Linguaacoustic resources for belarusian speech synthesizers

Zianouka E.
The Center for belarusian culture, language and literature
National Academy of Sciences
Minsk, Republic of Belarus
evgeniakacan@gmail.com

***Abstract*—The article describes an algorithm for creating a linguaacoustic data array for Belarusian text-to-speech systems, which consists of 350 communicative-syntactic units of a text corpus with coverage of all possible syntactic structures of statements and full punctuation of the Belarusian language. It also gives a reason for creating such resources for synthesizers, depicts its constituents and the principle of material processing.**

***Keywords*—computational linguistics, linguaacoustic resources, data array, text-to-speech synthesizers**

I. Introduction

Today computer technology is used in almost all spheres of human activity. One important area of its application is the development of text-to-speech systems (TTS), which automatically convert an electronic text to speech. High-quality TTS have wide potential in various fields of economy, science, culture, medicine, education and others [1]. There are a huge number of synthesizers that handle different languages. Speech synthesis and recognition laboratory of UIIP NASB for the past 55 years has been developing belarusian text-to-speech and recognition systems. Today, stationary platform of synthesizer "Multiphone-4", speech synthesizer for mobile platforms and its internet-version are built up [2]. The main feature of these developments, in particular an internet synthesizer, is an open free access following the link http://www.corpus.by/tts3/ [3]. The functionalities of the synthesizer are quite high and diverse, but there are some drawbacks. Therefore, the relevance for a linguaacoustic data array is explained by the necessity of improving the quality of information processing, adding supplementary functions and testing existing ones through linguistic resources for such systems in the Belarusian language [4].

II. The relevancy of linguaacoustic data array

Modern computer technologies allow for collecting and analyzing linguaacoustic resources, building them in TTS through regular formal rules and algorithms that have been outlined by natural language. Under linguaacoustic resources is meant a set of organized in a certain way speech and language data which are saved in machine storage media and are used in various fields of practice (education, industry, economy, culture and art, etc.) [5]. Large linguistic data files are used for creating and developing efficient systems of text and speech processing. In TTS under linguaacoustic resources is understood an array of language data, which analyze and test the input text (see figure 1). The general scheme of the speech synthesizer is as follows[6]: an input text is primarily processed by the text processor, in which word stress, letter-phonemic processing, its splitting into syntagmas, the choice of intonation type for each syntagma are taken place. Then marked phonemic text is fed to the input of two processors: prosodic and phonetic. The phonetic processor generates positional and combinatorial allophones of phonemes. The prosodic processor determines the current values of the amplitude and duration of sounds, as well as the frequency of the fundamental tone. It also answers for the right and emphatic intonation of voiced text (see figure 2). Prosodic processor requires more rework in the Belarusian text-to-speech synthesizers, then under linguaacoustic data array is meant voiced marked dataset with all the intonation signs, types and constructions. Therefore, the main concept in this paper is voiced data with intonation marking.

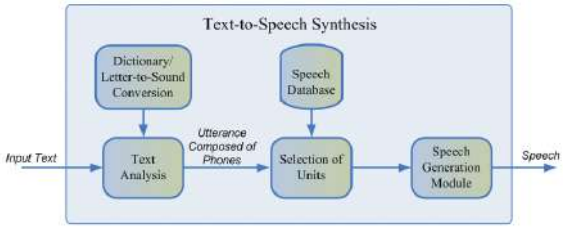

Figure 1. The structure of tex-to-speech systems

III. Criteria for material collection

Due to the fact that the linguaacoustic data array is being created for TTS to process text into emphatic speech, we must collect the material which formal algorithm could analyze. Since Belarusian synthesizers (BTTS) don't have syntactic parser, the machine will be

able to analyze an incoming text only by formal indication. These are syntactic structure of the sentence and punctuation. BTTS dataset should include sentences or expressions with various schemes of syntactic structures and punctuation. The syntactic structure of the sentence the Belarusian literary language is divided into [7]:

1) Simple sentences:
   - one-member sentences;
   - two-member sentences.
2) Complex:
   - compound;
   - complex;
   - conjunctionless sentences;
   - mixed complex sentences with include different types of complex sentences.

