

MULTIMEDIÁLNÍ A HYPERMEDIÁLNÍ SYSTÉMY

5)
Uložení a komprese zvuku

Petr Lobaz, 12.3.2013

ZMĚNY OPROTÍ VIDEO 2010/2011

- 18 – upřesněn odstup signál/šum
- 19 – nový snímek, výpočet odstupů signál/šum
- 27 – nový snímek, diskuse latence
- 37 – přidán rádek HE-AAC
- 46 – nový snímek, moderní trendy

MHS – Uložení a komprese zvuku

ULOŽENÍ ZVUKU

ANALOGOVÉ

- mechanický záznam – gramofon
- magnetický záznam – magnetofon, film
- optický záznam – zvuková stopa filmu
- rozhlas

DIGITÁLNÍ

- příznakové – „noty“
- vzorkovaný záznam

MHS – Uložení a komprese zvuku

2 / 47

ULOŽENÍ ZVUKU

RĚŽIMY

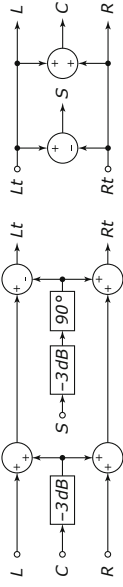
- mono
- stereo
- surround, stereo kompatibilní (matrix)
 - Dolby Surround (pasivní dekodér, špatná separace Center a Surround kanálů)
 - Dolby Pro Logic (aktivní dekodér)
 - Dolby Pro Logic II (aktivní dekodér; lepší separace než u Pro Logic pro zvuky mezi C a L/R)
 - využití pro 2.0 – 4.0, 5.1 → 7.1 atd.
- surround, diskrétní kanály

MHS – Uložení a komprese zvuku

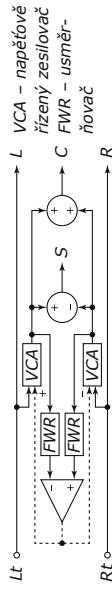
3 / 47

ULOŽENÍ ZVUKU

Dolby Surround – kodér



Dolby Pro Logic – dekodér



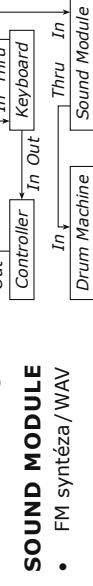
ULOŽENÍ ZVUKU

MIDI

- Musical Instrument Digital Interface
- MIDI Controller, MIDI Sound Module, MIDI Sequencer
- konektory MIDI In, MIDI Out, MIDI Thru

MIDI MESSAGES

- 31,25 kbit/s, 1-3 byte, v jednom toku 16 kanálů
- System Messages – pro všechny kanály
- Channel Messages



SOUND MODULE

- FM syntéza / WAV

ULOŽENÍ ZVUKU

Pulse Code Modulation

diskretizace spojitého signálu

– v čase: vzorkování – převod $x(t)$ na $x[i]$

– v hodnotách: kvantizace – převod na celá čísla

– PCM záznam vnáší do signálu chybu (šum)

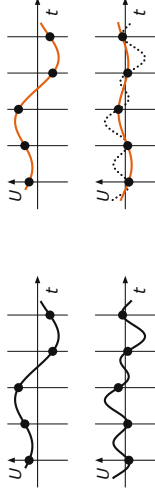
– snaha o minimalizaci



PCM

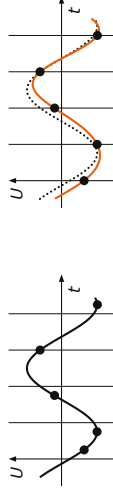
VZORKOVÁNÍ

- vzorkovací frekvence nejméně 2x max. frekvence, resp. šířka pásma
- vyšší frekvence v signálu se projevují jako alias
- rozsah lidského sluchu 20 – 20 000 Hz
- šířka pásma pro lidský hlas cca 4 kHz



PCM

- typické vzorkovací frekvence:
 - telefonie 8 kHz
 - rádio 32 kHz
 - CD 44,1 kHz
 - profesionální AV aplikace DV, DVD 48 kHz
 - v aplikacích zpracování zvuku vyšší (až 192 kHz)
- potíže s hodinami v A/D nebo D/A převodníku – jitter



