

# Kompresce zvuku

Ing. Jan Přichystal, Ph.D.

PEF MZLU v Brně

14. března 2011

# Úvod

Komprimace umožňuje efektivní digitální reprezentaci zdrojového signálu jako je text, obraz, zvuk nebo video, použitím redukovaného počtu prvků digitální informace, než má originál. Musí však umožňovat, pokud má být efektivní, reprodukci komprimované informace v požadované kvalitě.

Pro komprimaci dat v současných informačních a komunikačních systémech hovoří tři základní důvody:

- ▶ rozsáhlé paměťové nároky multimediálních dat,
- ▶ relativně pomalá paměťová zařízení, která neumožňují přehrávání multimediálních dat v reálném čase,
- ▶ často nedostatečná šířka pásma současných sítí pro přenos zvuku v reálném čase.

# PCM, DPCM a ADPCM

PCM je nejjednodušší způsob kódování dat. Nepoužívá kompresi. Navzorkovaná data jsou ukládána jako posloupnost celočíselných dat.

Diferenciální PCM kóduje hodnoty vzorků jako rozdíl oproti předchozí hodnotě. Dosahuje až 20 % úspory.

Adaptivní DPCM navíc dovolují proměnnou velikost vzorku.

# Komprimace audio signálu

Kvalitní digitální stereozáznam používá vzorkovací frekvenci 44,1 kHz, což odpovídá datovému toku 176 400 bajtů za sekundu.

Komprimace zvuku se uplatňuje v oblasti přenosu lidského hlasu, kde lze použít nižších vzorkovacích frekvencí a predikce chování akustického signálu. Druhou oblastí je hudební signál, kde se naopak využívají vysoké vzorkovací frekvence a uplatňuje se psychoakustický model sluchu.

Myšlenka vytvoření ztrátového kompresního algoritmu pro zvuková data se zrodila v roce 1987 ve Fraunhofer Institut Integrierte Schaltungen (IIS).

# MPEG formáty

MPEG (Moving Picture Experts Group) pracuje pod vedením International Standards Organization (ISO) a International Elektro-Technical Commission (IEC) a zabývá se kódováním videa a audia. MPEG standardy se dělí:

- ▶ MPEG 1 kódování videa a audia pro uložení na digitálních médiích, datový tok do 1,5 Mbit/s
- ▶ MPEG 2 kódování při nižších datových tocích, poloviční vzorkovací frekvence
- ▶ MPEG 3 původně plánováno pro HDTV, později spojeno s MPEG 2
- ▶ MPEG 4 kódování audiovizuálních objektů (např. pro media objects, syntézu zvuku)

# Vrstvy MPEG 1

Každý standard obsahuje několik částí, které popisují kódování audia, videa, synchronizačních dat a formáty uložení kódovaných dat. MPEG standard obsahuje několik vrstev Layer-I–III, které popisují kódovací schémata. Od Layer-I do Layer-III roste komplexnost a efektivita komprese zvuků, ale klesá rychlost kódování a dekodování. Zvuková schémata se dělí:

- ▶ Layer-I nejjednodušší schéma, původně je určeno pro Digital Compact Cassette (DCC)
- ▶ Layer-II kompromis mezi kvalitou, rychlostí a kompresním poměrem
- ▶ Layer-III od začátku vytvářeno pro nízké bitové proudy, vylepšené kódování

# Layer-I – kompresní algoritmus

- ▶ Kódovaný zvuk se rozdělí na 32 frekvenčních pásem. Každé toto pásmo se kóduje zvlášť. Jednotlivá pásma jsou široká 625 Hz. Pásma jsou na nízkých frekvencích užší a na vyšších širší (kvůli citlivosti ucha na různé frekvence).
- ▶ Je použita ztrátová komprese, tj. kvantizace.
- ▶ Kvantizační koeficienty se vypočítávají dynamicky podle dosažené akustické hladiny zvuku tak, aby šum vzniklý použitím kvantizace byl pod rozlišovací schopností ucha. Využívá se zde psychoakustického modelu.
- ▶ Vypočtený signál se kóduje entropickou metodou, např. Huffmanovým kódováním.

