

# Dialogové systémy

## Rozpoznávání řeči

Luděk Bártek

Laboratoř vyhledávání a dialogu, Fakulta Informatiky Masarykovy Univerzity,  
Brno

jaro 2020

- Zvuk je neměnný pouze na krátkých časových úsecích – metody krátkodobé analýzy.
- Tento interval se nazývá mikrosegment – velikost 10 — 40 ms.
- Metody krátkodobé analýzy:
  - V časové oblasti – zpracovávají se přímo hodnoty jednotlivých vzorků.
  - Ve frekvenční oblasti – ze vzorků se získávají frekvenční charakteristiky, které jsou následně zpracovány.
- Modelování funkce Cortiho ústrojí – pomocí diferenciálních rovnic se simuluje rezonance na určitých vlákénkách Cortiho ústrojí.

# Zpracování digitalizovaného signálu

## Váhové okénko

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Při krátkodobé analýze předpokládáme, že signál je v okolí mikrosegmentu periodický se stejnou periodou jako uvnitř.
- Vzniklá chyba se kompenzuje použitím „okénka”.
- Okénko – posloupnost vah pro vzorky v mikrosegmentu.
- Tyto váhy by měly odpovídat tomu, jak je daný vzorek ovlivněn okolím mikrosegmentu.
- Nejčastěji používané typy okének:
  - pravoúhlé okénko
  - Hammingovo okénko

# Zpracování digitalizovaného signálu

## Hammingovo okénko

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

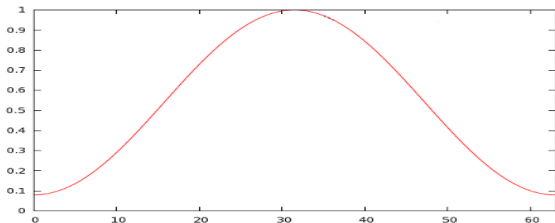
Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Vychází z předpokladu, že čím jsou vzorky blíže středu mikrosegmentu, tím méně jsou ovlivněny okolím.
- Pro výpočet vah se používá vzorec:

$$w(n) = \begin{cases} n = 0 \dots N - 1 & 0,54 - 0,46 \cos\left(\frac{2\pi n}{N-1}\right) \\ n < 0 \vee n \geq N & 0 \end{cases}$$

- Průběh vah okénka na mikrosegmentu:



- Vychází se z předpokladu:
  - 1 vzorky mikrosegmentu nejsou pro naše potřeby ovlivněny okolím mikrosegmentu
  - 2 všechny vzorky mikrosegmentu jsou ovlivněny stejně.
- Všechny vzorky mikrosegmentu mají shodnou váhu.

$$w(n) = \begin{cases} 0 \leq n < N & 1 \\ n < 0 \vee n \geq N & 0 \end{cases}$$

# Analýza digitalizovaného signálu v časové oblasti

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Vychází přímo z hodnot vzorků, nikoliv z hodnot spektra.
- Používané metody:
  - funkce krátkodobé energie
  - funkce krátkodobé intenzity
  - funkce středního počtu průchodů nulou
  - diference 1. řádu
  - autokorelační funkce
  - ...

# Analýza v časové oblasti

## Funkce krátkodobé energie

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Využívá funkci průměrné energie v rámci segmentu:

$$E(n) = \sum_{k=-\infty}^{\infty} (s(k)\omega(n-k))^2$$

- $s(k)$  – vzorek v čase  $k$
- $\omega(n-k)$  – váha odpovídajícího okénka pro čas  $k$
- Výstupem je průměrná energie v daném okénku.
- Druhá mocnina zvyšuje dynamiku zvukového signálu.
- Použití:
  - automatické oddělení ticha řeči (signálu)
  - příznaky v jednoduchých klasifikátorech slov
  - oddělení znělých a neznělých částí promluvy.

# Analýza v časové oblasti

## Funkce krátkodobé intenzity

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Funkce intenzity signálu v daném okénku.

$$I(n) = \sum_{k=-\infty}^{\infty} |s(k)|\omega(n-k)$$

- $|s(k)|$  – absolutní hodnota vzorku v čase  $k$
- $\omega(n-k)$  – váha odpovídajícího okénka pro čas  $k$
- Použití – stejné jako funkce krátkodobé energie.
- Oproti krátkodobé energii nezvýrazňuje tolik dynamiku řečového signálu.



# Analýza v časové oblasti

Krátkodobá funkce středního počtu průchodu nulou

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Počítá změny znaménka digitalizovaného signálu.

$$Z(n) = \sum_{k=-\infty}^{\infty} |\operatorname{sgn}[s(k)] - \operatorname{sgn}[s(k-1)]| \omega(n-k)$$

- Varianta – počet lokálních extrémů.
- Obě metody mohou být negativně zatíženy šumem zvukového pozadí.
- Použití:
  - detekce ticha
  - detekce začátku a konce i zašuměné promluvy
  - přibližné určení základního hlasivkového tónu a formantů
  - příznaky jednodušších klasifikátorů slov

# Analýza v časové oblasti

## Autokorelační funkce

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Vrací podobnost úseků daného mikrosegmentu (čím větší výsledná hodnota, tím podobnější úseky posunuté o  $m$  vzorků).

$$R(m, n) = \sum_{k=-\infty}^{\infty} (s(k)\omega(n-k))(s(k+m)\omega(n-k+m))$$

- Je-li signál periodický s periodou  $P$ ,  $R(m, n)$  nabývá maxima pro  $m=0, P, 2P, \dots$
- Předpokládá délku mikrosegmentu aspoň  $2P$ .
- Použití:
  - Používá se k zjišťování periodicity signálu základního tónu řeči.
  - Základ pro výpočet koeficientů LPA.

