New Slovak Unit-Selection Speech Synthesis in ARTIC TTS System

Jindřich Matoušek, Daniel Tihelka, and Josef Psutka

***Abstract*—ARTIC (Artificial Talker in Czech) is a corpus-based text-to-speech (TTS) system that enables to synthesise an arbitrary text, mainly for the Czech language. Basically, two versions of ARTIC are available—a single unit instance system (also known as fixed-inventory synthesis) with the quality of resulting speech limited by the fixed inventory, and multiple unit instance system with the quality profitting from employing a unit-selection algorithm to select the longest suitable sequence of phonetic units from many units available. In this paper, a process of building a new Slovak voice for the unit-selection version of ARTIC is presented. All steps in the design, from the preparation of a suitable speech corpus to the creation of an acoustic unit inventory of the new Slovak voice and its use in the ARTIC system will be described. Text processing and speech production modules will be mentioned as well.**

***Index Terms*—text-to-speech, corpus-based speech synthesis, unit selection, Slovak language.**

I. Introduction

TEXT-TO-SPEECH (TTS) synthesis is one of the most important tasks of computer speech processing. Nowadays, corpus-based synthesis is the most widely used approach to speech synthesis. The current trend in this approach is to use large speech corpora and acoustic unit inventories to catch as many speech phenomena (i.e. spectral variations, prosodic variations, etc.) in segments of speech as possible (see e.g. [1] or [2]). In the case of such large acoustic unit inventories, an automation of the inventory creation process is very helpful, especially for multilingual and/or multi-voice TTS systems. Thanks to the automation, different inventories/voices can be created very quickly within a framework of a single TTS system.

In our previous work, ARTIC, a modern TTS system was developed to synthesize primarily Czech speech [3]. Two corpus-based approaches to speech synthesis are currently supported in the system, resulting in two versions of TTS systems: a *single unit instance* (SUI) system (also known as *fixed-inventory synthesis* or *diphone-based synthesis* [1]) with the quality of resulting speech limited by the fixed inventory, and *multiple unit instance* (MUI) system with the quality profitting from selecting (employing *unit-selection algorithm* [1], [4]) from many unit instances available or from modelling properties of speech statistically (using Hidden Markov models, HMMs, in *HMM-based speech synthesis* [5], [6]). ARTIC was primarily designed to synthesise Czech speech; two Czech voices for SUI and four high-quality Czech voices for MUI are currently available [7]. While focused on the Czech language, some experiments to synthesise speech in other languages, German [8] and Slovak [3], were also carried out. However, as previous-generation SUI version of the system was employed in these experiments, the quality of synthetic speech in these languages was limited and appears now to be insufficient in modern applications. Therefore, a new Slovak MUI system with a new unit-selection compatible voice was built from a new large speech corpus within the framework of ARTIC as described further in the paper.

Since corpus-based speech synthesis (both unit selection and HMM based) is very popular today and was shown to be able to produce synthetic speech of a high quality, new corpora in many languages (of course, in major languages like English [11] or French [12] but also in minor languages like Czech [7], [13], [14] or Slovak [15]) have been intensively designed. Thanks to data-driven approaches in various phases of the preparation of a new corpus, a new voice derived from this corpus can be designed relatively quickly and easily. Following the principles on which the Czech corpora were built (and which also proved to be useful for automatic recognition of the Slovak language [16], [17]), a new Slovak corpus for unit-selection speech synthesis was designed and is described in this paper.

All steps in the design of a new Slovak unit-selection speech synthesis system are described in the paper. Section II briefly introduces the ARTIC text-to-speech system. In Section III, the process of the preparation of a new Slovak speech corpus and the creation of a new Slovak acoustic unit inventory, including the description of a Slovak phonetic alphabet and its differences from Czech, are introduced. Text-processing issues, including examples of Slovak phonetic transcription rules, are shown in Section IV. Speech production utilising a unit-selection algorithm is depicted in Section V. Finally, Section VI concludes the paper and

