

Úvod do počítačového zpracování řeči

Luděk Bártek¹

1

Obsah

- 1 Syntéza řeči v časové oblasti
- 2 Syntéza řeči - postprocessing

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 $2n$, 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 $n3$ (n – počet fonémů)
- Difóny
 - začínají uprostřed jednoho fonému a končí uprostřed následujícího
 - počet $n2$
 - často využívané pro syntézu i rozpoznávání:
 - MBrola
(<http://tcts.fpms.ac.be/synthesis/mbrola.html>)
- Trifóny
 - začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního
 - počet $n3$
 - 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. sed**m**
 - 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 jsem 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>`.

Prozodie

- Výstupem syntézy je monotóní hlas bez intonace a přízvuku – zní nepřirozeně
- Doplnění prozodie
 - základní prozodické prvky:
 - výška
 - hlasitost
 - doba trvání
 - nositelem je slabika
 - Větná intonace (prozodie) – závisí na typu věty:
 - otázky zjišťovací (odpověď ano/ne) – rostoucí
 - oznamovací, tázací doplňovací, rozkazovací – klesající
 - řeší se modulací F_0
 - Doplnění přízvuku/důrazu
 - modifikace F_0 a intenzity
 - lokální modifikace větné melodie

Prozodie – ukázky větné intonace

- Originální promluva (data/masse.wav)
- Oznamovací věta (data/masse-ozn.wav)
- Otázka zjišťovací (data/masse-dotaz.wav)

Výška základního tónu

- Výška základního tónu odpovídá formantu F_0 .
- Průběh F_0 na vokalickém jádru bývá nelineární.
- Změna intonace není pouhou změnou F_0
 - nutno modifikovat i vyšší formanty.
- Na základě důležitosti F_0 se jazyky dělí na:
 - tónové (čínština, vietnamština, ...)
 - čínské slovo -ma- v závislosti na průběhu F_0 může znamenat matka, konopí, kůň, nadávat
 - jazyky s melodickým přízvukem (srbština, slovinština, litevština, norština, švédština, ...)

Další prozodické vlastnosti

- Intenzita (hlasitost):
 - fyzikální pohled – intenzita signálu v daném časovém okamžiku
 - fyziologický pohled – reakce vnitřního ucha (cortiho ústrojí) na vnímaný zvuk.
 - Tato hlediska se různí.
 - Subjektivní vnímání zvuku neodpovídá ani v prvním přiblížení fyzikální intenzitě signálu.
- Doba trvání:
 - Slabika může mít různou dobu trvání v různém kontextu.
 - Drobné odchylky mohou být i ve stejném kontextu.
 - Typická doba trvání slabiky 50 — 200 milisekund.

Další prozodické vlastnosti

- Kvalita hlasu
 - chvění hlasu (jitter)
 - nepravidelné výchylky v amplitudě F_0 (shimmer)
 - zbarvení tónu
 - ochraptělost
 - níra znělosti
 - ...
- Rychlost řeči
 - Lze chápat jako převrácenou hodnotu průměrné délky slabiky
 - Lze měřit i jinými způsoby:
 - počtem vyslovených textových znaků za jednotku času (vyhodnocování syntetizérů řeči).

Další prozodické vlastnosti

Pokračování

- Pauza
 - tichá
 - vyplněná – obsahuje nějaký charakteristický zvuk (např. eeh)
 - ztížená detekce – hlavní format je blízký formantům samohlásek "a", "e".
- Zaváhání
 - Přímo vypovídá o pragmatice projevu.
 - Důležitý např. pro modifikaci dialogové strategie u dialogových systémů.
 - Typický případ informace obsažené zejména v prozodické vrstvě jazyka.

Základní odvozené prozodické vlastnosti

- Rytmus (časování):
 - Prozodický prvek odvozený z dob trvání
 - slabik
 - pauz v daném časovém úseku.
- Slovní přízvuk
 - Je odvozen ze všech základních atributů.
 - Je výrazně jazykově závislý:
 - umístění přízvuku ve slově/přízvučné jednotce
 - míra použití prozodických prostředků k jeho vyjádření zejména použití hlasitosti proti výšce.
- Větný přístup (intonační centrum):
 - zjednodušeně jde o prozodické zvýraznění jádra výpovědi věty

Základní odvozené prozodické vlastnosti (2.)

- Intonace
 - nejobecnější – časový průběh zvukového spektra hlasu
 - za určující pro melodii se obvykle považuje základní hlasová frekvence – lze zobrazit grafem v závislosti na čase
 - časová závislost základní hlasové frekvence
 - související terminologie:
 - melodie
 - kadence
 - intonační kadence
 - melodém
 - průběh F0
- Emotivní zabarvení hlasu
 - projevuje se:
 - rychlými změnami hlasitosti a základní frekvence
 - Často přesahují hranici věty.
 - Detekce je důležitá např. pro dialogové systémy – umožňuje zvolit vhodnou dialogovou strategii.

Základní odvozené prozodické vlastnosti (3.)

- **Emfatický přízvuk**
 - Vytvářen emotivním zbarvením hlasu.
 - Vyskytuje se např. ve větách pronesených v situacích s výrazným emocionálním kontextem, např.
 - To je tedy opravdu **neslýchané**.
 - Bolí to jak **čert**.
- **Kontrastní přízvuk**
 - snaha o zdůraznění slova nebo slabiky v kontrastu s jiným slovem nebo slabikou během promluvy nebo dialogu:
 - "řekl jsem do **Šakvic** ne **Rakvic**"
 - "**byte** ne **bit**"

Základní odvozené prozodické vlastnosti (4.)

