

Úvod do počítačového zpracování řeči

Luděk Bártek¹

1

Obsah

- 1 Digitalizace akustického signálu
- 2 Zpracování digitalizovaného signálu

Digitalizace zvuku

- Cíl - převod spojitého signálu na posloupnost digitálních hodnot vhodných pro uchování v počítači.
- Postup digitalizace:
 - 1 Vzorkování - převod reálných vstupních hodnot na posloupnost diskrétních reálných čísel.
 - 2 Kvantizace - převod posloupnosti reálných čísel na posloupnost celých čísel.
 - 3 Kódování - způsob uložení a kódování posloupnosti celočíselných hodnot získaných v kroku 2.

Vzorkování

- Transformace spojitého časové závislého signálu $s(t)$ na časově diskrétní posloupnost $s(nT) = 0, 1, 2, \dots$
 - T - perioda vzorkování.
 - Pokud nemá dojít ke ztrátě informace, musí být vzorkovací frekvence aspoň dvojnásobkem nejvyšší frekvence, která je signálu obsažena.
- Po čase T je sejmuta a dána na výstup (ke kvantizaci) hodnota ze vstupního snímače.
 - většinou okamžitá úroveň napětí nebo proudu na vstupu.

Vzorkování

Oblasti použití

- Digitální zpracování zvuku:
 - audio CD
 - mp3 - navíc použita ztrátová komprese
 - miniDisc - navíc použita ztrátová komprese ATRAC
 - DAT
 - ...
- Digitální zpracování signálu obecně:
 - digitalizace dat z různých analogových měřících zařízení
 - digitální zpracování obrazu
 - ...

Ukázka digitalizovaného signálu

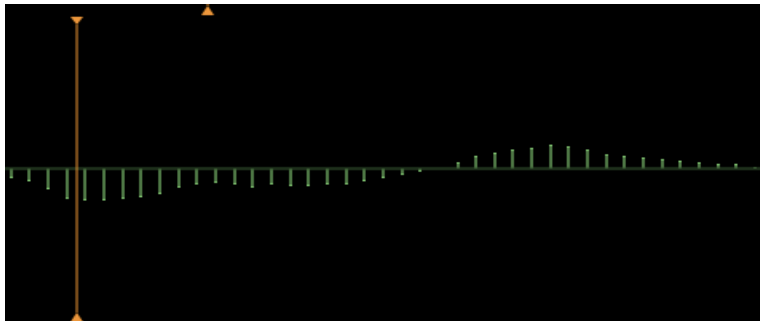


Figure: Ukázka digitalizovaného zvuku

Shanonův vzorkovací teorém

- Analogový signál $s(t)$ lze rekonstruovat z hodnot vzorků $s(nT)$ následovně:

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT) \frac{\sin(\pi(\frac{t}{T} - n))}{\pi(\frac{t}{T} - n)}$$

právě tehdy když je vzorkovací frekvence alespoň dvojnásobkem nejvyšší frekvence obsažené ve vstupním signálu.

- Důsledky:
 - Vzorkovací frekvence by měla být alespoň dvojnásobkem nejvyšší frekvence vstupního signálu.
 - Je-li menší dochází ke zkreslení složek vyšších frekvencí.
 - Spor příznivců a odpůrců audio CD - je 44kHz dostačující vzorkovací frekvence pro hudbu?

Kvantizace

- Převod reálných navzorkovaných hodnot na celočíselné hodnoty.
- Počet celočíselných hodnot = počet úrovní kvantování
 - 256
 - 65 536
 - 16 777 216
- Kvantizační krok - reálný interval přiřazený kvantizované jednotce.
 - Na vstupu je signál s amplitudou 128 mA (-128 - 128 mA).
 - 8bitová kvantizace - 256 kvantizačních úrovní
 - kvantizační krok = $256 \text{ [mA]} / 256 \text{ [kvantizačních úrovní]} = 1 \text{ [mA]}$.

Kvantizace

pokračování

- Běžně používané kvantizace:
 - 8 bitů
 - 16 bitů
 - 24 bitů
- Realizováno pomocí A/D převodníků
 - součást zvukových karet
 - mobilních telefonů
 - ...