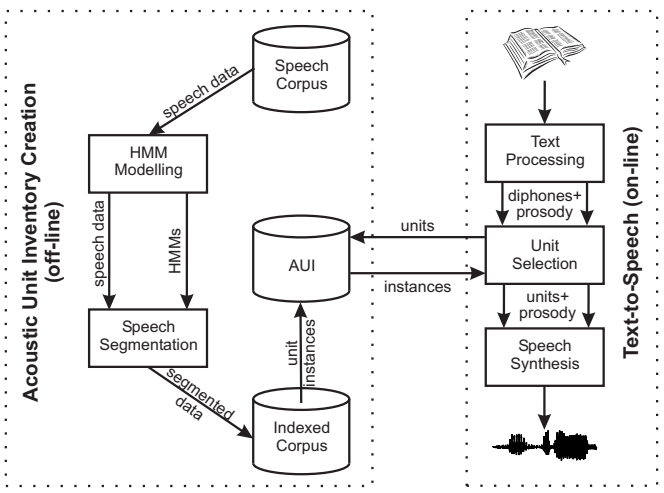

Fig. 1. Block diagram of ARTIC TTS system.

Manuscript received June 28, 2011; revised August 11, 2011. This work was supported by the Ministry of Industry and Trade of the Czech Republic, project No. MPO FR-TI1/518.

The authors are with the Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic, e-mail: {jmatouse, dtihelka, rompi, psutka}@kky.zcu.cz.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

TABLE I
SLOVAK PHONETIC INVENTORY WITH 54 PHONES USED IN ARTIC (SYMBOLS ARE IN SAMPA NOTATION [9], [10]).

Phone	Word	Trans.	Phone	Word	Trans.	Phone	Word	Trans.
a	mama	mama	d	dom	dom	Z	žena	Zena
e	pes	pes	c	ťava	cava	x	chata	xata
i	pivo	pivo	J\	háďa	h\a:J\a	h\	had	h\at
o	bok	bok	k	oko	oko	G\	nechže	JeG\Ze
u	bubon	bubon	g	guma	guma	r	rak	rak
a:	páv	pa:f	?	áno	?a:no	r=	vrch	vr=x
e:	želé	Zele:	m	mama	mama	r=:	vŕba	vr=:ba
i:	víno	vi:no	F	amfiteáter	aFfitea:ter	l	loď	loc
o:	katalóg	katalo:k	n	nos	nos	l=	vlk	vl=k
u:	múr	mu:r	N	banka	baNka	l=:	vĺča	vl=:t_Sa
{	päť	p{c	J	vaňa	vaJa	L	ľad	Lat
i_^a	piatok	pi_^atok	N\	Slovensko	sloveN\sko	j	jama	jama
i_^e	mier	mi_^er	f	figa	figa	u_^	pravda	prau_^da
i_^u	paniu	paJi_^u	w	vdova	wdova	i_^	kraj	krai_^
u_^o	kôň	ku_^oJ	v	vlak	vlak	t_s	cena	t_sena
p	prak	prak	s	osa	osa	t_S	oči	ot_Si
b	bod	bot	z	zima	zima	d_z	medza	med_za
t	vata	vata	S	šek	Sek	d_Z	džungľa	d_ZuNgLa

outlines our future work in synthesis of Slovak speech.

II. ARTIC TTS SYSTEM

ARTIC (Artificial Talker in Czech) is a Czech text-to-speech system developed since 1997 [18]. It is a corpus-based system which is based on a large carefully designed speech corpus (annotated on orthographic, phonetic and prosodic levels [19], [7]). From the very beginning, automatic HMM-based approach to acoustic unit inventory creation was used. Two speech synthesis methods, fixed-inventory synthesis (a SUI approach) and unit-selection synthesis (which could be, in fact, viewed as a sort of a MUI approach), are currently implemented [18]. Experiments with HMM-based speech synthesis method (a MUI approach) have been carried out recently as well [6]. The block diagram of the ARTIC TTS system is shown in Figure 1.

The basic speech units used in the SUI system are *triphones*. Since only one instance of each triphone is employed in the system, the representative instances are selected off-line. Due to the fixed inventory, compact inventories can be utilized (tens of MBs), but the quality of output speech is limited. As a result, the SUI version of ARTIC is suitable for low-resource devices.

