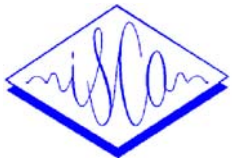

SPECOM'2004: 9th Conference
Speech and Computer
St. Petersburg, Russia
September 20-22, 2004

Polish TTS in Multi-Voice Slavonic Languages Speech Synthesis System

Edward Shpilewski, Bozhena Piurkowska, Janush Rafalko*
Boris Lobanov, Vitaly Kiselov, Liliya Tsirulnik**
**Institute of Computer Sciences, University of Bialystok,*
Sosnowa str. 64, 15-887 Bialustok, Poland.
E-mail: edszp@ii.uwb.edu.pl
***United Institute of Information Problems, Nat. Ac. of Sc. of Belarus,*
Surganov str. 6, 220012 Minsk, Belarus
E-mail: lobanov@newman.bas-net.by

Abstract

The report describes the Polish TTS and a set of Polish language-specific resources utilized by TTS. It presents both vocabulary and grammar, acoustic databases and a set of specific rules for word-stress position marking, letter-to-phoneme conversion and others.

1. Introduction.

Slavonic language and speech systems, in particular, those of Belarusian, Polish, Russian and Ukrainian, have very much in common. It refers to their phonetic, lexical, morphological and syntactic structure. The fact enables the researchers to set the creation of an integrated algorithm of multi-language TTS conversion and the construction of a new TTS system common for all these languages as the objective. One may expect that such a system will be also applicable to other Slavonic languages, such as Czech, Slovak, Serbo-Croatian, Slovenian, Bulgarian and Macedonian. At present, only a few TTS systems of Polish and Russian speech generation are available in the market. However, the quality of the synthesized speech is still far from natural, and the number of synthetic voices is very restricted. Belarusian and Ukrainian TTS systems are not presented in the market at all.

The report describes a part of an International research project "Multi-Lingual and Multi-Voice TTS-Synthesis System for Slavonic Languages" [1], namely TTS for Polish. The system has common structure for all Slavonic languages but it uses different linguistic and acoustical resources for each language. The project objective is development of a high-quality multi-lingual and multi-voice TTS-system on a common platform for many Slavonic languages. The objective can be achieved using original algorithms of multi-language and multi-voice TTS synthesis, which were developed in UIIP NAS Belarus before. The Synthesis of phonemic speech characteristics is based on the Allophones Natural Waves (ANW) method of speech signal concatenation. The basic principle of synthesizing the prosodic speech features is division of an utterance into accentual units and formation on their basis of entire tonal, rhythmical and dynamic contours of a syntagm and an utterance as a whole [2]. Using Data Driven (DD) approach, the TTS system will adress to a vast multilingual set of ANW and prosodic features databases for the synthesis of speech sounds and intonation. In order to synthesize prosodic features the system will also resort to deep morphological and syntactic analysis of sentences [3]. The two modules operating jointly are expected to achieve a high quality of synthesized speech.

2. General structure of the TTS-synthesizer

General structure of the multi-lingual and multi-voice TTS-synthesizer looks the following way (see Fig.2.1).
The incoming orthographic text undergoes a number of successive analytical operations carried out with the help of specialized processors. A ***textual*** processor is devised to transform the incoming orthographic text into a marked phonemic one. The processor performs the following tasks:

- transforming of numbers, abbreviations, shortenings;
- placing word stress;
- dividing orthographic text into accentual units (AU);
- dividing orthographic text into syntagms;
- marking intonation type within the syntagms;
- phonemic transcribing of the text;

The marked phonemic text is then sent to two processors: prosodic and phonetic.

The ***prosodic*** processor performs the following tasks:

- splitting AU into the elements of accentual units (EAU): pre-nuclear, nuclear and post-nuclear parts;
- measuring of amplitude (A), phoneme duration (T) values as well as fundamental frequency (F0) for each EAU.

