MPEG audio (Layer 1,2 a 3)

Dušan SOVIČ, 2000

soviar@post.sk
http://www.pakuj.host.sk

O čo tu ide

Tento dokument obsahuje základné informácie o štandarde MPEG pre kompresiu zvuku. Sú tu naznačené základné dôvody, ktoré vedú ku kompresiu zvukového signálu. Pri kompresii zvukového signálu v štandarde MPEG sa využívajú poznatky o tom, ako ľudské ucho vníma zvuk (psychoakustický model, maskovací efekt). Uvedené poznatky sú využité v kompresných algoritmoch Layer 1,2 a 3. Nasledujúci text vám stručne priblíži funkciu kompresných algoritmov Layer.

Obsah

1. Úvod
2. Psychoakustický model a vlastnosti ľudského sluchu
3. Kompresné algoritmy Layer 1,2 a 3
4. Slovko na záver
5. Literatúra

1. Úvod

Skratka **MPEG** (Motion Picture Expert Group) vyjadruje názov skupiny expertov ktorá vznikla v roku 1988 a pracuje pod spoločným vedením ISO (International Standards Organization) a IEC (International Electro-Technical Commision). Skupina pripravuje medzinárodne uznávané štandardy pre kódovanie pohyblivého obrazu (video) a zvuku (audio). Ďalej sa budem venovať len kódovaniu zvuku a kompresným algoritmom Layer 1,2 a 3 a štandardom MPEG-1 a MPEG-2 v ktorých sa používajú.

Krátka špecifikácia štandardov MPEG:

MPEG-1

- prijatý za štandard v roku 1992
- kódovanie video a audiozáznamov s dátovým tokom do 1,5 Mbit/s
- podporované frekvencie vzorkovania: 32kHz, 44.1kHz a 48kHz
- jeden alebo dva audio kanály
- módy: mono, dual chanel, stereo a joint stereo

MPEG-2

- prijatý za štandard v roku 1994
- spätná kompatibilita s MPEG-1
- s dátovými tokmi až 80 Mbit/s pre HDTV (High Definition Television)
- podporované frekvencie vzorkovania: 16kHz ,22.05kHz, 24kHz, 32kHz, 44.1kHz a 48kHz
- viackanálový zvuk

(Pozn.: Pre úplnosť treba dodať, že sú definované aj štandardy MPEG-4 a MPEG-7)

Prečo komprimovať?

Napríklad digitálny zvukový signál v CD-kvalite (odpovedajúci svojou kvalitou záznamu CD-DA, digitálneho zvukového signálu na kompaktných diskoch) s nasledujúcimi parametrami:

- 2 kanály (stereo)
- 16 bitové vzorky
- vzorkovacia frekvencia 44,1 kHz

,vyžaduje dátový tok 1,4 Mbit/s. Len tak pre zaujímavosť v MPEG-1 je maximálny dátový tok 1,5 Mbit/s a z toho je 1,2 Mbit/s vyhradené pre video dáta a 0,3 Mbit/s pre audio dáta. Použitím kompresného algoritmu Layer 1 znížime dátový tok z pôvodných 1,4 Mbit/s na 384 kbit/s, pri Layer 2 na 256-192 kbit/s a s použitím Layer 3 na 128-112 kbit/s pri zachovaní "CD" kvality. (Kompresné algoritmy Layer 1,2 a 3 budú popísané a vysvetlené nižšie).

