

Syntéza řeči

**Přednáška z předmětu Neuronové sítě pro humanitní studia
(NEUH)**

Jindřich Matoušek

Osnova

- q Úvod
- q Historie syntézy řeči
- q Akustický model vytváření řeči
- q Formantová syntéza
- q Konkatenáční syntéza
- q Artikulační syntéza
- q Syntéza řeči z textu (TTS)

Osnova

ü Úvod

- q Historie syntézy řeči
- q Akustický model vytváření řeči
- q Formantová syntéza
- q Konkatenáční syntéza
- q Artikulační syntéza
- q Syntéza řeči z textu (TTS)

Motivace

- q řeč – nejpřirozenější forma komunikace mezi lidmi, činnost člověku vlastní a přirozená
- q syntéza řeči – důležitá oblast zpracování řečového signálu
- q syntéza řeči = proces umělého vytváření řeči (počítačem)
- q počítačová syntéza řeči si klade za cíl „zpřirozenit“ komunikaci člověka s počítačem
- q konečný cíl: vytvářet řeč v takové formě a kvalitě, aby nebyla rozpoznatelná od řeči člověka

Lidská komunikace

- q **písmo** – psaná podoba komunikace
 - § věty, slova, písmena
- q **řeč** – mluvená podoba komunikace
 - § akustika
 - vytváření a vnímání řeči
 - akustické vlastnosti řeči (formanty, způsob a místa tvoření řeči,...)
 - § fonetika a fonologie (promluvy, slova, hlásky, fonémy, alofóny)
 - § lingvistika (věty, gramatika, syntaxe, sémantika, ...)
 - § prozodie (melodie/intonace, trvání/rychlost, hlasitost/energie)
- q **fonetická** informace (posloupnost hlásek)
 - § *jaká* řeč se má vytvořit (význam)
- q **prozodická** informace (melodie, trvání/rychlost, hlasitost promluvy)
 - § *jak* se má řeč vytvořit (věta oznamovací, tázací, ...)

Úrovně popisu mluvené řeči

- q **artikulační** – činnost hlasivek, pohyby artikulátorů, způsob a místa tvoření řeči
- q **akustická**
 - § vytváření a vnímání zvuku
 - § akustické vlastnosti řeči (formanty, antiformanty)
- q **fonetická**
 - § výslovnostní podoba jazyka; bez ohledu na význam, který řeč přenáší
 - § „zvuky“ řeči – hlásky (phones)
- q **fonologická**
 - § výslovnostní podoba jazyka + souvislost s významem promluvy
 - § „most“ mezi akustikou a lingvistikou vyšší úrovně
 - § základní jednotky: fonémy, alofóny
- q **morfologická** – skladba slov (odděleně od ostatních slov)
- q **syntaktická** – struktura vět
- q **sémantická** – význam promluvy
- q **pragmatická** – souvislost významu sdělení a záměru řečníka

Syntetizér řeči

- q zařízení pro umělé vytváření řeči
- q jádro každého systému konverze textu na řeč (text-to-speech – TTS)
- q systém na základě vstupní informace vytváří řeč
- q **vstup**: fonetická a prozodická informace
- q **výstup**: řeč



Základní přístupy k syntéze řeči

q artikulační syntéza

§ komplexní řešení, modelování celého procesu vytváření řeči

§ prakticky se zatím nevyužívá

q formantová syntéza

§ zjednodušené modelování hlasového traktu pomocí formantů

§ praktické aplikace TTS (60-80. léta)

q konkatenáční syntéza (řetězení)

§ řetězení segmentů řeči, využívá inventář řečových jednotek

§ současné TTS

Rozdělení metod syntézy řeči

q podle principu

§ systémový přístup (artikulační syntéza)

§ signálový přístup (formantová a konkatenáční syntéza)

q podle míry zapojení člověka:

§ syntéza podle pravidel (ruční nastavení parametrů syntetizéru)

– většinou formantová a artikulační syntéza

§ syntéza řízená daty (automatické nastavení parametrů syntetizéru na základě reálných dat)

– většinou konkatenáční syntéza

q podle oblasti, v níž syntéza probíhá:

§ časová oblast (neparametrická)

§ frekvenční oblast (parametrická)

Osnova

ü Úvod

ü Historie syntézy řeči

q Akustický model vytváření řeči

q Formantová syntéza

q Konkatenáční syntéza

q Artikulační syntéza

q Syntéza řeči z textu (TTS)

Vývoj syntézy řeči

- q mechanické syntetizéry
- q elektrické syntetizéry
- q digitální syntetizéry

Osnova

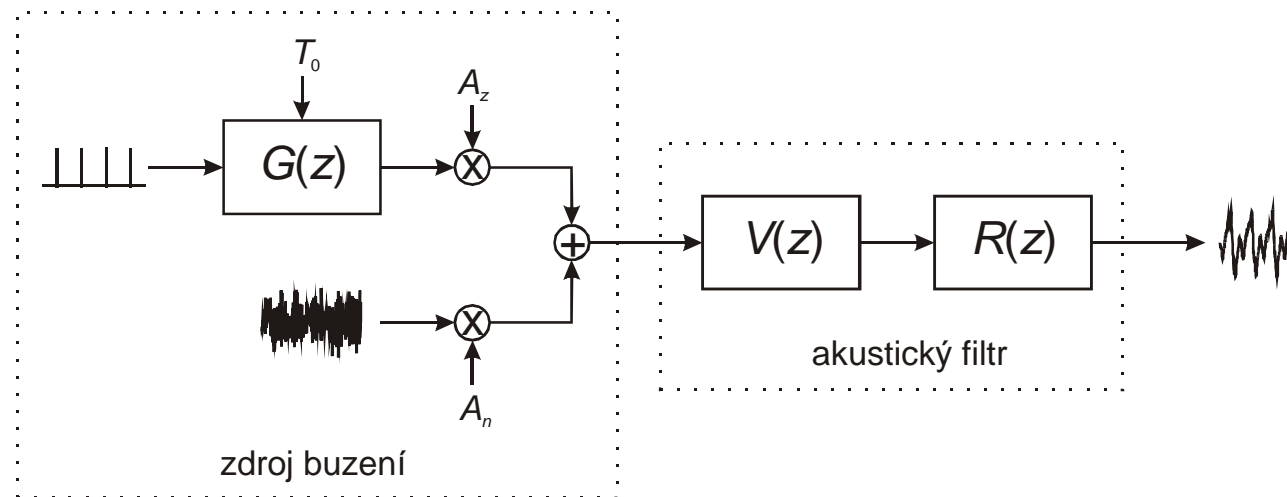
- ü Úvod
- ü Historie syntézy řeči
- ü Akustický model vytváření řeči
 - q Formantová syntéza
 - q Konkatenáční syntéza
 - q Artikulační syntéza
 - q Syntéza řeči z textu (TTS)

