MULTIMEDIÁLNÍ A HYPERMEDIÁLNÍ SYSTÉMY

5)
Uložení a komprese zvuku

Petr Lobaz, 10. 3. 2015

ZVUKOVÝ ZÁZNAM

ANALOGOVÝ

- záznam akustického tlaku (resp. napětí) spojitou veličinou
- citlivý na stav záznamového média
  - médium časem degraduje
  - ale: plošná hustota záznamu poměrně malá
  - se vzrůstajícím poškozením média „rovnoměrně" roste zkreslení (poškození) záznamu
  - dá se chápat jako nevýhoda i jako výhoda (archivnictví!)
- citlivý na kvalitu záznamového i přehrávacího zařízení
- běžně pro monofonní a stereofonní záznam
- vícestopé záznamy komplikovanější – stopy musí být synchronní, kvalita všech čtecích hlav stejná, …

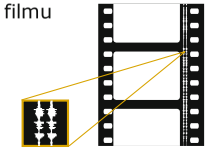
⇒ nástup digitálních formátů

ZVUKOVÝ ZÁZNAM

- některé principy zavedené v analogovém záznamu se používají stále (úmyslně i neúmyslně)
- mechanický
  - hrací strojky, orchestriony, … ⇒ princip podobný notovému záznamu ⇒ digitální záznamy typu MIDI
  - fonograf, gramofon – záznam průběhu napětí na mikrofonu hloubkou drážky ⇒ digitalizace vede na PCM záznam apod.
  - některá technická omezení se vžila (ve stereofonní gramofonové nahrávce nesměl být jeden kanál výrazně hlasitější než druhý ⇒ umístění bicích na střed stereofonní báze – používá se bez technického důvodu i dnes)

ZVUKOVÝ ZÁZNAM

- optický – záznam napětí „průhlednosti" filmu
  - pro zvukové filmy – zvuková stopa proužek (proužky) proměnlivé šířky vedle obrazu
  - na film časem přibývaly další typy záznamu zvuku (magnetický, optický komprimovaný), ale základní stereostopa zůstala pro případ problému s „lepšími" formáty



- magnetický – záznam magnetizací materiálu
  - magnetofon, magnetická stopa na filmovém pásu
  - často vícestopé záznamy
  - magnetická média se v digitálním světě používají nadále (páska – archivace, pevný disk – externí paměť)

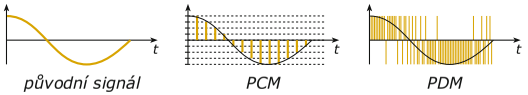
ZVUKOVÝ ZÁZNAM

DIGITÁLNÍ

- záznam akustického tlaku (resp. napětí) veličinou diskrétní v času i v hodnotě (vzorkování, kvantizace)
- umožňuje velkou plošnou hustotu zázamu
  - náchylný k poškození
  - ⇒ záznam typicky opatřený samoopravným mechanismem
  - ⇒ při malém poškození média zvuk stále perfektní, při větším poškození prudká degradace (až zničení)
- při záznamu je třeba A/D převodník, při reprodukci D/A
  - kvalita digitálního zvuku záleží především na nich
  - samotné uložení záznamu, distribuce, zpracování v jistém smyslu „bezšumové"

ZVUKOVÝ ZÁZNAM

- příznakový
  - kódování „not" (MIDI) – obdoba vektorové grafiky
  - kódování charaketru zvuku, např. při kódování hlasu
- vzorkovaný záznam
  - digitalizace průběhu napětí na mikrofonu
  - nejčastěji PCM (pulse code modulation)
  - existují i alternativy, ale ve světě multimédií se užívají spíše zřídka: PDM (pulse density modulation), resp. ΣΔ (sigma-delta modulation) – signál rozložen do série pulsů (např. Super Audio CD)



původní signál *PCM* *PDM*

VÍCESTOPÝ ZÁZNAM

ODDĚLENÉ KANÁLY

- perfektní oddělení signálů
- přehrávač musí příslušný počet kanálů podporovat
  - záznam není zpětně kompatibilní se staršími přehrávači
- stereo – užití digitální i analogové
- surround (5.1 apod.) – užití obvykle jen digitální
  - analogová podoba prakticky jen ve zvukových studiích, jistou dobu na 70mm filmovém pásu

ZPĚTNĚ KOMPATIBILNÍ ODDĚLENÉ KANÁLY

- např. FM rádio: vstup L, R; přenos L+R, L−R
- monofonní přijímač umí přijmout jen L+R ⇒ kvalitní mono
- stereofonní přijímač z L+R, L−R dekóduje L, R

VÍCESTOPÝ ZÁZNAM

MATICOVÉ ULOŽENÍ

- uložení několika kanálů do menšího počtu stop
  (např. left, right, center, surround → $left_{TOTAL}$, $right_{TOTAL}$)
  (dále jen zkratky L, R, C, S, L_T, R_T)
- nezvyšuje kapacitu přenosového kanálu,
  dobrá zpětná kompatibilita
- neumožňuje dokonalé oddělení původních kanálů
- typický příklad:
  Dolby Motion Picture (MP) kodér: L, C, R, S → L_T , R_T
  Dolby Surround dekodér (pasivní): L_T , R_T → L, R, S
  Dolby ProLogic dekodér (aktivní): L_T , R_T → L, R, C, S

VÍCESTOPÝ ZÁZNAM

DOLBY MP KODÉR

- low-pass filtrace: S → S‘
- L_T = L{0°} + 0,707 C{0°} + 0,707 S‘{90°}
  R_T = 0,707 C{0°} + R{0°} – 0,707 S‘{90°}
  (složené závorky: dodatečný fázový posuv)
  (násobení 0,707: útlum –3 dB ⇒ polovina energie)

- stereofonní přehrávání L_T, R_T – zkoumáme směr originálního a reprodukovaného zvuku:
  - originál z L (R) ⇒ reprodukce vlevo (vpravo)
  - originál z C ⇒ virtuální stereofonní zdroj uprostřed
  - originál z S ⇒ v reprodukci v L_T, R_T s opačnou polaritou ⇒ obtížná lokalizace virtuálního zdroje