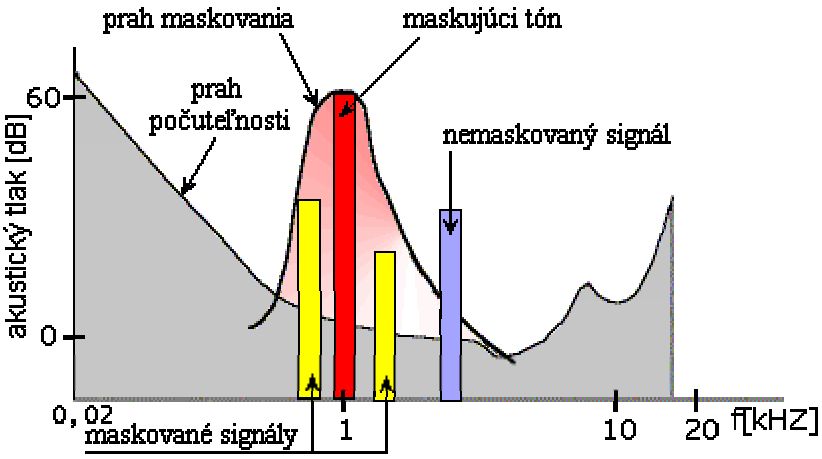
2. Psychoakustický model a vlastnosti ľudského sluchu

Vlastnosti ľudského sluchu sú zhrnuté v psychoakustickom modeli. Tento model vychádza zo spôsobu, akým ľudský sluch vníma zvukové informácie. Využitie tohto modelu pri kompresii zvuku nám umožňuje odstrániť dáta ľudským sluchom nepočuteľné a teda signálov redundantných, bez znateľnej újmy na kvalite reprodukovaného signálu. Samotnou podstatou psychoakustického princípu je **maskovací efekt**. Maskovací efekt je založený na dvoch základných princípoch maskovania: frekvenčného a časového.

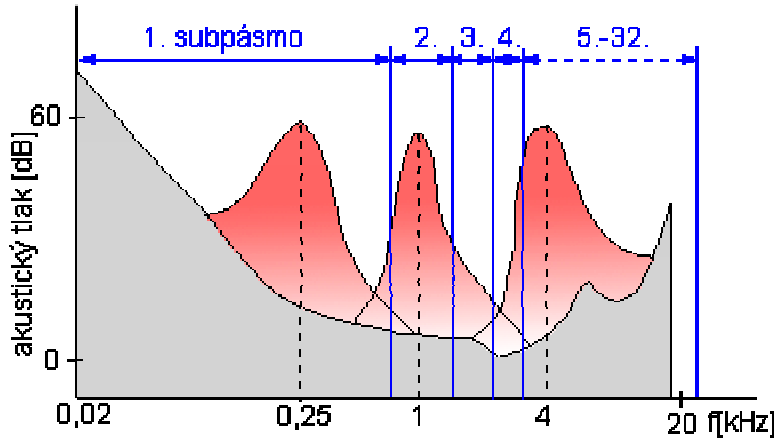
Frekvenčné maskovanie *(frequency masking)*

Využívame nelineárnu citlivosť ľudského sluchu. Jav maskovania spočíva v tom, že jeden silnejší zvuk zakrýva (maskuje) slabší zvuk (tento je maskovaný) aj keď ich frekvencie nie sú rovnaké.

Napríklad: Na obrázku č.1 máme zvukový tón s frekvenciou 1000 Hz s hladinou akustického tlaku približne 60 dB. Tento signál maskuje signály s frekvenciami 990 Hz a 1100 Hz s hladinami akustických tlakov nižšími o 20 resp. 30 dB ako maskujúci signál. Signál s úrovňou tiež nižšou o 20 dB, ale s frekvenciou 2000 Hz už nieje maskovaný.



Obr.1 Princíp frekvenčného maskovania tónom s frekvenciou 1000 Hz a hladinou 60dB.



