INFORMATICA, 2018, Vol. 29, No. 3, 487–498

DOI: http://dx.doi.org/10.15388/Informatica.2018.177

Lithuanian Speech Corpus Liepa for Development of Human-Computer Interfaces Working in Voice Recognition and Synthesis Mode

Sigita LAURINČIUKAITĖ[1]*, Laimutis TELKSNYS[1],
Pijus KASPARAITIS[2], Regina KLIUKIENĖ[3], Vilma PAUKŠTYTĖ[3]

[1] *Institute of Data Science and Digital Technologies, Vilnius University*
Akademijos 4, Vilnius, Lithuania
[2] *Faculty of Mathematics and Informatics, Vilnius University, Didlaukio 47, Vilnius, Lithuania*
[3] *Faculty of Philology, Vilnius University, Universiteto 5, Vilnius, Lithuania*
e-mail: sigita.lau@gmail.com, laimutis.telksnys@mii.vu.lt, pkasparaitis@yahoo.com, vilma.paukstyte@gmail.com

Received: May 2017; accepted: March 2018

Abstract. The problem of speech corpus for design of human-computer interfaces working in voice recognition and synthesis mode is investigated. Specific requirements of speech corpus for speech recognizers and synthesizers were accented. It has been discussed that in order to develop above mentioned speech corpus, it has to consist of two parts. One part of speech corpus should be presented for the needs of Lithuanian text-to-speech synthesizers, another part of speech corpus – for the needs of Lithuanian speech recognition engines. It has been determined that the part of speech corpus designed for speech recognition engines has to ensure the availability to present language specificity by the use of different sets of phonemes. According to the research results, the speech corpus Liepa, which consists of two parts, was developed. This speech corpus opens possibilities for cost-effective and flexible development of human-computer interfaces working in voice recognition and synthesis mode.

Key words: speech corpus, speech annotation, speech synthesis, speech recognition, human-computer interfaces.

1. Introduction

A standard control system receives a command by keyboard, mouse or equipment-generated electrical signals and gives a response in the format of image, sound, and movement of equipment or electrical signal. Meanwhile, a system controlled by a voice receives a task and gives a response in the form of speech signal or voice. Users that give commands by voice are of different age, in diverse emotional and physical conditions or a particular acoustical environment. Prerequisite for such system is the ability to capture, distinguish, analyse, and make decisions about characteristics of speech signals and to pronounce the

*Corresponding author.

answer by voice correctly. Speech corpus is one of the instruments of this kind, which opens possibilities to create and to build tools as speech recognizer and text-to-speech synthesizer for systems controlled by voice. The systems controlled by voice raise large-scale demands to speech corpora such as availability of annotated speech recordings in large quantities, diversity of speakers, availability of pronunciation dictionaries, mistake-free and high quality material. The process of development of speech corpora is usually standard and always involves time-consuming handwork. Some part of development process could be facilitated by introduction of extra tools and techniques.

The investigators of under-resourced languages are always trying to find solutions to the problems of how to overcome the scarcity of speech resources. Researchers investigate the choice of "useful" training examples (Axelrod *et al.*, 2015), produce a reliable pronunciation dictionary from limited available resources (Takahashi *et al.*, 2016), and exploit resources from a closely related language (Samson *et al.*, 2014). Nevertheless, the development of speech resources of a particular language is crucial for the progress of implementation of human-computer interfaces working in voice recognition and synthesis mode.

Lithuanian speech resources applicable to implementation of such task are scarce. Many speech corpora are created and used mainly for investigative purposes. Further we discuss the issue of development of speech corpus that applies to the needs of implementation of human-computer interfaces working in voice recognition and synthesis mode.

2. Related Works

A demand for a speech corpus can arise from different sources. It could be used for applications in industry, for scientific researches, for storage of national acoustic space. Different demands generate speech corpora with different attributes. Speech corpora differ in content of speech and in characteristics such as the number of speakers, level of annotation, structure of a corpus. The most valuable characteristic of a speech corpus is annotation of speech, and the most important task of corpora is to meet demands, fulfil requirements and generate expected outcomes. Unfortunately, digital support for 21 of the 30 European languages was defined as "non-existent" or "weak" (Meta-Net, 2018), Lithuanian with other 5 languages has "weak" support in the category of speech and text resources. Considering this group of languages, only Latvian speech corpus (Pinnis *et al.*, 2014) reach duration of 100 hours. Other languages that could be referred as supported "fragmentary" have more speech corpora, sometimes of a smaller size. If one of the largest Bulgarian speech corpus (Hateva *et al.*, 2016) reach up to 32 hours of duration, Romanian speech corpus (Stan *et al.*, 2017), Icelandic speech corpus (Language Resources in Icelandic, 2008) reach up to 21 hours of duration, other speech corpora as Norwegian language (Amdal *et al.*, 2008), Slovenian language (Zgank *et al.*, 2006; Zwitter *et al.*, 2013) have respectively 77, 100 and 120 hours of duration.

