PHONETIC ALPHABET FOR SPEECH RECOGNITION OF CZECH

Jan NOUZA[1], Josef PSUTKA[2], Jan UHLÍŘ[3]
[1]*Department of Electronics and Signal Processing, Technical University of Liberec, Hálkova 5, 461 17 Liberec, Czech republic*
[2]*Department of Cybernetics, University of West Bohemia, Univerzitní 22, 306 14 Plzeň, Czech republic*
[3]*Department of Circuit Theory, Czech Technical University of Prague, Technická 2, 166 27 Praha 6, Czech republic*

Abstract

In the paper we introduce and discuss an alphabet that has been proposed for phonemicly oriented automatic speech recognition. The alphabet, denoted as a PAC (Phonetic Alphabet for Czech) consists of 48 basic symbols that allow for distinguishing all major events occurring in spoken Czech language. The symbols can be used both for phonetic transcription of Czech texts as well as for labeling recorded speech signals. From practical reasons, the alphabet occurs in two versions; one utilizes Czech native characters and the other employs symbols similar to those used for English in the DARPA and NIST alphabets.

Keywords

speech recognition, phonetic transcription, speech signal labeling.

1. Introduction

Only simple speech recognition systems, namely those oriented on isolated words belonging to small or medium-size vocabularies, can suffice with a single, word based level. The systems that are to operate with larger vocabularies (with more than several hundred items) and/or with continuous speech must adopt another strategy; a strategy that allows for building up words as well as phrases or sentences from a small set of basic speech units. These units should fulfill two principal demands: 1) to correspond with the textual form of a language, and 2) to have a close relation to the acoustic signal of speech. In most of the state-of-art systems, these basic units are phonemes (or phoneme-based composites, such as diphones or triphones). The phonemes are language specific units and thus each language needs a declaration of its own phonetic alphabet.

The Czech language has traditionally had its own phonetic inventory that is described in details, for example, in [1]. However, the standard inventory as it is used in phonetics and phonology, may not be always optimal for automatic speech recognition (ASR) tasks. The reasons are at least twofold. First, the definitions of the phonemes mostly follow articularly based attributes of speech rather than acoustic characteristics of speech sounds. Second, a lot of the symbols used in phonetics are not well suited for the use in computer environment. In many languages, the latter problem has been already solved by transforming the established International Phonetic Alphabet (IPA, [2]) into alphabets that are „more friendly" to a computer. It is the case, for example, of the SAMPA alphabet [3] developed for the languages of the European Union within the ESPRIT project and DARPA or NIST alphabets [4] used mainly for American English in the USA.

According to the authors' knowledge, there has been no published work that would define a similar alphabet suitable for the ASR in Czech, up to now. In order to overcome this gap we have decided to introduce the Phonetic Alphabet for Czech (PAC). Our primary aim was to establish a common base (a norm) for the further research in speech recognition in the Czech Republic. The secondary goal of this work is to present the phonetic inventory of Czech to non-native researchers.

The layout of the paper is as follows. In the next section we try to summarize most of the requirements that an ASR alphabet should meet. The complete definition of the PAC together with a brief discussion is presented in section 3. Some technical and application remarks are mentioned in section 4.

2. Requirements on the phonetic alphabet

Any alphabet designed for the ASR purposes must fulfill one essential demand. It should provide a recognition system by elements and symbols that allow to link and match - as precisely as possible - the textual and acoustic forms of spoken language. The units of the alphabet must meet several practical requirements. They should be:

1.*Consistent*, so that different instances of a unit have similar characteristics.

2.*Economic*, to make the whole inventory as small as possible.

3.*Comprehensive*, so that most of the contextual effects are contained within the units.

4.*Suitable for concatenation,* so that larger language elements such as words or sentences can be composed and recognized as a sequence of the units.

5.*Applicable for computer processing,* so that the alphabet symbols can be easily entered and processed.

6.*Applicable for real acoustic signal,* so that also non-speech acoustic events can be handled.