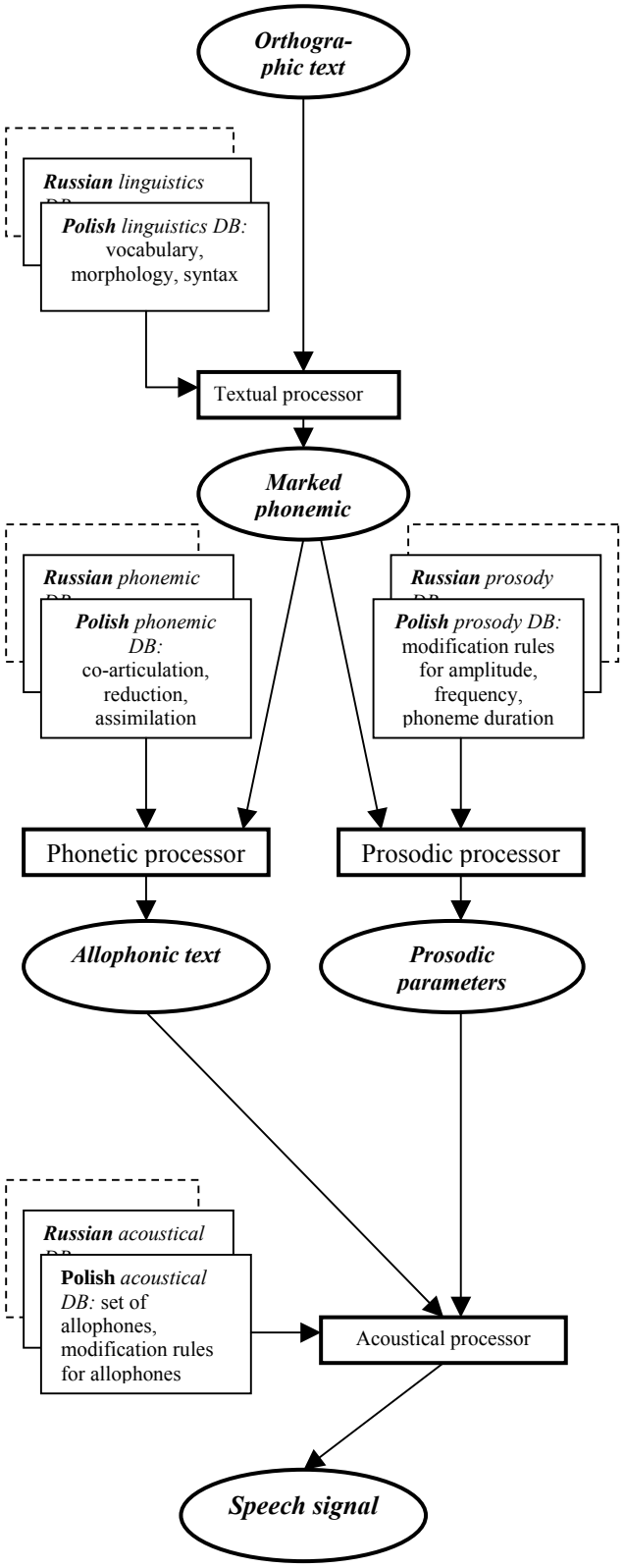

Fig 2.1. General structure of the multi-lingual TTS - synthesizer

The ***phonetic*** processor performs the following tasks:

- generating positional allophones;
- generating combinatory allophones from the in-coming phonemic text.

The ***acoustical*** processor uses the ***phonetic*** and ***prosodic*** processors information to determine:

- the allophones that are necessary to synthesize;
- the prosodic characteristics to be ascribed to each allophone.

Finally, it generates the speech signal by concatenating portions of allophone sound waves and their modifications in accordance with the required current values of F0, A, T. The report describes a set of Polish language-specific resources utilized by TTS and compared with Russian. It presents linguistic resources (both vocabulary and grammar), multi-voice acoustic databases and a set of specific rules for word-stress position marking, letter-to-phoneme conversion and others.