# Analýza signálu ve frekvenční oblasti

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Transformuje digitální řečový signál z časové oblasti do frekvenční oblasti.
- Využívá k tomu nejčastěji Fourierovu transformaci.
- Nejčastěji používané druhy analýzy ve frekvenční oblasti:
  - krátkodobá Fourierova transformace
  - krátkodobá diskrétní Fourierova transformace
  - rychlá Fourierova transformace
  - keprstrální analýza
  - lineární predikce
  - ...

# Analýza signálu ve frekvenční oblasti

## Krátkodobá Fourierova transformace

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Vychází z Fourierovy transformace:

$$S(\omega, t) = \sum_{k=-\infty}^{\infty} s(k)h(t - k)e^{-i\omega k}$$

- Obyčejnou Fourierovu transformaci získáme fixací času  $t$ .
- $|S(\omega, t)|$  – amplituda složky akustického spektra odpovídající frekvenci  $\omega$  v čase  $t$ .
- $h(n)$  – váhová funkce okénka.
- Předpokládá na vstupu periodickou funkci – zvuk je periodický na krátkých časových úsecích.
- Při jejím použití se předpokládá, že zpracováváný mikrosegment se periodicky opakuje.

# Analýza signálu ve frekvenční oblasti

## Diskrétní Fourierova transformace

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Používá se pro vyjádření spektrálních vlastností periodických posloupností s periodou  $N$  vzorků resp. konečných posloupností délky  $N$  vzorků.
- Výpočet koeficientů  $X(k)$  DFT:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} x(n)W_N^{-kn}$$

- $|X(k)|$  – intenzita  $k$ . spektrálního koeficientu, frekvence závisí na velikosti mikrosegmentu  $N$  a vzorkovací frekvenci.
- $x(n)$  –  $n$ . vzorek daného mikrosegmentu
- $W_n = e^{i\frac{2\pi}{N}} = \cos(2\pi/N) + i \cdot \sin(2\pi/N)$ .
- Výpočet  $n$ . vzorku na základě hodnot  $X(k)$  – IDFT:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{i\frac{2\pi}{N}kn} = \frac{1}{N} \sum_{k=0}^{N-1} X(k)W_N^{kn},$$

# Analýza signálu ve frekvenční oblasti

## Rychlá diskretní Fourierova transformace

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Výpočet spektrálních koeficientů pomocí DFT –  $n^2$  operací nad komplexními čísly.
- Pomocí FFT –  $N \cdot \log_2 \frac{N}{2}$  operací násobení.
- FFT požaduje, aby délka analyzovaného segmentu byla mocninou 2.
  - využívá metodu rozděl a panuj pro optimalizovaný výpočet DFT
  - zvláště se provádí výpočet lichých a sudých členů sumy
  - předchozí lze chápat jako transformaci dvou vektorů  $(x_0, x_2, \dots, x_{N-2})$  a  $(x_1, x_3, \dots, x_{N-1})$ , lišících se pouze členem  $(e^{-i\frac{2\pi}{N}})^k$ , a vlastní transformace se neliší.

# Analýza signálu ve frekvenční oblasti

## Kepstrální analýza

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Vychází z modelu činnosti hlasového ústrojí.
- Řečové kmity lze modelovat jako odezvu lineárního systému na buzení sestávající ze sledu pulzů pro znělou řeč a šumu pro neznělou.
- Kepstrum –  $X(k) = IFFT(\log|FFT(x(k))|)$
- Kepstrální analýza umožňuje z řeči oddělit parametry buzení a parametry hlasového ústrojí.
- Využití:
  - ocenění fonetické struktury řeči – znělost, perioda základního tónu, formanty, ...
  - rozpoznávání slov
  - verifikace a identifikace mluvčího
  - ...

# Analýza signálu ve frekvenční oblasti

## Lineární prediktivní analýza

Dialogové  
systémy

Luděk Bártek

Zpracování di-  
gitalizovaného  
signálu

Zpracování v časové  
oblasti

Zpracování ve  
frekvenční oblasti

- Jedna z nejefektivnějších metod analýzy akustického signálu – zajišťuje velmi přesné odhady parametrů při relativně malé zátěži.
- Vychází z předpokladu, že  $s(k)$  lze popsat jako lineární kombinaci  $N$  předchozích vzorků a buzení  $u(k)$ :

$$s(k) = - \sum_{i=1}^N a_i s(k-i) + Gu(k)$$

kde  $G$  je koeficient zesílení a  $N$  řád modelu.

- Použití:
  - určování spektrálních charakteristik modelu hlasového ústrojí
  - z chyby predikce lze odvodit poznatky o znělosti a určit frekvenci základního hlasivkového tónu
  - koeficienty  $a_i$  nesou informaci o spektrálních vlastnostech – lze je použít jako příznaky pro rozpoznávání řeči.