PCM

VZORKOVÁNÍ SIGNÁLU

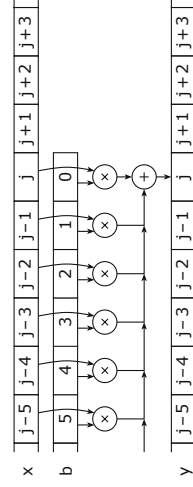
- analogová dolní propust + vzorkování
 - komplikovaný návrh analogového obvodu
 - náchylné k rušení
 - obtížná implementace lineární fázové charakteristiky
 - velmi kvalitní
- vzorkování vysokou frekvencí + digitální dolní propust + převzorkování na běžnou vzorkovací frekvenci

PCM

DIGITÁLNÍ FILTRY

- s konečnou odezvou na impuls (nerekurzivní) – FIR

$$y[j] = \sum_{k=0}^P b[k]x[j - k]$$



PCM

DIGITÁLNÍ FILTRY

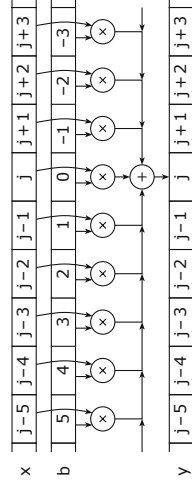
- příklad: dolní propust 0,1 vzorkovací frekvence (40 tap)
 - 0,0027 -0,0054 -0,0073 -0,0062 0,0000 0,0113
 - 0,0242 0,0315 0,0250 -0,0000 -0,0400 -0,0811
 - 0,1017 -0,0794 0,0000 0,1340 0,3011 0,4648
 - 0,5845 0,6283 0,5845 0,4648 0,3011 0,1340
 - 0,0000 -0,0794 -0,1017 -0,0811 -0,0400 -0,0000
 - 0,0250 0,0315 0,0242 0,0113 0,0000 -0,0062
 - 0,0073 -0,0054 -0,0027 -0,0000



PCM

- varianta FIR s pamětí

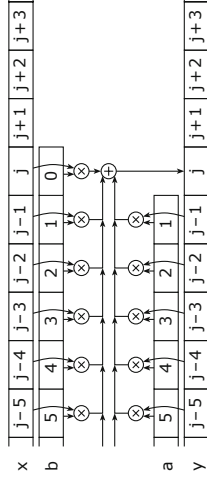
$$y[j] = \sum_{k=-P}^Q b[k]x[j-k]$$



PCM

- nekonečnou odezvou na impuls (rekurzivní) – IIR

$$y[j] = \sum_{k=0}^P b[k]x[j-k] + \sum_{k=1}^Q a[k]y[j-k]$$



PCM

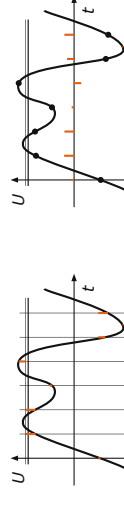
NÁVRH FIR FILTRU

- obecný tvar filtru – konvoluce
$$y[j] = \sum_{k=-\infty}^{\infty} b[k]x[j-k] = (b \otimes x)[j]$$
- pro konvoluci platí
$$b \otimes x = x \otimes b$$
$$\text{DFT}(b \otimes x) = \text{DFT}(b) \cdot \text{DFT}(x)$$
$$\text{DFT}(b \cdot x) = \text{DFT}(b) \otimes \text{DFT}(x)$$
- konstrukce frekvenční propusti (nástin jedné z metod):
 - návrh filtru B ve frekvenční oblasti
 - v časové oblasti $b = \text{DFT}^{-1}(B)$
 - problémy kvůli aproximaci konečným součtem

PCM

KVANTIZACE

- přidání „kvantizačního šumu“ k signálu
- jemnost podle míry akceptovatelného šumu
- 1 bit \approx 6 dB (tj. 16 bitů \approx 96 dB)
 - délka slova delší o 1 bit
 - \Rightarrow dvojnásobná max. amplituda
 - \Rightarrow dynamický rozsah $= 20 \log(2A_{\text{max}}/A_{\text{max}}) = +6,02$ dB



