



Úvod do metód spracovania zvuku v súčasnom multimedialnom prostredí

Diplomová práca

Pavol Adam

Univerzita Komenského, Bratislava
Fakulta matematiky, fyziky a informatiky
Katedra aplikovanej informatiky

študijný odbor: informatika

Konzultant: Ľubomír Lúčan, CSc.

Bratislava 2006

Pod'akovanie

Ďakujem:

- dobrému Bohu, že nám dal život a múdrosť,
- mojim rodičom, že ma s láskou vychovali a dali mi zázemie,
- môjmu konzultantovi Ľubomírovi Lúčanovi, CSc. za cenné rady, vedenie, usmerňovanie,
- Doc. RNDr. Milanovi Ftáčnikovi, CSc. a doc. RNDr. Andrejovi Ferkovi, PhD. za cenné rady a povzbudenie,
- mojej milovanej polovičke Majke, rodine a ďalším blízkym za trpezlivosť a podporu,
- Maťovi, Jankovi a Dominikovi za to, že mi vytvorili priestor na prácu,
- Šibovi, Mirovi, Peťovi, Ľubke a p. Zajacovi za rôzne rady a nájdené chyby.

Ďestné prehlásenie

Prehlasujem, že som túto prácu vytvoril samostatne, vedený radami a usmernením konzultanta, čerpajúc z uvedenej literatúry, zdrojov dostupných na internete a skúseností priateľov.

.....
Pavol Adam

Abstrakt

Práca – ako knižná publikácia, i vo forme internetovej učebnice – ponúka úvod do metód spracovania zvuku vo svete výpočtových technológií. Oboznamuje čitateľa s postupmi pri stratovej a bezstratovej kompresii zvuku, trendmi v oblasti záznamu, spracovania a reprodukcie zvuku. Súčasťou práce je aj súhrnný výpočet jednotlivých zvukových formátov a ich porovnanie.

Kľúčové slová: metódy spracovania zvuku, kompresia zvuku, zvukové formáty

Obsah práce

1 Úvod	7
2 Analógový svet zvuku	8
2.1 Zvuk ako fyzikálny fenomén	9
2.1.1 Frekvencia a vlnová dĺžka	9
2.1.2 Hlasitosť zvuku	10
2.2 Psychoakustika	12
2.2.1 Fyziológia ľudského ucha	12
2.2.2 Vnímanie hlasitosti	13
2.2.3 Vnímanie kvality, harmonické tóny	15
2.2.4 Maskovanie zvukov	17
2.2.5 Vnímanie frekvencie	18
2.2.6 Hudobný vnem	19
2.3 Charakteristiky zvuku	21
2.3.1 Druhy zvukov	21
2.3.2 Zvuky hudobných nástrojov	22
2.3.3 Ľudský hlas	23
2.3.4 Úprava a zmena zvuku	24
2.4 Analógové spracovanie zvuku	26
2.4.1 Záznam zvuku	26
2.4.2 Spracovanie zvuku	27
2.4.3 Reprodukcia zvuku	27
2.5 Priestor vo zvuku	30
2.5.1 Problémy monofonického záznamu a reprodukcie	31
2.5.2 Stereofonický záznam zvuku	31
2.5.3 Súčasné trendy	33
3 Digitálny svet zvuku	34
3.0.1 Definícia pojmov	35
3.1 Problematika vzorkovania zvuku	39
3.1.1 Vzorkovanie	39
3.1.2 Kvantovanie	41
3.1.3 Problémy digitálneho spracovania zvuku	43
3.2 Digitálne kódovanie čísel	44
3.2.1 Kompresia a kompakcia, redundancia a irelevancia	44
3.2.2 Huffmanov kód	46
3.2.3 Aritmetické kódovanie	47
3.2.4 Modifikácie Huffmanovho a aritmetického kódovania	48
3.3 Princípy stratovej kompresie zvuku	50
3.3.1 Diskrétné ortogonálne transformácie	50
3.3.2 Fourierova transformácia, kosínusová transformácia	51
3.3.3 Ortogonálne transformácie v praxi	52
3.3.4 Vektorové kvantovanie (VQ).....	56
3.3.5 Kompresia stereofonického záznamu	57

3.4	Princípy bezstratovej kompresie zvuku	59
3.4.1	<i>Bezstratové transformačné kódovanie</i>	59
3.4.2	<i>Kódovanie lineárnou predikciou</i>	60
3.4.3	<i>Hybridné kódovanie</i>	62
3.4.4	<i>Kompresia stereofonického záznamu</i>	63
3.5	Zvuk v súčasnom svete elektroniky	64
3.5.1	<i>Štruktúra zvukového systému PC</i>	64
3.5.2	<i>Priestorový zvuk a zvukové kanály</i>	66
3.5.3	<i>Priestorový zvuk v softvéri</i>	69
3.5.4	<i>Zvuková syntéza a MIDI</i>	70
3.5.5	<i>Hudobné sekvencery, moduly</i>	73
3.5.6	<i>Ochrana duševného vlastníctva</i>	74
3.5.7	<i>Úložné médiá</i>	77
3.5.8	<i>Zvuk a video</i>	80
3.5.9	<i>Rozhrania na prenos zvuku</i>	82
3.5.10	<i>MPEG</i>	83
3.5.11	<i>Predpokladaný vývoj</i>	85
4	Prehľad stratových zvukových formátov	87
4.1	ADPCM	88
4.2	MPEG-1/2 Audio (MP3)	89
4.2.1	MP1, MP2	89
4.2.2	MP3	90
4.2.3	MP3pro	90
4.2.4	MP3 Surround	91
4.2.5	Zhrnutie	92
4.3	MPEG-2/4 Audio (AAC)	94
4.3.1	HE-AAC (AACplus)	96
4.3.2	Zhrnutie	96
4.4	Windows Media Audio (WMA)	98
4.5	RealAudio (RA)	100
4.6	ATRAC	102
4.7	Dolby Digital (DD, AC-3)	104
4.8	Coherent Acoustics (CAC, DTS)	106
4.9	Ogg Vorbis	108
4.10	Musepack (MPC)	111
4.11	Kompresia hovoreného slova	112
4.11.1	<i>Speex</i>	113
5	Prehľad bezstratových zvukových formátov	115
5.1	Shorten	117
5.2	WavPack	119
5.3	Monkey's Audio (MA, APE)	121

5.4	FLAC	123
5.5	Lossless Audio (LA)	125
5.6	True Audio (TTA)	126
5.7	OptimFROG	127
5.8	MPEG-4 ALS	128
5.9	MPEG-4 SLS	130
5.10	Meridian Lossless Packing (MLP)	131
6	Porovnanie zvukových formátov	133
6.1	Porovnanie stratových kompresných schém	135
6.2	Porovnanie bezstratových kompresných schém	139
7	Záver	142
8	Zoznam použitej literatúry	143

Zoznam obrázkov

Obr. 2.1	- Krivky rovnakej subjektívnej hlasitosti v závislosti od frekvencie tónu ...	14
Obr. 2.2	- Rôzne tvary zvukových vln: sinusoida, štvorcová vlna, trojuholníková vlna a pílová vlna	21
Obr. 3.1	- Schéma stratovej kompresie zvuku – po transformácii (A) nasleduje kvantovanie (Q), výsledné čísla sa kódujú entropickým kódom (EC)	55
Obr. 3.2	- Schéma bezstratového transformačného kódovania	60
Obr. 3.3	- Demonštrácia spojitého charakteru zvuku	60

Zoznam tabuliek

Tab. 2.1	- Porovnanie hladiny akustického tlaku (L_p) s akustickým tlakom (p), ktoré generujú rôzne zdroje zvuku	13
Tab. 6.1	- Prehľad vlastností vybraných formátov na stratovú kompresiu zvuku	138
Tab. 6.2	- Prehľad vlastností vybraných formátov na bezstratovú kompresiu zvuku	141

1 ÚVOD

Sluch je jedným z piatich základných zmyslov človeka. Má dôležité miesto nielen v signalizácii. Pri zachovaní typických štruktúr a charakteristík získava význam komunikačný (reč), ba i estetický (spev, hudba). Snaha o reprodukciu počutého zvuku spôsobuje, že sa dieťa naučí hovoriť. Človek oddávna na rovnakom princípe vytváral návnady na zver. Spev a hudobné nástroje sú súčasťou snád' každej kultúry.

V súčasnej dobe (či už z dôvodov vojenských, komerčných či iných) vyvstáva potreba zvuk zaznamenávať, spracúvať a reprodukovať. Moderné technológie to umožňujú vykonávať digitálne. V našej práci sa pokúsime zhrnúť rôzne metódy na zápis dát reprezentujúcich zaznamenaný zvuk.

V Kapitole 2 (*Analógový svet zvuku*, str. 8) čitateľa oboznámime so zvukom v prirodzenom, analógovom prostredí, popíšeme objektívne charakteristiky zvuku z pohľadu fyziky a poskytneme základné informácie o vnímaní zvuku človekom – poslucháčom.

Kapitola 3 (*Digitálny svet zvuku*, str. 34) je jadrom práce. Čitateľa v nej oboznámime s technikami digitalizácie zvuku za účelom počítačového spracovania zvukovej informácie. Poskytneme prehľad metodík používaných pri kódovaní zvuku, predovšetkým stratovej a bezstratovej kompresie. Na dôvažok načrtujeme aktuálny stav a trendy v používaní „digitálneho zvuku“.

Kapitola 4 (*Prehľad stratových zvukových formátov*, str. 87) a kapitola 5 (*Prehľad bezstratových zvukových formátov*, str. 115) poskytnú výpočet jednotlivých zvukových formátov rozšírených v súčasnom multimediamiálnom prostredí. Ich porovnanie uvedieme v kapitole 6 (*Porovnanie zvukových formátov*, str. 133).

V závere (kapitola 7, str. 142) uvedieme prínosy našej práce.

Veríme, že naša práca uľahčí mnohým čitateľom spoznať základné pojmy a metódy z oblasti spracovania zvuku a že i pre ľudí s pokročilými znalosťami z tejto oblasti bude dobrou referenciou, zachytávajúcou aktuálne postavenie zvuku a zvukových formátov v multimediamiálnom prostredí na prelome rokov 2005 / 2006.

Autor

Pozn: Od čitateľa očakávame poznatky z matematiky a informačných technológií aspoň na úrovni stredoškolského vzdelania.

2 ANALÓGOVÝ SVET ZVUKU

Aby bolo možné uvažovať o kvalitách zvukových formátov a princípoch ich činnosti, musíme čitateľa oboznámiť so základnými vlastnosťami zvuku v prostredí, v ktorom existuje – toto prostredie je spojené, teda analógové. Nasledujúce časti práce sú stručným prierezom charakteristík zvuku, jeho vnímania a spracovania v analógovom chápaní tohto fenoménu. Bližšie sa im venujú vedné odbory akustika (a jej odvetvia) a elektrotechnika (najmä náuka o spracovaní signálu¹).

Zvuk je v prvom rade jav skutočného, fyzikálneho sveta, preto v časti 2.1 *Zvuk ako fyzikálny fenomén* (str. 9) zhrnieme poznatky o ňom na úrovni, ktorá bude potrebná na ďalšie pochopenie textu.

Časť 2.2 *Psychoakustika* (str. 12) sa venuje vnímaniu zvuku z pohľadu človeka. On je, koniec koncov, ten, čo zvuk nielen tvorí, zaznamenáva, spracúva a reprodukuje, ale najmä konzumuje. Tomuto účelu treba podriadiť ostatné spomenuté články reťazca.

Zvuky, ktoré treba zaznamenať a reprodukovať, nie sú náhodné javy, ale zväčša hudba, spev či hovorené slovo. Každý z nich má dôležité i menej výrazné atribúty, o ktorých hovoríme v časti 2.3 *Charakteristiky zvuku* (str. 21).

Aby sme so zvukom mohli pracovať na osobných počítačoch, musíme ho vedieť zaznamenať a spracovať, neskôr aj reprodukovať. Do týchto procesov vstupujú vlastnosti elektronických súčiastok a zariadení, ktoré pri tom použijeme. Podrobnejšia analýza je v časti 2.4 *Analógové spracovanie zvuku* (str. 26).

Človek (v role poslucháča) má dve uši, preto okrem zvuku samotného vníma aj smer, z ktorého prichádza. O princípoch lokalizácie zdroja zvuku a o stereofonickom vneme hovorí viac časť 2.5 *Priestor vo zvuku* (str. 30).

¹ Zvuk a elektrický signál, najmä vo fáze zaznamenávania a spracúvania, majú veľa spoločného, preto budeme v práci často hovoriť o zvuku alebo jeho zápise či reprezentácii (analógovej alebo digitálnej) ako o signáli.

2.1 Zvuk ako fyzikálny fenomén

Na laickej úrovni možno o zvuku hovoriť ako o vlnení prostredia, ktoré človek potenciálne môže vnímať. Keďže prirodzeným prostredím človeka, ktoré ho obklopuje, je vzduch, ide najmä o **vibráciu vzduchu**. (Do úvahy prichádza napríklad aj voda – človek pod hladinou vody tiež vníma zvuk, ale o tomto špeciálnom prípade nebudeme ďalej uvažovať. Z pohľadu fyzikálnych zákonitostí sa však zvuk šíri všetkými tuhými, kvapalnými a plynými látkami.) Na vneme **počutia** (vnímanie zvuku poslucháčom) má podiel predovšetkým ľudský sluch.

Odborne sa dá zvuk stručne popísať ako „zmena tlaku, rozloženia častíc alebo ich rýchlosti, ktorá sa šíri pružným materiálom.“². Fakticky ide o pozdĺžne vlnenie (v aspoň trochu stlačiteľnom médiu), teda o vlny, ktoré vibrujú v smere svojho šírenia. **Vlna**, ako akási jednotka vlnenia, je potom lokálna zmena tlaku vzduchu v čase, zložená z **deja kompresie** prostredia (stlačenia zmrštenia, zvýšenia tlaku) a **dekompresie** prostredia (roztiahnutia, rozpínania, zníženia tlaku).

Treba podotknúť, že zvuk nie je pohyb média, ale zmena jeho vnútorného stavu. Tento princíp sa opisne vysvetľuje na papierovej loďke na povrchu jazera pri pohľade zhora. Ak hodíme do vody kameň, vlny sa budú šíriť radiálne od miesta dopadu a keď zasiahnu loďku, táto bude stúpať a klesať podľa momentálnej **fázy**³ vlny. Pohyb vody je však len zdanlivý, loďka totiž – odhliadnuc od vertikálnych pohybov – ostane na tom istom mieste. Ak teda abstrahujeme rozmer „výšky“, ide o šírenie pozdĺžnej vlny. Na objasnenie (jednorozmerného) pozdĺžneho vlnenia možno použiť aj napr. dlhú pružinu v pokojovom (nestlačenom a neroztiahnutom) stave, ktorú na jednom okraji upevníme a druhým okrajom náhle pohneme smerom k prvému okraju. Pružina sa na tomto mieste „neprirodzene“ stlačí, pričom táto tlaková vlna postupuje ďalej. Za ňou zas nasleduje „neprirodzené“ roztiahnutie.

V tejto časti práce si spomenieme základné fyzikálne vlastnosti zvuku. Obsahuje dva odseky: 2.1.1 *Frekvencia a vlnová dĺžka* (str. 9) a 2.1.2 *Hlasitosť zvuku* (str. 10).

2.1.1 Frekvencia a vlnová dĺžka

Zvuk je prejav zložený z rôznych vlín. Ak je však tento jav periodický, podľa Fourierovej vety⁴ ho možno rozložiť na sínusové vlnenia s rôznymi kmitočtami a amplitúdami. Je to práve sinusoida, na ktorú ľudské ucho reaguje ako na „čistý tón“ nejakej výšky. Každá sínusová zvuková vlna, teda dej kompresie i dekompresie, nastáva s určitou pravidelnosťou. Rýchlosť tejto oscilácie sa označuje **frekvencia**

2 Cit. OLSON, Harry F. *Acoustical Engineering*. 1957. Ďalej v texte budeme tento jav jednoducho označovať **zmena tlaku**.

3 Stupeň priebehu (časové vymedzenie priebehu) vlny medzi kompresiou a dekompresiou prostredia.

4 P. odsek 3.3.2 *Fourierova transformácia, kosínusová transformácia*, str. 51.

vlnenia (ekvivalentným termínom je **kmitočet**) a vyjadruje sa v Hertzoch ($\text{Hz} = 1/\text{s}$), teda počte vln za sekundu. Zmenu tlaku $P(t)$ oproti okolitému tlaku na danom mieste v priestore (v ďalšom texte považujeme za médium na šírenie zvuku implicitne vzduch) v závislosti od času t možno vyjadriť ako:

$$P(t) = P_0 \sin(2\pi f t + \varphi_0)$$

kde P_0 je **amplitúda vlnenia** (najväčšia veľkosť zmeny tlaku oproti okolitému tlaku na danom mieste⁵), f je frekvencia vlnenia a φ_0 je počiatočná fáza.

Vlna sa šíri v priestore, teda možno zmerať vzdialenosť medzi dvoma najbližšími bodmi s najvyššou (resp. najnižšou) zmenou tlaku. Táto vzdialenosť – vlnová dĺžka λ – je zjavne rovná

$$\lambda = \frac{c}{f}$$

kde c je **rýchlosť** šírenia **zvuku** vo vzduchu. Táto rýchlosť pri štandardných podmienkach ($20\text{ }^\circ\text{C}$) predstavuje 343 m/s . Treba poznamenať, že na rýchlosť zvuku, ktorá sa odvíja aj od rýchlosti častíc vzduchu (pozri odstavec o intenzite zvuku nižšie), nemá vplyv tlak vzduchu ani frekvencia vlnenia. Vzduch je pre človekom počuteľné zvuky nedisperzným médium (tzn. šíria sa bez rozptylu). Pre vysoké frekvencie (nad 5 kHz) dochádza pri dlhšej dráhe zvuku k miernemu útlmu (vysokofrekvenčné kmitanie ľahšie stráca energiu).

Zvuk sa šíri všetkými smermi, ak narazí na prekážku, väčšina energie sa odrazí, čiastočne dochádza k pohlteniu (šíreniu zvuku v hmote prekážky, resp. premene akustickej energie na tepelnú energiu). Pomer odrazenej a pohltenej energie závisí od frekvencie a od vlastností povrchu. Odraz zvuku od prekážky spôsobuje prirodzený **dozvuk** (ľudského hlasu, nástroja); na základe času dozvuku je možné odhadnúť rozmery miestnosti.

2.1.2 Hlasitosť zvuku

Keďže zmeny tlaku sú v porovnaní s atmosférickým tlakom zanedbateľné a navyše samotné číslo vyjadrujúce zmenu tlaku nevytvára nič o „hlasitosti“ vlnenia, čo je subjektívne kritérium, ujali sa aj iné spôsoby merania úrovne hlučnosti zvuku. Smerodajná odchýlka⁶ v tlaku vzduchu, ktorého zmena je zapríčinená daným vlnením, sa označuje ako **akustický tlak** a meria sa v Pascaloch. Práve akustický tlak zohráva hlavnú rolu pri vnímaní hlasitosti.

Testy potvrdili, že človek vníma zvuk až od istého (tzv. prahového) akustického tlaku. Pre frekvenciu 1 kHz je tento **prah počuteľnosti** $p_0 = 20,4\text{ }\mu\text{Pa}$ ⁷. Táto hodnota bola zvolená ako štandard; skutočný, tzv. **absolútny prah počuteľnosti** sa od neho

5 Alebo ináč: polovica rozdielu tlaku medzi najväčšou kompresiou a dekompresiou.

6 Pri meraní v krátkom časovom okamihu je fakticky zhodná so samotnou zmenou tlaku vzduchu.

7 SIBBALD, Alastair (Sensaura). An Introduction to Sound and Hearing. 2001. s. 4.

môže líšiť (v závislosti od kmitočtu vlnenia a veku poslucháča) a venujeme sa mu v odseku 2.2.2 *Vnímanie hlasitosti* (str. 13). Akustický tlak s rastúcou vzdialenosťou od zdroja zvuku klesá nepriamo úmerne.

Intenzita zvuku je potom množstvo zvukovej energie, ktorá prejde danou plochou za jednotku času (W/m^2). Možno ju vypočítať ako súčin akustického tlaku a rýchlosti častíc vzduchu pri šírení zvuku. Rýchlosť častíc je logicky závislá aj od akustického tlaku, mierny vplyv na ňu má dokonca teplota (so vzrastajúcou teplotou mierne rastie rýchlosť) a vlhkosť vzduchu. Intenzita zvuku teda klesá so štvorcom vzdialenosti od zdroja zvuku. Celkové množstvo energie vo Wattoch, ktoré daný zdroj vyžiari do okolia, sa nazýva **akustický výkon**.

Hladina akustického tlaku (anglický termín *sound pressure level*, SPL) je potom logaritmickým vyjadrením „úrovne hluku“ skúmaného zvuku v závislosti od prahového akustického tlaku p_0 podľa vzťahu

$$L_p = 20 \log\left(\frac{p_1}{p_0}\right),$$

kde p_1 je akustický tlak zvuku. Meriame ju v desiatkach belov (dB, decibel, správne sa označuje dB(SPL), keďže dB nie je jednotka sama osebe, ale vždy sa vzťahuje na nejakú referenčnú hodnotu, akou je p_0 v tomto prípade). Opodstatnenie logaritmického merania hlasitosti je namieste: človek zachytí veľký rozsah akustických tlakov, preto by číselné vyjadrenie lineárnymi jednotkami (napr. akustický tlak p) viedlo k príliš dlhému číselnému zápisu (príliš veľké alebo príliš malé čísla).

Podobne meriame aj **hladinu intenzity zvuku** podľa vzťahu

$$L_I = 10 \log\left(\frac{I}{I_0}\right),$$

kde $I_0 = 10^{-12} \text{ W}/\text{m}^2$ je **prahová intenzita** zvuku. Meriame ju v dB (používané označenie dB(SIL)).

Pre grafické zobrazovanie zvuku, teda zmeny tlaku vzduchu na danom mieste v závislosti od času, sa ujalo zakresľovanie do grafu (**vlnový diagram**), v ktorom os x predstavuje postupujúci čas v kladnom význame, os y vyjadruje zmenu tlaku, kde kladné hodnoty predstavujú stlačenie molekúl vzduchu (zvýšenie tlaku), záporné hodnoty zas ich roztiahnutie (zníženie tlaku).

Bližší popis subjektívneho vnímania hlasitosti uvádzame v nasledujúcej časti.

2.2 Psychoakustika

Zvuk má mnoho objektívne merateľných vlastností, tie však nemusia podávať dobrý obraz o tom, čo človek ozaj zachytí sluchom. Vzduch je spojité prostredie, teda jeho molekuly môžu teoreticky prenášať ľubovoľnú zmes frekvencií, každú s osobitou amplitúdou a fázou. Faktom však je, že nie všetko je pre ľudské ucho dôležité, a preto to nie je potrebné zaznamenávať. Je napríklad všeobecne známou skutočnosť, že človek zachytí zvuky od 20 Hz (nižšie kmitočty pri dostatočnej hlasitosti vníma skôr v podobe vibrácie tela) do 20 kHz (táto hranica klesá s dospievaním a starnutím na asi 16 kHz).

