

# Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta informatiky  
Masarykova univerzita

podzim 2024

# Obsah

- 1 Syntéza řeči
- 2 Syntéza ve frekvenční oblasti
  - Syntéza formantového typu
  - LPC Syntéza
- 3 Syntéza řeči v časové oblasti

# Úvod

- Úkol:
  - Převod psaného textu na mluvenou řeč.
  - Co nepřirozenější řeč - ideálně k nerozeznání od člověka:
    - správná intonace
    - správné umístění přízvuků
    - správná koartikulace
    - správný rytmus
    - ...

# Druhy syntézy řeči

- Druhy syntézy řeči
  - ve frekvenční oblasti
  - v časové oblasti
    - korpusová
    - problémově orientovaná syntéza (hlášení nádražního rozhlasu, automatizované linky telefonické podpory) – často syntéza postavená na rámcích.

# Fáze syntézy řeči

- 1 Fonetický přepis.
- 2 Syntéza fonetické transkripce
- 3 Případný postprocessing:
  - intonace
  - správné časování - modifikace délky fonémů, ...
  - větné přízvuky
  - ...

# Fonetický přepis

- Slouží k přesnému, jednoznačnému zápisu mluvené řeči.
- Využívá fonetickou abecedu:
  - mezinárodní fonetická abeceda - IPA (součást standardu UNICODE): ma:f se d'obr,e
  - 7bitový přepis IPA pomocí ASCII - SAMPA/Česká SAMPA: ma:S se dobr'e / ma:S se dobRe
    - Syntetizéry/hlasy v syntetizérech mohou používat modifikace.
- Nelze si pamatovat fonetický přepis každé promluvy - nutno zabezpečit automatický přepis:
  - fonologická pravidla
- Při transkripci češtiny se některé české znaky nevyužívají:
  - ch - x
  - w - v
  - y/ý - i/í
  - q - kv
- Koartikulace

# Pravidla fonetického přepisu češtiny

- ch → x
- ů → ú
- w → v
- q → kv
- y → i
- ý → í
- ě → je /po b,p,f,v
- dě, tě, ně, mě
  - dě → d'e
  - tě → t'e
  - ně → ňe
  - mě → m'ne

# Pravidla fonetického přepisu češtiny

- di, ti, ni
  - di → d'i
  - ti → t'i
  - ni → ňi
- X:
  - x → ks | začátek slova před samohláskou, mezi samohláskami nebo před neznělou souhláskou a nebo na konci slova, s výjimkou ex<samohláska> → egz
  - x → gz | před znělou souhláskou



# Změny na při spojování souhlásek

- Dochází k nim při spojování souhlásek.
- Způsobeny přenastavováním mluvidel.
- 2 druhy:
  - spodoba znělosti - změna znělosti párových souhlásek
    - ZPS → ~ ZPS
    - NPS → ~ NPS
    - dub → dup
    - zpěv → spjev
    - sběr → zbjer
    - když → gdiš
  - spodoba artikulační - při spojení dvou souhlásek s různou artikulací
    - banka, tango
    - tramvaj, nymfa
    - punt'a, pindík
    - odpovědně, sto dní, vodní
    - ts → c, tš → č
    - ds → c, dš → č

# On-line přístupné ukázky syntézy řeči

- AT&T Labs Natural Voices© Text-To-Speech demo at NetUp technologies
- IBM Watson Text to Speech Voices
- Free demo to create avatars using TTS by SitePal
- Cepstral Text-to-Speech
- Festival Online Demo
- Speechech s.r.o.

# Syntéza ve frekvenční oblasti

- Emulace funkce hlasového ústrojí pomocí FM syntezátoru.
- Nutno uchovávat:
  - frekvenční charakteristika použitého hlasu
  - parametry buzení.
- Využívá:
  - systém frekvenčních generátorů - simulují hlasivky
  - filtry a zesilovače - simulace rezonance v dutinách
  - Tyto komponenty ovládány parametry modelu.
- Nejběžněji použité způsoby kódování zdroje:
  - Řečová syntéza formantového typu - uchovávají se parametry průběhu jednotlivých formantů a buzení.
  - LPC řečová syntéza - uchovávají se  $F_0$ , příznak znělosti, amplituda budícího signálu  $G$  a koeficienty LPC,
  - syntéze postavená na HMM,
  - ...

# Syntéza ve frekvenční oblasti

- Výhody
  - menší paměťové nároky - uchovávají se pouze parametry modelu.
- Nevýhody:
  - oproti syntéze v časové oblasti může být výsledek méně přirozený - „robotické“ hlasy
  - Softwarová - výpočetně relativně náročná - lze implementovat přímo na úrovni HW
    - skládání jednotlivých frekvencí, které tvoří příslušné fonémy
    - řešení koartikulace
    - ...
  - Neexistuje dostatečně přesný matematický model

# Využití syntézy ve frekvenční oblasti

- Využití dříve:
  - malé paměťové nároky
  - domácí počítače (Amiga, Atari, ...)
  - syntéza realizována většinou hardwarově
- Dnes:
  - Syntéza na zařízeních s nedostatkem paměti.
  - Syntéza realizovaná hardwarově pomocí zákaznických obvodů.
- Doplnění syntézy v časové oblasti o prozodické jevy:
  - Větná intonace
  - ...
  - Realizováno programově pomocí modifikace  $F_0$  a formantů.

