



První experimenty s náhradou hlasu pro pacienty po LE

Markéta Jůzová¹

1 Úvod

Řeč je považována za jednu z běžných součástí našeho života, jedná se také o nejpřirozenější způsob komunikace mezi lidmi. Pozitivní nález rakoviny hrtanu a jeho následné odstranění (*totální laryngektomie*, LE) však může člověka o hlas připravit. Spolu s hrtanem jsou totiž odstraněny i hlasivky, které jsou jeho součástí.

V rámci projektu zkoumajícím možnosti pacientů po totální laryngektomii se ve spolupráci s klinikou ORL Fakultní nemocnice v Motole snažíme o zlepšení kvality jejich života (Romportl et al. (2015)). Jednou z možností, jak těmto pacientům umožnit přirozený způsob komunikace, tedy řeč znějící jako jejich ztracený hlas, je *konzervace hlasu* ještě před radikální operací a následná tvorba syntézy řeči (Matoušek et al. (2004)). Tu je pak možné využívat k běžné komunikaci s okolím prostřednictvím mobilního telefonu, tabletu či počítače.

2 Postup tvorby syntézy řeči

Prvním krokem tvorby syntézy řeči je příprava *textového korpusu*. Jedná se o sadu vět připravených tak, aby obsahovaly co nejvíce různých řečových jednotek s vysokou četností (řečovou jednotkou může být např. foném nebo difón, Matoušek et al. (2008)), a to v různých prozodických pozicích (Romportl (2006)). Tato sada v případě obecné syntézy (tedy ne pro pacienty před LE) obsahuje obvykle kolem 10 000 vět a je nahrávána profesionálním řečníkem ve zvukové komoře. Nahrávky jsou rozděleny na jednotlivé řečové jednotky, z nichž se pak spojováním tvoří syntetická řeč. Uvedme tři základní typy syntézy používané na našem pracovišti:

- *HMM syntéza* - jedná se o statistické modelování vlastností řečových jednotek, syntetizovaná řeč vzniká spojováním natrénovaných modelů
- *Konkatenáční syntéza s jedním reprezentantem* (SUI) - pro každou řečovou jednotku je vybrán průměrný reprezentant, pro dosažení požadované prozodie se signál modifikuje
- *Konkatenáční syntéza s více reprezentanty* (MUI) - pro každou řečovou jednotku uložíme všechny reprezentanty, při syntéze se vybírá jejich optimální posloupnost; je obecně pokládána za nejpřirozenější syntézu nejvíce podobnou skutečné řeči člověka, ale vyžaduje dostatečné pokrytí všech jednotek v korpusu

2.1 Modifikace obecného postupu

Pro účely úlohy *konzervace hlasu* před radikální operací bylo třeba obecný postup částečně modifikovat, neboť se objevilo několik závažných problémů způsobených specifickými podmínkami a požadavky. Prvním, a také největším z nich je jistě nedostatek času - max. 1-2 týdny od diagnózy po operaci. V tomto krátkém časovém období je třeba pacienta nahrát. Jelikož ne všichni mají možnost přijet na naše pracoviště, je možné nahrávat v nemocnici, testujeme

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

i nahrávání doma. Nicméně oba alternativní způsoby vedou ke zhoršení kvality nahrávek a tím i samotné syntézy.

Pacienti také nebývají zkušenými řečníky, někteří z nich mají problémy s ovládním počítače (jedná se často o starší osoby), takže je pro ně nahrávání dosti komplikované. Z toho důvodu byl pro ně připraven nový, speciální textový korpus, obsahující pouze jednodušší a kratší věty (max. 8 slov). Při výběru vět byl zohledněn i fakt, že množství vět, které bude konkrétní pacient schopen nahrát, není předem známé. Proto výběr probíhal v několika krocích, aby bylo zajištěno maximální možné pokrytí řečových jednotek pro libovolný počet nahraných vět. Podrobněji je příprava takového textového korpusu popsána v Jůzová et al. (2015).

Velký vliv na kvalitu výsledné syntézy řeči mají řečnické schopnosti jednotlivých pacientů. Jak již bylo zmíněno výše, většinou se nejedná o zkušené řečníky, a tak dochází během nahrávání k častým přerukům, zadrhávání v řeči, nádechům uprostřed vět. Důležitým krokem před samotnou segmentací je proto ruční kontrola nahrávek, tj. zda text přesně odpovídá nahraným větám, a případná oprava všech nesrovnalostí. Pacienti mají často také bolesti při mluvení, mohou mít chrapot, velmi patrná je i únava hlasu pacientů v průběhu nahrávacího sezení.

3 Zhodnocení dosavadního průběhu projektu

V rámci projektu bylo zatím nahráno několik pacientů a byla pro ně rovněž vytvořena jejich vlastní syntéza. Vzhledem k různému (často nízkému) počtu vět nebylo vždy možné použít MUI metodu, tu lze využít jen u pacientů, kteří nahrají alespoň 1 000 promluv. Pokud pacient nahrál jen několik set vět, zkoušely se aplikovat další dvě zmíněné metody. Obecně nelze říct, zda je v takovém případě lepší SUI nebo HMM metoda, záleží vždy na konkrétním hlase, jeho vlastnostech a také na subjektivním vnímání každého z nás.

První pacienti již používají syntézu řeči na svých mobilních zařízeních v běžném životě. Na základě našich dosavadních zkušeností je syntéza řeči přijímána pozitivně a pomáhá pacientům překlenout obtížné období a usnadnit jim komunikaci s okolím. Kladné reakce dostáváme i od rodinných příslušníků, kteří oceňují to, že jejich blízký si s nimi povídá hlasem velmi podobným jeho ztracenému hlasu.

Poděkování

Výzkum vedoucí k těmto výsledkům byl financován z prostředků Norského finančního mechanismu na období 2009-14 a Ministerstva školství, mládeže a tělovýchovy v rámci projektové smlouvy č. MSMT-28477/2014, projekt č. 7F14236, a grantovým projektem SGS-2013-032.

Literatura

- Romportl, J., Řepová, B., and Betka, J., 2015. Vocal Rehabilitation of Laryngectomised Patients by Personalised Computer Speech Synthesis. *Phoniatics*, 2015 in press.
- Matoušek, J., Romportl, J., Tihelka, D. and Tychtl, Z., 2006. Recent Improvements on ARTIC: Czech text-to-speech system *INTERSPEECH 2004 - ICSLP, proceedings of the 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, 2006.
- Matoušek, J., Tihelka, D. and Romportl, J., 2006. Building of a Speech Corpus Optimised for Unit Selection TTS Synthesis *LREC 2008, proceedings of 6th International Conference on Language Resources and Evaluation*, Jeju Island, ELRA, 2006.
- Romportl, J., 2006. Structural Data-Driven Prosody Model for TTS Synthesis. *Proceedings of the Speech Prosody 2006 Conference*, Dresden, 2006.
- Jůzová, M., Romportl, J., and Tihelka, D., 2015. Speech Corpus Preparation for Voice Banking of Laryngectomised Patients. *Accepted on TSD 2015*.