Since the development of speech recognizers and synthesizers involves application of complex methods, creation of speech corpora is time and labour consuming. It involves manual work, the aim to attain accuracy, to manage requirements that are complex, challenging and changing. Development of speech corpora follows the same sequence of activities: specification of a corpus, recording of speech and annotation. Difference in activities makes an impact on quality of a corpus and time consumption. The investigators try to discover automatic tools that could facilitate and shorten the process of development of a speech corpus (Petursson *et al.*, 2016), develop tools for annotation (Glavatskih *et al.*, 2015), incorporate already existing tools for automatic annotation (Pinnis *et al.*, 2014; Kamandulytė-Merfeldienė, 2017). Nonetheless, important researches are carried out in the field of annotation of speech, which is the most labour and time consuming activity. The majority of investigators demonstrate abilities in transcription of broadcast speech (Esteve *et al.*, 2010; Lileikytė *et al.*, 2016; Mansikkaniemi *et al.*, 2017), extent of annotation and linkage (Johannessen *et al.*, 2007).

Another direction of investigations is the development of speech resources for multimodal human-computer interaction. Interaction-based multimodal interfaces include typing, speech, lip-reading, eye-tracking and face recognition in combinations (Vo and Waibel, 1993). For these purposes multimodal corpora are developed (Grishina, 2010; Giraudel *et al.*, 2012; Czyzewski *et al.*, 2017), which try to annotate different artifacts as morphology, semantics, accent, speech act, gesture, segments of patterns, etc. However, specific corpora, which comprise one entity for speech recognition and synthesis purposes, are rare and lack specifically processed data for speech synthesis (Martins *et al.*, 1998). Developers of human-computer interfaces are constrained to use different speech resources to overcome this problem, but are confronted with difficulties to integrate speech resources of various characteristics.

The availability of Lithuanian speech corpora for investigative purposes is satisfactory. Speech corpora represent Lithuanian acoustic space, usually are of about 20 hours of duration, precisely annotated by human at phonemic level, and usually comprise spoken words, phrases, syllables, names of cities or persons (Kazlauskienė and Raškinis, 2013; Vaičiūnas *et al.*, 2016). A corpus of large extent would not usually have qualities, which could be obtained only by manual work, and this lack of quality results in a possible impairment of scientific investigations. Larger Lithuanian speech corpora are of about 50–100 hours of duration at the moment (Kazlauskienė and Raškinis, 2013; Kamandulytė-Merfeldienė, 2017). Speech corpora for synthesis purposes exist, but they are not publicized and are integrated in speech synthesis engines. Unfortunately, there is a lack of continuous speech corpora and corpora of large extent (Rudžionis *et al.*, 2014), corpora for synthesis purposes, which hamper specific investigations. With the aim to address above mentioned problem the corpus Liepa was developed.

3. Statement of the Problem

A human-computer interface requires two-way communication. Therefore, a computer should be able to solve two problems, i.e. to recognize a speech and to pronounce an answer. Two different problems are solved applying different techniques, which use speech resources of different characteristics. It is more practical to keep speech resources in one place and to consider these resources as one package, which is prerequisite and sufficient for development of automated speech recognition and synthesis engines.

The purpose of this research was to develop structure for speech corpus, which opens possibilities for implementation of human-computer interface using Lithuanian language. This research came alongside of the aim to develop the Lithuanian speech corpus. The corpus, which is presented in this article, is sufficiently large, has two different parts and is not too complex to be applied it in the construction of speech human-computer interfaces.