PCM

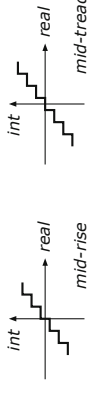
KVANTIZAČNÍ ŠUM

- velký počet kvantizačních úrovní
⇒ kvantizační šum „nezávislý“ na signálu
⇒ zvukový charakter jako bílý šum
- malý počet kvantizačních úrovní
⇒ kvantizační šum závislý na signálu
⇒ zvukový charakter není jako bílý šum
⇒ vhodnější je hovořit o kvantizačním zkreslení

PCM

KVANTIZÉRY

- mid-rise $\text{floor}(x) + 0.5$
 - 2^n úrovní se dělí symetricky na kladné a záporné
 - neobsahuje kód pro ticho – nevhodné pro zvuk
- mid-tread $\text{floor}(x + 0.5)$
 - lichý počet úrovní – nevhodné pro hrubou kvantizaci
 - sdružování několika vzorků do jednoho kódu
 - př.: 3 výstupní úrovně (0, 1, 2) ⇒ 2 bity na vzorek
lépe: $x[0] + 3x[1] + 3^2x[2]$ ⇒ 5 bitů na 3 vzorky



PCM

UNIFORMNÍ KVANTIZACE

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

PCM

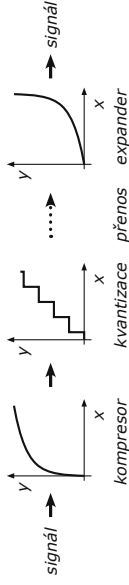
KOREKTNÍ VÝPOČET ODSTUPU SIGNÁL-ŠUM

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

PCM

NEUNIFORMNÍ KVANTIZACE

- velikost kvantizačního intervalu různá
- pro tiché zvuky jemná kvantizace, pro hlasité hrubá
- companding – komprese dynamického rozsahu, uniformní kvantizace, expanze
- telefonie, kódování pro malý počet (8) bitů



PCM

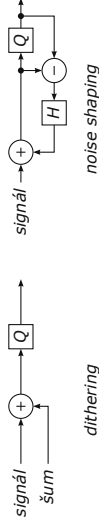
STANDARDNÍ NEUNIFORMNÍ KVANTIZACE

- μ -law (USA, Japonsko)
$$y = \frac{\ln(1+\mu x)}{\ln(1+\mu)}$$
$$\mu = 255$$
- A-law (mezinárodní telefonie)
– větší dynamický rozsah, větší zkreslení
$$y = \frac{1+\ln(Ax)}{1+\ln A} \quad \text{pro } 1/A \leq x \leq 1$$
$$y = \frac{Ax}{1+\ln A} \quad \text{pro } 0 \leq x \leq 1/A$$
$$A = 87,6$$

PCM

POTLAČENÍ KVANTIZAČNÍHO ŠUMU

- dithering – přidání šumu $\pm 0,5$ bitu
- objektivní zhoršení SNR cca o 3–5 dB
- subjektivní zlepšení SNR, linearita, konstantní šum
- kvalitní ditheringový šum respektuje citlivost ucha
- tvarování spektra šumu (noise shaping)
– přidávání šumu kompenzujícího kvantizační chybu
- kvalitní tvarovací filtr (H) respektuje citlivost ucha



COMPACT DISC

- ochranná vrstva, odrazivá vrstva, plastový nosič
- čtecí laser 780 nm (DVD 635/650 nm laser)

SPECIFIKACE

- Red Book – audio CD (CDDA)
44100 Hz, 16bit stereo, 72 minut, 1520 kbit/s
- Yellow Book – data CD
- Green Book – CD-i
- Orange Book – CD-MO, CD-R, CD-RW
- White Book – VideoCD (VCD)
- Blue Book – CD Extra (CDDA + data)

COMPACT DISC

- pits, lands; kódování NRZI
- minimálně 2, maximálně 10 nul za sebou
- byte – EFM (14/8) kódování
- mezi skupinami po 14 bitech 3 spojovací bity
- rámeček (frame) 33 byte – 24 data, 1 subkód, 8 parita
- před rámečkem 24 bitů synchronizace
- po stranách dat 3 spojovací bity ⇒ rámeček = 588 bitů
- F3 frame – ořezaný rámeček – 33 byte
- F2 frame – 32 byte dat
 - dekódování CIRC (Cross-Interleave Reed-Solomon)
 - korekce C1 a C2
- F1 frame – 24 byte uživ. dat (subkód není CIRC)