Běžně používané parametry digitalizace zvuku

- Vzorkovací frekvence:
 - 8 kHz - telefonní kvalita
 - 16 kHz - běžná řeč
 - 22 kHz - rozhlasová kvalita
 - 44 kHz - audio CD
 - 48 kHz - DVD
- kvantizace:
 - 8 bitů
 - 16 bitů
 - 24 bitů
- počet audio kanálů
 - 1
 - 2
 - 4
 - 6 (5.1, 5 směrových kanálů + basy)

Způsoby kódování signálu

- PCM - přímé ukládání hodnot získaných kvantizací.
- výhody:
 - jednoduché na zpracování
 - nedochází k další ztrátě informací
- nevýhody:
 - Často malé rozdíly mezi hodnotami sousedních vzorků - značná redundance dat.
 - Konstantní hodnota kvantizačního kroku (závisí na parametrech AD převodníku);,
 - V případě malé amplitudy vstupního signálu - ztráta informace (signál nepřekročí kvantizační krok).
 - V případě velké amplitudy - hodnota překročí rozsah - zkreslení signálu.
 - Oba případy brání kvalitní rekonstrukci původního signálu.

Způsoby kódování signálu

Způsoby řešení nedostaků kódování PCM

- Diferenční PCM - uchovávání rozdílů sousedních vzorků místo uchovávání jejich hodnot. Hodnota rozdílu bývá podstatně menší než hodnota vzorku - lze uchovat pomocí méně bitů.
- Adaptivní PCM - kvantizační krok se určuje na základě amplitudy vstupního signálu.

Metody krátkodobé analýzy

- Zvuk je "periodický" pouze na krátkém intervalu.
- Zpracování signálu na krátkém časovém intervalu, kde se nepředpokládají výraznější dynamické změny (*mikrosegment*).
 - 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 prvky mikrosegmentu
- Nejběžněji používané typy okének:
 - hammingovo
 - pravoúhlé

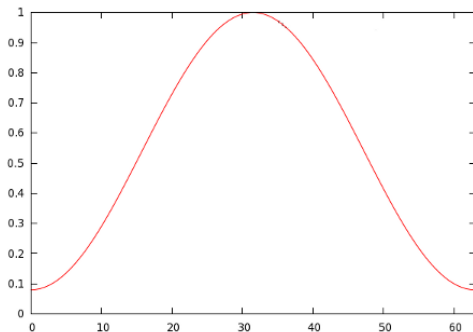
Hammingovo okénko

- Pro výpočet n -té váhy se využívá vztah

$$w(n) = \begin{cases} n = 0 \dots N - 1 & 0.54 - 0.46 \cos(2\pi n / (N - 1)) \\ n < 0 \vee n \geq N & 0 \end{cases}$$

N - počet vzorků v mikrosegmentu

- Hammingovo okénko pro mikrosegment délky 64



Pravouhlé okénko

- Přiřadí každému prvku mikrosegmentu váhu 1:

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

N - délka mikrosegmentu

Analýza digitalizovaného signálu v časové oblasti 1.

- Vychází se přímo z hodnot vzorků, nikoliv z hodnot spektra.
- Používá se funkce krátkodobé energie:

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

- Ukázka výpočtu funkce krátkodobé energie v Octave (octave/ste.m)
- $s(k)$ - vzorek v čase k , $w(n-k)$ - 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.

Analýza digitalizovaného signálu v časové oblasti 2.

- Funkce krátkodobé intenzity:

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

- Používá se např. pro detekci ticha.