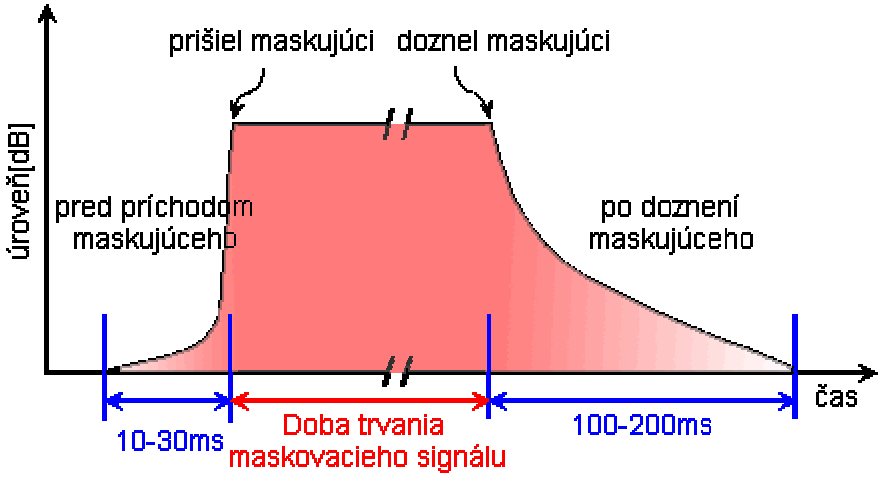
Obr.2 Krivky súčastnej počuteľnosti pri maskovaní jednotlivých frekvencií (0,25; 1; 4kHz)

Zvukový signál sa skladá zo spektra zložiek rôznych frekvencií. Na obrázku č.2 sú znázornené tri najsilnejšie signály s frekvenciami 0,25; 1 a 4 kHz s hladinou akustického tlaku 60 dB. Príslušné krivky maskovacích prahových hodnôt udávajú rozhranie, pod ktorým nie sú ostatné zvuky (príslušných frekvencií a akustických hladín) počuteľné. Túto vlastnosť využijeme pre maskovanie šumu tak, že šumové spektrum budeme "tvarovať" podľa najsilnejších zložiek zvukového signálu. Z premennej hladiny šumu vychádza minimálny počet bitov pre kvantovanie. Inak povedané, budeme zvyšovať šumovú úroveň okolo silného signálu, pretože bude pre ucho nepočuteľný. Zvyšovať šumovú úroveň je to isté ako znižovať počet bitov pre záznam, čo v konečnom dôsledku predstavuje kompresiu. V praxi sa frekvenčný rozsah 20 Hz až 20 kHz rozdelí na 32 subpásiem (subband). V

každom subpásme zvolíme optimálny počet bitov na vzorku, pri ktorom je kvantovací šum ešte maskovaný a teda nedochádza k počuteľnému zníženiu kvality zvuku. Počet prideľovaných bitov pre kvantovanie sa teda adaptívne mení, čím sa výrazne obmedzí dátový tok audiosignálu.

Časové maskovanie *(temporal masking)*

V predchádzajúcom prípade sme uvažovali, že maskovací a maskovaný signál pôsobili dlhšie ako 200 ms. V časovom maskovaní využívame zotrvačnosť ľudského sluchu. Maskovanie nastáva aj pri doznení (zániku) maskovacieho signálu a chvíľu trvá (zhruba v intervale 200 ms, po 200 ms maskovanie úplne zaniká pozri obrázok č.3), než začneme vnímať súčastne pôsobiaci slabší signál. Veľmi zaujímavá je aj skutočnosť, že rovnako môže byť maskovaný krátky zvukový impulz, ktorý príde zhruba 5 ms pred maskujúcim signálom. To je evidentné porušenie v prírode inak striktne dodržovaného zákona kauzality.



Obr.3 Princíp časového maskovania

3. Kompresné algoritmy Layer 1,2 a 3

Pri kompresii audiosignálov v štandardoch MPEG sa využívajú výkonné komprimačné algoritmy označované ako Layer (vrstva). Štandard MPEG zahŕňa tri takéto úrovne kompresie: Layer 1, Layer 2 a Layer 3. Čím vyššia úroveň tým je vyšší kompresný pomer pri rovnakej audiokvalite. Je zachovaná horizontálna kompatibilita týchto úrovní. Teda dekóder pre Layer 3 dokáže dekódovať nižšie vrstvy Layer 1 a Layer 2. Opačným smerom to neplatí. Všetky tri vrstvy používajú tú istú základnú štruktúru. Vychádza sa z vnímateľnej úrovne šumu. Kóder najprv analyzuje spektrálne komponenty audio signálu a potom aplikuje psychoakustický model k odhadu vnímateľnej šumovej úrovne.

