

Multimediální systémy

Jan Outrata



KATEDRA INFORMATIKY
UNIVERZITA PALACKÉHO V OLMOUCI

přednášky



Komprese audia

- Audio CD nekomprimované, ale šíření audia po sítích s omezenou rychlostí (streaming, telefonní sítě) a médiích s omezenou kapacitou (přehrávače), narůstající požadavek kvality ⇒ **potřeba komprese**
- např. 44.1 kHz, 16 bps, stereo (CD) ~ 1.3 Mbps \rightarrow 160 – 230 kbps (MP3), 192 kHz, 16 bps, 5.1 surround ~ 18 Mbps \rightarrow 256 – 2048 kbps (AC-3)
- např. standardy (ISO) MPEG, (ITU) G.7xx, Dolby AC-x aj.
- audio: **širokopásmové** = hudba a **úzkopásmové** = mluvené slovo (řeč)
- zvuk reprezentován jako 1D signál, ale **méně redundance** a lidský zvukový systém citlivější na šum a zkreslení než u obrazu a videa, vnímání závisí na intenzitě a frekvenci zvuku
- různé **odlišné metody**: predikce (DPCM) a bezztrátové (odstranění statistické redundance), využití psychoakustických modelů, modelů zdroje zvuku, sémantické modely (strukturované audio)

- = **relativní** jednotka pro měření amplitudy (intenzity) signálu
- = $10 \log \frac{I}{I_0}$, kde I (I_0) je intenzita (referenčního) signálu v W/cm^2
- referenční signál: experimentálně zjištěný sotva slyšitelný zvuk s $I_0 = 10^{-6} W/cm^2$
- např. konverzace 40 dB, doprava 70–90 dB, práh bolesti 120–130 dB

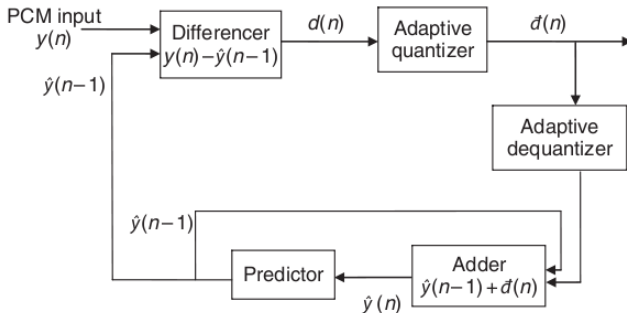
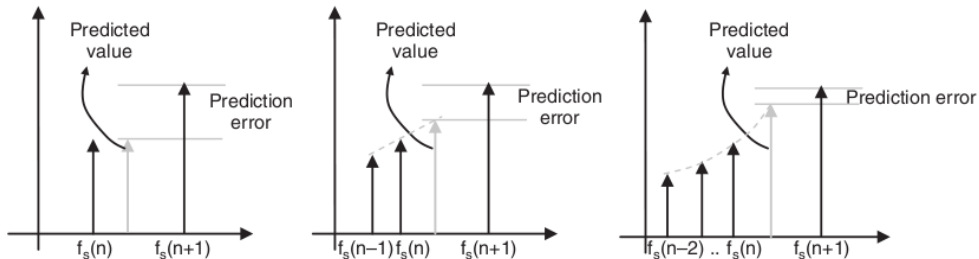
(A)DPCM

- predikovaná hodnota = předchozí hodnota nebo aproximace z **více předchozích hodnot** – s (predikčními) koeficienty, menší chyba

Obrázek: Obr. DPCM z více předchozích hodnot

- ADPCM (Adaptive DPCM): adaptace počtu bitů pro kvantizaci chyby (rozdílu)
 - řídicí bity navíc zvyšují entropii → dva módy: pro nízké a vysoké frekvence

Obrázek: Obr. ADPCM



Delta modulace

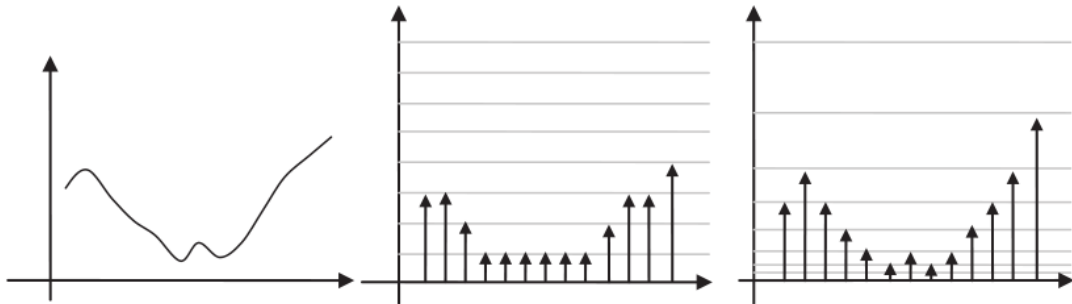
- jednoduchá, jako DPCM, ale chyba (rozdíl) kódována jedním bitem – 0 při zvýšení nebo 1 při snížení hodnoty o **konstantu (delta)**
- delta je pevná, určená na základě vzorkovací frekvence
- pro pomalu se měnící signál (řeč, do 8 kHz), jinak velké chyby, kontrolovaná bit rate

A/ μ -law

- **logaritmický kvantizační rozsah (companding)** – empiricky častěji nízké intenzity signálu, např. u řeči \rightarrow rozdíl kvantizačních úrovní přímo úměrný intenzitě, častější nízká intenzita = menší kvantizační chyba

Obrázek: Obr. Uniformní a logaritmická kvantizace

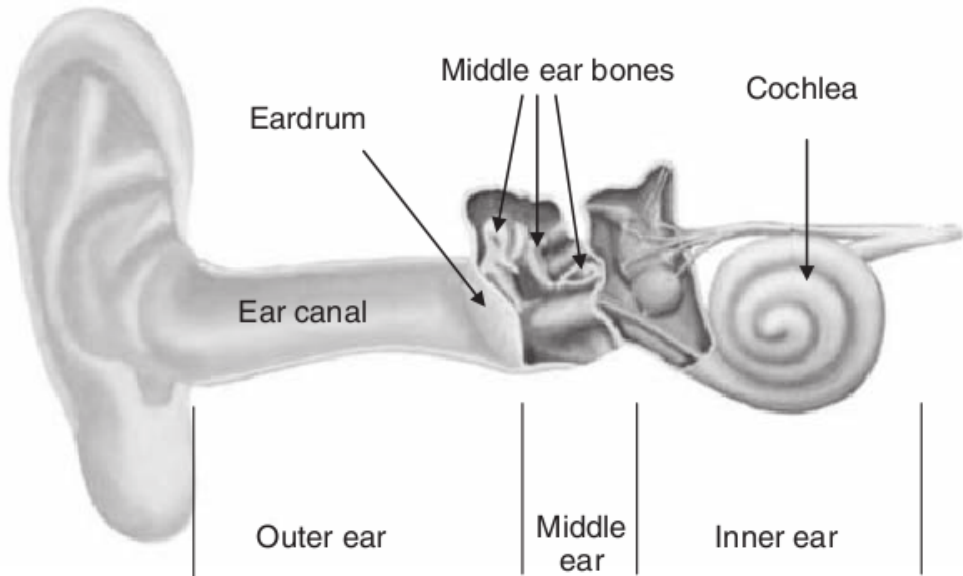
= ITU standardy pro **kompresi řeči**: až 60 dB, ucho citlivější na nízké intenzity, 13/14 bps uniformně \rightarrow 8 bps logaritmicky



- = **psychologie vnímání zvuku** lidským zvukovým systémem → využití omezení pro filtraci neslyšitelných zvuků a kvantizaci
- ucho: **zvuk** jako tlakové vibrace vzduchu je přes bubínek převeden na pohyb kůstek a přes membránu na tlakové vibrace tekutiny hlemýždě snímané jako změny frekvence vlásenkovými receptory – nelineární (tvar hlemýždě, distribuce receptorů aj.), uvažované (**zjednodušeně!**) jako **1D signál**