To je len jeden zo záverov **psychoakustiky**, teda náuky o vnímaní zvuku človekom, o ktorej si bližšie porozprávame v tejto časti práce. Obsahuje šesť odsekov: 2.2.1 *Fyziológia ľudského ucha* (str. 12), 2.2.2 *Vnímanie hlasitosti* (str. 13), 2.2.3 *Vnímanie kvality, harmonické tóny* (str. 15), 2.2.4 *Maskovanie zvukov* (str. 17), 2.2.5 *Vnímanie frekvencie* (str. 18) a 2.2.6 *Hudobný vnem* (str. 19).

2.2.1 Fyziológia ľudského ucha

Ucho sa skladá z vonkajšieho, stredného a vnútorného ucha. Vonkajšie ucho má dve časti - ušnicu a zvukovod. Zvukové vlny zo zdroja zvuku sú sústredované ušnicou do zvukovodu. Jej nesymetrický tvar spôsobuje, že zvukové vlny prichádzajúce z rôznych smerov sú mierne odlišné (odraz zvuku od ľubovoľného povrchu, najmä mäkkého, ho pozmeňuje vo fáze, amplitúde i frekvenčnom spektre), čo pomáha pri lokalizácii zdroja (p. časť 2.5 *Priestor vo zvuku*, str. 30). Zvukové vlny pokračujú zvukovodom, jednoduchým valcovým otvorom (ten však mierne zosilňuje zvuky v rozsahu 3 až 12 kHz) až k bubienku.

Bubienok tvorí hranicu medzi vonkajším uchom a stredným uchom. Aj keď sa v strednom uchu tiež nachádza vzduch, zvuk už prúdi prostredníctvom kostičiek, pomenovaných kladivko, nákovka a strmienok. Tie privádzajú vibrácie vo forme tlakových vln na oválne okienko.

Oválne okienko je vstupnou bránou do vnútorného ucha, teda najmä slimáka (kochlea), v ktorom sa už nachádza nie vzduch, ale zmes tekutín (endolymfa). Dno slimáka tvorí **bazilárna membrána**, na ktorej je uložený tzv. Cortiho orgán. Na ňom sa nachádzajú vláskové bunky. Bazilárna membrána reaguje na chvenie vnútornej tekutiny rozkmitaním na danom mieste membrány (v závislosti od frekvencie). Na tom mieste dochádza k vzájomnému pohybu vláskových buniek, čo spôsobuje zmenu ich elektrického potenciálu; chvenie vnútornej tekutiny teda premieňajú na elektrické impulzy, ktoré spracúva mozog – a vzniká sluchový vnem.

2.2.2 Vnímanie hlasitosti

V predošlej časti práce (2.1) sme zaviedli logaritmický spôsob merania hladiny akustického tlaku (aj hladiny intenzity zvuku – p. odsek 2.1.2 *Hlasitosť zvuku*, str. 10), teraz je načase vysvetliť jeho opodstatnenie. Ľudské ucho je úžasne citlivý orgán. Zachytí už veľmi tiché zvuky, ale vie si poradiť i so zvukmi vysokej hlasitosti. Pomer akustického tlaku medzi týmito druhmi zvukov môže byť až jeden k miliónu (a keďže intenzita zvuku sa v porovnaní s akustickým tlakom správa kvadraticky, je v jej prípade pomer medzi najtichším a najhlučnejším bezbolestne počuteľným zvukom až jeden k biliónu). Vtedy nám príde vhod spomenutý logaritmický výpočet, ktorý dáva rozsah 120 dB (v oboch prípadoch). Takýmto spôsobom môžeme zachytiť malými číslami veľký rozsah, pričom na začiatku škály je citlivosť (presnosť) vysoká, na vyšších číslach zas nižšia. Logaritmické vyjadrenie akustickej úrovne umožňuje lepšie porovnanie hlasitosti dvoch signálov. Ak vzrastie hlasitosť zvuku napr. z 25 dB na 35 dB, človeku sa ten nárast subjektívne zdá porovnateľný s nárastom zo 70 dB na 80 dB. Prevláda názor, že nárast o 10 dB je subjektívne vnímaný ako „zdvojnásobenie hlučnosti“ (čo sa, samozrejme, len ťažko dá kvantifikovať). Ak rátame ako referenčný akustický tlak prah počuteľnosti, môžeme si urobiť obraz o jednotlivých úrovniach hlasitosti na základe tabuľky 2.1⁸.

L_p [dB(SPL)]	p [Pa]	zdroj zvuku, udalosť s danou hlasitosťou
0	0,000 020	<i>prah počuteľnosti (zdravého jedinca, 1 kHz)</i>
10	0,000 063	ľudský dych zo vzdialenosti 3 metrov
30	0,000 632	ticho v divadle, noc na púšti
40	0,002 000	ticho na sídlisku v noci, ľudský šepot
50	0,006 325	tichá kaviareň
60	0,020	zvyčajná hlasitosť normálnej konverzácie
80	0,200	vysávač zo vzdialenosti 1 meter
90	0,632	nákladné auto zo vzdialenosti 1 meter
100	2	diskotéka, televízor alebo veža pri najvyššej hlasitosti
120	20	rockový koncert, vzlet prúdového lietadla (100 m) telo cíti vibrácie pri nízkych frekvenciách
130	63	vojenský dychový orchester zblízka <i>prah bolesti pre všetky frekvencie</i>
150	632	prúdový motor (30 m) bolesť v hrudníku, rozmazané videnie <i>poškodzovanie sluchu pri všetkých frekvenciách</i>
180	20 000	motor rakety (30 m), dunenie vráskavca ozrutného (1 m) Výbuch sopky Krakatau (r. 1883) zo vzdialenosti 160 km!

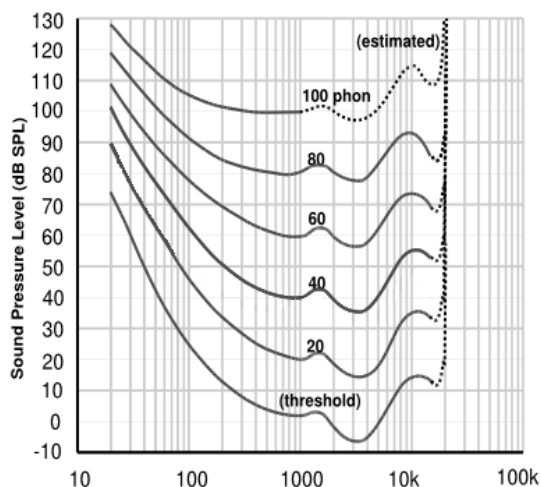
Tab. 2.1 – Porovnanie hladiny akustického tlaku (L_p) s akustickým tlakom (p), ktoré generujú rôzne zdroje zvuku

8 Tabuľka 2.1 je skompilovaná z rôznych zdrojov.

V laboratórnych podmienkach je človek schopný rozlíšiť rozdiel v hlasitosti pre tón danej frekvencie už od 1 dB, pre všeobecné zvuky a prostredie je táto úroveň 3 dB a viac.

Keďže však človek nevníma rôzne frekvencie pri zhodnom akustickom tlaku rovnako, Medzinárodná elektrotechnická komisia (*International Electrotechnical Commission*, IEC) definuje aj tzv. **A-váhovanie** (angl. *A-weighting*), ktoré na základe štatistických prieskumov upravuje zmeranú hladinu akustického tlaku v závislosti od frekvencie zdroja. Ľudské ucho je totiž najcitlivejšie na frekvencie v rozsahu 3 kHz až 6 kHz, pričom pre frekvencie nad 20 kHz a pod 20 Hz nastáva úplný útlm (ľudský sluch ich nezachytí). Toto váhovanie sa opiera o výsledky zostavené Fletcherom a Munsonom z Bell Laboratories v r. 1933 (ich výsledky boli neskôr – v r. 2003 – upresnené normou ISO 226:2003), ktorí krivkami (**izofónami**) vyjadrili subjektívne vnímanie rovnakej úrovne hlasitosti (obr. 2.1⁹) v závislosti od frekvencie (os x) a hladiny akustického tlaku (os y) pre referenčný kmitočet 1 kHz. Zatiaľ čo pre 1 kHz je prahom počuteľnosti akustický tlak p_0 (teda 0 dB), pre 20 Hz je to vyše 3000-krát vyššia hodnota (70 dB). Najnižšia krivka teda vyjadruje **absolútny prah počuteľnosti**, ktorý je závislý od frekvencie. Najvyššia krivka je zas **absolútny prah bolesti**. (Pri vysokej hlasitosti dochádza k tzv. akustickému reflexu, keď svaly vo vnútornom uchu odtiahnu kladivko od bubienka a strmienok od slimáka, aby lepšie chránili sluch pred poškodením.)

A-váhovanie je potom prenosová funkcia totožná s krivkou pre 1 kHz a 40 dB. Takto vyjadrená hlasitosť sa kedysi merala v jednotkách **fón** (phon), v súčasnej dobe sa o nej skôr hovorí ako o A-váhovanej hlasitosti, ktorú možno označiť aj $L_A = x \text{ dB(SPL)}$, čo vyjadruje použitie referenčnej hodnoty p_0 , predsa však pri aplikácii frekvenčného filtra¹⁰ A-váhovania.¹¹



Obr. 2.1 - Krivky rovnakej subjektívnej hlasitosti v závislosti od frekvencie tónu

Treba poznamenať, že výsledky Fletchera a Munsona boli namerané použitím slúchadiel, ktoré, nasadené na ušiach, takmer bez skreslenia prenášajú až na bubienok ucha nízke frekvencie, zatiaľ čo pri vysokých frekvenciách vznikajú v malých priestoroch medzi reproduktorom a zvukovodom rôzne ozveny a rezonancie. Podobný výskum ako Fletcher a Munson robili v r. 1956 Robinson a Dadson, ale použili jeden

9 Autor obrázku: SKIRROW, Peter. Zdroj [cit. 2006-02-14]: <http://upload.wikimedia.org/wikipedia/en/thumb/4/47/Lindos1.svg/400px-Lindos1.svg.png>.

10 P. odsek 2.3.4 *Úprava a zmena zvuku*, str. 24.

11 Ešte presnejšou jednotkou môže byť 1 són (angl. sone, navrhnutý americkým psychológom menom Stanley Smith Stevens), definovaný ako hlasitosť 1 kHz tónu pri 40 dB(SPL). Zvýšenie váhovanej hlasitosti o 10 dB znamená zdvojnásobenie hodnoty – napr. tón frekvencie 1 kHz pri 70 dB(SPL) má hlasitosť 8 sónov.

reproduktor umiestnený pred poslucháčom – výsledky boli aj preto trochu odlišné, aj keď ich hodnovernosť je tak či tak často spochybňovaná, keďže významné rozdiely oproti Fletcherovi a Munsonovi nemožno vysvetliť len spôsobom reprodukcie. Odlišné výsledky by zrejme boli namerané aj pri použití dvojice reproduktorov.

Práca Fletchera a Munsona navyše rátala len s jednoduchým (čistým) tónom, nie so šumom alebo inými zvukmi. V šesťdesiatych rokoch 20. storočia urobilo preto rozsiahlejší prieskum výskumné centrum BBC. Na základe ich výsledkov zostavil Rádiokomunikačný sektor Medzinárodnej telekomunikačnej únie (ITU-R) štandard **váhovania ITU-R 468**. Táto prenosová funkcia zosilňuje najmä pásmo okolo 6 kHz, keďže na šum s touto frekvenciou je ľudský sluch najcitlivejší. Vo všeobecnosti lepšie vyjadruje subjektívny vnem hlasitosti pre ľubovoľné zvuky, je však na elektronickej úrovni ťažšie konštruovateľný (čo v súčasnej digitálnej ére už nie je podstatné). Ujal sa najmä v Európe, predovšetkým v profesionálnej sfére, zatiaľ čo v USA ostáva rozšírené „nedokonalé“ A-váhovanie, vhodné skôr pre tiché zvuky (40 dB). Nezabudnime však na fakt, že všetky tieto váhovania sú len štatistickou aproximáciou, a aj keď vnímanie hlasitosti je pre ľudí podobné, predsa je pre každého z nich jedinečné.

2.2.3 Vnímanie kvality, harmonické tóny

Aby sme vôbec mohli zvažovať kvalitu zvuku, jeho prenosu či reprodukcie, musíme sa podrobnejšie oboznámiť s rozšírenými spôsobmi výpočtu skreslenia zvuku. Vec komplikuje fakt, že na to, čo chápeme ako „kvalitné“, neexistuje žiadne skutočne objektívne meradlo. Matematický aparát preto môžeme použiť len ako „barličku“.

V ideálnom prípade má zariadenie prenášajúce alebo spracujúce zvuk (zväčša vo forme elektrického prúdu a napätia) lineárnu **prenosovú funkciu**¹², ktorá je navyše časovo invariantná (nemenná). Miera skreslenia zvuku indikuje, do akej miery sa skutočná prenosová funkcia líši od ideálnej. Začneme však niekoľkými potrebnými pojmami, ktoré treba objasniť.

Základný tón (angl. *fundamental tone*) je nosný tón zvuku (zväčša hlasu, hudobného nástroja), teda tón, ktorý pri počutí zvuku subjektívne vnímame ako hlavný. Táto základná frekvencia je zároveň najnižšia frekvencia vo zvuku, pričom ostatné prítomné frekvencie sú jej násobkami. Táto charakteristika platí pre väčšinu prirodzených zdrojov zvuku, ako bližšie objasníme v odseku 2.3.2 *Zvuky hudobných nástrojov* (str. 22).

Frekvencie prítomné vo zvuku, ktoré sú násobkami základnej frekvencie, sa nazývajú **alikvotné tóny** (angl. *overtone*). Ak je tento násobok celočíselný, ide o **harmonický tón**. Prvý harmonický tón je zhodný so základným tónom, druhý

¹² Prenosová funkcia je funkcia vyjadrujúca rozdiel medzi vstupným a výstupným signálom.

harmonický (alebo prvý alikvotný) tón má dvojnásobnú frekvenciu, tretí harmonický (druhý alikvotný) tón má trojnásobnú frekvenciu atď. Ak násobok nie je celočíselný, ide o tzv. **čiasťkový tón**.

Vyššie harmonické tóny (2, 3, ...) vznikajú v prostredí alebo pri prenose elektrických signálov často samovoľne¹³ a ľudské ucho je na ne zvyknuté, dokonca s nimi „ráta“ a na základe ich prítomnosti a frekvencie dedukuje aj frekvenciu základného tónu¹⁴, aj keď tento ani nemusí byť vo zvuku prítomný. Vtedy hovoríme o **chýbajúcom základnom tóne**.

Celkové harmonické skreslenie (angl. *total harmonic distortion*, **THD**) potom vyjadruje skreslenie, ktoré nastane neželaným vznikom vyšších harmonických tónov. Je to podiel súčtu ich energií P_2, P_3, \dots, P_n a energie základného tónu P_1 , teda:

$$\text{THD} = \frac{P_2 + P_3 + \dots + P_n}{P_1}.$$

Keďže energia rastie kvadraticky s akustickým tlakom, môžeme ju vyjadriť aj priemerným napätím, ktoré bolo zaznamenané, resp. použité pri reprodukcii uvedených frekvenčných zložiek:

$$\text{THD} = \frac{\sqrt{V_2^2 + V_3^2 + \dots + V_n^2}}{V_1}.$$

Skreslenie sa zväčša vyjadruje priamo ako THD (v percentách), alebo po aplikovaní desiatkového logaritmu v dB. Počet vyšších harmonických tónov zahrnutých do výpočtu (n) určuje presnosť merania.

Možno sa stretnúť tiež s jednotkou THD+N, čo je skreslenie vrátane šumu (angl. *noise*). Počíta sa ako podiel súčtu energií vyšších harmonických frekvencií a šumu (teda celkovej energie prijatého zvuku po odfiltrovaní tónu, ktorý sme poslali na vstup) a energie tónu na vstupe. Navyše, všetky neželané zložky výstupu možno jednoducho pokladať za šum – hovoríme o **odstupe signálu od šumu** (angl. *signal-to-noise ratio*, **SNR**).

Aby sme však boli objektívni, musíme myslieť aj na subjektívnu stránku, veď tieto miery boli navrhnuté pre ľudí. Faktom je, že vyjadriť skreslenie ako jedno číslo je nedostačujúce. Pri jednom počuteľnom „tuku“ v tichej pasáži zvuku a vernom prenose zvyšku signálu môže vyjsť pri výpočte THD+N nízke číslo, zatiaľ čo pri signáli s nepočuteľným nepretržitým šumom vyjde číslo vysoké. Podobne, symetrická nelinearita prenosového kanála zapríčiňuje vznik nepárnych harmonických tónov, ktoré sú sluchu príjemnejšie než párne, spôsobené nesymetrickou nelinearitou. Vznikli preto aj iné testy (napr. od Európskej rozhlasovej únie, alebo od Lindos Electronics), ktoré zahŕňajú sekvenciu testovacích vzoriek a metodiku ich vyhodnocovania. Ich popis však zďaleka presahuje rámec tejto práce.

¹³ Viac v odseku 2.3.2 *Zvuky hudobných nástrojov*, str. 22.

¹⁴ P. tiež odsek 2.4.3 *Reprodukcia zvuku*, str. 27.

Otázna je navyše potreba „dokonalého zvuku“. Samozrejme, často možno odlíšiť zvuk kvalitnejší od menej kvalitného. (Stačí, ak k zvukovej karte počítača, alebo inému elektronickému zdroju zvuku pripojíme najprv malé reproduktory a potom kvalitnejšiu reproduktorovú sústavu; rozdiel je jasne počuteľný, druhé riešenie označíme ako „kvalitný zvuk“ v porovnaní s prvým; ten zas bude „nekvalitný“.) Ale existuje niečo ako „dokonalý zvuk“? Túto otázku budeme ešte zvažovať v časti 2.4 *Analógové spracovanie zvuku* (s. 26). Tu stačí spomenúť, že napr. na prenos hovoreného slova v telefónnom signáli sa používa frekvenčný rozsah 400 Hz až 3 400 Hz. Možno sa to zdá málo, ale ľudskému sluchu to spokojne stačí na porozumenie hovoreného slova.

2.2.4 Maskovanie zvukov

V úplnom tichu človek začuje aj pád ihly na koberec, ak však stojí na hlučnej ulici plnej áut, nepočuje ani zvonenie svojho mobilného telefónu. Tento jav je v psychoakustike známy ako **maskovanie zvukov** (v angličtine sa používa termín *masking*). Zvuk s väčšou inenzitou zhoršuje vnímanie iného prítomného zvuku s nižšou intenzitou, alebo ho úplne pokrýva, takže tichší zvuk ani nepočuť. Ľudský sluch totiž zachytáva zvuk ako súčet všetkých zvukových signálov (frekvencií, amplitúd, fáz), ktoré dorazia na ušný bubienok, a vďaka adaptivite na rôzne úrovne hlasitosti sa mení aj presnosť (rozlíšenie), s ktorou ich zachytí.

Ak sú vo zvuku prítomné dva signály, v ktorom jeden maskuje ten druhý, hovoríme o **frekvenčnom** (alebo simultánnom) **maskovaní**. Či a do akej miery nastáva jav maskovania, závisí od frekvencie (tóny s blízkou frekvenciou sú náchylné na jav maskovania; prítomnosť veľmi odlišných frekvencií, ak nejde o harmonické násobky, je ľahko počuteľná), tvaru signálu (čistý sínusový signál ťažko pokrýva zvuk podobný šumu, zvuk podobný šumu ľahko pokrýva nevýraznú sinusoidu) a, samozrejme, pomeru hlasitostí maskovacieho a maskovaného signálu. Dôvodom je najmä konštrukcia ľudského ucha – rozkmitanie bazilárnej membrány¹⁵ na jednom mieste (v dôsledku dominantnej frekvenčnej zložky zvuku) spôsobuje aj jej chvenie v blízkom okolí. To zapríčiňuje fyziologickú detekciu tónov s frekvenciami blízkymi dominantnej, ktoré však mozog ignoruje (keďže v pôvodnom zvuku nie sú prítomné). Na dodanie tónov kmitočtom blízkych (v rámci tzv. **kritického pásma**) dominantnému tónu je preto poslucháč len slabo citlivý¹⁶. Jednou z hojne používaných mierok tejto tolerancie je **Barkova mierka** (angl. *Bark scale*), deliaca počuteľné spektrum na 24 kritických pásiem¹⁷.

15 P. odsek 2.2.1 *Fyziológia ľudského ucha*, str. 12.

16 Alebo tieto pridané tóny nezachytí vôbec. Samozrejme, hovoríme o „čistých tónoch“ (sínusových).

17 Hraničné frekvencie sú 0, 100, 200, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4400, 5300, 6400, 7700, 9500, 12000, 15500. Zdroj: http://en.wikipedia.org/wiki/Bark_scale.

Špeciálny prípad javu maskovania je aj **časové maskovanie** (angl. *temporal masking*). Sluch nezachytí veľmi tichý zvuk znejúci po hlasnom zvuku (do času najviac 50 ms, tento jav sa nazýva **dopredné maskovanie**, angl. *forward masking, post-masking*), dokonca ani tichý zvuk znejúci tesne pred hlasným signálom (asi 10 ms; **spätne maskovanie**, angl. *backward masking, pre-masking*). Na vine je oneskorená adaptácia sluchu na rôzne hlasitosti, zapríčinená najmä chvením bazilárnej membrány.

2.2.5 Vnímanie frekvencie

Jednou z objektívnych mierok zvuku je kmitočet. Tu by však bolo vhodnejšie hovoriť o „výške tónu“, keďže práve tá odzrkadľuje subjektívny vnem poslucháča. Faktom však je, že „výšku“ počutého zvuku nemožno dobre kvantifikovať. Ovplyvňuje ju najmä frekvenčný rozklad zvuku, ale aj jeho hlasitosť a trvanie. Strojovo sme však schopní spracúvať len merateľné vlastnosti zvuku – rozhodujúcim atribútom je spomenutý frekvenčný rozklad zvuku, najmä základná frekvencia.

Človek však nárast či pokles vo frekvencii tónu vníma logaritmicky – nepočujeme skutočnú frekvenciu tónu (lineárne), ale jej logaritmus¹⁸. Ak pomer medzi frekvenciami dvoch tónov je zhodný s pomerom medzi frekvenciami iných dvoch tónov, frekvenčný nárast alebo pokles ľudskému sluchu pripadá rovnaký. Tento princíp odzrkadľovali mnohé systémy ladenia a radenia tónov, vrátane pytagorejského ladenia, ktoré bolo neskôr v európskej kultúre úplne nahradené súčasným **rovnomerne temperovaným ladením**. Tón o oktávu vyššie než referenčný tón má dvakrát vyššiu frekvenciu (teda je jeho prvým alikvotným tónom) – charakteristikou akoby išlo o ten istý tón, len s frekvenciou o „stupeň“ vyššou. Oktáva sa navyše rovnomerne delí na 12 poltónov, čo znamená, že každý ďalší poltón má frekvenciu oproti predošlému o $\sqrt[12]{2}$ vyššiu. Tón o oktávu vyššie ($(\sqrt[12]{2})^{12}=2$) je teda druhý harmonický tón, ďalšia oktáva znamená štvornásobnú (2×2) frekvenciu, ďalšia osemnásobnú (ôsmy harmonický tón). (Teda „výška“ – ako subjektívne vnímanie frekvencie – harmonických tónov je stále bližšie, hustejšie pri sebe.) Podľa normy má tón a^1 frekvenciu 440 Hz.