Vstupné dáta sa delia na rámce obsahujúce 384 vzoriek. Štruktúra jednotlivých vrstiev:

- **Layer 1** najjednoduchší, psychoakustický model používa iba frekvenčné maskovanie, najmenší dosahovaní kompresný pomer, pôvodne určený pre DCC (Digital Compact Cassette)
- **Layer 2** deliaci filter pracuje s troma rámcami naraz (predchádzajúci, aktuálny a nasledujúci=1152 vzoriek), v psychoakustickom modely je použité aj jednoduché časové maskovanie
- **Layer 3** najvýkonnejší algoritmus, dosahuje najvyššieho kompresného pomeru v porovnaní s vrstvami 1 a 2, lepší deliaci filter s premennou frekvenciou, psychoakustický model využíva úplné časové maskovanie

Najväčšiu zásluhu na rozvoji aplikácií MPEG-1 Layer 3 má Frauenhoferov inštitút pre integrované obvody (Fraunhofer Institut Integrierte Schaltungen-IIS) ležiaci v nemeckom Erlagene, ktorý sa v roku 1987 zapojil do projektu EUREKA v oblasti kódovania zvuku na základe vnímania (perceptual audio coding). V spolupráci s univerzitou v Erlagene, IIS vyvinul efektívny algoritmus, ktorý bol štandardizovaný ako ISO MPEG Audio Layer-3.

Príklad použitia kompresných algoritmov pri kompresii audiosignálu v "CD-kvalite" a dosahované výsledky boli uvedené vyššie. Efektívnosť jednotlivých algoritmov v danom príklade ukazuje nasledujúca tabuľka.

Tabuľka 1 Efektívnosť jednotlivých algoritmov

Pomer	Layer	Požadovaná rýchlosť prenosu
1:4	1	384 kbit/s pre stereo signál
1:6 – 1:8	2	256 - 192 kbit/s pre stereo signál
1:10 – 1:12	3	128 - 112 kbit/s pre stereo signál

Tabuľka 2 Typické hodnoty dosahované algoritmom MPEG Layer-3

Kvalita	Šírka pásma	Mód	Rýchlosť prenosu	Faktor kompresie
Telefónna	2,5 kHz	mono	8 kbit/s	96:1
Rozhlas na DV	4,5 kHz	mono	16 kbit/s	48:1
Rozhlas na SV	7,5 kHz	mono	32 kbit/s	24:1
Rozhals na VKV	11 kHz	stereo	56-64 kbit/s	26-24:1
Blízka CD kvalite	15 kHz	stereo	96 kbit/s	16:1
CD kvalita	> 15 kHz	stereo	112-128 kbit/s	14-12:1

Princíp činnosti algoritmov Layer 1,2 a 3

Vývojový diagram zvukového kóderu MPEG Audio Layer 2 je na obrázku č.4. Nasledujúci text vysvetľuje zjednodušene princíp činnosti kóderov MPEG Audio Layer 1,2 a 3.

V banke filtrov sa prevedie **dekompozícia signálu na 32 subpásiem** (je to prevod z časovej do frekvenčnej oblasti). V layer 3 je ešte za bankov filrov v kaskáde zaradený obvod s MDCT (modifikovaná diskrétna kosínusová transformácia) pre dosiahnutie jemnejšieho rozlíšenia frekvencií. Pre všetky vrstvy je pri dekompozícii na subpásma použité tzv. kritické vzorkovanie tj. počet všetkých vzoriek na jednotku času v každom subpásme krát počet subpásiem je rovnaký ako počet vzoriek signálu pred dekompozíciou.