Obrázek: Obr. Struktura ucha

- **frekvenční omezení:** slyšitelný rozsah asi 20 Hz až 20 kHz → filtrace, dynamický rozsah asi 120 dB
- **časová omezení:** různé zvuky audio signálu v čase jsou vnímané odděleně, pokud jsou aspoň asi 30 ms od sebe, jinak je vnímání jako u frekvencí

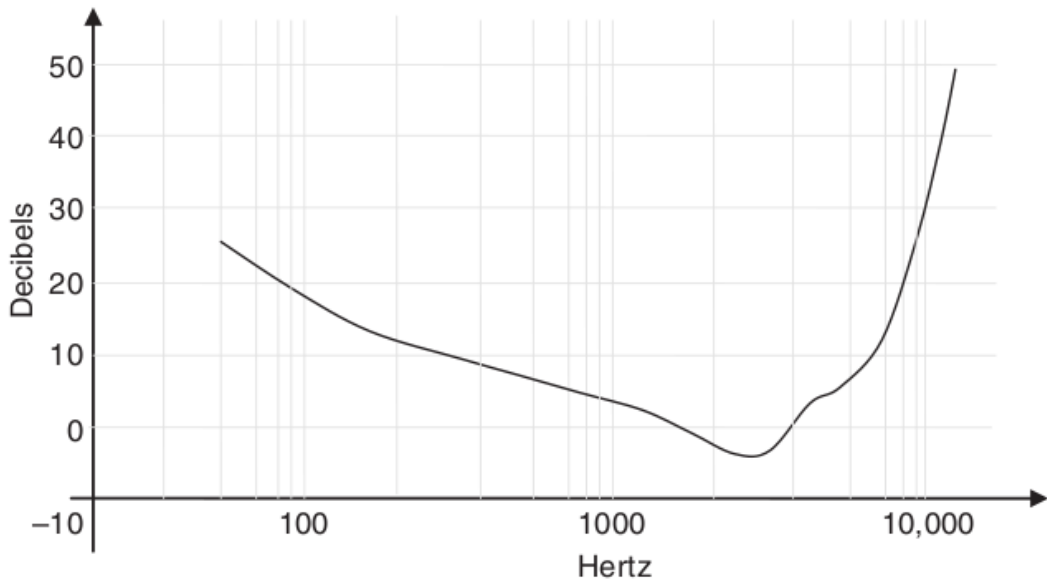


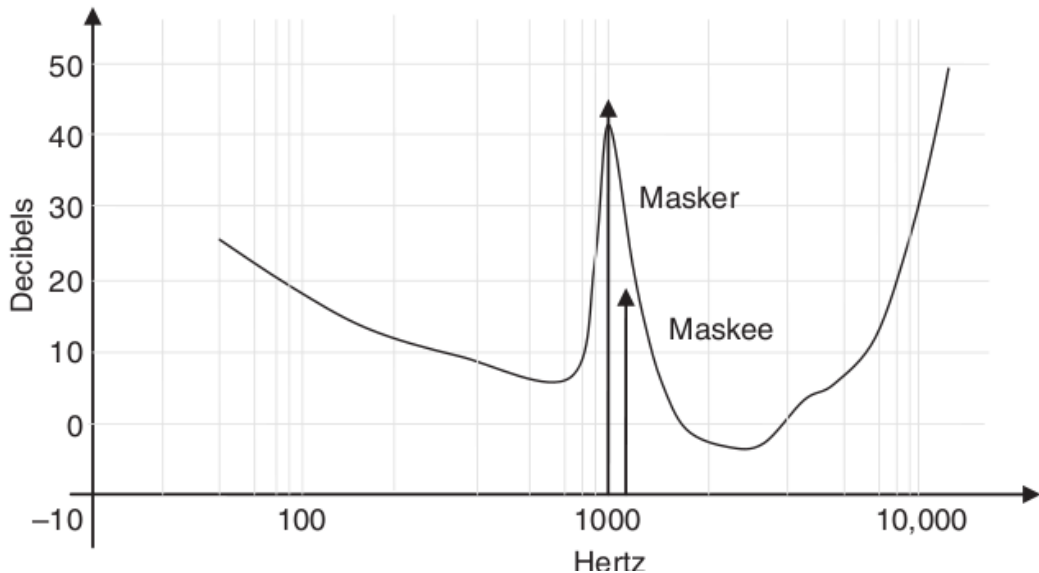
- každá frekvence (tón) má samostatně jiný **práh slyšitelnosti**

Obrázek: Obr. Práh slyšitelnosti frekvencí

- = frekvence přítomná (s nějakou intenzitou) v signálu maskuje jiné frekvence, které tak nejsou vnímány = **frekvenční maskování**
- = přítomnost frekvence (nad prahem slyšitelnosti) zvyšuje práh slyšitelnosti pro jiné frekvence – změny zjištěny experimentálně
- maskované frekvence = sousední menší (do menší vzdálenosti) a větší (do větší vzdálenosti) s menší intenzitou, rozsah maskovaných frekvencí a jejich intenzit závisí (přímo, nelineárně) na frekvenci i intenzitě maskující frekvence

Obrázek: Obr. Změna prahu slyšitelnosti v přítomnosti frekvence





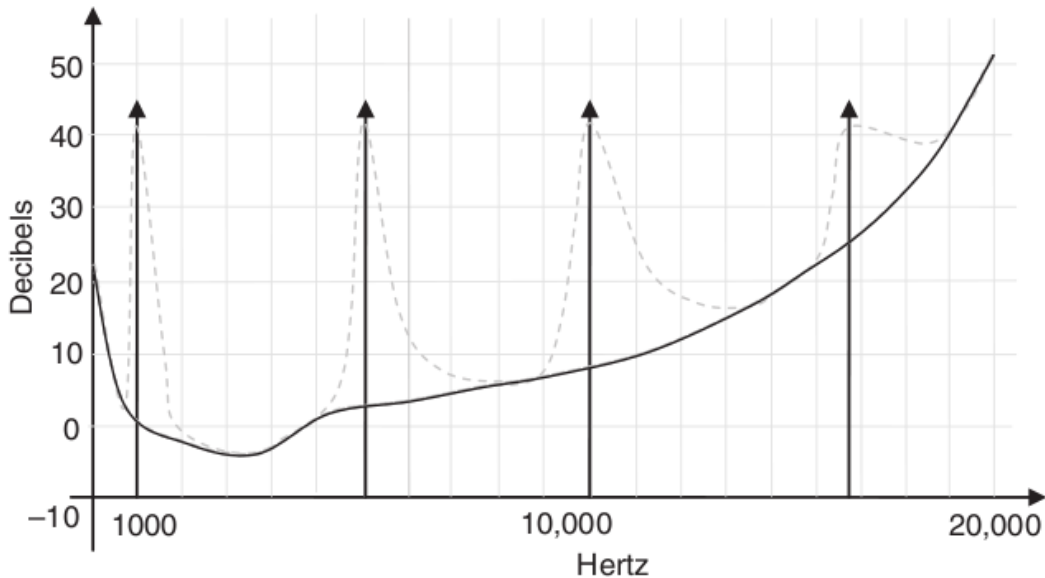
- rozsah maskovaných frekvencí je pro každou maskující frekvenci jiný