4. Description of Structure and Development of Speech Corpus Liepa

4.1. *Requirements for Speech Corpus*

Requirements for a speech corpus are usually derived from ideal characteristics of a corpus (Patil and Basu, 2009), but differences between methodologies of speech recognition and speech synthesis present different requirements. Speech recognition rarely uses the corpus directly; various methods transform the corpus in statistical models as neural networks or hidden Markov models. Whereas speech synthesis uses the corpus in most cases directly, various methods are used to combine small speech parts to form comprehensible utterance. The primary requirement is availability of high-quality data. The main differences are the following:

- Requirements for phonetic coverage. Speech data must include all features of phonemes of Lithuanian language, i.e. they have to reflect all vowels, consonants, fricatives, liquids, pure diphthongs and mixed diphthongs. Good phonetic coverage in speech corpora enables a more profound acoustic modelling.
- Quantity of speech data. Basically, the statistical methods that are used by speech recognition imply usage of as much data as possible and are able to distinguish and group similar features of speech. Speech data are usually used indirectly after speech signal processing as sampling, quantization, filtering, feature extraction and application of various methods involving Hidden Markov Models, neural networks, language models or adaptation techniques. Speech synthesis requires less speech data and uses them directly.
- Quality of speech data. As a consequence that speech synthesis uses speech data directly, data must be of the highest quality. Speech recognition with application of various preprocessing techniques and methods is able to cope with or manage noise and other artifacts and to capture features of speech data.
- Quantity of speakers. Speaker independent speech recognition systems requires gathering of data of many speakers. A primary criterion in selection of speakers is to be a native speaker with expressive pronunciation. Speakers from different age groups and gender have to be selected proportionally. Speech synthesis for one application uses data of one or few speakers.
- Intrusion of noise and other artifacts. The corpus should include all possible artifacts (cough, laugh, bang, etc.). The assumption is that low level noises do not impact an accuracy of speech recognition process. Speech data without significant noise, phonetic distortions of words have to be used. Speech synthesis excludes possibility to have any of mentioned artifacts in data.

Considering the given differences, further the requirements are given for Part 1 – a part of speech corpus for speech recognition purpose and Part 2 – a part of speech corpus for speech synthesis purpose:

- Size of speech corpus: 100 hours of speech data for Part 1 and 13 hours of speech data for Part 2. The size of speech data for Part 1 was chosen considering potential use of speech recognition method, which employs speech data for development of statistical models, quantity of potential speakers and aim to accumulate enough data for building speaker independent speech recognizer. The size of speech data for Part 2 was obtained from previous results of established speech data gathering method. Each speaker has to record the same amount of speech data that comprise all features of Lithuanian acoustic space. Speech data of all speakers compose the size of Part 2.
- Requirements for speakers. The primary criterion in selection of speakers is to be a native speaker with expressive pronunciation. Part 1 should include not less than 300 of speakers. Speakers from different age groups and gender have to be selected proportionally. Considering potential applications for which Part 2 could be used, as well the demand to stay attractive to different age groups, 2 men and 2 women of a youth and older age should be selected.
- Requirements for quality. For Part 1 speech data without significant noise, phonetic distortions of words have to be used. The aim is to collect the records as clean from noise as possible and this way not to make an impact on the modelling of speech recognition systems. The assumption is taken that at this early stage of modelling of speech recognition systems the main efforts should be given to achievement of high accuracy of speech recognition and not to coping with problems of noise elimination. For Part 2 of the corpus, speech data of the highest quality are chosen.
- Technical requirements. For Part 1 speech data should be gathered from different environments, with different technical characteristics of equipments for recording, but unvarying technical characteristics of speech records. For Part 2 technical characteristics of equipments and speech are fixed.
- Other requirements. Requirements regarding formats and naming of files, duration or introduction of pauses, ranking of words in speech recording are to be set at the beginning as unvarying characteristics of speech corpus; some changes could be made in the course of processes according to indications of speech recognition investigators.

4.2. *Description of Development of Part 1 of Speech Corpus*

The process of development of Part 1 is presented in Fig. 1.

It is similar to processes of many developers of corpora and to the particular process, given in Kazlauskienė and Raškinis (2013). The difference of this particular and presented process is the extent of handwork. The process depicts the main phases in sequential order. In practice, all the phases are interlaced; different stages ran at the same time. The unique components in the given process are Nos. 9, 10 and 11. The components Nos. 9 and 10 could be applied to any language as they include a fixed set of the verification rules and