VÍCESTOPÝ ZÁZNAM

DOLBY SURROUND DEKODÉR

- $L_{OUT} = L_T$
  $R_{OUT} = R_T$
  S_{OUT} = zpožděný, lowpass filtrovaný signál $L_T – R_T$
- jaký je výstup S_{OUT} reproduktoru?
  originál z C ⇒ v S_{OUT} dokonalé odečtení
  originál z S ⇒ objeví se v S_{OUT}
- zpoždění signálu ($L_T – R_T$) ⇒ využití Haasova jevu – zvuk je stále primárně vnímán zepředu (z obrazu) – při sledování filmu méně rušivé než zvuky „zezadu“
- lowpass filtrace signálu ($L_T – R_T$) ⇒ zvuk zní vzdáleněji, lze jej hůře lokalizovat
- fázový posun S{90°} ⇒ oddělení současných zvuků v C a S

VÍCESTOPÝ ZÁZNAM

DOLBY PRO LOGIC DEKODÉR

- pokud byl originální zvuk např. virtuálně mezi L a C, nedokáže pasivní dekodér korektně kanály oddělit:
  např. v $S_{OUT} = L_T – R_T$ už není dokonalé odečtení
- řešení:
  $L_{OUT} = a_{L1}L_T + a_{L2}R_T$
  $R_{OUT} = a_{R1}L_T + a_{R2}R_T$
  $S_{OUT} = a_{S1}L_T + a_{S2}R_T$
  $C_{OUT} = a_{C1}L_T + a_{C2}R_T$
- parametry a_{XY} (tj. hodnota, znaménko) určovány průběžně podle povahy L_T , R_T
  - lepší oddělení kanálů, v plném frekvenčním rozsahu
  - umožňuje oddělit i C_{OUT} kanál

VÍCESTOPÝ ZÁZNAM

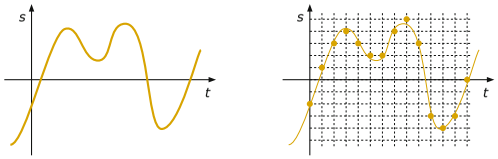
- maticové uložení se používá např. pro uložení 6.1, 7.1 zvuku apod. do 5.1 digitálního zvukového záznamu
  - standardní 5.1 přehrávač přehraje 5.1 zvuk
  - vylepšený přehrávač dekóduje zbylé kanály a přehraje 6.1 apod. (Dolby Digital EX)

MIDI

- Musical Instrument Digital Interface
- definuje příznakový systém záznamu zvuku, např.:
  - v čase t_0 prudce stisknuta klávesa C5
  - v čase t_1 prudce stisknuta klávesa E5
  - v čase t_2 pomalu uvolněna klávesa C5
  - …
- MIDI příkazy generuje tzv. controller (hudební nástroj, např. klávesy) nebo tzv. sequencer (např. počítač)
- MIDI příkazy přijímá tzv. sound module – generuje reálný zvuk
- alternativně může MIDI zařízení přijmout MIDI příkaz, generovat další a odesílat jinému zařízení, …
- typický datový tok 31,25 kbit/s

PCM

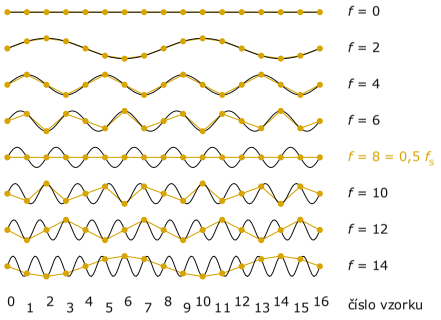
- Pulse Code Modulation
- diskretizace spojitého signálu
  - v čase: vzorkování – převod $s(t)$ na $s[i] = s(i \times \Delta t)$
  - v hodnotách: kvantizace – „převod na celá čísla"
- PCM záznam vnáší do signálu chybu (šum)
  - snaha o minimalizaci



původní signál *PCM záznam*

VZORKOVÁNÍ

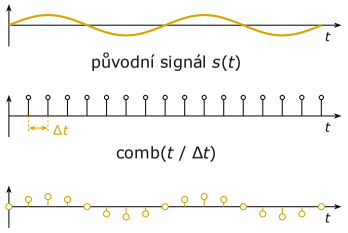
- vzorkování funkce $s(t) = \sin(2\pi f t)$ pro $t \in [0, 1]$
- $\Delta t = 1/16$
  $\Rightarrow f_s = 16$
- A/D převod: zjištění funkční hodnoty
- D/A převod: lineární interpolace
- nefunguje pro $f \geq 0{,}5 f_s$



VZORKOVÁNÍ

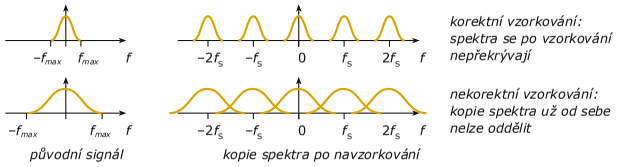
PROCES VZORKOVÁNÍ TEORETICKY

- vstupní signál: $s(t)$
- vzorkovaný signál: $s(t) \times \mathrm{comb}(t / \Delta t)$
- comb(t)
  - rovno nule mimo celočíselná t
  - v celočíselných t se „blíží nekonečnu"
  - není to klasická funkce, matematické podrobnosti vynecháme



VZORKOVÁNÍ

- spektrum signálu $s(t)$ je dáno FT{ $s(t)$ } = $S(f)$
- spektrum navzorkovaného signálu:
  FT{ $s(t)$ × comb(t / Δt) }
  = FT{ $s(t)$ } ⊗ FT{comb(t / Δt) }
  = $S(f)$ ⊗ comb($f\,\Delta t$) × $|\Delta t|$
- konvoluce $S(f)$ hřebenovou funkcí s roztečí 1 / Δt
  ⇒ součet kopií $S(f)$ posunutých o k / Δt = $k \times f_S$

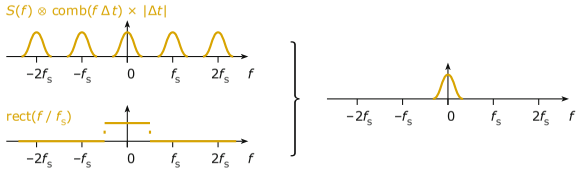


korektní vzorkování: spektra se po vzorkování nepřekrývají

nekorektní vzorkování: kopie spektra už od sebe nelze oddělit

VZORKOVÁNÍ

- je-li vzorkovací frekvence $f_S \geq 2f_{max}$,
  lze po navzorkování rekonstruovat původní spektrum:
  $S(f) = \left(S(f) \otimes \text{comb}(f\,\Delta t) \times |\Delta t|\right) \times \text{rect}(f / f_S)$
- funkce $\text{rect}(f) = \begin{cases} 1 & \text{pro } |f| < 0{,}5 \\ 0{,}5 & \text{pro } |f| = 0{,}5 \\ 0 & \text{pro } |f| > 0{,}5 \end{cases}$