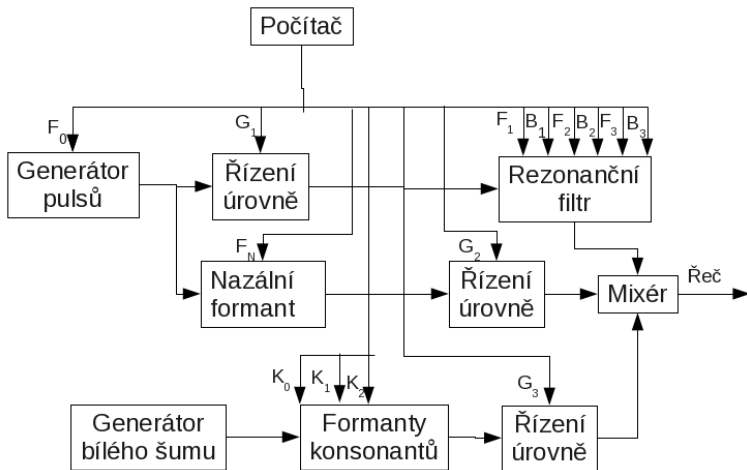
# Způsoby kódování parametrů syntezátoru ve frekvenční oblasti

- Způsoby kódování parametrů syntetizéru ve frekvenční oblasti:
  - syntéza formantového typu
  - LPC syntéza
  - a další – např. HMM

# Syntéza formantového typu

- Parametry modelu:
  - jednotlivé formanty, jejich intenzity a šířky pásem
    - formanty –  $F_0, F_1, F_2, F_3, F_N, K_0, K_1, K_2$
    - intenzity – parametry zesílení –  $G_1, G_2, G_3$
  - Průběh syntézy:
    - Samohlásky – generátor pulzů vygeneruje  $F_0$ , signál je následně zesílen a je přidán nazální format, v rezonančním filtru jsou přidány formanty, řízení úrovně zesílí signál s nazálním formantem.
    - Souhlásky – základem je bílý šum, který je obohacen o formanty konsonant a zesílen.
    - Dojde k smísení těchto tří signálů a výsledkem je odpovídající hláska, . . .

# Schéma syntetizéru formantového typu

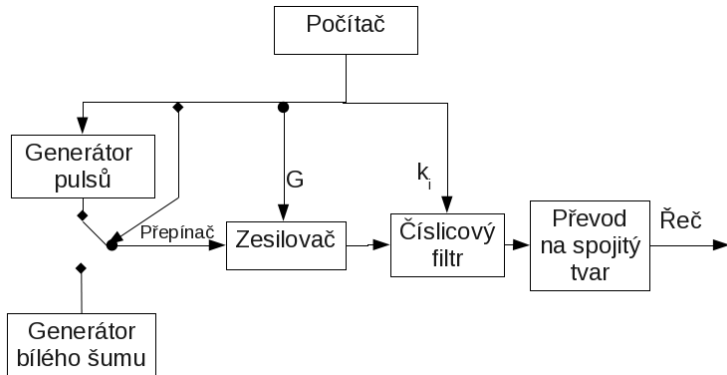




# LPC Syntéza

- Parametry syntetizéru – získány pomocí LPA:
  - příznak znělosti hlásky – ovládá přepínač mezi generátorem pulsů a generátorem bílého šumu
  - parametr buzení pro generátor pulsů pro generátor pulsů
  - parametr  $G$  – zesílení signálu
  - parametry spektrální obálky  $k_i$
- Vlastní syntéza
  - 1 vygenerován základní signál hlásky
    - 1 znělé hlásky – generátor pulsů vygeneruje budící signál
    - 2 neznělé hlásky – jako základní signál je použit bílý šum
  - 2 signál z předchozího kroku je zesílen v zesilovači – ovládán parametrem  $G$
  - 3 doplnění vyšších harmonických frekvencí v číslicovém filtru – ovládán parametry  $k_i$
  - 4 výsledná posloupnost vzorků je převedena na spojitý signál.

# Schéma LPC syntetizéru



# 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 (<http://tcts.fpms.ac.be/synthesis/mbrola.html>)

# Používané řečové segmenty

## Pokračování

- 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/dvojhláska) - ú - kol
  - KV (souhláska - samohláska) - vo - da
  - KVK - jed-not-ka
  - KK - tr-sy
  - KKV - dna
  - KKVK - dmout
- Tvoří více než 95 % slabik
- Umožňují automatickou segmentaci textu.
- Používají se např. v syntetizéru Demosthénés (doc. Kopeček LAF (LSD) FI)