Piotr KŁOSOWSKI

Silesian University of Technology

Improving speech processing based on phonetics and phonology of Polish language

Abstract. *The article presents methods of improving speech processing based on phonetics and phonology of Polish language. The new presented method for speech recognition was based on detection of distinctive acoustic parameters of phonemes in Polish language. Distinctivity has been assumed as the most important selection of parameters, which have represented objects from recognized classes. Speech recognition is widely used in telecommunications applications.*

Streszczenie. *W artykule zaprezentowano metody usprawnienia przetwarzania mowy wykorzystując do tego celu wiedzę z zakresu fonetyki I fonologii języka polskiego. Przedstawiona innowacyjna metoda automatycznego rozpoznawania mowy polega na detekcji akustycznych parametrów dystynktywnych fonemów mowy polskiej. O dystynktywności cech decydują parametry niezbędne do klasyfikacji fonemów. (**Usprawnienie przetwarzania mowy w oparciu o fonetykę i fonologię języka polskiego)*.

Keywords: speech processing, speech recognition, speech synthesis.
Słowa kluczowe: przetwarzanie mowy, rozpoznawanie mowy, synteza mowy.

Introduction

Division of Telecommunication, a part of the Institute of Electronics Silesian University of Technology, for many years specializes in advanced fields of telecommunication engineering [1-6]. One of them is speech processing applications [7-9]. Main research areas on this field are: speech synthesis, speech recognition and speaker verification and identification systems. Typical speech processing applications are presented on Figure 1 [10].

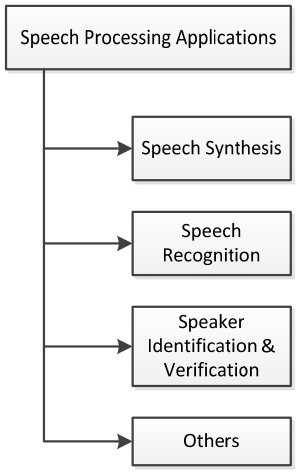

Fig.1. Speech processing application

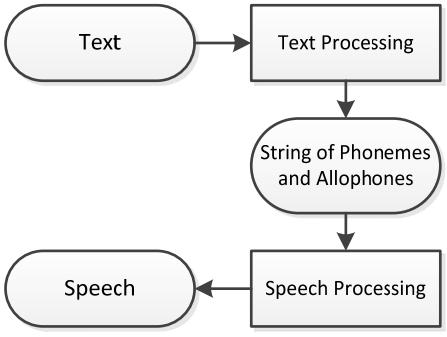

Fig.2. Structure of text-to-speech synthesis system

Results of research on speech synthesis

At present, the speech synthesis is widely used in many applications, the first of all in telecommunication [11-13]. In Institute of Electronics Silesian University of Technology ware developed two generation of speech synthesizer for polish based on TTS (Text to Speech) technology. Structure of TTS synthesis system is presented on Figure 2.

The full TTS system converts an arbitrary ASCII text to speech. The first task of the system is to extract the phonetic components of the required message realized in text processing unit shown in Figure 3.

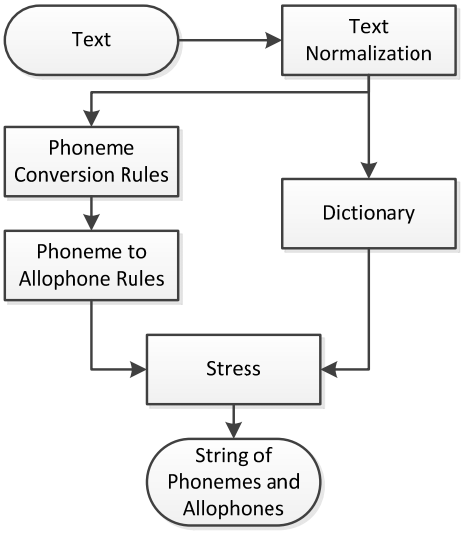

Fig.3. Structure of text processing unit

The output of this stage is a string of symbols representing sound-units (phonemes or allophones), boundaries between words, phrases and sentences along with a set of prosody markers (indicating the speed of speech, the intonation etc.). The second part of the process is to match the sequence of symbols up with items stored in the phonetic inventory, link them together and send to a voice output device. This task is realized in speech processing unit shown on Figure 4.