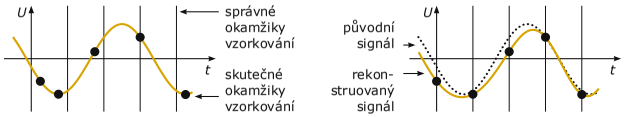


VZORKOVÁNÍ

- neplatí-li $f_S \geq 2f_{max}$ (tj. platí $f_{max} \geq 0{,}5f_S$),
  navzorkovaný signál nelze korektně rekonstruovat
  vzniklá chyba: *aliasing*
- v praxi je nutné před A/D převodník zařadit dolní propust (low-pass filtr) – ze signálu odstraní frekvence $|f| \geq 0{,}5f_S$
- konstrukce kvalitní analogové dolní propusti složitá (frekvence $|f| < 0{,}5f_S$ ponechá beze změny, vyšší potlačí)
- řešení:
  - filtrace méně kvalitní dolní propustí
    (ponechá $|f| < 0{,}5f_S$, potlačí $|f| > 0{,}5f_{OS}$)
  - vzorkování frekvencí $f_{OS} > f_S$ (oversampling, např. $f_{OS} = 4f_S$)
  - digitální lowpass filtrace pro odstranění $|f| \geq 0{,}5f_S$
  - vynechávání vzorků ⇒ redukce na vzorkování f_S

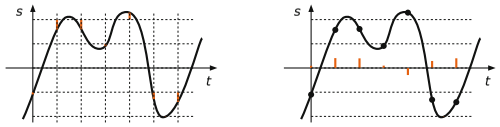
VZORKOVÁNÍ

- nejčastější vzorkovací frekvence f_S pro zpracování zvuku:
  - telefonie 8 kHz (stačí pro srozumitelnost hlasu)
  - rádio 32 kHz (kompromis mezi kvalitou a objemem dat)
  - Audio CD 44,1 kHz (plný záznam 20 Hz – 20 kHz)
  - zvukové stopy k filmu 48 kHz
  - v profesionálních aplikacích pro zpracování zvuku vyšší (až 192 kHz) – snazší konstrukce filtrů
- kvalita A/D i D/A převodu záleží také na přesnosti hodin generujících pulsy – problémy s nepřesností označujeme *jitter*



KVANTIZACE

- přidání „kvantizačního šumu" k signálu
  - „šum" = zaokrouhlená hodnota – skutečná hodnota
- jemnost kvantizace podle míry akceptovatelného šumu
- odstup signál–šum: 1 bit ≈ 6 dB (tj. 16 bitů ≈ 96 dB)
  - při n-bitové kvantizaci maximální amplituda $A_{max} = 2^{n-1}$
  - amplituda kvantizačního šumu $A_Q = 1 / 2$
  - ⇒ dynamický rozsah = $20 \log(A_{max} / A_Q) = 20 \log 2^n$
    $= n \times 20 \log 2 = n \times 6{,}02$ dB



KVANTIZACE

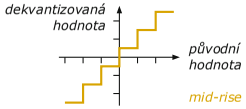
KVANTIZAČNÍ ŠUM

- velký počet kvantizačních úrovní
  - kvantizační šum „nezávislý" na signálu
  - zvukový charakter jako bílý šum
    (každá hodnota kvantizační chyby stejně pravděpodobná)
  - hodnota „odstup signál–šum" vypovídá o kvalitě signálu
- malý počet kvantizačních úrovní
  - kvantizační šum závislý na signálu
  - zvukový charakter není jako bílý šum
  - hodnota „odstup signál–šum" příliš nesouvisí s vnímanou kvalitou zvuku
  - vhodnější je hovořit o kvantizačním zkreslení

KVANTIZACE

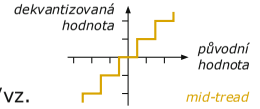
KVANTIZÉRY

- mid-rise: $q[i] = \mathrm{floor}(As(\mathrm{i} \times \Delta t))$ $\quad s(t) \in (-1, 1)$
  - $q[i]$ = číslo kvantizační hladiny, interpretace = dekvantizace
  - A je kvantizační hloubka, typicky $A = 2^{n-1}$
  - např. pro $As(t) \in [0, 1)$ je kvantizovaná hodnota 0
  - ⇒ chyba kvantizace ∈ [0, 1)
  - při interpretaci „kvantizační hladina 0" ⇔ „hodnota 0,5"
    je chyba ∈ [–0.5, 0.5)
  - ⇒ dekvantizace: $s'[i] = [\mathrm{floor}(As(\mathrm{i} \times \Delta t)) + 0{,}5] / A$
  - 2^n úrovní se dělí symetricky
    na kladné a záporné hodnoty
  - neobsahuje kód pro ticho
  - ⇒ pro zvuk se používá zřídka



KVANTIZACE

- mid-tread: $q[i] = \mathrm{floor}(As(\mathrm{i} \times \Delta t) + 0.5)$
  - obvykle chceme stejný počet kladných a záporných kódů
  - navíc kód pro 0
  - ⇒ lichý počet úrovní ⇒ neefektivní uložení pro malé n
  - příklad: $n = 2$ bity na vzorek
    převod kód → hladina: 00 → 0, 01 → 1, 10 → –1
    kód 11 nepoužit ⇒ 25 % kapacity kódu nepoužito
  - sdružování několika vzorků do jednoho kódu
  - příklad: kódy 00, 01, 10 interpretovány jako 0, 1, 2
    kódování vzorků $s[0]$, $s[1]$, $s[2]$
    jedním kódem
    $s[0] + 3s[1] + 3^2 s[2]$
    $\in [0, 26]$
    ⇒ 5 bitů na 3 vzorky ≐ 1,67 bit/vz.