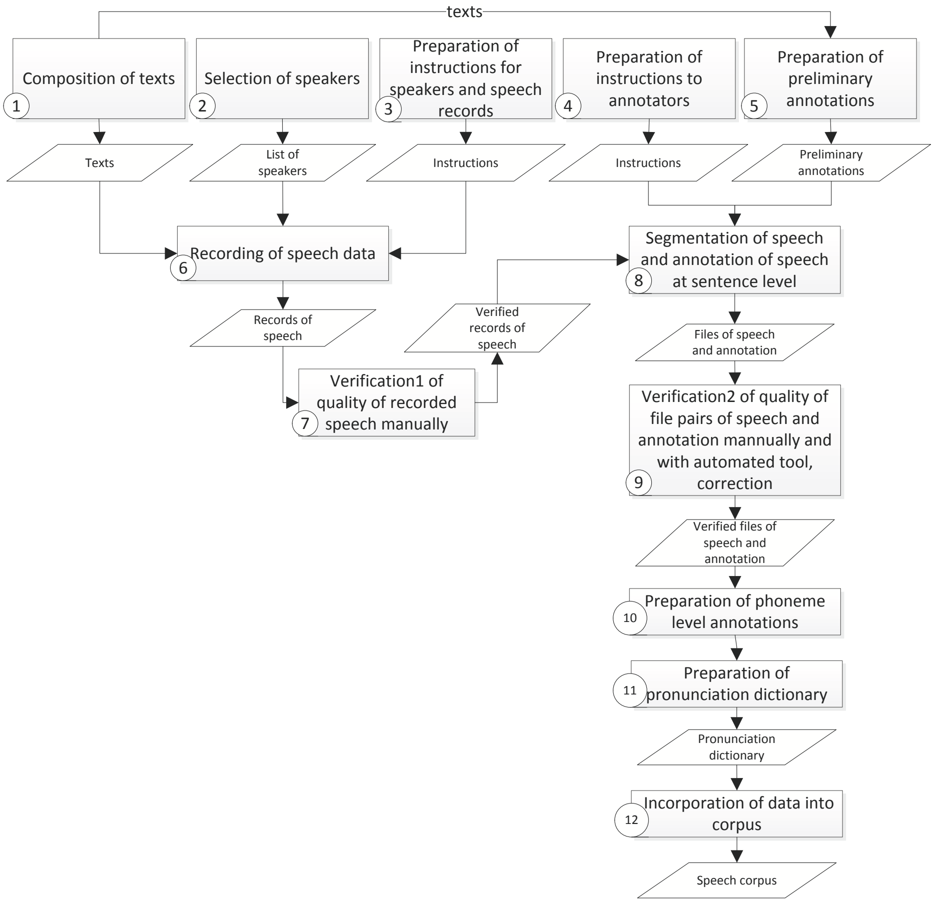
texts
Composition of texts
1
Selection of speakers
2
Preparation of instructions for speakers and speech records
3
Preparation of instructions to annotators
4
Preparation of preliminary annotations
5
Texts
List of speakers
Instructions
Instructions
Preliminary annotations
Recording of speech data
6
Segmentation of speech and annotation of speech at sentence level
8
Records of speech
Verified records of speech
Verification1 of quality of recorded speech manually
7
Files of speech and annotation
Verification2 of quality of file pairs of speech and annotation mannually and with automated tool, correction
9
Verified files of speech and annotation
Preparation of phoneme level annotations
10
Preparation of pronunciation dictionary
11
Pronunciation dictionary
Incorporation of data into corpus
12
Speech corpus

Fig. 1. The process of development of Part 1 of speech corpus.

a construction of the primitive speech recognition system. However, the component No. 11 is specific to a language and susceptible to a change of the phonemic representation.

The first block concerns the composition of texts. The input from linguists is a prerequisite for the achievement of texts that fully represent uniqueness of the phonemic space of a particular language. It becomes a challenge to represent very rare phonemes and diphthongs of a particular language and it is even more difficult to cope with the intrusion of phonetic space of other languages. The input from the potential users of a corpus is sought as well as it gives direction and defines the corpus. The second block is the selection of speakers. The distribution of age and gender of speakers should be equal. Additional criteria such as ethnicity make a corpus multifunctional. The activity of the sixth block is the recording of speech data. The seventh block is the first stage of verification, which covers manual verification of the correspondence of annotation to the speech signal. The generate feedback of this block is the recommendations how to improve the recording of speech data. Segmentation of speech data and annotation of speech data at sentence level

