SAMPA V SLOVENČINE A JEJ VÝZNAM Z POHĽADU VIACJAZYČNÝCH SYSTÉMOV NA ROZPOZNÁVANIE REČI

Jozef Ivanecký

IBM Voice Systems, European Speech Research,
Vangerowstr. 18, D-69115 Heidelberg
ivanecky@de.ibm.com

Technická Univerzita Košice
Fakulta elektrotechniky a informatiky
Katedra kybernetiky a umelej inteligencie

ABSTRAKT

The paper deals with the phonetic alphabet SAMPA. In a brief survey of the development of the SAMPA transcription we are indicating possible problems in the formation of the SAMPA transcription for new languages and the significance of the SAMPA transcription for computer speech recognition from the multilingual systems‘ view. The core of the paper is the procedure used in the formation of the SAMPA transcription for Slovak as well as for a current status of the Slovak version of the SAMPA transcription.

1. ÚVOD

Význam fonetickej transkripcie v počítačovom rozpoznávaní reči nie je potrebné zdôrazňovať. Transkripcia je použitie príslušných znakov abecedy na reprezentáciu zvukov reči. Systém na počítačové rozpoznávanie reči rozpoznáva na najnižšej úrovni jednotlivé zvuky, ktoré sú označované znakmi fonetickej abecedy zvolenej pre danú fonetickú transkripciu.

Keďže ide o počítačové spracovanie, znakmi sú vo väčšine prípadov kvôli jednoduchosti a prenositeľnosti 7 bitové ASCII znaky. Priradenie ASCII znakov jednotlivým zvukom – t. j. kódovanie – je záležitosť tvorcu daného systému, pričom medzi konkrétnymi transkripČnými systémami sú – aj vzhľadom na fonetické rozdiely v spracúvaných jazykoch – väčšie či menšie rozdiely.

Tento stav viacmenej pretrváva dodnes, aj keď potreba jednotného kódovania v poslednom desaťročí narastá. Dôvodom je nielen potreba vzájomnej kompatibility (transkripčné systémy) a uľahčenie komunikácie pri výmene skúseností, no v poslednom čase aj nárast viacjazyčných systémov na rozpoznávanie reči. Na tento účel sa dnes takmer výhradne používa kódovanie SAMPA (Speech Assestment Methods – Phonetic Alphabet).

2. SAMPA – VÝVOJ A SÚČASNOSŤ

Transkripcia SAMPA vznikla na základe kódovania použitého v niekoľkých projektoch na rozpoznávanie reči na konci deväťdesiatych rokov. Každý z týchto projektov sa zaoberal jednojazyčným systémom, no pre rôzne jazyky. Kódovanie použité v týchto projektoch zobrazuje tabuľka 1.

Výsledkom dohôd v rámci SAM – European collaborative speech technology assessment research – boli odporúčania pre štandardný transkripčný systém aplikovateľný na mnohé európske jazyky. Pojem *transkripčný systém* má v tomto prípade oveľa širší význam. Problematika bola riešená v júni 1987 na University College v Londýne v rámci práce na *Labelling, Transcription a Management Methods for speech databases.* [1].

Súčasťou týchto prác bolo aj vytvorenie jednotného kódovania, dnes známeho ako SAMPA abeceda. Z tabuľky č. 1 je evidentné, že kódovacie tabuľky COST 209, Esprit 291 a Alvey majú veľa spoločného. Na základe toho boli postavené základy pre európsky štandard. Z princípov formulovaných v projekte Esprit 291 vyplýva nasledujúce: ASCII znaky v IPA tabuľke ostávajú nezmenené. Nie ASCII znakom – do tej miery, do akej je to možné – sú priradené veľké ASCII znaky. Použitím týchto pravidiel na tabuľku č. 1 dostaneme ASCII reprezentáciu pre nasledujúce znaky z fonetickej abecedy IPA:

<table>
<tr><th></th><th>IPA</th><th></th><th>COST</th><th>Esprit</th><th>Alvey</th><th>VL</th><th>DIN</th><th>ISO</th><th>Allen</th></tr>
<tr><td>1</td><td>ɑ</td><td>!</td><td>A</td><td>A</td><td>A</td><td>A</td><td>42</td><td></td><td>a (aa)</td></tr>
<tr><td>2</td><td>æ</td><td></td><td>{ 123</td><td>\</td><td>&</td><td></td><td>44</td><td>71</td><td>ae</td></tr>
<tr><td>3</td><td>ɐ</td><td></td><td>7</td><td>R \</td><td>6</td><td></td><td>41</td><td></td><td></td></tr>
<tr><td>4</td><td>ɒ</td><td></td><td>C</td><td></td><td>Q</td><td></td><td>43</td><td></td><td>ao</td></tr>
<tr><td>5</td><td>ɔ</td><td>!</td><td>O</td><td>O</td><td>O</td><td>O</td><td>66</td><td>a5A</td><td>O</td></tr>
<tr><td>6</td><td>ç</td><td></td><td>c</td><td>C</td><td>C</td><td></td><td>48</td><td></td><td>c-</td></tr>
<tr><td>7</td><td>ð</td><td>!</td><td>D</td><td>D</td><td>D</td><td></td><td>4A</td><td>73</td><td>D</td></tr>
<tr><td>8</td><td>ɛ</td><td>!</td><td>E</td><td>E</td><td>E</td><td>E</td><td>4D</td><td>a37</td><td>E</td></tr>
<tr><td>9</td><td>ə</td><td>!</td><td>@ 64</td><td>@</td><td>@</td><td>@</td><td>4C</td><td>a38</td><td>6</td></tr>
<tr><td>10</td><td>ɜ</td><td></td><td>3</td><td>@/</td><td>3</td><td></td><td>4E</td><td></td><td>3</td></tr>
<tr><td>11</td><td>ɤ</td><td>!</td><td>G</td><td>G</td><td>G</td><td>G</td><td>54</td><td>a3C</td><td>g-</td></tr>
<tr><td>12</td><td>ɪ</td><td>!</td><td>I</td><td>I</td><td>I</td><td></td><td>5A</td><td>a3F</td><td>I</td></tr>
<tr><td>13</td><td>ʎ</td><td></td><td>l,</td><td>L</td><td>F</td><td></td><td>60</td><td></td><td>lj</td></tr>
<tr><td>14</td><td>ɲ</td><td></td><td>n,</td><td>N˜</td><td>J</td><td></td><td>64</td><td>a56</td><td>nj</td></tr>
<tr><td>15</td><td>ŋ</td><td>!</td><td>N</td><td>N</td><td>N</td><td>N</td><td>63</td><td>a57</td><td>N</td></tr>
<tr><td>16</td><td>ø</td><td></td><td>|124</td><td>q</td><td>{ 123</td><td>&</td><td>67</td><td>79</td><td>oe</td></tr>
<tr><td>17</td><td>œ</td><td></td><td>\92</td><td>Q</td><td>9</td><td>@</td><td>68</td><td></td><td>oE</td></tr>
<tr><td>18</td><td>Œ</td><td></td><td>&</td><td></td><td></td><td></td><td></td><td></td><td>OE</td></tr>
<tr><td>19</td><td>θ</td><td>!</td><td>T</td><td>T</td><td>T</td><td></td><td>73</td><td></td><td>T</td></tr>
<tr><td>20</td><td>ʃ</td><td>!</td><td>S</td><td>S</td><td>S</td><td>S</td><td>71</td><td>a5F</td><td>S</td></tr>
<tr><td>21</td><td>ʊ</td><td>!</td><td>U</td><td>U</td><td>U</td><td></td><td>77</td><td>a73</td><td>U</td></tr>
<tr><td>22</td><td>ʌ</td><td></td><td>V</td><td>ˆ</td><td>V</td><td></td><td>79</td><td></td><td>A</td></tr>
<tr><td>23</td><td>Y</td><td>!</td><td>Y</td><td>Y</td><td>Y</td><td></td><td>3C</td><td></td><td>Y</td></tr>
<tr><td>24</td><td>ɥ</td><td></td><td></td><td>H</td><td>H</td><td></td><td>57</td><td></td><td></td></tr>
<tr><td>25</td><td>ʒ</td><td>!</td><td>Z</td><td>Z</td><td>Z</td><td>Z</td><td>3E</td><td>a77</td><td>Z</td></tr>
<tr><td>26</td><td>?</td><td>!</td><td>?</td><td>?</td><td>?</td><td></td><td>3F</td><td>a79</td><td>?</td></tr>
<tr><td>27</td><td>ʁ,R</td><td></td><td>R</td><td>R</td><td>K</td><td></td><td>6D,6E</td><td></td><td>R</td></tr>
<tr><td>28</td><td></td><td></td><td>E ˜</td><td>E˜</td><td>E+˜</td><td></td><td>4D22</td><td></td><td>E,n</td></tr>
<tr><td>29</td><td>aI</td><td></td><td>A>I</td><td>a/</td><td>aI</td><td>(Y)</td><td>405A</td><td></td><td>ai</td></tr>
<tr><td>30</td><td>ʧ</td><td></td><td>I ˆS</td><td>T</td><td>IS</td><td></td><td></td><td></td><td></td></tr>
<tr><td>31</td><td>:</td><td>!</td><td>:</td><td>:</td><td>:</td><td></td><td>2A</td><td></td><td>:</td></tr>
<tr><td>32</td><td>’</td><td></td><td>(’)</td><td>’</td><td>”</td><td></td><td>31</td><td></td><td>”</td></tr>
<tr><td>33</td><td>,</td><td></td><td>(‘)</td><td>”</td><td>%</td><td></td><td>32</td><td></td><td>’</td></tr>
</table>