KVANTIZACE

UNIFORMNÍ KVANTIZACE

- uniformní = šířka kvantizačních intervalů stejná
- CD, DVD, …
- kódování pro 12 a více bitů
- odhad dynamického rozsahu 16bitové kvantizace (odstup signál / šum):
  16 × 6,02 dB ≈ 96 dB
- při korektním výpočtu je dynamický rozsah n-bitového bipolárního signálu s rovnoměrně rozloženou kvantizační chybou (1,76 + 6,02n) dB
- pro unipolární signál (tj. např. pro obraz) je to (7,78 + 6,02n) dB

KVANTIZACE

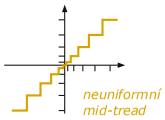
KOREKTNÍ VÝPOČET ODSTUPU SIGNÁL-ŠUM

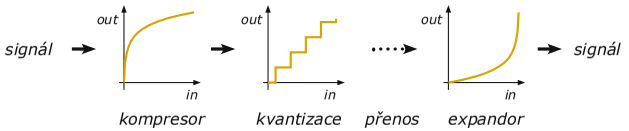
- n-bitový signál
  velikost kvantizačního intervalu q
  maximální amplituda signálu: $A_{max} = q \times 2^{n-1}$
  výkon signálu: $P = {}^{1}/_{2\pi} \int_0^{2\pi} (A_{max} \sin(t))^2 \, dt = {A_{max}}^2 / 2$
- maximální kvantizační chyba Δ: $\Delta_{max} = q / 2$
- rovnoměrné rozložení ppsti kvantizační chyby ⇒
  střední hodnota výkonu (čtverce) kvantizační chyby:
  $Q = {}^{1}/_{q} \int_{-q/2}^{q/2} \Delta^2 \, d\Delta = q^2 / 12$
- maximální poměr signál / šum: $P / Q = 2^{2n} \times 1{,}5$
  v dB: $10 \log(P / Q) = n \times 20 \log 2 + 10 \log 1{,}5 =$
  $= (6{,}02n + 1{,}76)$ dB

KVANTIZACE

NEUNIFORMNÍ KVANTIZACE

- velikost kvantizačního intervalu různá
  – v tichém zvuku je i malý šum patrný
  ⇒ pro nízké úrovně kvantizace jemná, pro vysoké úrovně hrubá
- praktická implementace: companding – komprese dynamického rozsahu, uniformní kvantizace, expanze
- vhodné při malém počtu (např. 8) bitů na vzorek (telefonie)



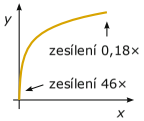


KVANTIZACE

STANDARDNÍ NEUNIFORMNÍ KVANTIZACE

- vstupní signál $x(t) \in (-1, 1)$, přenosové funkce liché
- μ-law (telefonie USA, Japonsko)

$$y = \frac{\ln(1+\mu x)}{\ln(1+\mu)} \quad \mu=255,\ 0 \le x \le 1$$

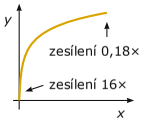


- A-law (mezinárodní telefonie)
  – menší zesílení šumu při $|x| \doteq 0$

$$y = \frac{1+\ln(Ax)}{1+\ln A} \quad \text{pro } 1/A \le x \le 1$$

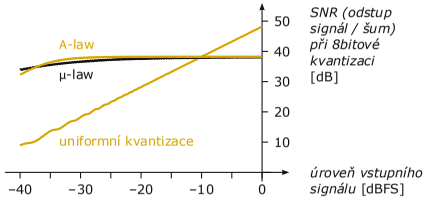
$$y = \frac{Ax}{1+\ln A} \quad \text{pro } 0 \le x \le 1/A$$

A=87,6



KVANTIZACE

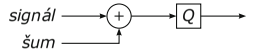
- obecná analýza odstupu signál / šum komplikovaná
- tvary funkcí A-law, μ-law umožňují zjisit vlastnosti kvantizace prostředky matematické analýzy
- alternativně: numerická simulace kvantizace, numerický výpočet (výkon signálu) / (výkon šumu)
- uniformní kvantizace: tišší signály mají horší SNR
- neuniformní kvantizace: SNR pro tiché i hlasité zvuky srovnatelný



KVANTIZACE

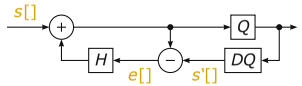
POTLAČENÍ KVANTIZAČNÍHO ŠUMU

- kvantizovaný signál není lineární:
  kvantizace($a \times s(t)$) ≠ $a \times$ kvantizace($s(t)$)
- linearita důležitá pro správnou činnost filtrů
- dithering
  - přidání šumu ±0,5 kvantizační úrovně před kvantizérem
  - kvantizovaný signál s ditheringem je statisticky lineární (tj. ne pro konkrétní čas t, ale průměrně na intervalu)
  - objektivní zhoršení poměru signál-šum (SNR, signal to noise ratio) cca o 3–5 dB
  - subjektivní zlepšení SNR
  - kvalitní ditheringový šum respektuje citlivost ucha



KVANTIZACE

- tvarování spektra šumu (noise shaping)
  - hodnota před kvantizací: $s[i]$ — 1,3
    hodnota po dekvantizaci: $s'[i] = Q(s[i])$ — 1
    chyba kvantizace: $e[i] = s[i] - s'[i]$ — +0,3
    před kvantizací: $s[i+1]$ — 1,4
    další vzorek: $s'[i+1] = Q(s[i+1] + e[i])$ — 2
  - přidávání šumu $e[i]$ kompenzujícího kvantizační chybu
  - šum $e[i]$ lze filtrovat filtrem H, aby respektoval citlivost ucha
  - lze kombinovat s ditheringem



NEKOMPRIMOVANÝ ZVUK

- nejvyšší kvalita, největší datový tok
- standard pro studiové zpracování
- např. Audio Compact Disc (CD)
  - 16 bitů na vzorek
  - stereofonní nahrávka (= 2 kanály)
  - vzorkovací frekvence f_S = 44 100 Hz
  ⇒ datový tok 16 × 2 × 44 100 = 1,41 Mbit/s
- pro srovnání: nekomprimovaný datový tok HDTV (1920 × 1080 pixelů, 25 snímků za vteřinu): 622 Mbit/s
- při dnešním výpočetním výkonu není problém nekomprimovaný zvuk zpracovávat ani archivovat

- rozhodnutí komprimovat/nekomprimovat ovlivněno jinými než technickými faktory

KOMPRESE ZVUKU

- bezztrátová – odstranění redundance v datech
  - po dekompresi požadujeme přesný duplikát vstupních dat
- ztrátová – odstranění neslyšitelných částí zvuku (a odstranění redundance v datech)

