

SIN04: Řečová interakce a sociální sítě

Luděk Bártek

Fakulta Informatiky
Masarykova Univerzita

podzim 2013

Obsah

- 1 Syntéza řeči
- 2 Syntéza v časové oblasti
- 3 Syntéza ve frekvenční oblasti

Úvod

- Cíle:
 - Převod textu na mluvenou řeč.
 - Výsledek by měl znít co nejpřirozeněji.
- Přirozená řeč by měla obsahovat:
 - správnou výslovnost (spodoba znělosti, koartikulace, ...)
 - správnou intonaci:
 - správná větná melodie
 - správné časování
 - správné umístění a intenzita přízvuků (větných, slovních).

Druhy syntézy řeči

- Syntéza v časové oblasti:
 - Založena na spojování segmentů řeči do větších celků.
 - Zvláštní druhy:
 - korpusová syntéza
 - problémově orientovaná syntéza.
- Syntéza ve frekvenční oblasti:
 - simuluje chování řečového ústrojí.

Fáze syntézy řeči

- 1 Fonetický přepis syntetizovaného textu - zahrnuje:
 - spodobu znělosti - dochází k ní na hranici mezi znělou a neznělou souhláskou a na hranici mezi souhláskou a samohláskou.
 - přepis čísel
 - přepis cizích slov
 - přepis zkratk
 - ...
- 2 Syntéza foneticky přepsaného textu.
- 3 Případný postprocessing - doplnění intonace, přízvuků, ...

Fonetický přepis textu

- Psaná a mluvená podoba textu nejsou shodné.
- Slouží k přesnému a jednoznačnému zápisu syntetizované promluvy.
- Využívá fonetickou abecedu:
 - mezinárodní fonetická abeceda IPA (součást standardu UNICODE)
 - SAMPA - 7bitový přepis IPA, který se využívá v různých syntetizérech.
- Ukázka zápisu pomocí SAMPA:
 - tSeScina je kra:sni: jazik
- Stejná promluva pomocí IPA:
 - tʃ'ɛfcina je kr'a:sni: j'azik

Fonetický přepis textu (pokračování)

- Nelze si pamatovat přepisy všech promluv - nutno zabezpečit automatický přepis.
- Pravidla fonetického přepisu:
 - Mohou mít regionální charakter (výslovnost na shledanou Čechy vs. Morava).
 - Obecně přepis nemusí využívat všechny znaky abecedy (y=i, ě=je/e, ů=ú, ...)
 - Zohledňuje koartikulaci (spodobu znělosti na rozhraní znělé a neznělé souhlásky).

Syntéza v časové oblasti

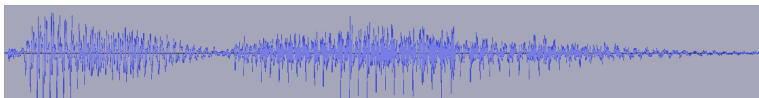
- Cíl - převod obecného textu na mluvenou řeč.
- Postavena na spojování segmentů řeči.
- Využívají se různé délky základních segmentů:
 - Nejpoužívanější typy segmentů - alofóny, difóny, trifóny, slabičné segmenty, ...
 - Delší segmenty:
 - lepší možnost modelování prozodických jevů (větné, slovní přízvuky, větná melodie, ...)
 - větší paměťové nároky - potenciálně až m^n segmentů (n - délka segmentu, m - počet segmentů)
 - příklady segmentů - slova, části vět, věty, ...
 - Kratší segmenty:
 - horší možnost modelování prozodie
 - menší paměťové nároky - méně segmentů.

Používané řečové segmenty

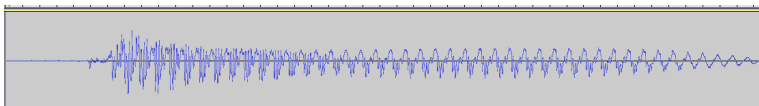
- Nelze použít přímo fonémy - koartikulace.
- Alofóny - poziční varianta segmentu
 - obsahuje foném a okolí ovlivněné koartikulací (n^3 alofónů, kde n je počet fonémů).
- Difóny - začínají uprostřed jednoho fonému a končí uprostřed následujícího (n^2 difónů).
 - často používané pro syntézu i pro rozpoznávání (např. syntetizér MBrola)
- Trifóny - začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního fonému. (n^3 trifónů).
 - Často používané pro rozpoznávání a syntézu řeči.
- Slabičné segmenty - umělá obdoba slabik.
 - Jaké je struktura slabiky?
 - Délka 1 - 3 fonémy.
 - Využívají se např. v TTS Demosthénés.

Ukázka vlivu koartikulace

- Spojení fonémů "p" a "á"



- Slovo pá



Slabičné segmenty

- Definovány uměle - řešení nejednoznačnosti hranice slabiky.
- Frekventované slabičné typy:
 - V (samohláska/dvojhlaska)
 - KV (souhláska - samohláska)
 - KVK
 - KK
 - KKV
 - KKVK
- Tyto segmenty tvoří 95 % slabik.
- Umožňují automatickou segmentaci textu.
- Využívá např. syntetizér Demosthénés.

Průběh syntézy v časové oblasti

- 1 Fonetický přepis textu.
- 2 Segmentace textu podle délky použitých segmentů.
- 3 Výběr odpovídacích akustických segmentů.
 - odpovídá textu
 - pokud možno spojitě a hladce navazuje na předchozí segment
 - ...
- 4 Spojení segmentů
 - Nutné pokud možno spojitě hladké navázání segmentů.
 - Vhodné zohlednit při výběru segmentů.
- 5 Postprocessing
 - Doplnění prozodie.