In most of the languages it is the phonemes that fulfill the basic requirements 1) to 4). The set of internationally recognized phonemic symbols is known as the International Phonetic Alphabet (IPA). Its most recent version has been published in 1992 [2]. Unfortunately, the IPA does not comply with the last two requirements because it uses many special symbols not available on common computers. That is why several computer-oriented versions of the IPA have been introduced. In many EU countries, the SAMPA (SAM Phonetic Alphabet, created within the Speech Assessment Methods project) has been widely used recently. In USA and some other countries, DARPA and NIST alphabets are used for speech labeling. There are some differences in the EU and US approaches. While the SAMPA prefers a single symbol for a phoneme (including digits and non-letter characters), both the US alphabets, that are very similar each to other, utilize symbols composed from one to three letters.

None of the above mentioned alphabets is directly applicable to Czech. It is because some sounds that are specific for Czech (not only the well-known „ř" but also some others, e.g „d", „ť", „ň") are not included there. That is why it is necessary to define a Czech phonetic alphabet. It should meet all the above requirements together with the following demands:

a) Full coverage of all Czech phonemes as they are declared in [1].

b) Easy legibility of transcribed texts, so that even non-specialists can read and check the texts and labels.

c) Alternative transcription in DARPA style that will allow for compatibility with existing speech processing systems and that will give also foreigners a chance to read transcribed Czech texts.

d) Consistency with the DARPA/NIST alphabet, so that the PAC alternative symbols and original symbols of the US alphabet can be used together without overlapping.

3. PAC - the Phonetic Alphabet for Czech

3.1 The alphabet

The alphabet proposed for automatic speech recognition of Czech is listed in Table 1. It has been named the Phonetic Alphabet for Czech (PAC). It includes all the Czech phonemes declared in [1] together with some auxiliary symbols. The additional symbols have been added to allow for some alternations (e.g. if one wants to handle diphthongs as individual phonemes) and also to cover some non-speech events, which is an essential demand in processing of real speech signals.

In accordance with the above requirements a) to d), the PAC occurs in two versions. One, denoted as PACcz, employs characters of the Czech native alphabet (including the letters with diacritics). As a rule, a phoneme is transcribed by a lower-case 1-letter symbol. In most cases the symbol corresponds with the letter that represents the sound in written Czech, as it can be observed from the first and second column in Table 1. If there is no such letter (e.g. in case of velar „n", labiodental „m" and non-voiced „ř") or if two letters would be needed to characterize one phoneme („ch", „dz" and „dž") we use the symbol of the closest phoneme written in upper-case (e.g. N, M, Ř, X, C and Č). In accord with [1], we consider long vowels as separate phonemes that have been given their own symbols (vowel letters with diacritics). The main reason why the PACcz has been introduced, is its easy use for Czech users. As a matter of fact, a message transcribed by the PACcz still resembles the original text and is thus well-legible (see section 3.3). Moreover, the symbols can be easily remembered since there is a direct correspondence in nearly 90% of all the cases. The transcribed texts are compact, no longer than the original ones and keep the word structure of the texts.

Yet, there are reasons that led us to the introduction of alternative symbols that are referred to as PACal symbols. The first reason is quite practical. The Czech letters with diacritics are not available on some computer systems or within some speech processing packages. The second reason has an international aspect. We consider very useful to have an alphabet that would allow us to transcribe Czech texts into a form understandable for a foreign expert. Since it is the DARPA/NIST alphabet that is worldwide known to most speech experts, our PACal has been derived on that base. Some symbols have been taken directly from the DARPA and many new were added. We strictly kept the rule that the added symbols did not overlap with those in the original DARPA set. It means that both the alphabets could be used together (for example, to transcribe mixed Czech and English texts) without a danger of confusion. The drawbacks of the PACal are evident: more letters (and space) needed for transcription (which may be inconvenient, particularly, in speech signal labeling) and non-compact texts without explicit word boundaries.