Ešte presnejšie možno **frekvenčný interval**, teda pomer frekvencií, deliť na centy – 1 cent = $\sqrt[1200]{2}$. Výskumy ukázali, že ľudia s citlivým sluchom sú schopní zachytiť rozdiel v dvoch tónoch s intervalom už od 5 centov vyššie¹⁹. Človek sluchom

18 Ide o Weberov-Fechnerov psychologický zákon – p. napr. HALUŠKA, Ján. The Mathematical Theory of Tone Systems. 2004. s. 5. Pre muzikálny vnem však nie je dôležitá absolútna hodnota frekvencie alebo jeho logaritmu, ale najmä kontext ostatných tónov; triedu tónov človek vníma ako vektorový priestor (tamže). Tiež treba spomenúť, že logaritmický vnem sa uplatňuje najmä pre frekvencie nad 800 Hz. Na začiatku frekvenčnej škály, pri nižších frekvenciách (najmä pod 500 Hz) vníma poslucháč nárast frekvencie takmer ako lineárny nárast výšky tónu.

19 Tamže. Treba však uviesť, že v prípade hudobného vnemu je poslucháč omnoho tolerantnejší – akceptuje intervaly až do 20-30 centov (p. tamže, Garbuzovove zóny, s. 57).

vedome vníma len frekvenciu zvuku, nie fázu, aj keď fáza vlnenia je tiež spracovaná mozgom a aj na jej základe dokáže poslucháč lokalizovať zdroj zvuku.

Pri súčasnom posluhu dvoch tónov však sme schopní zistiť rozdielnosť ich frekvencií už pri omnoho nižších rozdieloch vďaka **záznejom** (angl. *beat*), ktoré vznikajú interferenciou dvoch vlnení s veľmi blízkou frekvenciou. Výsledný tón ponesie akoby jednoliatu frekvenciu (tóny od seba neodlíšime, zachytíme priemer ich kmitočtu), ale jeho intenzita (hlasitosť) bude kmitať na frekvencii zhodnej s rozdielom frekvencií pôvodných tónov. Tento jav je badateľný od rozdielu niekoľkých desiatín Hz až po 15 Hz. Samozrejme, interferencia vznikne aj pri použití alikvotných tónov miesto základného tónu. Zázneje sa využívajú napr. aj pri ladení gitár.

Ľudský sluch má dobrú citlivosť na čas zachytenia vysokých tónov, v prípade tónov s nízkofrekvenčnou charakteristikou je vnímanie času zachytenia tónu menej presné. Naopak, frekvenčné rozlíšenie nízkych tónov v dôsledku uvedeného logaritmickeho vnímania je omnoho presnejšie než v prípade vysokých tónov.

2.2.6 Hudobný vnem

Hudba v tom najširšom zmysle je ťažko definovateľná, ale medzi jej základné charakteristiky patria rytmus, tempo, melódia, príp. harmónia. Pre nás bude podstatná len harmónia tónov. Harmonické tóny vznikajú v hudobných nástrojoch samovoľne²⁰ a ľudský sluch v nich vidí súzvuk. Vezmime si ako príklad tón s frekvenciou 440 Hz (komorné *a* – v medzinárodnom značení tón A4). Druhý harmonický tón má kmitočet 880 Hz (a^2), tretí harmonický tón 1 320 Hz. Frekvencia e^3 má v rovnomerne temperovanom ladení kmitočet 1 318,5 Hz, čo je takmer totožný tón. Aj v dôsledku toho tóny *a* a *e*, ak sú vo zvuku prítomné aj ich harmonické kmitočty, znejú poslucháčovi akosi „zladene“, keďže zdieľajú spoločné harmonické frekvencie.

Pomer frekvencií tónov e^3 a a^2 je teda v ideálnom prípade 3:2. Podobnými úvahami možno získať i ideálny pomer d^3 a a^2 (4:3), ktorý by vyžadoval od d^3 frekvenciu 1 173 Hz (v súčasnom rovnomerne temperovanom ladení je to 1 175 Hz) a ostatné tóny oktávy, všetky s jednoduchým racionálnym pomerom k základnému tónu. Zdá sa teda, že by pytagorejské ladenie mohlo byť výhodnejšie. Opak je však pravdou – možno s ním vytvoriť úžasný súzvuk v jednom akorde, z pohľadu iného akordu (ak by sme začínali trebárs od tónu *h*) by miesto „pekných“ malých pomerov vychádzali podiely dvoch vysokých čísel, ktoré by vytvárali disharmonický zvuk. Súčasné rovnomerne temperované ladenie je teda nielen vhodnou matematickou konštrukciou, ale aj praktickou výhodou pre všetkých hudobníkov (umožňuje ľahko transponovať hudobný zápis o ľubovoľný počet poltónov). Nevýhodou je disharmónia, ktorá je však príliš slabá (rozdiely do 1% frekvencie), aby ju ľudské ucho vnímalo natoľko kriticky. Tá spôsobuje (z pohľadu následnej frekvenčnej

²⁰ P. odsek 2.3.2 *Zvuky hudobných nástrojov*, str. 22.

analýzy, potrebnej pri stratovej kompresii zvuku) „zbytočné“ obohatenie frekvenčného spektra hranej hudby (ktoré ináč pri väčšine súčasnej hudby nie je veľmi bohaté).

2.3 Charakteristiky zvuku

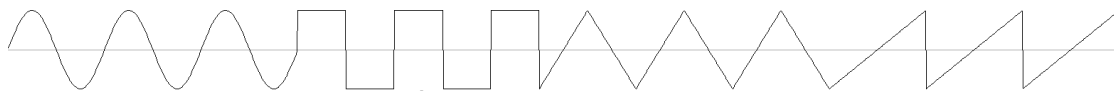
V tejto časti práce rozoberieme oproti časti 2.1 (*Zvuk ako fyzikálny fenomén*, str. 9) vlastnosti zvuku trochu podrobnejšie a z iného pohľadu. Prenesieme sa ponad jednoduché stavebné prvky zvuku, aby sme ich vnímali komplexnejšie – ako súzvuk tónov, ktoré môžeme vytvárať (hlasom, nástrojmi, ...) a zachytávať. Spomenieme tiež rozličné javy, ktoré môžu pri práci so zvukom nastať.

Táto časť práce obsahuje štyri odseky: 2.3.1 *Druhy zvukov* (str. 21), 2.3.2 *Zvuky hudobných nástrojov* (str. 22), 2.3.3 *Ludský hlas* (str. 23) a 2.3.4 *Úprava a zmena zvuku* (str. 24).

2.3.1 Druhy zvukov

Ako sme si spomenuli v časti 2.1, zvuk ako vlnenie má vlastnosti, ktorými ho možno charakterizovať. Jednou z týchto vlastností je frekvencia, ktorá vyjadruje, ako často sa striedajú jednotlivé deje zvukovej vlny. Zvuková vlna (tzv. čistého tónu) má teda tvar **sinusoidy**. Ak máme napríklad v prostredí dve zariadenia generujúce nejaké čisté tóny, obe vplývajú na pohyb molekúl vzduchu, teda ich vplyv na zmenu tlaku vzduchu v ľubovoľnom bode merania sa sčítuje. Pri zachytávaní (napr. meracím prístrojom) dvoch „čistých tónov“ by teda šlo o súčet dvoch sínusových funkcií. Takýto rozbor kmitočtových zložiek zvuku sa tiež nazýva **frekvenčná charakteristika** (alebo **frekvenčný rozklad**, **analýza**) zvuku.

V praxi sa používajú aj iné tvary zvukových vln. Medzi najrozšírenejšie patrí tzv. **štvorcová vlna**, kde sa strieda dej najväčšieho stlačenia a najväčšieho roztiahnutia (ide o nespojitú funkciu, teda v reálnom svete ju možno len aproximovať), **trojuholníková vlna**, v ktorej sa strieda (vzhľadom na čas) lineárna kompresia a lineárna dekompresia vzduchu a **pílová vlna**, v ktorej po lineárnej kompresii dochádza k okamžitému zníženiu tlaku na minimum a ďalšej následnej kompresii (tiež nespojitá funkcia). Frekvenčné charakteristiky trojuholníkovej i štvorcovej vlny vykazujú prítomnosť nepárnych harmonických tónov²¹, pílová vlna obsahuje všetky svoje harmonické zložky.



Obr. 2.2 – Rôzne tvary zvukových vln: sinusoida, štvorcová vlna, trojuholníková vlna a pílová vlna

Celou kategóriou zvukov sú **šumy**. Vo všeobecnosti ide o neusporiadanú zmenu tlaku vzduchu, ktorá nie je periodická. Šum je často neželaný a vzniká samovoľne nielen v každom fyzikálnom prostredí (molekuly vzduchu sa v dôsledku tepelnej energie neusporiadane hýbu), ale aj v každej fáze manipulácie so zvukom²² (záznam

21 P. odsek 2.2.3 *Vnímanie kvality, harmonické tóny*, str. 15.

22 Hovoríme o zvuku v analógovej podobe.

zvuku, prenos elektrického signálu, jeho spracovanie... – kvantové javy, ktoré sa prejavajú ako zvuk nízkej intenzity; následným spracúvaním môže byť intenzita zosilnená). V závislosti od príčin vzniku má šum zväčša nejaký charakteristický frekvenčný rozklad. Ak sú dominujúce frekvencie vysoké, ide o vysokofrekvenčný šum (angl. *hiss*), ak nízke, ide o nízkofrekvenčný šum (občas „brum“, angl. *hum*; najznámejší príklad je 50 Hz brum zapríčinený neuzemneným nulákom v elektrickom napájaní zariadení).

Špeciálnym prípadom šumu je tzv. **biely šum**. Ide o úplne náhodnú zmenu tlaku v čase (signál bez autokorelácie), ktorá, ak rátame s neobmedzene spojitým priestorom, má navyše zhodne vysokú energiu vo všetkých frekvenčných pásmach (analógia s bielym svetlom), teda teoreticky má nekonečnú energiu. Vzhľadom na to, že v praxi sa stále pracuje s obmedzeným frekvenčným pásmom, aj frekvenčné spektrum bieleho šumu je obmedzené. Ďalším špeciálnym prípadom šumu je tzv. **ružový šum**, ktorý sa od bieleho šumu líši logaritmickým poklesom energie vzhľadom na frekvenciu, pričom úbytok je 3 dB na jednu oktávu (teda každú oktávu klesá energia danej frekvencie na polovicu²³ – teda napr. medzi kmitočtami 600 a 800 Hz je vo frekvenčnom rozklade rovnaké množstvo energie ako medzi 6000 a 8000 Hz). Spomenúť možno ešte tzv. **sivý šum**, ktorý má frekvenčné spektrum také, aby každá frekvencia bola v signáli prítomná s rovnakou psychoakustickou hlasitosťou²⁴.

2.3.2 Zvuky hudobných nástrojov

Hudobných nástrojov existuje nepreberné množstvo a delia sa do rôznych skupín (strunové, sláčikové, kovové dychové, ...), no nie je tu priestor, aby sme ich jednotlivito opísali. Pokúsime sa venovať vlastnostiam, ktoré majú podobné alebo totožné.

Keďže zvuk je vlnenie (chvenie vzduchu), musí toto vlnenie vzniknúť nejakým spôsobom vo vnútri alebo na povrchu nástroja (zväčša oba prípady vplyvajú na celkový zvuk nástroja). Typickým príkladom prvého je flauta. Zvuk rezonuje medzi otvorom pri zobci a ústím flauty. Prvotný impulz – stlačenie vzduchu – zabezpečí fúknuť do zobca nástroja. Dĺžka tubusu hudobného nástroja (v ktorého vnútri sa nachádza stĺpec vzduchu) určuje vlnovú dĺžku základného tónu. Otváraním dierok (vďaka konštantnej rýchlosti zvuku) meníme jeho vlnovú dĺžku, teda i frekvenciu. V tubuse vznikajú navyše stojaté vlnenia, ktoré delia jeho dĺžku na polovicu, tretinu, ... – v závislosti od rezonančných vlastností nástroja. V jeho zvukovom prejave sa tak objavujú harmonické frekvencie, ktorých hlasitosť nielen dotvára, ale priam určuje charakteristický zvuk (farbu) toho-ktorého nástroja, vďaka ktorému ho vieme odlišiť od iných. Ak intenzitou prevažujú párne harmonické tóny

²³ $\log 2 = 3,01$ dB.

²⁴ Teda podľa vhodného váhovania – p. odsek 2.2.2 *Vnímanie hlasitosti*, str. 13.

nad nepárnymi, znamená to ostrejší zvuk nástroja (napr. plechové dychové nástroje), zvýšená intenzita nepárnych dodáva zvuku mäkkosť (napr. drevené dychové nástroje). Vo všeobecnosti intenzita s rastúcim harmonickým násobkom klesá.

Medzi inštrumenty, ktorých povrch sa chveje, a tak rozkmitáva vzduch vo svojom okolí, patria medzi iným strunové nástroje, napríklad gitara. Brnknutím do struny sa táto rozozvučí svojou typickou frekvenciou. Keďže ide o kmitanie, vznikajú na strune i stojaté vlnenia, dodávajúce zvuku gitary harmonické tóny. Samozrejme, nielen chvejúca sa struna vytvára akustický vnem gitarového zvuku. Vlnenie zosilnené rezonanciami v ozvučnici sa prenáša sa na celú konštrukciu nástroja, čím prispieva k celkovému dojmu zo zvuku.

Existujú i nástroje s čiastkovými alikvotnými tónmi vo svojom frekvenčnom rozklade – najznámejším príkladom je asi činela, tvoriaca celé spektrum alikvotných tónov od základnej frekvencie (danej priemerom činely a použitým materiálom).

S vyššie popísanými vlastnosťami zvuku však ešte nie je definícia hudobných nástrojov úplná. Pri začiatku hrania (napr. fúknutie do zobca klarinetu) vzniká chvíľkový jav **nábehu** (angl. *attack*), so špecifickou frekvenčnou charakteristikou a rastúcou intenzitou; nasleduje krátky **útlm** intenzity (*decay*), až po ňom prichádza pravidelná frekvenčná charakteristika **držania** zvuku (*sustain, hold*). Pri ukončení hrania (napr. hráč prestane fúkať vzduch) sa objaví krátky zvuk **doznenia** (*release*). Táto schéma funguje prakticky pri všetkých hudobných nástrojoch, aj pri ľudskom hlase.

2.3.3 Ľudský hlas

Hlasivky človeka fungujú ako nezvyčajný hudobný nástroj, ich hlavné využitie je však v medziľudskej komunikácii. Vďaka vnútornej členitosti vzniká v ich vnútri okrem základného tónu i mnoho alikvotných tónov (harmonické i čiastkové). Tým hlas získava „farbu“, vďaka ktorej možno odlíšiť hlas dvoch spevákov spievajúcich rovnaký tón. Hlasivky navyše dokážu meniť svoj tvar a veľkosť a tým upravovať výšku a farbu tónu. Zvuk generovaný hlasivkami prechádza hrtanom, nosohltanom, a ústnou dutinou, ktoré spolu s chrupom a perami menia jeho frekvenčnú charakteristiku (fungujú teda ako akýsi frekvenčný filter – p. odsek nižšie). Charakteristické frekvenčné zložky takto upraveného zvuku (zväčša ešte posilnené rezonančnými vlastnosťami hlasového traktu) sa nazývajú **formanty**. Okrem rôznych základných frekvenčných zložiek však ľudský hlas obsahuje aj šumové charakteristiky (napr. sykavky), impulzné zvuky (vzniknuté turbulenciou vzduchu, ako napr. hlásky „p“, „b“), alebo kombinácie šumu a tónu (hláska „z“).

Prax ukázala, že pre porozumenie záznamu hovoreného slova stačí počet frekvenčný rozsah pribl. 400 Hz až 4 000 Hz. Väčšina energie zo zvukového záznamu hlasu sa nachádza vo frekvenčnom spektre pod 2 000 Hz.

2.3.4 Úprava a zmena zvuku

Zvuk v tej podobe, v akej vzniká, môže málokedy zachytiť poslucháč bez zmeny. Samozrejme, ak sa nachádza blízko zdroja zvuku, počuje všetko v podstate verne. Ak ho však od zdroja delí napr. veľká vzdialenosť alebo prekážka, zvuk sa zmení. My musíme ísť ešte ďalej, keďže zvukové formáty pracujú so zvukom zaznamenaným a následne reprodukoványm. Pri tomto procese nastanú niektoré javy či procesy, ktorým sa nemôžeme (alebo ani nechceme) vyhnúť. V nasledujúcich riadkoch sa s nimi oboznámime.

Orezávanie (angl. *clipping*): Každé záznamové zariadenie má nejaké rozlíšenie v zápise zmien tlaku vzduchu – má najnižšiu a najvyššiu zaznamenateľnú hodnotu. Fyzikálny svet však nepozná obmedzenie na maximálnu zmenu tlaku (môže byť teoreticky ľubovoľne veľká). Ak teda zariadenie spracúva údaj, ktorý je väčší než povolené maximum, nahradí ho maximom. Tak je sinusoida, ktorá sa svojou amplitúdou zmestí do rozlíšenia zariadenia, prenesená verne, zatiaľ čo trochu silnejší záznam toho istého zvuku má orezané **špičkové** (angl. *peak*) **hodnoty** – najväčšie kompresie a dekompresie vzduchu. Aj keď mohla byť zmena v absolútnych číslach minimálna, vznik „parazitných“ frekvencií v zázname, ktoré v pôvodnom zvuku neboli, bude enormný. V praxi sa môžeme stretnúť aj s tzv. mäkkým orezaním, keď od istej úrovne signálu dochádza k jeho značnej kompresii (p. ďalší odstavec). (Treba povedať, že každý zosilňovač, reproduktor, mikrofón má nejaký limit, od ktorého už nastáva orezávanie.)

Často je spomenutý jav orezávania žiaduci. Niektorí gitaroví hráči (napr. v rocku) preferujú zvuk tzv. gitarového skreslenia, ktoré dosiahnu použitím poddimenzovaného zosilňovača a následným upravením frekvenčnej charakteristiky. Mnohí zas používajú elektrónkové zosilňovače, charakteristické mäkkým orezaním (pri preťažení vzniká väčšie harmonické spektrum), nie zosilňovače tranzistorové, ktoré majú vlastnosti objektívne ideálne, pre neprítomnosť špecifického skreslenia však medzi poslucháčmi subjektívne nedosahujú taký „pekný“ zvuk (pri preťažení – ktoré sa často pre subjektívnu výhodu hlasnejšieho zvuku používa – je v nich zosilnená najmä tretia harmonická frekvencia²⁵). V súčasnosti sa však používajú najmä tranzistorové zosilňovače, keďže majú väčšiu životnosť a menšie rozmery. Navyše, vhodné skreslenie možno umelo dotvoriť počas mixáže.

Kompresia akustickej hladiny (angl. *audio level compression*): Ak je rozdiel medzi tichou pasážou (napr. sólo hoboja) a veľmi hlasnou pasážou (burácajúci orchester) v zázname príliš veľký, prichádza na rad kompresia. Zaznamenaný akustický tlak je spracovaný logaritmickou prenosovou funkciou. Dôsledkom je zmenšenie rozdielu medzi zvukmi s nízkou amplitúdou a zvukmi s amplitúdou

25 HAMM, Russel O. Tubes vs. Transistors: Is There an Audible Difference? 1972.

vysokou. Keďže teraz záznam zaberá menší dynamický rozsah, možno ho celý zhlasiť. Dôsledkom je zvýšenie akustického tlaku záznamu pri zachovaní jeho maximálnej amplitúdy. (Treba spomenúť aj subjektívnu výhodu kompresie a následného zhlásenia: hlasnejší záznam – v porovnaní s tichším – je poslucháčom vnímaný ako kvalitnejší. Posilnením tichších pasáží vzniká dojem, že nástroj alebo rečník, na ktorého záznam bola aplikovaná kompresia, sa nachádza bližšie k poslucháčovi; rozdielnou kompresiou tak možno jednotlivé nástroje „rozložiť“ okolo poslucháča.)

Frekvenčný filter. Zvuk na vstupe tohto filtra je akoby separovaný na jednotlivé frekvenčné zložky, pričom sa na ich intenzity aplikuje prenosová funkcia závislá od frekvencie. Ide zdanlivo o konvolúciu hodnôt intenzít jednotlivých frekvenčných zložiek vstupu a frekvenčnej charakteristiky filtra. Medzi najrozšírenejšie filtre patria:

- **Vysokofrekvenčný priepust** (angl. *hi-pass filter*): frekvenčné zložky vyššie než zvolená frekvencia prepúšťa bez zmeny, nižšie utlmuje; od istej frekvencie nižšie už zvuky vôbec neprepúšťa.
- **Nízkofrekvenčný priepust** (angl. *lo-pass filter*): funguje obdobne, ale prepúšťa len nízke frekvencie.
- **Pásmový priepust** (angl. *band-pass filter*): prepúšťa len frekvenčné zložky v istom rozpätí. Zložky frekvenčného spektra mimo tohto pásma utlmuje, v ideálnom prípade vôbec neprepúšťa.
- **Pásmový uzáver** (angl. *band-stop filter*): frekvenčná charakteristika tohto filtra je opakom pásmového priepustu – neprepúšťa frekvencie v danom rozsahu.
- **(Grafické) Ekvalizéry** (angl. *graphic equaliser*) rozdeľujú frekvenčné pásmo na niekoľko zón. Tieto zóny sa čiastočne prekrývajú (útlm jedného filtra v závislosti od frekvencie je vyrovnávaný nárastom priepustnosti ďalšieho filtra, aby ich súčet bol jednotkový). Ak sú tieto filtrované zóny znova spojené bez zmeny v hlasitosti, dostávame pôvodný zvuk. Nezávislým zvyšovaním a znižovaním hlasitosti výstupu jednotlivých zón pred zmiešaním zvuku môžeme doceliť zvýraznenie alebo útlm niektorých kmitočtov.

Frekvenčné filtre sú silným nástrojom, možno nimi napr. zosilniť v nahrávke basové alebo výškové spektrum, ktoré je napr. pre horší druh mikrofónu alebo podmienky miestnosti v pôvodnom zvuku utlmené. Takisto ich možno použiť na odfiltrovanie šumu, sykaviek (napr. kompresiou akustickej hladiny určitého frekvenčného spektra) atď. Filtre²⁶ sa dajú konštruovať na úrovni elektroniky (vďaka rôznym frekvenčným a fázovým charakteristikám súčiastok) alebo digitálne (uvádzame v odseku 3.01, str. 35).

²⁶ Nielen spomínané druhy frekvenčných filtrov, ale aj napr. polyfázové kvadrátne filtre a kvadrátne zrkadlové filtre (p. odsek 3.3.3 *Ortogonalne transformácie v praxi*, str. 52).

2.4 Analógové spracovanie zvuku

O zvuku možno veľa hovoriť, ale ak s ním chceme pracovať, musíme myslieť na jeho záznam a reprodukciu. V tejto časti práce spomenieme aspekty, ktoré vstupujú do hry pri analógovom spracovaní zvuku. Má tri odseky: 2.4.1 *Záznam zvuku* (str. 26), 2.4.2 *Spracovanie zvuku* (str. 27), 2.4.3 *Reprodukcia zvuku* (str. 27).

