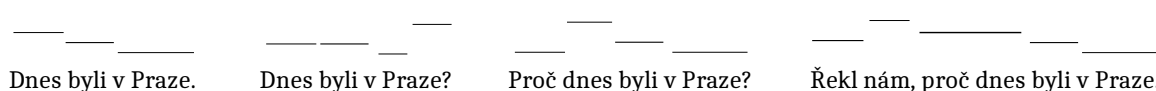


Doplňovací otázky v syntéze řeči

Markéta Jůzová¹

1 Úvod

Pojmem *intonace* se označují změny tónu v průběhu řeči. Tento jev není používán pro rozlišení slov (s výjimkou tónových jazyků), ale rozlišuje různé typy vět, postoje řečníka apod. Protože čeština nemá pevně daný slovosled, je intonace (stoupavá/klesavá) někdy jediným faktorem, podle kterého lze poznat, o jaký typ věty se jedná – zvláště se to týká rozlišení zjišťovací otázky a oznamovací věty (viz obr. 1).



Obrázek 1: Schématicky znázorněný průběh intonace oznamovací věty, zjišťovací (stoupavá intonace) a doplňovací otázky (klesavá intonace) a souvětí; převzato z Palková (1994)).

Doplňovací otázky (začínající jedním z tázacích zájmen) se v neutrální řeči vyznačují specifickým intonačním průběhem – na začátku věty je tzv. *vrchol* (stoupavě klesavá intonace, Palková (1994)), na konci věty se vyskytuje intonace klesavá.

2 Větná intonace v syntéze řeči

Ukazuje se, že v případě syntézy řeči (angl. *text-to-speech*, TTS) je větná intonace (např. klesání na konci oznamovací věty nebo stoupání na konci zjišťovacích otázek) pro posluchače velmi důležitá a významná odchylka od předpokládané, lingvisticky správné melodie bývá vnímána velmi negativně. Proto v metodě syntézy řeči výběrem jednotek jako jeden z příznaků pro výpočet optimální posloupnosti jednotek používáme *prozodém* (Romportl a Kala (2007)), abstraktní prozodický příznak, který odpovídá určité komunikační funkci. V současné verzi TTS ARTIC se používají 4 základní typy prozodémů:

- *P1* - poslední prozodické slovo (Romportl a Kala (2007)) oznamovacích vět (před tečkou)
- *P2* - poslední prozodické slovo otázek (před otazníkem)
- *P3* - poslední prozodické slovo před čárkou
- *P0* - jinde

Je tedy zřejmé, že současný TTS systém zatím nerozlišuje různé typy otázek, všechny jsou syntetizovány se stoupavou koncovou intonací.

Protože je popis intonace doplňovacích otázek složitější, nejsou při syntéze jednotky z těchto otázek používány, přestože byly nahrány řečníkem při pořizování řečového korpusu a tvoří asi 6 % nahraných dat. Naším cílem bylo zjistit, zda je možné používat i jednotky z nahraných doplňovacích otázek při syntéze, aniž by se zhoršila kvalita výsledné syntetizované řeči, a jestli není lepší syntetizovat doplňovací otázky vždy s klesavou intonací. Úlohu jsme

¹ studentka doktorského studijního programu Aplikované vědy a informatika, obor Kybernetika, specializace Počítačová syntéza řeči, e-mail: juzova@kky.zcu.cz

rozdělili do dvou experimentů. Podrobnější popis problému a výsledky poslechových testů lze nalézt v Jůzová a Tihelka (2016).

3 Experimenty

V prvním experimentu jsme vytvořili nový TTS systém přidáním doplňovacích otázek. Jelikož byla řečová data nahrávána profesionálem, mají všechny tyto otázky stejný průběh intonace – klesavý, a proto jsme jednotky v posledních prozodických slovech označili prozodémem *P1*. Poté jsme původním i novým TTS systémem vysyntetizovali velké množství vět a náhodně vybrané promluvy s velkými rozdíly v jednotkách použili v poslechových testech (hodnotilo je 7 posluchačů), jejichž výsledky jsou prezentované v tab. 1. Pozitivní hodnota skóre ukazuje, že při použití TTS s doplňovacími otázkami došlo ke zlepšení celkové kvality syntetizované řeči.

V druhém experimentu jsme zjišťovali preference posluchačů týkající se koncové intonace v doplňovacích otázkách. Vysyntetizovali jsme velké množství doplňovacích otázek ve dvou variantách (se stoupavou a klesavou intonací) a opět vytvořili poslechové testy. Pro větší objektivitu testy vyhodnocovalo celkem 22 posluchačů. Počty odpovědí i hodnota skóre (viz tab. 1) ukazují, že posluchači v doplňovacích otázkách preferují klesavou intonaci .

TTS s doplň. otázkami je lepší	stejná kvalita	původní TTS je lepší	skóre
152	138	74	0.214
klesavá varianta je lepší	nelze rozhodnout	stoupavá varianta je lepší	skóre
355	170	245	0.143

Tabulka 1: *Výsledky poslechových testů.* Nahoře – ovlivnění kvality syntézy jednotkami z doplňovacích vět, dole – preference stoupavé/klesavé intonace.

4 Závěr

Experimenty potvrdily, že je možné pro syntézu řeči využít doplňovací otázky, které byly pořízeny během nahrávání řečového korpusu, a to se stejným koncovým prozodémem jako oznamovací věty. Poslechový test týkající se preferencí stoupavé/klesavé intonace na konci doplňovacích otázek ukázal, že posluchači upřednostňují variantu klesavou. Výběr stoupavé intonace u některých vět zdůvodňovali tím, že varianta promluvy s klesavou intonací zněla spíše jako konec oznamovacího souvětí (na začátku chyběl vrchol, viz obr. 1). Dalším plánovaným experimentem je tedy zavedení nového prozodému označujícího část promluvy, ve které by se vrchol měl nacházet.

Poděkování

Příspěvek byl podpořen grantovým projektem číslo SGS-2016-039.

Literatura

- Jůzová, M., a Tihelka, D., 2016. Difficulties with wh-questions in Czech TTS System. Odesláno na TSD 2016.
- Palková, Z., 1994 *Fonetika a fonologie češtiny: s obecným úvodem do problematiky oboru.* Karolinum, Praha.
- Romportl, J. a Kala, J., 2007. Prosody Modelling in Czech Text-to-Speech Synthesis. *Proceedings of the 6th ISCA Workshop on Speech Synthesis*, s. 200-205.