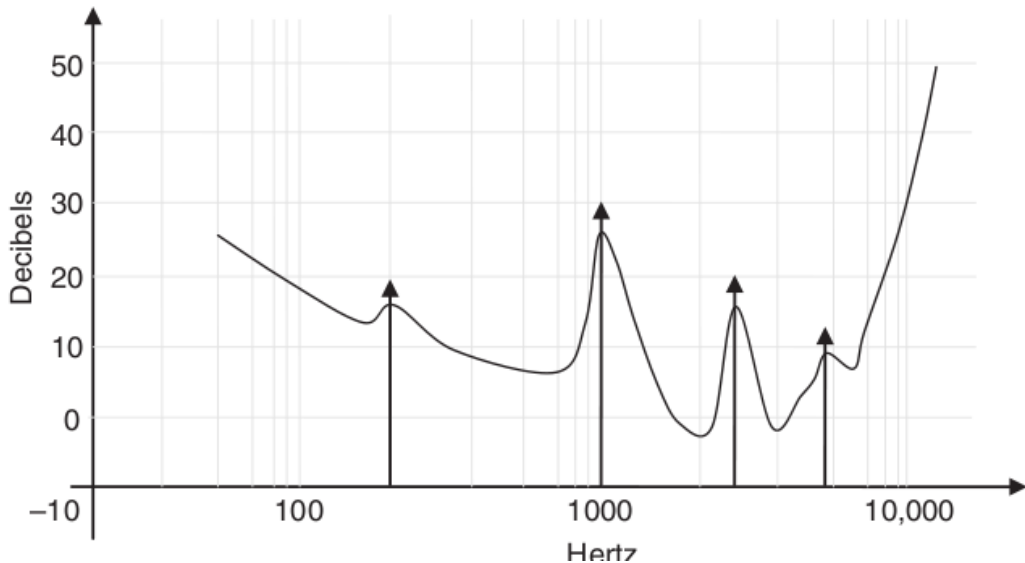
Obrázek: Obr. Změna prahu slyšitelnosti pro různé přítomné frekvence

- rozsah maskovaných frekvencí pro danou maskující frekvenci je jiný i v případě přítomnosti dalších maskujících frekvencí a práh slyšitelnosti frekvencí se celkově zvyšuje

Obrázek: Obr. Změna prahu slyšitelnosti v přítomnosti více frekvencí

- ⇒ lidský zvukový systém vytváří a filtruje **pásma** (nerozliší frekvence v jednom pásmu) – šířky pásem 100 Hz pro frekvence pod 500 Hz, pro zvyšující se frekvence se lineárně zvyšuje až do 6 kHz, zjištěno experimentálně



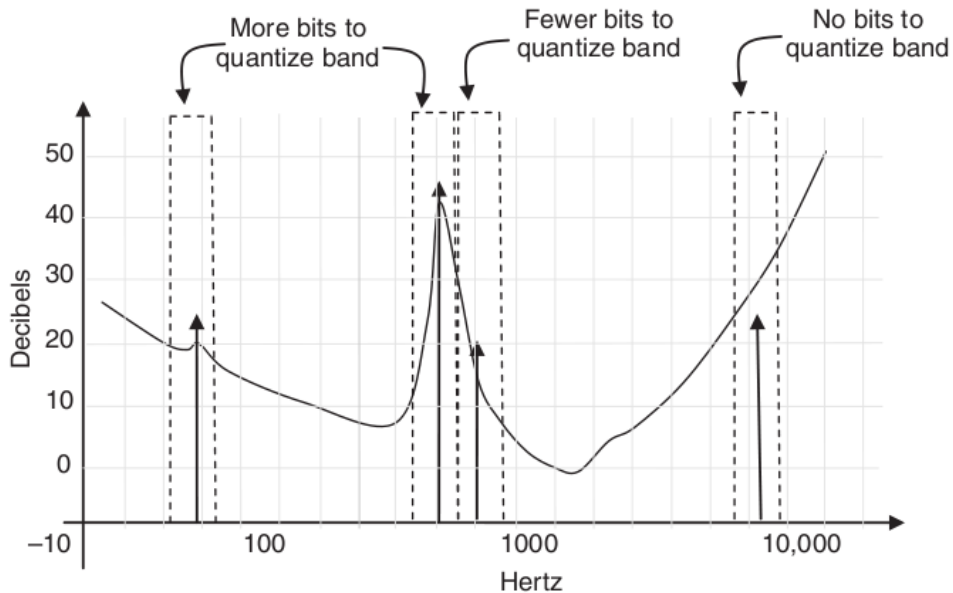




- **temporální maskování** = maskující frekvence maskuje maskované frekvence po nějakou dobu i po svém vymizení, s postupným zmenšováním rozsahu maskování, doba závisí na rozdílu a hodnotě intenzit (přímo) a vzdálenosti frekvencí (nepřímo)
- **psychoakustický model** = model maskovaných frekvencí v závislosti na přítomnosti (a vymizení) maskujících frekvencí

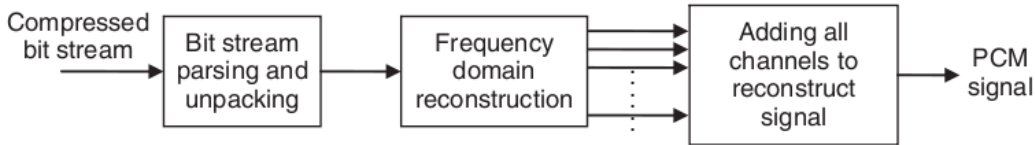
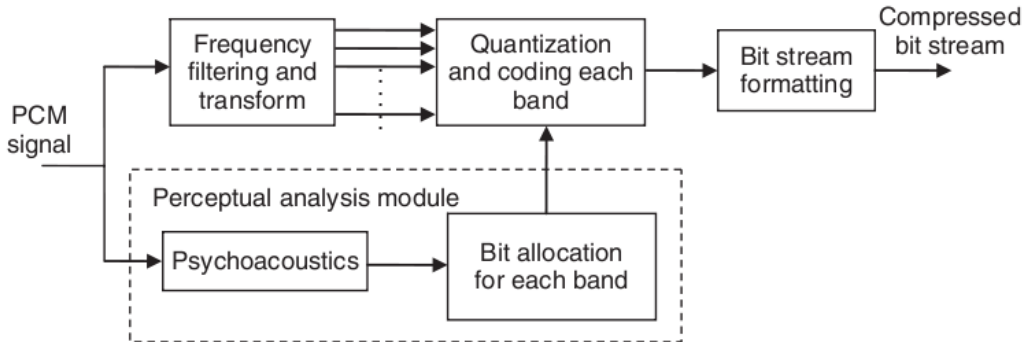
- 1 **rozdělení signálu do oken** s typicky od 512 do 4096 vzorky, nezávislé zpracování oken
 - 2 **transformace z časové do frekvenční domény**
 - 3 **filtrace** – odstranění neslyšitelných frekvencí a rozdělení slyšitelných **do pásem**
 - 4 **paralelně kódování signálu a úpravy psychoakustického modelu** podle signálu
- kódování signálu = **kvantizace** frekvenčních koeficientů pásem – narozdíl od statických tabulek u obrazu (dominantní = nižší frekvence) dynamická na základě **dynamického** psychoakustického modelu (dominantní = maskující frekvence, měnící se v čase)

Obrázek: Obr. Distribuce bitů v pásmech



- úpravy psychoakustického modelu = **úpravy křivky prahu slyšitelnosti** frekvencí pásem **a kvantizačních tabulek**, na základě analýzy signálu

Obrázek: Obr. Komprese a dekomprese





- komprese = analýza a **modelování (parametrizace) zdroje zvuku ze signálu**, parametry např. frekvence (výška), délka, amplituda (intenzita), periodičita, barva aj.
- dekomprese = **syntetizace zvuků** na základě modelu
 - použitelné pouze pro **známé zdroje zvuku**, např. hlas, siréna
 - hlas \approx fyzikální proces průchodu zvuku z plic hrdlem přes hlasivky a ústy – model rezonující trubky/píšťaly, parametry **rezonance** a residuální frekvence a amplituda