3.2 Discussion

The proposed PAC uses 40 phonemes that can be alternatively augmented by 3 diphthongs. The inventory includes also some phonemes that are not so frequent and that are sometimes considered as variants of one phoneme (e.g. „ř" and „Ř" or „m" and „M"). In concrete applications it is possible to merge such cases into one

Table 1: *PAC - Phonetic Alphabet for Czech (both versions) compared with closest DARPA/NIST symbols*

	Phoneme (in plain Czech)	PACcz	PACal	DARPA	NIST	Example
1	„a“	a	a	ah	ah	táta
2	„á“	á	aa			táta
3	„b“	b	b	b	b	bába
4	„c“	c	c			ocel
5	„dz“	C	dz			leckde
6	„č“	č	ch	ch	ch	čichá
7	„dž“	Č	dzh			rádža
8	„d“	d	d	d	d	jeden
9	„ď“	ď	dj			dělat
10	„e“	e	e	eh	eh	lev
11	„é“	é	ee			méně
12	„f“	f	f	f	f	fauna
13	„g“	g	g	g	g	guma
14	„h“	h	h	hh	hh	aha
15	„ch“	X	x			chudý
16	„i“ or „y“	i	i	ih	ih	bil, byl
17	„í“ or „ý“	í	ii	iy	iy	vítr, lýko
18	„j“	j	j	y	y, j	dojat
19	„k“	k	k	k	k	kupec
20	„l“	l	l	l	l	dělá
21	„m“	m	m	m	m	máma
22	„m“	M	mg			tramvaj
23	„n“	n	n	n	n	víno
24	„n“	N	ng	ng	ng	banka
25	„ň“	ň	nj			koně
26	„o“	o	o	oh	oh	kolo
27	„ó“	ó	oo			óda
28	„p“	p	p	p	p	pupen
29	„r“	r	r	r	r	bere
30	„ř“	ř	rzh			moře
31	„ř“	Ř	rsh			keř
32	„s“	s	s	s	s	sud
33	„š“	š	sh	sh	sh	duše
34	„t“	t	t	t	t	dutý
35	„ť“	ť	tj			kutil
36	„u“	u	u	uh	uh	duše
37	„ú“ or „ů“	ú	uu			růže
38	„v“	v	v	v	v	láva
39	„z“	z	z	z	z	koza
40	„ž“	ž	zh	zh	zh	růže
Additional symbols						
41	seam (šev)	_	_			na_uka
42	„ou “	ou	ow	ow	ow	pouze
43	„au“	au	aw	aw	aw	auto
44	„eu“	eu	eu	eu	eu	euforie
45			sil	sil	pau	silence
46			sp	sp	sp	short pause
47		„ „	/			word separator
48		!	!			glottal plossive
49		[event]	[event]	[event]		non-speech event

class provided it does not degrade the recognition system performance. Similarly, also the symbol „!" used for a glottal plosive (hit) occurring before an initial vowel can be considered as an event whose marking may or may not be compulsory.

The PAC includes also three different symbols for separating speech and text flow. The „sil" is supposed to denote mostly „silence" at the beginning and end of a recorded utterance, while the „sp" should be used for annotating shorter pauses between words or within words. The slash „/" is just an auxiliary symbol that has been introduced in the PACal to mark the boundaries between words. (In the PACcz this is achieved by using standard space character as a word delimiter.) The „/" symbol in the PACal and the „ „ in the PACcz are thus corresponding symbols that are usful for automatic translation between both the PAC versions.

The pair of brackets [] is used for identifying and specifying non-speech and noise events (e.g. [cough], [lip_smack], [door_slam]) as they occur in real speech recordings. A detailed specification of the event is left on the annotator.

3.3. Examples of a transcribed text

In order to demonstrate how a transcribed Czech text can look like we have chosen two examples of written and recorded utterances.

An example of a written utterance: „Hledal v trávě ztracený desetihaléř."

Transcription in PACcz: hledal f trávje stracení deseťihaléŘ

Transcription in PACal: / h l e d a l / f / t r aa v j e / s t r a c e n ii / d e s e tj i h a l ee rsh /

2. An example of a spoken utterance: „Otevři dveře."

Transcription in PACcz: !otevři dveře [hluk]

Transcription in PACal: sil / ! o t e v rzh i / d v e rzh e / sil [door_noise]

4. Some remarks to the practical application of the PAC

In speech processing (apart from the recognition itself), the phonetic alphabet is used in two different areas:

phonetic transcription of texts (i.e. vocabularies, sample sentences, etc.),

phonetic labeling of speech recordings used for training purposes.