On the other hand, many instances of each speech unit (*diphones* in this case) are employed in the MUI version, and the optimal instances are selected on-line using a unit-selection algorithm. As a result, high-quality speech can be produced in this way at the expense of employing large acoustic unit inventories (hundreds of MBs). In this paper, a process of building of a new Slovak voice for the MUI approach with unit selection is described.

In the following sections, a process of building of a new voice in a new language, Slovak, within the ARTIC TTS system framework will be described. Unlike [3], focus will be given to the MUI version of the system (incorporating the unit-selection algorithm). Spoken form of Slovak (and mainly phonetics and phonology) will be dealt with too. As Czech is the main language the TTS system ARTIC has been designed for, we will also mention some differences between Czech and Slovak. Furthermore, all steps in the design, including the preparation of a suitable acoustic unit inventory of the new Slovak voice, text processing and speech production itself, will be detailed.

III. SLOVAK ACOUSTIC UNIT INVENTORY

We used the same *phonetic inventory* as described in [3] and shown in Table I. To create an *acoustic unit inventory* (AUI) for a unit-selection system, a new large speech corpus of a Slovak voice was designed first. It comprised 12,070 utterances (approx. 19 hours of speech, see Table II). The texts of utterances were downloaded from Internet news portals, checked to comprise all phones in sufficient number of occurrences and recorded by a semi-professional voice talent in an anechoic chamber using high-quality recording devices. The recorded utterances were then carefully annotated following the principles described in [19], and phonetic transcripts were obtained automatically by applying rules described in Section IV. Beside speech waveforms, glottal signals were also recorded using an electroglottograph and used as input signals to glottal pulses (pitch-marks) detection algorithm [20], [21], [22]. Pitch-marks are used for an accurate computation of fundamental frequency (F0) in join cost (see Section V) and also as consistent concatenation points during speech synthesis.

TABLE II
OVERVIEW OF UTTERANCES IN THE NEW SLOVAK CORPUS.

Type of utterances	**Number**	**Length**
Declarative	10,000	∼17 hours
"Yes/No"-questions	1,013	∼55 mins
"Wh"-questions	1,057	∼63 mins
Total	**12,070**	**∼19 hours**

When comparing Czech and Slovak, it should be said that, being Slavic languages, both languages are very similar in all linguistic aspects (unlike e.g. German [8]). However, despite the similarity, there are some differences both in orthographic and phonetic forms. These differences should be taken into account when building a Slovak voice in the ARTIC TTS system. As for orthography, there are some Slovak letters which are not used in written Czech (namely *ä, ô, ľ, ŕ, Ĺ*). Acoustic unit inventories also do differ. Standard Slovak phonetic alphabet (denoted using SAMPA [9], [10]) consists

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

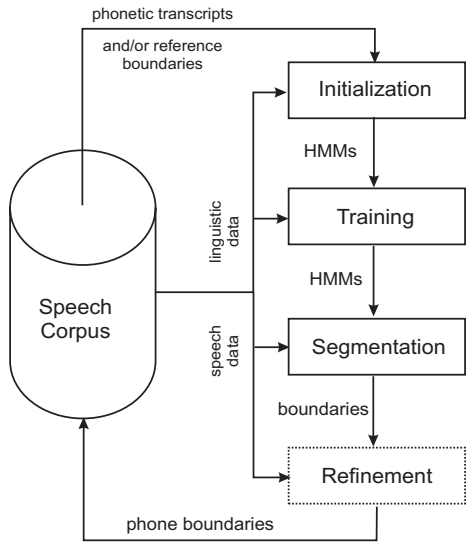

Fig. 2. Scheme of automatic acoustic inventory creation.

of 54 phones (see Table I) while 48 phones are usually used for Czech [23]. Here is a comparison between Slovak and Czech phonetic inventories:

Vowels. There are almost no distinctions between Czech and Slovak vowel systems—basically there are 5 short [a, e, i, o, u] and 5 long [a:, e:, i:, o:, u:] vowels in both languages. The only exception is an "additional" Slovak short vowel [{] which can rarely appear in spoken Slovak (often is pronounced as [e]).

Diphthongs. 4 diphthongs [i_^a, i_^e, i_^u, u_^o] occur in Slovak. None of them exists in Czech. On the contrary, there are three Czech diphthongs [o_u, a_u, e_u] that do not occur in Slovak.