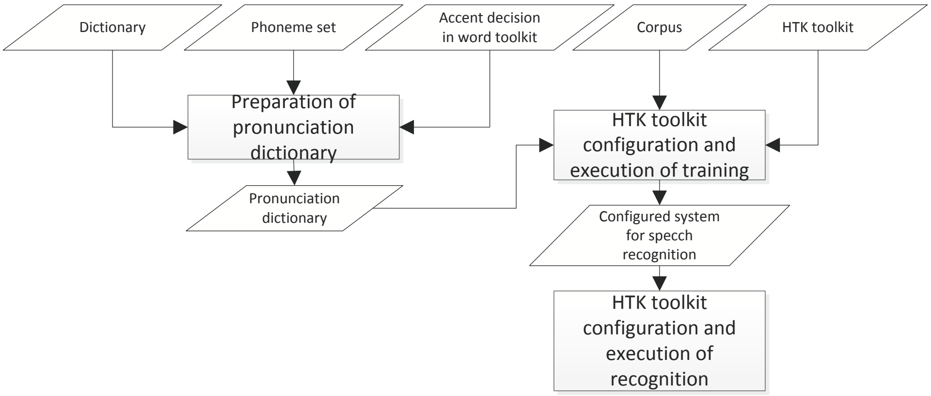
Dictionary
Phoneme set
Accent decision in word toolkit
Corpus
HTK toolkit
Preparation of pronunciation dictionary
Pronunciation dictionary
HTK toolkit configuration and execution of training
Configured system for specch recognition
HTK toolkit configuration and execution of recognition

Fig. 2. Automatic verification of Part 1 of corpus.

comprise activities of the eight block. An annotator has to listen and to make corrections to a preliminary annotation adjusting it to the audio recording. The second stage of verification in the ninth block includes automatic verification. This stage of verification is shown in Fig. 2 and consists of the construction of a primitive speech recognition system. Automatic toolkits help to track the correspondence between speech data and their annotations, allow examining the formats of files and structure of file names, the content of annotations, etc.

The tenth block offers a technique, which is described in Laurinčiukaitė *et al*. (2009). It consists of a realignment of speech data in the speech recognition process with the purpose to develop the phoneme-level annotations. The realignment of speech data is a known technique, used by many investigators, which takes training data of speech recognition system and performs an iterative process to find time boundaries of phonemes in speech data that match trained phoneme models the best. Application of this technique in development of speech corpus helps to automate an annotation process. The eleventh block helps to make pronunciation dictionary in a form of the word-to-phoneme transcriptions. The chosen phoneme set would likely have an impact on speech recognition results and for this reason a preliminary research on the phoneme set could help. One of the possible solutions is to use an already developed toolkit for word-to-phoneme transcriptions and to offer converting tools to compile the different phoneme sets.

4.3. *Description of Development of Part 2 of Speech Corpus*

Development of Part 2 of speech corpus mostly involves handwork. The first activity is composition of texts, which could represent Lithuanian acoustic space. The second activity involves the selection of speakers with the perfect pronunciation. Recording of speech is carried out in especially quite room. And the last activity is manual annotation at the level of diphones (Kasparaitis, 2005).

Following described processes we constructed the speech corpus Liepa.

5. Description of Speech Corpus Liepa

The quantity of texts for Part 1 was 78, 33 of them covered words and phrases, 45 of them covered continuous speech. The speakers had to read 5–6 texts. The lists of words and phrases spanned mostly exact commands, required by the group of speech recognition research. The texts of continuous speech spanned descriptions of UNESCO objects, protected animals and food. There were a lot of names of foreign origin, and that caused problems for speakers, annotators and even during the preparation of pronunciation dictionary, since the transcription of foreign names required use of elements of another phonetical system than Lithuanian. Elimination of the problem could be to give speakers the exact transcription of the pronunciation of a foreign name. The quantity of sentences for Part 2 was 5000. The speakers had to read all of them.

There were 376 speakers for Part 1: 116 speakers from schools and 260 from the main site (university students and invited speakers); 248 females and 128 males. The problem was to attract males and middle-aged or elderly people to recording site. In future, the strategy of attraction of people from outside of the university should be better developed. Four speakers were selected for Part 2.

The speech recordings for Part 1 were gathered in 6 locations (one main site and 5 schools) with different equipments and at a different time during 2 years. A group of speech recording was supported with texts, instructions, and technical requirements. The best quality of speech recordings was achieved at the main site, which was designed for this specific purpose and supervised by more experienced staff. These speech recordings included fewer mispronunciations, the speech was clearer; the speech signal was of a higher quality. The speech recordings for Part 2 were collected at the main site.

