

Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta infomatiky
Masarykova univerzita

podzim 2019

Obsah

1 Syntéza řeči v časové oblasti

Syntéza v časové oblasti

- Princip
 - spojování navzorkovaných řečových segmentů uložených v databázi.
- Využívají se různé typy základních segmentů:
 - větší
 - lépe se modelují některé další charakteristiky jako intonace, přízvuky
 - větší nároky na paměť – větší množství segmentů (potenciálně až 2^n , kde n je délka segmentu)
 - příklady – slova, části vět
 - menší
 - menší paměťové nároky – menší množství segmentů
 - horší možnost modelování větné intonace, přízvuků, ... (viz oblasti spektrální stacionarity řeči).

Používané řečové segmenty

- Alofóny
 - poziční varianty fonémů – obsahuje i části okolních fonémů
 - počet n^3 (n – počet fonémů)
- Difóny
 - začínají uprostřed jednoho fonému a končí uprostřed následujícího
 - počet n^2
 - často využívané pro syntézu i rozpoznávání:
 - MBrola, ...
- Trifóny
 - začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního
 - počet n^3
 - často využívané pro rozpoznávání a syntézu
- Slabičné segmenty.
- Segmenty proměnné délky získané z korpusu.
- Rámce

Slabiky

- Slabika
 - Slabikovat se učí už děti v první třídě.
 - Nejmenší jednotka organizační jednotka řeči.
 - Nelze odvodit strukturu slabik – nejednoznačnost dělení některých slov na slabiky
 - funk-ční vs funkč-ní.
 - Počet slabik – uvádí se cca 10000.
 - Struktura slabiky
 - preatura (onset)
 - nukleus (vokalické jádro) – bývá to samohláska, příp. dvojhlaska, sonora – např. **krk**, frikativa – např. **pst**, nazála – např. **sedm**
 - koda – nemusí se vyskytovat
 - nukleus + koda jsou považovány za základ slabiky
 - svahy – preatura a koda; jedná se většinou o jednu nebo více souhlásek

Slabičné segmenty

- Definovány uměle
- Řešení nejednoznačnosti hranice slabiky.
- Frekventované slabičné typy:
 - V (samohláska/dvojhlaska) – ú – kol
 - KV (souhláska – samohláska) – vo – da
 - KVK – jed-not-ka
 - KK – tr-sy
 - KKV – dna
 - KKVK – dmout
- Tvoří více než 95
- Umožňují automatickou segmentaci textu.
- Používají se např. v syntetizéru Demosthénés (doc. Kopeček LAF (LSD) FI)

Vlastní syntéza

- 1 Fonetický přepis.
- 2 Segmentace dle použitých řečových segmentů.
- 3 Výběr odpovídajících akustických segmentů
 - databáze segmentů.
- 4 Spojení segmentů
 - nutné, aby odpovídala F_0 – jinak se vyskytnou různé ruchy (lupnutí, ...)
 - vhodné řešit už při vytváření db segmentů.
- 5 Případný postprocessing

Korpusová syntéza

- Konkatenativní syntéza v časové oblasti.
- Jako db segmentů využívá řečový korpus.
- Nutno doplnit značky pro syntézu:
 - fonetický přepis
 - hranice řečových segmentů
 - průběh F_0
 - ...
- Umožňuje přesnější výběr segmentů
 - snižuje výpočetní složitost spojování a postprocessingu.
- Příklad – viz dizertační práce dr. Batůška v knihovně FI.

Syntéza na bázi rámců

- Většinou se jedná o problémově orientovanou syntézu.
- Syntéza se skládá z:
 - rámců – neměnicí se části vět
 - slotů – měnicí se části promluvy
- Výhoda:
 - rámce jsou dopředu namluveny a mohou obsahovat intonaci
 - syntetizuje se pouze obsah slotů
 - omezená množina
 - lze použít celá slova
- Příklady:
 - hlášení nádražního rozhlasu:
 - Osobní vlak číslo <číslo_vlaku> ze směru <seznam_stanic> přijede k <číslo_nástupiště>. nástupišti v <čas>.