Plosives. There are no differences between 9 Slovak and Czech plosives: [p, b, t, d, c, J\, k, g, ?]. [?] stands for glottal stop.

Affricates. 4 Slovak affricates are the same as the Czech ones: [t_s, t_S, d_z, d_Z].

Nasals. There are 5 "basic" nasals [m, F, n, N, J] in both Slovak and Czech. Moreover, another nasal [N\] can be pronounced in some contexts in Slovak.

Fricatives. There are 9 fricatives in "basic" Slovak [f, w, v, s, z, S, Z, x, h\]. They are the same as the Czech ones with the exception of [w] being an important variant of [v]. Moreover, due to voice assimilation "voiced *ch*" [G\] can be pronounced alternately with [h\] in both languages.

Liquids. In fact, 3 liquids occur in Slovak [r, l, L]. But there are also their significant allophones which express the syllabicity [r=, r=:, l=, l=:]. Symbol [=] denotes the syllabicity, [:] stands for "long" duration. "Long" syllabic phones [r=:, l=:] (written as *ŕ*, *ĺ*) and "soft" [L] (written as *ľ*) do not exist in Czech.

Glides. There are 3 glides in Slovak [j, u_^, i_^]. Just [j] occurs in Czech.

The process of the automatic acoustic unit inventory creation is illustrated in Figure 2. Hidden Markov models (HMMs) and ASR-based training procedures were employed to align HMMs with speech data, producing boundaries between phones in each source utterance. Optionally, the boundaries can be refined as described e.g. in [24], [25]. As diphones are used as basic speech units in ARTIC unit-selection system, diphone boundaries were then derived from the phone boundaries. More details about the automatic segmentation and acoustic unit inventory creation process can be found e.g. in [26], [27].

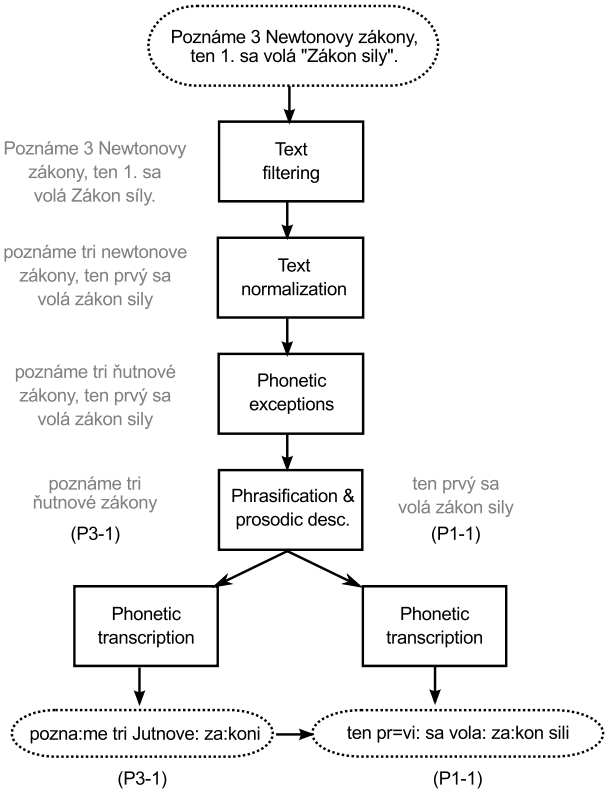

Fig. 3. Schematic view of text processing in ARTIC TTS system. For clarity, a single sentence is shown in the figure, phones-to-diphones conversion is not shown, and prosodic description is limited to prosodemes. Prosodeme P3-1 occurs in a non-terminating part of a sentence, while prosodeme P1-1 occurs in a declarative sentence (for list of all prosodemes, see e.g. [33]). In the real system more detailed prosodic description is used [34].

IV. Text Processing

Being a Slavic language, Slovak is similar to Czech in all linguistic aspects. Hence, very similar text-processing techniques as for Czech were carried out also for Slovak with the aim to reveal phonetic and prosodic aspects of an input text. Due to a complexity of such a task, current text processing in the ARTIC system is somewhat simplified to several main steps (see Figure 3):

- sentence boundary detection [28];
- text normalisation of "non-standard" words (digits, abbreviations, acronyms, etc.) [29], [30], [31];
- detailed rule-based phonetic transcription, including pronunciation dictionary of "exceptional" words (mostly foreign words) [32];
- phones-to-diphones conversion;
- prosodic description ("phrasification") in terms of prosodic symbols (prosodic clauses, phrases, prosodemes, etc.) using prosodic phrase grammar described in [33], [34].