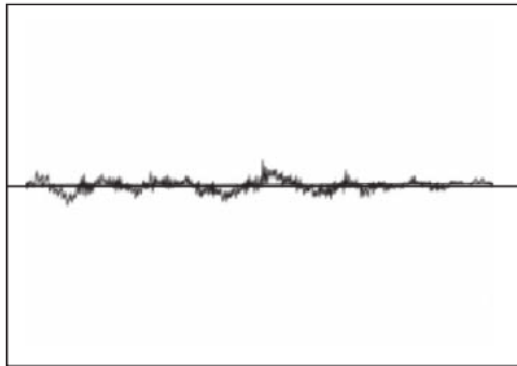
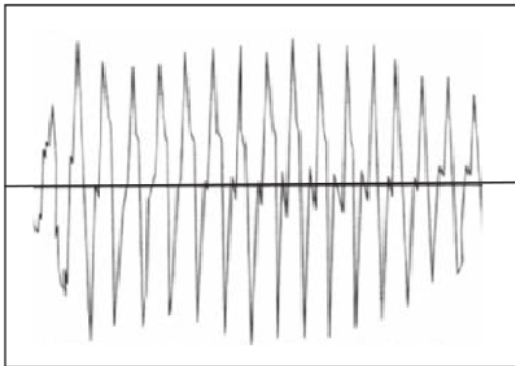
- = metoda **analýzy řeči**, použití pro imitaci (syntetizace)
- rozdělení signálu do bloků
- analýza na **hlas = samohlásky** = nízké frekvence a vyšší amplituda, a **šum = souhlásky** = vyšší frekvence a nízká amplituda

Obrázek: Obr. Samohlásky a souhlásky v řeči

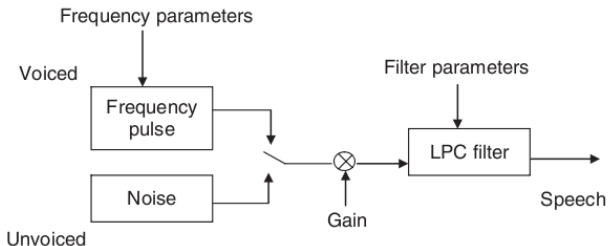
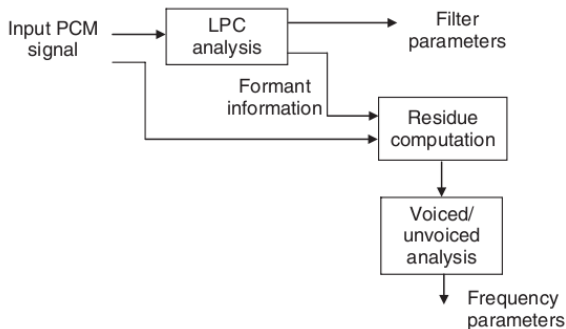
- kódování parametrů: typ, frekvence nebo šum, amplituda, filtrační pro reziduum

Obrázek: Obr. Komprese a dekomprese

Linear Predictive Coding (LPC)

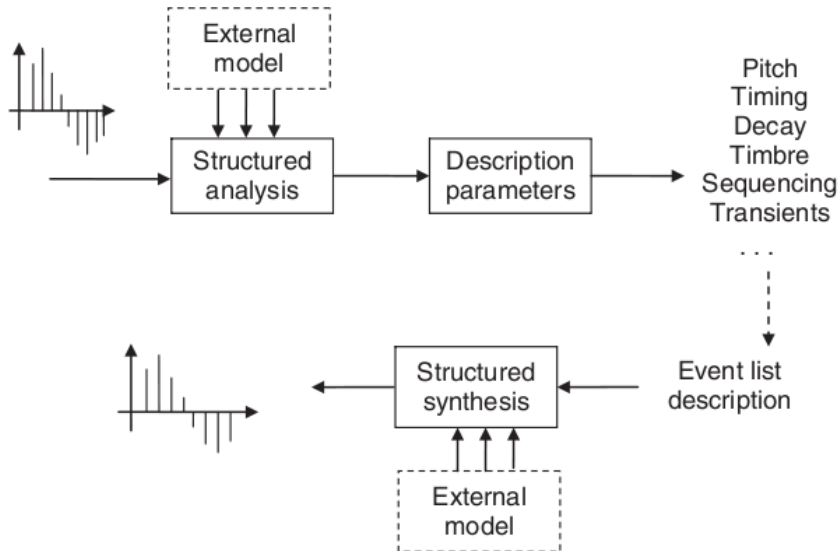


Linear Predictive Coding (LPC)



- komprese = analýza a sémantický popis strukturovaného audia jako posloupnosti **kombinací zvuků generovaných nástroji**, tj. hudby, hudební parametry frekvence (noty), časování (délka), amplituda (síla), barva (nástroj), mixážní efekty atd.
- kódování posloupnosti **hudebních parametrů**
- dekomprese = **syntetizace** zvuků podle parametrů interpretovaných/generovaných **externím modelem (nástroj, emulace oscilátory)**, např. piano

Obrázek: Obr. Komprese a dekomprese





- **aditivní syntéza** – skládání základních frekvencí, s amplitudou proměnných v čase, pro (kvazi)periodické zvuky (neimpulsivních) nástrojů
- **subtraktivní syntéza** – filtrace bohatě harmonického zvuku z oscilátorů, jako modelování hlasu (samohlásek)
- **frekvenční modulace** – modulace nosné (vyšší) frekvence pomocí modulační frekvence zvuku
- **fyzické modelování** – emulace fyzické struktury zdroje a zvuku, parametry popisují materiál a konstrukci zdroje a jeho použití, např. buben a bouchnutí



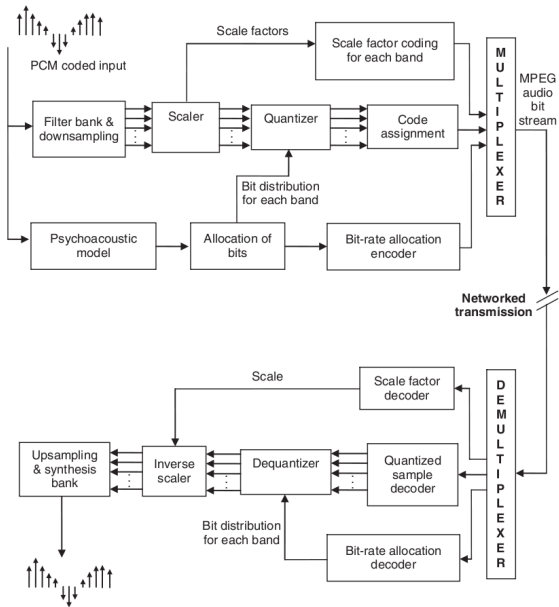
- úspornost – velice nízké bit rate
- snadná tvorba a modifikace hudby – formální vysokoúrovňový popis, flexibilita, podobné **komponování hudby**
- interaktivita – např. **automatické generování hudby** „na pozadí“ při akci člověka, např. hraní
- sémantický popis – analýza, **kategorizace** apod. **hudby**

- některé standardizované – ISO MPEG-x, ITU G.7xx, Dolby AC-x, MIDI
- spolupráce s výzkumnými institucemi a průmyslem

MPEG-1

- 3 zpětně kompatibilní **vrstvy**: **Layer I (MP1), II (MP2) a III (MP3)**
- 4 **módy**: mono, stereo, dual channel a joint (intensity) stereo
- filtrace do 32 stejně širokých pásem, podvzorkování 1 : 32, bloky s 32×12 normalizovanými vzorky a dynamická kvantizace vzorků bloků
- paralelně aplikace psychoakustického modelu pro dynamickou kvantizaci – 512/1024-prvková **FFT** u Layer I/II, jedna z 15 kvantizačních tabulek

Obrázek: Obr. MPEG1 Layer I a II



MPEG-1 Layer III (MP3)

