

SIN04: Řečová interakce a sociální sítě

Luděk Bártek

Fakulta Informatiky
Masarykova Univerzita

podzim 2018

Obsah

- 1 Digitalizace zvuku
 - Vzorkování
 - Kvantizace
 - Kódování průběhu vlny
- 2 Zpracování digitalizovaného zvuku
 - Analýza zvuku v časové oblasti
 - Analýza zvuku ve frekvenční oblasti

Digitalizace zvuku

- 1 Vzorkování – snímání aktuální výchylky akustického signálu
 - převod spojitého signálu na posloupnost diskretních reálných hodnot.
- 2 Kvantizace – převod reálných hodnot na celočíselné.
- 3 Kódování průběhu vlny.

Vzorkování

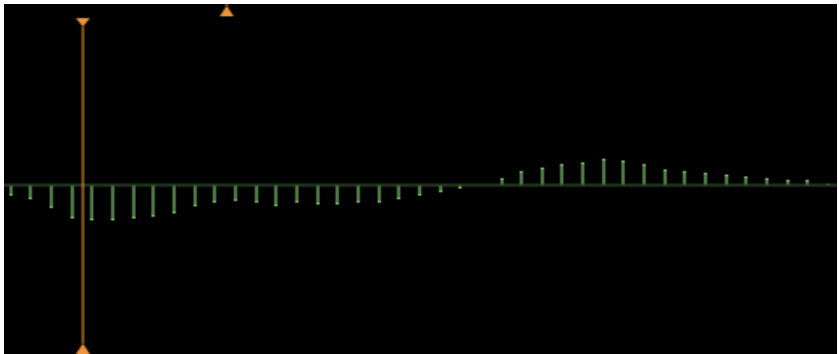
- Převod spojitého signálu $s(t)$ na posloupnost diskretních hodnot $s_n(t)$.
 - V daném okamžiku se sejme hodnota zvolené veličiny vstupního signálu (napětí, proud, ...).
 - Vzorkování se děje s periodou T .
 - Vzorkovací frekvence $f = \frac{1}{T}$.
 - Takto získané hodnoty jsou následně kvantizovány.
- Pokud nemá dojít ke ztrátě informace obsažené v signálu, pak vzorkovací frekvence musí být minimálně dvojnásobkem nejvyšší frekvence, která je v signálu obsažena (Shannonův vzorkovací teorém).

Běžně používané vzorkovací frekvence

- 8 kHz – telefonní kvalita
- 16 kHz
- 22 050 Hz – rozhlasová kvalita
- 44 100 Hz – CD kvalita
- 48 kHz – DVD kvalita

Vzorkování

Ukázka



Kvantizace

- Metoda převodu spojitých hodnot na diskrétní (reálných na celočíselné, ...).
- Princip:
 - Chceme kvantizovat vstupní hodnoty z intervalu $< min, max >$
 - Spočítáme kvantizační krok $step = \frac{max-min}{N}$ kde N je počet různých výstupních hodnot.
 - Pokud vstupní hodnota překročí k-násobek kvantizačního kroku, na výstup jde hodnota k.
- Kvantizační chyba
 - zaokrouhlovací chyba způsobená velikostí kvantizačního kroku
 - je přímo úměrná velikosti kvantizačního kroku.

Běžně používané kvantizace

- Celočíselné:
 - 8 bitů – 256 úrovní
 - 16 bitů – 65 536 úrovní
 - 24 bitů – 16 777 216 úrovní
 - 32 bitů – 4 294 967 296 úrovní – používá se hlavně pro zpracování obrazu.
- Reálné:
 - 32 bitů: 24 bitů znaménková mantisa, 8 bitů exponent
 - 64 bitů: 52 bitů znaménková mantisa, 11 bitů exponent

Kvantizace

Dokončení

- Využití:
 - zpracování zvuku na počítači
 - audio stopa na blu-ray discích (kódování MPEG-4)
 - ...

Určení kvantizačního kroku

Ukázka

- Rozsah hodnot vstupního signálu $\langle -127mV, 128mV \rangle$
 - velikost intervalu vstupních hodnot 256 mV
- 8bitová kvantizace
 - 256 úrovní signálů
- Kvantizační krok
 - $\frac{256mV}{256} = 1mV$ – změna vstupní úrovně napětí o 1 mV – změna výstupní hodnoty o 1.
 - např. změna vstupního napětí z 0,5 mV na 1,1 mV – změna hodnoty na výstupu z 0 na 1.

