

OÙ SONT LES LIMITES PHONOSTYLISTIQUES DU TCHÈQUE SYNTHÉTIQUE

Marie Dohalská - Zichová, Jana Mejvaldová

*Institut de Phonétique, Faculté des Lettres, Université Charles Prague
nám. J. Palacha 2, 116 38 Prague 1, République Tchèque
e-mail: marie.dohalska@ff.cuni.cz, jana.mejvaldova@ff.cuni.cz*

Abstract: *Exploring Phonostylistic Limits of Synthetic Czech.*

Results of practical TTS diphone synthesis applications employed especially in aids for handicapped seem to suggest the need to refine and „humanize“ the prosodic quality of synthetic Czech.

Our experiments concentrate on analyzing relations and recompensations in intensities (I), durations (T), and F_0 in both natural and synthetic speech. In fairly neutral sentences, we observe intensity drop at final syllables as low as 54% of the TTS value, whereas the durations extend up to 370% - all while keeping F_0 at the level same or close to the TTS. *Initial* syllables show a reverse trend: intensity may rise to 113%, and duration only 77% of the TTS values.

Results of pilot auditory testing of Czech listeners (45 philology students) reveal that substantial changes of (I) and (T) in relation to TTS values get significantly positive response with 84% tested persons.

Our current research focuses on verifying limits for possible (acceptable) changes in (I) and (T) in synthesized model sentences with varied phonostylistic tenor.

Mots-clés: synthèse, tchèque, prosodie, variantes phonostylistiques, intensité, durée, fréquence fondamentale

Dans notre contribution, nous nous sommes basées sur une recherche de longue date du tchèque parlé naturel et synthétique. Les règles prosodiques employées à l'origine étaient élémentaires et ne prenaient en compte que la fréquence fondamentale / F_0 / en négligeant la durée /T/ et l'intensité /I/.

Les expériences sorties de l'analyse profonde de la parole *naturelle* ont été appliquées à la synthèse de la parole (**TTS LPC 1994**). L'acceptabilité et le caractère naturel de ces

locutions synthétiques ‘neutres’ ainsi que l’influence des changements des relations mutuelles des trois composantes **T**, **I**, **F_o** ont été vérifiés par des tests auditifs.

La simulation en laboratoire aspire à atteindre un modèle de la prosodie synthétique tchèque la plus naturelle et ‘humaine’ que possible. On nous a confirmé que les handicapés et notamment les non-voyants auxquels avant tout la synthèse est consacrée, ont la tendance à ne pas employer la prosodie qui se trouve être souvent fatigante et stéréotypée. Il est alors nécessaire de créer une prosodie la plus naturelle et variable que possible. En plus, il faut avoir la possibilité de la changer d’après la nature de la communication présentée par les différents services d’information (le météo, l’annuaire électronique, le programme culturel ou l’audition du journal ou de la littérature). Les non-voyants revendiquent aussi que la parole synthétique avec une prosodie naturelle puisse être accélérée.

Cette recherche forme une partie du projet COST 258 de l’Union Européenne *Naturalness of Synthetic Speech* auquel nous participons en coopération avec **l’Institut de la Radiotéchnique et Electronique** de l’Académie des Sciences de la République Tchèque.

La partie de la recherche présentée ici s’attache à la recherche déjà publiée (voir la bibliographie). Nos expériences récentes se concentrent sur l’analyse plus profonde des relations et des compensations mutuelles des changements d’intensité /I/, de durée /T/ et de fréquence fondamentale /F_o/ dans la parole naturelle et de ses applications à la parole synthétique.

Dans les phrases synthétiques plus ou moins neutres nous avons pu constater, dans la première phase de nos recherches, la chute de l’intensité dans la syllabe finale jusqu’à 54% des valeurs automatiques. Par contre, les expériences actuelles aboutissent aux constatations suivantes:

- il est souhaitable que l’intensité soit réduite jusqu’à 25 - 33 % des valeurs de la prosodie automatique;
- parallèlement à la chute de l’intensité, nous avons observé une augmentation significative de la durée dans la syllabe finale - jusqu’à 370% pour une voyelle longue;
- pour une voyelle finale brève, l’augmentation de la durée se limite à un interval de 145% (pour une syllabe ouverte) à 162% (pour une syllabe fermée).

Ces constatations concernent les syllabes finales des phrases neutres qui ne dépassent pas 9 - 12 syllabes.

Dans cette phase, nous nous sommes concentrées surtout sur les syllabes initiales et finales des phrases brèves, dont les valeurs peuvent être directement appliquées aux phrases plus longues.