- obecná
  - libovolný realistický zvuk
  - obvykle testováno pro hudbu, hlas (ne pro obecný šum!)
  - obvykle požadujeme maximální věrnost
- speciální
  - hlas – obvykle požadujeme jen srozumitelnost
  - kódování stovek podobných zvuků pro MIDI Sound Module (všechny tóny klavíru apod.)
  - …

KOMPRESE ZVUKU

- konstantní datový tok (CBR)
  - snadná manipulace, snadný odhad potřebného výpočetního výkonu, paměťových nároků,
  - „ticho" komprimováno stejným datovým tokem jako složitý zvuk
- proměnný datový tok (VBR)
  - „ticho" kódováno velmi úsporně, složitý zvuk patřičným množstvím bitů
  - dá se chápat jako „konstantní kvalita zvuku"
  - paměťové nároky se nedají příliš odhadnout
- průměrný datový tok (ABR)
  - kompromis CBR × VBR
  - lokálně VBR, průměr přes dlouhý interval CBR
  - datový tok CBR také kolísá, ale proti ABR velmi málo

KOMPRESE ZVUKU

POŽADAVKY NA KOMPRESI

- pro archivaci – co nejlepší kvalita
- pro komunikaci – co nejmenší zpoždění
  - např. kodér pracující po blocích 384 vzorků při vzorkování 48 kHz zanáší zpoždění nejméně 8 ms
  - další zpoždění způsobené filtrací, prediktivním kódováním, …
  - např. mp3 typické zpoždění 130 ms, HE-AAC 320 ms
  - pro telefonii tolerovatelné 150 ms, v hudební produkci 30 ms
  - speciální low-latency kodéry CELT 5 ms, Opus 22 ms, AAC LD 35 ms, …
- obecně: co nejjednodušší (= nejlevnější) dekodér

BEZZTRÁTOVÁ KOMPRESE

- vstup: kvantizované vzorky $s[i]$ monofonní zvukové stopy (případně několik monofonních stop, např. dvě pro stereo)
- pokud se některé hodnoty vyskytují pravděpodobněji, je vhodné je uložit menším počtem bitů
  - ⇒ ideální je kódovat hodnotu, o které víme, že přijde (nemusíme uložit ani bit)
  - ⇒ snahou je $s[i]$ převést do podoby, kde několik hodnot má značnou pravděpodobnost, ostatní velmi malou
- pravděpodobnost hodnoty (symbolu) h je p
  - její zakódování vyžaduje $-\log_2 p$ bitů
    např. hodnota s ppstí $p = 0{,}25 = 2^{-2}$ vyžaduje 2 bity
  - v celém signálu délky N vzorků zabere hodnota h celkem $-Np \times \log_2 p$ bitů

BEZZTRÁTOVÁ KOMPRESE

- signál (zpráva) $s[]$ délky N složený z *nezávislých* vzorků $s[i]$, $s[i] \in \{h_0, h_1, \ldots\}$, pravděpodobnosti hodnot (symbolů) $h_0, h_1, \ldots$ jsou $p_0, p_1, \ldots$, zabere

  $N \times \sum_k -p_k \log_2 p_k$ bitů

  - hodnoty s pravděpodobností 0 se ze sumy vynechávají
  - suma ve výrazu se označuje pojmem *entropie*
  - výraz označuje teoretický limit, snažíme se mu přiblížit
- jsou-li vzorky $s[i]$ na sobě závislé, můžeme entropii podstatně snížit
- obecný postup:
  - dekorelace: odstranění závislostí mezi vzorky $s[i]$
  - uložení (zakódování) upraveného dekorelovaného signálu

BEZZTRÁTOVÁ KOMPRESE

PŘÍKLAD

- signál $s[]$ složený z hodnot {0, 1, 2, 3}
  0 0 0 1 1 1 2 2 2 2 3 3 3 3
- standardní uložení
  00 00 00 01 01 01 10 10 10 10 11 11 11 11
  (14 × 2 = 28 bitů)
- $p_0 = 3/14$, $p_1 = 3/14$, $p_2 = 4/14$, $p_3 = 4/14$
  teoreticky lze uložit na 27,79 bitu
- pozorování: hodnoty se často opakují
  dekorelace: uložení rozdílů mezi vzorky
  0 0 0 1 0 0 1 0 0 0 1 0 0 0
- pouze hodnoty {0, 1} ⇒ lze uložit jen 14 bity
  $p_0 = 11/14$, $p_1 = 3/14$ ⇒ teoreticky lze uložit na 10,49 bitu

BEZZTRÁTOVÁ KOMPRESE

KÓDOVÁNÍ HODNOT (SYMBOLŮ)

- hodnota (symbol) h_i má ppst p_i – kód má ideálně $-\log_2 p$ bitů
  ⇒ variable length coding (VLC)
- Huffmannovo kódování
  - symbol h_i kódován kódem (v jistém smyslu) optimální délky (celočíselným počet bitů – slabina Huffmanova kódu)
  - vyžaduje znalost pravděpodobností, kódování a dekódování vyžaduje tabulku kódů
  ⇒ vyžaduje preprocessing pro zjištění ppstí (nevhodné pro realtime kódování), nebo je tabulka kódů pevně daná
  - příklad: zpráva A A B C B B A A A EOF (10 symbolů)
    $p_A = 5/10$, $p_B = 3/10$, $p_C = 1/10$, $p_{EOF} = 1/10$
    kódy: A → 0, B → 10, C → 110, EOF → 111
    kód zprávy: 00101101010000111 (17 bitů)

BEZZTRÁTOVÁ KOMPRESE

- Riceovo kódování
  - nevyžaduje tabulku kódů, vhodné pro kódování signálu s exp. rozdělením ppsti hodnot: p(x) ~ exp(−|x|)
  - kód pro celé číslo I:
    bit pro znaménko I, m nejméně důležitých bitů I,
    z bitů 0 reprezentujících ostatní bity I, bit 1
  - příklad (m = 2)

0 (+00000) → 0 00 1	8 (+01000) → 0 00 00 1
1 (+00001) → 0 01 1	9 (+01001) → 0 01 00 1
2 (+00010) → 0 10 1	10 (+01010) → 0 10 00 1
3 (+00011) → 0 11 1	12 (+01100) → 0 00 000 1
4 (+00100) → 0 00 0 1	13 (+01101) → 0 01 000 1
5 (+00101) → 0 01 0 1	
7 (+00111) → 0 11 0 1	−17 (−10001) → 1 01 0000 1