- Opakování
 - prozodický atribut silně svázaný s mluvčím.
 - Opakování bývá často variantou výplňkových částí promluvy – mluvčí si ji často ani neuvědomuje (nezaměňovat s koktáním – porucha řeči).
 - Může se jednat o formu zdůraznění – v krajním případě může být považováno za vadu řeči.
- Výplňkové části
 - kromě výplňkové funkce mohou charakterizovat
 - styl mluvčího: „Byl jsi včera na akci, **viď?**”
 - nářečí resp. slang: „**Vole**, ta včerejší spářka byla hustá, že **vole?**”

Základní odvozené prozodické vlastnosti (5.)

- Přerušení:
 - častý jev v mluvené řeči na úrovni:
 - vyšších celků (výpověď/promluva, věta, prozodická fráze, ...)
 - uvnitř slov.
 - Mívá návaznost na další prosodické prvky:
 - zaváhání
 - opakování
 - vyplněnou pauzu
 - ...
 - Zvyšuje obtížnost rozpoznávání mluvené řeči – nutno s ním počítat.
- Korekce částí promluvy:
 - Častý jev a to vzhledem k rozdílným částem.
 - Příčiny vzniku:
 - důsledek přeřeknutí,
 - upřesnění předchozí části promluvy,
 - oprava předchozí části promluvy.
 - Často následuje přerušení nebo další prozodické jevy.

Prozodické segmenty mluvené řeči

- Prozodické segmenty mluvené řeči:
 - Promluva.
 - Prozodická fráze
 - Skupina slov vytvářející jednotný intonační celek.
 - Představuje základní, z prozodického hlediska kompaktní strukturu.
 - Členění do prozodických frází ve velké míře souvisí se syntaktickou strukturou odpovídající věty.
 - Přízvukový takt
 - skupina slabik podřízená jednomu slovnímu přízvuku.
 - V češtině typicky slovo nebo slovo a jednoslabičné slovo.
 - Slabika

Standardy pro syntézu řeči

- Snaha sjednotit jazyky pro popis promluvy pro řečové syntetizéry.
- Definují značkování postihující:
 - prozódii
 - rychlost řeči
 - F_0
 - zdůraznění části promluvy
 - pauzu
 - hlasitost
 - ...
 - mluvčího
 - pohlaví
 - věk
 - ...
 - ...
- Používané standardy:
 - SABLE
 - SSML

SABLE

- Vývoj započat v 2. polovině 90. let
- aplikace XML/SGML
- snaha o zkombinování 3 značkovacích jazyků pro syntézu řeči:
 - SSML – Speech Synthesis Markup Language (<http://www.w3.org/TR/2010/REC-speech-synthesis11-20100907/>) (W3C, 1999)
 - STML – Spoken Text Markup Language (<http://www.bell-labs.com/project/tts/stml.html>) (CSTR Edinburgh University, Lucent Technologies, 1997)
 - JSML – Java Synthesis Markup Language (<http://www.w3.org/TR/jsml/>) (Sun Microsystems, 2000)
- <http://www.bell-labs.com/project/tts/sable.html>

SABLE

Základní značky

- SABLE – kořenová značka
- div – slouží k logickému členění dokumentu (odstavec, věta)
- prozodické:
 - EMPH – zdůraznění části promluvy
 - PITCH – výška promluvy
 - VOLUME – úroveň hlasitosti
 - RATE – rychlost
 - BREAK – pauza
- popis hlasu:
 - SPEAKER – popisuje pohlaví a věk mluvčího
- fonetické
 - PRON – výslovnost – fonetický přepis
 - SAYAS – způsob fonetického přepisu (datum, telefon, url, poštovní adresa, ...)
 - LANGUAGE – jazyk promluvy

SABLE – ukázka

```
<SABLE>
  <DIV TYPE="paragraph">
    <VOLUME LEVEL="quiet">Šepot.</VOLUME>
    <VOLUME LEVEL="medium">
      <RATE SPEED="fast">Rychlá věta.</RATE>
      <PITCH BASE="+50%">Vysoko posazená věta</PITCH>
    </VOLUME>
  </DIV>
</SABLE>
```

SSML

- Vývoj započat v koncem 90. let
- součást W3C Voice Browser Activity
(<http://www.w3.org/Voice>)
- Aktuální verze 1.0 (září 2004)

SSML

Základní značky

- kořenový element – speak
- strukturní elementy
 - p – odstavec
 - s – věta
- fonetické:
 - say-as – způsob fonetického přepisu (výslovnosti, datum, telefon, url, číslo, ...)
 - phoneme – fonetický přepis dané promluvy
 - sub – substituce (např. přepis zkratek, ...)
- popis hlasu:
 - voice – popis hlasu, kterým se má text přečíst (pohlaví, věk, ...)
- prozodie:
 - emphasis – zdůraznění částí promluvy
 - break – pauza
 - prosody – ovlivňuje prozodické jevy: výšku, průběh základní frekvence, rychlost, item délka trvání promluvy, hlasitost.

SSML

Ukázka

```
<?xml version="1.0"?>  
<ssml>  
  <voice gender="female">Female voice.</voice>  
  <voice gender="male">Male voice.</voice>  
  <emphasis level="soft">Soft emphasis</emphasis>  
  <p>Speech with 5 seconds <break time="5s"/> break.</p>  
  <prosody volume="+6dB">Speech at double volume.</prosody>  
  <prosody volume="-6dB">Speech at half volume.</prosody>  
</ssml>
```