Kódování průběhu vlny

- Pulsní kódová modulace
 - přímo ukládá hodnoty, které jsou výstup z kvantizace.
- Nevýhody:
 - Relativně pomalé změny zvukového signálu \Rightarrow relativně malé rozdíly sousedních vzorků \Rightarrow velká redundance dat. – řešení diferenční PCM – ukládají se rozdíly mezi sousedními vzorky.
 - V případě příliš velkých změn amplitudy signálu problém s nastavením kvantizačního kroku.
 - příliš velký krok – ztráta informace o částech s malou amplitudou
 - příliš malý krok – přetečení hodnot v částech s velkou amplitudou.
 - Řešení – adaptivní PCM – kvantizační krok se určí v závislosti na amplitudě signálu.

Zpravování digitalizovaného signálu

- Zvuk mívá velkou dynamiku
 - Většina charakteristik zvuku je neměnná pouze v rámci krátkých časových úseků – metody krátkodobé analýzy.
- Mikrosegment
 - časový interval, na kterém předpokládáme neměnnost charakteristik zvuku.
 - používaná velikost 10 — 40 ms – závisí na použité metodě
- Metody krátkodobé analýzy
 - v časové oblasti – pracují přímo s hodnotami vzorků
 - ve frekvenční oblasti – z hodnot vzorků se získají frekvenční charakteristiky, které jsou následně zpracovány.
- Modelování funkce Coortiho ústrojí – matematická simulace rezonance vybraných vláček Coortiho ústrojí.

Váhové okénko

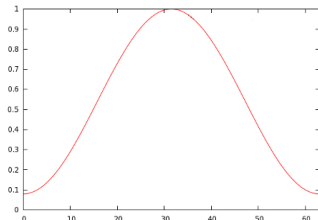
- Pro účely krátkodobé analýzy předpokládáme, že je signál signál v okolí mikrosegmentu periodický se stejnou periodou jako uvnitř mikrosegmentu.
- Vzniklou chybu lze kompenzovat použitím „okénka“.
- Okénko - posloupnost vah pro jednotlivé vzorky mikrosegmentu.
- Váhy odpovídají tomu, jak je vzorek pro účely dané metody ovlivněn okolím mikrosegmentu.
 - Čím více je vzorek ovlivněn okolím mikrosegmentu, tím má přiřazenu nižší váhu.
- Nejčastěji používané typy okének:
 - pravoúhlé okénko
 - Hammingovo okénko.

Pravoúhlé okénko

- Vychází z předpokladů:
 - 1 Vzorky uvnitř mikrosegmentu nejsou pro naše potřeby ovlivněny okolím.
 - 2 Všechny vzorky uvnitř mikrosegmentu jsou ovlivněny stejně.
- Všechny vzorky uvnitř mikrosegmentu mají stejnou váhu $w(s) = 1$.
- Váha vzorků mimo mikrosegment $w(s) = 0$.

Hammingovo okénko

- Vychází z předpokladu, že čím je vzorek blíž okraji mikrosegmentu, tím více je ovlivněn okolím.
- Váha vzorků uvnitř mikrosegmentu
 $w(s_n) = 0,54 - 0,46\cos\left(\frac{2\pi n}{N-1}\right)$
 - N - počet vzorků v mikrosegmentu.
- Váha vzorků mimo mikrosegment $w(s) = 0$.



Analýza digitalizovaného zvuku v časové oblasti

- Při analýze se vychází přímo z hodnot vzorků, nikoliv z hodnot spektra.
- Používané metody:
 - metoda krátkodobé energie
 - metoda krátkodobé intenzity
 - funkce středního počtu průchodu nulou
 - difference 1. řádu
 - autokorelační funkce
 - ...

Metoda krátkodobé energie

- Využívá funkci průměrné energie v segmentu

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

- $s(k)$ – vzorek v čase k
- $\omega(n-k)$ – váha okénka pro vzorek v čase k .
- Druhá mocnina zvyšuje dynamiku zvukového signálu.
- Použití:
 - automatická detekce ticha a promluvy
 - tvorba příznaků pro jednoduché klasifikátory slov
 - oddělení znělých a neznělých částí promluvy.

Metoda krátkodobé intenzity

- Využívá funkci krátkodobé intenzity na daném mikrosegmentu

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

- $s(k)$ – vzorek v čase k
- $\omega(n-k)$ – váha okénka pro vzorek v čase k
- Použití – stejné jako u průměrné energie.
- Oproti krátkodobé energii nezvýrazňuje tolik dynamiku řečového signálu.

Funkce středního počtu průchodů nulou

- Počítá změny znaménka digitalizovaného signálu na daném mikrosegmentu.