The basic phoneme set of the corpus Liepa described in Kasparaitis (2005) was used. The set included 92 phonemes: long and short vowels, soft and hard consonants, diphthongs (vowel-vowel) and affricates with later obtained accent information. This phoneme set reflects main attributes of Lithuanian language and includes accent information, which could be rarely obtained without specific language susceptible tool. The whole set of the attributes of the phonemes of Lithuanian language was included to present researchers the possibility to investigate different attributes and their impact to the accuracy of speech recognition. This phoneme set enables researchers to move smoothly from one phoneme set to another by removing conventional marks and combining phonemes. The dictionary was prepared using the accent decision toolkit and implementing the following approach, given in Kasparaitis (1999, 2000).

For verification of Part 1, the primitive speech recognition system was developed using HTK tools, described in HTK toolkit (2017). The training of hidden Markov models for 92 phonemes was performed on the speech data of the whole speech corpus. The hidden Markov models with 1-component Gaussian mixture were obtained, trained and used for the realignment process, which resulted in generation of the phoneme-level annotations for speech data.

Speech data of two different parts were kept separately. Speech data were grouped by a speaker. For Part 1 the structure of the corpus could be easily changed, since the construction of data file names enables the re-grouping of speech records and their annotation files.

Table 1
The characteristics of the speech corpus Liepa.

Criterion	Characteristics of corpus	
	Part 1	Part 2
Speech type	Continuous	
Corpus size	100 hours	13 hours
Speech content	Read words, phrases, sentences, texts	Sentences
Annotation	Sentence-level and time-aligned phoneme-level annotations	Diphone-level annotations
Number of speakers	376	4
Sampling	22 kHz	
Quantization	16 b	
Channels	Mono	
Phone set	92	

The name of each file encoded different information attributes, which made the structure of speech corpus flexible. The corpus was partitioned into training and evaluation data sets. The evaluation data set was used only by the group of speech recognition investigators and was divided into small parts according to the applications tested. Speakers for training and evaluation sets were different. The corpus could be used for further research without attention to its artificial division for temporary usage.

The main characteristics of the speech corpus Liepa are given in Table 1.

6. Applications

The speech corpus Liepa as a whole was used in four applications and had different patterns of usage. The Part 1 was transformed into format, used by the specific speech recognition system. Changes were applied to a phoneme set and composition of a pronunciation dictionary by a specific software (Greibus *et al.*, 2017). This example demonstrates that every speech corpus can be used according to the needs of speech researchers and processed in different ways. Part 2 was used and should be used in the future completely unchanged.

Subsequent use of the speech corpus Liepa as a whole or as some part could include research of various speech recognition systems and practical implementations. The speech corpus Liepa has sufficient amount of data for scientific research, such as investigation of attributes of Lithuanian speech, development of recognition systems. The simplest case of application of the speech corpus Liepa is acoustic modelling.

Some issues to be solved by investigators are: definition of training, development, and evaluation data sets, construction of language model, development of software for the transformation of a phoneme set and a pronunciation dictionary.

To date the corpus Liepa is not employed in speech-based services, despite being free of charge.

7. Conclusions

1. The problem of development of speech corpus structure, which could fit the needs of development of human-computer interfaces working in voice recognition and synthesis mode, was investigated.
2. In order to develop economical and comfortable speech corpus for development of human-computer interfaces working in voice recognition and synthesis mode, the speech corpus was divided into two parts: part one of speech corpus for development of text-to-speech synthesizers; part two of speech corpus for development needs of speech recognition engines.
3. It has been determined that the part of speech corpus designed for speech recognition engines has to ensure the availability to present speech specificity by the use of different sets of phonemes.
4. According to the investigation results, Lithuanian speech corpus, which consists of two parts, was developed for human-computer interfaces working in voice recognition and synthesis mode:

   - Part 1 of the speech corpus Liepa, duration of which is 100 hours, for development of speech recognition engines,
   - Part 2 of the speech corpus Liepa, duration of which is 13 hours, for development of text-to-speech synthesis engines.

Acknowledgements. The development of speech corpus has been supported by the European Structural and Investment Funds. Project VP2-3.1-IVPK-12-K-01-001.

References

Amdal, I., Strand, O., Almberg, M.J., Svendsen, T. (2008). RUNDKAST: an annotated Norwegian broadcast news speech corpus. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation*, LREC'08, Morocco, pp. 1907–1913.