Similarly as for Czech, phonetic form of Slovak is similar to orthographic form. Thus, relatively simple phonetic transcription rules can be utilised to convert Slovak letters to

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

phones. Based on phonetic rules defined by Slovak phoneticians in [35], approximately 150 rules in a form [32]

$$A \rightarrow B / C\_ \; D \tag{1}$$

were defined in our system. The rule can be read as follows: letter sequence A with both left context C and right context D is transcribed as phone sequence B. Some examples of these rules are given here (CON means a consonant, PAU stands for a pause, [|] denotes a word boundary, other notation is explained in Table III):

$$\text{VPC} \rightarrow \text{UPC} \; / \; * \_ \; \langle \text{NPC}, |\text{NPC}, |\text{PAU}, |? \rangle \tag{2}$$
$$\text{UPC} \rightarrow \text{VPC} \; / \; * \_ \; \langle \text{VPC}, |\text{VPC}, |\text{SON}, |\text{VOC}, |\text{v} \rangle \tag{3}$$
$$\text{TDNL} \rightarrow \text{ALV} \; / \; * \_ \; \langle \text{i}, \text{i :}, \text{e} \rangle \tag{4}$$
$$\text{r} \rightarrow \text{r} = \; / \, \text{CON\_} \; \langle \text{CON}, | \rangle \tag{5}$$
$$\text{Í} \rightarrow \text{l} =: \; / \; * \_ * \tag{6}$$
$$\text{ia} \rightarrow \text{i\_\^a} \; / \; * \_ * \tag{7}$$

TABLE III
NOTATION USED IN THE EXAMPLES OF PHONETIC TRANSCRIPTION RULES.

Abbreviation	Description	Phones
VPC	voiced paired consonant	b w d z d_z Z d_Z J\ g G\
UPC	unvoiced paired consonant	p f t s t_s S t_S c k x
TDNL	consonants "t d n l"	t d n l
ALV	alveopalatal consonants	c J\ J L
VOC	vocals and diphthongs	a a: e e: i i: o o: u u: { i_^a i_^e i_^u u_^o
SON	sonorant consonants	m F n N J N\ l l= l=: L r r= r=: j

The examples include special inter- and cross-word voice assimilation rules (2, 3), rules for "softening" (or palatalisation) of consonants (4), rules for transcribing short (5) and long (6) syllabic consonants, rules for transcribing diphthongs (7), etc. For instance, rule (4) can be read as "written *t, d, n, l* are pronounced as alveopalatal [c, J\, J, L] in front of [i, i:, e]". Multiple transcriptions of ambiguous Slovak contexts are supported as well. Pronunciation dictionary of "exceptional" words (foreign words but also some Slovak native words not obeying the transcription rules) is also employed.

Prosodic analysis includes punctuation-driven sentence clause detection, rule-based word stress detection and symbolic prosodic description adopted from the Czech language [33]. Symbolic features based on a prosodic phrase grammar, like prosodic sentence, prosodic clause, prosodic phrase, prosodic word, and prosodeme, were used to describe prosodic characteristics and to express prosodic structure of to-be-synthesised texts. The same parsing algorithm as for Czech was employed [34].

V. Speech Production

In MUI version of ARTIC, resulting speech is generated by a unit-selection algorithm (see e.g. [4], [36]). Its principle is to smoothly concatenate (according to *join cost*) speech segments (diphones in our case), extracted from natural utterances using the automatically segmented boundaries, from a large speech unit inventories according to phonetic and prosodic criteria (*target cost*) imposed by the *input specification* given by the synthesised utterance. As there are usually many instances of each speech segment, there is a need to select the optimal (with respect to both target and join costs) instances dynamically during synthesis run-time (using a unit selection technique). A scheme of unit-selection approach to speech synthesis is shown in Figure 4. Since the unit-selection framework is, in fact, language-independent, the same setup as for Czech was employed respecting the phonetic and prosodic nature of the Slovak language in the input specification.