MHS – Uložení a komprese zvuku

24 / 47

COMPACT DISC

- sektor – 98 rámců = 98×24 B data + 98 B subkód = 2352 + 98 B = 2450 byte, frekvence 75 Hz
- descrambler
 - ⇒ 2352/4 = 588 stereo vzorků/sektor (= 13,3 ms)
- datová CD – sektor složitější
 - mode 1 – 2048 byte data
 - mode 2 – 2336 byte data (CD-ROM/XA)
- subkódy P–W
 - P – začátek stopy
 - Q – TOC, čísla stop, indexy, čas
 - R – S Red Book nedefinuje

MHS – Uložení a komprese zvuku

25 / 47

KOMPRES ZVUKU

- ztrátová – odstranění neslyšitelných částí zvuku
- bezztrátová – odstranění redundance v datech
- konstantní datový tok (CBR) – snadná manipulace
- proměnný datový tok (VBR) – úspěšnější
- průměrný datový tok (ABR) – kompromis CBR x VBR
- obecná – libovolný zvuk
- speciální – hlas, WAV pro MIDI Sound Module

MHS – Uložení a komprese zvuku

26 / 47

KOMPRES ZVUKU

- pro archivaci – co nejlepší kvalita
- pro komunikaci – co nejmenší zpoždění
 - např. kódér pracující po blocích 384 vzorků při vzorkování 48 kHz zanaší 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, ...

MHS – Uložení a komprese zvuku

27 / 47

BEZTRÁTOVÁ KOMPRESI

- příklad principu (Monkey Audio)
 - 1) odstranění redundance mezi kanály
 $X = (L + R) / 2, Y = L - R$
 - 2) predikce
 $P = (2X_{-1}) - X_{-2}$
rozdíl predikce a signálu adaptivní:
 $D = X - (P \cdot m / 1024)$
m se upraví do dalšího kroku
 - 3) Riceovo kódování dat
- FLAC – Free Lossless Audio Codec
- pro Unix, podobné kroky jako Monkey Audio
- Dolby TrueHD, DTS-HD Master Audio (Blu-ray disc)

ZTRÁTOVÁ KOMPRESI

PRINCIPI

- hrubší kvantizace, podvzorkování
 - využívá se v kombinaci s jinými metodami
- komprese a dekomprese dynamiky (kompanđer)
- odhad následujícího vzorku, kódování rozdílu skutečného vzorku oproti odhadu
- odhad spektra, kódování rozdílu oproti skutečnému rozdělení zvuku na více frekvenčních pásem
 - informace většinou pouze v několika z nich
- kódování transformovaného signálu (MDCT)

ZTRÁTOVÉ KOMPRESI

DPCM (DIFFERENTIAL PCM)

- vstup $x[j]$
- odhad vzorku $\hat{x}[j] = \sum_{k=1}^P a[k] \tilde{x}[j - k]$
- chyba predikce $\epsilon[j] = x[j] - \hat{x}[j]$
- přenos kvantizované chyby $\tilde{\epsilon}[j]$
- v dekóderu výpočet odhadu $\hat{x}[j]$ jako v kodéru, oprava podle přenesené chyby $\tilde{x}[j] = \hat{x}[j] + \tilde{\epsilon}[j]$

ADPCM (ADAPTIVE DPCM)

- adaptace kvantizéru nebo koeficientů $a[k]$ podle vstupu

ZTRÁTOVÁ KOMPRESI

SUB-BAND KÓDOVÁNÍ

- stejný princip jako LFE (low frequency effect channel)
- rozdělení zvuku na více frekvenčních pásem
 - informace většinou pouze v několika z nich
 - optimální dělení podle kritických pásem
- datový tok před a za sub-band filtrem stejný
- snadné využití frekvenčního maskování
- dynamický rozsah největší v malém počtu pásem, snadná analýza slyšitelného šumu
 - ⇒ vhodné využití (proměnné) hrubé kvantizace pásem

ZTRÁTOVÁ KOMPRESE

IMPLEMENTACE SUB-BAND KÓDOVÁNÍ

- rozdělení pomocí horní a dolní propusti na dvě pásma
 - podvzorkování pásma na poloviční frekvenci
 - při rekonstrukci převzorkování pásma, 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.