Axelrod, A., Resnik, P., He, X., Ostendorf, M. (2015). Data selection with fewer words. In: *Proceedings of the Tenth Workshop on Statistical Machine Translation*, Lisbon, pp. 58–65.

Czyzewski, A., Kostek, B., Bratoszewski, P., Kotus, J., Szykulski, M. (2017). An audio-visual corpus for multimodal automatic speech recognition. *Journal of Intelligent Information Systems*, 49(2), 167–192.

Esteve, Y., Bazillon, T., Antoine, J.-Y., Bechet, F., Rarinas, J. (2010). The EPAC corpus: manual and automatic annotations of conversational speech in French broadcast news. In: *Proceedings of the 7th International Conference on Language Resources and Evaluation*, LREC'10, Malta, pp. 1686–1689.

Giraudel, A., Carre, M., Mapelli, V., Kahn, J., Galibert, O., Quintard, L. (2012). The REPERE Corpus: a multimodal corpus for person recognition. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation*, LREC 2012, pp. 1102–1107.

Glavatskih, I., Platonova, T., Rogozhina, V., Shirokova, A., Smolina, A., Kotov, M., Ovsyannikova, A., Repalov, S., Zulkarneev, M. (2015). The multi-level approach to speech corpora annotation for automatic speech recognition. In: *Proceedings of 17th International Conference of Speech and Computer*, SPECOM 2015, Athens, pp. 438–445.

Greibus, M., Ringelienė, Ž., Telksnys, L. (2017). The phoneme set influence for Lithuanian speech commands recognition accuracy. In: *Proceedings of Open Conference of Electrical, Electronic and Information Sciences*, eStream, Vilnius. pp. 1–4.

Grishina, E. (2010). Multimodal Russian corpus (MURCO): first steps. In: *Proceedings of the 7th Language Resources and Evolution Conference*, LREC 2010, pp. 2953–2960.

Hateva, N., Mitankin, P., Mihov, S. (2016). BulPhonC: Bulgarian speech corpus for development of ASR technology. In: *Proceedings of the 10th International Conference on Language Resources and Evaluation*, LREC 2016, pp. 771–774.

HTK toolkit. http://htk.eng.cam.ac.uk/. Last view online: 2017.

Johannessen, J.B., Hagen, K., Priestley, J., Nygaard, L. (2007). An advanced speech corpus for Norwegian. In: *Proceedings of the 16th Nordic Conference of Computational Linguistics*, NODALIDA-2007, pp. 29–36.

Kamandulytė-Merfeldienė, L. (2017). Grammatically coded corpus of spoken Lithuanian: methodology and development. *Engineering and Technology International Journal of Cognitive and Language Sciences*, 11(4), 853–857.

Kasparaitis, P. (1999). Transcribing of the Lithuanian text using formal rules. *Informatica*, 10(4), 367–376.

Kasparaitis, P. (2000). Automatic stressing of the Lithuanian text on the basis of a dictionary. *Informatica*, 11(1), 19–40.

Kasparaitis, P. (2005). Diphone databases for Lithuanian text-to-speech synthesis. *Informatica*, 16(2), 193–202.

Kazlauskienė, A., Raškinis, G. (2013). Principles of development of the intonational annotated spoken corpus. *Žmogus ir žodis: didaktinė lingvistika*, 15(1), 101–110 (in Lithuanian).

Language Resources in Icelandic: Parliament Speech Corpus.
http://www.malfong.is/index.php?pg=althingi&lang=en. Last view online: 2018.

Laurinčiukaitė, S., Filipovič, M., Telksnys, L. (2009). Lithuanian continuous speech corpus LRN 1: an improvement. *Information Technology And Control*, 38(3), 203–207.

Lileikytė, R., Gorin, A., Lamel, L., Gauvain, J.-L., Fraga-Silva, T. (2016). Lithuanian broadcast speech transcription using semi-supervised acoustic model training. In: *Proceedings of 5th Workshop on Spoken Language Technologies for Under-Resourced Languages*, SLTU-2016, Yogyakarta, pp. 107–113.

Mansikkaniemi, A., Smit, P., Kurimo, M. (2017). Automatic construction of the finnish parliament speech corpus. In: *Proceedings of INTERSPEECH 2017*, Stockholm, pp. 3762–3766.

Martins, C., Mascarenhas, M.I., Meinedo, H., Neto, J.P., Oliveira, L., Ribeiro, C., Trancoso, I., Viana, C. (1998). Spoken language corpora for speech recognition and synthesis in European Portuguese. In: *Proceedings of 10th Portugese Conference on Pattern Recognition*, RECPAD'98, Lisboa.