3 Characteristic properties of Russian and Polish texts processing.

The first operation performed by the textual processor is text preprocessing, i.e. transforming of symbols, numbers, abbreviations and shortenings.

To transform symbols into text the symbols transformation list is used. A part of the list for Polish and Russian languages is shown in Table 3.1.

Symbol	Polish text	Russian text
#	Hasz	Решетка
&	Oraz	Амперсанд
@	Małpa	Собачка
*	pomnóż razy	Умножить на

Table 3.1. Symbols transformation list for Polish and Russian languages

To transform numbers into numerals a special transformation list is created, as it is shown in Table 3.2.

Number	Numeral	Number	Numeral
0	Zero	19	Dziewiętnaście
1	jeden	20	Dwadzieścia
2	dwa	30	Trzydzieści
3	Trzy	40	Czterdzieści
4	Cztery	50	pięćdziesiąt
5	Pięć	60	sześćdziesiąt
6	Sześć	70	siedemdziesiąt
7	Siedem	80	osiemdziesiąt
8	Osiem	90	dziewięćdziesiąt
9	Dziewięć	100	sto
10	Dziesięć	200	dwieście
11	Jedenaście	300	trzysta
12	Dwanaście	400	czterysta

13	Trzynaście	500	pięćset
14	Czternaście	600	sześćset
15	Piętnaście	700	siedemset
16	Szesnaście	800	osiemset
17	Siedemnaście	900	dziewięćset
18	Osiemnaście	1000	Tysiąc

Table 3.2. Numbers to numerals transformation list for Polish language.

When transforming numbers including thousands and millions the case of 'thousand' and 'million' words is taken into account. The requisite case for thousands depends on the last digit and can be defined using the following rules:
1 – *tysiąc*,
2,3,4 – *tysiące*,
5,6,7,8,9 – *tysięcy*.
The requisite case for millions can be defined using the rules below:
1 – *milion*,
2,3,4 – *miliony*,
5 and more – *milionów*.
To transform shortenings and abbreviations into a text a shortenings transformation list is used. Parts of the lists for Polish and Russian languages are shown in Table 3.3 and Table 3.4 correspondingly.

Abbreviation and shortenings	**Text representation with stress marks**	**Abbreviation expansion**
Au	zło+to	Złoto
MSW	emeswu+	Ministerstwo Spraw Wewnętrznych
RP	erpe+	Rzeczpospolita Polska
UW	uwu+	Uniwersytet Warshawski
Zł	zło+ty	Złoty

Table 3.3. Shortenings transformation list for Polish language.

Abbreviation and shortenings	**Text representation with stress marks**	**abbreviation expansion**
и т.д.	итэдэ+	и так далее
МВД	эмвэдэ+	Министерство внутренних дел
пос.	посе+лок	поселок
РБ	эрбэ+	Республика Беларусь
см.	смотри+	смотри

Table 3.4. Shortenings transformation list for Russian language.

When the abbreviation or shortening from the text is found in the list, it is transformed into text representation with stress mark or abbreviation expansion, depending on transformation settings.
If the abbreviation is not contained in the list, it is spelled out according to the language-characteristic rules.
The next step of the textual processor is to place a word stress.
The word stress position in Russian language is independent and there are no rules ensuring a correct word stress placing in every situation. The word stress database for Russian language was created. It is a word storage, where every word is associated with its stress position and grammar context.
The word stress in Polish language is usually on the penultimate though there are some exceptions:
1. The fourth from the end of a word syllable can be marked stressed. This occurs in first and second person conditional mood verbal forms, e.g. *pojecha+libyśmy* (we would go), *da+libyście* (you would give)
2. The stress can lie on a third from the end of the word syllable. This occurs in the following cases:
- some Polish nouns, e.g. *oko+lica*, *rzeczpospo+lita*;
- some foreign nouns and their derivatives, e.g. *dże+ntelmen* (gentleman), *wu+nderkind* (child prodigy), *indywi+duum* (individual), *grama+tyca* (grammar), *hipe+rbola* (hyperbola);
- some noun forms, e.g. *o+gólem* (generally), *szcze+góły* (details) ;
- some numeral forms, e.g. *czte+rysta* (four hundred), *sie+demset* (seven hundred);
- some past tense and conditional mood forms, e.g. *by+liśmy* (we were), *mie+liby* (they would have);
- tetrasyllable forms of the adjective 'powinien' (guilty) : *powi+nniśmy, powi+nnyśmy, powi+nniście, powi+nnyście*;
- some loan proper nouns and geographical names: *Napo+leon, Ha+nnibal, Wa+szyngton*.