Další typy syntézy v časové oblasti

- Korpusová syntéza.
 - Jako databázi segmentů využívá řečový korpus.
 - Označovaná rozsáhlá db mluvené řeči.
 - Značkování obsahuje - fonetický přepis dané řeči, hranice řečových segmentů, průběh F_0 a příp. dalších formantů.
 - Umožňuje přesnější výběr řečových segmentů - nižší náročnost spojování segmentů.
- Syntéza na bázi rámců - problémově orientovaná syntéza.
 - Využívá:
 - rámce - neměnicí se části vět
 - sloty - měnicí se části promluvy.
 - Využití - hlášení nádražního rozhlasu, automatické telefonní systémy např. telekomunikačních operátorů, ...

Ukázky syntézy řeči v časové oblasti dostupné na webu

- AT&T Labs Natural Voices Text-to-Speech Demo
- Free demo to create avatars using Text-to-Speech (TTS) by SitePal
- Cepstral Text-to-Speech
- Festival Online Demo
- SpeechTech Demo TTS
- MBrola Home Page a MBrola Christmas song

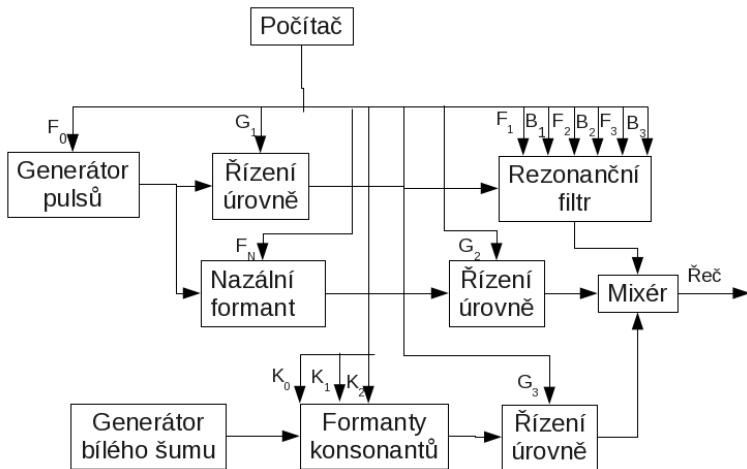
Syntéza řeči ve frekvenční oblasti

- Simuluje tvorbu hlasu v řečových orgánech.
- Uchovává se:
 - frekvenční charakteristika hlasu použitého pro syntézu
 - parametry buzení
- Princip:
 - Emulace hlasových orgánů s využitím:
 - frekvenčních generátorů
 - generátorů šumu
 - filtrů
 - zesilovače
 - Komponenty jsou ovládány parametry modelu.
- Využívají se dva typy kódování zdroje:
 - řečová syntéza formantového typu
 - LPC řečová syntéza.

Řečová syntéza formantového typu

- Rekonstruuje formanty hlasového traktu pomocí spojení několika rezonančních obvodů.
- Jejich frekvence a šířky pásma jsou ovládány elektronicky.
- Parametry syntetizéru:
 - F_0 - základní frekvence
 - F_i - formanty
 - F_N - nazální formant
 - B_i - pásmové filtry pro formanty F_i
 - G_i - parametry řízení zisku (zesílení)
 - K_i - formanty pro konsonanty.

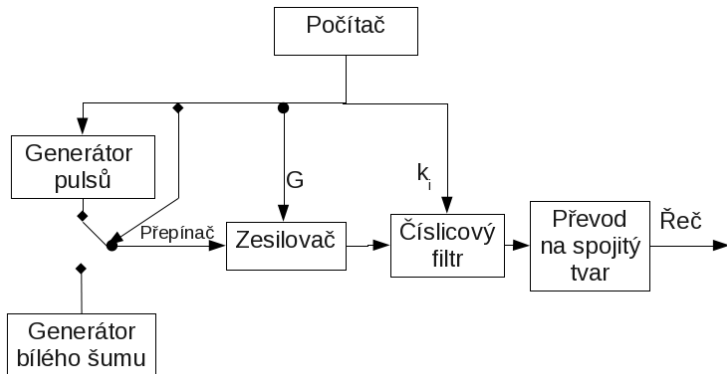
Schéma syntetizéru formantového typu



LPC syntetizér

- Charakteristiky pro LPC syntetizér:
 - perioda základního hlasivkového tónu T_0
 - charakteristika hlásky - znělé/neznělé
 - amplituda budícího signálu G
 - koeficienty číslicového filtru.
- Způsob získání koeficientu číslicového filtru:
 - vrcholy v LPC spektrální obálce analyzovaného mikrosegmentu
 - kořeny charakteristické rovnice zdrojového filtru
 - reflexní koeficienty.

Schéma LPC syntetizéru



Shrnutí

- Výhody syntézy ve frekvenční oblasti:
 - malé paměťové nárok - model mluvčího
 - syntézu lze realizovat hardwarově.
- Nevýhody:
 - Hlas bývá méně přirozený oproti syntéze v časové oblasti.
 - Problém přesnosti matematického modelu.
 - Softwarová syntéza ve frekvenční oblasti bývá výpočetně náročnější než syntéza v časové oblasti.
- Obvyklé využití:
 - doplnění syntézy v časové oblasti o prozodické faktory.
 - občas pro syntézu na zařízeních, která nedisponují dostatečnou kapacitou paměti (mobilní telefony, PDA, ..)