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

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

Autokorelační funkce

- Vrací podobnost úseků daného mikrosegmentu posunutých o m vzorků – čím větší hodnota, tím jsou si vzorky podobnější.

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

- Je-li funkce periodická s periodou P , potom, $R(m, n)$ nabývá maxima pro $m=P, 2P, \dots$
- Předpokládá délku mikrosegmentu aspoň $2P$.
- Použití:
 - zjištění periodicity řeči a určení základního tónu řeči
 - základ pro výpočet koeficientů LPA.

Analýza zvuku ve frekvenční oblasti

- Ze vstupních vzorků získává akustické spektrum.
- Nejpoužívanější metody:
 - krátkodobá Fourierova transformace
 - krátkodobá diskrétní Fourierova transformace
 - rychlá Fourierova transformace
 - kepstrální analýza
 - lineární prediktivní analýza.

Krátkodobá Fourierova transformace

- Vychází z Fourierovy transformace:
 - Krátkodobá Fourierova transformace

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

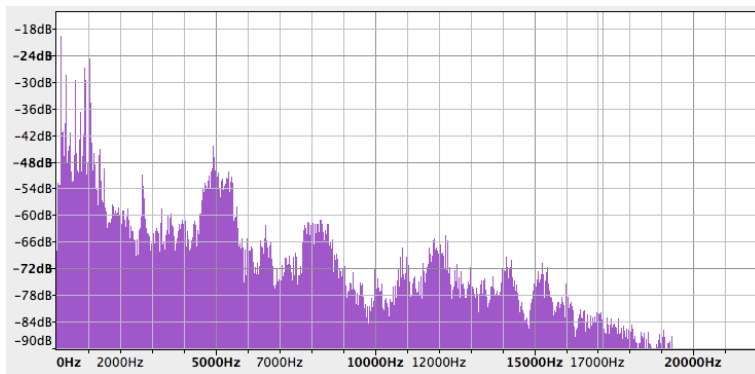
- Fixací času t získáme klasickou Fourierovu transformaci.
 - $|S(\omega, t)|$ – amplituda akustického spektra odpovídajícího frekvenci ω v čase t .
 - h – váhová funkce okénka.
- Předpokládá na vstupu periodickou funkci – zvuk je periodický na krátkých časových úsecích.
- Předpokládá periodické opakování daného mikrosegmentu.

Krátkodobá diskrétní Fourierova transformace a rychlá Fourierova transformace

- Používá se na výpočet spektra periodických posloupností s periodou N , resp. posloupností délky N .
- Frekvence odpovídající spektrálním koeficientům závisí na délce mikrosegmentu a vzorkovací frekvenci.
- Diskrétní Fourierova transformace
 - výpočetně časově náročná – n^2 výpočtů nad komplexními čísly
 - prakticky nelze použít pro výpočty v reálném čase.
- V praxi se používá Rychlá Fourierova transformace (FFT).
 - složitost $n \log(\frac{n}{2})$ operací násobení
 - algoritmus postaven na metodě rozděl a panuj
 - vyžaduje, aby délka mikrosegmentu byla mocninou dvou.

Spektrum zvuku á

Figure: FFT spektrum hlásky á

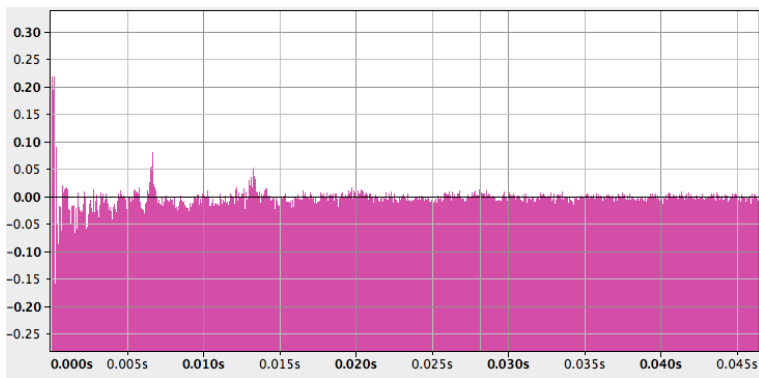


Kepstrální analýza

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

Kepstrum hlásky á

Figure: Kepstrum hlásky á



Lineární prediktivní analýza

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

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

- Použití:
 - Určení charakteristik modelu hlasového ústrojí.
 - Z chyby predikce lze:
 - odvodit poznatky o znělosti
 - určit frekvenci základního hlasivkového tónu.
 - Získané koeficienty lze použít jako příznaky pro rozpoznávání řeči – nesou informaci o spektrálních