Tabuľka 1: Prehľad počítačovo reprezentovateľných fonetických transkripcií. Znak ’!’ v treťom stĺpci poukazuje na dohodu medzi COST 209, Esprit 291 a Alvey tabuľkami. Kódovanie v DIN a ISO stlpci sú v hexa formáte.

IPA znaky	ɑ	ɔ	ð	ɛ	ə	ɤ	ɪ	ŋ	θ	ʃ	ʊ	ʏ	ʒ	ʔ	ː
kódované ako ASCII	A	O	D	E	@	G	I	N	T	S	U	Y	Z	?	:

Na základe tohto a ďalších pravidiel bola formovaná SAM-PA fonetická abeceda, prvýkrát publikovaná v [10]. Táto fonetická abeceda, predstavujúca prvý – počítačovo jednoducho reprezentovateľný – kódovací systém dohodnutý medzi predstaviteľmi ôsmich európskych krajín spolupracujúcich na projekte SAM, je v tabuľke č.2 a č.3. Prvá verzia SAMPA pokrývala tieto jazyky: angličtina, dánčina, francúzština, nemčina, španielčina a taliančina.

V čase jeho vzniku nebol do projektu SAMPA zahrnutý žiaden zo slovanských jazykov, z čoho vyplýva značná limitácia použitia SAMPA abecedy pre slovanské jazyky. Napriek týmto obmedzeniam sme v roku 1995 vytvorili prvú neoficiálnu verziu SAMPA abecedy pre slovenčinu [7].

V posledných piatich rokoch bola v rámci niekoľkých projektov SAMPA značne rozšírená a dnes už pokrýva mnohé – nie len indoeurópske – jazyky. Z priestorových dôvodov nie je možné, aby súčasná verzia SAMPA abecedy bola súčasťou tohto príspevku. Aktuálny stav SAMPA abecedy je možné nájsť na [11].

Napriek tomu, že zo slovanských jazykov dnes existujú oficiálne verzie SAMPA abecedy pre bulharčinu, chorvátčinu, poľštinu, ruštinu a slovinčinu, oficiálna verzia pre slovenčinu a rovnako ani pre češtinu dodnes nie je k dispozícii.

3. SAMPA A SLOVENČINA

Slovenskú fonetickú transkripciu ovplyvnili najmä dva činitele:

- úzus v zapisovaní slovenských nárečových textov
- technické vybavenie tlačiarní

Vývin našej fonetickej transkripcie súvisí aj s inými okolnosťami: s užším využívaním transkripcie IPA u nás a s ťažkosťami pri jej používaní na zápis prejavov v slovanských jazykoch podobne ako aj v mnohých orientálnych jazykoch [3]. Zostavovatelia transkripcie IPA sa totiž opierali predovšetkým o poznanie neslovanských jazykov. Vývin slovenskej transkripcie