ZTRÁTOVÁ KOMPRESE

KÓDOVÁNÍ TRANSFORMOVANÉHO SIGNÁLU

- rozdělení signálu na bloky
- transformace bloku pomocí DFT, DCT, nejčastěji MDCT
- využití kompresních technik
 - odhad amplitudy jiné frekvence v bloku (intra kód.)
 - odhad amplitudy v následujícím bloku (inter kód.)
 - kvantizace amplitud
- kvůli impulsním zvukům jen na krátké bloky zvuku
 - pre-echo efekt



MDCT

- Modified Discrete Cosine Transform
- speciální typ DCT (typ DFT)
- konverze n hodnot na n/2 hodnot:

$$X[m] = \sum_{k=0}^{n-1} f[k]x[k] \cos\left(\frac{\pi}{2n}(2k+1+n/2)(2m+1)\right)$$

- x[k] – signál, k = 0 ... (n-1)
- X[m] – frekvenční charakteristika, m = 0 ... (n/2 - 1)
- f[k] – okno
- okna se z poloviny překrývají
- TDAC – Time Domain Alias Cancellation

KOMPRESE HLASU

- omezený frekvenční a dynamický rozsah
 - 200 – 3200 Hz, vzorkování 8 kHz, 12bitů
- cílem srozumitelnost hlasu
- využití principu vzniku hlasu
 - ⇒ vokodéry nevhodné pro obecný zvuk
- rozdělení signálu na fragmenty (25 ms)
 - charakterizace fragmentu – šum nebo tón
 - výpočet spektra X[k]
 - predikce X[k+1] na základě X[k], X[k - 1], ... pomocí FIR filtru – LPC (Linear Predictive Coding)
 - kódování koeficientů filtru
- datový tok několik kbit/s (např. 2,4 kbit/s)

VORBIS

- často ve spojení s transportním proudem Ogg
- neobsahuje detekci chyb, synchronizaci
- inicializace kodeku, dekodování může začít kdekoliv
- rámec se může oříznout a je stále dekódovatelný
- kódování po kanálech nebo párování kanálů
- využívá psychoakustiky (kritická pásma)
- rámce transformovány pomocí MDCT
- floor – hrubá reprezentace spektra v rámci
- reziduum – rozdíl skutečného a floor spektra, speciální kvantizace

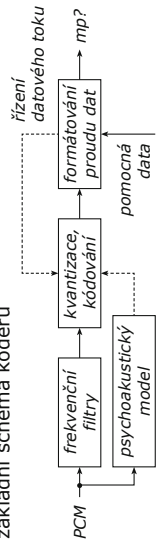
MPEG AUDIO

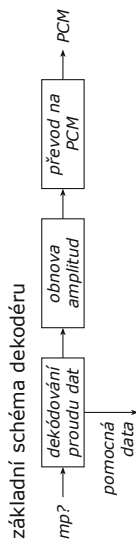
- **MPEG-1**
 - vzorkovací frekvence 32, 44,1 a 48 kHz
 - režimy mono, dual mono, stereo, joint stereo
 - dual mono – nezávislé mono kanály
 - stereo – kanály podobné, ale citlivé na chybu fáze
- **MPEG-2**
 - doplňuje další vzorkovací frekvence
 - kompatibilní a nekompatibilní (AAC) vícekanálový zvuk
- **MPEG-4**
 - doplňuje kompresi hlasu, syntézu zvuku
 - HE-AAC schéma

MPEG-1 AUDIO

- MPEG-1 Audio – základní kompresní schéma
- Layer 1 (mp1)
 - nejjednodušší, datový tok > 128 kbit/s
- Layer 2 (mp2)
 - střední složitost, datový tok ≥ 128 kbit/s
 - rozšíření Layer 1
 - VideoCD, DVD
- Layer 3 (mp3)
 - zavedení transformace signálu pomocí MDCT
 - nejsložitější, ≥ 64 kbit/s
 - datový tok 32 – 224 kbit/s
 - komprese 2,7 – 24 : 1