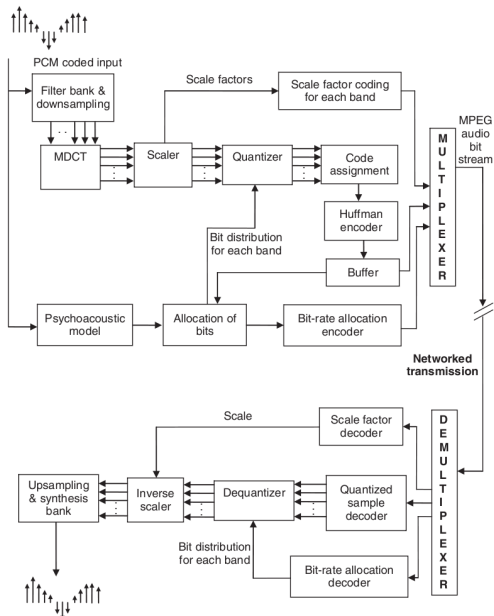
- 18-prvková **Modifikovaná DCT (MDCT)** bloků před normalizací – řešení problému nespojitosti signálu na hranicích oken (\rightsquigarrow „lupání“) pomocí 50% překrytí bloků:

$$F(u) = 2 \sum_{x=0}^{x=N-1} f(x) \times \cos \left[\frac{2\pi}{N} \left(x + \frac{N/2 + 1}{2} \right) \left(u + \frac{1}{2} \right) \right], \text{ pro } u = 0, \dots, N/2 - 1$$

- **Huffmanovo kódování** neuniformně kvantizovaných frekvenčních koeficientů \rightarrow variabilní bit rate

Obrázek: Obr. MPEG1 Layer III

- kvalita transparentní při bit rate 384/296/96 kb/s pro Layer I/II/III na kanál – sluchem nerozeznatelné od originálu

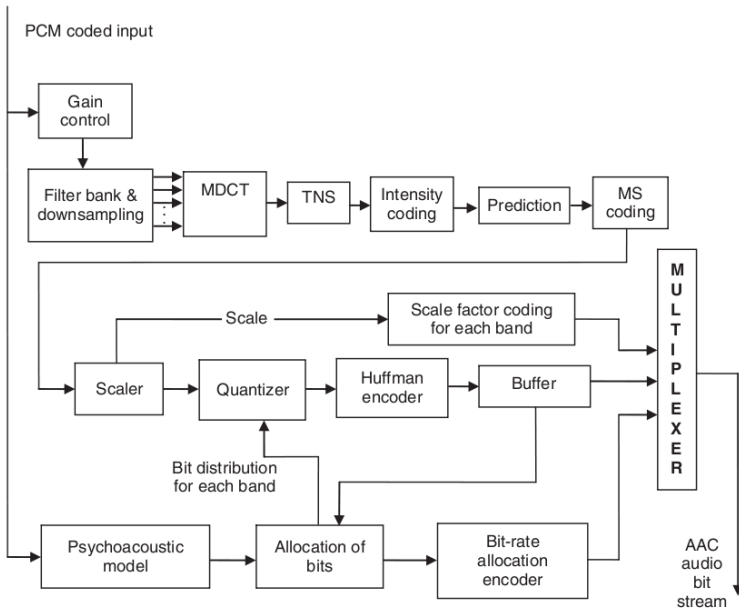


MPEG-2 (AAC)

- podpora 5.1 **surround zvuku**
- 2 skupiny: **BC** (zpětně kompatibilní s MPEG-1, ale pro různé vzorkovací frekvence) a **AAC (Advanced Audio Coding)** – pro DVD-Audio
- za MDCT: temporal noise shaping (pro redukci kvantizačního šumu), stereo intensity coding (redukce irelevancí mezi kanály pomocí kombinace kanálů ve vyšších frekvencích), predikce korelovaných sousedních bloků z předchozího, M/S coding (kódování součtu a rozdílu levého a pravého kanálu)

Obrázek: Obr. MPEG-2 AAC

- vysoká kvalita při bit rate 128 kb/s pro stereo, 320–430 kb/s pro 5.1 (méně než polovina BC)



MPEG-4

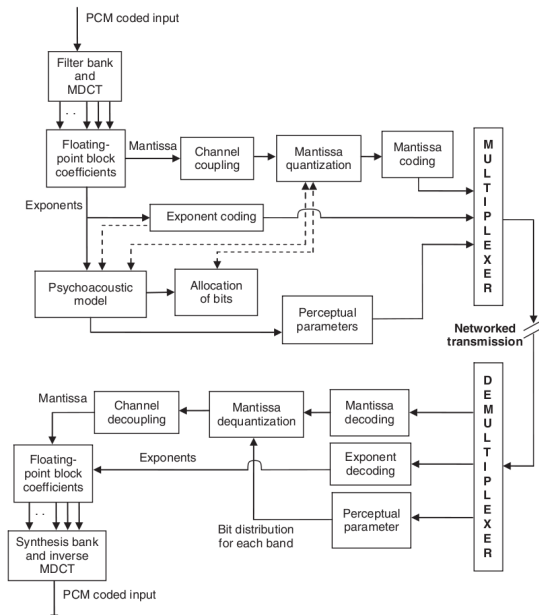
- více než 20 částí: i video, 2D/3D grafika, formát aj. („multimedia standard“), část 3 pro **řeč, hudbu i stukturované audio (SAOL a SASL)**
- řeč: **HVXC** (2 kb/s) a **CELP** (Code Excited Linear Prediction, podobné LPC), syntetizační modul
- hudba: vylepšené verze **MPEG-2 AAC – low complexity (LC, původní), high efficiency (HE, replikace pásem), scalable sample rate (SSR, dvojúrovňová 4/32** filtrace do pásem, transparentní kvalita při 64 kb/s), bit sliced aritmetické kodování (BSAC)

Dolby AC-2/3

- reprezentace **frekvenčních koeficientů v plovoucí řádové čárce** (s více mantisami) po filtraci a MDCT
- využití exponentů v psychoakustickém modelu a kvantizace mantis – dekodovatelná z exponentů a parametrů modelu

Obrázek: Obr. Dolby AC-3

- bit rate 128–192/32–640 kb/s pro AC-2/3 na kanál



ITU G.711/G.722/G.721,6,7

- pro telefonní signál řeči a hudby vzorkovaný na 3/7 kHz, bit rate 64/48–64/16–40 kb/s
- neuniformní kvantizace PCM signálu, podobná A/ μ -law (G.711), rozdělení signálu na pásma nízkých a vyšších frekvencí kódovaných různě (G.722), ADPCM (G.721, G.726, G.727)

ITU G.723,9/G.728

- pro signál 8 kHz po telefonu (s modemem, G.723 – spolu s H.263 pro video H.324, nebo ISDN, G.728), bit rate 5–8/16 kb/s
- aplikace modelů hlasu (CELP), kombinace s ADPCM (G.728)
- **robustní** vůči ztrátám při přenosu a šumu (G.728)



- 1980 (spolu s digitálním audiem a CD), standard pro **elektronické nástroje produkující digitální/elektronickou hudbu** (syntetizátory, sekvencery aj.) – specifikace HW i SW protokolů pro komunikaci
- sémantický model strukturovaného audia, hudební parametry jako **MIDI zprávy**, systémové a kanálové – pro 16 kanálů = nástrojů, zprávy nota zapnuta/vypnuta, pitch-wheel = změna frekvence, (poly)pressure = intenzita (aftertouch, změna při notě), kontrolní (tremolo, vibrato, sustain atd.), změna nástroje aj.
- nevýhody: jen instrukce pro tvorbu zvuků (z uložených vzorků), ne digitální signál ⇒ **různý signál** z různých zařízení (jako různí hudebníci), omezení intenzity nástroje (7b) a časování (sekvenční nesoučasná interpretace zpráv např. plynule měnícího se zvuku flétny) – vhodnější pro diskrétní nástroje jako např. piano