A combination of linguistic analysis must be done in the first stage which involves: converting abbreviations and special symbols (decimal points, plus, minus, etc.) to spoken form.

On Institute of Electronics, Silesian University of Technology was developed two generation of text-to-speech synthesis system for polish. The first system was created to simulate the human vocal tract, dedicated for blind persons. System allows proper word pronunciation

and word stress by means of full phoneme transcription. Speech synthesis was made on the phoneme level. The next one is based on allophonic speech synthesis level. Allophonic speech synthesis quality is better than quality of phoneme speech synthesizer. This software provides natural-sounding, highly intelligible text-to-speech synthesis.

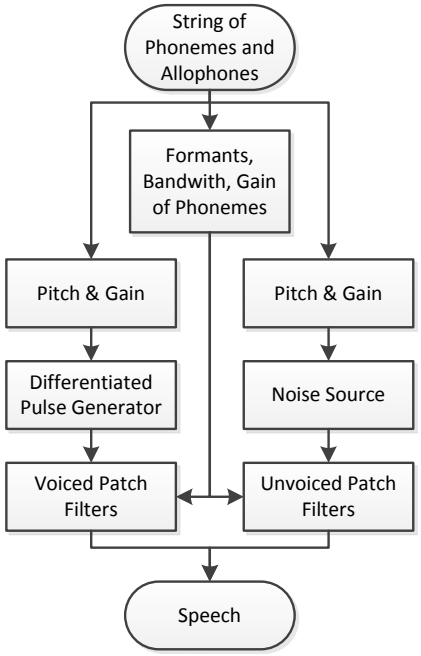

Fig.4. Structure of speech processing unit

Results of research on speech recognition

Speech recognition is a conversion from an acoustic waveform to a written equivalent of the message information [14]. The nature of speech recognition problem is heavily dependent upon the constraints placed on speaker, speaking situation and message context. Speech recognition process is realized in two steps presented on Figure 5

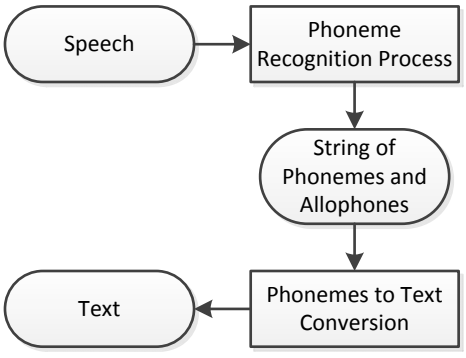

Fig.5. Two steps of speech recognition process

In the first step speech signal is processed by phonemes recognition system. The result of this process is sequence of phonemes or allophones. This sequence is processed by phonemes to text conversion unit with elements of speech understanding system. Final result is this process is text. Detailed structure of speech recognition process is shown on Figure 6.

The second major of research in speech communication applications is speech recognition and particularly improving speech recognition process of polish language using linguistic knowledge (phonetics and phonology) [15]. This idea is presented on Figure 7.