Pre každé subpásmo sa stanovia **činitele merítka** (napr. ako maximálna amplitúda z 12 vzoriek v jednom subpásme pre layer 1, v layer 2,3 je v subpásme viac vzoriek ako 12 a podľa potreby sa určí aj viac činiteľov merítka). Činitele merítka majú rozhodujúcu úlohu pri stanovení maskovacích prahov v každom subpásme. V

časti, ktorá reprezentuje **psychoakustický model** sa porovnávajú činitele merítka s hodnotami získanými pri experimentoch a výsledkom je pomer signálu k maskovaciemu prahu **SMR** (Signal to Mask Ratio), teda hodnota pomeru užitočného signálu k šumu, pod ktorou môžeme očakávať zamaskovanie skreslenia zvuku. Podľa SMR sa pridelí v každom subpásme vždy pre dvanásť vzoriek určitý počet bitov *b* ktorými budú **kvantované**. Tento počet je rovnaký pre uvedenú skupinu vzoriek, ale odlišný pre každé subpásmo. Zmenšenie počtu bitov (2-15) sa priaznivo prejaví na prenosovej rýchlosti výsledného bitového toku.
Ďalšou úlohou činiteľov merítok je **normovanie signálu** vo všetkých 32 subpásmach. Normovanie sa deje tak, že signál sa delí príslušným činiteľom merítka. Z toho vyplýva, že maximálna amplitúda v subpásme je normovaná na jednotku. Následný kvantizér dostáva rovnakú max. úroveň, čo znamená, že jeho maximálna prebuditeľnosť je v jednotlivých rámcoch rovnaká a takto môžeme zamerať kvantizáciu na oveľa užšie rozmedzie dynamického rozsahu vzoriek. Prideľovanie počtu bitov na kvantovanie bolo už vysvetlené vyššie. Informácie o merítkach ScFSI (Scale Factor Select Information) sú zakódovaná pomocou 6 bitov.

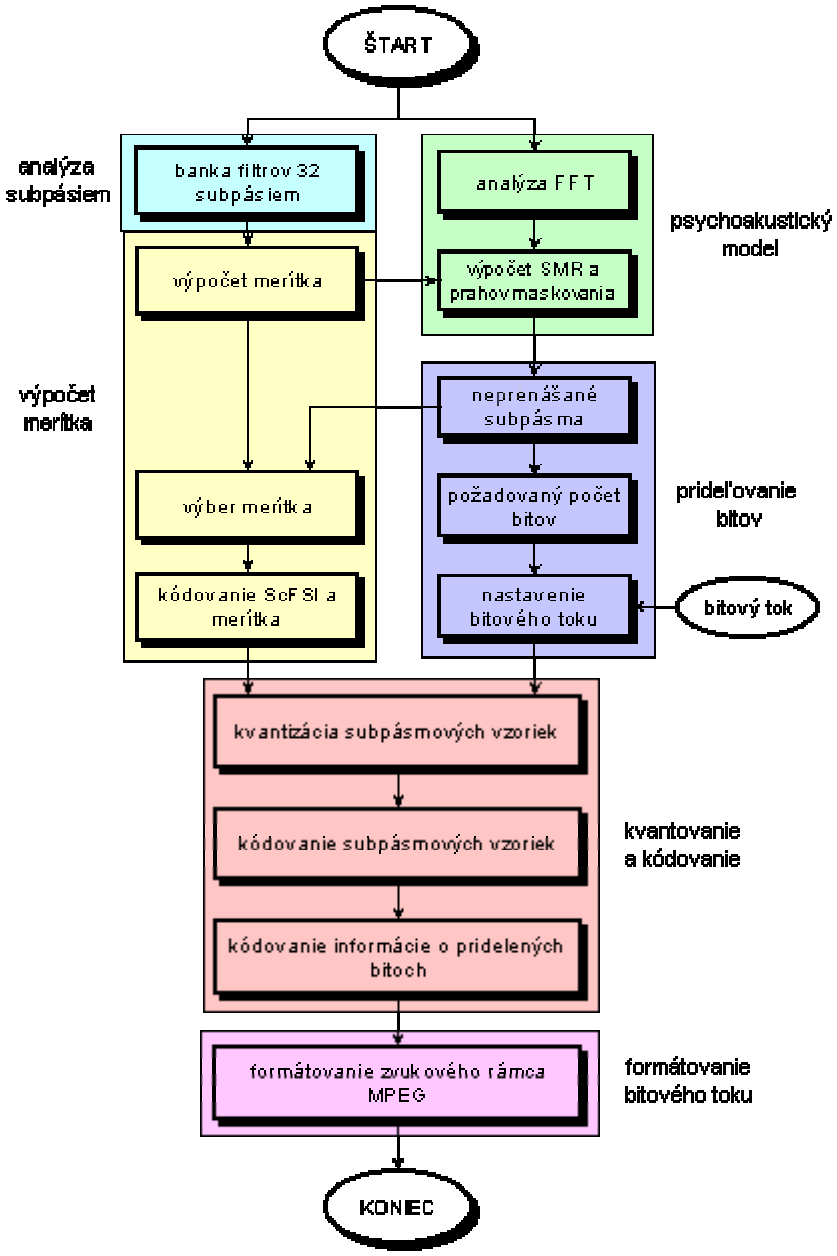
Treba ešte spomenúť, že psychoakustický model nedostáva informácie len o veľkosti vzoriek, ale aj o tom, či ide o tónový alebo netónový signál. Túto informáciu potrebuje preto, lebo priebeh maskovacieho prahu závisí aj na časovej postupnosti, tzn. či je tón sínusový alebo ide len o ojedinelí ráz. Preto vstupný signál PCM podrobujeme **FFT** (rýchla Fourierova transformácia) , ktorá určí o aký z uvedených signálov ide. Podľa týchto informácií potom v obvode psychoakustického modelu "modelujeme" tvar maskovacej krivky.