3. The last syllable can be stressed. This occurs in the next cases:
- some loan words, mainly of French origin: *atelie+r* (studio)*, menu+* (menu)*, turnee+* (tour);
- some compound words, consisting partly of roots *arcy, eks, wice*: *wicekro+l* (viceroy), *eksmąż* (ex-husband), *arcyło+tr* (ultra- villain).

With due regard of these rules and exceptions the word stress database for Polish language was created, where all

the words-exceptions were included. Every word in the database is associated with its stress position.

After a word is extracted from the text, the textual processor looks for it in words-exceptions stress database and places the stress, or, if it is a regular stress word, places the stress to a penultimate.
The last task of textual processor is phonemic transcribing of the text. This task is language phonetic system dependent, and it differs for the Russian and Polish languages.

4. Characteristic properties of Polish phonetic system.

The Polish alphabet consists of 32 letters. There are 23 regular letters among them: *A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, R, S, T, U, W, Y, Z*, and 9 additional: *Ó, Ś, Ż, Ź, Ć, Ń, Ą, Ł, Ę*. More over, three additional letters: Q, V, X appear in loan words, for example *quorum, veto, xero*.

For phonemes reproduction besides single letters the following digraphs are used: *Cz, Dz, Dż, Dź, Rz, Sz, Ch*.

There are 51 phonemes in the Polish language. The phonetic set consists of 8 vowels and 43 consonants. The vowels and their pronunciation criterion are the following:

U **–** velar, high, labial;
O **–** velar, low, labial;
A **–** velar, low;
Ą **–** velar, low, labial, nasal;
Ę **–** dorsal, low, nasal;
Y **–** dorsal, high;
E **–** dorsal, low;
I **–** blade, high.

The consonants with different pronunciation criterion, their transcription, and corresponding Russian consonants, if exist, are shown in Table 4.1.

		stops		**affricates**		**fricatives**		**sonorants**			
		unvoi-ced	**voiced**	**unvoi-ced**	**voiced**	**unvoi-ced**	**voiced**	**laterals**	**nasals**	**vibrants**	**glades**
velar	**soft**	***K'*** [k'] K'	***G'*** [g'] G'			***H'*** [h'] H'					***J'*** [j'] J'
	hard	***K*** [k] K	***G*** [g] G			***H*** [h] H					
dorsal	**soft**			***Ć*** [ch'] -		***Ś*** [sh'] -	***Ź*** [zh'] -				
	hard			***Cz*** [ch] -	***Dż*** [dzh] -	***Sz*** [sh] sh	***Ż*** [zh] zh				
blade	**soft**	***T'*** [t'] T'	***D'*** [d'] D'	***C'*** [c'] -	***Dź*** [dz'] -	***S'*** [s'] S'	***Z'*** [z'] Z'	***L'*** [l'] L'	***N'*** [n'] N'	***R'*** [r'] R'	
	hard	***T*** [t] T	***D*** [d] D	***C*** [c] C	***Dz*** [dz] -	***S*** [s] S	***Z*** [z] Z	***L*** [l] L	***N*** [n] N	***R*** [r] R	
labial	**soft**	***P'*** [p'] P'	***B'*** [b'] B'			***F'*** [f'] F'	***W'*** [v'] V'		***M'*** [m'] M'		
	hard	***P*** [p] P	***B*** [b] B			***F*** [f] F	***W*** [v] V		***M*** [m] M		***Ł*** [w] -

Table 4.1. Consonants of Polish language, their transcription, Russian adequacy

For the purpose of the text phonemic transcribing the sequences of letters, phonemes and the additional sequence were specified, as shown in Tables 4.2, 4.3, 4.4.