Princip

- q akustický model lineární produkce řeči
- q je na něm založena formantová syntéza a většina současných parametrických metod syntézy řeči
- q spíše přístup k modelování řeči než konkrétní model
- q nepopisuje interakce mezi zdrojem buzení a filtrem reprezentujícím hlasový trakt
- q v praxi jisté vzájemné interakce existují
 - § např. hlasový trakt může zpětně působit na vibrace hlasivek (a naopak)
- q někdy těžké oddělit charakteristiky hlasového traktu od zdroje
 - § zvláště ženské a dětské hlasy (frekvence základního hlasivkového tónu F_0 a prvního formantu F_1 mohou být podobné)

Vytváření řeči

- q vytváření řeči modelováno 2 navzájem nezávislými složkami (source-filter theory)
- q zdroj buzení:
 - § kvaziperiodický sled hlasivkových pulsů pro znělé zvuky
 - § náhodný šum pro neznělé zvuky
 - § možnost smíšeného buzení
- q lineární akustický filtr reprezentující frekvenční odezvu hlasového traktu



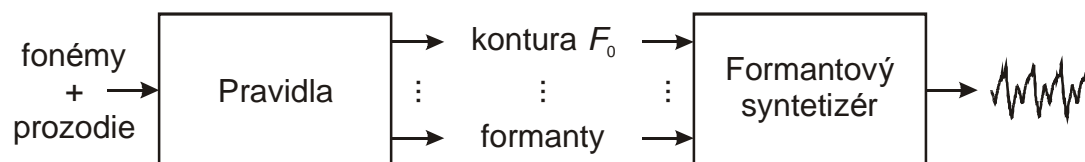
Osnova

- ü Úvod
- ü Historie syntézy řeči
- ü Akustický model vytváření řeči
- ü Formantová syntéza
- q Konkatenáční syntéza
- q Artikulační syntéza
- q Syntéza řeči z textu (TTS)

Princip

- q založena na akustické teorii vytváření řeči
- q zjednodušená simulace procesu vytváření řeči člověkem:
 - § **zdroj buzení**: generátor impulsů pro znělé zvuky a šum pro neznělé zvuky (+ smíšené buzení)
 - § **hlasový trakt**: modelování pomocí filtru, jehož parametry jsou spjaty zejména s formanty hlasového traktu
- q syntéza podle pravidel – parametry se nastavují na základě manuálně nalezených pravidel
- q dříve úspěšná a používaná metoda syntézy řeči
- q dnes se téměř nepoužívá (výjimka: DECtalk)

(OVE, Fant 1953)



Pravidla

- q pravidla pro nastavení vývoje kontury základního hlasivkového tónu (F_0), formantových frekvencí a pásem, trvání, atd.
- q pravidla odhadována z reálných dat
- q časově velmi náročný proces – pravidla odhadují experti (fonetika, fonologie, akustika)
- q pravidla se navzájem ovlivňují – obtížné najít optimální sadu pravidel
- q potřeba rozsáhlého testování

Vytváření řeči

q buzení

- § zdroj znělých zvuků – generátor impulsů s frekvencí F_0
- § zdroj neznělých zvuků – generátor šumu

q hlasový trakt

- § filtr modeluje hlasový trakt pomocí **rezonátorů** či **antirezonátorů**, které simulují **formanty** a **antiformanty** hlasového traktu

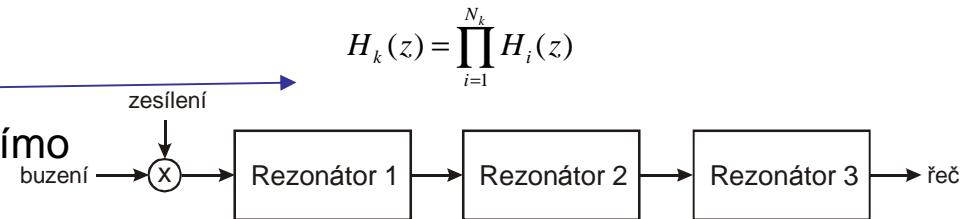
Modelování hlasového traktu

q typicky 4 – 6 (anti-) formantů

q **kaskádní (sériové) zapojení**

§ výstup jednoho rezonátoru aplikován přímo na vstup dalšího

§ vhodné zejména pro vytváření zvuků se silnou harmonickou strukturou (vokály)



q **paralelní zapojení**

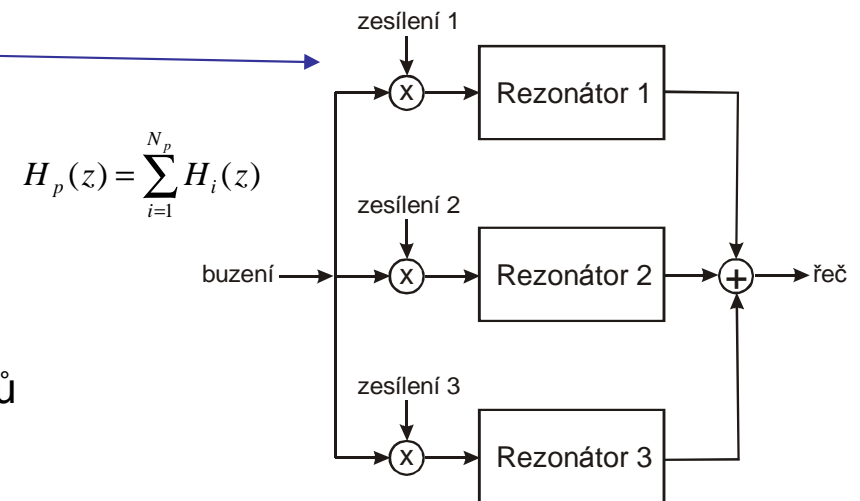
§ buzení aplikováno na všechny rezonátory paralelně