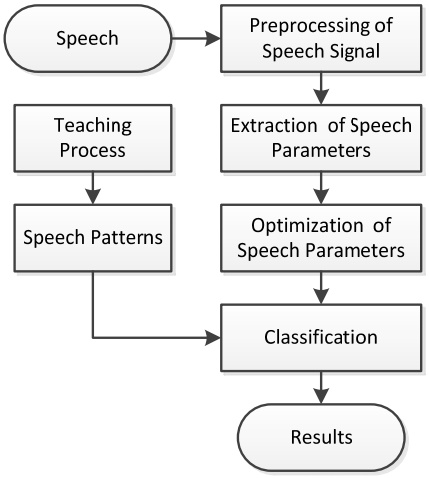

Fig.6. Detailed structure of speech recognition process

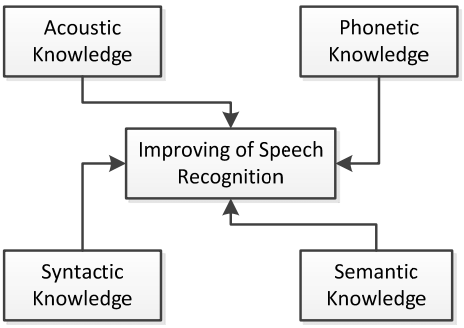

Fig.7. Improving speech recognition process by using language knowledge

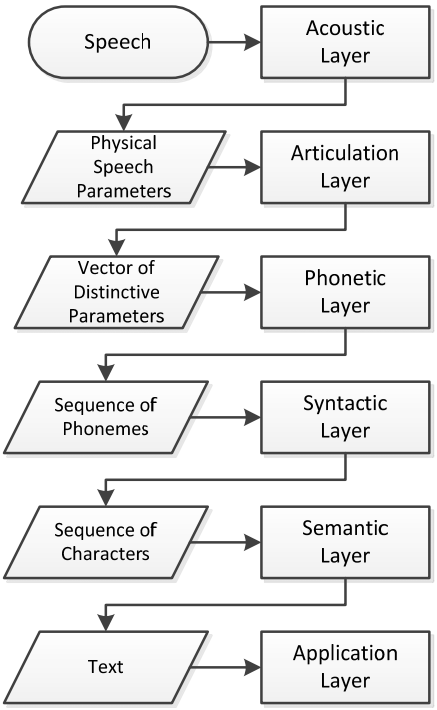

Fig.8. Multilayer speech recognition system with elements of speech understanding

Improving speech recognition process is realized by using acoustic, phonetic, syntactic and semantic knowledge of polish language. Result of this research was creation of multilayer speech recognition system. Each layer realizes one step of speech recognition process. There are: acoustic layer, articulation layer, phonetic layer, syntactic layer, semantic layer and application layer. Model of multilayer speech recognition system is shown on Figure 8.

The first acoustic layer provides physical parameters of speech. Second articulation layer provides vectors of distinctive parameters of speech. Phonetic layer on the basics of these vectors generates sequence of speech phonemes. Syntactic layer using dictionary of pronunciation rules provide orthographical notation of speech. Semantic layer establishes of meaning orthographical sequents of characters and provides sentences in polish language. Task of application layer depends on destination of speech recognition system. The new used method for speech recognition was based on detection of distinctive acoustic parameters of phonemes in polish language. Distinctivity has been assumed as a most important selection of parameters which have represented objects from recognized classes of phonemes.

Improving recognition process of polish phonemes

Phonemes are sound units determined meaning of words. Effective phonemes recognition, sound units of each language allow to effective recognizing continuous speech. Improving phonemes recognition process is possible using phonetics and phonology of polish language [16]. The new method of speech recognition was based on detection of distinctive acoustic parameters of phonemes in polish language. Each phoneme is specific by vector of distinctive parameters of speech signal. This vector has the form:

$$X = [SA, MA, CD, SW] \tag{1}$$

where: SA – class of phoneme, MA – place of phoneme articulation feature, CD – additional feature of phoneme, SW – method of articulation.

Table 1. The values of distinctive features and their meanings

Distinctive feature	Description	Value	Meaning
SA	Class of phoneme	SA=1	stop/affricate phoneme
		SA=2	fricative phoneme
		SA=3	semi-fricative phoneme
		SA=4	vowel phoneme
		SA=5	nasal phoneme
		SA=6	retroflex phoneme
		SA=7	lateral phoneme
		SA=8	semi-vowel phoneme
MA	place of phoneme articulation	MA=1	labial phoneme
		MA=2	labial-dental phoneme
		MA=3	dental phoneme
		MA=4	palatal phoneme
		MA=5	front part of tongue in articulation
		MA=6	middle part of tongue in articulation
		MA=7	rear part of tongue in articulation
CD	additional feature of phoneme	CD=0	no feature
		CD=1	high position of tongue in articulation
		CD=2	low position of tongue in articulation
SW	method of articulation	SW=0	no feature
		SW=1	voiced phoneme
		SW=2	unvoiced phoneme