2.4.1 Záznam zvuku

Na záznam sa vo všeobecnosti používajú **mikrofóny**. Ich úlohou je, laicky povedané, premeniť akustickú energiu na energiu elektrickú. Presnejšie: menia svoj elektrický odpor (a tým aj veľkosť napätia a prúdu, ktoré nimi pretekajú) na základe aktuálnej zmeny tlaku vzduchu. Líšia sa v konštrukcii a iných vlastnostiach.

Podľa konštrukcie možno mikrofóny deliť na rôzne druhy. Najrozšírenejšími sú **kapacitné** (nazývané aj **kondenzátorové**) a **dynamické** mikrofóny. Prvé sú konštrukčne náročnejšie a potrebujú osobitné napájanie, poskytujú však lepšiu citlivosť a frekvenčnú vernosť. Druhé sú pre nižšiu cenu rozšírenejšie, vhodné sú zas na záznam zvuku s vysokou hlasitosťou (spev zblízka, bicie, ...).

Podľa citlivosti v závislosti od polohy zdroja zvuku možno mikrofóny deliť na **gul'ové** (všesmerové), **osmičkové** (dvojsmerné) a **kardioidné** (alebo s gul'ovou, osmičkovou či kardioidnou smerovou charakteristikou). Nasleduje popis ideálneho správania sa jednotlivých druhov:

- Gul'ové mikrofóny verne snímajú zvuk bez ohľadu na to, odkiaľ prichádza.
- Osmičkové mikrofóny majú dobrú citlivosť, ak zvuk prichádza spredu alebo zozadu, pri zmene polohy snímajú slabšie, z boku nesnímajú vôbec.
- Kardioidné mikrofóny najlepšie snímajú spredu, z boku slabšie, zozadu vôbec.

Tento ideálny stav je možný len pre nižšie frekvencie; čím je frekvencia vyššia, tým sú výsledky menej podobné želanému stavu. Na vine je fyzická konštrukcia mikrofónu, navyše aj jeho uchytenie môže predstavovať prekážku zvuku. (Poznámka: nízke frekvencie malé prekážky „obídu“ ľahšie než vyššie frekvencie.)

Podľa elektrického odporu možno mikrofóny deliť aj na vysokoimpedančné (zvyčajný druh) a nízkoimpedančné (štúdiová technika).

Záznam zvuku sa v ideálnom prípade vykonáva v štúdiu, ktoré je odhlučnené od vonkajšieho prostredia a má želané vlastnosti (veľkosť miestnosti, odrazivosť povrchu stien atď.). Je potrebné vyvarovať sa známeho „pukania“ vo zvukovom zázname (angl. *popping*), ktoré vzniká pri vyslovovaní hlások „p“ a „b“, resp. sykotu (angl. *essing*) pri hláskach „s“, „š“. Sú spôsobené rýchlym pohybom vzduchu, ktorý mikrofón

zaznamená ako zvuk (veľmi hlučný, preto je ním preťažený), aj keď to zvuk nie je²⁷. Podobné problémy spôsobuje aj vietor. Riešením je vrstva molitanu, v lepšom prípade nylonová či kovová sieťka, ktorá bráni pohybu vzduchu, predsa sa však cez ňu šíri zvuk. Pomôcť môže aj neskoršia úprava frekvenčnými filrami²⁸.

2.4.2 Spracovanie zvuku

Zvuk sa po zázname ďalej spracúva. V prvom rade sa naň aplikuje orezávanie²⁹ (teda obmedzenie rozsahu signálu – neznamená to však, že k samotnému orezaniu dôjde), aby sa privysokým signálom nepoškodila elektronika a zachovali korektné výsledky. Hneď potom sa (zväčša na úrovni elektroniky) odstránia kmitočty vyššie než 20-22 kHz (ľudské ucho ich už nezachytí) a nižšie než 20 Hz (podobne). Zo záznamu sa odfiltruje aj frekvencia napájania (50 Hz), ktorá sa v ňom často objavuje z dôvodu výkyvov v elektrickom napájaní prístrojov.

Nahrávacie štúdiá sa snažia o relatívne vyrovnaný priebeh hlasitosti, aj za cenu čiastočného potlačenia vernosti dynamiky nahrávky. Aby mohli na nahrávku úspešne aplikovať kompresiu akustickej hladiny, je potrebné zabezpečiť jej neskreslenú (neorezanú) dynamiku, tzn. je potrebné sledovať jej hlasitosť už počas záznamu. Na to slúžia rôzne zariadenia merajúce úroveň signálu zvuku. Ak sa to vykonáva vo väčšom časovom intervale (časovom „okne“), ide o meranie aktuálnej priemernej hlasitosti. Príkladom sú britsko-americké VU metre (*volume unit*) – sú nastavené na 300 ms. V prípade krátkeho času nábehu zariadenia ide o meranie špičky, napr. britské PPM (*peak programme meter*) reagujú v čase 10 ms. Kratšie intervaly nie sú potrebné, harmonické skreslenie, ktoré by kvôli orezaniu vzniklo, trvá príliš krátko, aby ho ľudský sluch zachytil.

2.4.3 Reprodukcia zvuku

Na reprodukciu zvuku sa využívajú dve hlavné metódy. Prvou sú slúchadlá. Reprodukory slúchadiel sú len nepatrne vzdialené od ušného bubienka, preto môže byť ich výkon veľmi malý. To umožňuje zmenšiť ich rozmer a dať im kvalitatívne atribúty za omnoho nižšiu cenu, v porovnaní so samostatnými reproduktormi. Ak sú uzavreté, pôsobia navyše ako hluková bariéra, filtrujúca zvuky prichádzajúce zvonka. Ide vtedy o asi najkvalitnejší umelý posluš, ktorý možno súčasou technikou dosiahnuť. Nevýhodou je rýchla „únava“ uší (sluchová i zdravotná) po dlhšom čase posluchu z dôvodu zastavenia cirkulácie vzduchu vo zvukovode. Otvorené slúchadlá tento problém do značnej miery riešia, no za cenu vyššej priepustnosti hluku zvonku.

27 Ako sme si povedali v časti 2.1 *Zvuk ako fyzikálny fenomén* (str. 9), pri šírení zvuku sa vzduch nehýbe.

28 P. odsek 2.3.4 *Úprava a zmena zvuku*, str. 24.

29 Tamže.

Vo všeobecnosti si treba dávať pozor na vyššiu hlasitosť posluchu – blízkosť reproduktorov k uchu oberá o objektívny úsudok o hlasitosti reprodukcie a môže spôsobiť poškodenie sluchu.

Najrozšírenejší spôsob reprodukcie sú samostatné reproduktory. Nachádzajú sa na stenách, na stoloch, v mobilných telefónoch či na koncertných pódiami. Každý z nich je čímsi špecifický. Konštrukcia membrány reproduktora spôsobuje, že niektoré frekvencie dokáže reprodukovat' veľmi dobre, niektoré horšie alebo vôbec. Napr. veľmi malý reproduktor môže mať vysoký výkon, ale fyzická konštrukcia mu bráni dosiahnuť vysokú amplitúdu (a teda aj akustický tlak) na to, aby v nízkych frekvenciách prekročil čo i len prah počuteľnosti. Veľmi veľkému reproduktoru zas zotrvačnosť membrány bráni meniť svoju polohu podľa signálu príliš rýchlo (prípád vysokých kmitočtov), pri nízkych frekvenciách však dokáže vytvoriť vysoký akustický tlak. (Ak je reproduktor malý, dokáže „nahradit“ nízke frekvencie akustickou ilúziou – **chýbajúci základný tón** mimo spektra reproduktora napodobňuje prítomnosťou jeho harmonických násobkov. P. odsek 2.2.3 *Vnímanie kvality, harmonické tóny*, str. 15.)

Z dôvodu komplementarity rôznych veľkostí reproduktorov sa teda tieto často inštalujú vo väčšom množstve (2-4), zväčša blízko seba, v jednom³⁰ šasi. A tak sa napr. pre stereofonický posluš nepoužijú dva reproduktory (ľavý a pravý), ale rovno dve sústavy reproduktorov. Využitie nie je len v posilnení akustického výkonu (potrebné napr. pre vystúpenia kapiel naživo), ale najmä v rozdelení frekvenčného spektra. Každý reproduktor v sústave spracúva tú časť, v ktorej má najmenšie skreslenie. Súčasťou reproduktorovej sústavy je teda aj elektronika, ktorá filtruje, rozdeľuje, upravuje signál (frekvenčná výhybka, angl. *crossover*); nastavená je výrobcom tak, aby výsledný zvuk bol čím vernejší, podobnejší originálu, pri čo najširšom spektre (zväčša uvádzané ako napr. „odstup ± 3 dB v rozsahu 45 Hz – 19 kHz“). O kvalite reprodukcie tak rozhodujú všetky komponenty reproduktorovej sústavy, pričom výsledok závisí od najslabšieho článku. Okrem reproduktorov s celým frekvenčným rozsahom (**celorozsahové**, *full-range*) existujú i dedikované **basové** (*woofer*, ešte basovejší sa volá **subwoofer**), **strednopásmové** (*mid-range*) a **výškové** (*tweeter*) reproduktory.

Všeobecný trend medzi domácimi používateľmi je používať reprodukčné sústavy, ktoré majú charakteristický zdvih na vysokých frekvenciách (dodáva pocit farby – posilnením prirodzených harmonických tónov) a frekvenciách nízkych (dodáva hutnosť vďaka vysokej amplitúde tlaku vzduchu), inštaláciou niekoľkých vysokofrekvenčných reproduktorov a zväčša jedného subwooferu (často ešte doplneného tzv. bass-reflexom). Posluš na takýchto sústavách môže znieť subjektívne dobre, z objektívneho hľadiska je však veľmi skreslený.

30 Aby sa zabránilo efektu hrebeňového filtra – p. odsek 2.5.2 *Stereofonický záznam zvuku*, str. 31.

Reproduktory menia len časť vstupnej energie na energiu akustickú³¹, zvyšok sa mení na teplo. Ich efektívnosť sa udáva v dB/W. Konštrukčne bezpečný elektrický príkon (značený **RMS**, *root mean square*; ide o priemerný príkon, ktorý sa dá vypočítať zo štandardnej odchýlky v napätí pri zachovaní odporu obvodu) sa udáva vo Wattoch (občas sa objavuje aj označenie **PMPO** – maximálny špičkový príkon zariadenia; často ho reproduktor nevydrží dlhšie než pár milisekúnd a ide skôr o „marketingové“ číslo).

Vzhľadom na popísaný proces záznamu, spracovania a reprodukcie zvuku môžu čitatelia vidieť, že zvuková „vernosť“ je v každom tomto procese veľmi ťažko dosiahnuteľná, často dokonca neželaná – zvuk, ktorý nie je verný, ale prejde mnohými úpravami skúsených zvukárov, znie v konečnom dôsledku občas „lepšie“ než originál.

31 Tzv. **budič** reproduktoru (angl. *driver*) mení elektrickú energiu na kinetickú – pohyb membrány, ktorá následne rozkmitáva vzduch v okolí.

2.5 Priestor vo zvuku

Doteraz sme hovorili o zvuku ako fenoméne, ktorý je v závislosti od času a miesta jedinečný. Je to pravda. Človek má však dve uši, preto počujeme akoby dva vnemy naraz. Ich porovnávaním mozog zisťuje, z akého smeru zvuk prichádza. V úvode tejto časti práce si spomenieme základné indicie pre priestorový vnem, nasledujúce tri odseky nesú názvy 2.5.1 *Problémy monofonického záznamu a reprodukcie* (str. 31), 2.5.2 *Stereofonický záznam zvuku* (str. 31) a 2.5.3 *Súčasný trendy* (str. 33).

Prvotným aspektom pomáhajúcim lokalizovať zvuk je **rozdiel v hlasitosti** medzi vnemom zachyteným oboma ušami. Tento aspekt je však sám osebe nedostatočným ukazovateľom umiestnenia zdroja, preto sa kombinuje s ostatnými aspektmi.

Ďalšou indíciou pre ľudský mozog je **časové oneskorenie** zvukov. Ak zvuk (najlepšie impulzného charakteru) vznikol vpravo od poslucháča, dorazí najprv do pravého ucha, do ľavého (vzdialeného asi 15 cm vzdušnou čiarou – v skutočnosti musí prejsť asi 23 cm okolo hlavy) dorazí o 0,68 ms neskôr. Podobná situácia nastáva, ak je zdroj zvuku vľavo. Ak sa zdroj nachádza priamo pred (za) poslucháčom, nedochádza k oneskoreniu. Obmedzenia ľudského sluchu spôsobujú, že ak dorazí do uší najprv zvuk z jedného smeru a potom ten istý zvuk (z iného zdroja – alebo len napr. ozveny pôvodného) zo smeru iného, ak sa to všetko odohrá za čas kratší než 15 ms, sú poslucháčom vnímané ako jednoliaty celok prichádzajúci zo smeru prvého zvuku³², a to takmer bez ohľadu na intenzitu druhého (a ďalších) zvukov, ktorá môže byť i vyššia ako v prípade prvého zvuku (tzv. Haasov efekt). (Niektorí autori³³ uvádzajú, časový odstup aj 50 ms až 100 ms; musí však byť väčší ako 2 ms až 10 ms, ináč sluch zachytí ako zdroj zvuku bod na spojnici oboch zdrojov zvukov.) Vo všeobecnosti je človek citlivejší na čas zachytenia vyšších frekvencií, pri nízkych kmitočtoch vníma čas s istým fyziologickým oneskorením. Inými slovami: človek vykazuje väčšiu časovú toleranciu pri nízkych tónoch.

Tretím aspektom je **fázový posun**. Kvôli časovému oneskoreniu je ľubovoľný konštantne znejúci tón vnímaný oboma ušami korektne, aktuálna fáza zvuku je však odlišná. Keďže za normálnych okolností obe uši počujú zvuk z toho istého zdroja (len neskôr alebo s fázovým posunutím), je mozog na základe spomenutých odchýlok schopný veľmi presne lokalizovať zdroj zvuku. Táto indícia však podlieha tzv. efektu hrebeňového filtra³⁴. Ako dôsledok ľudský sluch slabo lokalizuje zdroj veľmi hlbokých tónov (prichádzajú v takmer totožnej fáze). pri vysokých tónoch

32 SIBBALD, Alastair (Sensaura). An Introduction to Sound and Hearing. 2001. s. 11.

33 Napríklad v článku Wikipédie na stránke: <http://de.wikipedia.org/wiki/Präzedenz-Effekt>.

34 P. odsek 2.5.2 Stereofonický záznam zvuku, str. 31.

(nad 2 kHz³⁵) je zas fázový rozdiel priveľký, preto ho ľudský sluch ignoruje (a do hry vstupuje najmä rozdiel v hlasitosti).

Ak nastáva zmena vzájomnej polohy poslucháča a zdroja zvuku, dochádza k **frekvenčnému posunu**, ako dôsledku známeho Dopplerovho efektu.

Nakoniec, vďaka **zmene frekvenčnej charakteristiky** (zapríčinennej najmä ušnými boltcami, ale aj nesúrodým tvarom ľudskej hlavy) je mozog schopný odlíšiť od seba zvuky prichádzajúce spredu alebo zozadu, zhora či zdola.

Všetky tieto aspekty sa dajú odborne vyjadriť v tvare „**prenosovej funkcie súvisiacej s hlavou**“ (*head related transfer function*, HRTF). Vedci z mnohých oblastí neustále pracujú na jej upresňovaní a zdokonaľovaní.

2.5.1 Problémy monofonického záznamu a reprodukcie

Pri porovnaní ľudského sluchu a techniky na záznam zvuku stojíme pred dilemou: ak sa snažíme zaznamenávať zvuk mikrofónom, strácame jeho priestorovú hodnotu. Poznáme akustické zmeny „tu a teraz“, ale nezachytíme informácie o priestore (tie sa však čiastočne prejavia v čase dozvuku miestnosti). Nakoniec získa zvuk priestorovú informáciu pri reprodukcii – lokalizácia jeho zdroja poslucháčom bude totožná s polohou reproduktora.

Vhodné by bolo mať dva zvukové záznamy a dva reproduktory. aby sme mohli dosiahnuť stereofonický vnem. Pravdou však je, že bez ďalších úprav by sme zrejme dosiahli lokalizáciu zdroja zvuku len na spojnici medzi oboma reproduktormi (v tomto prípade určenú najmä pomerom hlasností daného zvuku v oboch reproduktoroch).

Ponúka sa riešenie: Vytvoriť model hlavy a do jej vnútra, na miesta, kde sa nachádzajú ušné bubienky, umiestniť mikrofóny. Záznam by mal byť dokonalý. Problém je s reprodukciou. Ak by sme použili dva reproduktory, dostali by sme nezvyčajný a ťažko lokalizovateľný zvuk – záznam a reprodukcia sú na celkom odlišných miestach. Nutne by bolo treba zvuk reprodukovať dvoma vysokokvalitnými reproduktormi zasunutými do zvukovodov uší. Okrem cenovej náročnosti a nízkeho pohodlia poslucháča má tento prístup ďalšiu nevýhodu: ignoruje rozdielnosť poslucháčov čo do vzdialenosti uší medzi sebou a tvaru ušného boltca.

2.5.2 Stereofonický záznam zvuku

Pri zvukovej nahrávke (kapely, zboru) sa vo väčšine prípadov nevie, akým spôsobom bude poslucháč zvuk reprodukovať. Môže použiť pár vysokokvalitných

35 P. napr. http://www.chiariglione.org/MPEG/tutorials/papers/icj-mpeg4-si/09-natural_audio_paper/gacoding.html.

reproduktorov, malé reproduktory so slabou reprodukciou basov, slúchadlá, ... Alebo môže mať k dispozícii len monofonickú reprodukciu (napr. tranzistorové rádio), prípadne je vzájomná vzdialenosť reproduktorov v porovnaní so vzdialenosťou od poslucháča taká malá, že sa správajú takmer ako jeden reproduktor.

Pri posledných dvoch prípadoch vzniká problém. Ak chceme zaznamenať zvuk na dvoch miestach, vznikajú dva rôzne záznamy. Reprodukovaný je však ich súčet. Vzhľadom na fázový posun nahrávok môže dôjsť k fenoménu nazývanému **efekt hrebeňového filtra**. Predstavme si, že zdroj zvuku je v jednej línii s oboma mikrofónmi, na ktorý je zaznamenávaný. Tie nech sú vzdialené od seba 1 m. Tóny s vlnovou dĺžkou 1 m (343 Hz) budú tak reprodukované dvojnásobne vyšším signálom (súčet vlnení vo fáze), tóny s vlnovou dĺžkou 2 m však vôbec nebude počuť (stretnú sa v opačnej fáze – dôjde k vynulovaniu). Podobne pre tóny frekvencie $343 \times n$ Hz, resp. $171,5 + 343 \times n$ Hz. Dôsledkom je silné skreslenie frekvenčnej charakteristiky zaznamenaných zvukov. Podľa možnosti je vhodné poradiť si s týmto problémom už pri nahrávaní.

Predstavíme si tri rôzne spôsoby stereofonického záznamu zvuku, ako ich predstavuje Vlachý³⁶:

Systém XY (koincidenčný pár mikrofónov). V tomto prípade sú použité dva kardioidné (možno použiť i osmičkové) mikrofóny umiestnené v tesnej blízkosti, pričom vzájomné vychýlenie osí ich snímania je 90° . Jeden sníma pravú polovicu miestnosti, druhý ľavú (stred leží medzi osami mikrofónov). Vďaka veľmi blízkej polohe mikrofónov nenastáva problém s efektom hrebeňového filtra. Nevýhodou je najmä skreslené podanie zvuku zo stredu – útlm kardioidných mikrofónov je závislý nielen od uhla, ale aj od frekvencie, preto majú zvuky zo stredu stlmené výšky (čo čiastočne simuluje útlm pri prirodzenom posluchu zvukov, spôsobený ľudskou hlavou). Pre blízke umiestnenie mikrofónov tiež zaniká všetka fázová informácia, ktorá by vznikla pri prirodzenom posluchu vďaka vzájomnej vzdialenosti uší. Tieto nevýhody sú však mierne, preto je systém XY najčastejším druhom snímania.

Systém MS (*middle & side*, teda stred a strana). Vychádza z predpokladu, že stereofonickú informáciu tvorí najmä rozdiel medzi „ľavým“ a „pravým“ záznamom. Použije sa pár mikrofónov v tesnej blízkosti, pričom jeden (guľový) sníma stred miestnosti, druhý (osmičkový) je namierený na strany (teda osi sú kolmé na seba). Výsledné dva kanály sa získajú ako súčet zvuku zo stredu a zvuku zo strán v pôvodnej fáze (ľavý kanál) alebo protifáze (pravý kanál), resp. naopak. Výhodou je úplná kompatibilita s mono reprodukciou (odlišné fázy sa vynulujú a ostane len stred) a verný záznam zvuku zo stredu miestnosti. Čím je však zdroj zvuku bližšie k stranám, tým znie neprirodzenejšie. Ak však odchýlky od stredu nie sú príliš veľké, možno meniť šírku stereofonického vnemu i dodatočne (zvýšením váhy z postranného mikrofónu).

36 VLACHÝ, Václav. Praxe zvukové techniky. 2000. s. 46-50.

Systém AB. Použijú sa dva mikrofóny, ktoré sú od seba dostatočne vzdialené. Vzniká verný a silný stereofonický vnem, s ním však aj zreteľný efekt hrebeňového filtra. Používa sa skôr pri zázname väčších priestorov (orchester) na dodanie „celkového dojmu“, v kombinácii s ďalšími „bodovými“ mikrofónmi, ktoré zaznamenávajú jednotlivé hudobné nástroje.

2.5.3 Súčasné trendy

Do skúmania trojrozmerného vnímania zvuku sa investuje množstvo peňazí, keďže ide o komerčne zaujímavú tému. Faktom je, že v súčasnosti existujú systémy (čisto softvérové³⁷ i softvérovo-hardvérové³⁸), ktoré dokážu vytvoriť veľmi hodnovernú ilúziu skutočného zvuku (tzv. 3D audio efekt), pričom sú schopné pracovať nielen so slúchadlami, ale aj s reproduktormi, dokonca umožňujú nakonfigurovať systém podľa tvaru ušnice poslucháča. V prípade reproduktorového posluchu sa používajú rôzne proprietárne technológie na odstránenie presluchu, teda duplicitného počutia zvuku z jedného reproduktora v oboch ušiach (angl. *transaural acoustic crosstalk cancellation*).

Popri viacerých reproduktoroch a slúchadlách so psychoakustickým modelom jestvuje i tretia metóda vytvárania dojmu priestorového zvuku. Využíva sa Huygensov princíp, ktorý hovorí, že každý bod média, do ktorého dorazí vlna, môže byť považovaný za nový zdroj vlnenia. S použitím výpočtovo náročných postupov a množiny reproduktorov možno zostrojiť dva efekty, akési analógie hologramu, len vo svete zvuku. Prvým je **ambisonický zvuk**, ktorý na určenom mieste v priestore vytvára miesto dokonalého priestorového posluchu (angl. *sweet point* – poslucháč na tom mieste počuje dokonalý trojrozmerný zvuk; zmenou polohy sa zhoršuje lokalizácia domnelého zdroja). Druhým je **syntéza vlnového poľa** (angl. *Wave Field Synthesis*), pri ktorom sa vytvára dojem z trojrozmerného zvuku v celom zvolenom priestore (nedosahuje však takú presvedčivosť ako prvá metóda). Prvý prístup je menej náročný na hardvér, softvér a výpočtovú kapacitu a existuje mnoho produktov (hardvérových i softvérových, komerčných i voľne dostupných), ktoré ho realizujú. Druhý prístup rozvíjajú napr. komerčné riešenia firiem Sonic Emotions a Iosono. Aj v tejto oblasti je výskum stále aktívny.