Komprese grafiky

- 1990, na interaktivních CD a ve hrách 2D grafika a animace, dnes narůstající využití 3D grafiky – hry, filmy (efekty, reklama), design, GUI, prezentace (modelování) informací ... → **realistické modelování objektů** z reálného světa
- (přesná) vektorové reprezentace, (lepší) možnosti interakce, skenovací a modelovací (authoring) nástroje, výkonný grafický HW ⇒ rozvoj **počítačové (3D) grafiky**
- v poslední době nárůst objemu (pro přenos/zobrazování) a složitosti geometrické reprezentace (vizuální detaily vs. čas renderování) ⇒ **potřeba komprese** – závislé na reprezentaci
- více reprezentací, nejběžnější **polygonální** pomocí sítě trojúhelníků – de facto standard
- např. postava v 3D hře až 5 tis. polygonů ~ 100 kB (10× víc u modelů v 3D efektech filmů), CAD/CAM modely až 2 mil. polygonů ~ 30 MB – jen geometrie bez obrazových textur!
- např. formáty DODELAT



- z jednoduchých tvarů (např. rovina, krychle, koule aj.) pomocí **modelovacích operací** (geometrické transformace, množinové, vyhlazení aj.)
- **skenování** = vzorkování povrchu objektu, vzorky = body sestaveny do reprezentace

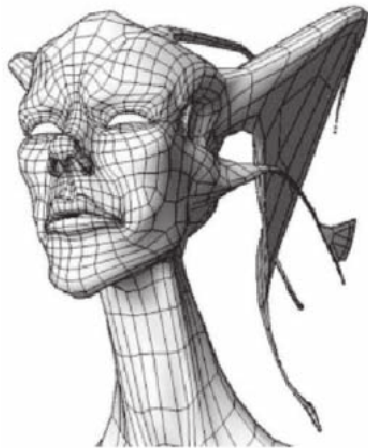
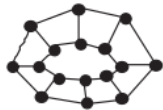
- více, v závislosti na aplikaci a požadované přesnosti
- **vektorové**, pro zobrazení **renderované do rastrového obrazu**
- body: souřadnice (x, y, z) + souřadnice normálového vektoru, barva/souřadnice v textuře, stínování, materiál atd. pro renderování
- 2D oblasti: hraniční body na řádcích souřadnicové mřížky, **quad-tree** = hierarchie nejmenších čtverců postupně dělených na 4 podčtverce obsahujících celý a pak část regionu, hraniční křivky
- křivky: (první) bod a relativní pozice dalších bodů (8/26 pozic v souřadnicové mřížce), **aproximace** propojenými polynomickými (**interpolačními** nebo **spline**) podkřivkami různých stupňů – seznam řídicích bodů (získaných iterativně přidáváním čar nebo rekurzivně dělením čar mezi body křivky)

Polygonální síť (Polygonal meshes)

= síť polygonů – seznam bodů, hran (= dvojice bodů), nebo stěn (= seznam bodů nebo hran)

Obrázek: Obr. Polygonální síť

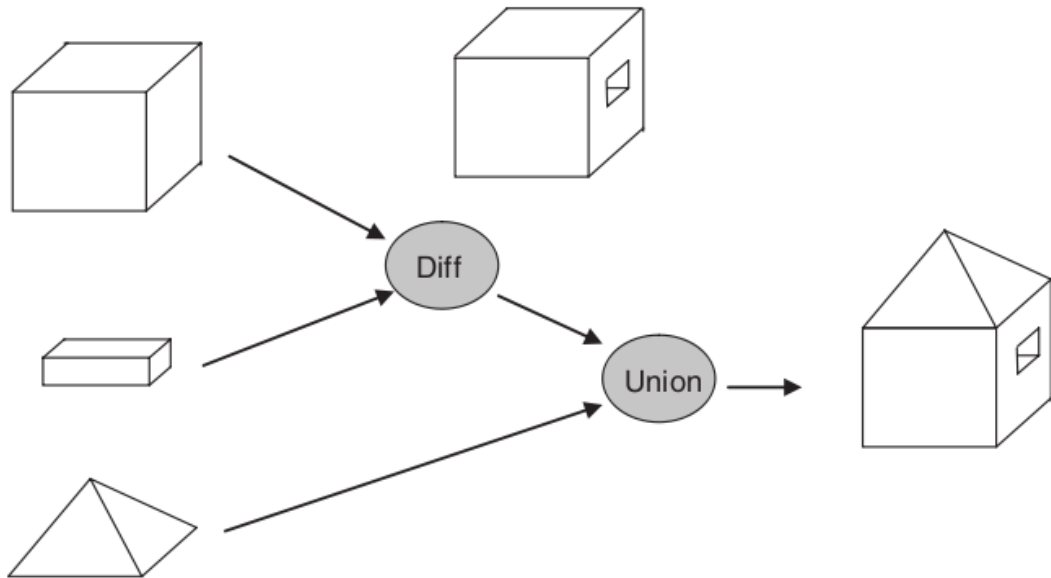
- polygony nejčastěji **trojúhelníky** (popř. čtyřúhelníky, ale lze převést) = **trojúhelníkové síť (triangular meshes)** – de facto standard pro modelování povrchů
- nejpoužívanější, podporována grafickým HW a SW (polygony = atomické jednotky)
- plochy: **aproximace** propojenými polynomickými podplochami (**Bézier**, **NURBS**) různých stupňů – řídící polygonální síť



Konstruktivní geometrie těles (Constructive Solid Geometry)

- pro **přesné modelování** těles v CAD/CAM, nevhodné pro animace
- = úpravy základních parametricky popsaných těles (např. krychle, koule, kvádr, kužel, jehlan atd.) pomocí **množinových operací**

Obrázek: Obr. Konstruktivní geometrie těles



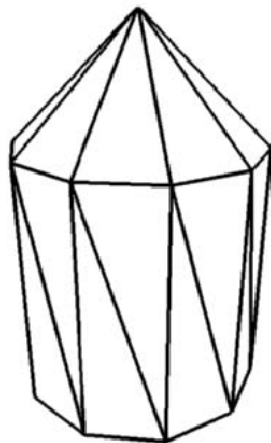
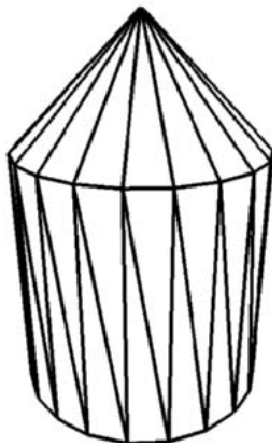
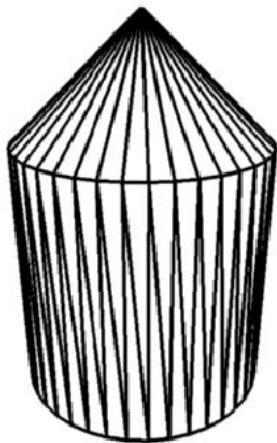


- relativně nová oblast (nárůst objemu a složitosti 3D grafiky), reprezentace objektů trojúhelníkovou sítí
- informace: **pozice a propojení bodů do trojúhelníků**
- redundance:
 - **reprezentační** – vícenásobné reprezentace společných hran a bodů trojúhelníků
 - **povrchová** – body nepřispívající k detailnější aproximaci (při skenování), např. příliš trojúhelníků v rovině

- bezztrátová (přesnost v CAM/CAD) a ztrátová (zobrazování, animace):
 - **komprese parametrů** – kvantizace parametrů bodů, RLE posloupností bodů
 - **komprese kódování konektivity** – redukce informací o propojení trojúhelníků, využití koherence, reprezentační, např. topological surgery
 - **zjednodušení mnohostěnu** = zrušení bodů, hran a trojúhelníků při zachování (vnímání) geometrického tvaru (minimalizace zkreslení), povrchová, např. progresivní síť

Obrázek: Obr. Zjednodušení mnohostěnu

- **podpásmová povrchová komprese** – kvantizace frekvence trojúhelníků na různých úrovních detailů, např. waveletová

