

# MULTIMEDIÁLNÍ A HYPERMEDIÁLNÍ SYSTÉMY

4)  
Zvuk, epizoda II

Petr Lobaz, 10.3.2004

---

# ULOŽENÍ ZVUKU

---

## **ANALOGOVÉ**

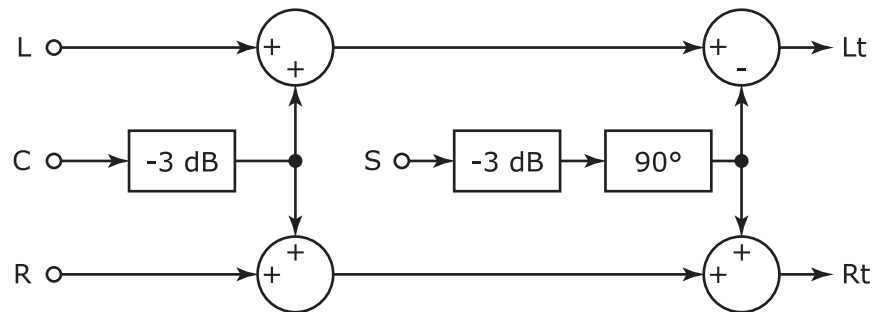
- gramofon, magnetický záznam, optický záznam
- rozhlas

## **REŽIMY**

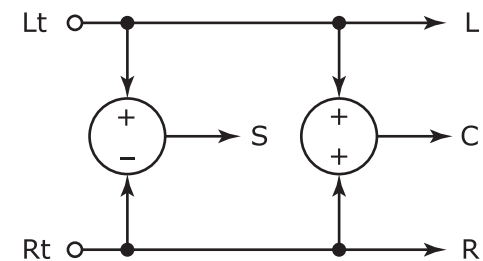
- mono
- stereo
- surround
  - Dolby Surround (pasivní dekodér)
  - Dolby Pro Logic (aktivní dekodér)
  - Dolby Pro Logic II (aktivní dekodér)

# ULOŽENÍ ZVUKU

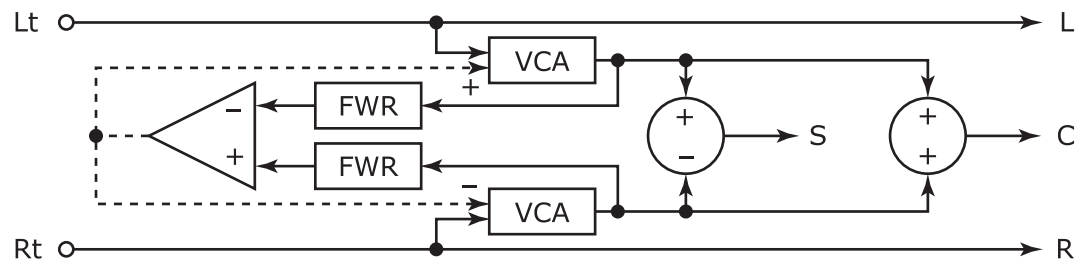
## Dolby Surround – kodér



## – dekodér



## Dolby Pro Logic – dekodér



---

# ULOŽENÍ ZVUKU

---

## **MIDI**

- Musical Instrument Digital Interface
- MIDI Controller, MIDI Sound Module, MIDI Sequencer
- 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 – Voice Messages, Mode Messages

## **SOUND MODULE**

- FM syntéza / WAV

---

# ULOŽENÍ ZVUKU

---

## **DIGITÁLNÍ**

- PCM (Pulse Code Modulation)
- vzorkovací frekvence
  - $2 \times$  max. frekvence, v praxi vyšší (až 192 kHz)
  - jitter
- rozlišení
  - 1 bit  $\approx$  6 dB (16bit  $\approx$  96 dB)
  - kvantizace
  - dithering – přidání šumu  $\pm 0,5$  bitu
  - zhoršení SNR o 4,77 dB

---

# PCM

---

- uniformní kvantizace – CD, DVD
  - kódování pro 12 a více bitů
- neuniformní kvantizace (companding)
  - kódování pro 8 bitů

