

Dialogové systémy

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,
Brno

jaro 2021

- Cíl – převod psaného textu na mluvenou řeč.
 - Výsledná řeč by měla znít co nejpřirozeněji.
- Přirozená řeč by měla obsahovat:
 - správnou intonaci
 - správné umístění přízvuků
 - slovní
 - větný
 - korektní koartikulaci
 - správný rytmus (časování)
 - ...

Druhy syntézy řeči

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Syntéza ve frekvenční oblasti – simuluje chování řečového ústrojí.
- Syntéza v časové oblasti – spojování řečových segmentů do větších celků (věta, promluva, ...)
- Korpusová – varianta syntézy v časové oblasti – jako databáze řečových segmentů slouží řečový korpus.
- Problémově orientovaná syntéza:
 - varianta syntézy v časové oblasti
 - využívá větší celky – věty, ...
 - příklady:
 - hlášení nádražního rozhlasu
 - automatizované linky telefonické podpory
 - ...

Fáze syntézy řeči

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- 1 Fonetický přepis textu.
- 2 Syntéza foneticky přepsaného textu:
 - Syntéza ve frekvenční oblasti – volba průběhu parametrů syntézy (F_0 /generátor šumu, vyšší harmonické frekvence, jejich intenzita, ...)
 - Syntéza v časové oblasti – výběr vhodných segmentů a jejich spojení.
- 3 Případný postprocessing:
 - doplnění intonace
 - doplnění přízvuků
 - ...

- Slouží k přesnému, jednoznačnému zápisu mluvené řeči.
- Využívá fonetickou abecedu:
 - mezinárodní fonetická abeceda (IPA) – součást standardu UNICODE
 - SAMPA (Speech Assessment Method Phonetic Alphabet)
 - sedmibitový přepis IPA
 - navržena v 80. letech
 - používá se v různých TTS
 - příklad:

tSeSTina je kra:sni: jazik

- ...

CONSONANTS (PULMONIC)

	LABIAL		CORONAL				DORSAL			RADICAL		LARYNGEAL				
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Epi-glottal	Glottal				
Nasal	m	ɱ	n			ɳ	ɲ	ŋ	ɴ							
Plosive	p b	ɸ β	t d			ʈ ɖ	c ɟ	k ɡ	q ɢ				ʔ	ʔ		
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ	ʕ	ħ	ʕ	h	ɦ	
Approximant		ʋ	ɹ			ɻ	j	ɰ								
Trill	ʙ		r						ʀ				ʀ			
Tap, Flap		ⱱ	ɾ			ɽ										
Lateral fricative			ɬ ɮ		ɮ	ɬ	ɬ	ɮ								
Lateral approximant			l		ɭ	ɭ	ʎ	ʎ								
Lateral flap			ɺ		ɻ	ɻ										

Where symbols appear in pairs, the one to the right represents a modally voiced consonant, except for murmured ɦ. Shaded areas denote articulations judged to be impossible. Light grey letters are unofficial extensions of the IPA.

- Nelze si pamatovat fonetické přepisy všech promluv:
 - Nutno zabezpečit automatický přepis.
- Pravidla fonetického přepisu:
 - mohou mít regionální charakter.
 - Příklad – výslovnost na shledanou v ČR:
 - Čechy – naschledanou
 - Morava – nazhledanou.
 - Obě varianty jsou spisovné.
 - Obecně přepis nemusí používat všechny znaky dané abecedy (i/y = i, c = ts, ...)
- Zohledňuje koartikulaci (spodobu znělosti).

Pravidla fonetického přepisu češtiny

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- $ch \rightarrow x$, $ů \rightarrow ú$, $w \rightarrow v$, $q \rightarrow kv$, $y \rightarrow i$, $ý \rightarrow í$
- ě:
 - $bě \rightarrow bje$, $pě \rightarrow pje$, $fě \rightarrow fje$, $vě \rightarrow vje$
 - $dě \rightarrow d'e$, $tě \rightarrow t'e$, $ně \rightarrow ňe$, $mě \rightarrow mňe$
- i/í:
 - $di/í \rightarrow d'i/í$, $ti/í \rightarrow t'i/í$, $ni/í \rightarrow ňi/í$
- X:
 - $x \rightarrow ks$ — začátek slova před samohláskou, mezi samohláskami nebo před neznělou souhláskou a nebo na konci slova.
 - $x \rightarrow gz$:
 - *exsamohláska*
 - před znělou souhláskou