§ výstupy se sčítají

§ nezávislé řízení amplitudy každého formantu

§ více parametrů

§ vhodné zejména pro generování konsonantů



q **smíšené**

§ pro kvalitní modelování řeči je zapotřebí využívat obě konfigurace hlasového traktu – kaskádní i paralelní větve

§ nutno efektivně řídit přepínání mezi jednotlivými větvemi

Výhody a nevýhody

- + malý počet parametrů (40 – 60)
- + jednoduchý, jasný koncepční model
- + snadné řízení prozodických charakteristik
- + konstantní kvalita
- ± spjatost s procesem vytváření řeči člověkem
- ± koartikulační jevy zachyceny v pravidlech (**obtížné!**)
- ± závislost i nezávislost na konkrétním hlasu (pro změnu hlasu **pravidla!**)
- ± změny hlasu a emoce – možno řídit podle pravidel (**pravidla!**)
- ± schopnost vytvářet plynulou kvalitní řeč (ale: **pravidla!**)
- pracné hledání a nastavování pravidel (koartikulace, dynamické zvuky)
- pravidla jsou závislá na realizaci fonému (alofónová pravidla)
- vzájemná interakce mezi hodnotami parametrů
- časová náročnost vývoje systému
- složité vytváření některých zvuků podle pravidel (např. plozivy)
- nízká přirozenost syntetické řeči (vyšší kvalita vyžaduje složitější pravidla – ty je však téměř nemožné určit)

Osnova

- ü Úvod
- ü Historie syntézy řeči
- ü Akustický model vytváření řeči
- ü Formantová syntéza
- ü Konkatenáční syntéza
- q Artikulační syntéza
- q Syntéza řeči z textu (TTS)

Princip

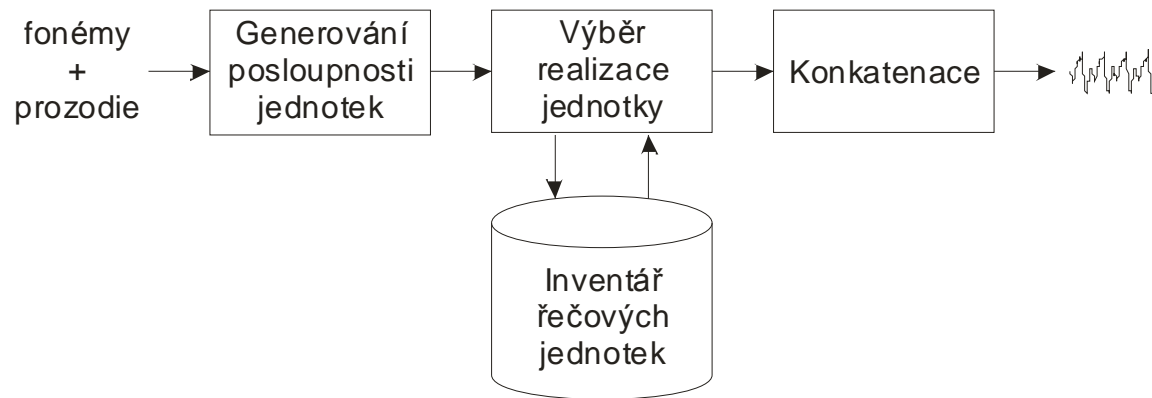
- q používá přímo části přirozeného řečového signálu
- q předpokládá, že řeč se skládá z řečových (akustických) jednotek
- q řeč je pak možné rozdělit na segmenty odpovídající těmto jednotkám a uložit je do **inventáře řečových jednotek**
- q řeč se vytváří řetězením (konkatenací) řečových segmentů uložených v inventáři řečových jednotek
- q syntetická řeč napodobuje řečníka z inventáře

Vlastnosti

- q vytváření inventáře řečových jednotek:
 - § ruční vytváření
 - § automatické vytváření
- q způsob reprezentace řečových jednotek:
 - § neparametrická (přímo vzorky řeči)
 - § parametrická (LPC, kepstrální, HNM)
- q spektrální/prozodické modifikace jednotek:
 - § bez modifikací (pouhé řetězení)
 - § s modifikacemi (snaha o minimalizaci nespojitostí na hranici řetězených jednotek)
- q možnosti generování řeči:
 - § s omezeným slovníkem – věty ze specifické oblasti
 - § s neomezeným slovníkem – libovolné věty

Základní schéma

- q generování posloupnosti řečových jednotek
- q výběr vhodné realizace řečové jednotky
- q vlastní řetězení (konkatenace)
- q syntéza řízená daty – parametry syntetizéru se na nastavují automaticky z řečových dat



Řečové jednotky

q řečová jednotka

§ abstraktní termín pro pojmenování stejného typu řečových zvuků – např. foném (hláska)

q zástupce (reprezentant) řečové jednotky

§ pojmenování konkrétní (akustické) realizace konkrétní řečové jednotky

Volba řečových jednotek

- q **maximální pokrytí koartikulačních jevů**
 - § snaha volit delší jednotky obsahující co nejvíce koartikulační informace
- q **bezproblémové řetězení**
 - § minimalizace spektrálních a prozodických nespojitostí při konkatenaci → delší jednotky
- q **zobecnitelnost**
 - § použitelnost jednotek v obecném TTS → kratší jednotky
- q **malý počet řeč. jednotek**
 - § snaha volit kratší řečové jednotky
- q při výběru řeč. jednotek je třeba zohlednit všechna kritéria a podle typu aplikace mezi nimi učinit vhodný kompromis