- $\mu$ -law:  $y = \ln \frac{1 + \mu x}{1 + \mu}$        $\mu = 255$

- A-law:  $y = \frac{1 + \ln(Ax)}{1 + \ln A}$       pro  $1/A \leq x \leq 1$

- $y = \frac{Ax}{1 + \ln A}$       pro  $0 \leq x \leq 1/A$

$A = 87,6$

---

# 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  $\Rightarrow$  rámeček = 588 bitů
- F3 frame – ořezaný frame – 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)



---

# COMPACT DISC

---

- sektor – 98 rámců = 24 byte dat + 98 byte subkódu = 2352 byte
  - descrambler
- ⇒  $2352/4 = 588$  stereo vzorků/sektor
- 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

---

# COMPACT DISC

---

## **OCHRANA DAT**

- Macrovision SafeAudio – umělé chyby  
přehrávač interpoluje, digitální čtení problém
- SunComm MediaCloQ – disk 2 session, špatný TOC  
MediaMax CD3 – rezidentní sw :)
- Midbar Cactus Data Shield – nestandardní TOC  
vkládání špatných rámců
- Key2Audio / Sony DADC – ?  
CD-ROM mechanika nerozpozná disk

---

# KOMPRESSE ZVUKU

---

- ztrátová – odstranění zbytečností
- bezztrátová – odstranění redundance
  
- obecná – libovolný zvuk
- speciální – hlas, WAV pro MIDI Sound Module

---

# BEZZTRÁTOVÁ KOMPRESSE

---

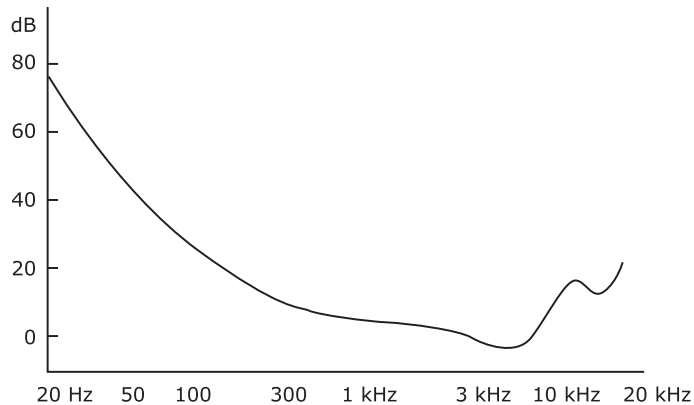
- Monkey Audio
  - 1) odstranění redundance mezi kanály
$$X = (L + R) / 2, Y = L - R$$
  - 2) predikce
$$P = (2X_{-1}) - X_{-2}$$
rozdíl predikce a signálu adaptivní:
$$D = X - (P \cdot m / 1024)$$
m se upraví do dalšího kroku
  - 3) rice kódování dat
- FLAC – Free Lossless Audio Codec  
pro Unix, podobné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

---

# PSYCHOAKUSTIKA

---

- frekvenční rozsah 20 Hz – 20 kHz
- maximální citlivost 1 – 4 kHz  
(hlas typicky 500 – 2000 Hz (samohlásky, souhlásky))
- dynamický rozsah 96 dB
- citlivost – Fletcher-Munsonova křivka