The first task, the transcription of texts written in Czech, is not so difficult problem, particularly, when compared with other languages (e.g. with English). A program for automatic transcription must take into account some tens of rules that deal mainly with changes in voicing. Most of these rules are described in detailed algorithmic form in [5]. However, since many words used in Czech are of foreign origin, the program must check with a large list of pronunciation exceptions.

The second task solves the problem of correct alignment between a recorded speech signal and its phonetic transcription (see, for example [6]). The sequence of labels (alphabet symbols) is a priori known. What is searched are the time instants when each phoneme starts and ends. This can be accomplished either manually (by visual investigation of the signal and its spectrogram), or with an aid of a segmentation program. Our investigations showed that phoneme boundaries can be successfully tracked by means of spectral variation functions (SVF), i.e. by functions that measure the level of changes in signal spectrum. In [7] we defined an SVF based on the following formula:

$$SVF(t) = \sum_{p=1}^{P} \left(\frac{1}{L} \sum_{i=1}^{L} x_{t-i}^{p} - \frac{1}{L} \sum_{i=1}^{L} x_{t+i}^{p} \right)^2$$

where is the p-th cepstral coefficient estimated in time t, P is the number of the coefficients and L is the length of the SVF window. We demonstrated that more than 80% of the peaks (maxima) in the SVF contour can be used for marking the phoneme boundaries. A program [8] that employs the SVF for speech segmentation and label alignment tasks is shown in Fig.1.

5. Conclusions

The aim of this paper was to define an inventory of elementary speech units that could be applied for phonetic-acoustic analysis of the Czech language. The proposed alphabet, referred to as the PAC (Phonetic Alphabet for Czech), consists of 48 items covering all major phonetic events in spoken Czech. The alphabet symbols have been chosen so that either Czech native or international coding can be used alternatively. The definition of the PAC establishes a common base for further research in automatic speech recognition in the Czech Republic.

Acknowledgment:

The authors wish to thank Daniel Hájek, Vlasta Radová and Václav Hanžl for their fruitfull contribution into the discussion on the topic of the article.

The paper has been prepared under project „Theory and Applications of Speech Communication in Czech" supported by complex grant No.102/96//K087 of the Czech Grant Agency (GAČR).