Ukázka řečových jednotek

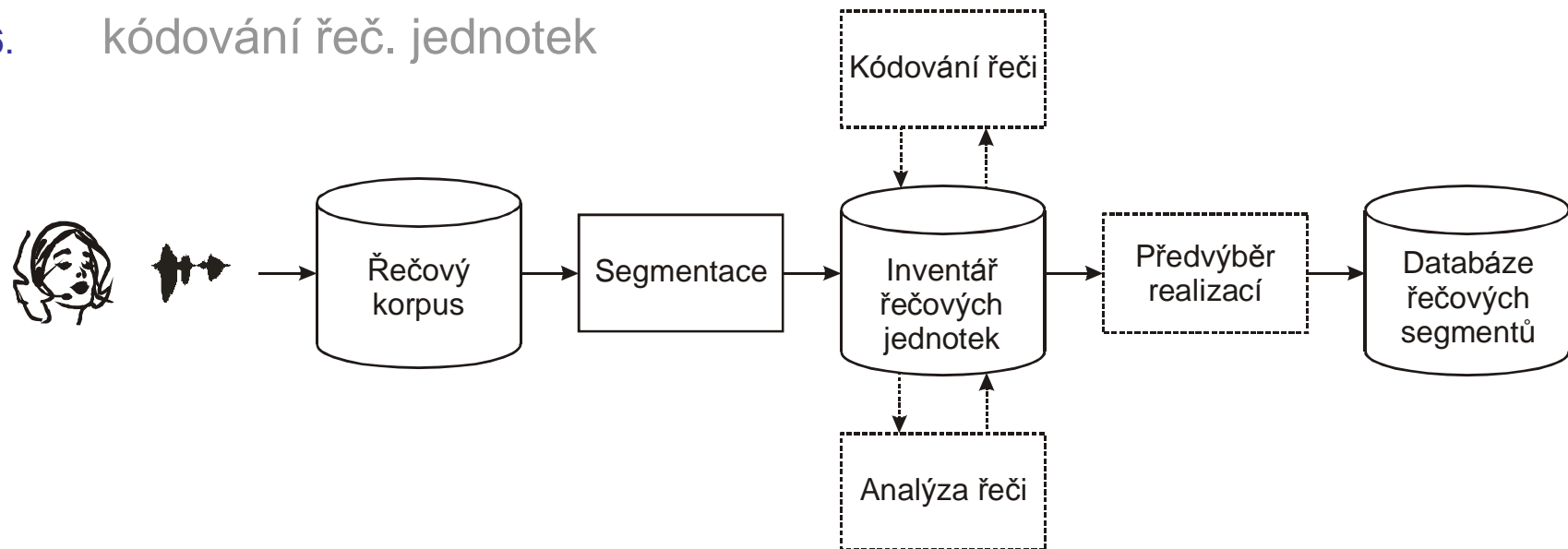
slova	vánoce										
slabiky	vá			no			ce				
demislabiky	#vá	ván	áno	noc	oce	ce#					
difóny	#-v	v-á	á-n	n-o	o-c	c-e		e-#			
fonémy	v	á	n	o	c		e				
trifóny	#-v+á	v-á+n	á-n+o	n-o+c	o-c+e		c-e+#				
půlfóny	v1	v2	á1	á2	...	o1	o2	c1	c2	e1	e2

Typy řečových jednotek

- q **slovní jednotky**
 - § věty, fráze, slova
 - § používá se pro syntézu s omezeným slovníkem
- q **slabičné jednotky**
 - § slabiky
 - § demislabiky
- q **fonémové jednotky**
 - § fonémy (hlásky)
 - § kontextově závislé fonémy (trifóny, polyfóny)
- q **„přechodové“ jednotky**
 - § difóny
 - § trifóny (polyfóny)
- q **subfonémové jednotky (půlfóny)**
- q **kombinované jednotky**

Vytvoření databáze řeč. jednotek

1. volba typu řečových jednotek
2. vytváření řečového korpusu
3. segmentace řečového korpusu
4. „předvybrání“ zástupců řeč. jednotek
5. parametrizace řeč. jednotek
6. kódování řeč. jednotek



Řečový korpus

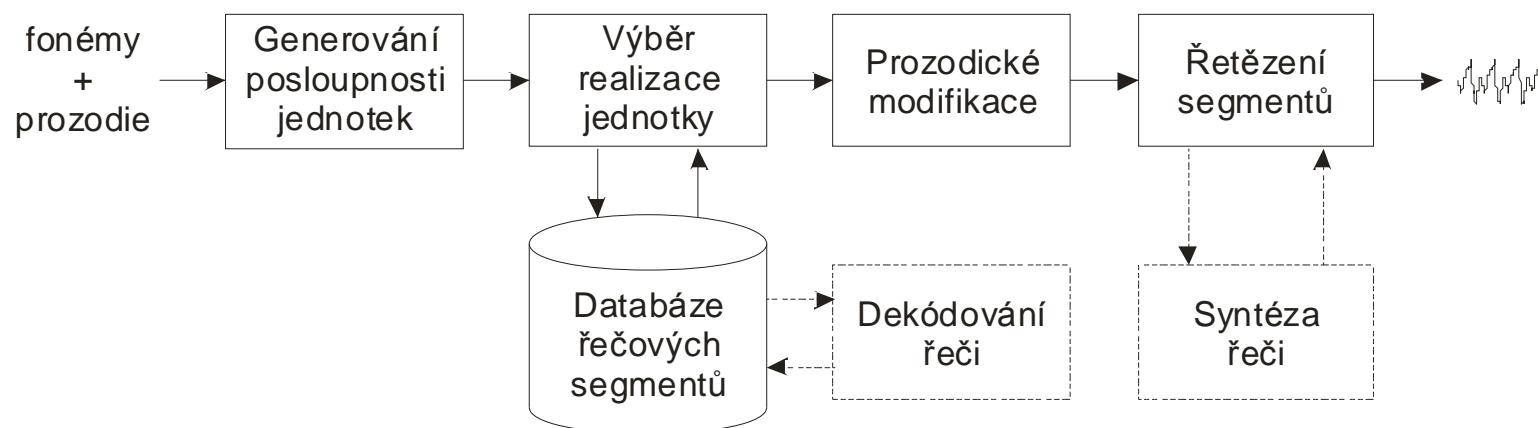
- q základní řečový materiál pro konkatenční syntézu
- q namlouvání korpusu:
 - § monotónně → systémy s 1 reprezentantem řeč. jednotky
 - § přirozeně neutrálně (bez emocí) → systémy s více reprezentanty
 - § přirozeně spontánně (s emocemi)
- q typy promluv:
 - § souvislá řeč → systémy s více reprezentanty
 - § izolovaná slova → systémy s 1 reprezentantem
 - § logatomy (nesmyslně vyslovované slabiky) → systémy s 1 reprezentantem
- q nejlépe namlouvání profesionálním řečníkem v nahrávacím studiu
- q ideální případ: každá jednotka zastoupena několikrát
- q realita: různý kontext a prozodické vlastnosti → v konečném korpusu nelze skladovat všechny realizace každé jednotky