---

# PSYCHOAKUSTIKA

---

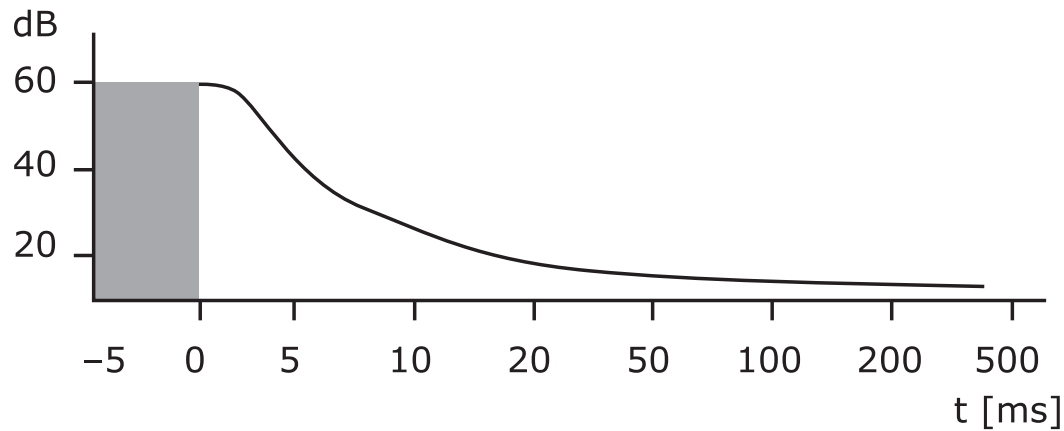
- frekvenční maskování
  - maskování slabšího, frekvenčně podobného tónu
- kritické pásmo
  - frekvenční rozsah se stejným rozlišením frekvencí
  - bark (Barkhausen) - jednotka pro frekvenci psychologicky stejného rozlišení
  - 1 bark = šířka kritického pásma
  - pro  $f < 500$  Hz       $1 \text{ bark} \approx f/100$
  - pro  $f > 500$  Hz       $1 \text{ bark} \approx 9 + 4 \log(f/1000)$

---

# PSYCHOAKUSTIKA

---

- časové (temporal) maskování
  - po ukončení hlasitého tónu chvíli trvá, než se stane slabší tón slyšitelný



---

# ZTRÁTOVÉ KOMPRESSE

---

- DPCM – differential PCM
- ADPCM – adaptive DPCM
  - rozlišení (bitová hloubka) rozdílu se mění
- LPC (Linear Predictive Coding)
  - pro hlas
  - rozdělení signálu na fragmenty (25 ms)
  - charakterizace fragmentu – šum nebo tón
  - určení dalších informací – výška, koeficienty filtru
  - 200–3200 Hz, vzorkování 8 kHz, 12bit, 2,4 kbit/s
- CELP (Code Excited Linear Predictive)
  - používá codebook, odvozeno z LPC, 4,8 kbit/s
  - použití v MPEG-4



---

# VORBIS

---

- často ve spojení s transportním proudem Ogg
  - neobsahuje detekci chyb, synchronizaci
  - inicializace kodeku, dekódování 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

---

# MDCT

---

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

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

$x_t(k)$  – signál,  $k = 0 \dots (n-1)$

$X_t(k)$  – frekvenční charakteristika,  $k = 0 \dots (n/2 - 1)$

$f(k)$  – okno

- okna se překrývají

---

# MPEG

---

## **MPEG-1**

- vzorkovací frekvence 32, 44,1 a 48 kHz
- režimy mono, dual mono, stereo, joint stereo

## **MPEG-2**

- doplňuje další frekvence
- vícekanálový zvuk

## **MPEG-2.5**

- není standard (Fraunhofer Institute)
- další vzorkovací frekvence

## **MPEG-4**

- doplňuje kompresi hlasu, syntézu zvuku

---

# MPEG - 1

---

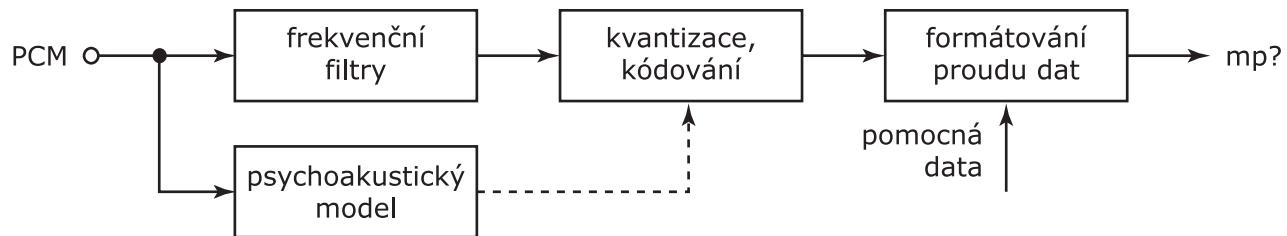
- základní kompresní schéma
- Layer 1 (mp1)
  - nejjednodušší, datový tok  $> 128$  kbit/s
- Layer 2 (mp2)
  - střední složitost, datový tok  $\geq 128$  kbit/s
  - rozšíření Layer 1
  - VideoCD
- Layer 3 (mp3)
  - nejsložitější,  $\geq 64$  kbit/s
  - datový tok 32–224 kbit/s
  - komprese 2,7–24 : 1