BEZZTRÁTOVÁ KOMPRESE

- aritmetické kódování
  - poskytuje teoreticky nejlepší kód
  - výpočetně nejnáročnější
  - v kódování zvuku se používá zřídka (DVD-Audio, MPEG-4)
  - detaily později

DEKORELACE

- lineární predikce vzorků: $p[i] = a_1 x[i-1] + a_2 x[i-2] + \cdots$
  kódování rozdílu predikce $p[i]$ a skutečného vzorku $x[i]$
- hledání opakujících se sekvencí, transformace signálu, …
- dá se použít libovolná technika ze ztrátové komprese dat
  - výstup z kodéru ztrátově komprimovaný (dekorelovaný) zvuk + odchylka od původního signálu

BEZZTRÁTOVÁ KOMPRESE

BĚŽNÉ KODÉRY

- SHORTEN, FLAC (Free Lossless Audio Codec)
  lineární predikce vzorku FIR filtrem
- DVD-Audio (Meridian Lossless Packing)
  predikce vzorku IIR filtrem
- Dolby TrueHD, DTS-HD Master Audio (Blu-ray disc)

- obvyklý kompresní poměr 2 : 1 až 4 : 1
  pro srovnání: u ztrátových kodérů běžně 10 : 1 až 25 : 1

BEZZTRÁTOVÁ KOMPRESE

PŘÍKLAD PRINCIPU (FLAC)

1. rozdělení vstupního signálu na bloky
   výběr optimálních parametrů kódování bloku
2. odstranění redundance mezi kanály L, R
   např.: $x = (L + R) / 2$, $y = L - R$
   alternativně $x = L$, $y = R$ nebo $x = L$, $y = L - R$ nebo …
3. lineární predikce vzorku $x[i]$ (resp. $y[i]$)
   $p[i] = x[i-1] + (x[i-1] - x[i-2]) = 2x[i-1] + x[i-2]$
   alternativně jiné prediktory, např. $p[i] = 0$, …
4. určení chyby predikce (rozdílu predikce $x[i]$ a signálu $p[i]$):
   $e[i] = x[i] - p[i]$
5. uložení parametrů kódování bloku
   uložení $e[i]$ Riceovým kódem

ZTRÁTOVÁ KOMPRESE

PRINCIP

- dekorelace signálu
- *odstranění psychoakusticky nevýznamné informace*
- kódování zpracovaného signálu
- *řízení datového toku* – průběžné řízení parametrů mechanismu odstranění nevýznamné informace

- vzorky dekódovaného zvuku nejsou totožné se vstupem
- příklad
  vstup: náhodná čísla s jistým rozložením pravděpodobnosti
  výstup: jiná (!) náhodná čísla se stejným rozložením ppsti
  zvuk zní stejně, ačkoliv data jsou naprosto odlišná

ZTRÁTOVÁ KOMPRESE

METODY

- hrubší kvantizace, podvzorkování
  – využívá se v kombinaci s jinými metodami
- komprese a dekomprese dynamiky (kompandér)
- odhad následujícího vzorku,
  kódování rozdílu skutečného vzorku oproti odhadu
- odhad spektra, kódování rozdílu oproti skutečnému
  (čili technika „odhad následujícího vzorku" ve frekvenční oblasti)
- rozdělení zvuku na více frekvenčních pásem
  kódování každého z nich samostatně
- kódování transformovaného signálu
  (fourierovské transformace apod.)

DPCM

- differential PCM
- vstup $x[i]$
- predikce vzorku $p[i] = \sum_{k=1}^{M} a[k]\tilde{x}[i-k]$

  (kvalita predikce závisí na zvolených koeficientehc $a[]$)
- chyba predikce $e[i] = x[i] - p[i]$
- přenos *kvantizované* chyby $\tilde{e}[i]$ ⇒ redukce datového toku
- v dekodéru výpočet odhadu $p[i]$ jako v kodéru

  $p[i] = \sum_{k=1}^{M} a[k]\tilde{x}[i-k]$
- rekonstrukce $\tilde{x}[i] = p[i] + \tilde{e}[i]$
- dekodér rekonstruuje poškozené vzorky $\tilde{x}[i]$
  ⇒ kodér musí při výpočtu predikce $p[i]$ pracovat s tím, co bude mít dekodér k dispozici (tj. se vzorky $\tilde{x}[i]$)

DPCM

PŘÍKLAD

- prediktor $p[i] = \tilde{x}[i-1]$, $\tilde{x}[-1] = 0$
- kvantizace chyby „na násobky 5"

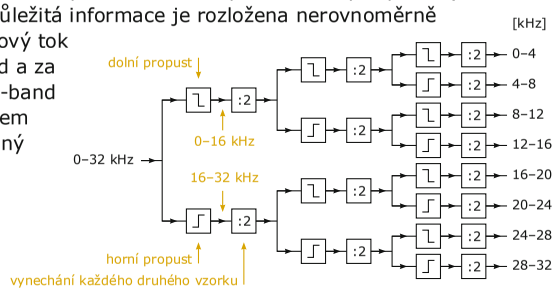
vstup $x[i]$	predikce $p[i]$	chyba $e[i]$	kvantizovaná chyba $\tilde{e}[i]$	rekonstrukce $\tilde{x}[i]$
0	0	0	0	0
3	0	3	5	5
3	5	−2	0	5
1	5	−4	−5	0

ADPCM

- adaptive differential PCM
- průběžná adaptace kvantizéru nebo prediktoru (koeficientů $a[k]$) podle vstupu
- dekodér i kodér se adaptují podle stejného algoritmu
  ⇒ není třeba přenášet parametry prediktoru/kvantizéru
- používáno samostatně i jako doplněk dalších metod (stejně jako DPCM)

SUB-BAND KÓDOVÁNÍ

- stejný princip jako LFE (low frequency effect channel)
- rozdělení zvuku na více frekvenčních pásem (optimální počet a velikost podle kritických pásem)
  – důležitá informace je rozložena nerovnoměrně
- datový tok před a za sub-band filtrem stejný