Změny při spojování souhlásek

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Dochází k nim při spojování souhlásek.
- Způsobeny přenastavováním mluvidel.
- Dva druhy:
 - spodoba znělosti – změna znělosti párových souhlásek:
 - ZPS → ∅ ZPS: dub → dup, zpěv → spjev
 - NPS → ∅ NPS: sběr → zbjer, když → gdiš
 - spodoba artikulační – při spojení dvou souhlásek s různou artikulací:
 - nk/ng – banka, tango
 - mv/mf – tramvaj, nymfa
 - nt'/nd – punťa, pindík
 - dň – odpovědně, sto dní, vodní
 - ts → c
 - tš → č
 - ds → c
 - dš → č

Syntéza řeči ve frekvenční oblasti

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
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
 - filtrů
 - zesilovače (zesilovačů).
 - Tyto komponenty jsou ovládány parametry modelu.
- Využívají se následující způsoby kódování zdroje:
 - řečová syntéza formantového typu
 - LPC řečová syntéza
 - syntéza postavená na HMM
 - ...

Řečová syntéza formantového typu

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Rekonstruuje formanty hlasového traktu pomocí sériových a paralelních 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 F_i
 - G_i – parametry řízení zisku/zesílení
 - K_i – formanty pro konsonanty

Schéma sériového formantového syntetizéru

Dialogové systémy

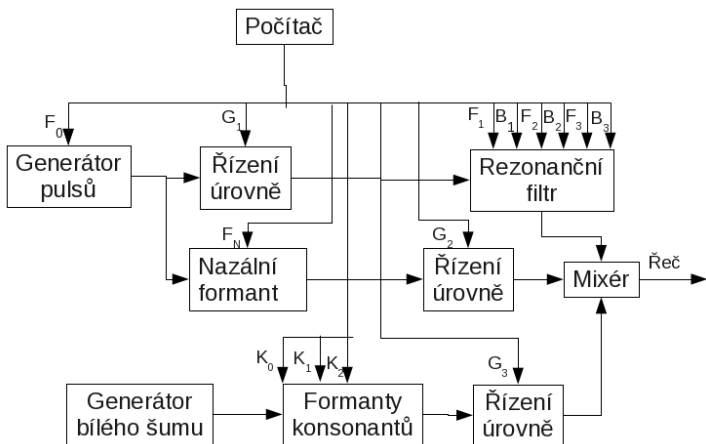
Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti



Obrázek: Blokové schéma sériového formantového syntetizéru

- Charakteristiky pro LPC syntetizér:
 - perioda základního hlasivkového tónu T_0
 - charakteristika hlásky – znělá/neznělá
 - 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

Dialogové
systémy

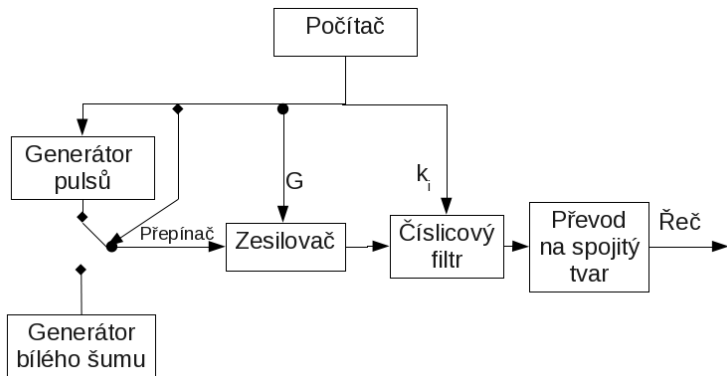
Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti



Obrázek: Blokové schéma LPC syntetizéru

Syntéza ve frekvenční oblasti

Shrnutí

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Výhody a nevýhody syntézy ve frekvenční oblasti:
 - + Malé paměťové nároky – model použitého mluvčího.
 - + Syntézu lze realizovat hardwarově.
 - 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:
 - větnou intonaci
 - větný a slovní přízvuk
 - další prozodické faktory.
 - Občas pro syntézu na zařízeních, která nedisponují dostatečnou kapacitou paměti (mobilní telefony, PDA, ...).
 - Občas pro multilinguální syntézu.
- Více viz např. J. Psutka – Komunikace s počítačem mluvenou řečí.