MPEG-1 AUDIO

- základní schéma kodéru

```
graph LR
    PCM[PCM] --> Filters[frekvenční filtry]
    Filters --> Model[psychoakustický model]
    Model --> Quant[kvantizace, kódování]
    Model --> Aux[pomocná data]
    Quant --> Rate[řízení datového toku]
    Rate --> MP3[formátování proudu dat mp?]
```
- základní schéma dekodéru

```
graph LR
    MP3[mp?] --> Rate[pomocná data]
    Rate --> Amp[obnova amplitud]
    Amp --> PCM[převod na PCM]
```

MPEG-1 AUDIO

PRINCIP KOMPRESI

- rozdělení PCM na 32 stejně velkých frekvenčních pásem – počet kompromis
- rozdělení toku na bloky ($12 \times 32 = 384$ vzorků), zesílení scale factorem na max. amplitudu (kompanér) (rozsah ± 8 vyžaduje 5 bitů \Rightarrow využít 53 % rozsahu)
- 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

MPEG-1 AUDIO

PŘÍKLAD ALOKACE BITŮ

- vstup: pásmo po průchodu kompanérem
- | | | | | | | | | | | | |
|-------------|---|---|----|----|---|---|----|----|----|----|-----|
| 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 \Rightarrow 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
můžeme kódovat s 2bitovou kvantizační chybou (= 12 dB)

MPEG-1 AUDIO

PRINCIP KOMPRESI

- rozdělení PCM na 32 stejně velkých frekvenčních pásem – počet kompromis
- rozdělení toku na bloky ($12 \times 32 = 384$ vzorků), zesílení scale factorem na max. amplitudu (kompanér) (rozsah ± 8 vyžaduje 5 bitů \Rightarrow využít 53 % rozsahu)
- 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

MPEG-1 AUDIO

VYLEPŠENÍ (LAYER 2)

- charakter pásem se nemění příliš často
 \Rightarrow určení scale faktorů pro 3 bloky najednou
 \Rightarrow 1–3 scale faktory na 3 bloky
- snižuje datový tok na cca polovinu
- nízké frekvence – 15 různých délek slova
- střední frekvence – 7 různých délek slova
- vysoké frekvence – 3 různé délky slova
- sdružování krátkých slov do jednoho kódu
- přesnější FFT (1024 vzorků) pro psychoakustický model

MPEG-1 AUDIO

VYLEPŠENÍ (LAYER 3)

- výstup sub-band filtrů transformován MDCT
 \Rightarrow přesnější rozlišení frekvencí
- \Rightarrow možnost větší kvantizace, neuniformní kvantizace
- délka okna MDCT se mění kvůli impulsním zvukům
- Huffmanovo kódování koeficientů MDCT

MPEG-1 AUDIO

KOMPRESSE STEREOFONNÍHO ZVUKU

- kódování dvou kanálů s přihlédnutím k jejich fázovým posunům
- vysoké frekvence de facto mono ⇒ joint stereo
- intensity (Layer 1/2/3)
 - pro frekvence > 2 kHz se stereo informace získává z obálky, ne z mikrodynamiky
- ⇒ u vyšších frekvencí se kanály sčítají, scale factors jsou různé pro L/R kanál
- MS (Layer 3)
 - middle/side
 - tj. kódování součtového a rozdílového signálu

MPEG-2 AUDIO

- pro nižší datové toky vhodnější hrubší vzorkování
 - frekvence 0,5x MPEG-1 (16; 22,05; 24 kHz)
- kompatibilní rozšíření na více kanálů
 - typicky 5 kanálů
 - v základních MPEG-1 datech matrix Lt, Rt
 - 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) 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

DALŠÍ FORMÁTY

- AC-3 (A-52, Dolby Digital)
 - 5.1, další kanály matrix
 - 6.1 – Dolby Digital EX
 - typické datové toky 448, 384 kbit/s
- DTS (Digital Theater System)
 - sub-band kódování, MDCT
 - 5.1, 6.1
 - až 2304 kbit/s, 24 bitů/vzorek
 - vzorkovací frekvence až 192 kHz
 - ADPCM
- Dolby TrueHD, DTS-HD bezztrátové kodéry
- RA (RealAudio), WMA (Windows Media Audio), ...