37 Jedna z firiem, ktorá sa tejto oblasti profesionálne venuje, je *Sensaura* (<http://www.sensaura.co.uk>). Na ich stránke možno nájsť množstvo materiálov z oblasti trojrozmerného vnímania zvuku. Kedysi hrala dôležitú úlohu aj firma *Aureal* so svojou technológiou A3D, po finančných problémoch však bola odkúpená firmou Creative.

38 Príkladom môže byť zvuková karta s ovládačmi *X-Fi* od firmy *Creative* (<http://www.creative.com>), založená práve na výskumoch firiem *Sensaura* a *Aureal*.

3 DIGITÁLNY SVET ZVUKU

V predošlej kapitole (2) sme si predstavili zvuk v jeho prirodzenom, spojitom prostredí. Je to forma, v ktorej ho môžeme počuť, zaznamenávať, reprodukovať. Dokonca aj uchovávať.

Je niekoľko analógových spôsobov archivácie zvuku. Napríklad gramofónová LP platňa alebo magnetická páska. Tým, že sú analógové, dokážu v ideálnom prípade zo svojej podstaty zachytiť zvyčajnú zvukovú informáciu ako celok, tzn. bez nutnej straty dynamiky, frekvenčného spektra atď. To je však len teória.

V praxi sa stretávame s obmedzením na technologickej úrovni. Kvantové javy nám diktujú nutnosť existencie najmenšieho zaznamenateľného signálu, konštrukcia úložného zariadenia zas obmedzuje maximálny signál. Použitá elektronika, ktorá upravuje signál, aby bol vhodný na záznam, má svoju frekvenčnú charakteristiku, harmonické skreslenie, odstup signálu od šumu, medzikanálový presluch, nelineárny útlm signálu, ... Dôsledkom je, že aj keď sa snažíme o „dokonalý“ prenos, vždy vznikajú chyby, skreslenia. Každé viacnásobné spracovanie (napr. mixáž zvukov v štúdiu) skreslenie len zvyšuje, odchýlka od originálu sa „hromadí“, je stále viac počuteľná. Prirodzenou snahou je teda používať čím kvalitnejšie zariadenia, aby bolo skreslenie minimálne – to však znamená predraženie celkového riešenia.

Druhým nepriateľom je šum, ktorý vzniká na (analógových) záznamových médiách nielen používaním, ale aj samovoľne, časom. Znova platí, čím kvalitnejšie (a drahšie) zariadenia, tým sú starosti so šumom menšie, no celkom sa mu vyhnúť nedá.

Hľadané riešenie musí uchovávať údaje o zvuku bez zmeny, teda zvuk musí byť presne popísateľný, nesmie byť len „javom“. Navyše, zariadenia, ktoré s ním pracujú, musia mať „kontrolovateľnú“ kvalitu, ktorú možno pri ich návrhu jednoducho podľa potreby zvyšovať, bez nutnosti predraženia riešenia. Práca s ním by mala byť jednoduchá, takmer bezplatná, bez využitia ďalších a ďalších obvodov. Riešenie musí byť univerzálne, číslicové, digitálne.

V časti 3.1 *Problematika vzorkovania zvuku* (str. 39) rozoberieme spôsoby, ktorými meníme analógový zvuk na digitálny, načrtneme problémy a technologické aspekty digitalizácie zvuku.

Keď sme už zvuk zmenili na sled číslic, musíme sa naučiť tieto číslice zapísať. S niekoľkými spôsobmi a metódami čitateľa oboznámime v časti 3.2 *Digitálne kódovanie čísel* (str. 44).

V častiach 3.3 *Princípy stratovej kompresie zvuku* (str. 50) a 3.4 *Princípy bezstratovej kompresie zvuku* (str. 59) sa bližšie oboznámime s metódami, ktoré používajú tieto dva odlišné postupy, ktoré nie sú natoľko konkurenčné, ako sa môže zdať.

Zvuk, hudba či hovorené slovo – v súčasnosti rozšírenejšie ako kedysi, najmä však celkom inou formou. Na trhu je nepreberné množstvo rôznych elektronických zariadení a počítačov, ktoré umožňujú zvuk v digitálnej podobe zaznamenávať, reprodukovat', spracúvať. V časti 3.5 *Zvuk v súčasnom svete elektroniky* (str. 64) objasníme základnú stavbu a použitie zvukového systému vo svete PC, popíšeme jednotlivé rozšírené spôsoby záznamu, prenosu a „použitia“ zvuku v praxi, čitateľa oboznámime s trendmi a inováciami nielen v minulosti, ale načrtne aj vývoj v blízkej budúcnosti.

Kvôli zrozumiteľnosti je však namieste definovať niektoré pojmy z digitálneho sveta zvuku.

3.0.1 Definícia pojmov

V súčasnej dobe sa používateľ pri práci so zvukovými formátmi môže stretnúť s mnohými termínmi, ktoré mu nemusia byť jasné. Niektoré, spomínané tiež v tejto práci, si zasluhujú bližší výklad.

Kompresný algoritmus (resp. len **algoritmus**) je program, ktorý (zvukové) dáta v nekomprimovanom formáte¹ nejakým matematickým postupom prevádza do menšieho objemu, teda upravuje zápis dát tak, aby potrebovali menší priestor. Keďže sa dopĺňa s **dekompresným algoritmom**, často s ním tvorí jeden program (alebo súbor programov), laicky tiež nazývaný **kodek** (z angl. *compress-decompress*).

Pod **zvukovým formátom** (resp. len **formátom**) budeme rozumieť **tok** (sekvenciu) **dát** a / alebo **špecifikáciu**, na základe ktorej tieto dáta vznikajú (resp. na základe ktorej ich treba interpretovať), pričom pôvod týchto dát bude vo výstupe kompresného algoritmu². Dáta sú zväčša združené do atomických jednotiek – **rámcov** – blokov dát tvoriacich menší časový rozsah zvukového záznamu³.

Spomínané dáta sú však zriedka uložené ako sekvencia čísel, skôr sú súčasťou **kontajnera**, čo je **súbor** počítačového systému (resp. jeho **špecifikácia**), ktorého úlohou je poskytnúť možnosť operovania nad pôvodným „surovým“ tokom dát zvukového formátu. Kontajner je zvyčajne navrhnutý priamo pre potreby daného formátu. Kontajner môže obsahovať (a zvyčajne aj obsahuje) viac tokov dát, ktoré po častiach (tzv. paketoch) zmiešava (angl. *mux*) dokopy, aby boli uložené synchronne (príkladom môže byť zvuk a obraz uložené v kontajneri AVI, ktoré sú nezávislými tokmi dát, predsa však musí byť zabezpečená ich synchronnosť). (Poznámka: je aj mnoho zvukových súborov – napr. súbory typu .mp3 – ktoré sú v skutočnosti

1 Teda vo forme PCM – p. časť 3.1 *Problematika vzorkovania zvuku*, str. 39.

2 Teda pojem „formát“ vystihuje aj samotný kompresný algoritmus v kontexte toku dát, ktoré prijíma ako vstup, alebo vytvára ako výstup, alebo v kontexte vlastností tohto toku dát.

3 Obsah rámcov, predovšetkým pri stratovej kompresii zvuku, sa extrahuje z tzv. okien, ktoré sa môžu navzájom prelínať – p. odsek 3.3.2 *Fourierova transformácia, kosínusová transformácia*, str. 51.

pôvodným tokom dát, často s pridanou identifikačnou značkou⁴, nejde teda o kontajnery v pôvodnom zmysle slova.)

Jednoduchým príkladom kontajnera môže byť súbor podľa špecifikácie RIFF (angl. *Resource Interchange File Format*). Ide o tzv. metaformát: umožňuje vložiť do jedného súboru niekoľko nezávislých blokov dát (v angl. sa používa termín *chunk*), ktoré sú uvedené hlavičkou obsahujúcou názov bloku (4 bajty – v praxi 4 písmená) a jeho veľkosť (tiež štvorbajtové číslo). Pre svoju jednoduchosť je hojne používaný (napr. MIDI súbory), označenie blokov sa zas stalo inšpiráciou na označovanie jednotlivých formátov v iných súboroch - napr. dvojbajtové číslo v kontajneri .wav, (ktoré je samo osebe RIFF súborom) identifikujúce použitý kodek (aj keď sa zväčša používa PCM⁵), alebo štvorbajtový „FOURCC“ kód označujúci obrazový kodek v súboroch typu .avi atď.

Kontajner často vsúva medzi dáta, ktoré obsahuje, rôzne ďalšie dátové „značky“ na ochranu pred stratou signálu (pokračuje sa od najbližšej značky), resp. na umožnenie pretáčania (náhodného prístupu, angl. *seek*) v súbore (v kombinácii s tabuľkami ukazovateľov na rôzne časti súboru uľahčujúcimi náhodný prístup, tzv. **seek tables**).

Súčasťou zvukových súborov (najmä ak sú použité na archiváciu či prenos hudby) býva aj **identifikačná značka** (angl. *ID tag*). Umiestňuje sa spravidla na koniec alebo začiatok súboru. Jej úlohou je v prvom rade sprostredkovať bližšie metaúdaje o zvukovom zázname – číslo piesne, album, autora hudby, speváka, atď., aby uľahčila archiváciu, vyhľadávanie a získavanie informácií o nahrávke. Môže obsahovať aj informácie štatistického rázu – napr. najvyššia hodnota signálu v zázname a priemerná hlasitosť nahrávky, čo umožňuje **zhlásenie pri reprodukcii** (angl. *replay gain*) – úpravou hlasitosti reprodukcie tej-ktorej skladby možno doceliť vyrovnanie priemernej hlasitosti vybranej množiny piesní, odstrániac tak disproporcie v dynamikách rozličných skladieb. Medzi najrozšírenejšie identifikačné značky patria:

- **ID3** (ID3 tag). Jeho prvú verziu (teraz označovanú *ID3v1*) stvoril roku 1996 programátor menom Eric Kemp. Mala umožniť zaznamenať jednoduché informácie o skladbe v súboroch typu .mp3. Jednotlivé polia majú obmedzenú veľkosť (napr. 30 znakov názov piesne, 30 znakov na autora atď.), celková veľkosť bloku identifikačnej značky je 128 bajtov, podporuje len ASCII. O značné rozšírenie tohto „štandardu“ sa postaral Martin Nilsson s ďalšími ľuďmi z internetovej komunity, ktorý definovali konvenciu pre **ID3v2**. Ide o samostatný kontajner umiestnený v súbore (zväčša .mp3), v ktorom môžu byť zaznamenané rozličné polia, texty so špeciálnymi znakovými sadami (UTF-8), dokonca sprievodné binárne dáta ako obrázky a iné. ID3v2 pre svoju striktnú štruktúru nemôže byť uložená v štandardných kontajneroch (Ogg, Matroska, WMA, ...),

4 P. text nižšie.

5 Kontajner .wav je vo svete MS Windows štandardne používaný práve na ukladanie zvuku vo formáte PCM, podobne ako .aiff, .au/.snd v prípade iných platforiem.

keďže bola od začiatku navrhovaná ako súčasť formátu MP3. Viac informácií možno nájsť na oficiálnej stránke <http://www.id3.org>.

- **Vorbis comment**⁶. Vyvinutý organizáciou Xiph.org Foundation pre formát Vorbis, neskôr používaný aj pre ďalšie formáty z ich dielne (FLAC, Speex). Umožňuje voľné používanie ľubovoľných textových položiek (dvojíc “*identifikátor=hodnota*“), pričom definuje len konvenciu ich zápisu, nie ich vnútornú štruktúru a reprezentáciu. Spolieha sa teda na zahrnutie do kontajnera (zväčša Ogg). Povoľuje viacnásobné použitie toho istého identifikátora (napr. pri dvoch spevákoch, dvoch skladateľoch atď.). Oficiálna špecifikácia je dostupná na <http://www.xiph.org/vorbis/doc/v-comment.html>.
- **APEv2**. Vznikol ako rozšírenie pôvodnej identifikačnej značky formátu Monkey's Audio programátorom Frankom Klemmom, ktorý ju navrhol pre svoj formát Musepack. Má striktnejšiu štruktúru ako Vorbis comment, umožňuje skladovať viac hodnôt pre jeden identifikátor, binárne údaje, texty k piesňam atď. Okrem Monkey's Audio (do ktorého bola neskôr tiež začlenená) ju používajú formáty Musepack, OptimFROG a WavPack. Špecifikácia je napr. na stránke http://wiki.hydrogenaudio.org/index.php?title=APEv2_specification.

Vyššie spomínané rozlišovanie medzi algoritmom, formátom a kontajnerom je nutné, keďže nie vždy bude z kontextu jasné, o čom hovoríme. Príkladom môže byť skratka MP3, ktorá môže označovať kompresný algoritmus podľa špecifikácie konzorcia MPEG, alebo zvukový formát (tok dát), ktorý táto špecifikácia tiež podrobne opisuje, alebo známy rozšírený súbor typu .mp3 (ktorý síce nie je kontajner, môže sa v ňom však nachádzať napríklad identifikačná značka), alebo všetky tri možnosti naraz (keďže algoritmus, formát a kontajner tvoria spolu organický celok).

Rýchlosť toku dát, na základe ktorého možno rekonštruovať istý časový úsek zvuku, budeme označovať **dátový tok** (angl. *bitrate*, bitová prenosová rýchlosť) a počítat' v tisíckach bitov potrebných na rekonštrukciu jednej sekundy zvukového záznamu (jednotka kbps⁷ – z angl. *kilobits per second*).

Predovšetkým stratové kompresné algoritmy sú obvyčajne navrhnuté s úmyslom dosiahnutia istého (tzv. typického) dátového toku, ktorý vyplýva z ich návrhu a určenia. Ak dátový tok potrebný na zápis jedného rámca je konštantný, resp. možno ho vybrať z diskkrétnej množiny hodnôt, hovoríme, že formát dosahuje (resp. má, umožňuje atď.) **konštantný dátový tok** (angl. *constant bitrate*) a označovať skratkou **CBR**. Ak návrh algoritmu umožňuje ľubovoľný dátový tok, budeme hovoriť o **variabilnom dátovom toku** (angl. *variable bitrate*, **VBR**). Úloha VBR je jasná – kým pri CBR musí kompresný algoritmus vyprodukovať pre jednoduché a tiché pasáže kódovaného zvuku rovnaké množstvo dát ako pre náročné, frekvenčne či informačne bohaté signály, s VBR môže jednotlivým rámcom pridelit' taký dátový

⁶ Občas sa označuje aj „*Xiph comment*“.

⁷ Správne by sme mali písať *kbit/s* a *Mbit/s* (v prípade miliónov bitov), prikloníme sa však k zaužívanému a viac rozšírenému označeniu *kbps* a *Mbps*.

tok, ktorý je potrebný na zachovanie danej „kvality“, zväčša vyjadrenej odstupom maskovacieho signálu od šumu, signálu od maskovacieho signálu⁸, presnosťou kvantovania atď.

Mnohé zvukové formáty (napr. WMA) sú schopné zaručiť⁹ **priemerný dátový tok** (angl. *average bitrate*, **ABR**) vzhľadom na celý kódovaný záznam napriek premenlivosti aktuálneho dátového toku v danej časti záznamu (zachovávajúc tak výhody VBR).

Na dôvažok treba spomenúť **filtre**, presnejšie ich digitálnu¹⁰ konštrukciu. Rozšíreným riešením ich konštrukcie je použitie **filter s konečnou impulzovou charakteristikou**. Ide o matematickú konštrukciu, pri ktorej novú hodnotu vzorky¹¹ vypočítame ako súčet súčinov hodnôt niekoľkých predošlých vzoriek s vhodne zvolenými koeficientmi¹². Počet použitých vzoriek určuje rád filtra. Ak sumáciu obohatíme aj o súčin hodnôt novovypočítaných vzoriek s vhodne zvolenými koeficientmi, hovoríme o **filtru s nekonečnou impulzovou charakteristikou**. Bližší popis metód konštrukcie filtrov prekračuje rámec tejto práce. Poznámka: v prípade frekvenčných filtrov sa ponúka aj iné, triviálne riešenie: vykonať frekvenčnú dekompozíciu (analýzu) signálu¹³, úpravu vo frekvenčnej doméne a následnú syntézu.

8 P. odsek 3.3.3 *Ortogonalne transformácie v praxi*, str. 52.

9 Záruka však vyplýva z *dvojprechodového* spracovania zvuku – prvý raz sa zhodnotí náročnosť vstupného signálu, vypočíta sa potrebná úroveň kvantovania, v druhom prechode prebieha samotná kompresia.

10 Možno ich realizovať i analógovo – p. odsek 2.3.4 *Úprava a zmena zvuku*, str. 24.

11 P. nasledujúcu časť (3.1) o vzorkovaní zvuku.

12 Ide teda o konvolúciu hodnôt predošlých vzoriek s číselným vektorom reprezentujúcim charakteristiku filtra. To je konštrukčne zhodné s lineárnou predikciou – p. odsek 3.4.2, str. 60.

13 Napr. použitím ortogonálnych transformácií, p. odsek 3.3.2, str. 51.

3.1 Problematika vzorkovania zvuku

Konverzia analógového zvuku na digitálny a späť je v podstate to najcitlivejšie miesto pri digitálnom spracovaní zvuku. Ak zmeníme analógový zvuk na digitálny hneď „na začiatku“, potom už pracujeme s diskretnými údajmi, číslicami. Všetko ďalej je už len jednoduchá aritmetika, ktorou môžeme zvuk¹⁴ skôr či neskôr ľubovoľne upraviť, archivovať atď. Potom, kdesi na konci procesu, môžeme zmeniť zvuk digitálny späť na analógový – bez toho, aby sme dovedy museli bojovať s problémami spomenutými v úvode kapitoly 3 (str. 34).

Srdcom digitalizácie je teda analógovo-digitálny prevodník (označuje sa tiež **A/D prevodník**, angl. *A/D converter*, **ADC**), ktorý prevádza analógový signál na digitálne vyjadrenie. Na reprodukciu je zas potrebný digitálno-analógový (**D/A**) prevodník (**DAC**) s opačným efektom. Tie sa často z praktických dôvodov kombinujú v jednom čipe, ktorý sa označuje **kodek** (kóduje a dekoduje zvuk z jednej formy do druhej). (Keďže zhodným slovom možno označiť aj softvérový program prevádzajúci zvuk z jednej formy do druhej kompresiou a dekompresiou, budeme toto slovo vždy používať v kontexte.) Vyrobiť dobrý prevodník je „umenie“. Ak už zvuk máme v digitálnej forme, v práci s ním ide „len“ o vedu.

A/D prevodník **vzorkuje** analógový signál do číslicovej podoby. Inými slovami: v diskretných časových intervaloch (akoby pulzoch) meria úroveň analógového signálu (zväčša napätia, ktoré vzniká zmenou odporu mikrofónu pri zmene tlaku vzduchu v jeho okolí) a s danou presnosťou zaznamenáva, **kvantuje** jej hodnotu (teda zaokrúhľuje spojité signály na diskretné číslo). Zdanlivo ide o moduláciu (konvolúciu) pôvodného signálu signálom s pravidelnou pulznou charakteristikou, z čoho pochádza aj označenie výsledného číselného zápisu takto získaného zvuku – **PCM** (*pulse-code modulation*, teda **pulzne kódová modulácia**), ktorý prvý raz popísal britský vedec Alec Reeves v r. 1937.

D/A prevodníky fungujú na opačnom princípe ako A/D prevodníky. Keďže jadrom našej práce sú digitálne formáty (a preto aj skôr digitalizácia zvuku, nie prevod do analógovej podoby), nebudeme sa im bližšie venovať.

Táto časť práce obsahuje tri odseky: 3.1.1 *Vzorkovanie* (str. 39), 3.1.2 *Kvantovanie* (str. 41) a 3.1.3 *Problémy digitálneho spracovania zvuku* (str. 43).

3.1.1 Vzorkovanie

Vzorkovanie sa opiera o práce dvoch vedcov z amerického Bell Laboratories (vedci z tohto podniku boli viac-menej zodpovední za väčšinu už spomínaných technológií, ako mikrofón a reproduktor). Harry Nyquist najprv ukázal, že pri použití

¹⁴ V podstate nejde o zvuk v pravom zmysle slova, skôr o nejaký symbolický spôsob jeho zápisu; predsa však o ňom budeme hovoriť ako o (digitálnom) zvuku.

zvolenej vzorkovacej frekvencie budú v signáli zaznamené všetky kmitočty menšie než polovičná hodnota vzorkovacej frekvencie (tzv. **Nyquistova frekvencia**). Problém však môže nastať pri reprodukcii tohto záznamu: v dôsledku aliasingu¹⁵ vznikajú kmitočty, ktoré v pôvodnom zázname neboli prítomné. Napr. pri vzorkovacej frekvencii 10 kHz má 3 kHz sinusoida reprezentáciu zhodnú so 7 kHz sinusoidou (obidve sú vzdialené o rovnaký diel od Nyquistovej frekvencie – zrkadlenie), 13 kHz sinusoidou (vzorkovacia frekvencia + 3 kHz) atď. Claude Shannon preto navrhol odfiltrovať kmitočty vyššie než Nyquistova frekvencia (ešte pred záznamom a aj pri samotnej reprodukcii), čím by sa zabránilo vzniku nežiaducich javov.

Frekvencie však nemožno od istého momentu len tak „odrezať“, elektronické súčiastky dovolili len spojitú frekvenčnú charakteristiku filtra, tzn. od Nyquistovej frekvencie by dochádzalo k útlmu, úplné vynulovanie signálu by však prišlo až s podstatným zvýšením kmitočtu signálu. Pre tento plynulý priebeh sa miesto vzorkovacej frekvencie 40 kHz (dvakrát hodnota najvyššej frekvencie, ktorú človek môže počuť) vyberali vyššie hodnoty – zväčša 44,1 kHz¹⁶. Frekvenčný filter mal potom pokles charakteristiky v rozsahu 21 - 23 kHz, teda prípadné deformácie v signáli vznikali už v nepočuteľnej oblasti ultrazvuku.