IPA	SAM-PA	ASCII	IPA	SAM-PA	ASCII
Samohlásky			Spoluhlásky		
a	a	97	b	b	98
ɑ	A	65	c	c	99
æ	{	123	ç	C	67
ɐ	6	54	d	d	100
ɒ	Q	81	ð	D	68
ɔ	O	79	f	f	102
e	e	101	g	g	103
ɛ	E	69	ɣ	G	71
ə	@	64	h	h	104
ɜ	3	51	j	j	106
i	i	105	k	k	107
ɪ	I	73	l	l	108
o	o	111	ʎ	L	76
ø	2	50	m	m	109
œ	9	57	n	n	110
ɶ	&	38	ɲ	J	74
u	u	117	ŋ	N	78
ʊ	U	85	p	p	112
ʉ	}	125	r	r	114
ʌ	V	86	ʁ,R	R	82
y	y	121	s	s	115
Y	Y	89	ʃ	S	83
			t	t	116
			θ	T	84
			v	v	118
			w	w	119
			x	x	120
			ɥ	H	72
			z	z	122
			ʒ	Z	90
			’	?	63

Tabuľka 2: SAM abeceda (SAMPA) pre fonetickú transkripciu - 1987.

IPA	SAM-PA	ASCII	Komentár
Dvojznakové symboly			
œ̃	9	57,126	Príklad na nosovku
aI	aI	97,73	Príklad na diftong
ʧ	tS	116,83	Príklad na afrikatív
n̩	=n	61,110	Príklad na slabičné /n/
	E/	69,47	Príklad na archifonému

IPA	SAM-PA	ASCII	Komentr
Hraničné a prozodické vlastnosti			
	:	58	Znak pre kvantitu
	"	34	Znak pre hlavný prízvuk
	%	37	Znak pre vedľajší prízvuk
	-	45	Znak pre úroveň tónu
	'	39	Znak pre stúpavý tón
	`	96	Znak pre klesavý tón
	`'	96,39	Znak pre stúpavo–klesavý tón
	'`	39,96	Znak pre klesavo–stpavý tón
	$	36	Zank pre slabičnú hranicu
	|	124	Znak pre slabičnú hranicu vnútri skupiny samohlások
	-	45	Rozdeľovač

Tabuľka 3: SAM abeceda (SAMPA) pre fonetickú transkripciu - 1987.

závisel ďalej od toho, že staršie výskumné ciele nevyžadovali takú techniku (podrobnosť) zápisu, akú umožňuje IPA. Zápisy slovenských textov boli určené predovšetkým domácim bádateľom. Z tohto dôvodu sa vystačilo s jednoduchým systémom založeným na gramatickej sústave jazyka, v ktorom sa pre prepis hlásky volí to písmeno, ktoré je v ortografickej forme jeho základným ekvivalentom.

V tabuľke č. 4 je uvedený zoznam fónov spisovnej slovenčiny kódovaných znakmi používanými v slovenskej fonetickej transkripcii podľa [5]. Takáto fonetická abeceda je síce plne vyhovujúca domácemu použitiu, avšak v žiadnom prípade nevyhovuje pri použití na medzinárodnom poli, na čo upozorňuje už v roku 1966 [3]. Tento stav v oblasti slovenskej fonetickej transkripcie pretrváva dodnes.

Pri pokuse vytvoriť inventár slovenských fónov kódovaných pomocou SAMPA fonetickej abecedy sme narazili na dva základné problémy:

- Nekompatibilita slovenskej fonetickej transkripcie s tranksripciou IPA. Riešenie tohto problému malo pre nás veľký význam z toho dôvodu, že SAMPA je popisovaná práve na základe znakov IPA. Predpokladá sa rozšírenosť IPA systému a na základe toho jednoduchá konverzia zo znakov IPA na znaky SAMPA.

- SAMPA abeceda v čase svojho vzniku pokrývala len dosť malú časť repertoáru IPA. S určitými obmedzeniami a zavedením dvoch nových znakov ju bolo možné použiť na vytvorenie fonematickej abecedy slovenčiny [4], avšak na vytvorenie jej fonetickej abecedy nebola vhodná vôbec. Tento stav bolo možné zovšeobecniť pre väčšinu slovanských jazykov.