# Parametry dosažené komprese

- ▶ Záznam v kvalitě CD je zaznamenán při bitrate 256–384 kbps. Stupeň komprese je tedy až 1:4.
- ▶ Dekodér velmi jednoduchý, kodér přibližně 1,5–3 krát složitější.
- ▶ Nejmenší teoreticky dosažitelný delay (zpoždění reprodukováného signálu za originálním; je způsobeno zpracováním po blocích) je 19 ms. Obvyklá hodnota odpovídá zpracování 4 frames, což při 48 kHz dává 100 ms.
- ▶ Velikost rámce (nejmenší komprimované jednotky zvuku) je 384 bitů.



# Layer-II – kompresní algoritmus

- ▶ Komprimuje se najednou větší množství dat, délka rámce je 1152 bitů.
- ▶ Algoritmy vycházejí z Layer-I
- ▶ V každém pásmu se kvantizovaná data kódují 0–15 bity

# Parametry dosažené komprese

- ▶ Záznam odpovídající kvalitě CD je možno získat při 192–256 kbps. Dosažený stupeň komprese je tedy 1:6–8.
- ▶ Dekodér asi o 25 % složitější než u Layer-I, složitost kodéru roste.
- ▶ Nejmenší delay je 35 ms.
- ▶ Velikost rámce je 1152 bitů.

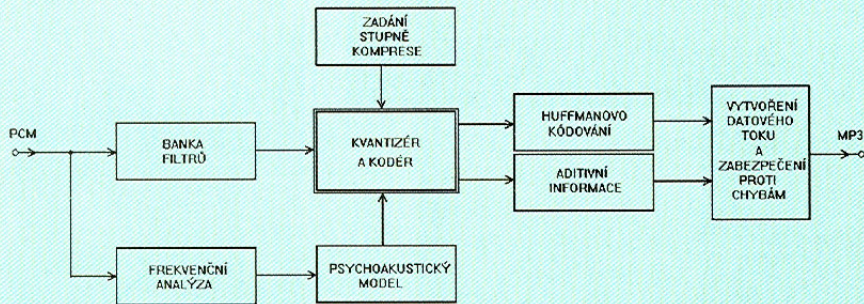
# Použití

- ▶ CD-i, Video-CD, DVD, kabelová a satelitní televize.
- ▶ Pro profesionální studiové použití je doporučen právě tento formát. Ačkoliv nemá tak vysoké kompresní poměry jako Layer-III, vyniká nejlepší kvalitou zvuku a složitost jeho kompresních algoritmů je optimální. Studiové kvalitě odpovídá 384 kbps, pro Joint Stereo stačí pouze 192 kbps.

# Layer-III (MP3)

1. Kódovaný zvuk se rozdělí na 32 frekvenčních pásem. Každé toto pásmo se kóduje zvlášť. Jednotlivá pásma jsou široká 625 Hz. Pásma jsou na nízkých frekvencích užší a na vyšších širší (kvůli citlivosti ucha na různé frekvence).
2. Je použita ztrátová komprese – kvantizace. Kvantizační koeficienty se vypočítávají dynamicky podle dosažené akustické hladiny zvuku tak, aby šum vzniklý použitím kvantizace byl pod rozlišovací schopností ucha. Využívá se psychoakustického modelu.
3. Kromě klasického psychoakustického modelu je využit temporal masking, který počítá s tím, že hlasitý tón zastíní blízké frekvence nejenom v době, kdy sám zní, ale i chvíli potom (asi 100ms). Zajímavé je, že díky jisté malé prodlevě při zpracování zvuku lidským uchem platí maskování frekvencí i 2–5 ms předem.
4. Vypočtený signál se kóduje Huffmanovým kódováním.
5. Výsledná kvalita je určena datovým tokem.

# Formát MP3



Blokové schéma kodéru MP3

# Psychoakustický model

Umožňuje odstranění dat lidským sluchem nepostižitelných a tedy v signálu redundantních, bez znatelné újmy na kvalitě reprodukovaného signálu.