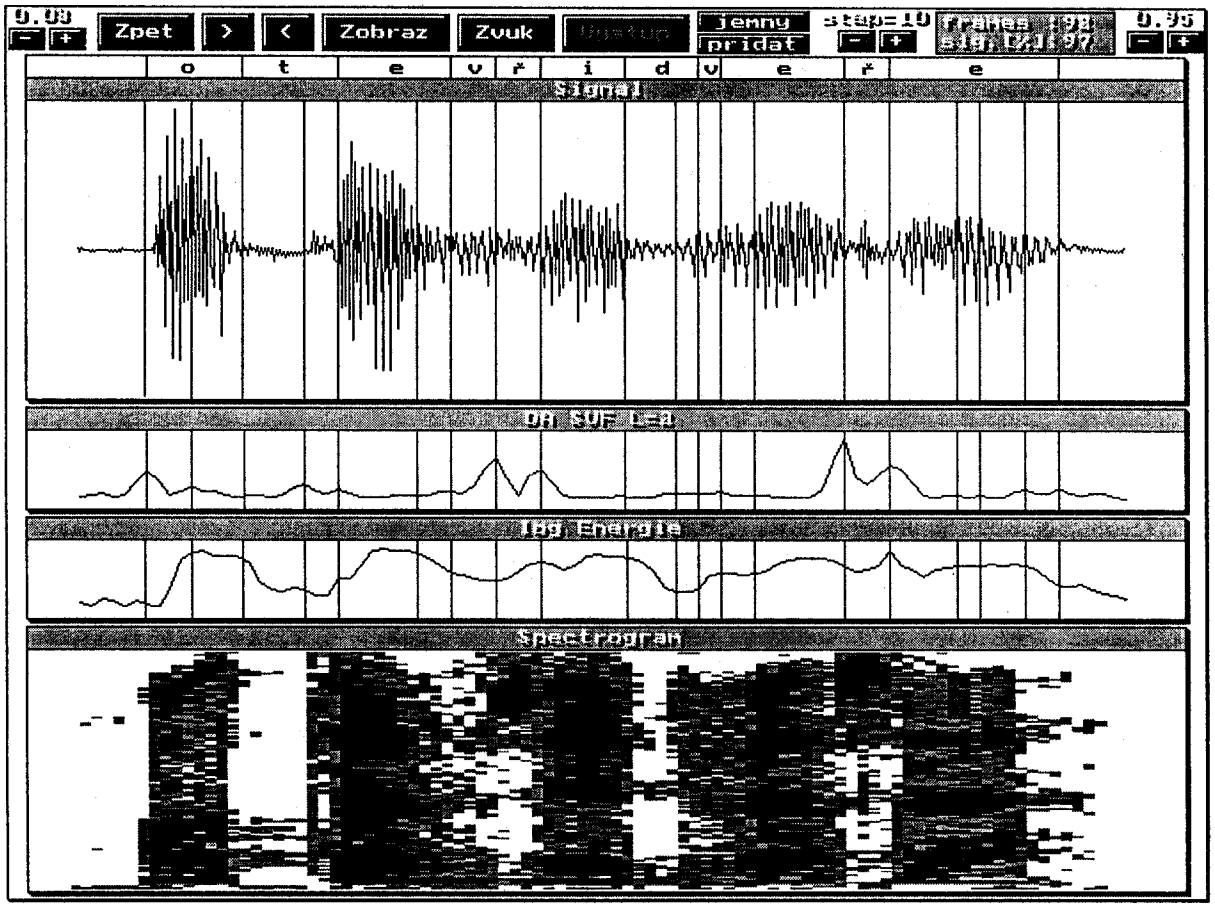

Fig. 1. A program for speech segmentation and phonetic label alignment based on a spectral variation function.

References:

[1] Palková Z.: Fonetika a fonologie češtiny. Univerzita Karlova, Praha 1992.

[2] The International Phonetic Alphabet. Journal of the Phonetic Association, vol.19, no.12, Dec. 1989.

[3] Wells, J.C. et al.: Standard Computer-Compatible Transcription. Final report of ESPRIT Project 2589 (SAM), 1992.

[4] Waves+ manual (version 5.1). Entropic Research Laboratory, 1996.

[5] Psutka J.: Komunikace s počítačem mluvenou řečí. Academia, Praha, 1995.

[6] Kvale K.: Segmentation and labelling of speech. Dissertation, Norwegian Institute of Technology, Trondheim, 1993.

[7] Nouza J.: Spectral Variation Functions Applied to Accoustic-Phonetic Segmentation of Speech Signals. Forum Phoneticum, Goethe University, Frankfurt, to appear.

[8] Teiml M.: Akusticko-fonetická segmentace signálu řeči. Diploma thesis. TU of Liberec, 19[8] Teiml M.: Akusticko-fonetická 95.

About authors...

Jan NOUZA was born in 1957. He received MSc and PhD level degrees at the Czech Technical University in Prague, in 1981 and 1986, respectively. Currently he is an associate professor at the Department of electronics and signal processing at the Technical University of Liberec. His research interests are mainly in the speech processing area with a special focus on automatic speech recognition.

Josef PSUTKA was born in 1951. He received MSc and PhD level degrees at the Czech Technical University in Prague, in 1974 and 1980, respectively. Now he is a professor at the Faculty of Applied Science of the West Bohemia University in Pilsen. His professional interest is aimed at the artificial intelligence and pattern recognition problems and especially at the speech recognition.

Jan UHLÍŘ was born in Prague in 1940. He received MSc and PhD level degrees at the Czech Technical University in Prague in 1962 and 1973, respectively. He is currently a professor at the Department of Circuit Theory at the Faculty of Electrical Engineering, Czech Technical University in Prague. He is interested in digital signal processing, mainly in speech recognition.