Úvod do počítačového zpracování řeči

Luděk Bártek

Fakulta Informatiky
Masarykova Univerzita

podzim 2021

Obsah

1 Zpracování digitalizovaného signálu
- Metody krátkodobé analýzy
- Váhová okénka
- Zpracování signálu v časové oblasti
- Zpracování signálu ve frekvenční oblasti

Metody krátkodobé analýzy

- Zvuk je periodický pouze na krátkém intervalu.
- Zpracování signálu na krátkém časovém intervalu (*mikrosegmentu*), kde se nepředpokládají výraznější dynamické změny.
  - velikost od 10 do 40 ms
- Metody krátkodobé analýzy:
  - v časové oblasti,
  - ve frekvenční oblasti.

Krátkodobá analýza

- Nevýhoda použití mikrosegmentu:
  - Chyba způsobená předpokladem, že zvuk v okolí okénka zůstává periodický s periodou okénka.
  - Tuto chybu lze kompenzovat použitím okénka.
- Okénko - posloupnost vah pro vzorky v mikrosegmentu.
- Nejběžněji používané typy okének:
  - hammingovo,
  - pravoúhlé.

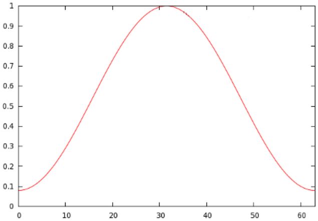
Hammingovo okénko

- Pro výpočet n. váhy se využívá vztah

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

  N - počet vzorků v mikrosegmentu

- Hammingovo okénko pro mikrosegment délky 64



Pravoúhlé okénko

- Přiřadí každému prvku mikrosegmentu váhu 1:

$$\omega(n) = \begin{cases} 1 & n = 0 \dots N-1 \\ 0 & n < 0 \vee n \geq N \end{cases}$$

N - délka mikrosegmentu

Analýza digitalizovaného signálu v časové oblasti 1.

- Vychází se přímo z hodnot vzorků, nikoliv z hodnot spektra.
- Používají se:
  - funkce krátkodobé energie
  - funkce krátkodobé intenzity
  - krátkodobá funkce středního počtu průchodů nulou
  - diference 1. řádu

Funkce krátkodobé energie

- Výpočet podle vzorce:

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

  - s(k) - vzorek v čase k, $\omega(n-k)$ - váha odpovídajícího okénka pro čas k
  - Výstupem je průměrná energie v rámci segmentu.
  - Značně citlivá na velké změny úrovně signálu v rámci segmentu.
  - Druhá mocnina zvyšuje dynamiku zvukového signálu.
  - Ukázka výpočtu funkce krátkodobé energie v Octave.
- Využití:
  - detekce ticha a promluvy
  - příznaky pro jednoduché klasifikátory slov
  - oddělení znělých a neznělých částí promluvy
  - ...

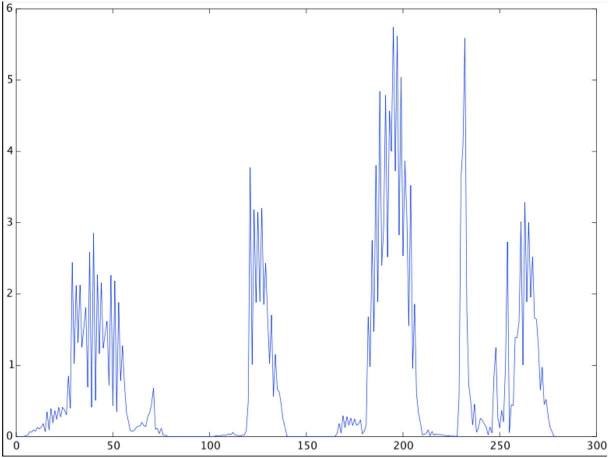
Funkce krátkodobé intenzity

- Funkce krátkodobé intenzity:

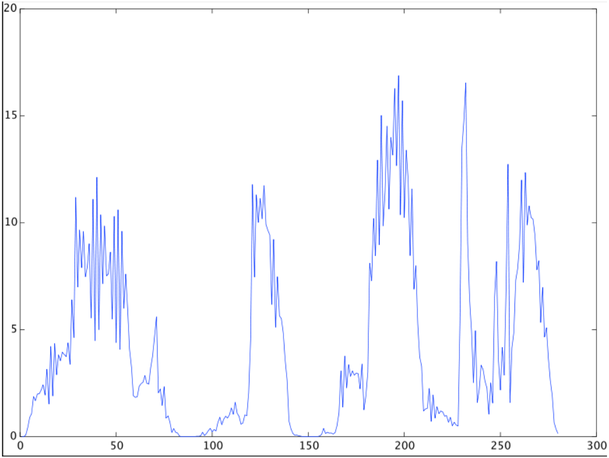
$$I(n) = \sum_{k=-\infty}^{\infty} |s(k)\omega(n-k)|$$

    - Použití - shodné s funkcí krátkodobé energie.
- Ukázka implementace pro systém Octave.

Ukázka průběhu funkce krátkodobé energie



Ukázka průběhu funkce krátkodobé intenzity



Krátkodobá funkce středního počtu průchodů nulou

- Krátkodobá funkce středního počtu průchodu nulou:
  - součet všech průchodů signálu nulou

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

  - varianta - počet lokálních extrémů
  - obě mohou být negativně ovlivněny šumem zvukového pozadí
- Využití:
  - detekce začátku a konce slova (i zašuměného)
  - určení základního hlasivkového tónu
  - přibližné určení formantů
  - ...