Table 2. Set of distinctive parameters of polish phonemes with articulation probability and number of distinctive parameters required to recognize each phoneme

k	Polish phoneme	Probability of k-th phoneme articulation	Vector of distinctive parameters	Number of distinctive parameters required to recognize k-th phoneme
1	b	0.013	[1,1,0,1]	3
2	p	0.027	[1,1,0,2]	3
3	d	0.019	[1,5,0,1]	3
4	t	0.038	[1,5,0,2]	3
5	ǵ	0.001	[1,6,0,1]	3
6	ḱ	0.006	[1,6,0,2]	3
7	g	0.013	[1,7,0,1]	3
8	k	0.023	[1,7,0,2]	3
9	v	0.030	[2,2,0,1]	3
10	f	0.013	[2,2,0,2]	3
11	z	0.015	[2,3,0,1]	3
12	s	0.026	[2,3,0,2]	3
13	ž	0.010	[2,4,0,1]	3
14	š	0.017	[2,4,0,2]	3
15	ź	0.002	[2,6,0,1]	3
16	ś	0.013	[2,6,0,2]	3
17	χ	0.009	[2,7,0,0]	2
18	ʒ	0.007	[3,3,0,1]	3
19	c	0.013	[3,3,0,2]	3
20	ǯ	0.0005	[3,4,0,1]	3
21	č	0.010	[3,4,0,2]	3
22	ʒ́	0.002	[3,6,0,1]	3
23	ć	0.011	[3,6,0,2]	3
24	i	0.034	[4,5,1,0]	3
25	y	0.035	[4,5,2,0]	3
26	e	0.088	[4,6,1,0]	3
27	a	0.080	[4,6,2,0]	3
28	u	0.029	[4,7,1,0]	3
29	o	0.078	[4,7,2,0]	3
30	m	0.030	[5,1,0,0]	2
31	n	0.034	[5,5,0,0]	2
32	ń	0.022	[5,6,0,0]	2
33	ŋ	0.007	[5,7,0,0]	2
34	r	0.007	[6,0,0,0]	1
35	l	0.018	[7,0,0,0]	1
36	j	0.039	[8,5,0,0]	2
37	u̯	0.019	[8,7,0,0]	2

The first distinctive parameter means class of phoneme. Second means place of phoneme articulation. Last two parameters are: additional feature and method of phoneme articulation. The values of distinctive features and their meaning are presented on Table 1.

For the physical parameters of the speech signal used in the extraction of distinctive features include:

- formants frequency,
- relative amplitudes of formants,
- anti-formants frequency,
- fundamental frequency of laryngeal tone F0,
- waveform amplitude envelope or energy,
- extracted waveforms in the frequency domain parameters such as fundamental frequency of laryngeal tone and trajectory of formants,
- spectrograms and sonograms of the speech signal.

Average number of distinctive parameters required to recognize one phoneme equals 2.71, and was estimate using formula (2):

(2) $$N = \sum_{k=1}^{M} p_k \cdot N_k = \sum_{k=1}^{37} p_k \cdot N_k = 2{,}71$$

where: N – average number of distinctive parameters of speech, M – number of phonemes, p_k – probability of k-th – phoneme articulation , N_k – number of distinctive parameters required to recognize k-th phoneme.
Set of polish phonemes is presented on Table 2 and Table 3 presents set of distinctive parameters of Polish phonemes with articulation probability.

Examples of language processing in Polish

We can distinguish two very important tasks in Polish language processing:

- text-to-phoneme and phoneme-to-speech conversions in speech synthesis and
- speech-to-phoneme and phoneme-to-text conversion in speech recognition process.

The letter-to-phoneme conversion changes ASCII text sequences to phoneme sequences. The phoneme-to-letter conversion performs reverse operations. It is based on implementation and employment of rule-based system and the dictionary for exceptions. This is very crucial fragment of the code within the entire speech processing software. Pronunciation of Polish language words is not very complicated.