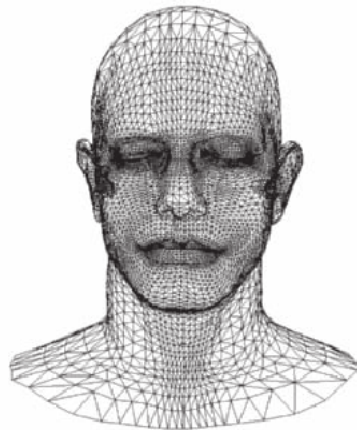
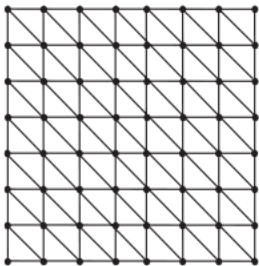


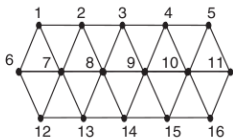
- neexistuje formule pro počet trojúhelníků jako funkce počtu bodů
- **trojúhelníkové mřížky** – strukturované (implicitní) propojení trojúhelníků, např. pro jednoduché parametrické plochy (rovina, koule), terény v 3D kartografii

Obrázek: Obr. Trojúhelníkové mřížky vs. obecná plocha

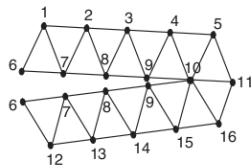
- obecné 3D plochy – potřeba **explicitní** reprezentace propojení trojúhelníků pomocí 3 vrcholů, redundance
- **pruhy trojúhelníků** – implicitní propojení (trojúhelník = tři vrcholy „cikcak“ po sobě), reprezentace seznamem vrcholů, použití např. v OpenGL

Obrázek: Obr. Pruhy trojúhelníků

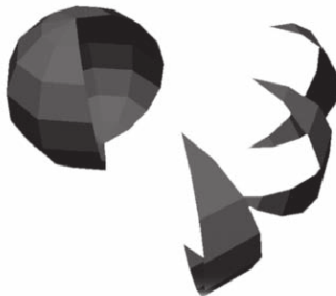
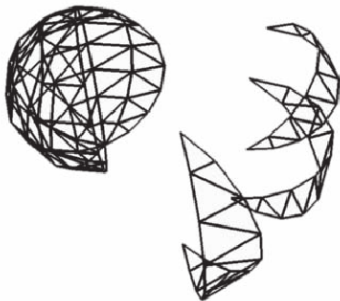




Triangle mesh specified by 16 points and 18 triangles. Each triangle is a triplet of indices (1,6,7) (7,1,2) (2,7,8) (8,2,3) ... (13,8,7) (7,12,13) (12,7,6)—a total of 54 indices.

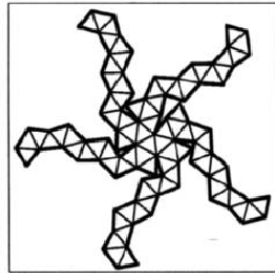
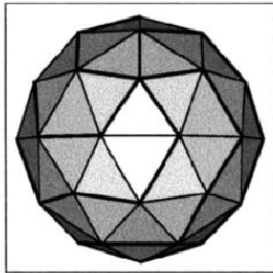
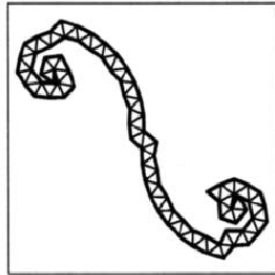
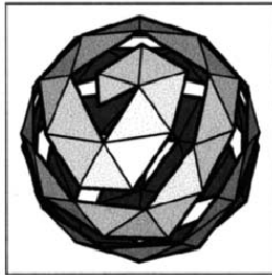
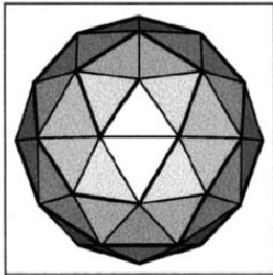


Same mesh is cut open and approximated by a triangular strip with indices 6,1,7,2,8,3,9,4,10,5,11,16,10,15,9,14,8,13,7,12,6—a total of 21 indices. The triangles are implicitly defined as a moving window of three indices—6,1,7 – 1,7,2 – 7,2,8.



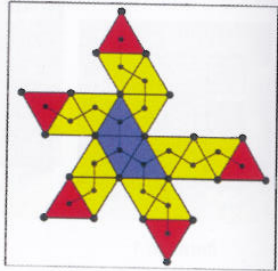
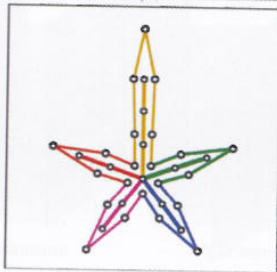
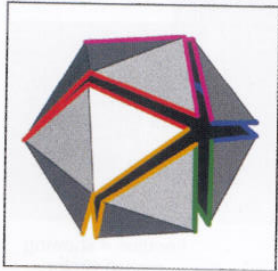
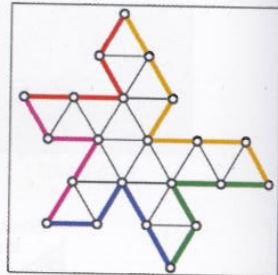
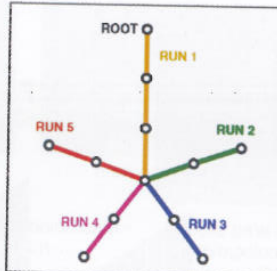
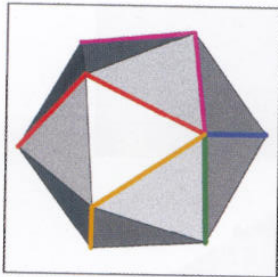
- metoda bezztrátové komprese kódování konektivity, součást MPEG-4 (3D komprese)
- = **dekompozice (rozřezání) trojúhelníkové sítě na pruhy a kódování pruhů** – ideálně 1 pruh, prakticky graf propojených pruhů
- 1** konstrukce **spanning stromu vrcholů** – **heuristické strategie** průchodu vrcholů a výběru následujícího vrcholu (vrcholy mají více sousedů), určuje míru komprese
- 2** **rozřezání sítě** podle cest ve stromu vrcholů od kořene k listům do grafu pruhů = **stromu trojúhelníků, minimalizace počtu pruhů** (a maximalizace jejich délky) u strategie průchodu vrcholů v konstrukci spanning stromu vrcholů, např. do hloubky, šířky, kombinovaný (A* search)

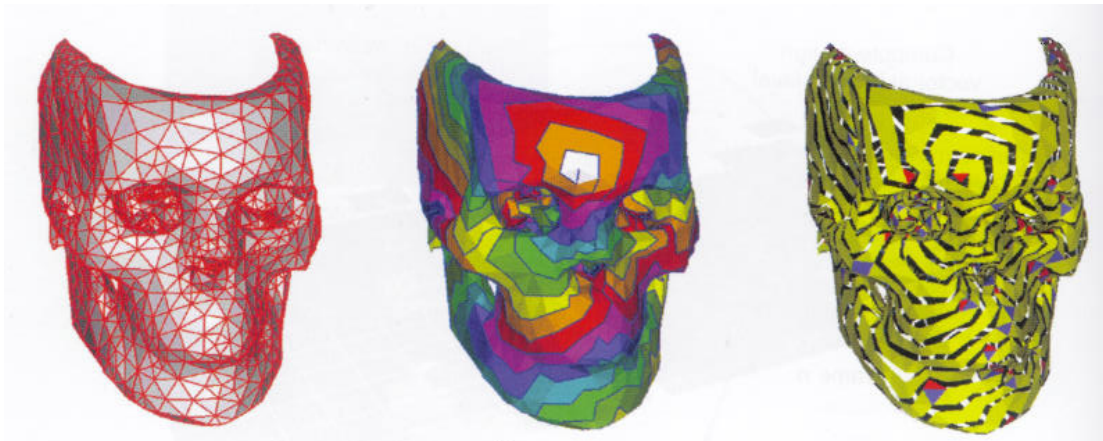
Obrázek: Obr. Strategie rozřezání sítě trojúhelníků