Skonstruovať takýto filter by však bolo neúmerne náročné. V rozsahu 20 Hz až 20 kHz treba zachovať aspoň približnú linearitu, potom musí nasledovať náhly pokles. Riešením je **prevzorkovanie** (angl. *oversampling*), teda zdanlivá zmena vzorkovacej frekvencie. Pomocou interpolácie možno zo vzorkovaného záznamu získať hodnoty v čase, keď ku vzorkovaniu nedošlo – použitím hodnôt susedných vzoriek. Ak sa takýmto spôsobom zvýši zdanlivá vzorkovacia frekvencia, hovoríme o **nadvzorkovaní** (angl. *upsampling*), ak zníži, ide o **podvzorkovanie** (*downsampling*). Vyfiltrovanie vysokých frekvencií teda môže prebiehať nasledovne: záznam navzorkovaný frekvenciou 44,1 kHz sa interpoláciou nadvzorkuje na štvornásobnú frekvenciu (176,4 kHz), použije sa nízkofrekvenčný priepust s takmer lineárnou frekvenčnou charakteristikou (v praxi neexistuje ideálny filter) pod 20 kHz a s poklesom od 20 kHz po 156,4 kHz. Parazitické frekvencie od 156,4 kHz do 176,4 kHz (prvý, zrkadlový aliasingový obraz počuteľného spektra) a vyššie

15 **Aliasing** je jav pri vzorkovaní signálu (obrazu, zvuku), keď sa rôzne druhy signálov stanú od seba neodlíšiteľnými. V tomto prípade by kmitočty presahujúce Nyquistovu frekvenciu mohli byť po vzorkovaní zhodné s frekvenciami nižšími, ktoré neboli v pôvodnom signáli. Ako príklad môže poslúžiť snímanie kamerou saka s husto mriežkovanou vzorkou – na televíznych obrazovkách sa toto sako zrazu akoby „hmýri“ čiernymi a bielymi útvarmi.

16 Zvláštnosťou je pôvod tejto frekvencie. Na začiatku osemdesiatych rokov boli na trhu prakticky len analógové médiá a jediný spôsob výmeny digitálneho (PCM) záznamu zvuku v analógovej forme bolo použiť štandardnú videokazetu (keďže takýto záznam potreboval vysoký frekvenčný rozsah – a videokazeta poskytovala vysoký dátový tok). Zariadenia (používané len v profesionálnej sfére) označované ako „PCM adaptory“ konvertovali analógový zvuk do PCM reprezentácie a späť do vysokofrekvenčnej analógovej formy. Do jedného obrazového riadku sa uložilo 96 bitov informácií, čo pri počte riadkov (294) a polí za sekundu (50) v štandarde PAL znamenalo 44 100 vzoriek na každý kanál pri 16-bitovej hĺbke. Pri návrhu formátu zvukového CD sa rátalo s tým, že medzi nahrávacím štúdiom a lisovňou CD sa nahrávka preniesie práve na takejto videokazete.

(napr. 176,4-196,4 kHz – druhý aliasingový obraz počuteľného spektra, 332,8-352,8 kHz – tretí, zrkadlový obraz atď.) teda budú zo signálu odstránené.

Podobne aj pri podvzorkovaní treba zamedziť vplyvu aliasingu. Riešením je odfiltrovanie frekvencií vyšších než novozamýšľaná Nyquistova frekvencia. Táto kombinácia nízkofrekvenčného priepustu a podvzorkovania sa nazýva tiež **decimácia**.

Zvyčajne sa používajú vzorkovacie frekvencie 44,1 kHz a 48 kHz, v novších, najmä profesionálnych zariadeniach je to 96 kHz, dokonca i 192 kHz (v kombinácii s 24-bitovým kvantovaním sa takýto zvuk nazýva **zvuk s vysokým rozlíšením**, angl. *High Definition Audio*, **HD Audio**).

3.1.2 Kvantovanie

Samotné zmeranie aktuálnej úrovne signálu (teda získanie **vzorky** v danom časovom okamihu, vyjadrenie jej kvantity) sa musí vykonať s istou presnosťou, keďže chceme získať jej číslícový zápis, aby sme tak dostali zvuk z analógového (spojitého) prostredia do digitálneho. Súčasný digitálne systémy sú binárne, preto je potrebné zaznamenať údaj binárnym číslom nejakej dĺžky. Táto dĺžka sa nazýva **rozlíšenie vzorky** alebo **bitová hĺbka vzorky**. Prvotné zvukové systémy mali rozlíšenie 8 bitov (teda zaznamenaná úroveň bola vyjadrená celými číslami v rozsahu -128 a 127, dokopy $256 = 2^8$ úrovni). Súčasný systémy používajú skôr 16 bitov (65 536 úrovni), lepšie 20 bitov (vyše milióna úrovni) a 24 bitov (vyše 16 miliónov úrovni), profesionálne dokonca 32 bitov (vyše 4 miliardy úrovni). Vo všeobecnosti možno tieto úrovne chápať ako rovnomerné delenie intervalu $\langle -1, +1 \rangle$, vyjadrujúceho hraničné hodnoty signálu podporované zariadením (**úplný rozsah**, angl. *full scale*). Zjavne platí, že každý ďalší bit zdvojnásobuje počet hodnôt v rozsahu, teda každý ďalší bit rozlíšenia znamená rozšírenie dynamického rozsahu o 6 dB¹⁷ (akustický výkon a intenzita zvuku sú kvadraticky závislé od akustického tlaku, teda aj napätia v obvode). Dostávame tak rozsah 48 dB pre osembitové, 96 dB¹⁸ pre šestnásťbitové zariadenia. (Teda najtichší zaznamenateľný zvuk má intenzitu -96 dB oproti úplnému rozsahu.) Treba spomenúť aj existenciu zápisov s pohyblivou desiatinnou čiarkou. Najrozšírenejší je tzv. *IEEE float 32* (ANSI/IEEE Std. 754-1985), čo je tridsaťdvabitové číslo (24 bitov na mantisu), existuje i formát s dĺžkou 64 bitov. Používa sa však len v špecifických prípadoch (kvalita 24 bit je normálne dostačujúca), hlavne pre ľahké spracovanie počítačom (v procesore sú registre na prácu s číslami tohto formátu).

Pri kvantovaní je aktuálna úroveň zaokrúhľená k najbližšej hodnote prítomnej v danom delení (rozlíšení). Vzniká chyba zo zaokrúhľenia – **kvantovací šum** (angl.

¹⁷ $2 \log 2 = 3,01$ dB.

¹⁸ Skutočný dynamický rozsah sa zväčša využíva len v záznamoch klasickej hudby či vystúpení virtuózov. V populárnej hudbe sa výsledný dynamický rozsah znižuje použitím akustickej kompresie.

quantization noise), keďže sa táto chyba prejavuje v každom (diskrétnom) časovom okamihu a vytvára vo všeobecnom prípade biely šum¹⁹. Ak ide o rovnomerné delenie (väčšina prípadov), odstup jednotlivých kvantovacích hodnôt (alebo rozdiel medzi nimi) sa nazýva **kvantovací krok** alebo **kvantovací interval**.

Historicky pôvodný spôsob kvantovania signálu je použiť **komparátor**, ktorý porovnáva úroveň vstupného signálu s referenčnou hodnotou. Ak je signál väčší, zvýši referenčnú hodnotu (počnúc najvyšším, najvýznamnejším bitom hodnoty – ten nastaví na jednotku), ak nižší, zníži ju (vynuluje bit). Porovnávanie sa teda realizuje mnohokrát za sekundu, v závislosti od rozlíšenia vzorky a vzorkovacej frekvencie. (Objavuje sa aj problém nestability signálu počas určovania hodnoty bitov jednej vzorky.)

Predstavme si však, že komparátor porovnáva vstupný signál s hodnotou predtým vyjadrenej úrovne signálu. Výsledkom sú dva stavy: 1, ak treba hodnotu oproti predošlej zvýšiť (keďže vstupný signál je už vyšší), 0, ak znížiť (naopak). „Predtým vyjadrenú úroveň“ pritom vypočítame ako súčet binárnych hodnôt výstupu vo zvolenom časovom okne. Výhody sú jasné – jednoduchá konštrukcia umožňuje vyššiu vzorkovaciu frekvenciu, navyše máme rýchlejší cyklus (jedno porovnávanie na vzorku miesto 8 alebo 16). Výsledkom kvantovania je však **tok jednobitových informácií**²⁰, teda teoreticky je kvantovací šum na úrovni -6 dB oproti úplnému rozsahu (½ bitu), čo je veľmi vysoké číslo! Treba však myslieť na to, že energia bieleho šumu je rovnomerne rozložená medzi frekvenciami menšími než Nyquistova frekvencia. Pri zvolení vysokej vzorkovacej frekvencie (zvyčajne 64×44,1 kHz, teda 2 822,4 kHz) je drvivá väčšina šumu prítomná v ultrazvukových frekvenciách. Na výstupe zariadenia je decimátor, ktorý podvzorkuje signál na zvolenú vzorkovaciu frekvenciu a rozlíšenie. Tento spôsob kvantovania, ktorý v súčasnosti (vlastne už od roku 1989) prevažuje nad pôvodným použitím komparátora, sa nazýva **delta-sigma modulácia**.

Navyše, pri kvantovaní sa používa aj **rozptyľovanie** (angl. *dithering* zo staroanglického *didderen*, triasť sa). Kvantovací šum nemusí byť totiž len šumom. Ak je vstup nejakým spôsobom korelovaný (vo väčšine prípadov je), kvantovací šum sa prejavuje skôr ako harmonické skreslenie. Rozptyľovanie znamená náhodne zmeniť úroveň signálu v rozsahu kvantovacieho kroku ešte tesne pred jeho zmeraním. Biely šum sa teda do zdroja dodáva pred kvantovaním, a to úmyselne. Že táto technika prináša dobré výsledky, vidno v oblasti grafiky. Miesto toho, aby bolo konkrétne číslo vždy zaokrúhlené tým istým smerom, je toto zaokrúhľované medzi susedné hodnoty tak, že štatisticky je interpretované svojou pôvodnou hodnotou. Pri technike rozptyľovania možno využiť i spätnú väzbu a nasledujúcu dodanú chybu upraviť

19 P. odsek 2.3.1 *Druhy zvukov*, str. 21.

20 Štandardné techniky kódovania takýchto tokov zahŕňajú *Pulse Width Modulation* (PWM) a *Pulse Density Modulation* (PDM) odlišnej konštrukcie, predsa s možnosťou rovnakého dekódovania. Bližšie sa im nebudeme venovať.

podľa chyby predošlej vzorky, na čom je založené tzv. **tvorovanie šumu** (*noise shaping*), ktoré bližšie nebudeme rozoberať. Využíva sa v spomenutých delta-sigma modulátoroch, aby frekvenčnú charakteristiku kvantovacieho šumu ešte viac potlačilo smerom k ultrazvuku.

3.1.3 Problémy digitálneho spracovania zvuku

V digitálnom svete zvuku sa nepotrebujeme pasovať s problémami, ktoré sme mali vo svete analógovom, aj keď sa im v konečnom dôsledku úplne nevyhneme, keďže potrebujeme zvuk prevádzať z jedného „sveta“ do druhého. Práve pri tom prevode sa objavujú problémy preň špecifické.

Prvým je šum. Vzniká v každom obvode, preto aj v prevodníkoch (A/D aj D/A). Pôsobí nezávisle od kvantovacieho šumu.

Druhým je **časové chvenie**, nestálosť (angl. *jitter*). Časový odstup medzi jednotlivými pulzmi vzorkovacej frekvencie nie je konštantný (závisí od kvality kryštálového oscilátora v A/D či D/A prevodníku, no žiaden nie je dokonalý). Vznikajú preto skreslenia vo frekvenčnej charakteristike výsledku.

Ďalším je presnosť vzorkovacej frekvencie. Ak si dve digitálne zariadenia vymieňajú dáta, unikátnosť kryštálových oscilátorov spôsobuje **synchronizačné problémy**. Jedno zariadenie musí občas čakať na dáta toho druhého, čím vznikajú medzery v signáli, resp. nestíha za druhým zariadením – vtedy zas dochádza k zahodeniu časti dát, akoby preskakovaníu. Treba podotknúť, že tento problém sa dá riešiť systematicky vhodnými časovacími alebo synchronizačnými postupmi (na úrovni komunikácie zariadení), navyše sa vyskytuje zriedka (väčšinou prebieha spracovanie zvuku na jednom komplexnom zariadení, napr. počítači, bez nutnosti komunikácie).

Aspekty spomenuté v predošlých odstavcoch možno merať – a to je asi jediné, čo ich so skresleniami prítomnými v analógovom svete zvuku spája. Ostatné aspekty, „problémy“, sú skôr vecou špecifikácie (rozlíšenie vzorky, vzorkovacia frekvencia).

3.2 Digitálne kódovanie čísel

Výstupom všetkých algoritmov na kompresiu zvuku je **tok dát** (v angl. sa používa aj termín *bitstream*, teda tok binárnych čísel) podľa špecifikácie daného formátu, v ktorom sú zakódované informácie o prenášanom zvuku. Tieto informácie sú prevažne (no nielen) číselné. Dôvod je prostý – už digitalizácia zvuku, o ktorej sme hovorili v predošlej časti práce (3.1), je zároveň „kvantifikáciou“ zvuku, premenou vlastností zvuku (v tomto prípade zmeny tlaku vzduchu) na číselný zápis. (Inými slovami: zvuk sa z analógového zmenil na diskretný, v časovej i hodnotovej doméne.) Kompresný algoritmus môže analyzovať digitalizovaný zvuk z viacerých hľadísk, no jeho výstupom (popri rôznych iných dátových položkách, určených na úspešnú dekompresiu) je v prvom rade sled čísel, vyjadrujúci vlastnosti zvukového signálu. **Kódovanie**, teda zápis týchto čísel, je často rozhodujúcim faktorom, vplývajúcim na efektívnosť formátu, a teda aj jeho úspešnosť na trhu.

Táto časť práce pozostáva zo štyroch odsekov. Najprv čitateľovi predstavíme problematiku symbolického zápisu dát vo všeobecnosti (3.2.1 *Kompresia a kompakcia, redundancia a irelevancia*, str. 44), jadro potom budú tvoriť jednotlivé druhy kódovania, ktoré sú v súčasnom svete zvukových formátov hojne používané (3.2.2 *Huffmanov kód*, str. 46; 3.2.3 *Aritmetické kódovanie*, str. 47; 3.2.4 *Modifikácie Huffmanovho a aritmetického kódovania*, str. 48).

3.2.1 Kompresia a kompakcia, redundancia a irelevancia

Pod slovom **kompresia** budeme rozumieť postup transformácie súboru alebo toku dát do inej reprezentácie, ktorá je svojím objemom menšia ako pôvodná reprezentácia. (Podotýkame, že všetky dáta, o ktorých uvažujeme, sú v skutočnosti spracúvané a ukladané binárne, teda aj ich objem sa hodnotí ako veľkosť ich binárneho zápisu; ten sa v praxi realizuje skôr po osmiciach bitov, teda bajtoch. Pod pojmom reprezentácia myslíme skôr vnútornú štruktúru a význam týchto dát.) Platí, že z novej reprezentácie (skomprimovaných dát) možno pôvodné dáta, alebo ich aproximáciu, získať tiež nejakým algoritmom. Ak sme na základe návrhu kompresného algoritmu vždy schopní získať pôvodné dáta späť bez straty, hovoríme o **bezstratovej kompresii (zvukových dát)**, alebo len o **kompakcii**. Ak dochádza k nejakej strate, ide o **stratovú kompresiu (zvukových dát)** (čo je pôvodný význam slova kompresia).

Inými slovami, kompresia funguje na základe redukcie **redundancie** (teda nadbytočných dát – nižšie popisujeme kódovú a medzisymbolovú redundanciu) a / alebo **irelevancie** (dát, ktoré sú z nejakých dôvodov považované za nepodstatné – nižšie hovoríme o tzv. psychoakustickej redundancii). Ak dochádza k redukcii irelevancie, ide o nevratný proces, teda stratovú kompresiu, ináč o kompresiu

bezstratovú, vratný proces. Úspešnosť kompresie sa dá vyjadriť rôznymi spôsobmi, v tejto práci budeme používať skôr zápis jej neúspešnosti – pomeru veľkostí výstupného a vstupného súboru.

Pod **kódovou redundanciou** myslíme zápis dát, v ktorom je výskyt jednotlivých **vstupných symbolov** (zväčša uvažujeme o bajtoch alebo iných väčších informačných jednotkách) štatisticky nerovnomerne rozdelený, tzn. pravdepodobnosť výskytu niektorých symbolov je omnoho vyššia ako pre iné symboly. Riešením by bol algoritmus prevádzajúci vstupné symboly na **kódové slová** (tiež vyjadrené číselným zápisom) taký, že kódové slová prislúchajúce symbolom s malým výskytom v zdroji budú mať väčšiu dĺžku ako kódové slová prislúchajúce symbolom často používaným. To by prinieslo redukciu veľkosti zápisu dát. Toto kódovanie možno vyjadriť **slovníkom**, zoznamom kódových slov prislúchajúcich jednotlivým vstupným symbolom. **Entropia** (prvého rádu) je potom vyjadrenie neusporiadanosti systému, zároveň dolná hranica dĺžky binárneho zápisu. Vypočítame ju ako

$$H_2(P) = \sum_{i=0}^{n-1} p_i \log_2 p_i^{-1},$$

kde $P = \{p_0, p_1, \dots, p_{n-1}\}$ je rozdelenie pravdepodobností výskytu jednotlivých n symbolov v danom súbore dát (teda ich početnosť, ak ju poznáme). V nasledujúcich odsekoch si predstavíme niekoľko **entropických kódov**, teda kódov, ktoré využívajú kódovú redundanciu zdroja. Ide zároveň o kódy s variabilnou dĺžkou slova.

Pod **medzisymbolovou redundanciou** budeme rozumieť štatistickú závislosť dát. Prax ukazuje, že väčšina dát vykazujú nejakú štatistickú závislosť (**korelovanosť**) so sebou samým. Zvukové dáta nie sú výnimkou. V časti 2.3 sme hovorili o charakteristikách zvukov – tie sa prejavujú práve v korelovanosti. Ak znie vo zvukovom zázname napr. určitý tón, má graf zvuku nejaký charakteristický tvar, ktorý sa v čase opakuje. Pri kódovaní je dôležité popísať jestvujúcu korelovanosť a odstrániť ju zo zdroja, čím sa značne zredukuje množstvo redundantných dát. Presnejšie, definujme celkovú (binárnu) entropiu zdroja (v tomto prípade kanálu s pamäťou) ako limitnú hodnotu

$$H'_2(P) = \lim_{m \rightarrow \infty} \sum_{\substack{i_1, i_2, \dots, i_m \in \{0, \dots, n-1\}}}^{n-1} p_{i_1} p_{i_2} \dots p_{i_m} \log_2 \frac{(p_{i_1} p_{i_2} \dots p_{i_m})^{-1}}{m}.$$

Odstránenie korelovanosti (**dekorelácia**) znamená premeniť medzisymbolovú redundanciu na redundanciu kódovú (teda priblížiť entropiu prvého rádu celkovej entropii zdroja, ktorú znížiť nemožno)²¹. S touto sa potom možno vysporiadať entropickým kódovaním.

Pod **psychoakustickou redundanciou** chápeme prítomnosť dát, ktoré sú vzhľadom na používateľa (či už pre obmedzenie sluchu, alebo pre požiadavky kladené na reprodukciu) nepodstatné (irelevantné), resp. sú menej podstatné ako iné,

²¹ Zdroj: LIEBCHEN, Tilman. Realisierung einer verlustlosen Transformationscodierung zur Datenkompression von Mono- und Stereo-Audiosignalen. 1998. str. 4.

dominantné. Stratová kompresia je vždy kompromisom medzi kvalitou a výslednou veľkosťou, predsa však existujú oprávnené prípady, kedy možno dáta redukovať (napr. frekvencie nad 20 kHz, alebo zbytočné 24-bitové kvantovanie vzoriek, keď je zvuk reprodukován malým reproduktorom na mobile atď.)

3.2.2 Huffmanov kód

Entropické kódovanie, ktoré v r. 1952 navrhol Američan David A. Huffman²², sa stalo synonymom pre efektívny kód odstraňujúci redundanciu – v prípade kódovej redundancie je dokonca optimálny. Mnohé ďalšie formy kódovania, ktoré si predstavíme neskôr, z neho akýmsi spôsobom vychádzajú.

Proces kódovania prebieha v dvoch krokoch. Podľa možnosti sa vykoná prvý „**prechod**“ vstupnými dátami – zistenie početnosti vstupných symbolov. Následne sa zostaví tabuľka kódov (tzv. kódový strom). Všetkým vstupným symbolom sa na začiatku priradí „aktívny“ vrchol. Pre dva aktívne vrcholy s najmenšou frekvenciou (početnosťou, pravdepodobnosťou výskytu) sa vytvorí nový aktívny vrchol, s ktorým sa spoja; jeho frekvencia je rovná súčtu pôvodných vrcholov, teraz už vyradených zo zoznamu aktívnych vrcholov. Nový vrchol je stromom reprezentujúcim pôvodné dva vrcholy, pričom číslo 0 určuje prvý z nich, číslo 1 ten druhý. Znova sa vyberú dva aktívne vrcholy s najmenšou frekvenciou a postup sa opakuje, kým nevznikne jeden súvislý strom. Jediný aktívny vrchol je potom koreň, každý list je vrcholom vyjadrujúcim vstupný symbol, pričom kódové slovo, ktoré mu je priradené, je binárny zápis „cesty“ (sled núl a jednotiek – hodnoty bitov) od koreňa k nemu. Dá sa ukázať, že tento kód je **úplný**, tzn. pre ľubovoľnú postupnosť bitov (núl a jednotiek) je táto buď prefixom práve jedného kódového slova, alebo práve jedno kódové slovo je jej prefixom. (Alternatívou k postupu uvedenému v tomto odstavci je použiť statický kódový slovník – v tom prípade možno rovno pristúpiť k druhému kroku, no optimálnosť kódu závisí od zhody predpokladanej a skutočnej distribúcie vstupných symbolov.)

Druhým krokom je samotné kódovanie, realizované pri druhom „**prechode**“ dátami. Výsledkom kódovania je kódový slovník, teda tabuľka kódových slov priradených k vstupným symbolom (resp. frekvencie vstupných symbolov, z ktorých ju možno skonštruovať) a samotné zakódované dáta.

Dá sa ukázať, že priemerná dĺžka kódového slova výstupu (pri zachovaní rozdelenia pravdepodobností, na základe ktorej bola vypočítaná tabuľka kódu) je nanajvýš o jednotku horšia oproti binárnej entropii. Pri vyššej entropii zdroja je to veľmi dobrý výsledok, pri nízkej je to nebezpečné. Ako vidno, medzisymbolovú redundanciu Huffmanov kód vôbec nevyužíva. Prítom napr. celková binárna entropia zdroja (teda entropia, ktorá ráta aj s medzisymbolovou redundanciou) pre anglický

22 HUFFMAN, David A. A Method for the Construction of Minimum-Redundancy Codes. 1952

jazyk je 1 bit na znak²³! Riešením oboch problémov by mohlo byť rozšírenie Huffmanovho kódu (združovanie vstupných symbolov), alebo použitie efektívnejšieho, napr. aritmetického kódovania.

3.2.3 Aritmetické kódovanie

Toto kódovanie oddeľuje pravdepodobnostný model zdroja od samotného procesu kódovania²⁴. To nám umožňuje prejsť od statického riešenia k dynamickému (tiež „spätne adaptívnemu“) riešeniu (tzn. dáta stačí spracovať len raz – v jednom „prechode“ dochádza priamo ku kódovaniu; navyše, so zmenou rozdelenia pravdepodobnosti vstupných symbolov a zmenou medzisymbolovej redundancie sa menia aj kódové slová – rovnako na strane kodéra i dekodéra). Ďalším pozitívom je, že sa vstupným symbolom akoby pridružuje neceločíselný počet bitov kódového slova – vďaka tomu každý ďalší vstupný symbol môže predĺžiť výstup kodéra o menej než jeden bit.