Table 3. Set of polish phonemes

Nr	Phoneme	Example of word	Nr	Phoneme	Example of word
1	i	wi**c**i	20	ž	k**a**żdy
2	y	s**y**ty	21	ś	**si**ano
3	e	s**e**rce	22	ź	**zi**arno
4	a	b**a**ba	23	χ	**h**igiena
5	o	**o**ko	24	p	**p**raca
6	u	b**u**k	25	b	**b**aba
7	i̯	**j**ajo	26	t	**t**rawa
8	u̯	**ł**ysy	27	d	**d**udek
9	r	**r**ok	28	k	**k**ot
10	l	**l**ato	29	g	mo**g**ą
11	m	**m**a**m**a	30	ḱ	**ki**no
12	n	**n**oc	31	ǵ	ma**gi**czny
13	ń	ko**ń**	32	c	**c**a**c**ko
14	ŋ	rę**k**a	33	ʒ	na**dz**y
15	f	**f**ala	34	č	**cz**arny
16	v	**w**ada	35	ǯ	dro**żdż**e
17	s	**s**ok	36	ć	**ci**asto
18	z	ko**z**a	37	ʒ́	**dzi**e**dzi**c
19	š	**sz**y**sz**ka			

Even though the letter-to-phoneme conversion has more than 90 pronunciation rules, which requires an exception dictionary. Each phoneme is actually represented by a structure that contains a phonemic symbol and phonemic attributes that include duration, stress, and other proprietary tags that control phoneme synthesis.

This scheme is used for handling allophonic variations of a phoneme. The term phoneme refers either to this structure or to the particular phone specified by the phonemic symbol in this structure. Figure 9 and Figure 10 present examples of this process.

Summary

The research on speech recognition is continued. At present efforts concentrate in creation efficient speech recognition system based on multilayer speech recognition model using distinctive parameters of speech. The second major of effort is creation speaker verification and identification system and implement some speaker identification algorithms in speech recognition system. Future goal of research is construction of full speech dialog system with elements speech understanding based on artificial intelligence technology.

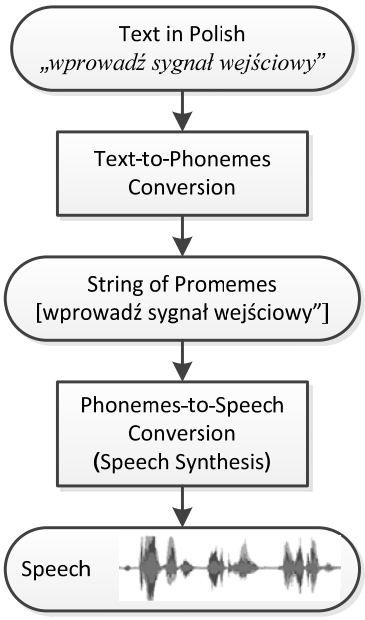

Fig. 9. Example of Text-to-Speech Conversion

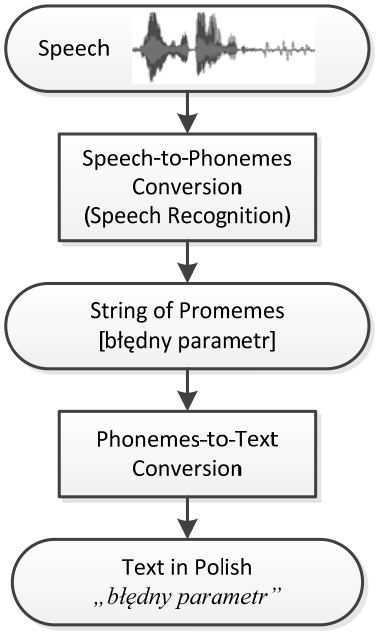

Fig. 10. Example of Speech-to-Text Conversion

Knowledge and experience in the field of speech processing can also be used in development of automatic speech translation systems. Automatic speech translation is the process by which conversational spoken phrases are instantly translated and spoken aloud in a second language. It is a technology enables speakers of different languages to communicate [17-19]. Automatic speech translation systems can play a critical role on empowering people to communicate with speakers of a different language and to access or present information in a cross-lingual way.

Acknowledgements

This work was supported by The National Centre for Research and Development (www.ncbir.gov.pl) under Grant number POIG.01.03.01-24-107/12 (Opracowanie innowacyjnej metody identyfikacji mówcy dla podniesienia stopnia bezpieczeństwa systemów teleinformatycznych).

REFERENCES

Dziwoki G., Analiza nienadzorowanej korekcji kata fazowego w systemach z modulacją kwadraturową (An analysis of the unsupervised phase correction method in quadrature amplitude modulation systems), Przegląd Elektrotechniczny, vol.88, nr 7a, 2012, ss.245–249.