To calculate the target cost, a prosodic structure of the to-be-synthesised utterance is estimated, and a comparison between prosodic symbolic features (plus some positional features, like position of a diphone in a prosodic word, and contextual factors like immediate left and right phone) in the utterance and in the unit inventory is carried out. Join cost is evaluated as a distance between spectral features (mel-frequency cepstral coefficients, MFCCs, are currently used) and F0 around the concatenation point of two potentially neighbouring speech units.

After selecting the optimal sequence of (diphone) speech segments, neither prosodic nor spectral modifications are made in the ARTIC system except for simple smoothing at concatenation points. To cope with high CPU power and memory cost typical for unit-selection systems, a computational optimisation was proposed in [36].

VI. Conclusion and Future Work

In this paper, a new Slovak voice suitable for a multiple-unit-instance version (with unit-selection algorithm) of the ARTIC TTS system was presented. The new Slovak voice was built on the same principles as the existing Czech voices within the framework of the ARTIC system. All steps in the design including the preparation of an appropriate acoustic unit inventory of the new Slovak voice, text processing and speech production modules were mentioned in the paper. According to informal listening tests the new MUI (unit-selection) version of Slovak clearly outperformed the SUI version. In fact, speech synthesised by the MUI version was never rated as worse than speech synthesised by the SUI version.

In our future work, we plan to focus mainly on text processing issues, especially on advanced text pre-processing, fine-tuning of phonetic transcription rules and updating of the pronunciation dictionary of exceptional words. Application of HMM-based speech synthesis method for Slovak language is also under consideration.

References

[1] T. Dutoit, "Corpus-based speech synthesis," in *Springer Handbook of Speech Processing*, J. Benesty, M. Sondhi, and Y. Huang, Eds. Dordrecht: Springer, 2008, pp. 437–455.
[2] P. Taylor, *Text-to-Speech Synthesis*. Cambridge: Cambridge University Press, 2009.
[3] J. Matoušek and D. Tihelka, "Slovak text-to-speech synthesis in ARTIC system," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004, vol. 3206, pp. 155–162.
[4] D. Tihelka and J. Matoušek, "Unit selection and its relation to symbolic prosody: a new approach," in *Proc. INTERSPEECH*, Pittsburgh, USA, 2006, pp. 2042–2045.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