For letter-phoneme correspondence some phonemes occur several times in the sequence.
The rules for letter-to-phoneme transformation are shown in Table 4.5, where in left table column the values of

item number i of phoneme sequences are shown, in the upper table line the values of item number i of letter sequences are shown. The conditions of letter-to-phoneme conversion are shown in corresponding column and line intersection, where y stands for the next after analyzed symbol of text.

Sequence appellation	**Sequence content**
vowels LV	A, Ą, E, Ę, I, O, Ó, U, Y
unpaired consonants LUP1	J, Ń, Ł
unpaired consonants LUP2	L, M, N, R, H
paired consonants LP	(B,P); (DZ,C); (D,T); (W,F); (G,K); (Z,S); (Ź,Ś); (Ż,SZ); (DŹ,Ć); (DŻ,CZ)
voiced consonants LV	B, DZ, D, W, G, Z, Ź, Ż, DŹ, DŻ
unvoiced consonants LUV	P, C, T, F, K, S, Ś, SZ, Ć, CZ
consonants combination LS	(CH,RZ); (K,RZ); (T,RZ); (P,RZ); (T,RZ); (T,W); (K,W); (S,W); (SZ,W); (Ś,W);(CH,W); (C,W); (CZ,W); (Ć,W)

Table 4.2. Sequences of letters for Polish language

Sequence appellation	**Sequence content**
Ph1	A, E, I, Y, O, U, O", O", E"
Ph2	J', N', L'
Ph3	L, M, N, R, X
Ph4	L', M', N', R', X'
Ph5	P, C, T, F, K, S, S', SH, C', CH
Ph6	B, DZ, D, V, G, Z, Z', ZH, DZ', DZH
Ph7	B', DZ', D', W', G', Z', Z', ZH', DZ', DZH'
Ph8	P', C', T', F', K', S', SH', C', CH'
Ph9	B, DZ, D, W, G, Z, Z, ZH, DZ, DZH
Ph10	P, C, T, F, K, S, S, SH, C, CH
Ph11	(XRZ'); (KRZ'); (TRV'); (PRV'); (TRZ'); (TV'); (KV'); (SV'); (SHV'); (S'V'); (XV'); (CV'); (CHV'); (C'V')
Ph12	(XRZ); (KRZ); (TRV); (PRV); (TRZ); (TV); (KV); (SV); (SHV); (S'V); (XV); (CV); (CHV); (C'V)

Table 4.3. Sequences of phonemes for Polish language

Sequence appellation	**Sequence content**
pause and signs S	. , ? ! ; ' " ( ) [ ] <space>

Table 4.4. Additional sequence for phonemic transcribing of text

	LV(i)	LUP1(i)	LUP2(i)	LP(i)	LUV(i)	LS(i)	LS(i)
Ph1(i)	ever						
Ph 2(i)		ever					
Ph 3(i)			y = "I"				
Ph 4(i)			y ≠ "I"				
Ph 5(i)				y ∈ S || y ∈ LUV			
Ph 6(i)				y ∈ LV			
Ph 7(i)	y = "I"						
Ph 8(i)					y = "I"		
Ph 9(i)	y ∈ LV& y ≠ "I"						
Ph 10(i)					y ∈ LV& y ≠ I		
Ph11(i)						y = "I"	
Ph12(i)							y ≠ "I"

Table 4.5. Letter-to-phoneme transformation rules for Polish language.

5. Forming of prosodic text characteristics for Polish language.

To mark out accentual units (AU) the sets of enclitics and proclitics are used. The enclitics in the Polish language are:

- monosyllable words after ***nie*** particle, f.e. *nie+ wiem, nie+ jedz*;
- monosyllable nouns and pronouns after preposition: *do+ mnie, na+ wsi, coś za+ coś*;
- short forms of personal pronouns ***cię, go, ich, im, ją, je, jej, mi, mu, nas, was***, e.g. *zapro+śmy ich, mówi+łem jej, dora+dzę mu*;
- ***się, no, by*** particles, e.g. *urzą+dziłbym się, da+j no mi, mo+żna by*.