Meta-Net, White Paper Series: Press Release. http://www.meta-net.eu/whitepapers/press-release. Last view online: 2018.

Patil, H.A., Basu, T.K. (2009). Development of speech corpora for speaker recognition research and evaluation in Indian languages. *International Journal of Speech Technology*, 11(1), 17–32.

Petursson, M., Klüpfel, S., Gudnason, J. (2016). Eyra – speech data acquisition system for many languages. In: *Proceedings of 5th Workshop on Spoken Language Technologies for Under-Resourced Languages*, SLTU-2016, Yogyakarta, pp. 53–60.

Pinnis, M., Auzina, I., Goba, K. (2014). Designing the Latvian speech recognition corpus. In: *Proceedings of the 9th Edition of the Language Resources and Evaluation Conference*, LREC'14, Reykjavik.

Rudžionis, V., Raškinis, G., Ratkevičius, K., Rudžionis, A., Bartišiūtė, G. (2014). Medical – pharmaceutical information system with recognition of Lithuanian voice commands. In: *Proceedings of 6th International Conference: Human Language Technologies – The Baltic Perspective HLT*, Riga, pp. 40–45.

Samson, J.S., Besacier, L., Lecouteux, B., Tan, T.-P. (2014). Using closely-related language to build an ASR for a very under-resourced language: Iban. In: *Proceedings of Co-Ordination and Standardization of Speech Databases and Assessment Techniques* (COCOSDA), Phuket, pp. 1–5.

Stan, A., Dinescu, F., Ţiple, C., Meza, S., Orza, B., Chirila, M., Giurgiu, M. (2017). The SWARA speech corpus: a large parallel romanian read speech dataset. In: *Proceedings of the 9th Conference on Speech Technology and Human-Computer Dialogue*, Bucharest.

Takahashi, N., Naghibi, T., Pfister, B. (2016). Automatic pronunciation generation by utilizing a semi-supervised deep neural networks. In: *Proceedings of the 17th Interspeech 2016* (submitted on 15 Jun 2016).

Vaičiūnas, A., Raškinis, G., Kazlauskienė R. (2016). Corpus-based hidden Markov modelling of the fundamental frequency of Lithuanian. *Informatica*, 27(3), 673–688.

Vo, M.T., Waibel, A. (1993). Multimodal human-computer interaction. In: *Proceedings of the International Symposium on Spoken Dialogue*, ISSD'93.

Zgank, A., Rotovnik, T., Grasic, M., Kos, M., Vlaj, D., Kacic, Z. (2006). SloParl – Slovenian parliamentary speech and text corpus for large vocabulary continuous speech recognition. In: *Proceedings of Interspeech 2006*, pp. 197–200.

Zwitter, V.A., Zemljaric, M.J., Krek, S., Stabej, M., Erjavec, T. (2013), Spoken corpus Gos 1.0, Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1040. Last view online: 2018.

S. Laurinčiukaitė received her PhD degree from Vilnius Gediminas Technical University. From 2000 to 2008 she worked at the Institute of Mathematics and Informatics. Currently she is working with development of speech corpora. Her research field is HMM based methods for Lithuanian speech recognition, development of speech corpora.

L. Telksnys, professor, doctor habilitatis in informatics, doctor honoris causa of the Kaunas University of Technology, member of Lithuanian Academy of Sciences, senior research fellow of Recognition Processes Department at the Institute of Mathematics and Informatics, Vilnius University, Lithuania. He is the author of an original theory of detecting changes in random processes, investigator and developer of a computerized system for statistical analysis and recognition of random signals. His current research interests are in analysis and recognition of random processes, cardiovascular signals and speech processing.

P. Kasparaitis graduated from Vilnius University (Faculty of Mathematics) in 1991. He became a PhD student at Vilnius University in 1991. In 2001 he defended his PhD thesis. Presently he is an associate professor at Vilnius University. Current research includes text-to-speech synthesis and other areas of computer linguistics.

R. Kliukienė received her PhD degree from Vilnius University. She worked at Vilnius University, the Faculty of Philology. She supervised the construction of the Lithuanian speech corpus Liepa.

V. Paukštytė received her master's degree in Lithuanian linguistics from the Vilnius University in 2012. She worked with the Lithuanian speech corpus Liepa.