Pre jednoduchosť si popíšeme len adaptívny model prvého rádu (tzn. model bez využitia medzisymbolovej redundancie). Na začiatku máme základné rozdelenie pravdepodobnosti vstupných symbolov $P = \{p_1, p_2, \dots, p_{n-1}\}$ (môže byť napr. $p_i = 1/n$, alebo môže vychádzať z vlastností zdroja). Rozdelením intervalu $\langle a, b \rangle$ budeme rozumieť lineárne zobrazenie intervalu $I = \langle 0, p_0, p_0 + p_1, \dots, p_0 + p_1 + \dots + p_{n-1} \rangle$ na interval $\langle a, b \rangle$. Nech je počiatočný interval $\langle 0, 1 \rangle$. Rozdelíme ho podľa distribúcie symbolov P . Kód vstupného symbolu p_i vyjadruje obraz intervalu $\langle p_0 + p_1 + \dots + p_{i-1}, p_0 + p_1 + \dots + p_{i-1} + p_i \rangle$ (teda symbol p_0 je vyjadrený obrazom intervalu $\langle 0, p_0 \rangle$, symbol p_1 obrazom intervalu $\langle p_0, p_0 + p_1 \rangle$ atď.). Po zakódovaní prvého vstupného symbolu získavame interval, ktorý opäť rozdelíme podľa distribúcie P . Keď dokončíme kódovanie vstupných dát, získavame dve reálne čísla – krajné hodnoty intervalu vyjadrujúceho zakódované dáta. Vstupné dáta teraz možno vyjadriť ako desatinný rozvoj reálneho čísla (ktoré je logicky v rozsahu $\langle 0, 1 \rangle$) nachádzajúceho sa v tomto intervale. Distribúciu P následne aktualizujeme podľa spracovaných symbolov.

Tento popis je zjednodušený. Zdanlivo potrebuje nekonečnú presnosť počítania. V praxi sa však používajú rôzne metódy, ako aritmetické kódovanie realizovať pri obmedzených schopnostiach počítačov. Ráta sa len s istou presnosťou, pričom ak sú nejaké číslice (teda bity – binárny aritmetický kodér) výsledku (teda všetkých čísel v intervale) nemenné, tieto sa zapíšu na výstup, „zabudnú sa“ a miesto nich získavame nové bity na zvýšenie presnosti. Druhý problém je, že algoritmus by musel preniesť

23 SHANNON, Claude E. A mathematical theory of communication. Bell System Technical Journal 27. 1948. In MACKAY, D. J. C. Information Theory, ... 2005. str. 113.

24 OLEJÁR, D., STANEK, M. Úvod do teórie kódovania. 2006. str. 59 (kapitola 4.4). Alebo: MACKAY, D. J. C. Information Theory, ... 2005. str. 111 (Chapter 6.2). Obe spomenuté kapitoly podrobnejšie rozoberajú konštrukciu aritmetických kódov.

informáciu o počte zakódovaných symbolov (dekódovanie by ináč prebiehalo donekonečna). V praxi sa to skôr rieši dodaním symbolu <EOD> (*end of data*) medzi vstupné symboly, ktorý indikuje ukončenie prenosu.

Spomínanými výhodami aritmetického kódovania sú vysoká úspešnosť a najmä dynamickosť, nevýhodou potom časová náročnosť a čiastočná patentovanosť v rámci USA²⁵. Ide o **prúdové kódovanie**, keďže je vhodné na kódovanie dát prichádzajúcich sekvenčne, v reálnom čase, na rozdiel od Huffmanovho kódu, ktorý potrebuje dva prechody, aby zostavil tabuľku vhodnú pre zdroj. Aritmetické kódovanie vďaka adaptívnosti žiadnu tabuľku prenášať nemusí.

3.2.4 Modifikácie Huffmanovho a aritmetického kódovania

Kódovania spomenuté v predošlých dvoch odsekoch sú veľmi kvalitné, predsa však akýmsi spôsobom nepohodlné. V tomto odseku si predstavíme tri ich modifikácie alebo špeciálne prípady.

Začneme **dynamickým** (resp. **adaptívnym**) **Huffmanovým kódom**. Vysvetlíme si však najprv tzv. súrodenecké kritérium: očísľujme vrcholy v binárnom strome kódovej tabuľky vzostupne zľava doprava, zdola (od najväčšej hĺbky) hore. Každý vrchol spomedzi vrcholov rovnakej hĺbky má číslo menšie než ľubovoľný vrchol o úroveň vyššie (väčšie než ľubovoľný vrchol o úroveň nižšie). Zároveň má každý rodič väčšie číslo než obaja jeho potomci. Ak platí, že s rastúcim číslom vrchola rastie aj frekvencia znaku, ktorý reprezentuje, tento binárny strom (vyjadrujúci kódový slovník) spĺňa súrodenecké kritérium. Princíp dynamického Huffmanovho kódu je nasledovný: Na začiatku je tabuľka prázdna a existuje spôsob dodávania nových symbolov. Prijatím už známeho znaku sa zvýši jeho frekvencia. Aktuálny kódový slovník sa mení priebežne pre každý prijatý symbol, nezostavuje sa však celý nanovo, ale v každom kroku sa len upraví tak, aby spĺňal súrodenecké kritérium. Odpadá tak potreba dvojnásobného prechodu zdroja (na zistenie frekvencie pred kódovaním) a prenosu tabuľky spolu so zakódovanými dátami. Efektívnosťou sa tento kód podobá aritmetickému kódovaniu.

Ďalšou spomenutou modifikáciou je **rozsahové kódovanie** (angl. *range encoding*), ktoré je len transformáciou aritmetického kódovania do sféry celých čísel. Miesto intervalu $(0,1)$ sa používa interval medzi nulou a ohraničene „najväčším“ kladným číslom (binárne samé jednotky). Ako pribúdajú známe (nemenné) číslice na výstupe, rozširuje sa aj aktuálny rozsah dodaním ďalších číslic (analógia so zvyšovaním presnosti v aritmetickom kódovaní). Výhodami sú menšia výpočtová náročnosť a nezaťaženosť patentmi. Viac informácií možno nájsť v článku MARTIN, 1997.

25 Väčšinu patentov vlastní firma IBM. Pozri napr. http://en.wikipedia.org/wiki/Arithmetic_coding.

Do tretice, **Golombovo kódovanie** nesie meno po svojom autorovi, známom americkom profesorovi Solomonovi Wolfovi Golombovi. Je určené na kódovanie zdrojov s malými kladnými (nenulovými) číslami, pričom s rastúcim číslom n klesá pravdepodobnosť jeho výskytu. (V ideálnom prípade ide o geometrické rozloženie – pravdepodobnosť výskytu čísla n je $P(n)=(1-p)^{n-1}p$ pre nejaké $0 \leq p \leq 1$.) Schéma používa nastaviteľný parameter b . Ak pravdepodobnosť vyšších čísel klesá rýchlo, b je nízke, ináč je vyššie. (Presnejšie: volí sa také b , aby

$(1-p)^b + (1-p)^{b+1} \leq 1 < (1-p)^{b-1} + (1-p)^b$.) Vypočíta sa $q = \left\lfloor \frac{n-1}{b} \right\rfloor$, $r = n - qb - 1$ a $c = \lceil \log_2 b \rceil$. Kód čísla n je potom unárny zápis čísla $q+1$ (tzn. q binárnych jednotiek a jedna nula – bity s vyššou hodnotou), za ktorým nasleduje binárny zápis čísla r (bity s nižšou hodnotou). Číslo r je zapísané „orezaným binárnym kódovaním“, pri ktorom sa na prvých $2^{c+1} - r$ hodnôt použijú c -bitové čísla (vzostupne od nuly), na zvyšné sa použijú $c+1$ -bitové čísla. (Např. pre $b=5$ a $c=2$ sú to postupnosti 00, 01, 10, 110, 111.) Ak chceme kódovať celé čísla, stačí vytvoriť ich bijektívne zobrazenie na množinu kladných čísel. Pôvodný autorov popis kódovania možno nájsť v článku GOLOMB, 1966²⁶.

V prípade Golombových kódov takých, že b je mocnina dvojky, hovoríme o **Riceovom kódovaní** (resp. Golombove-Riceove kódy). Výhodou je ľahká implementovateľnosť v prostredí počítačov (miesto delenia a násobenia sa používajú binárne posuny).

26 Text je však zložitý a popretkávaný vtipným príbehom. Na internete možno nájsť aj jednoduchšie opisy (např. http://urchin.earth.li/~twic/Golomb-Rice_Coding.html a i.).

3.3 Princípy stratovej kompresie zvuku

Ako sme si spomínali na začiatku predošlej časti práce (3.2 *Digitálne kódovanie čísel*, str. 44), existuje viacero algoritmov na kompresiu zvukového záznamu v PCM formáte. Tieto algoritmy sa delia na stratové a bezstratové²⁷. Vedomosti zo psychoakustiky priam vnukajú možnosť zapisovať zvukové dáta takým spôsobom, aby sme síce zachovali subjektívnu kvalitu výsledného zvukového záznamu, predsa však zanedbaním precíznosti hodnôt získali významnú úsporu úložného priestoru (resp. dátového toku). Možno hovoriť o psychoakustickej redundancii pôvodných dát. Vďaka vysokej kompresii (ktorú možno zvoliť ľubovoľne podľa potreby) a v súčasnej dobe už zanedbateľnej výpočtovej náročnosti sú práve stratové kompresné formáty v centre záujmu používateľov i firiem.

V tejto časti práce si objasníme spoločné črty stratovej kompresie zvuku a prístupy k nej. Pozostáva z piatich odsekov: 3.3.1 *Diskrétne ortogonálne transformácie* (str. 50), 3.3.2 *Fourierova transformácia, kosínusová transformácia* (str. 51), 3.3.3 *Ortogonálne transformácie v praxi* (str. 52), 3.3.4 *Vektorové kvantovanie (VQ)* (str. 56) a 3.3.5 *Kompresia stereofonického záznamu* (str. 57).

3.3.1 Diskrétne ortogonálne transformácie

Úlohou ortogonálnych transformácií je zmenšiť korelovanosť dát. Z praktických dôvodov ide o bloky dát určitej veľkosti, ktoré možno chápať aj ako vektory (napr. dĺžky n). Vzájomné závislosti v hodnotách ich prvkov sa dajú vyjadriť matematicky, a to použitím báz, ktoré tieto závislosti budú odzrkadľovať. Ak vektory nebudeme vyjadrovať v jednotkovej báze, teda v báze e_1, e_2, \dots, e_n (kde $e_i = \{0, \dots, 0, 1, 0, \dots, 0\}$ s jednotkou na i -tom mieste), čo je princíp PCM, ale v báze obsahujúcej napr. sinusoidy rozličných frekvencií (ktoré viac odzrkadľujú štruktúru zvuku), môže dôjsť k značnej redukcii medzisymbolovej redundancie. Táto zmena vyjadrenia vektorov je vlastne zmenou súradníc, známou z algebry, ktorú možno vyjadriť **maticou prechodu** A obsahujúcou zápis pôvodnej (jednotkovej) bázy vyjadrenej v novej báze. Keďže ide o bázu, jej jednotlivé vektory sú kolmé, transformácia súradníc je teda špeciálnym prípadom **ortogonálnej transformácie**.

Presnejšie, pre vstupný signál vyjadrený n -rozmerným (stĺpcovým) vektorom x a transformačnú maticu $A_{n \times n}$ (ide o maticu prechodu medzi jednotkovou bázou a novou bázou) vypočítame dekorelovaný vektor ako $t = Ax$. Ak je matica A **ortonormálna** (jej vektory majú jednotkovú veľkosť a platí $A^{-1} = A^T$), nedochádza k zmene energie vektorov, teda $t^T \cdot t = x^T \cdot x$. Prvky vektora t možno vyjadriť aj z jednotlivých riadkov a_0, a_1, \dots, a_{n-1} matice A ako

²⁷ Porov. odsek 3.2.1 *Kompresia a kompakcia, redundancia a irelevancia*, str. 44.

$$t_k = \sum_{j=0}^{n-1} x_j a_{k,j} \text{ pre } k=1, \dots, n.$$

Inverzná transformácia je potom $x = A^{-1}t$.

3.3.2 Fourierova transformácia, kosínusová transformácia

Fourierova transformácia je špeciálny prípad ortogonálnej transformácie. Nesie meno po francúzskom matematikovi a fyzikovi menom Jean Baptiste Joseph Fourier, ktorý na začiatku 19. storočia ukázal, že každú funkciu periodickú s istou frekvenciou možno vyjadriť ako súčet sínusových vln harmonických násobkov tejto frekvencie s rôznymi amplitúdami a počiatočnými fázami. V diskretnej doméne času (teda nespojitej, čo je náš prípad) sa využíva **diskrétna Fourierova transformácia**, ktorou možno dekomponovať vstupný signál na frekvenčné zložky. Tieto sú vyjadrené komplexnými číslami, ktoré zachytávajú nielen amplitúdu (veľkosť prvku), ale aj fázu vlnenia danej frekvencie (vyjadrenej ako arkus tangens podielu reálnej a imaginárnej zložky prvku). Dopredná a inverzná transformácia majú tvar:

$$t_k = \sum_{j=0}^{n-1} x_j e^{-2\pi i j k / n}, \text{ resp. } x_k = \frac{1}{n} \sum_{j=0}^{n-1} t_j e^{2\pi i j k / n}$$

(pričom podľa Eulerovej formuly $e^{is} = \cos(s) + i \sin(s)$). Táto transformácia má zjavne výpočtovú náročnosť $O(n^2)$. Existuje však mnoho algoritmov patriacich do skupiny FFT (*fast Fourier transform*), ktoré ju redukovujú na $O(n \log n)$. Ich presný popis prekračuje rámec tejto práce.

Vstup je v našom prípade podmnožinou oboru reálnych (nie komplexných) čísel. Pozornému čitateľovi preto neušlo, že najvyššia frekvencia skúmaná algoritmami FFT je zhodná s frekvenciou vzorkovania, teda dvojnásobne prekračuje Nyquistovu frekvenciu²⁸. Ak vieme, že v zdroji nie sú takéto vysoké frekvencie prítomné, dochádzame k záveru, že merané veličiny sú prejavom aliasingu. Inými slovami, aspoň polovica výsledných dát je (pre reálny vstup) redundantná. Jedným zo spôsobov, ako túto redundanciu odstrániť, je miesto sínusu a kosínusu (ktoré majú rovnaký tvar a líšia sa len fázou) pracovať len s jedným z nich, a to tiež v obore reálnych čísel (čím stratíme informáciu o fáze frekvenčných zložiek). Aby sme docielili postupný rast skúmaných frekvencií od najnižšej po Nyquistovu frekvenciu, rozšírime vstupný blok údajov na dvojnásobnú veľkosť.

Diskrétna kosínusová transformácia (DCT, angl. *discrete cosine transform*) je špeciálny prípad diskretnej Fourierovej transformácie; keďže kosínus je párna funkcia, DCT je vhodné predovšetkým na frekvenčný rozklad párných funkcií. Existuje niekoľko rozšírených druhov definícií DCT (najznámejšie sú označované DCT I, DCT II, DCT III, DCT IV). V kompresii obrazu je najpoužívanejšia DCT II

²⁸ P. odsek 3.1.1 *Vzorkovanie* (str. 39).

(označovaná často len ako „DCT“), ktorej inverzná funkcia je DCT III pre násobená vhodným číslom (často len „IDCT“). Pri kompresii zvuku sa však viac osvedčila *modifikovaná DCT* (MDCT), založená na DCT IV. Tá funguje na princípe prekrývania okien (skúmaných blokov vzoriek) – každá vzorka patrí do dvoch okien, pričom druhá polovica predchádzajúceho okna sa prekrýva s prvou polovicou nasledujúceho okna. Počet koeficientov je však oproti veľkosti okna polovičná, čo toto zdvojnásobenie dát kompenzuje. Výpočet sa realizuje nasledovne:

$$\text{MDCT: } t_k = \sum_{j=0}^{2n-1} a_j x_j \cos \left[\frac{\pi}{n} \left(j + \frac{1}{2} + \frac{n}{2} \right) \left(k + \frac{1}{2} \right) \right]$$

$$\text{IMDCT: } y_k = \frac{1}{a_k} \frac{2}{n} \sum_{j=0}^{n-1} t_j \cos \left[\frac{\pi}{n} \left(k + \frac{1}{2} + \frac{n}{2} \right) \left(j + \frac{1}{2} \right) \right].$$

V prípade IMDCT získame dva výsledky pre každú vzorku – tie jednoducho sčítame a získame želané x_j . Táto technika prekrývania sa tiež nazýva TDAC (angl. *time domain aliasing cancellation*, teda anulovanie aliasingu v časovej doméne). Zmysel škálovacieho vektora $a = [a_0, a_1, \dots, a_{2n-1}]$ je práve v prekrývaní okien – tieto faktory upravujú výsledky po aplikovaní MDCT tak, aby výsledná energia súčtu oboch výsledkov z IMDCT bola jednotková (presnejšie pre konštantnú veľkosť bloku n platí $a_i^2 + a_{i+n}^2 = 1$ – Princenova-Bradleyho podmienka).

V praxi sa používa symetrický škálovací vektor (teda párna funkcia), ktorý sa líši od formátu k formátu. Nižšie uvedené príklady používajú formáty MP3 a MPEG-2 AAC (a_i) a Vorbis (b_i). AC-3 používa tzv. Kaiser-Besselovo derivované okno (s netriviálnym výpočtom), rovnako ho môže použiť i MPEG-4 AAC.²⁹

$$a_i = \sin \left[\frac{\pi}{2n} \left(i + \frac{1}{2} \right) \right], \quad b_i = \sin \left(\frac{\pi}{2} \sin^2 \left[\frac{\pi}{n} \left(i + \frac{1}{2} \right) \right] \right).$$

Prvky výsledného vektora t získané z IMDCT (korešpondujúce s amplitúdami skúmaných frekvencií³⁰) sa nazývajú **koeficienty**.

3.3.3 Ortogonálne transformácie v praxi

Treba si uvedomiť, že aj keď sú vstupné dáta ortogonálnych transformácií (v tomto prípade hovoríme najmä o MDCT) celočíselné, výsledné koeficienty budú vo všeobecnosti reálne. Ak teda chceme pracovať s frekvenčnými zložkami, musíme ich zaokrúhliť, kvantovať. Možnosť presnej rekonštrukcie pôvodných signálov teda nepripadá do úvahy. Koeficienty možno kvantovať s rôznou presnosťou, v závislosti od frekvencie, vzhľadom na psychoakustiku a želaný dátový tok. V praxi sa používa

29 Obširnejší rozbor podmienok, ktoré musí každý škálovací vektor spĺňať, možno nájsť napr. v SPORER, Th. et al. The Use of Multirate Filter Banks... 1992.

30 Koeficienty nevyjadrujú priamo amplitúdu sinusoidy danej frekvencie, ale reprezentujú intenzitu, s ktorou je signál daného kmitočtu v zázname „prítomný“, vzhľadom na konkrétnu metódu skúmania (v tomto prípade DCT).

logaritmické kvantovanie – rozlíšenie hodnôt blízko malých čísel je presnejšie než v prípade vyšších hodnôt. Ortogonálne transformácie sú teda viac než vhodné na stratovú kompresiu signálu.

Ďalším aspektom je „statický čas“ riešenia. Zatiaľ sme hovorili o jednom spracúvanom bloku dát, problémom je však sekvenčnosť zvukových dát; tie vo všeobecnosti nevykazujú periodickosť zhodnú s veľkosťou bloku. Riešením je napr. prekryvanie okien, ako sme si ho spomenuli vyššie. Objavuje sa problém zachovania fázovej informácie (ktorá v koeficientoch nie je vyjadrená) – možnosťou je napr. stanoviť počiatočnú fázu v „čase 0“, zabezpečiac tak súvislosť fázy pri prechode oknami pre ľubovoľnú frekvenciu. Zvolenie veľkosti okna tiež nie je jednoduché. Malé okno znamená dobré rozlíšenie v čase (ľahko sa vyjadří presný čas, kedy sa v zázname vyskytol daný zvuk), ale slabý frekvenčný rozsah (frekvencie s menšou vlnovou dĺžkou než veľkosť okna sa vo výpočte MDCT nezohľadňujú). Veľké okno naopak zvyšuje frekvenčný rozsah za cenu straty časovej informácie. Riešenia sú rôzne, napr. meniť adaptívne veľkosť okna vzhľadom na prevažujúce kmitočty, alebo rozdeliť frekvenčné pásmo na viac častí a každé spracúvať osve, s osobitnými veľkosťami okna. Výsledkom je lepšia presnosť výsledku vo frekvencii i čase, transformácia však ostáva naďalej silno stratová. Medzi známe prejavy tejto stratovosti patrí jav nazývaný **zvlňenie** (angl. *ringing*). Frekvenčnou charakteristikou signálu s náhlou zmenou priebehu je impulz so širokým frekvenčným spektrom a amplitúdou. Pri spätnej rekonštrukcii z frekvenčných zložiek sa v signáli objavajú predtým neprítomné zvlňenia pred a za impulzom (tzv. Gibbsov fenomén – znemožňuje napr. presnú rekonštrukciu štvorcovej vlny). Keďže okná pokrývajú istý časový úsek, frekvenčná dekompozícia a spätná kompozícia spôsobia utlmenie impulznej charakteristiky pôvodného signálu a jeho rozloženie na tento časový úsek. Príbuzným negatívnym fenoménom je aj tzv. **predozvena** (angl. *pre-echo*). Tiež sa prejavuje pri výskyte signálu s vysokou energiou. Kodér, v snahe udržať vyrovnaný dátový tok, zväčša znižuje kvalitu kvantovania pre celý korešpondujúci rámec. Výsledkom je šum prítomný v celom rámci; signál s vysokou energiou, ktorý by ho maskoval, sa však objaví až od istého časového okamihu. Obe spomenuté deformácie teda vznikajú časovo pred i za impulzom – hovorí sa však len o „predozvene“. Dôvodom je nižšia citlivosť sluchu na skreslenie za hlasným zvukom (impulzom), tzv. dopredné maskovanie³¹. Oba fenomény sa prejavujú najmä pri nízkych dátových tokoch a impulzných zvukoch, ako sú potlesk, bubny, výbuchy, kastanety ap.