Syntéza v časové oblasti

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Cíl – převod obecného textu na řeč.
- Postavena na spojování segmentů řeči.
- Využívají se různé délky základních segmentů:
 - Větší:
 - lze lépe modelovat prozodické charakteristiky řeči
 - větší paměťové nároky – větší množství segmentů (potenciálně až 2^n , kde n je délka segmentu).
 - příklady segmentů – slova, části vět, věty, ...
 - Menší:
 - horší možnost modelování prozodických jevů (větná intonace, přízvuky, ...)
 - menší paměťové nároky – menší množství menších segmentů.

Používané řečové segmenty

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Alofóny:
 - poziční varianty fonémů – obsahují
 - foném
 - okolí ovlivněné koartikulací.
 - počet alofónů – n^3 (n – počet fonémů).
- Difóny:
 - začínají uprostřed jednoho fonému a končí uprostřed následujícího fonému
 - počet difónů – n^2
 - často využívané pro syntézu i pro rozpoznávání (např. syntetizér MBrola)

Používané řečové segmenty

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Trifóny:
 - Začínají uprostřed levého sousedního fonému a končí uprostřed pravého sousedního fonému.
 - Počet – n^3 .
 - Často používané pro rozpoznávání a syntézu řeči.
- Slabičné segmenty:
 - Snaha, aby co nejvíce odpovídaly slabikám.
 - Délka – 1 — 3 fonémy.
 - Využívá se např. v TTS systému Demosthénés.

- Slabika:
 - Slabikovat se učí už děti v první třídě.
 - Nejmenší organizační jednotka řeči.
 - Nelze odvodit strukturu slabik – nejednoznačnost dělení některých slov na slabiky:
 - funk-ční vs. funkč-ní
 - Počet slabik – uvádí se cca 10 000.

- **Struktura slabiky:**
 - preatura (onset)
 - nukleus (vokalické jádro slabiky) – bývá to:
 - samohláska resp. dvojhláska
 - sonora – např. *krk*
 - frikativa – např. *pst*
 - nasála – 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.

Syntéza v časovém pásmu

Slabičné segmenty

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- 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 – tma
 - KKVK – dmout
- Tyto segmenty tvoří více než 95 % slabik.
- Umožňují automatickou segmentaci textu.
- Používají se např. v syntetizéru Demosthénés (doc. Kopeček, LSD FI)

- 1 Fonetický přepis.
- 2 Segmentace textu dle použitých řečových segmentů.
- 3 Výběr odpovídajících akustických segmentů z db segmentů.
- 4 Spojení segmentů
 - Nutné, aby bylo možné spojitě hladké navázání segmentů:
 - shodné nebo velmi blízké hodnoty konce a začátku po sobě jdoucích segmentů
 - shodné nebo velmi blízké hodnoty 1. derivace navazujících konců obou segmentů.
- 5 Případný postprocessing
 - doplnění prozódie.
 - ...

Syntéza v časové oblasti

Korpusová syntéza

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Konkatenativní syntéza v časové oblasti.
- Jako databází segmentů využívá řečový korpus.
 - Obsahuje označovanou mluvenou řeč.
 - Značkování obsahuje:
 - fonetický přepis dané řeči
 - hranice řečových segmentů
 - průběh F_0 a případně i dalších formantů.
 - Umožňuje přesnější výběr řečových segmentů:
 - snižuje výpočetní složitost spojování a postprocessingu.
- Algoritmus výběru segmentů:
 - 1 Výběr odpovídajícího segmentu podle požadovaného obsahu.
 - 2 Pokud je segmentů více zvolí se z nich ten, který nejlépe navazuje.

Syntéza v časové oblasti

Syntéza na bázi rámců

Dialogové
systémy

Luděk Bártek

Syntéza řeči

Fonetický přepis

Syntéza řeči ve
frekvenční oblasti

Syntéza řeči v časové
oblasti

- Většinou se jedná o problémově orientovanou syntézu.
- Syntéza se skládá z:
 - rámců – neměnicí se části vět
 - slotů – měnicí se částí promluvy.
- Výhody:
 - Rámce jsou dopředu namluveny a mohou obsahovat intonaci.
 - Syntetizuje se pouze obsah slotů:
 - velmi dobře specifikovaná množina slov
 - lze použít celá slova.
- Příklad:
 - hlášení nádražního rozhlasu:

Osobní vlak číslo *číslo vlaku* ze směru *stanice* přijede
k *číslo nástupiště* nástupišti v *čas*.