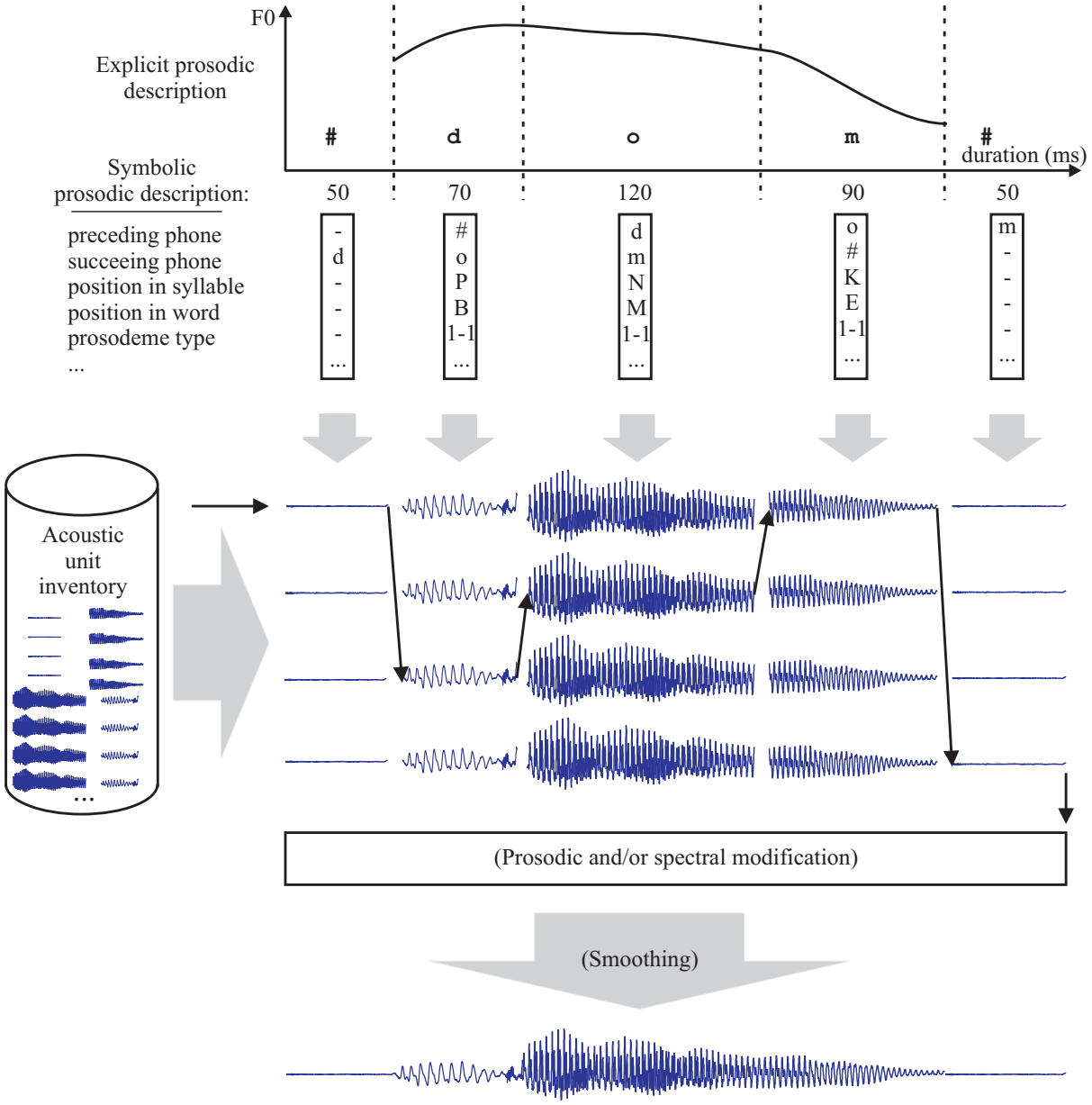

Fig. 4. A schematic view of unit-selection speech synthesis driven by symbolic prosody in the ARTIC TTS system (adapted from [1]). Position in syllables are denoted as (N)ucleus, (P)raetura, and (K)oda. Position in word is (B)eginning, (M)iddle, and (E)nd. Prosodeme 1-1 occurs in a declarative sentence (for list of all prosodemes, see e.g. [33]).

[5] H. Zen, K. Tokuda, and A. Black, "Statistical parametric speech synthesis," *Speech Commun.*, vol. 51, no. 11, pp. 1039–1064, 2009.
[6] Z. Hanzlíček, "Czech HMM-based speech synthesis," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Springer, 2010, vol. 6231, pp. 291–298.
[7] J. Matoušek, D. Tihelka, and J. Romportl, "Building of a speech corpus optimised for unit selection TTS synthesis," in *Proc. LREC*, Marrakech, Morocco, 2008.
[8] J. Matoušek, D. Tihelka, J. Psutka, and J. Hesová, "German and Czech speech synthesis using HMM-based speech segment database," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2002, vol. 2448, pp. 173–180.
[9] "Slovak SAMPA," http://www.ui.savba.sk/speech/sampa_sk.htm.
[10] J. Ivanecký and M. Nábělková, "Fonetická transkripcia SAMPA a Slovenčina (SAMPA transcription and the slovak language)," *Jazykovedný časopis*, vol. 53, no. 2, pp. 81–95, 2002, (in Slovak).
[11] J. Kominek and A. Black, "CMU ARCTIC speech databases, the," in *Proc. SSW*, Pittsburgh, USA, 2004, pp. 223–224.
[12] H. Francois and O. Boeffard, "Design of an optimal continuous speech database for text-to-speech synthesis considered as a set covering problem," in *Proc. INTERSPEECH*, Aalborg, Denmark, 2001, pp. 829–832.
[13] V. Radová and J. Psutka, "Recording and annotation of the Czech speech corpus," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2000, vol. 1902, pp. 319–323.
[14] J. Matoušek and J. Romportl, "On building phonetically and prosodically rich speech corpus for text-to-speech synthesis," in *Proc. CI*, San Francisco, USA, 2006, pp. 442–447.
[15] M. Rusko, M. Trnka, S. Daržágín, and M. Cerňak, "Slovak speech database for experiments and application building in unit-selection speech synthesis," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004, vol. 3206, pp. 457–464.
[16] J. Psutka, J. Hajič, and W. Byrne, "The development of ASR for Slavic languages in the MALACH project," in *Proc. ICASSP*, Montreal, Canada, 2004, pp. 749–752.
[17] J. Psutka, P. Ircing, J. V. Psutka, J. Hajič, W. Byrne, and J. Mírovský, "Automatic transcription of Czech, Russian, and Slovak spontaneous speech in the MALACH project," in *Proc. INTERSPEECH*, Lisbon, Portugal, 2005, pp. 1349–1352.
[18] J. Matoušek, D. Tihelka, and J. Romportl, "Current state of Czech text-to-speech system ARTIC," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, vol. 4188, pp. 439–446.
[19] J. Matoušek and J. Romportl, "Recording and annotation of speech