Except for the syntactic structure of the Belarusian language we are interesting in punctuation, as punctuation markers are formal indicators of the number of syntagmas and accentual units and the conditional indicators of intonation type in text. Therefore, we turn to the analysis of the following items [8]:

1) The communicative type of statement (narrative, question, imperative);
2) Punctuation used in this sentence (period, question mark, exclamation point, comma, semicolon, colon, dash, hyphen, parentheses, brackets, braces, apostrophe, quotation marks, and ellipsis);
3) Number of syntagmas in this statement.

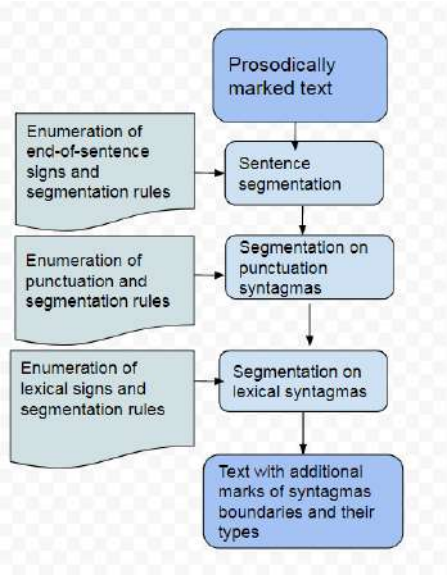

Figure 2. The prosodic processor structure

IV. The composition of linguaacoustic data array

The edition ""Bielaruskaja litaraturnaja spadcyna: antalohija u 2 tamach" was shoosen as bulk material for the data array. It equips belarusian masterpieces of XX - XXI cc. Their authors are A. Adamovich, G. Dolidovich, R. Borodulin, A. Vertinsky, J. Yanishchits, N. Gilewicz, A. Makayonok, P. Brovko et al [9]. As noted above, we have drawn attention to the main three points while collecting the data. Thus, 350 communicative-syntactic units have been defined, 100 of which will make linguaacoustic database, while the remaining 250 will be used for testing BTTS after inserting and verificating linguaacoustic database as part of a speech synthesizer. With 100 units selected 50 are narrative (about three, four statements for each syntactic unit of the text, taking into account all possible syntactic constructions), 50 occur in interrogative and imperative sentences. In the process of selecting interrogative sentences we have taken into account w-questions, general questions, questions which begin with participle and the number of syntagmas. Imperative statements were collected on the basis of emotional or intention of expression and the number of syntagmas [7].

Unfortunately, today there is no computerized software for the determination of the accent structure. This process is manual. Collected material is marked in accordance with the conventional signs, introduced by the author, such as stress (nuclear stress and secondary), boundaries(the border of phonetic syntagma and the border of phonetic phrase),tones Tags (rising tone, falling tone, rising-falling tone, falling-rising tone; rising + neutral + falling tone;rising - falling + neutral + rising - falling tone; falling-rising +neutral + falling-rising tone; intonation of incompleteness, understatement, which is understood (is used in the blank). A fragment of the markup text material for a database is shown in Table 1. It transmits the type of statement, calls the sentence structure, the number of syntagmas, the author remarks, syntactic unit and its marked-up version.

The next step is experimental recording of communicative-syntactic units with all intonation features of the Belarusian literary language. For this, the program Sonic Sound Forge 11 was chosen for the further audio recording, its editing and marking. Since this is a test recording and audio signal processing, adhere to the basic recording conditions: 8000 Hz 16 bit Mono on a simple or a built-in microphone system [10].