Pri vytváraní IPA verzie slovenskej transkripcie sme okrem deficitu zdrojov zaoberajúcich sa možnosťami aplikácie IPA systému na slovenčinu, narazili aj na isté malé rozdiely v [5] a [9] v popise transkripčného systému používaného pre slovenčinu. Na naše účely sme ako východiskový zdroj použili [5].

Z vyššie uvedených problémov vyplýva, že naším prvým logickým krokom malo byť vytvorenie IPA systému, na základe ktorého by sme mohli vytvoriť SAMPA systém. Nakoniec ale vzhľadom na problémy so získavaním informácií o IPA systéme – získať informácie o SAMPA systéme bolo pre nás nepomerne jednoduchšie – sme v mnohých prípadoch použili

Znak	Príklad	Znak	Príklad
a	p**a**pier	r	pa**r**a, And**r**ej
á	p**á**s	ṛ	p**r**st, v**r**ch
ä	p**ä**ť, m**ä**so	ṛ́	v**ŕ**ba
e	p**e**ro	l	ska**l**a, v**l**ak
é	dc**é**ra nov**é**	ḷ	v**l**k
i	p**i**vo, d**y**m	ḷ́	v**ĺ**ča
í	p**í**skať, nov**ý**	ň	va**ň**a, **n**iesť
o	p**o**p**o**l	ľ	**ľ**ad, **l**ipa, **l**iek
ó	p**ó**l	f	**f**ajka, **v**piť, **v**čela
u	p**u**to	v	slo**v**o, **v** oku
ú	p**ú**pava	w	**v**dova, **v**biť
ĭa	p**ia**tok		**v** zime, šé**f**lekár
ĭe	sp**ie**vať	ŭ	pra**v**da, otco**v**
ĭu	cudz**iu**		dlho**u** ulico**u**
ŭo	v**ô**ľa, k**ô**ň	s	o**s**em, **s**trom
p	**p**o**p**ol	z	vá**z**a, **z**bierať
b	ža**b**a	š	ko**š**eľa, prá**š**iť
t	va**t**a, **t**en**t**o	ž	ve**ž**a, **ž**aba
d	vo**d**a, **d**om	x	**ch**ata, po**ch**vala
ť	Ma**ť**o, pla**tiť**	h	Pra**h**a, **h**rýzť
ď	há**ď**a, **d**eti	γ	vr**ch** hory
k	pá**k**a, **k**r**k**	j	dva**j**a, pri**j**ímať
g	**g**á**g**ať, a**g**át	ĭ	kra**j**, kra**j**ský, z**j**esť
m	**mama**	c	**c**ena, po**c**ta
m̸	a**m**fiteáter, ta**m**von	ʒ	há**dz**ať, me**dz**a
n	ra**n**a, **n**os	č	o**č**i, ma**č**ka
n̸	pa**n**ský, i**n**žinier	ǯ	**dž**em, Ho**dž**a
ŋ	te**n**ký, ce**n**gať		

Tabuľka 4: Tabuľka znakov slovenskej fonetickej transkripcie

opačný postup. Najprv sme vytvorili prepis zo slovenského transkripčného systému do SAMPA systému. Znak IPA pre daný prepis sme na základe tohto prepisu už len vyhľadali v príslušnej tabuľke.

Napriek nekompatibilite slovenskej transkripcie s IPA systémom sa určité percento znakov prekrýva. Medzi zhodné znaky patria:

a, p, b, m, t, d, n, k

Týmto znakom sme mohli priamo priradiť aj rovnaké SAMPA znaky, čo vyplýva z princípov definovaných pre vytváranie SAMPA transkripčného systému. Ďalšie znaky, ktorých prepis bol relatívne bezproblémový napriek tomu, že sú pre ne použité v IPA systéme odlišné symboly, boli:

g, r, l, ľ, ň, v, f, g, x, s, z, š, ž, j, c, č, ʒ, ǯ

Z tohto zoznamu sa môže zdať, že prepis prebiehal bez ťažkostí. Musíme však poznamenať, že prvú verziu SAMPA pre slovenčinu sme vytvárali na fonologickej úrovni, čo značne zjednodušovalo riešenie. Dôvodom na to bola hlavne skutočnosť, že s prvou verziou SAMPA už táto úloha bola neriešiteľná bez pridania vlastných znakov. Hlavným problémom boli v tomto prípade palatálne konsonanty *ť* a *ď*. Pre ne sme v prvej verzie používali náhradné znaky. Znaky pre *ť* a *ď* boli do SAMPA zavedené až pri rozšírení SAMPA o gréčtinu, keďže žiaden z jazykov, na základe ktorých bola vytváraná prvá verzia SAMPA, tieto palatálne konsonanty nemá.
Znak pre palatalizáciu, ktorý bol zavedený skôr ako rozšírenie pre gréčtinu, je nevyhovujúci, napriek tomu, že sa používa v slovenskom rovnako ako aj českom transkripčnom systéme. Keďže slovenčina rovnako ako aj čeština obsahuje len palatálne konsonanty a neobsahuje palatalizované konsonanty, je v našich jazykoch používaný napriek zavádajúcej a milnej interpretácii pri konfrontácii so skutočnými palatalyzovanými konsonantami ake sa vyskytuju napríklad v ruštine.

Ďalším problémom pre nás boli samohlásky. V tomto prípade nešlo o nezrelosť SAMPA systému na popis samohlások, problém bol v rozchádzaní sa rôznych zdrojov na prepis slovenských samohlások do IPA systému. Nakoniec sme sa rozhodli pre použitie informácii zo zborníka IPA

ST	IPA	SAMPA	ST	IPA	SAMPA	ST	IPA	SAMPA
a	a	a	p	p	p	ň	ɲ	J
á	a:	a:	b	b	b	ľ	ʎ	L
ä	æ	{	t	t	t	f	f	f
e	ɛ	E	d	d	d	v	v	v
é	ɛ:	E:	ť	c	c	w	f̬	f_v
i	ɪ	I	ď	ɟ	J\	ŭ	u̯	u_^
í	i:	i:	k	k	k	s	s	s
o	o	o	g	g	g	z	z	z
ó	o:	o:	m	m	m	š	ʃ	S
u	u	u	~~m~~	ɱ	F	ž	ʒ	Z
ú	u:	u:	n	n	n	x	x	x
ǐa	i̯a	i_^a	~~n~~	?	n ?	h	ɦ	h\
ǐe	i̯ɛ	i_^E	ŋ	ŋ	N	γ	ɣ	G
ǐu	i̯u	i_^u	r	r	r	j	j	j
ŭo	u̯o	u_^o	r̥	r̩	r=	ǐ	i̯	i_^
			ŕ̥	r̩:	r=:	c	ts	ts
			l	l	l	ʒ	dz	dz
			l̥	l̩	l=	č	ʧ	tS
			ĺ̥	l̩:	l=:	ǯ	ʤ	dZ

Tabuľka 5: Porovnanie slovenskej transkripcie, IPA a SAMPA systému.

z roku 1999, v ktorom bola publikovaná aj IPA pre češtinu [2]. Problém dvojhlások bol vyriešený po zavedení znaku pre neslabičnosť.

V roku 2000 sme na základe aktualizovanej verzie SAMPA, ktorá už pokrývala podstatne väčšiu jazykovú skupinu ako na začiatku, rozhodli zrevidovať našu verziu SAMPA pre slovenčinu a previesť ju z fonologickej do fonetickej roviny. V tomto procese rozširovania sme narazili na jeden vážnejší problém, ktorý sa nám doteraz nepodarilo celkom uspokojivo vyriešiť. Je to prepis nazály *~~n~~*. Náš predpoklad je, že v rámci SAMPA rovnako ako aj IPA nie je možné pre tento prípad také podrobné rozlíšenie, a preto sa prepisujú vždy ako *n* . Túto skutočnosť je však potrebné overiť.

Tabuľka č. 5 obsahuje našu aktuálnu verziu SAMPA pre slovenčinu. Okrem toho sú v nej uvedené aj príslušné IPA znaky. Problematické znaky sú označené otáznikom. Keďže stále ide o pracovnú verziu, budeme vďační za akékoľvek pripomienky a návrhy, ktoré nám pomôžu ku konečnej realizácii SAMPA pre slovenčinu.