- 3 kódování spanning stromu vrcholů** – uspořádaný seznam vrcholů → diferenční a statistické kódování souřadnic, **minimalizace rozdílu (chyby predikce)** při výběru následujícího vrcholu v konstrukci spanning stromu vrcholů, např. nejbližší (lokální podobnost povrchu)
- 4 kódování stromu trojúhelníků** – větvící (3 sousedi), pruhové (2 sousedi) a listové (1 soused) trojúhelníky – tzv. **marching pattern**, kódování trojúhelníků jako vrcholů stromu

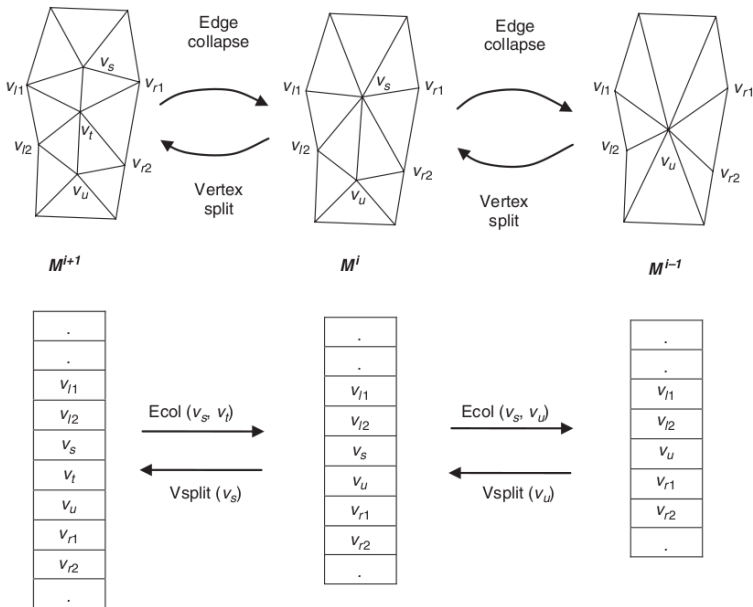
Obrázek: Obr. Topological surgery





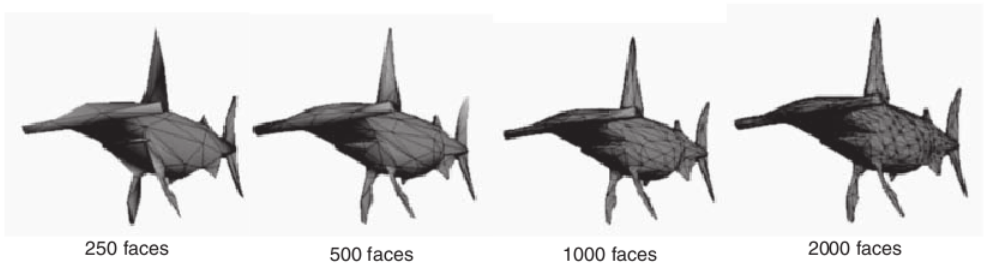
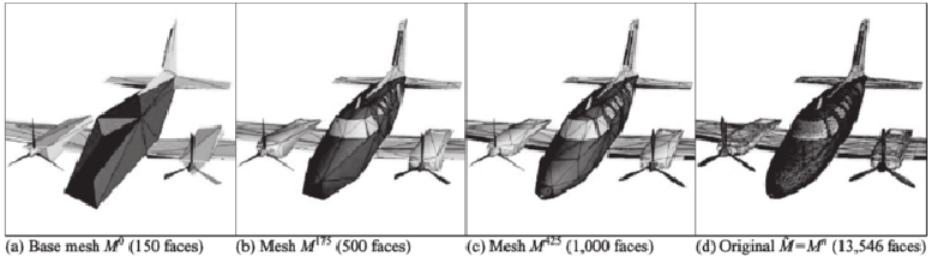
- bezztrátová i ztrátová metoda zjednodušení mnohostěnu
- = nejhrubější **aproximace trojúhelníkové sítě**, příp. spolu s informacemi pro **postupné** zjemnění do původní, transformace:
 - **rozdělení bodu (vertex split)** – přidání nového bodu propojeného s původním, případně posunutým, přidání 2 trojúhelníků sdílejících hranu, operace zjemnění pro dekompresi
 - **zhroucení hrany (edge collapse)** – sloučení bodu hrany s druhým, případně posunutým, zrušení 2 trojúhelníků sdílejících hranu, operace zhrubění pro kompresi, komplementární k rozdělení bodu

Obrázek: Obr. Transformace rozdělení bodu a zhroucení hrany



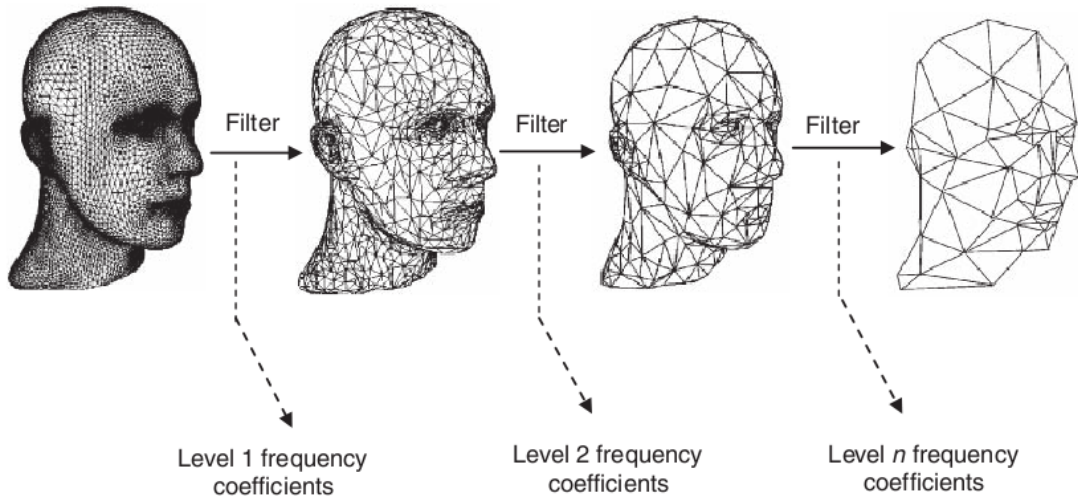
- diferenční kódování přidaného/zrušeného bodu a posunutí bodu, kódování trojúhelníků jako bodu a kombinace 2 susedů
- **minimalizace lokálního zkreslení sítě** při výběru hrany ke zhroucení – např. nejbližší body, ale zároveň nedůležité např. pro zobrazení při dekompresi (neviditelné, vzdálené apod.)

Obrázek: Obr. Progresivní síť



- globální bezztrátová i ztrátová komprese
- = také hrubá aproximace trojúhelníkové sítě spolu s informacemi pro postupné zjemnění, ale 3D frekvenční **(waveletové) transformace na celé síti**
- rozdělení (filtrace) sítě na části s nízkým a postupným vyšším rozlišením (frekvencí) trojúhelníků
- kvantizace frekvenčních (waveletových) koeficientů změn mezi pásmy

Obrázek: Obr. Waveletová komprese





- = uložení/přenos **nejhrubější aproximace** objektů spolu s informacemi pro **postupné zjemnění** do detailnějších reprezentací
- pro dosažení real-time přenosu a interaktivity/renderování, např. 3D síťové hry, vizualizace (vědecké, komerční aj.), animace – práce s hrubějšími objekty, pak renderování do jemnějších
- přirozeně např. progresivní síť (lokální zjemnění), waveletová komprese (globální)