SUB-BAND KÓDOVÁNÍ

- zjištění maximální amplitudy A_k v každém pásmu k
  ⇒ dynamický rozsah pásma k: $D_k = 20 \log A_k / A_Q$ [dB]
  (A_Q = velikost kvantizačního kroku)
  ⇒ vzorky pásma k je možné ukládat D_k / 6,02 bity na vzorek
  ⇒ redukce datového toku
- snadné využití frekvenčního maskování
  – je-li v pásmu k vysoká úroveň a v pásmu $k + 1$ nízká, může být pásmo $k + 1$ zamaskováno
  – maskuje-li pásmo k pásmo $k + 1$ až do úrovně S dB, můžeme pásmo $k + 1$ kvantizovat méně než D_{k+1} / 6,02 bity na vzorek, pokud bude kvantizační šum menší než S
- vstup je vhodné rozdělit na malé bloky, aby byl v rámci bloku charakter pásem stabilní

SUB-BAND KÓDOVÁNÍ

IMPLEMENTACE

- rozdělení pomocí horní a dolní propusti (FIR) na dvě pásma
  – podvzorkování pásem na poloviční frekvenci
  – při rekonstrukci převzorkování pásem, posun do správného frekvenčního pásma a součet
- vícefázový filtr (polyphase filter)
  – pásmová propust a podvzorkování najednou
  – pro vzorky 0 až n se použije FIR filtr 1 (n-tap)
  ⇒ 1 vzorek pro 1. pásmo
  pro vzorky 1 až (n + 1) se použije FIR filtr 2 (n-tap)
  ⇒ 1 vzorek pro 2. pásmo
  atd.
- návrh dobrých filtrů netriviální

TRANSFORMACE SIGNÁLU

- rozdělení signálu na bloky
- transformace bloku pomocí DFT (diskrétní Fourierova t.), DCT (diskrétní kosinová t.), *MDCT (modifikovaná DCT)*
- využití standardních dekorelačních technik
  – odhad amplitudy jiné frekvence v bloku (intra kód.) (podobné jako odhad následujícího vzorku v DPCM)
  – odhad amplitudy v následujícím bloku (inter kód.)
- kvantizace dekorelovaného signálu

- bloky dlouhé – dobré vyjádření tónů
  bloky krátké – dobré vyjádření impulsních bloků
  ⇒ délka bloků se průběžně mění (typicky 4 typy: dlouhý, krátký, blok mezi krátkým a dlouhým, blok mezi dlouhým a krátkým)

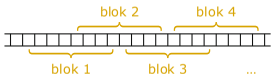
TRANSFORMACE SIGNÁLU

- typický artefakt: pre-echo
  - kvůli kvantizaci transformovaného signálu je chyba rozložena rovnoměrně v celém bloku
  - chvíli před a za hlasitou pasáží vzniká „echo"
  - artefakt za hlasitou pasáží typicky nevadí díky setrvačnosti bazilární membrány
  - velmi krátké pre-echo nevadí, viz časové maskování
    ⇒ v blízkosti prudké změny hlasitosti krátké bloky

vstupní signál *komprimovaný signál*

TRANSFORMACE SIGNÁLU

- potenciální artefakt: bloková struktura
  - původní signál spojitý
  - po rozdělení na bloky a nezávislé kompresi bloků může na hranici bloku vzniknout nespojitost
  ⇒ „lupnutí" s frekvencí odpovídající délce bloku
  - řešení: délka bloku $2N$ vzorků, rozteč bloků N vzorků (tj. bloky se z poloviny překrývají)
  aplikace „okna" $w[]$ na blok ⇒ na krajích útlum do ticha (tak, aby se po součtu bloků signál nezměnil)



rozdělení vstupu na bloky, N = 4

vzorek	útlum	vzorek	útlum
1	0,125	5	0,875
2	0,375	6	0,625
3	0,625	7	0,375
4	0,875	8	0,125

útlum vzorků bloku ($w[]$)

MDCT

- Modified Discrete Cosine Transform
- speciální typ DCT (podrobnosti k DCT později)
- konverze $2N$ hodnot signálu $x[]$ na N hodnot $X[]$:

$$X[k] = \sum_{j=0}^{2N-1} w[j]\,x[j]\cos\Big(\frac{\pi}{N}\,(j+\frac{1}{2}+\frac{N}{2})(k+\frac{1}{2})\Big)$$

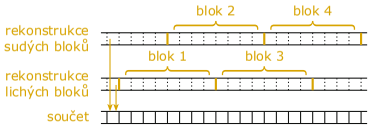
  $x[j]$ – signál, j = 0, …, $2N$ – 1
  $w[j]$ – okno
  $X[k]$ – frekvenční charakteristika, k = 0, …, N – 1
- bloky délky $2N$ se z poloviny překrývají
- konverze $2N$ vzorků $x[]$ na N koeficientů $X[]$ je nutně ztrátová

MDCT

- inverzní vztah (IMDCT):

$$x[j] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]\cos\Big(\frac{\pi}{N}\,(j+\frac{1}{2}+\frac{N}{2})(k+\frac{1}{2})\Big)$$

  pro j = 0, …, $2N$ – 1
- po IMDCT jednotlivých bloků, umístění na správnou pozici v čase a součtu zavlečená chyba zmizí (TDAC – Time Domain Alias Cancellation)



KOMPRESE HLASU

- omezený frekvenční a dynamický rozsah
  - 200 – 3200 Hz, vzorkování 8 kHz, 12 bitů/vzorek
- využití principu vzniku hlasu (proud vzduchu rozechvívá hlasivky)
  ⇒ vokodéry nevhodné pro obecný zvuk, cílem srozumitelnost
- rozdělení signálu na fragmenty (25 ms)
  - signál lze přibližně vyjádřit konvolucí: $x[] \doteq e[] \otimes f[]$
  - hledání excitace $e[]$ (např. položka ve slovníku excitací) a filtru $f[]$, které povedou na nejmenší chybu vyjádření
  - fragment zakódován koeficienty filtru (např. 10 čísel), parametry excitace (např. číslo položky ve slovníku), hlasitostí apod.
- datový tok několik kbit/s (např. 2,4 kbit/s)
- standardní kodéry: CELP (G.723.1, G.729 apod.)