Segmentace řeč. korpusu

- q proces hledání hranic řeč. jednotek v korpusu
- q **manuální** segmentace
 - § velice pracné a časově náročné
 - § trénování experti
 - § spíše pro systémy s 1 reprezentantem řeč. jednotek (malé inventáře řeč. jednotek)
- q **automatická** segmentace
 - § použití skrytých Markovových modelů (HMM)
 - § porovnávání se syntetickou řečí
 - § spíše pro systémy s více reprezentanty řeč. jednotek (obrovské inventáře řeč. jednotek)

Konkatenace

1. posloupnost fonémů + prozodie
2. odvození posloupnosti řeč. jednotek
3. výběr zástupce řeč. jednotky z databáze
4. dekódování řeč. jednotky
5. prozodické modifikace řeč. jednotek
6. spektrální vyhlazování řetězených jednotek (závislé na parametrizaci)
7. vytváření řeči na signálové úrovni – deparametrizace a vlastní konkatenace



Korpusově orientovaná syntéza

- q zvláštní případ konkatenáční syntézy
- q využití rozsáhlých foneticky a prozodicky pečlivě anotovaných řečových korpusů (řádově stovky MB)
- q více realizací každé řečové jednotky – v rozdílných fonetických, spektrálních i prozodických kontextech
- q plně automatická konkatenáční syntéza
- q všechny parametry se určují automaticky na základě dat z řeč. korpusu (včetně inventáře řeč. jednotek)
- q často tzv. **neuniformní** řečové jednotky (jednotky různého typu) – během on-line syntézy se vybere typ a realizace jednotky

= **syntéza výběrem jednotek**

Rozdělení metod

q metody pro systémy s 1 reprezentantem

§ probíhá mimo reálný čas

§ použití binárních rozhodovacích stromů pro výběr **typu** jednotky

§ **kontextově orientované shlukování (COC)**

- shlukování podobných akustických realizací daného fonému do „alofónových“ tříd na základě akustické podobnosti a fonetického/prozodického kontextu

§ **shlukování HMM pomocí fonetických stromů**

- rozšíření COC na práci s HMM
- shlukování podobných akustických realizací na úrovni HMM do „alofónových“ modelů s použitím fonetických otázek v každém uzlu stromu

q obecná úloha syntézy výběrem jednotek

§ pro systémy s více reprezentanty řeč. jednotek

§ výběr vhodného **typu** a **realizace** řečové jednotky probíhá v reálné čase syntézy

§ vhodné „předvybrat“ určitý počet relevantních reprezentantů pro každou jednotku

§ lze použít i ve spojení s výše uvedenými metodami

Prozodické a spektrální modifikace

- q přiblížení prozodických a spektrálních vlastností vybraných zástupců řeč. jednotek vlastnostem požadovaných v syntetické řeči
- q **prozodické modifikace**
 - § úprava prozodických vlastností řeč. jednotek z inventáře => přiblížení k požadovaným prozodickým vlastnostem syntetické řeči
 - § plně v režii konkrétní metody
- q **spektrální modifikace**
 - § úprava spektrálních vlastností syntetické řeči (v místech řetězení) za účelem vyhladit přechody mezi jednotkami
 - § dostačující většinou prostá lineární interpolace spektrálních parametrů (LPC, HNM)
- q žádné modifikace – teoreticky nejlepší kvalita (žádná degradace řeč. signálu → potřeba gigantických inventářů
- q s modifikacemi – větší pružnost systému → možno použít menší inventáře

Metody

- q přímá syntéza
- q LP syntéza
- q PSOLA
- q keprální syntéza
- q harmonický a šumový model vytváření řeči (HNM)

Výhody a nevýhody

- + nepotřebuje detailnější znalost procesu vytváření řeči
- + žádné ruční nastavování složitých pravidel
- + pracuje přímo s reálným řečovým signálem – problematické zvuky může zachytit v segmentech řeči (koartikulace)
- + lepší kvalita syntetické řeči (větší přirozenost)
- + rychlejší a jednodušší návrh syntetizéru (oproti formantové syntéze)

- ± kopíruje hlas řečníka z řečového korpusu

- těžkopádné změny hlasu (nová databáze)
- místa řetězení jednotek vždy potencionálním zdrojem problémů
- větší paměťové a výpočetní nároky (zejména v případě korpusově orientované syntézy)

Porovnání přístupů k syntéze řeči

q **formantová syntéza**

- nízká přirozenost
- + konstantní kvalita
- + nízké výpočetní nároky

q **konkatenační syntéza (1 zástupce řeč. jednotek)**

- ± lepší kvalita (vyšší přirozenost), ale stále nedostačující
- + konstantní kvalita
- + průměrné výpočetní nároky

q **konkatenační syntéza (výběr jednotek)**

- + velmi dobrá kvalita
- ± někdy vynikající, jindy ovšem (velmi) špatná
- vysoké výpočetní nároky (vyšší výkon PC à stále méně důležité)