4. VIACJAZYČNÉ SYSTÉMY

Prečo má fonetická abeceda použiteľná pre viaceré jazyky taký veľký význam? Niektoré odpovede sme naznačili už na začiatku. Ak sa na problematiku pozrieme z pohľadu počítačového rozpoznávania reči, najväčší prínos vidíme práve pri tvorbe viacjazyčných systémov. Systém, ktorý je schopný rozpoznávať viaceré jazyky, môže pracovať dvoma spôsobmi:

- paralelne pracujúce systémy – každý systém vie rozpoznávať práve jeden jazyk;

- jeden systém schopný rozpoznávať viaceré jazyky.

V prípade prvého systému s počtom jazykov neúmerne narastá požiadavka na výkon. Výhodou je, že takéto systémy dosahujú vyššiu úspešnosť rozpoznávania, avšak v prípade zlej začiatočnej identifikácie jazyka je celé rozpoznávanie chybné.

Pozrime sa teraz bližšie na druhú možnosť. V tomto prípade máme systém, ktorý zvláda rozpoznávať celý fonetický repertoár niekoľkých rôznych jazykov. Výkonovo je menej náročný ako paralelne pracujúce systémy, a aj v prípade chyby na začiatku môže pokračovať správne. Nevýhodou takýchto systémov je menšia úspešnosť rozpoznávania ako v predchádzajúcom prípade. Napriek tomu podľa nášho názoru patrí budúcnosť práve týmto systémom. Dôvodom je to, že vytvorenie kvalitného systému je časovo a aj finančne značne náročné, z čoho vyplýva, že pre jazyky, ktorými hovorí len malá skupina ľudí, je to neefektívne. Použiť v takomto prípade systém natrénovaný na iný jazyk a obsahujúci potrebný fonetický inventár môže byť oveľa schodnejšie. Z tohto dôvodu je podľa nášho názoru podpora a polupráca pri vytváraní počítačovo jednoducho kódovateľnej medzinárodnej fonetickej abecedy viac než potrebná.

5. ZÁVER

V príspevku sme sa zamerali na prezentáciu aktuálneho stavu pri vytváraní SAMPA abecedy pre slovenčinu a poukázali sme na dôležitosť tohto transkripčného systému z pohľadu viacjazyčných systémov na počítačové rozpoznávanie reči.

V blízkej budúcnosti sa chceme zamerať na potvrdenie správnosti nami navrhovanej verzie SAMPA pre slovenčinu a doplnenie kódovania pre poslednú chýbajúcu hlásku, poprípade jej vyradenie. V poslednom kroku chceme zaradiť slovenčinu do oficiálneho zoznamu jazykov popísaných systémom SAMPA.

LITERATÚRA

[1] Fourcin A., Harland G., Barry W., Hazan V.: *Speech input and output assestment–multilingual methods and standards*, Ellis Horwood Limited 1989.

[2] Dankovičová D.,*Czech IPA*, Handbook of the International Phonetic Assocation, Cambridge University Press, 1999.

[3] Isačenko A. V., Romportl M.:*Návrh fonetické a fonologické transkripce češtiny a slovenštiny*, Slovo a slovesnosť, 27, 1966.

[4] Ivanecký J.:*Automatizácia fonematickej transkripcie slovenčiny*, Varia VII. Bratislava, SJS pri SAV 1996.

[5] Kráľ A.:*Pravidlá slovenskej výslovnosti*, Slovenské pedagogické nakladateľstvo, Bratislava 1983, 632 s.

[6] Kráľ A., Sabol J.: *Fonetika a fonológia*, Slovenské pedagogické nakladateľstvo, Bratislava 1987, 392 s.

[7] Krokavec D., Filasová A., Ivanecký J.: *Basic Properties of the Phonetic Transcription and the Sentence Description in Slovak Language.*, In: "COST-249 Meeting". Kosice, FEI Technical University 1996.

[8] Pauliny E.:*Slovenská fonológia*, Slovenské pedagogické nakladateľstvo, Bratislava 1979, 213 s.

[9] Sabol J.:*Syntetická fonologická teória*, Jazykovedný Ústav Ľ. Štúra SAV, Bratislava 1988, 255 s.

[10] Wells J. C.:*Computer–coded phonetic transcription.* Journal of the International Phonetic Association 17, No. 2, s. 94–114, 1987.

[11] Wells J. C.:*URL: http://www.phon.ucl.ac.uk/home/sampa/home.htm*, SAMPA - official WWW pages.