Dôležitým krokom pri kódovaní je výber vhodného psychoakustického modelu. Každý výrobca kóderu môže použiť svoj vlastný model.

V layer 1 a 2 sú vzorky zakódované s pevnou dĺžkou a v layer 3 sú kódované premenlivým Huffmanovým kódovaním. Kódovanie layer 1 je veľmi podobné kódovaniu layer 2, rozdiel je v možnosti použitia premenlivého merítka pre skupiny 12 vzoriek, keď pomocou prídavnej informácie môžeme špecifikovať počet subpásiem, ktoré máme uvažovať a ktoré budeme ignorovať.

Výsledný bitový tok v jednom rámci potom obsahuje okrem hodnôt vzoriek aj ďalšie informácie ako je záhlavie, zabezpečenie proti chybám, pridelenie počtu bitov pre kvantovanie, počet činiteľov merítka, veľkosti činiteľov merítka a ďalšie prídavné dáta. Tieto dáta sú potrebné poskytnúť dekóderu aby vedel signál dekódovať do pôvodnej podoby.

Zvukový dekóder MPEG neobsahuje psychoakustický model ani procedúru na prideľovanie bitov pre jednotlivé vzorky v každom z 32 subpásiem a teda dekóder je oveľa jednoduchší a lacnejší ako kóder. Všetky potrebné informácie dostáva v zakódovanom tvare spolu s kódovaným signálom. Z týchto informácií a z dekódovaných vzoriek obnoví ich pôvodné hodnoty. Po potrebných frekvenčných presunoch realizovaným bankov filtrov , zlúčenie signálov jednotlivých subpásiem a prevode z paralelného na sériový tvar je na výstupe dekóderu opäť digitálny audiosignál v tvare ako na vstupe kóderu MPEG.



