delić, M. Sečujski, N. Jakovljević, M. Janev,.R. Obradović, D. Pekar, Speech Technologies for Serbian and Kindred South Slavic Languages. In: Advances in Speech Recognition. N. Shabtai (Ed.), 2010, pp. 141-164.

[10] T. Rotovnik, M. Maucec, Z. Kacix, Large vocabulary continuous speech recognition of an inflected language using stems and endings. Speech Communication, 49(6), 2007, pp. 437-452.

[11] P. Mitankin, S. Mihov, T. Tinchev. Large vocabulary continuous speech recognition for Bulgarian. In Proc. International Conference RANLP'09, Borovets, Bulgaria, 2009, pp. 246–250.

[12] V. Delić, M. Sečujski, D. Pekar, N. Jakovljević, D. Mišković. A Review of AlfaNum Speech Technologies for Serbian, Croatian and Macedonian. In Proc. Int. Language Technologies Conference IS-LTC'06, Ljubljana, Slovenia, 2006, pp. 257-260.

[13] N.T. Vu, T. Schlippe, F. Kraus, T. Schultz, Rapid Bootstrapping of five Eastern European Languages using the Rapid Language Adaptation Toolkit. In Proc. Interspeech'10, Makuhari, Japan, 2010, pp. 865-868.

[14] B. Tarjan, P. Mihajlik. On Morph-Based LVCSR Improvements. In Proc. 2nd International Workshop on Spoken Languages Technologies for Under-resourced Languages SLTU'10, Malaysia, 2010, pp. 10-16.

[15] H. Cucu, L. Besacier, C. Burileanu, A. Buzo. Enhancing Automatic Speech Recognition for Romanian by Using Machine Translated and Web-based Text Corpora. In Proc. 14th International Conference on Speech and Computer SPECOM'11, Kazan, Russia, 2011, pp. 81-88.

[16] D. Silingas, S. Laurinciukaite, L. Telksnys. A Technique for Choosing Efficient Acoustic Modeling Units for Lithuanian Continuous Speech Recognition. In Proc. SPECOM'06 International Conference, St. Petersburg, Russia, 2006, pp. 61-66.

[17] I. Auziņa. Towards Spoken Latvian Corpus: Current Situation, Methodology and Development. In Proc. 4th International Conference Baltic HLT'10, Riga, Latvia, 2010, pp. 39-44.

[18] T. Alumäe, E. Meister. Estonian Large Vocabulary Speech Recognition System for Radiology. In Proc. Baltic HLT'10, Riga, Latvia, 2010, pp. 33-38.

[19] T. Lyudovyk, V. Robeyko, V. Pylypenko. Automatic recognition of spontaneous Ukrainian speech based on the Ukrainian broadcast speech corpus. In Proc. Dialog'11 Conference, Moscow, Russia, 2011, pp. 540-551. (in Rus.)

[20] G. Dorohina, A. Ronzhin, I. Kagirov, D. Azarenko, T. Ermolenko. Development of Language Processing Means for a System of Bilingual Speech Decoding with an Extra-Large Vocabulary. Artificial Intelligence Journal, Donetsk, Ukraine, Vol. 4, 2007, pp. 352-356. (in Rus.)

[21] V. Pylypenko, V. Robeiko, M. Sazhok, N. Vasylieva, O. Radoutsky. Ukrainian Broadcast Speech Corpus Development. In Proc. SPECOM'11, Kazan, Russia, 2011, pp. 435-440.

[22] T. Lyudovyk, S. Brozinski, M. Noner, V. Robeiko, M. Sazhok. Speech synthesis applied to SMS reading. In Proc. SPECOM'09, St. Petersburg, Russia, 2009, pp. 300-305.

[23] R. Hoffmann, E. Shpilewsky, B. Lobanov, A. Ronzhin. Development of a multi-voice and multi-language text-to-speech (TTS) and speech-to-text (STT) conversion system (languages: Belorussian, Polish, Russian), In Proc. SPECOM'04, St. Petersburg, Russia, 2004, pp. 657-661.

[24] A. Ivanov, A. Petrovsky. Analysis of the IHC Adaptation for the Anthropomorphic Speech Processing Systems. EURASIP J. on Advances in Signal Processing, No. 9, 2005, pp. 1323-1333.

[25] E. Bovbel, D. Tsishkou. Belarussian Speech Recognition Using Genetic Algorithms. In Proc. 3rd International Workshop on Text, Speech and Dialogue TSD'2000, Brno, Czech Republic, 2000, pp. 301-306.

[26] M. Karabalaeva, A. Sharipbaev. Algorithms for phone-based recognition of Kazakh speech in the amplitude-time space. In Proc. 2nd All-Russian Conference "Knowledge-Ontology-Theories", Novosibirsk, Russia, 2009 (in Rus.)

[27] R. Potapova. Multilingual Spoken Language Databases in Russia. In Proc. SPECOM'11, Kazan, Russia, 2011, pp. 13-17.

[28] J. Smirnova. Compound Systems of Pretonic Vocalism after Palatalized Consonants in Russian Dialects: A Synchronic and Diachronic Analysis. In Proc. 17th International Congress of Phonetic Sciences ICPhS'11, Hong Kong, 2011, pp. 1870-1873.

[29] D. Vazhenina, I. Kipyatkova, K. Markov, A. Karpov. State-of-the-art Speech Recognition Technologies for Russian Language. In Proc. Joint International Conference on Human-Centered Computer Environments HCCE'12, Aizu, Japan, 2012, pp. 59-63.

[30] A. Karpov, I. Kipyatkova, and A. Ronzhin. Very Large Vocabulary ASR for Spoken Russian with Syntactic and Morphemic Analysis. In Proc. Interspeech'11, Florence, Italy, 2011, pp. 3161-3164.

[31] M. Saraclar. Pronunciation Modeling for Conversational Speech Recognition, PhD Thesis. Baltimore, USA, 2000, 143 p.

[32] I. Kipyatkova, A. Karpov. Creation of Multiple Word Transcriptions for Conversational Russian Speech Recognition. In Proc. SPECOM'09, St. Petersburg, Russia, 2009, pp. 71-75.

[33] D. Vazhenina, K. Markov. Phoneme Set Selection for Russian Speech Recognition. In Proc. 7th Int. Conf. on NLP and Knowledge Engineering NLP-KE'11, Japan, 2011, pp. 475-478.

[34] N. Shvedova, et al. Russian Grammar. Vol. 1, Moscow: Nauka, 1980, 783 p. (in Rus.)

[35] K. Bhanuprasad, M. Svenson. Errgrams - A Way to Improving ASR for Highly Inflective Dravidian Languages. In Proc. 3rd International Joint Conf. on Natural Language Processing IJCNLP'08, India, 2008, pp. 805-810.