Osnova

- ü Úvod
- ü Historie syntézy řeči
- ü Akustický model vytváření řeči
- ü Formantová syntéza
- ü Konkatenáční syntéza
- ü Artikulační syntéza
- q Syntéza řeči z textu (TTS)

Artikulační syntéza

- q komplexní modelování systému vytváření řeči člověkem
- q **artikulační model** zahrnuje modely jednotlivých řečových orgánů (artikulátorů) člověka
 - § hlasivky, rty, čelisti, jazyk, měkké patro, ...
- q matematická simulace šíření řečové „vlny“ v hlasovém traktu
- q artikulační parametry
 - § velikost a tvar retní štěrbiny, poloha jazyka, ...
- q parametry pro buzení
 - § stav hlasivek, velikost otvoru mezi hlasivkami, napnutí hlasivek, ...
- q nedostatek reálných dat
- q vysoká složitost – zatím prakticky nerealizovatelné
- q syntéza budoucnosti???



Osnova

ü Úvod

ü Historie syntézy řeči

ü Akustický model vytváření řeči

ü Formantová syntéza

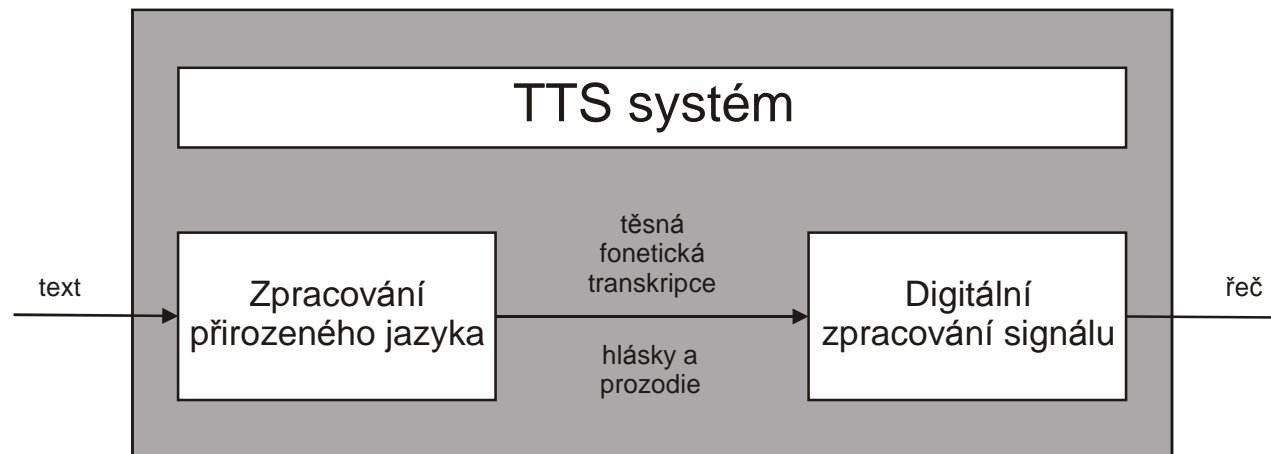
ü Konkatenáční syntéza

ü Artikulační syntéza

Ü Syntéza řeči z textu (TTS)

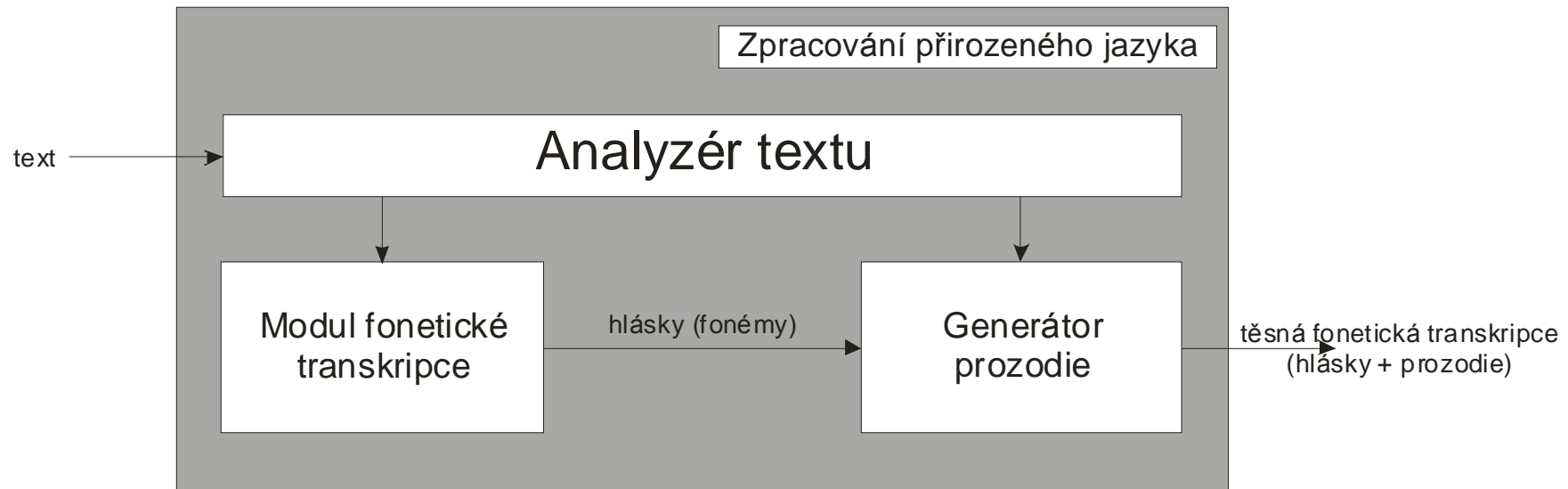
Syntéza řeči z textu (TTS)

- q nejobecnější úloha syntézy řeči: na vstupu text, výstupem řeč
- q cíl: generovat řeč z **libovolného** textu
- q **není možné uložit všechna slova (věty) do počítače, a pak je jen přehrávat!**
- q 2 základní moduly:
 - § modul pro zpracování textu
 - § syntetizér řeči