1. omezení frekvenčního rozsahu – využívá omezené citlivosti lidského ucha. Přenosového pásma se omezuje na šířku 20 Hz až 20 kHz.
2. maskování frekvencí – využívá nelinearity citlivosti lidského sluchu. V přítomnosti silnějšího signálu nedokážeme vnímat slabší signál, který zaniká a není třeba jej tedy uvažovat.
3. časové maskování – využívá setrvačnosti lidského sluchu. I po zániku silnějšího signálu chvíli trvá, než začneme vnímat současně působící slabší signál.

# Parametry dosažené komprese

- ▶ Záznam odpovídající kvalitě CD je možno získat při 112–128 kbps. Dosažený stupeň komprese je tedy 1:10–12.
- ▶ Kodér i dekodér velmi náročné na výpočetní výkon.
- ▶ Nejmenší delay je 59 ms.
- ▶ Velikost rámce je 1152.

# Použití

Tento formát je doporučen „do terénu“ – na reportáže, komentáře, atd. neboť exceluje na nejnižších bitrates. Pro vyšší kvalitu se hodí lépe Layer-II.

Nicméně v současné době se nejčastěji používá právě Layer-III, zvláště pak pro grabování CD.



# Ogg

Datový formát Ogg propagovaný nadací Xiph.org byl vytvořen jako výchozí materiál, který si klade za cíl vyvinout komponenty pro kódování a dekódování multimediálního obsahu, přičemž tyto komponenty budou svobodně dostupné a svobodně reimplementovatelné v softwaru.

Ogg je jen kontejnerový formát. Vlastní audio nebo video zakódované kodekem bude uloženo uvnitř Ogg kontejneru. Ogg kontejnery mohou obsahovat streamy zakódované s vícenásobnými kodeky, například, audio nebo video soubor může obsahovat data zakódovaná audio kodekem i video kodekem. Formát Ogg může uložit audio a video v různých formátech.

- ▶ Audio kodeky
  - ▶ ztrátová komprese
    - ▶ Speex: hlasová data v nízkém bitratu (8–32 kbps/kanál)
    - ▶ Vorbis: hlavní zvuková data ve střední až vysoké úrovni proměnného bitratu (16–500 kbps/kanál)
  - ▶ bezztrátová komprese
    - ▶ FLAC: pro archivní a vysoce přesná zvuková data
- ▶ Textový kodek
  - ▶ Writ: textový kodek k vkládání titulků nebo popisků
  - ▶ CMML: textový/aplikační kodek pro časová metadata, popisování, a formátování

- ▶ Video kodeky
  - ▶ Theora: konkuruje formátu MPEG-4 a kodekům RealVideo, Windows Media Video.
  - ▶ Tarkin: experimentální kodek využívající Diskrétní vlnkovou transformaci v třech dimenzích (šířka, výška a čas).
  - ▶ Dirac: experimentální kodek vyvinutý BBC jako základ nového kodeku pro přenos videa přes internet.
  - ▶ OggUVS: kodek pro ukládání nekomprimovaného videa.

# Ogg Speex

Speex je svobodný zvukový kodek, který slouží ke kompresi lidské řeči. Jeho specializace mu umožňuje dosahovat velmi dobrých kompresních poměrů při zachování srozumitelnosti. Není primárně určen na kompresi telefonních hovorů, ale na VoIP.

Speex byl navržen pro použití se třemi vzorkovacími frekvencemi 8 kHz, 16 kHz, a 32 kHz. Kvalita se nastavuje jako parametr od 0 do 10. CBR se nastavuje celočíselným parametrem, VBR reálným parametrem.

# Ogg Speex

Speex má několik zajímavých postupů, které jsou speciálně určeny pro kompresi lidské řeči:

- ▶ Voice Activity Detection (VAD) – detekce aktivity hlasu. Tato technologie má za úkol rozpoznat, zda řečník právě mluví, nebo dělá přestávku (mezi slovy, větami), nebo prostě jen čeká. V tom případě je produkováno jen velmi malé množství bitů, které stačí na generování šumu.
- ▶ Discontinuous Transmission (DTX) – rozšíření předchozí technologie, která detekuje delší odmlčení. V tom případě úplně zastavuje přenos dat.
- ▶ Perceptual enhancement – Snaží se zeslabit šum vytvořený kódováním/dekódováním. Objektivně (tedy měřením) tato technologie vzdaluje výstupní zvuk od originálu, ale subjektivně je zvuk lepší.