---

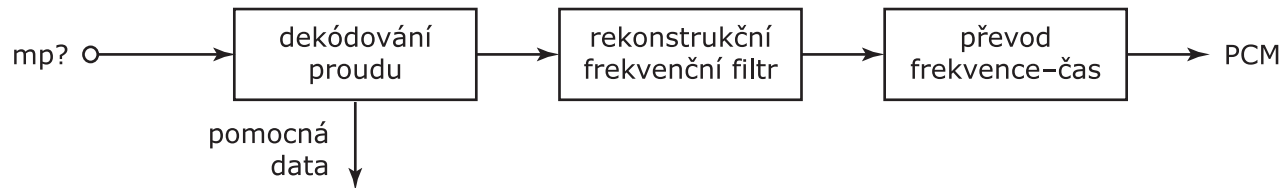
# MPEG-1

---

- základní schéma kodéru



- základní schéma dekodéru



---

# MPEG - 1

---

## FREKVENČNÍ FILTRY

- 32 filtrů o stejné šířce pásma
- polyphase – reagují na široké pásmo, překrývají se
- není bezztrátový (kolísání o -0,07 dB)
- reagují na každých 32 vzorků
- celková délka okna 512 vzorků

$$s_t[i] = \sum_{n=0}^{511} x[t-n] H_i[n]$$
$$H_i[n] = \pm C[n] \cdot \cos \frac{(2i+i)(n-16)\pi}{64}$$

$C[n]$  definuje standard – koeficienty okna pro analýzu

---

# MPEG - 1

---

## **PSYCHOAKUSTICKÝ MODEL**

- dva standardní modely
- frekvenční analýza v okně 512 nebo 1024 vzorků
- konverze výsledků frekvenční analýzy do pojmů kritických pásem
- rozdělení zvuku na tón a šum  
model 1 – striktní dělení tón/šum  
model 2 – tonality index
- určení hodnoty frekvenčního (šumového) maskování
- nastavení prahů slyšitelnosti
- výpočet SNR

---

# MPEG - 1

---

## **KVANTIZACE A KÓDOVÁNÍ**

- Layer 1
  - rámeček 384 vzorků (12×32 polyphase)
  - každá skupina 12 vzorků má alokovaný určitý počet bitů (0-15)
  - pro každou skupinu scale factor
  - využití frekvenčního maskování
- Layer 2
  - vylepšení Layer 1
  - rámeček 1152 vzorků (3×12×32 polyphase)
  - pro každou trojici skupin 0–3× scale factor
  - využití frekvenčního a částečně časového maskování



---

# MPEG - 1

---

## **KVANTIZACE A KÓDOVÁNÍ**

- Layer 3
  - výstup z polyphase filtrů prochází MDCT
  - do kvantizace 576 linek ( $18 \times 32$ )
  - bloky z MDCT krátké nebo dlouhé (6/18 vzorků)  
délky se mohou střídát
  - Huffmanovo kódování výstupu
  - bit reservoir – variable bitrate

---

# MPEG - 1

---

## ALOKACE BITŮ

- příklad

---

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

---

## **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

---

- vzorkovací frekvence MPEG-1 + 0,5 MPEG-1  
32; 44,1; 48; 16; 22,05; 24; kHz
- kompatibilní rozšíření na více kanálů
  - v základních MPEG-1 datech downmix
  - surround v pomocných datech
- nekompatibilní kompresní schéma AAC
  - Advanced Audio Coding
  - podobné MPEG-1

---

# 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
  - podobné MPEG-1
- DTS (Digital Theater System)
  - 5.1, 6.1
  - až 2304 kbit/s, 24 bitů/vzorek
  - vzorkovací frekvence až 192 kHz
  - podobné MPEG-1, používá ADPCM
- RA (RealAudio)
- WMA (Windows Media Audio)