After recording, the regions are accentuared (nuclear (n), prenuclear (p) and postnuclear (t)) in a sound track for visual perception of graphic contour, the quality of its recording. For this we use program Inton@trainer. Figure 3 illustrates the process of recording a narrative sentence Cioplaje pavietra pieralivalasia na so+ncy in Sonic Sound Forge. Then the audio has been divided into regions (nuclear (n), prenuclear (p) and postnuclear (t)) (see. Figure 4). This procedure is necessary to assess the correctness of statement's intonation marking to accent

TABLE 1. The fragment of the marked text corpus for linguaacoustic data array

Тып выказвання	*Структура сказа*	*Сказ*	*Размечаны сказ*	*Колькасць сінтагм*	*Тып інтанацыі*	*Аўтар выказвання*
апавядальны	назыўны	Раніца.	\ Ра+ніца ||	1	інт. завершанасці	А.Адамовіч
апавядальны	двухсастаўны	Цёплае паветра пералівалася на сонцы.	Цёплае /паве=тра | пералівалася на \со+нцы ||	2	інт. завершанасці	І. Чыгрынаў
апавядальны	просты + дзеепр. зв.	Ад спакойных, троху пахіленых каласоў дыхала прыемная радасць.	Ад /спако=йных | троху пахіленых /каласо=ў | дыхала прыемная \ ра+дасць ||	3	інт. незавершанас	І. Мележ
апавядальны	просты +адн. чл.	Супакой, ціша і прохаладзь панавалі тут.	/Супако=й | /ці=ша | і \\про+халадзь \панавалі тут ||	4	інт. пералічэння + інт. завершанасці	І. Мележ
апавядальны	смешаны	Колас — гэта сімвал жыцця, адраджэння, плён працы хлебароба, ён складаецца з некалькіх зярнятак...	/Ко+лас || гэта сі=мвал /жыцця+ | /адраджэ+ння || плё=н працы \\хлебаро+ба|| ён складаецца | з некалькіх \зярнятак||+	5	інт. удакл. + інт. пералічэння + інт. недаказанасці	І. Мележ
пабуджальны	форма ветл	Да сустрэчы!	Да \\ сустрэ+чы ||	1	інт. пабуджэння	П.Броўка

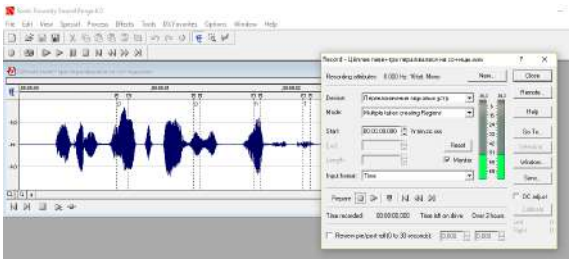

Figure 3. An audio recording and marking regions of the sentence Cioplaje pavietra pieralivalasia na so+ncy in Sonic Sound Forge

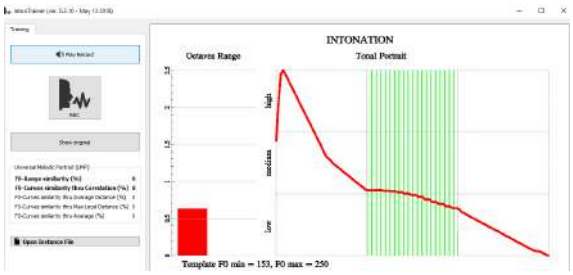

Figure 4. Graphic tonal countour of the sentence Cioplaje pavietra pieralivalasia na so+ncy in Intontrainer

units and sound quality which is wholesome for the further recording of full version by an expert. This procedure is made in the software "Inton@trainer" [11]. This Application is used as an instrument in a number of scientific and practical studies, namely the study of individual, emotional and stylistic features of intonation. Comparative evaluation of speech intonation in norm and pathology. Estimation of the intonational quality of synthesized speech. Figure 4 notes tonal intonation contour of Cioplaje pavietra pieralivalasia na so+ncy, where you can check the quality of sound, estimate the correct marking of regions (in the figure 4 nuclear region is mentioned by vertical lines).

