TEXT TO SPEECH SYSTEM

Jakub OCZKO, Master Degree Programme (5)
Dept. of Computer Graphics and Multimedia, FIT, BUT
E-mail: xoczko00@stud.fit.vutbr.cz

Supervised by: Ing. Igor Szöke

ABSTRACT

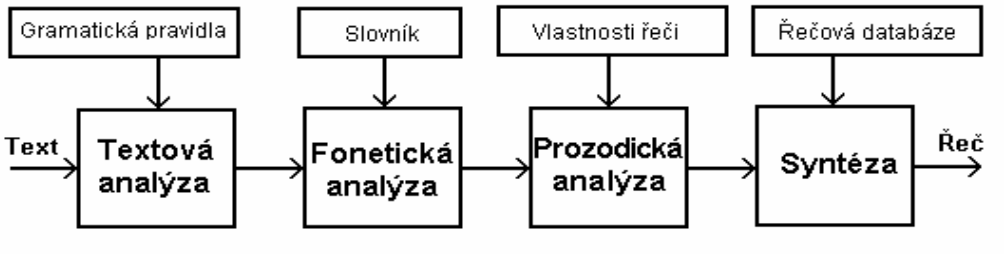
This paper presents a text to speech system for Czech language, based on a harmonic and noise model . All block of the system are described individually. The blocks have exactly specified inputs and outputs. The principles of processing input text, synthesis speech and modification of prosody in this system are described, and an example is shown in the paper. The system has modular design and will be used in others applications developed at the faculty.

1 ÚVOD

Člověk komunikuje s počítačem různými prostředky. Jedním z těchto prostředků je i řeč. Pomocí řeči se dá komunikovat dvěma směry, tedy jak člověk s počítačem tak počítač s člověkem. U komunikace počítače s člověkem se může jednat buď o čisté čtení textu (čtení knihy), čtení ovládacích prvků (ovládání počítače), nebo o interakci s uživatelem. Při syntéze řeči se snažíme přiblížit přirozenému charakteru řeči. Naším cílem bylo vytvořit systém pro převod textu na řeč v českém jazyce, který by byl modulární, dostatečně modifikovatelný a syntetizoval by řeč v co nejlepší kvalitě.

2 PŘEVOD TEXTU NA ŘEČ

Celý systém pro převod textu na řeč se skládá ze čtyř hlavních bloků (viz obr. 1).



Obr. 1 Blokové schéma systému pro převod textu na řeč

První tři bloky systému zpracovávají vstupní text tak, aby byl použitelný pro samotnou syntézu. Při komunikaci jednotlivých částí musí být jasně definovány vstupy a výstupy z každé části. Celý systém se dá rozdělit na dvě hlavní části a to na část *zpracování textu* a na *část syntézy*.

3 ZPRACOVÁNÍ TEXTU

V této části si popíšeme jednotlivé části určené pro zpracování textu, ukážeme podle jakých pravidel a kritérií se řídíme. Vstupem celého systému je jakýkoliv text. Může být strukturovaný, nebo se může jednat o prostý text. Výstupem zpracování textu je řetězec jednotek s řídícími údaji pro následnou změnu prozodie, které se syntetizují.

3.1 TEXTOVÁ ANALÝZA

Nejdříve se provádí substituce, tedy nahrazování jednoho slovního významu jiným (zkratky). Substituce se provádí jako první, neboť kdyby byla prováděna později bylo by obtížné v již upraveném textu najít dané podřetězce, které mohly být změněny nebo odstraněny ze vstupního textu. K substituci se využívá seznam pravidel, která určují, která slova mají být nahrazena za jiná. Pravidla jsou ve tvaru regulárních výrazů například `^VUT$` je vyhledáno v textu a nahrazeno textem `vysoké učení technické`. Výstupem substituční části je stále prostý text.

Další částí textové analýzy je normalizace. Normalizací rozumíme převod netextových informací na textové. Jedná se především o čísla a znaky. Na každou z těchto částí se aplikují pravidla ve tvaru regulárních výrazů, příklad pravidla pro přepis čísla `^10$` na text `deset`. Přepis je prováděn rekurzivně dokud lze aplikovat nějaké pravidlo, neboť při přepisu mohly vzniknout další netextové části.