The proclitics in the Polish language are:

- ***nie*** particle before multisyllable word, e.g. *nie poma+gaj*, *nie ta+ki,nie Ro+man*;
- monosyllable prepositions before multisyllable words, e.g. *pod mia+stem, za wo+lność*;
- monosyllable conjunctions, e.g. *Przydź, jak ze+chcesz*;
- ***co*** pronoun in word-combinations like *co ro+k, co dzie+n*.

The sets of enclitics and proclitics, as well as the attachment rules are described in Polish linguistics DB.

When dividing orthographic text into syntagms the syntax and morphology text structures are taken into account.

First, the bounds of macro-syntagma are determined:

- after punctuation marks;
- before conjunctions.

At the next step macro-syntagmas are divided into micro-syntagmas with the help of the morphology-based algorithm so that finally each micro-syntagma is composed of 4 AU at most.

Each micro-syntagma then is marked with its type, where the types are:

C – pending
P – final
E – exclamatory
Q – interrogative

The prosodic processor splits each incoming AU into elements of AU (EAU): pre-nuclear, nuclear and post-nuclear parts, where the stressed vowel is considered to be the nucleus.

Then every EAU is marked with the values of the amplitude (A), the phoneme duration (T) and the fundamental frequency (F0), which are language-specific and also depend on the EAU type and the type of micro-syntagma. Those values are contained in prosody database in parametrized form.

6. Acoustical processing of allophonic text.

Multi-voices acoustical database based on the methodology, developed at the United Institute of Information Problems of the National Academy of Sciences of Belarus (UIIP NASB) for 'cloning' of personal voice and phonetic peculiarities [4] was created. It contains a set of pitch marked allophone sound waves. The incoming allophonic text is used to concatenate portions of allophone waves and thus form monotonic speech signal.

To distinguish the inflection amplitude (A), phoneme duration (T) and fundamental frequency (F0) of allophones are changed according to the marks placed by prosodic processor.

The change of amplitude is attained by multiplication of speech signal by requested coefficient. The change of duration is attained by changing in number of periods. To change the fundamental frequency of pitch marked allophones the modified PSOLA algorithm is used.

7. Conclusion.

Speech synthesis system described in the article is designed so that all the differences are included into databases: linguistics DB, phonetics DB, prosody DB and acoustical DB, specified for Russian and Polish languages. Designed in this way, the speech synthesis system is flexible and language-independent.

The created system can be used as a synthesizer for some other Slavonic language on the stipulation that linguistic, phonetic, inflection and acoustical DB for the new synthesizing language will be added.

8. References.

1. Lobanov B. *Multi-Voice Text-to-Speech Synthesis and Large-Vocabulary Spoken Words Recognition for Slavonic Languages: Belarusian, Polish, Russian and Ukrainian***.** // Proceedings of the 6th International Workshop “Speech and Computer” – SPECOM’2003. Moscow, Russia, 2003, pp.123-128.
2. Lobanov B., Karnevskaya H. *MW Speech Synthesis from Text*. Proc. of the XII International Congress of Phonetic Sciences. Aix-en-Provense, France, 1991, pp. 406-409.
3. Boguslavsky I., Lobanov B. and Karnevskaya H**.** *Generation of Intonation and Accentuation of Synthetic Speech on the Base of Morpho-Syntactic Knowledge*, Proceedings of the International Workshop “Integration of Language and Speech”, Moscow, 1996, pp. 11-28.
4. Lobanob B. and Karnevskaya H**.** *TTS-Synthesizer as a Computer Means for Personal Voice (On the example of Russian).* In the Book: Phonetics and its Applications. Stuttgart: Steiner. 2002, pp. 445-452.