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

corpus for Czech unit selection speech synthesis," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin: Springer, 2007, vol. 4629, pp. 326–333.

[20] M. Legát, J. Matoušek, and D. Tihelka, "A robust multi-phase pitch-mark detection algorithm," in *Proc. INTERSPEECH*, Antwerp, Belgium, 2007, pp. 1641–1644.

[21] M. Legát, D. Tihelka, and J. Matoušek, "Pitch marks at peaks or valleys?" in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007, vol. 4629, pp. 502–507.

[22] M. Legát, J. Matoušek, and D. Tihelka, "On the detection of pitch marks using a robust multi-phase algorithm," *Speech Commun.*, vol. 53, no. 4, pp. 552–566, April 2011.

[23] "Czech SAMPA," http://www.phon.ucl.ac.uk/home/sampa/czech-uni.htm.

[24] S. S. Park and N. S. Kim, "On using multiple models for automatic speech segmentation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 8, pp. 2202–2212, 2007.

[25] J. Matoušek and J. Romportl, "Automatic pitch-synchronous phonetic segmentation," in *Proc. INTERSPEECH*, Brisbane, Australia, 2008.

[26] J. Matoušek, D. Tihelka, and J. Psutka, "Experiments with automatic segmentation for Czech speech synthesis," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2003, vol. 2807, pp. 287–294.

[27] J. Matoušek, "Automatic pitch-synchronous phonetic segmentation with context-independent HMMs," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Springer, 2009, vol. 5729, pp. 178–185.

[28] J. Romportl, D. Tihelka, and J. Matoušek, "Sentence boundary detection in Czech TTS system using neural networks," in *Proc. ISSPA*, Paris, France, 2003, pp. 247–250.

[29] J. Kanis and L. Müller, "Using the lemmatization technique for phonetic transcription in text-to-speech system," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004, vol. 3206, pp. 355–361.

[30] ——, "Automatic lemmatizer construction with focus on OOV words lemmatization," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, vol. 3658, pp. 132–139.

[31] J. Zelinka, J. Kanis, and L. Müller, "Automatic transcription of numerals in inflectional languages," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, vol. 3658, pp. 326–333.

[32] J. Psutka, L. Müller, J. Matoušek, and V. Radová, *Mluvíme s počítačem Česky (Talking with Computer in Czech)*. Prague: Academia, 2006, (in Czech).

[33] J. Romportl and J. Matoušek, "Formal prosodic structures and their application in NLP," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, vol. 3658, pp. 371–378.

[34] J. Romportl, "Prosodic phrases and semantic accents in speech corpus for Czech TTS synthesis," in *Text, Speech and Dialogue*, ser. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2008, vol. 5246, pp. 493–500.

[35] Ábel Kráľ, *Pravidlá slovenskej výslovnosti (Rules of Slovak Pronunciation)*, 2nd ed. Matica slovenská, 2009.

[36] D. Tihelka, J. Kala, and J. Matoušek, "Enhancements of Viterbi search for fast unit selection synthesis," in *Proc. INTERSPEECH*, Makuhari, Japan, 2010, pp. 174–177.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)