3.2 FONETICKÁ ANALÝZA

Fonetická analýza má za úkol převést text na posloupnost fonémů. Tento převod se nazývá fonetická transkripce. Zjednodušeně by se dalo říct, že se snažíme psaný text upravit do takové formy jak jej slyší člověk, například `včela` slyšíme jako `fčela`. Fonetická transkripce je určena k přesnému a jednoznačnému zápisu mluvené řeči. Přesný zápis je závislý na jazyce a zvolené abecedě v jaké tuto transkripci provádíme.

Při fonetické transkripci se používá obsáhlý soubor pravidel, která určují přepis jednotlivých fonémů a jejich spojení. Vstupní text je porovnáván ze seznamem pravidel a vyhovující části jsou přepsány na nové. V této části se provádí také převod jednotlivých fonémů do abecedy SAMPA.

3.3 PROZODICKÁ ANALÝZA

Souvislá řeč by zněla příliš monotónně, kdyby se skládala jen z hlásek, shluknutých v plynulé řetězce fonémů. Ke změně prozodie se využívá především dvou činitelů, výšky a hlasitosti. V prozodických vlastnostech češtiny se jedná o stručná pravidla a definice některých základních prozodických jevů, které se vyskytují při plynulé mluvě. Prozodickými jevy jsou, pauza, změna intonace, slovní a větný přízvuk a tempo řeči.

Při načítání věty se určil typ věty podle koncového znaku (tázací, oznamovací…). Dle

tohoto typu se určuje jaké mají být výšky základního tónu jednotlivých fonémů prvních a posledních slov. Je dána startovací a koncová hladinu výšky tónu pro jednotlivé typy vět a podle těchto hodnot jsou přiřazovány jednotlivým fonémům jejich výšky tónu. Výsledný text je vkládán do struktury, která nese informace o vlastnostech každého fonému ve větě.

4 SYNTÉZA

Abychom byli schopni vytvořit řečový signál, základem je mít řečovou databázi. Tato databáze obsahuje namluvený parametrizovaný text, který je vhodně navržen tak, aby obsáhl pokud možno co nejvíce řečových jednotek. Databáze řeči má linearizovaný průběh základního tónu, aby při spojování jednotek docházelo k co nejmenším nepřesnostem.

Nejprve se musí z databáze vybrat vhodné jednotky. Snaží se nalézt co nejdelší spojité části. Postupně jsou hledány dané jednotky od nejdelších částí (slova) po foném. Takto vybrané jednotky s ukazatelem na výběr z databáze a prozodickými vlastnostmi jsou předány syntéze.

V syntéze jsou jednotlivé jednotky načteny z řečové databáze. Při načítání přiřazujeme jednotlivým jednotkám jejich prozodické hodnoty a to podle toho, jakému fonému odpovídají. Následně se upravuje prozodie tak, že se upraví hodnotou jednotlivých jednotek a syntetizují. Syntéza řeči pracuje na HNM (harmonic and noise) modelu. Řečový signál se skládá ze znělé a neznělé části. Znělý řečový signál je rozdělen podle frekvenčního spektra na dvě části. Část spektra na nižších frekvencích je reprezentována hlavně harmonickou části signálu, zatímco část spektra na vyšších frekvencích je reprezentována šumovou složkou řečového signálu. Výsledný syntetizovaný signál je roven součtu harmonické a šumové složky.

5 ZÁVĚR

Byl implementován kompletní systém pro převod textu na řeč se změnou prozódie, kde vstupem tohoto systému je řeč a výstupem je řečový signál. Tento systém je implementován tak, aby byl snadno upravovatelný, tedy při dodržení vstupů a výstupů jednotlivých modulů, se dají jednotlivé části samostatně měnit. Další využití tohoto modulu je pro aplikace v rámci fakulty.

LITERATURA

[1] Oczko, J.: Systém pro převod textu na řeč. Brno, Semestrální projekt VUT v Brně 2006

[2] Psutka, J.: Komunikace s počítačem mluvenou řečí. Praha, Academia 1995, ISBN 80-200-0203-0

[3] Szöke, I.: Prozódie syntetické řeči. Brno, Diplomová práce. VUT v Brně, 2003

[4] Stylianou, I.: Harmonic plus Noise Models for Speech, combined with Statistical Methods for Speech a Speaker Modification. PhD Thesis. École nationale supérieure des Télécommunications (ENST), Paris , 1996