Ukázka průběhu funkce středního počtu průchodů nulou

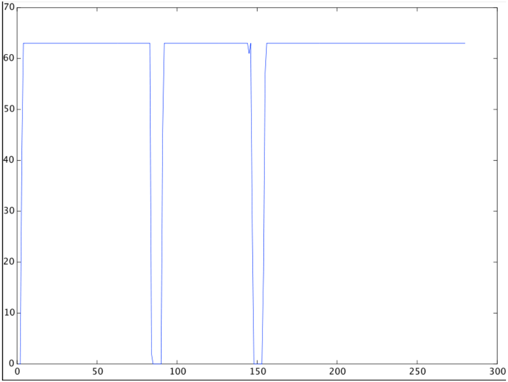


Figure: ZCR pro promluvu: ,,Jak se máš? “

Diferenční klasifikátory

- Diference prvního řádu

$$D_n = \sum_{k=-\infty}^{\infty} |s(k) - s(k-1)|\omega(n-k)$$

Krátkodobá autokorelační funkce

- Krátkodobá autokorelační funkce:

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

  - používá se při zjišťování periodicity signálu základního tónu řeči
  - je-li signál periodický s periodou T, R(m,n) nabývá maxima pro $m = 0, T, 2 \cdot T, \ldots$
  - předpokládá délku mikrosegmentu aspoň $2 \cdot T$

Zpracování signálu ve frekvenční oblasti

- Transformují hodnoty vzorků na různé frekvenční charakteristiky.
- Většinou je lze chápat jako spektrální charakteristiky.
- Nejvíce používané:
  - krátkodobá Fourierova transformace
  - kepstrální analýza
  - lineární prediktivní analýza

Fourierova řady

- f(x) - periodická funkce s periodou T, která má na intervalu T konečný počet extrémů a nespojitostí

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k cos(kx) + b_k sin(kx))$$

- Způsob výpočtu koeficientů a_i a b_i:
    - α, $\alpha + T$ - interval periodicity funkce f

$$a_k = \frac{2}{T}\int_{\alpha}^{\alpha+T} f(x)cos(kx)dx$$

$$b_k = \frac{2}{T}\int_{\alpha}^{\alpha+T} f(x)sin(kx)dx$$

- Nelze přímo použít - digitalizovaný zvuk není spojitý a je periodický pouze na omezených úsecích.

Diskrétní Fourierova Transformace (DFT)

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

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

  - $|X(k)|$ - intenzita k. spektrálního koeficientu; frekvence závisí na velikosti mikrosegmentu N a vzorkovací frekvence T
  - x(n) - n. vzorek daného mikrosegmentu.
  - $\omega = e^{i\frac{2\pi}{N}} = cos(\frac{2\pi}{N}) + i\,sin(\frac{2\pi}{N})$

Fourierova transformace

Výpočet hodnoty vzorku na základě hodnot X(k)

- Výpočet n. vzorku na základě hodnot X(k) - Inverzní diskrétní Fourierova transformace (IDFT):

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

Rychlá Fourierova transformace (FFT)

- Časová složitost výpočtu spektrálních koeficientů pomocí DFT - n^2 operací na komplexními čísly.
- Pomocí FFT - $N \, log_2(\frac{N}{2})$ operací násobení.
- FFT požaduje, aby délka analyzovaného segmentu byla mocninou 2.
- Algoritmus využívá:
  - periodicity členu ω_n^{-nk} ve výpočtu DFT
  - rekurzivní algoritmus metodou rozděl a panuj.

Kepstrální analýza

- Vychází z modelu činnosti hlasového ústrojí:
  - Řečové kmity lze modelovat jako odezvu lineárního systému na buzení sestávající ze sledu pulzů pro znělou hlásku 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
    - F_0, F_1, F_2, ...
  - rozpoznávání slov
  - verifikace a identifikace mluvčího
  - ...

Lineární predikce

- 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 koeficientem zesílení G:

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

Lineární predikce

Použití

- Určená spektrálních charakteristik modelu hlasového ústrojí.
- Z chyby predikce lze odvodit poznatky o znělosti a určit frekvenci základního tónu.
- Koeficienty a_i nesou informaci o spektrálních vlastnostech.
  - Lze je použít jako příznaky pro rozpoznávání řeči.

Další metody zpracování signálu ve frekvenční oblasti

- Pásmové filtry – obdoba pásmových filtrů z oblasti elektroniky, propouští rozsah frekvencí daný dvěma mezními frekvencemi.
- Dolní propust – nepropouští frekvence vyšší, než je mezní frekvence filtru, při jejím dosažení klesá intenzita signálu na cca 1/3.
- Horní propust – nepropouští nižší frekvence, než je mezní frekvence filtru, při jejím dosažení klesá intenzita signálu na cca 1/3.
- Modelování Cochlei – simuluje chování vybraných vláknének Cochlei (vybírá několik frekvencí).

Software pro analýzu signálu

- HTK - Hidden Markov Model Toolkit (Engineering Department of Cambridge University) - toolkit pro tvorbu rozpoznávačů řeči založených na skrytých Markovových modelech.
- ESPS toolkit
- NICO toolkit - toolkit pro vytváření umělých neuronových sítí, využívá se např. pro rozpoznávání řeči.
- Matlab - knihovny pro analýzu řeči
  - labrosa.ee.columbia.edu/matlab/
  - Audio processing in Matlab
  - ...
- Octave - opensource alternativa Matlabu
  - Měly by jít použít tytéž knihovny.
  - Viz též bakalářská práce L. Oroszlányho (FI, jaro 2012)
- SMP Tool