The above mentioned steps are performed for all 350 units of the text material on the basis of which further linguaacoustic database will be created. An updated table of text material comprises a reference to the audio file and displays the graphic contour of communicative-syntactic units (Table 2). According to this link, the user will automatically follow a link into a folder with the necessary material, which is stored in the cloud on Google drive of the author.

V. Conclusion

The algorithm for compiling linguaacoustic data array is shown above. The development of linguaacoustic resources for the belarusian text-to-speech synthesizer makes it possible to improve the quality of the system's functionality, continuous testing, analyse input, intermediate and final data for the development and improvement of the system, collect new linguistic resources. Linguaacoustic data array will allow to solve various problems. In particular, it makes possible to voice text information with a clear and distinct speech, which is similar to the human.

References

[1] Rechevye texnologii. programmy: sintez rechi [Speech technology. Programs: speech synthesis]. Avaliable at: http://speech - soft.ru/index.php?a=inf&inf=view&id%20razdel=1181017375 (acessed 15.03.2018).

[2] Hetsevich Yu. S. Alharytmy linhvistychnay apratsouki tekstau dlya sintezu maulennya na belaruskay i ruskay movakh: dys. kand. tekhn. navuk [Algorithms of linguistic processing for speech synthesis in Belarusian and Russian languages. Cand. Diss. Minsk, 2012. 198 p.

TABLE2.The fragment of the marked text corpus for linguaacoustic data array with audio links and tonal countors

Парад кавы нумар	*Тып выказвання*	*Структура сказа*	*Сказ*	*Размечаны сказ*	*Аўдыя*	*Танальны контур*	*Тып інтанацыі*	*Аўтар выказвання*
1	апавядальны	аднасастаўны	Раніца.	↘ Ра+ніца ||	1_апав._76кб.wav	1_апав.	інт. завершанасці	А.Адамовіч
2	апавядальны	двухсастаўны	Цёплае паветра пералівалася на соцы.	Цёплае ↗паве=тра | пералівалася на ↘со+цы ||	2_апав._128кб.wav	2_апав.	інт. завершанасці	І. Чыгрынаў
3	апавядальны	просты + дзеепр. зв.	Ад спакойных, троху пахіленых каласоў дыхала прыемная радасць.	Ад ↗спако=йных | троху пахіленых ↗каласо=ў | дыхала прыемная ↘ ра+дасць ||	3_апав._136кб.wav	3_апав.	інт. незавершанасці	І. Мележ
4	апавядальны	просты +адн. чл.	Супакой, ціша і прохаладзь панавалі тут.	↗Супако=й | ↗ці=ша | і ↘↘про+халадзь ↘панавалі тут ||	4_апав._183кб.wav	4_апав.	інт. пералічэння + інт. завершанасці	І. Мележ
5	апавядальны	смешаны	Колас — гэта сімвал жыцця, адраджэння, плён працы хлебароба, ён складаецца з некалькіх зярнятак...	↗Ко+лас || гэта сі=мвал ↗жыцця+ | ↗адраджэ+ння || плё=н працы ↘↘хлебаро+ба|| ён складаецца | з некалькіх ↘зярнятак||+	5_апав._270кб.wav	5_апав.	інт. удакл. + інт. пералічэння + інт. недаказанасці	І. Мележ
6	Пабуджальны	форма ветл	Да сустрэчы!	Да ↘↘ сустрэ+чы ||	6_апав._95кб.wav	6_апав.	інт. пабуджэння	П.Броўка
7	Пабуджальны	форма ветл	Дзень добры ў хату!	Дзень ↘↗до+бры ў хату ||	7_апав._154.wav	7_апав.	інт. ветлівасці	

[3] Sintezatar maulennya pa tekstse [Text-to-speech synthesizer]. Available at: http://corpus.by/TextToSpeechSynthesizer/ (accessed 01.07.2017).

[4] Zianouka Ja.S. Modern technology in text to speech transformation: text-to-speech synthesizers. Trudy BHTU [BSTU Proceedings]. Minsk, 2018, No 1 (207) Print and media technologies,, pp.77-83.