Obr. 4 Vývojový diagram zvukového kóderu MPEG Audio Layer 2

MPEG-1; štyri módy zahrnuté do štandardu:

- **mono**
- **dual chanel** *(duálne kanály)* - komprimácia dvoch nesúvisejúcich mono kanálov, bitrate(prenosová rýchlosť) je rovnomerne rozložená medzi oba kanály
- **stereo** - komprimácia dvoch súvisejúcich kanálov, možnosť prideliť niektorému kanálu väčšiu bitrate než druhému, tzn. bitrate sa nedelí rovnomerne (napr. v jednom kanály je ticho a druhému je pridelené viac bitov na jeho zakódovanie)
- **joint stereo** *(spojené stereo)* - tento mód využíva ku komprimácii vlastnosti ľudského sluchu pri vnímaní stereosignálu. Priestorová lokalizácia zdroja signálu závisí pri nízkych frekvenciách na amplitúdových a fázových charakteristikách, pri vysokých frekvenciách len na amplitúdových charakteristikách. To znamená, že pre vyššie frekvencie postačuje pre správnu lokalizáciu dosiahnuť správneho pomeru akustického výkonu pre pravé a ľavé ucho, bez ohľadu na fázu, zatiaľ čo pri nižších frekvenciách musíme uvažovať i fázové charakteristiky. Mód joint stereo umožňuje, aby sa pri kódovaní určilo medzné dielčie frekvenčné pásmo, nad ktorým sa merítka vzoriek signálu z ľavého a pravého kanálu budú počítať síce nezávisle, ale ich hodnoty po vydelení merítkom budú spriemerované. K uchovaniu či prenosu budú určené práve tieto

priemerné hodnoty a budú tiež použité pre rekonštrukciu signálu oboch kanálov. Tím sa odstráni relatívny rozdiel fázových charakteristík medzi ľavým a pravým kanálom. Pod určenou hranicou frekvencie subpásma prebehne kódovanie bezo zmeny. Táto metóda vedie pri typických aplikáciách k úspore nárokov na prenosovú rýchlosť rádovo v rozmedzí 5 až 10 %.

MPEG-2; viackanálový zvuk

Viackanálový zvuk je obvykle zložený z ľavého predného *(Left)* a pravého predného *(Right)* kanálu, stredového kanálu *(Center)* určeného pre dialóg, ľavého priestorového *(Left Surround)* kanálu a pravého priestorového *(Right Surround)* kanálu a kanálu pre nízkofrekvenčné efekty LFE *(Low Frequency Enhancement)* pre subwoofer.

4. Slovko na záver

Na záver by som chcel len dodať, že formáty Layer 1,2 a 3 nie sú jedinými formátmi pre záznam zvuku. Dolby Laboratories vyvinuli formát **AC-3** podobný s Layer 3, určený hlavne pre HDTV, surround sound a domáce kino, s tokmi pri päť kanáloch plus jednom nízkofrekvenčnom kanály okolo 320 kbit/s. Firma Yamaha s formátom **VQF** môže tiež plne konkurovať Layer 3 (vraj pri 80 kbit/s produkuje lepší zvuk ako Layer 3 pri 128 kbit/s).
Skupina MPEG pokračuje v tvorbe štandardov ďalej a výsledkom ich snaženia je štandard **AAC (Advanced Audio Coding)** "zdokonalené kódovanie zvuku", ktorý bude súčasťou špecifikácie MPEG-4, kde sa uvažuje aj s využitím formátu VQF pri prenosu zvuku s nižším tokom dát. Testy ukazujú, že AAC je asi o 1/3 lepší ako Layer 3. To znamená, že AAC s 96 kbit/s má lepšiu kvalitu ako Layer 3 s 128 kbit/s.

5. Literatúra

[1] Čapek, J., Fabian, P.: *"Komprimace dat principy a praxe"*, Computer Press, 2000
[2] Líška, D.: *"Digitálna terestriálna televízia DVB-T"*, Telekomunikace č.4/2000
[3] Toman, M.: http://home.zcu.cz/~mtoman/mp3.htm (Vinikajúca stránka o MP3 na i-nete)
[4] Bier, J.: *"Digital Audio Compression: Why,What and How"*, Berkely Design Technology, Inc., 2000