Ukázka průběhu funkce krátkodobé energie

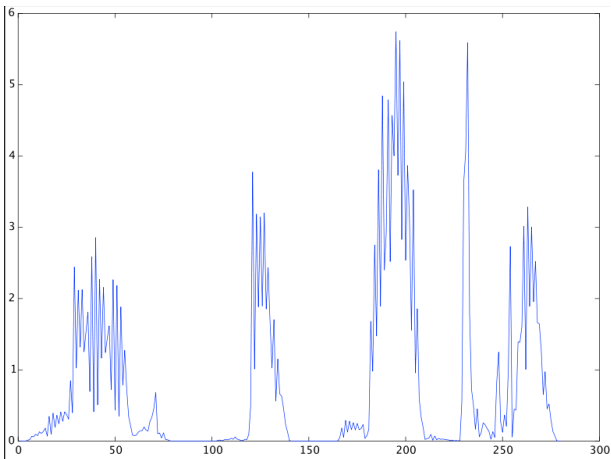


Figure: Ukázka průběhu funkce krátkodobé energie

Ukázka průběhu funkce krátkodobé intenzity

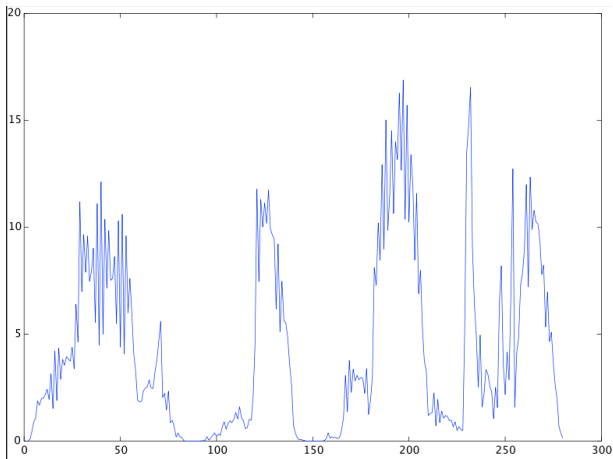


Figure: Ukázka průběhu funkce krátkodobé intenzity

Analýza digitalizovaného signálu v časové oblasti 2.

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

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

- varianta - počet lokálních extrémů
 - obě mohou být negativně ovlivněny šumem zvukového pozadí
- Diferenční klasifikátory:
 - difference prvního řádu

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

Ukázka průběhu funkce středního počtu průchodů nulou

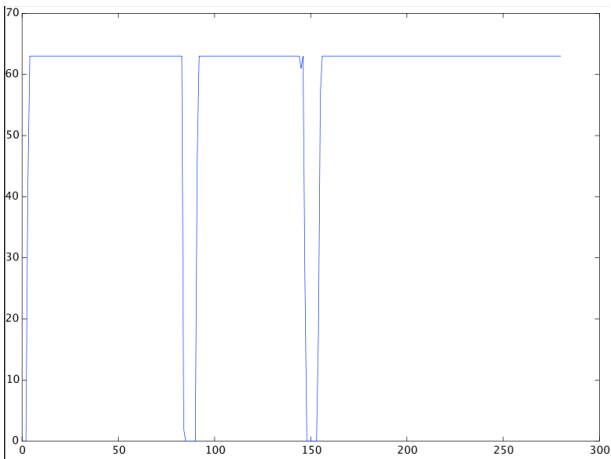


Figure: Ukázka průběhu funkce středního počtu průchodů nulou

Analýza digitalizovaného signálu v časové oblasti 3.

- Krátkodobá autokorelační funkce:

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

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

Zpracování signálu ve frekvenční oblasti

- Nejvíce používané:
 - krátkodobá Fourierova transformace
 - keprální analýza
 - lineární prediktivní analýza

Fourierova řady

- $f(x)$ - periodická spojitá funkce s periodou T

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

- Způsob výpočtu koeficientů a_i a b_j :
 - $\alpha, \alpha + T$ - 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:

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

- $|X(k)|$ - intenzita k. spektrálního koeficientu; frekvence závisí na velikosti mikrosegmentu N a vzorkovací frekvence T
- $x(n)$ - n. vzorek daného mikrosegmentu.
- $W_n = e_j * 2\pi/N = \cos(2\pi/N) + j * \sin(2\pi/N)$

Inverzní Diskrétní Fourierova Transformace (IDFT)

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

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

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.

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 - $X(k) = \text{IFFT}(\text{FFT}(x(k)))$
- Kepstrální analýza umožňuje z řeči oddělit parametry buzení a parametry hlasového ústrojí.
- Využití:
 - ocenění fonetické struktury řeči
 - znělost
 - F_0, F_1, F_2, \dots
 - 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 :

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

- Použití:
 - určování 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.

Software pro analýzu signálu

- HTK (<http://htk.eng.cam.ac.uk/>) - Hidden Markov Model Toolkit (Engineering Department of Cambridge University) - toolkit pro tvorbu rozpoznávačů řeči založených na skrytých Markovových modelech.
- ESPS toolkit (<http://www.speech.kth.se/software/#esps>)
- NICO toolkit (<http://nico.nikkostrom.com/>) - toolkit pro vytváření umělých neuronových sítí, využívá se např. pro rozpoznávání řeči.
- Matlab - knihovny pro analýzu řeči
 - labrosa.ee.columbia.edu/matlab/
(<http://labrosa.ee.columbia.edu/matlab/>)
 - Audio processing in Matlab
(http://www.umiacs.umd.edu/~ramani/cmssc828d_audio/Audio%20processing%20using%20Matlab.ppt)
 - ...
- Octave