[5] Proxorenko O.G. Linguistics, linguodidactics, linguoculturology: current issues and prospects of development. Materialy I respublikanskoj nauchno-prakticheskoj konferencii s mezhdunarodnym uchastiem [Proceedings of the I Republican scientific-practical conference with international participation]. Minsk, BSU Publ. center, 2017. 289 p.

[6] Lobanov B. M. Komp'yuternyy sintez i klonirovanie rechi [Computer synthesis and speech cloning]. Minsk, Belorusskaya nauka Publ., 2008. 357 p.

[7] Nikolaeva T.M. Frazovaya intonaciya slavyanskih yazykov [Phrasal intonation of Slavic languages]. Moskva, Nauka Publ., 1977. 278 p.

[8] Vigonnaya L.T. Intanacyja. Nacisk. Arfaepija [Intonation. Stress. Orthoepy]. Minsk, Nauka i tehnika Publ., 1991/ 215 p.

[9] Bielaruskaja litaraturnaja spadcyna : antalohija. U 2 kn. [Belarusian literary heritage: an anthology in 2 books]/ P.M. Lapo. Minsk, Belorusskaya nauka Publ., 2011. 944 p.

[10] Zianouka Ja.S. Modern technologies for intonation constructions processing. Lingvistika, lingvodidaktika, lingvokulturologiya: aktualnye voprosy i perspektivy razvitiya: materialy II mezhdunarodnoj nauchno- prakticheskoj konferkncii [Proceedings of the II Republican scientific-practical conference with international participation]. Minsk, BSU Publ. center, 2018. pp. 36-43.

[11] Intontrainer. Available at: https://intontrainer.by/ (accessed 12.03.2018).

[12] Dutoit T. An Introduction to text-to-speech synthesis. Norwell, MA, USA: Kluwer Academic Publishers, 1997. 420 p.

[13] Rusanova O.A. Modern technologies of oral speech synthesis. Available at: http://docplayer.ru/38896137-Sovremennye-tehnologii-sinteza-ustnoy-rechi.html (accessed 29.07.2017).

[14] Taylor P. Analysisand synthesis of intonation using the tiltmodel. J. Acoust. Soc. America, 2000, vol. 107, no. 3. pp. 1697–1714.

[15] Zakhar'ev V.A. Multi-voice text-to-speech synthesis for the construction of natural language interfaces of intellige9+nt systems. Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem: materialy mezhdunarodnoy nauchno-tekhnicheskoy konferentsii [Proceedings of the International Scientific and Technical Conference "Open Semantic Technologies for Intelligent Systems"]. Minsk, 2017, chapter 1, pp. 167-170.

ЛИНГВОАККУСТИЧЕСКИЕ РЕСУРСЫ ДЛЯ БЕЛОРУССКОЯЗЫЧНЫХ СИСТЕМ СИНТЕЗА РЕЧИ

Зеновко Е.С.

В статье описывается алгоритм по созданию лингвоакустического массива данных для белорусскоязычных систем синтеза речи, которая состоит из 350 коммуникативно-синтаксических единиц текстового корпуса с охватом всех возможных синтаксических структур высказываний и полной пунктуацией белорусского языка. Актуальность создания подобной базы данных обусловлена тем, что существующие системы синтеза речи на белорусском языке, несмотря на высокий уровень качества, далеки от совершенства. Одной из возможностей использования синтезированной речи является осуществление современных компьютерных технологий, позволяющих проводить сбор и совершенствование лингвистических ресурсов для подобных систем и улучшение их функциональности через формальные регулярные правила и алгоритмы, написанные на естественном языке. Разработка необходимых лингваакустычных ресурсов для синтезатора речи открывает возможность улучшения качества функциональности системы, постоянного тестирования системы, анализа входных, промежуточных и итоговых данных для доработки и совершенствования системы, сбора новых лингвистических ресурсов. Создание базы позволит решать разнообразные задачи. В частности, проводить озвучивание электронных текстов выразительной речью, подобной человеческой.

Received 29.12.18