MPEG AUDIO

MPEG-1

- vzorkovací frekvence 32, 44,1 a 48 kHz
- režimy mono, dual mono, stereo, joint stereo
  - dual mono: nezávislé mono kanály
  - stereo: kanály podobné, ale citlivé na chybu fáze
  - joint stereo: psychoakustický model stereofonie

MPEG-2

- doplňuje další vzorkovací frekvence
- kompatibilní a nekompatibilní (AAC) vícekanálový zvuk

MPEG-4

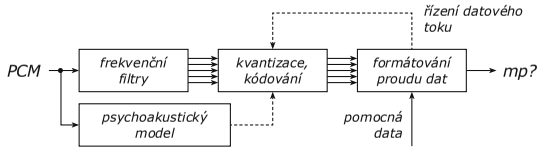
- doplňuje kompresi hlasu, syntézu zvuku
- HE-AAC schéma, bezztrátová komprese

MPEG-1 AUDIO

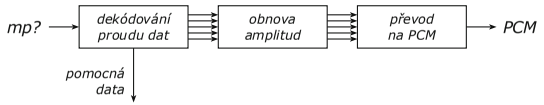
- MPEG-1 Audio – základní kompresní schéma
- Layer 1 (mp1)
  - nejjednodušší, datový tok > 128 kbit/s
  - sub-band kódování, využití frekvenčního maskování
- Layer 2 (mp2)
  - střední složitost, datový tok ≥ 128 kbit/s
  - rozšíření Layer 1, neobsahuje nové algoritmické prvky
  - VideoCD, DVD
- Layer 3 (mp3)
  - po sub-band kódovaní transformace signálu pomocí MDCT ⇒ lepší frekvenční rozlišení ⇒ psychoakustický model může lépe určovat, kterou informaci ztratit
  - nejsložitější, nejnižší datový tok (≥ 64 kbit/s)

MPEG-1 AUDIO

- základní schéma kodéru



- základní schéma dekodéru



PRINCIP KOMPRESE LAYER 1

- rozdělení PCM na 32 stejně velkých frekvenčních pásem (počet kompromis mezi složitostí a rozlišením pro psychoakustický model)
- rozdělení dat v pásmech na bloky po 12 vzorcích
- zesílení bloku scale factorem na max. amplitudu (příklad: rozsah ±8 vyžaduje 5 bitů ⇒ využití 53 % rozsahu)
- psychoakustický model: výpočet spektra z původního PCM (FFT z 512 vzorků) a určení maskování jednotlivých pásem, v úvahu se bere charakter zvuku (tón/šum/impuls)
- kvantizace pásem
- výstup délek slov (počet bitů na vzorek), scale factorů a kvantizovaných vzorků

PŘÍKLAD ALOKACE BITŮ

- vstup: pásma po zesílení scale factorem

pásmo	1	2	3	4	5	6	7	8	9	10 …
úroveň [dB]	0	8	12	10	6	2	10	60	35	20 …

8. pásmo 60 dB ⇒ podle psychoakustického modelu maskuje 12 dB v 7. pásmu, 15 dB v 9. pásmu

7. pásmo 10 dB (< 12 dB) – ignorovat
9. pásmo 35 dB (> 15 dB) – kódovat
   měli bychom kvantizovat na 35/6,02 = 6 bitů/vzorek, ale můžeme kódovat s 2bitovou kvantizační chybou (= 12 dB), tj. jen 4 bity/vzorek

VYLEPŠENÍ PRO LAYER 2

- charakter pásem se nemění příliš často
  ⇒ určení scale faktorů pro 3 bloky najednou
  ⇒ do výstupu 1 až 3 scale faktory na 3 bloky
  – snižuje datový tok na cca polovinu
- je třeba zaznamenat, kolik bitů/vzorek používá dané pásmo
  nízké frekvence – 15 různých délek slova
  střední frekvence – 7 různých délek slova
  vysoké frekvence – 3 různé délky slova
  jeden kód vyhrazen pro informaci „pásmo chybí"
- sdružování krátkých slov do jednoho kódu (viz mid-tread kvantizér, snímek 24)
- přesnější FFT (1024 vzorků) pro psychoakustický model

VYLEPŠENÍ PRO LAYER 3

- výstup sub-band filtrů transformován MDCT
  ⇒ přesnější rozlišení frekvencí
  ⇒ možnost větší kvantizace
- neuniformní kvantizace
- bloky pro MDCT krátké/dlouhé/přechodové
- Huffmanovo kódování koeficientů MDCT

MPEG-1 AUDIO

KOMPRESE STEREOFONNÍHO ZVUKU

- kódování dvou kanálů s přihlédnutím k jejich fázovým posunům
- vysoké frekvence de facto mono ⇒ redukce datového toku
- intensity (Layer 1/2/3)
  - pro frekvence > 2 kHz se stereo informace získává z obálky, ne z mikrodynamiky
  - ⇒u vyšších frekvencí se kóduje jen součet kanálů, scale factors jsou různé pro L/R kanál
- MS (Layer 3)
  - middle/side
  - tj. kódování součtového a rozdílového signálu
  - používá se automaticky tehdy, je-li rozdílový signál tichý

MPEG-2 AUDIO

- pro nižší datové toky vhodnější hrubší vzorkování
  - frekvence 0,5× MPEG-1 (16; 22,05; 24 kHz)
- kompatibilní rozšíření na více kanálů
  - typicky 5 kanálů
  - v základních MPEG-1 datech $\text{left}_{\text{TOTAL}}$, $\text{right}_{\text{TOTAL}}$ (viz maticové uložení vícestopého záznamu)
  - v pomocných datech 3 „čisté" kanály
    ⇒ L, R se dají zpětně vypočítat
- nekompatibilní schéma AAC (Advanced Audio Coding)
  - až 48 kanálů
  - kódování rozděleno na moduly
  - mnohem složitější než MPEG-1, principy zůstávají
  - navíc mj. predikce koeficientů v bloku a mezi bloky

DALŠÍ FORMÁTY

- HE-AAC (High-Efficiency AAC, MPEG-4 Part 3-Audio)
  - především pro nízké datové toky, založeno na AAC
  - dopočítávání harmonických frekvencí (neukládají se)
- Dolby Digital (AC-3, A-52)
  - de facto standard pro vícekanálový zvuk k filmu
  - sub-band kódování, MDCT
- snaha o sjednocení kodérů pro hlas a obecný zvuk
  - Opus – open source, nízká latence
  - Unified Speech and Audio Coding (USAC, MPEG-D Part 3), pro 12–64 kbit/s