# Ogg Vorbis

Vývoj kodeku se datuje do roku 1998, kdy Fraunhoferův institut ohlásil záměr zpoplatnit formát MP3.

Umožňuje používat CBR, ABR a VBR. Cílem je získat zvuk lepší kvality při menší velikosti než MP3.

Jde o dopředný, adaptivní monolitický kodek využívající DCT a psychoakustický model.

Dekódování je výpočetně méně náročné než u MP3, je však náročnější paměťově.

# Ogg Vorbis

kvalita	bitrate
-q-2	not available
-q-1	45 kb/s
-q0	64 kb/s
-q1	80 kb/s
-q2	96 kb/s
-q3	112 kb/s
-q4	128 kb/s
-q5	160 kb/s
-q6	192 kb/s
-q7	224 kb/s
-q8	256 kb/s
-q9	320 kb/s
-q10	500 kb/s

# Ogg Vorbis

Poslechové testy různých kodeků pro stejné datové toky (bitrates):

- ▶ Nízký bitrate ( $< 64$  kbps): lepší kvalita než WMA
- ▶ Střední až nízký bitrate ( $< 128$  kbps  $> 64$  kbps): lepší kvalita než MP3 a WMA
- ▶ Střední bitrate (128kbps): stejný jako MP3, WMA Pro a QuickTime AAC
- ▶ Vysoké bitrates ( $> 128$  kbps): znatelné rozdíly rozpozná jen školený posluchač, lepší než MP3



# FLAC – Free Lossless Audio Codec

- ▶ V současnosti nejrozšířenější bezztrátový audiokodek
- ▶ Navržen především pro kompresi zvukových dat (vzorky mají vysoký stupeň korelace)

# FLAC – vlastnosti

- ▶ Rychlost – symetrický z hlediska kódování a dekódování
- ▶ Podpora hardware – podporován na přenosných přehrávačích, domácích audiosystémech i PC
- ▶ Přizpůsobitelnost – metadata mohou být modifikována přehrávacím zařízením nezávisle na zvukové stopě
- ▶ Prohledávatelnost – rychlé přeskakování ve skladbě
- ▶ Streamování – každý rámeček je možné rozkódovat a přehrát nezávisle na ostatních
- ▶ Archivace – bezztrátová komprese umožňuje dekódování do stavu před kompresí
- ▶ Archivace CD – podporuje cue sheet, CD lze převést do jednoho souboru a skladby přehrávat jako na CD
- ▶ Odolný proti chybám – chybné rámce lze přeskaovat
- ▶ Neobsahuje DRM

# Postup – rozdělení do bloků

- ▶ vstupní signál rozdělen na bloky,
- ▶ délka bloku ovlivňuje výsledný kompresní poměr,
- ▶ minimální velikost bloku je 16 vzorků
- ▶ maximální velikost bloku je 65535 vzorků
- ▶ výchozí velikost bloku je 4096
- ▶ do budoucna se počítá s proměnnou délkou bloku
- ▶ stereo lze kódovat pomocí rozdílů

# Postup – aproximace

- ▶ signál je aproximován modelem zvuku
- ▶ po odečtení vzniká error signál, který vyžaduje méně bitů na kódování
- ▶ přenáší se i funkční parametry
- ▶ k aproximaci se používá polynomická a lineární predikce
- ▶ polynomická je rychlejší, ale ne tak přesná
- ▶ čím vyšší stupeň lineární predikce tím je proces pomalejší, ale přesnější

# Postup – kódování zbytku

- ▶ error signál je kódován Riceovým kódem (Huffman)
- ▶ každý výsledný rámeček je uvozen hlavičkou se synchronizačním kódem, vzorkovací frekvencí a počtem bitů na vzorek
- ▶ rámeček obsahuje i CRC kód pro kontrolu dat

# Výsledný formát

Ve zkomprimovaném souboru jsou odděleny části zvuku a metadata.

