

# Studentská Vědecká Konference 2012

## Příprava textového korpusu pro syntézu řeči z limitované oblasti

Markéta Jůzová<sup>1</sup>

### 1 Úvod

Syntéza řeči je úloha, která převádí text na vstupu na řeč na výstupu. Dnes již dokáží TTS systémy (angl. *text-to-speech*) syntetizovat téměř cokoliv. Ideální TTS systém by dokázal převést libovolný text do mluvené podoby v takové formě a kvalitě, aby nebyla vzniklá promluva rozpoznatelná od nahrávky stejného textu namluveného člověkem, tj. se správnou výslovností a přirozenou intonací, plynulá a srozumitelná.

Speciálním případem je *syntéza řeči z limitované (omezené) oblasti* (angl. *Limited Domain*, dále LD). Jedná se o syntetizování textu z předem známé omezené oblasti (např. zprávy o počasí, časová hlášení, telefonní automaty společností a pomůcky pro nevidomé a slabozraké s omezeným počtem vět). Obecnost systému bude tedy nižší, ale předpokládáme větší přirozenost promluv.

### 2 Syntéza řeči konkatenací metodou

Konkatenací metoda je založená na zřetězování úseků přirozené řeči. Postup při syntéze touto metodou je následující:

1. nahrání velkého množství vět zkušeným řečníkem (vznikne tak *řečový korpus*)
2. *segmentace* korpusu na krátké řečové úseky (fonémy, difony, trifony) a uložení těchto *řečových jednotek do inventáře*
3. hledání vhodné jednotky v inventáři a řetězení za sebe při syntéze

Nejlépe na sebe navazují jednotky pocházející ze stejné promluvy. Pokud se syntetizuje úsek, který je obsažen v korpusu, výsledná řeč je plynulá a nejsou slyšet nepřirozené artefakty. Konkatenací metoda se tedy nabízí využít u systémů syntézy z LD, kde předem známe nejčastější věty a fráze a můžeme je tak zahrnout do textového korpusu.

### 3 Textový korpus pro syntézu řeči z limitované oblasti

Limitované oblasti jsou charakteristické používáním malého počtu vět, z nichž některé jsou neměnné (pozdravy, chybová hlášení, instrukce) a v ostatních se mění jen některé části (tzv. *sloty*). Příkladem věty obsahující sloty je např.

- Upozorňujeme cestující, že vlaková souprava obsahuje  
vůz první|druhé třídy| jídelní vůz, který je zařazen  
na začátku | uprostřed | na konci soupravy.
- Chcete jet do stanice  
Plzeň, hlavní nádraží | Horažďovice, předměstí | Klatovy | ...?

Nahrávat věty několikrát se všemi možnými kombinacemi by bylo zbytečné a náročné pro řečníka, ačkoliv „syntéza“ by byla dokonalá (jednalo by se vlastně jen o přehrávání předem namluvených vět, ne o syntézu).

<sup>1</sup> student navazujícího studijního programu Aplikované vědy a informatika, obor Kybernetika a řídicí technika, specializace Umělá inteligence, e-mail: juzova@students.zcu.cz

Pro minimalizaci korpusu se obvykle nahrává celá věta pouze jednou, s malými pauzami před a za sloty, a poté samostatně zbylé výrazy. Při syntéze se úseky řeči za sebe napojují v místech pauzy. Taková syntetizovaná řeč ale není přirozeně plynulá.

Aby byla zachována přirozenost i minimálnost korpusu, navrhla jsem způsob, jak ze zadaných vět vygenerovat fráze (sloty s kontextem), které by se nahrávaly. Díky kontextu má algoritmus syntézy více možností napojení úseků řeči z různých promluv (viz obr. 1). Příkladem frází je např.

- ...do stanice \$STANICE?
- ...zařazen \$UMISTENI soupravy.

Chcete jet do stanice Praha, hlavní nádraží?

... do stanice Plzeň, Jižní předměstí?

**Obrázek 1:** Možnosti napojení frází

Výsledný textový korpus tedy bude obsahovat všechny neměnné věty, všechny věty s proměnnými rámci a poté všechny vygenerované fráze ideálně se všemi možnými výrazy, které lze do slotů dosadit. Korpus bude splňovat kritérium maximálního pokrytí LD při minimálním počtu nahrávaných vět a frází. Bude sice o něco větší, než kdyby se řetězilo v pauzách, ale očekáváme větší přirozenost.

#### 4 Problematické sloty

Ne vždy je možné nahrávat úplně všechny výrazy pro jejich velký počet. Je ale dobré vybrat únosný počet těch nejčastějších a algoritmus by pak do slotů dosazoval podle četnosti. Ve dvou LD (informace o přijímačkách, informace o vlakových spojeních), na kterých jsem algoritmus tvorby frází a dosazování výrazů do slotů testovala, jsem řešila:

- **Jména, příjmení** - četnosti jsem získala ze statistik Ministerstva vnitra
- **Názvy vlakových stanic** - četnosti jsem získala z přeepsaných dialogů na informacích Hlavního nádraží v Plzni
- **Různé druhy číslovek v různých pádech** - připravila jsem minimální počet číslovek, který byl potřeba pro pokrytí oblasti

#### 5 Shrnutí

Tvorba textového korpusu pro syntézu řeči z limitované oblasti za předpokladu znalosti častých vět a výrazů je poloautomatická a vyžaduje jen minimální interakci uživatele.

**Poděkování:** Příspěvek byl podpořen grantovým projektem SGS-2010-054.

#### Literatura

- [1] *Psutka J., Matoušek J., Müller L., Radová V.: Mluvíme s počítačem česky* Academia, Praha, 2006
- [2] *Taubr M.: Příprava korpusu pro syntézu řeči z limitované oblasti* Diplomová práce, ZČU, Plzeň, 2006