[1] Izydorczyk J., Izydorczyk M., Limits to microprocessor scaling, Computer , vol.43, no.8, pp.20-26, Aug. 2010.

[2] Sułek W., Pipeline processing in low-density parity-check codes hardware decoder, Bulletin of the Polish Academy of Sciences Technical Sciences, vol. 59, No. 2, 2011, pp. 149–155.

[3] Zawadzki P., Security of ping-pong protocol based on pairs of completely entangled qudits, Quantum Information Processing, vol. 11, No. 6, Dec. 2012, pp. 1419–1430

[4] Dziwoki G., Kucharczyk K., Sułek W., "Transmission over UWB Channels with OFDM System using LDPC Coding", Conference on Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments, Proceedings of SPIE, Vol. 7502, pp. 75021Q, 2009.

[5] Kucharczyk M., "Blind Signatures in Electronic Voting Systems", 17th International Conference Computer Networks, CCIS Vol. 79, pp. 349-358, Springer-Verlag, Berlin, Heidelberg, 2010.

[6] Dustor A., Speaker verification based on fuzzy classifier, International Conference on Man-Machine Interactions (ICMMI 2009), September 25-27, 2009. Chapter in book: Cyran K.A., Kozielski S., Peters J.F., Stańczyk U., Wakulicz-Deja A. (Eds.): Man-Machine Interactions, AISC 59, Springer-Verlag, Vol. 59, pp. 389–397, Berlin Heidelberg, 2009.

[7] Kłosowski P., Speech Processing Application Based on Phonetics and Phonology of the Polish Language, Proceedings of 17th International Conference of Computer Networks 2010, Ustroń, Poland, June 2010, Communications In Computer and Information Science, Springer-Verlag, Germany 2010, ISBN 1865-0929, pp.236–244.

[8] Kłosowski P., Pułka A., Polish Semantic Speech Recognition Expert System Supporting Electronic Design System , Prooccedings of The International Conference on Human Systems Interactions. HSI 2008, Kraków, Poland, May 25-27, IEEE Book Series: Eurographics Technical Report Series, 2008, pp. 479–484.

[9] Huang X., Acero A., Hon H.W., Spoken Language Processing. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[10] Rabiner L.R., Applications of Voice Processing to Telecommunications Proc. of the IEEE, vol.82, No.2, pp. 197-228 , Feb. 1994.

[11] Kłosowski P. Speech Communication Applications, The First Seminar – University of Central Florida and The Silesian University of Technology, Niedzica Castle, Poland 2002

[12] Kłosowski P. Speech Communication Applications, International Conference Programmable Devices and Systems PDS 2003 IFAC Workshop, Ostrava 2003, pp. 332-337.

[13] Rabiner L. R., Juang B. H., Fundamentals of speech recognition. Prentice Hall, 1993.

[14] Ostaszewska D.,Tambor J.,Podstawowe wiadomości z fonetyki i fonologii współczesnego języka polskiego Uniwercity of Silesia nr 488,Katowice 1993,(in Polish).

[15] Kłosowski P., Izydorczyk J. Base Acoustic Properties of Polish Speech, International Conference Programmable Devices and Systems PDS2001 IFAC Workshop, Gliwice 2001.

[16] Stuker, S.; Herrmann, T.; Kolss, M.; Niehues, J.; Wolfel, M., Research Opportunities In Automatic Speech-To-Speech Translation, *Potentials* , IEEE Volume: 31 , Issue: 3, 2012 pp. 26-33.

[17] Waibel A. , Fügen C., Spoken language translation—enabling crosslingual human-human communication, IEEE Signal Processing Mag., vol. 25, no. 3, pp. 70–79, May 2008.

[18] Hutchins J., International Association for Machine Translation compendium of translation software, 2010. [Online]. Available: http://www.hutchinsweb.me.uk/Compendium.htm

***Author**: dr inż. Piotr Kłosowski, Politechnika Śląska, Instytut Elektroniki, ul. Akademicka 16, 44-100 Gliwice, E-mail: pklosowski@polsl.pl.*