- ▶ 4 B textový řetězec „fLaC“ sloužící jako poznávací značka
- ▶ blok metadat STREAMINFO obsahující informace o vzorkování, počtu kanálů, min. a max. velikost rámců a MD5 otisk původních dat
- ▶ další metadata bloků (nepovinná část) s různými informacemi – např. PICTURE pro obal desky
- ▶ samotný komprimovaný audio záznam rozdělený do rámců

# Windows Media Audio (WMA)

- ▶ součástí programového vybavení Windows Media firmy Microsoft.
- ▶ zpočátku vyvíjen jako konkurence MP3.
- ▶ nabízí zvukovou kvalitu srovnatelnou s audio CD při datovém toku 64kbps, což představuje přibližně dvakrát účinnější kompresi oproti MP3.
- ▶ Ve spojení s ASF – Advanced Streaming Format může být WMA streamován.
- ▶ WMA formát podporuje celá řada přenosných hardwarových i softwarových přehrávačů.

# AAC – Advanced Audio Coding

- ▶ jedná se o nástupce MP3, není však zpětně kompatibilní
- ▶ lepší komprese při zachování srovnatelné kvality s MP3 (výsledkem jsou menší komprimované audio soubory),
- ▶ větší počet audio kanálů
- ▶ širší nabídku dostupných vzorkovacích frekvencí pro zajištění lepšího výsledného zvuku,
- ▶ součástí standardu MPEG-4.
- ▶ využíván firmou Apple, např. v přehrávačích iPod



# ATRAC3 – Adaptive Transform Acoustic Coding 3

- ▶ majetkem firmy Sony a představuje vylepšenou verzi formátu ATRAC, který je používán v přehrávačích MiniDisc.
- ▶ nabízí stejnou kompresi i zvukovou kvalitu jako MP3.

# ePAC – Enhanced Perceptual Audio Coder

- ▶ autorem firma Lucent Technologies
- ▶ při kompresním poměru 11:1 k nerozeznání od originálního CD (téměř na stejnou úroveň jako formát MP3)
- ▶ používá se pro přenosné MP3 přehrávače.

# Real Audio

- ▶ kompresním formátem pro streaming audio a video dat od RealNetworks
- ▶ po zvukové stránce při vyšším datovém toku (bitrate) srovnatelný s MP3

# Apple Lossless

Apple Lossless, také známý jako Apple Lossless Encoder (ALE), nebo Apple Lossless Audio Codec (ALAC) je audio kodek, vyvinutý společností Apple Computer pro bezztrátovou kompresi hudby.

Používá kontejner MP4. Není založen na AAC, ale používá lineární predikci podobně jako ostatní bezztrátové audio kodeky.

Podle specifikace firmy Apple audio soubory komprimované pomocí ALAC potřebují přibližně polovinu velikosti nekomprimovaných dat. Testy však prokázaly, že soubory průměrně dosahují 60 % původní velikosti, podobně jako ostatní bezztrátové formáty.

Na rozdíl od ostatních formátů, Apple Lossless není náročný na dekompresi a je tak použitelný i v přístrojích s omezeným zdrojem energie jako například iPod.

# Zvukové artefakty

Aplikováním kompresních algoritmů se v někdy objevují charakteristické zvukové projevy znehodnocující výsledek.

- ▶ pre-echo – slabá kopie zvuku, který teprve zazní
- ▶ reverb – kovově znějící ozvěna
- ▶ trylkování
- ▶ zhrubnutí
- ▶ ...

# Posuzování kvality kodeků

- ▶ posluchač se snaží identifikovat neznámý vzorek  $X$ , jestli jde o originál nebo komprimovaný zvuk,
- ▶ výsledek testu musí být statisticky významný,
- ▶ důležitý je předpoklad, že posluchač není ovlivněn svými očekávaními a hodnocení není výsledkem náhody,
- ▶ v případě správného odhadnutí posluchačem, může tento vzorku přiřadit subjektivní hodnotu kvality.

# Děkuji za pozornost

## Dotazy?