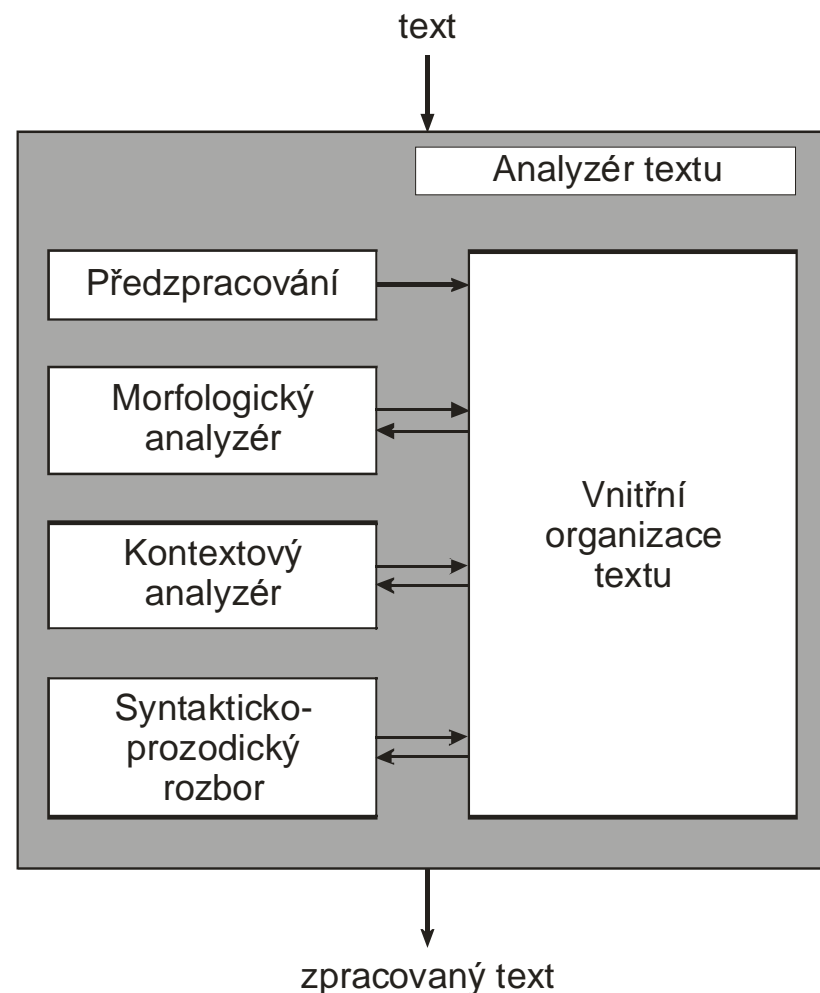
Zpracování textu

- q zpracování textu = zpracování přirozeného jazyka (*Natural Language Processing, NLP*)
 - § analýza textu
 - § fonetická transkripce
 - § generování prozodických charakteristik



Analýza textu

- q = morfologicko-syntaktický analyzátor
- q cíl: odstranit nejednoznačnosti z psaného textu
- q **předzpracování (normalizace)**
 - § přepis číslovek, zkratek, akronymů
- q **morfologický analyzátor**
 - § mluvnické kategorie slov izolovaně
- q **kontextový analyzátor**
 - § zkoumá slova v kontextu okolních slov
- q **syntakticko-prozodický rozbor (parsing)**
 - § detekce vět a větných (prozodických) úseků v textu



Předzpracování textu

q detekce slov

q detekce konců vět

§ pravidla

§ statistické metody
(neuronové sítě)

§ *Bylo nás 5. Skončil na
5. místě.*

q „normalizace“ textu

§ přepis textu do plné
slovní formy

§ slovníky

§ pravidla

§ kombinace

q přepis zkratk (ing., Dr., např. atd., ...)

q přepis akronymů (NATO, NASA, IBM, BBC)

q přepis symbolů (% , & , ...)

q přepis čísel

§ číslovky (*Skončil na 5. místě. Je tu 15
studentů.*)

§ datумы (1974, 1.1.2003, ...)

§ časové údaje (12:00)

§ peníze (1500 Kč, \$200, ...)

§ telefonní čísla (377 632 530, 377 63 2530)

§ ...

Morfologická analýza

- q navrhuje všechny možné mluvnické kategorie každého slova věty izolovaně bez kontextu okolních slov
- q slova se skládají z morfémů (nejmenší významotvorná jednotka jazyka)
- q slova neohebná
 - § předložky, spojky, ...
 - § slovníky
- q slova ohebná
 - § podst. jména, příd. jména, slovesa, ...
 - § různé tvary (pády, rody, čísla, časy, ...)
 - § tvoření slova: [předpona]+kmen+[přípona]+[koncovka] à pravidla, statistické metody
 - § slovníky pro kmeny, předpony, přípony, koncovky

Kontextová analýza

- q jednotlivá slova uvažuje v kontextu okolních slov
- q redukce seznamu možných mluvnických kategorií slov z morfologické analýzy → vytváří hypotézy

- q **Pravděpodobnostní metody**
 - § „přechodové pravděpodobnosti“ mezi sousedními mluvnickými kategoriemi 2 slov
 - § n-gramy, neuronové sítě (NN)
- q **Deterministické metody**
 - § pravidla typu ano/ne pomocí nichž se přijímá nebo odmítá určitá kombinace mluvnických kategorií
 - § klasifikační a regresní stromy (CART)

Syntakticko-prozodický rozbor

- q rozdělení věty na **větné úseky** (syntakticko-prozodické fráze) à souvisejí s očekávanou prozodickou realizací věty
 - § *Řekl jsem mu, ať tam jde.*
 - § *Poslechněte si / předpověď počasí / pro Českou republiku.*
 - § *Poslechněte si předpověď počasí / pro Českou republiku.*
- q velký význam pro správné generování prozodických vlastností syntetické řeči
- q **ručně odvozené heuristiky**
 - § pravidla pro detekce úseků pomocí interpunkce (čárky, středníky, pomlčky, závorky), spojek, ...
- q **gramatiky**
 - § rozšíření bezkontextových gramatik
- q **korpusově orientované metody**
 - § rozsáhlé označované textové korpusy (hranice mezi úseky)
 - § skryté Markovovy modely (HMM), klasifikační a regresní stromy (CART), neuronové sítě (NN)