En quête des valeurs limites du point de vue de *la tolérance perceptive*, nous sommes parties des valeurs numériques des phrases considérées par les tests auditifs comme les plus naturelles. Nous nous sommes concentrées sur l’application des valeurs dégagées dans deux groupes de phrases:

- 1° phrases interrogatives initiées par un interrogatif *kdo* (*qui*), *kdy* (*quand*), *kolik* (*combien*), *jak* (*comment*), etc. pour leurs fréquence communicatives;

2^e phrases avec la structure lexicale et syntaxique neutre qui n'aquièrent leur sens sémantique précis qu'en étant réalisées avec une prosodie particulière.

Les différentes variantes de la prosodie synthétique peuvent ajouter à la phrase des sens différents et même opposés. Ce phénomène propre sans doute à toutes les langues, se réalise, dans chaque langue, par une intonation spécifique.

La prosodie synthétique, réglée manuellement, peut donner aux phrases des sens très différents et différencier les nuances fines sémantiques.

Dans ce sens, nous sommes arrivées aux valeurs limites concrètes, au-dessus desquelles la parole synthétique devient inacceptable et sa perception désagréable.

Les valeurs de l'intensité /I/, de la durée /T/ et de la fréquence fondamentale /F_o/ ont été modifiées systématiquement de façon suivante:

a) F_o restant stable, nous avons modifié les valeurs de I:

Kolik je hodin? (Quelle heure est-il?)

	k	o	l	i	k	j	e	h	o	d'	i	n	?
F _o	112	110	120	126	112	97	92	97	94	78	72	67	60

L'intensité de la voyelle augmente dans la syllabe initiale jusqu'à 176 %.

b) Dans ce type de phrases, la composante *temporelle* est relativement stable et son changement ne procure pas de variations phonostylistiques.

c) La fréquence dans la première syllabe atteint 191 % pour un des deux types de question. Dans le deuxième type où la fréquence culmine dans la deuxième syllabe, la valeur limite est 171 %. Pour pouvoir augmenter la fréquence fondamentale, il a fallu conserver les valeurs de la durée et de l'intensité modérées:

	k	o	l	i	k	j	e	h	o	d'	i	n	?
T	94	100	73	79	48	50	50	76	95	94	138	96	89
I	85	95	83	79	57	62	60	57	45	37	35	33	25

To se ti povedlo. (C'est gagné!)

Valeurs 'neutres':

	t	o	s	e	t'	i	p	o	v	e	d	l	o	.
T	68	100	55	93	62	87	82	96	91	87	85	85	107	100
F _o	146	150	152	160	145	143	105	104	101	95	90	87	84	80
I	102	106	71	70	67	66	88	93	76	60	41	36	33	30

Différences entre les valeurs minimales et maximales de la durée des phonèmes dans la phrase *To se ti povedlo*:

	t	o	s	e	t'	i	p	o	v	e	d	l	o	.
min.	68	100	55	93	62	87	82	96	91	87	85	85	107	100
max.	98	165	89	147	99	141	138	153	148	144	142	110	164	157
Δ	30	65	34	54	37	54	56	57	57	57	57	25	57	57

Valeurs maximales obtenues des trois phrases (dans une phrase, une composante a été changé uniquement):

	t	o	s	e	t'	i	p	ø	y	e	d	l	ø	.
T	98	165	89	147	99	141	138	153	148	144	142	110	164	157
F ₀	195	207	152	160	145	143	105	104	101	95	50	87	84	80
I	172	176	71	70	67	66	88	92	72	56	38	32	29	25

Nous avons été obligées d'arrêter nos recherches au problème du timbre. Il serait peut-être possible d'aller plus loin en introduisant des diphones correspondant à certaines variantes phonostylistiques du tchèque. Lorsque nous avons augmenté la fréquence au maximum, nous avons atteint un changement qui pourrait être aussi perçu comme un changement du timbre..

Jusqu'ici, nous avons analysé les limites phonostylistiques sur des exemples de phrases provenant de différentes situations de communication et ayant donc différentes variantes prosodiques. Désormais, nous voudrions nous consacrer sur les valeurs qui forment les limites perceptifs entre une proposition indicative et interrogative.

Cette recherche est réalisée grâce au soutien de Grantová agentura České republiky qui a octroyé le projet N° 405/95/6223 'Časové členění mluvené češtiny'.

BIBLIOGRAPHIE

- Dohalská, M. - Ptáček, M. (1994) Quelques remarques sur la perception du tchèque synthétique, AUC Phonetica Pragensia VII, Karolinum, Charles University Prague, 76-126
- Dohalská, M. - Duběda, T. (1996) Rôle des changements de la durée et de l'intensité dans la synthèse du tchèque, XXI Journées d'Etudes sur la Parole, Avignon, 375 - 377
- Dohalská, M. (1996) Testing Intelligibility at Maximum Rate, AUC Phonetica Pragensia IX, Karolinum, Charles University Prague, 53 - 59
- Horák, P. - Mejvaldová, J. (1996) Design and Application of New Czech Prosody Modelling System, Speech Processing 6th Czech - German Workshop, Prague