Fonetická transkripce

- q **ortografická** (psaná) podoba jazyka (text = posloupnost písmen) à **fonetická** (výslovnostní) podoba (posloupnost fonémů)
- q 2 základní přístupy:
 - § **fonetický slovník** (analytické jazyky)
 - slovo a jeho výslovnost
 - morfémy (+ pravidla pro rozklad slova na morfémy)
 - pravidla pro spojování morfémů a slov
 - § **fonetická pravidla** (flexivní jazyky)
 - expertní systémy: produkční pravidla $A \rightarrow B / L_R$: podmínka
 - statistické přístupy na základě korpusu (HMM, NN)
- q kombinace
 - § pravidla + slovník (např. čeština: slovník výjimečných výslovností)
 - § slovník + pravidla (např. angličtina)
- q problém: cizí slova, jména, názvy měst, států, ...

Generování prozodie

- q prozodické charakteristiky řeči popisují intonaci, rychlost, hlasitost, přízvukování, rytmus a členění řeči
- q nevztahují se přímo k segmentům (fonémům), ale spíše ke slabikám a větším celkům
à suprasegmentální charakteristiky
- q informace o tom, **jak** má být daná promluva syntetizována
- q vyjadřují se pomocí 3 základních charakteristik
 - § perioda základního hlasivkového tónu (ZHT) – kontura F_0
 - § trvání segmentů řeči (fonémů)
 - § energie
- q **generátor prozodie**
 - § vstup: posloupnost fonémů, větné úseky (syntakticko-prozodické fráze), (text)
 - § výstup: **těsná fonetická transkripce** (doplnění posloupnosti fonémů o prozodické značky)
- q velký vliv na přirozenost syntetické řeči
- q generování přirozené prozodie (vč. emocí)
 - § jeden z hlavních problémů současných TTS
 - § praxe: detekce větných úseků, kontura F_0 , trvání, energie
- q tónové jazyky (čínština)
 - § prozodické charakteristiky (intonace) ovlivňují význam slov

Intonace

- q intonace vět, resp. větných úseků
 - § rozlišení typů vět (oznamovací, tázací, neukončené)
 - § vyjádření přízvuku (lokální vzestup F_0)
- q pravidla pro generování kontury F_0
 - § jednoduché, nedostačující
- q intonační modely
 - § akustické
 - Fujisakiho model, metody akustické stylizace
 - § percepční
 - IPO model, metody percepční stylizace
 - § lingvistické
 - teorie kontur výšky hlasu, teorie posloupnosti tónů (TOBI)
- q automatické trénování parametrů modelů z anotovaných řečových dat
 - § značkovací systémy (TOBI, ...)

Syntéza řeči z textu

Časování

- q ovlivňuje rychlost a celkový rytmus řeči
- q **trvání**
 - § základní jednotky – fonémy (slabiky)
 - § průměrná trvání + pravidla (pozice fonému ve slově, větném úseku, ...)
 - zpomalování řeči na konci věty, zrychlování na začátku delšího úseku
 - § statistické modely trvání
 - využívají anotované korpusy (pozice fonému ve slově, větném úseku, větě, typ větného úseku, přízvučnost/nepřízvučnost)
 - NN, CART, regresní techniky
- q **pauzy**
 - § souvisí s rychlostí řeči
 - § různé pauzy mezi různými větnými úseky
 - § delší pauzy pro delších větných úsecích, atd.
- q **přízvuky**
 - § realizují se pomocí všech 3 základních prozodických charakteristik
 - § slovní přízvuk
 - odlišuje slova v souvislé řeči
 - fonetické slovníky, pravidla + malé slovníky výjimek (čeština)
 - § větný přízvuk
 - nutno pochopit význam věty (syntaktická, ale lépe sémantická analýza)

Hlasitost

- q „nejméně“ důležitá prozodická charakteristika
- q základní jednotky – fonémy (slabiky)
- q průměrná hlasitost + pravidla
 - § pozice fonému ve slově
 - § pozice fonému ve větném úseku
 - § pozice fonému ve větě
 - § typ věty/větného úseku
 - § přízvučnost/nepřízvučnost
 - § ...
- q statistické modely – CART

Hodnocení kvality syntetické řeči

- q kvalita: srozumitelnost, přirozenost, plynulost, příjemnost, přijatelnost uživatelem
- q vzhledem ke komplexnosti řeči neexistují objektivní testy
- q poslechové testy – subjektivní hodnocení kvality (hodně posluchačů à „objektivnost“)
- q **testy srozumitelnosti**
 - § MRT (Modified Rhyme Test)
 - 50 skupin slov po 6, slova se liší v počátečním nebo koncovém fonému
 - např.: *pes – les – ves – bez – děs – rez*
 - § SUS (Semantically Unpredictable Sentences)
 - gramaticky správné, ale nesmyslné věty
 - nesrozumitelné slovo nelze odvodit z kontextu okolních slov
 - např.: *Ušatí komáři štěkali mokré diváky.*
- q **testy přirozenosti (celkové kvality)**
 - § MOS (Mean Opinion Score)
 - hodnocení kvality řeči: 5-vynikající, ..., 1-špatný
 - § CCR (Comparison Category Rating)
 - porovnání stejné věty generované 2 syntetizéry

Aplikace TTS systémů

- q pomůcky pro handicapované lidi
- q telekomunikační služby
- q automatické čtení (e-mail, SMS, ...)
- q hlasové monitorování
- q výuka jazyků
- q multimédia, komunikace člověk-počítač
- q mluvící hračky pro děti
- q výzkum (fonetika, lingvistika, akustika)