Najväčší problém je náročnosť výpočtu ortogonálnych transformácií. Je vhodné počítať ich pomocou matic (podobne ako pri kompresii obrazu), pri relevantných veľkostiach okien (napr. MP2 používa 512 vzoriek, čo je pri 44,1 kHz len 11,6 ms!) je to však veľmi náročné. Riešením môže byť **subpásmové kódovanie**. Pomocou vhodne zvolených digitálnych filtrov, zoskupených do tzv. **banky filtrov**, možno rozdeliť pôvodné skúmané okno na frekvenčné pásma (napr. 32 frekvenčných pásiem

31 Angl. post-masking. P. odsek 2.2.4 *Maskovanie zvukov*, str. 17.

rovnomerne deliacich počutelný frekvenčný rozsah zdroja – **subpásmová analýza**). Subpásma sú následne kriticky podvzorkované (ak je 32 subpásiem, tak faktorom 32 – pri veľkosti okna³² 512 vzoriek sa tak subpásma podvzorkuje na 16 vzoriek), čo umožní odstrániť zbytočnú redundanciu: všetky pôvodné frekvencie v subpásme ostanú po podvzorkovaní zachované vďaka aliasingu, v nepárnych subpásmach však budú „prevrátené“ – najmenšia frekvencia sa stane najvyššou a naopak. Na subpásma, vďaka malej veľkosti, možno neskôr efektívne aplikovať ortogonálnu transformáciu. Subpásmová analýza a **syntéza** (zloženie pôvodného zvuku zo subpásiem) vnáša do zvuku skreslenie spôsobené nedokonalosťou použitých filtrov. Navyše, rovnomerné rozdelenie frekvenčného rozsahu nekorešponduje s vnímaním frekvenčných rozdielov ľudským sluchom³³. Výhodou je možnosť osobitného spracovania signálu podľa jeho frekvenčných zložiek, a to bez transformácie z časovej do frekvenčnej domény. Najčastejšie používanými filtermi na subpásmovú dekompozíciu signálu sú **kvadrátúry zrkadlový filter**, deliaci frekvenčné spektrum pôvodného signálu na dva podvzorkované signály (kriticky, teda na polovicu, pričom v prvom subpásme sa nachádza nižšia polovica frekvenčného spektra, v druhom vyššia), a všeobecnejší **polyfázový kvadrátúry filter**, deliaci spektrum na niekoľko uniformne rozdelených subpásiem. Ich presnejší popis a konštrukcia prekračujú rámec tejto práce.

Po frekvenčnej analýze nasleduje rozbor rámca. Identifikujú sa dominantné frekvenčné zložky, na základe psychoakustického modelu sa vytvorí krivka – **prah maskovania** (akoby maskovací signál prítomný v zázname) pre rôzne frekvencie. Frekvenčné zložky s intenzitami pod prahom maskovania (preto „nepočuteľné“, aspoň podľa rozhodnutia psychoakustického modelu toho-ktorého algoritmu) možno celkom zanedbať. Zložkám len mierne prekračujúcim prah maskovania (majú nízky **odstup signálu od maskovacieho signálu** – *signal to mask ratio*, **SMR**) možno priradiť nižšie rozlíšenie pri kvantovaní (snaha udržať nízky **odstup šumu od maskovacieho signálu** – *noise to mask ratio*, **NMR**). Prah maskovania z predošlého rámca sa po miernom utlmení zvykne použiť pri spracúvaní ďalšieho rámca, simulujúc tak časové maskovanie. Kvantované hodnoty sa kódujú vhodným kódovaním, od nízkych frekvencií k vysokým³⁴.

Ak má byť želaný dátový tok nízky, znižuje sa presnosť kvantovania koeficientov, resp. postupne sa úplne potláčajú frekvencie na okraji psychoakustického pásma vnímania (najmä vysoké frekvencie). Niektoré formáty umožňujú umelo zvýšiť amplitúdu skupiny frekvenčných zložiek alebo subpásma – toto zhlásenie (angl. *gain*) sa prenáša ako dodatočná informácia. Keďže kvantovanie

32 O oknách má zmysel hovoriť najmä pri psychoakustickej analýze záznamu. Samotné kódovanie však prebieha po rámcoch získaných z týchto okien – v prípade MP2 je to 384 vzoriek, teda kódované subpásma zaberá 12 vzoriek.

33 P. odsek 2.2.5 *Vnímanie frekvencie*, str. 18.

34 Vysokofrekvenčné zložky majú vzhľadom na povahu zvuku (p. časť 2.3 *Charakteristiky zvuku*, str. 21) redšiu distribúciu a nižšie amplitúdy (po kvantovaní často zaokrúhlené k nule alebo veľmi nízkym číslam). Navyše je ľudský sluch háklivejší na zachovanie hlbokých fundamentálnych frekvencií.

zväčša prebieha podľa logaritmickej mierky (teda vysoká presnosť pre malé čísla, nižšie rozlíšenie pre veľké čísla), výsledkom je zníženie dátového toku.

Zvyšovať „kvalitu“ zaznamenaného zvuku zas možno presnejším kvantovaním koeficientov, použitím viacerých okien pre rôzne frekvenčné pásma, osobitným kódovaním jednotlivých kanálov atď. Treba si uvedomiť, že nemožno zvyšovaním dátového toku dosiahnuť bezstratovosť riešenia, aj keď, vzhľadom na návrh algoritmu, možno ľubovoľne zvyšovať psychoakustickú kvalitu, posluchovú „zhodnosť“ s originálom, tzv. **priesvitnosť záznamu** (v angličtine sa používa termín *transparency*, ak už je skomprimovaný záznam sluchom neodlíšiteľný od pôvodného, z ktorého stratovou kompresiou vznikol).

Medzi rôznymi ortogonálnymi transformáciami sa ujala predovšetkým DCT. Prax ukázala, že napriek svojej jednoduchosti ponúka najlepšie psychoakustické vlastnosti (i psychovizuálne, ak ide o kompresiu obrazu, kde sa tiež používa).

Schéma stratovej kompresie zvuku teda väčšinou vyzerá nasledovne (zjednodušený náčrt):



Obr. 3.1 – Schéma stratovej kompresie zvuku – po transformácii (*A*) nasleduje kvantovanie (*Q*), výsledné čísla sa kódujú entropickým kódom (*EC*)

Zvuk vo forme PCM (občas rozdelený na subpásma) je vstupným vektorom pre ortogonálnu transformáciu *A*. Na výstupné koeficienty sa aplikuje kvantovanie podľa potreby (na základe frekvenčnej a psychoakustickej analýzy zvuku), výsledok sa zakóduje entropickým kódom a pošle na výstup. Dekompresia funguje presne naopak: dekódovanie koeficientov, naškálovanie na pôvodné (reálne) hodnoty (so stratou presnosti), spätná transformácia (maticou A^{-1}) (prípadne nasleduje subpásmová syntéza) – na konci získavame dáta vo forme PCM.

S ortogonálnymi transformáciami a subpásmovým kódovaním úzko súvisí i tzv. *gapless playback* (dalo by sa preložiť ako súvislá, nepretržitá reprodukcia, záznam bez medzier). Veľkosť rámcov tej-ktorej kompresnej schémy nutne určuje najmenšiu časovú jednotku skomprimovaného záznamu. Riešením je doplnenie ticha („vypchatie“, angl. padding) na začiatku alebo konci zvukového záznamu, s cieľom dosiahnuť dĺžku deliteľnú veľkosťou rámca. Prehrávanie sledu záznamov, ktoré predtým tvorili súvislý celok (napr. skladby na niektorých CD³⁵ s plynulým prechodom z jednej skladby do druhej), je tak zrazu prerušovaný tichom – preto je vo výsledných súboroch často uložená informácia o trvaní doplneného ticha, zaručujú tak *gapless playback*.

Na dôvažok treba spomenúť, že zvuk sa „neodohráva“ len v časovej (PCM) alebo frekvenčnej doméne (MDCT), ale v oboch súčasne. Použitie okien znamená

³⁵ CD však majú tiež najmenšiu jednotku – sektor dĺžky 588 vzoriek (1/75 sekundy).

skombinovanie oboch domén. Svojou prirodzenosťou by však **wavelety**³⁶ omnoho lepšie vystihovali charakteristiku skutočného signálu. V praxi sa však (zatiaľ) s waveletovými kompresnými postupmi na kódovanie zvuku nestretávame.

3.3.4 Vektorové kvantovanie (VQ)

Najjednoduchším spôsobom zápisu vektorovej informácie je vymenovanie jednotlivých prvkov vektora³⁷. Ak je však signál zmysluplná informácia, obsahuje nejaké charakteristické prvky. V predošlom texte sme považovali signál za kompozíciu základných tónov, sinusoid. Transformáciou sme získali súbor koeficientov, ktoré vyjadrovali „mieru“ zastúpenia tej-ktorej frekvenčnej zložky. Keďže sa pohybujeme v diskretnej doméne, možno sa na jednotlivé frekvenčné zložky (sinusoidy) pozerat' ako na vektory. Výsledný signál je kompozíciou týchto vektorov vynásobených príslušnými „mierami“ (koeficientmi). Na ortogonálne transformácie (aj MDCT) sa teda dá pozerat' ako na špeciálny prípad **vektorového kvantovania** (v skratke budeme označovať **VQ** z anglického *vector quantisation*).

Vo všeobecnosti sa na princíp VQ dá pozerat' nasledovne: majme kódový slovník – súbor vektorov, ktorými aproximujeme vstupný vektor (signál). Rozmer vektorov je n , preto ich možno chápať ako body v n -rozmernom priestore. Skonstruujeme Voronoiove diagramy z kódových vektorov. Vstupný vektor spadá do jedného z diagramov, prislúchajúceho konkrétnemu kódovému vektoru. Preto mu je (vzhľadom na zvolenú metriku) najviac podobný (v danom priestore je tento kódový vektor vstupnému vektoru zo všetkých najbližší).

Pri kompresii zvuku metódou VQ sa obyčajne tento postup aplikuje na vektor koeficientov³⁸, teda vyjadrenie signálu vo frekvenčnej doméne (kde je oproti časovej doméne signálu ľahšie badať štruktúru a korelovanosť). Z kódového slovníka sa vyberie vektor, ktorý najlepšie reprezentuje vektor koeficientov. Keďže slovník oproti pôvodnému vektorovému priestoru obsahuje len obmedzený počet vektorov (vďaka čomu nastáva samotná kompresia, teda úspora v popise pôvodných dát), vzniká skreslenie (akási forma kvantovacieho šumu) – reziduálny signál, teda rozdiel medzi vstupným vektorom a vybraným kódovým vektorom. V prípade potreby možno ďalším krokom (tzv. prechodom) toto skreslenie zmenšiť, ak reziduálny signál znova podrobíme vektorovému kvantovaniu.

36 Jednotky signálu nesúce informáciu o frekvencii i čase signálu – teoreticky sú teda vhodnejšie na dekompozíciu zvuku.

37 Ak uvažujeme o vektore ako o výsledku vzorkovania (spojitého) signálu, ide vlastne o PCM reprezentáciu.

38 Vektor koeficientov sa zvykne normalizovať, aby jeho špičková hodnota bola zhodná s úplným rozsahom (najvyššou vyjadriteľnou hodnotou). Ak je tak, sú tomu prispôsobené i vektory v kódovom slovníku – každý z nich je normalizovaný. Informácia o potrebnom zosilnení signálu, samozrejme, tiež tvorí informáciu potrebnú na spätnú rekonštrukciu pôvodného vektora.

3.3.5 Kompresia stereofonického záznamu

V predošlom texte sme uvažovali o zvuku ako o monofonickom zázname. V praxi však používame stereofonický záznam, teda záznam dvoch (resp. viacerých) kanálov. Na podobnosť signálov v jednotlivých kanáloch sa dá pozerat' ako na (tzv. medzikanálovú) redundanciu, čo sa dá využiť pri kompresii. Máme na výber štyri prístupy:

- Kanály môžeme podrobiť **nezávislému kódovaniu kanálov** (angl. *independent stereo coding*), teda medzikanálovú redundanciu vôbec nevyužijeme.
- Vzhľadom na podobnosť zvukov v jednotlivých kanáloch sú koeficienty (intenzity frekvenčných zložiek vstupu) po nezávislom kódovaní kanálov zväčša silne korelované. To možno zúžitkovať technikou menom **združené kódovanie kanálov** (angl. *coupled channel coding*) a zefektívniť kódovanie (napr. koeficienty prvého kanála sa prenesú normálne, koeficienty druhého budú rozdielom hodnôt oproti prvému kanálu – **dekorelácia vo frekvenčnej doméne**). Výhodou je okamžité vyhodnotenie efektívnosti tohto postupu (v prípade že by sa zvýšil výstupný dátový tok, nevykoná sa žiadne odpočítavanie koeficientov). Problémom je však to, že už mierna rozdielnosť v PCM vyjadrení jednotlivých kanálov môže spôsobiť silné zmeny v koeficientoch, teda frekvenčnej charakteristike kanálov, čo znamená, že korelácia sa nie vždy dá nájsť. Mierne zlepšenie môže priniesť rozdelenie frekvenčného pásma na časti a osobitné rozhodovanie o efektívnosti dekorelácie každej z nich.
- Ďalšou alternatívou je využitie vzájomnej korelovanosti jednotlivých kanálov zvuku ešte pred aplikovaním ortogonálnej transformácie – **spojené kódovanie kanálov** (angl. *joint stereo coding*). Vtedy sa zvyčajne osobitne transformuje priemer (súčet) kanálov (resp. jeden z kanálov) a ich rozdiel (resp. rozdiel od priemeru; akási analógia so systémom nahrávania MS³⁹, občas sa preto označuje *M/S stereo coding*). Či takáto **dekorelácia v časovej doméne** priniesla ušetrenie dátového toku, možno zistiť len porovnaním s výsledkami nezávislého kódovania kanálov. Navyše, kanál vyjadrujúci rozdiel pôvodných kanálov často nemá štandardné charakteristiky zvukového záznamu (a žiadnu podobnosť so súčtovým kanálom), preto nemusí byť vôbec vhodné použiť naň rovnaký algoritmus stratovej kompresie ako na súčtový kanál. Veľkou výhodou tohto postupu je však to, že na rozdiel od prvej metódy nie sme nútení voliť rovnako veľké okná pre oba kanály.
- Pri starších stratových formátoch (napr. MP1, MP2, MP3, DTS) sa možno stretnúť aj so **stereofonickým kódovaním intenzity** (angl. *intensity stereo coding*). Využíva sa len na vysokofrekvenčné subpásma. Kóduje sa priemer oboch kanálov a jediná zaznamenaná informácia je pomer hlasitosti medzi kanálmi v danom frekvenčnom subpásme. Výsledkom je zníženie dátového toku i výslednej kvality stereofonického dojmu.

39 P. odsek 2.5.2 *Stereofonický záznam zvuku*, str. 31

Konkrétna realizácia sa líši od formátu k formátu. Často je výber medzi týmito technikami závislý od želaného dátového toku (napr. pri nižšom sa volí spojené alebo združené kódovanie kanálov, pri vyššom nezávislé kódovanie kanálov za účelom dosiahnutia vyššej vernosti záznamu). Jednotlivé kanály zväčša prenášajú tie isté tóny, len s odlišnou amplitúdou, fázovým posunom a časovým oneskorením. Tieto faktory, ktoré silno vplývajú na lokalizáciu zvuku a priestorový dojem z hudby, treba zväžiť už pri návrhu kompresného algoritmu. Vo väčšine prípadov sa využíva amplitúdová korelácia najmä vo vyšších frekvenciách (kde človek nie je taký citlivý na fázový posun medzi kanálmi), pre stredné a nižšie frekvencie sa používajú techniky na zaznamenanie fázového rozdielu medzi združenými kanálmi.

3.4 Princípy bezstratovej kompresie zvuku

Opodstatnenie kompresie zvuku je jasné: Zvukové dáta v PCM formáte zaberajú značné miesto, je preto namieste využiť ich vnútornú štruktúru, korelovanosť a pokúsiť sa zapísať ich efektívnejšie. (Rátame totiž s tým, že nejaká štruktúra vo väčšine zvukov existuje – ak nejde o šum, p. časť 2.3 *Charakteristiky zvuku*, str. 21.) V časti 3.3 (str. 50) sme si predstavili postupy pri stratovej kompresii zvuku. Tie dosahujú značný kompresný pomer – typicky 1:10 a lepší. Teraz je namieste povedať, že výsledný kompresný pomer bezstratových kompresných formátov sa typicky pohybuje nad 1:2 (tzn. skomprimované dáta zaberajú viac ako polovicu objemu pôvodných dát). Navyše sa bezstratové algoritmy vyznačujú vysokou výpočtovou náročnosťou. Predsa však uchovávajú pôvodné zvukové dáta bez zmeny, čo je pri stratových formátoch nedosiahnuteľné.

Táto časť práce pozostáva zo štyroch odsekov. Najprv si predstavíme dva základné princípy, s ktorými sa pri návrhu algoritmov na bezstratovú kompresiu zvuku v praxi stretávame – 3.4.1 *Bezstratové transformačné kódovanie* (str. 59) a 3.4.2 *Kódovanie lineárnou predikciou* (str. 60). V odseku 3.4.3 *Hybridné a škálovateľné kódovanie* (str. 62) si spomenieme ich modifikácie. Záver časti 3.4 tvorí odsek 3.4.4 *Kompresia stereofonického záznamu* (str. 63).

3.4.1 Bezstratové transformačné kódovanie

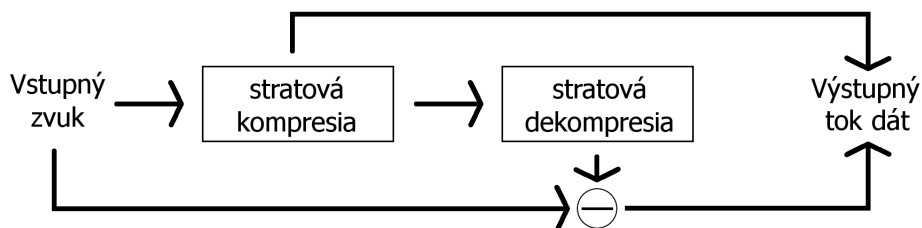
V časti 3.3 (*Princípy stratovej kompresie zvuku*, str. 50) sme si predstavili metódy stratovej kompresie zvuku. Uviedli sme si, že môžu často dosiahnuť dobrý kompresný pomer pri zachovaní vysokej vernosti, inými slovami, snažia sa skomprimovať vstup tak, aby bol veľmi podobný⁴⁰ pôvodným dátam. Ak teda existujú takéto úspešné algoritmy, zdá sa logické využiť ich ako základ bezstratovej kompresie zvuku. To, čo zostane, bude rozdiel medzi originálom a zvukom získaným po dekompresii dát zo stratovej kompresnej schémy – akási chybová zložka, zvyšok transmisie. Tento prístup rozvíja napr. Liebchen⁴¹ a nazýva ho **bezstratové transformačné kódovanie** (nem. „verlustlose Transformationscodierung“, angl. *lossless transformation coding*). Faktom je, že v praxi sa používa menej ako lineárna predikcia, ktorej sa venujeme v nasledujúcom odseku (str. 60).

Pri bezstratovom transformačnom kódovaní sú zvukové dáta najprv stratovo skomprimované, tak spätne dekomprimované, aby sa vyjadril rozdiel, skreslenie oproti originálu. Tento zvyšok sa nazýva **chybová zložka** (kóduje sa presne, bez straty) a spolu s koeficientmi zo stratovej kompresie (ktoré sa kvantujú len s istou

40 Samozrejme, ide o psychoakustickú vernosť (zhodnosť), ktorá nemusí byť dátovou zhodnosťou.

41 LIEBCHEN, Tilman. Realisierung einer verlustlosen Transformationscodierung zur Datenkompression von Mono- und Stereo-Audiosignalen. 1998.

presnosťou) tvorí výstupný tok dát (samozrejme, po použití vhodných entropických kódov). Schéma vyzerá nasledovne (obr. 3.2):



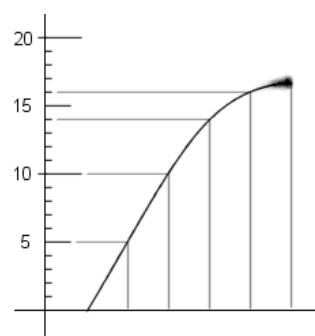
Obr. 3.2 – Schéma bezstratového transformačného kódovania

Pozornosť treba venovať vlastnostiam stratovej kompresie. V prípade nízkeho zvoleného dátového toku jej výstupu vzniká signál veľmi odlišný od pôvodného signálu, teda chybová zložka dosahuje vysoké hodnoty (pričom tieto sú silne korelované). Výstupný tok dát tak potom tvorí najmä chybová zložka. Pri zvolení vysokej kvality stratovej kompresie sa výrazne potláča veľkosť (i korelovanosť) chýb, výstup však zaberá zvýšený výstupný dátový tok z algoritmu stratovej kompresie. Vhodným kvantovacím faktorom α (číslo, ktoré vyjadruje najmenšiu jednotku škály pri zaokrúhľovaní koeficientov, teda stratovej zložky tohto kódovania) možno docíliť optimum. Liebchen a kol.⁴² ukazujú, že najlepšie výsledky možno docíliť, ak $0 < \alpha \ll 1$. Liebchen vo svojom algoritme LTAC⁴³ používa $\alpha = 0,25$.

Bezstratové transformačné kódovanie našlo uplatnenie nie ani tak v samotnej bezstratovej kompresii zvuku, ale najmä v *škálovateľných kompresných schémach* – p. odsek nižšie (3.4.3, str. 62).

3.4.2 Kódovanie lineárnou predikciou

Vieme, že digitálny zápis zvuku vznikol pôvodne zo spojitého materiálu – analógového záznamu zvuku. Tento záznam je výsledkom skladania rôznych jednoduchých funkcií, zväčša tvaru sinusoidy (inými slovami, je súzvukom rôznych tónov, vlnení). Diskrétné ortogonálne transformácie sa snažia blok dát rozložiť na frekvenčné zložky, ale ponúka sa aj iná možnosť: na základe aktuálneho priebehu zvukovej „krivky“ (samozrejme, toto spojité chápanie zvuku sa kvantovaním zmenilo na diskkrétne vyjadrenie, ktoré ho len aproximuje) predpovedať hodnotu ďalšej vzorky. Ilustrovať to možno na príklade (obr. 3.3). Ak majú namerané vzorky zvuku hodnoty 0, 5, 10,



Obr. 3.3 – Demonštrácia spojitého charakteru zvuku

42 LIEBCHEN, T., PURAT, M., NOLL, P. Lossless Transform Coding of Audio Signals. 1997

43 LTAC (Lossless Transform Audio Compression) je kodek vyvinutý Tilmanom Liebchenom z Technickej univerzity v Berlíne ako súčasť jeho diplomovej práce (LIEBCHEN, 1998). V praxi sa nedočkal veľkého úspechu, preto sa mu nebudeme ďalej venovať. Čitateľa odporúčame na stránky <http://www.nue.tu-berlin.de/wer/liebchen/ltac.html>.

14, 16, zo spojitosti pôvodného zvuku, z ktorého digitálny zápis vznikol, možno predpokladať, že hodnota ďalšej vzorky bude kdesi blízko čísla 17. Rozdiel oproti predpovedanej hodnote by potom bola jediná informácia, ktorú bude nutné preniesť. (Pravdepodobnosť prvej odchýlky od predpovedanej hodnoty je nízka – znamenalo by to, že v nahrávke sa náhle objavil zvuk s vysokou amplitúdou a širokým frekvenčným spektrom.)

V tomto odseku popisujeme metódy, ktoré sa snažia predpovedať hodnotu nasledujúcej vzorky na základe predošlých vzoriek, pričom pre zachovanie bezstratovosti kompresie prenášajú zvyšok, odchýlku od predpokladanej hodnoty, tzv. **reziduál**. Počet vzoriek použitých na predikciu budeme nazývať **rád predikcie**. V praxi sa na túto úlohu ukázali najvhodnejšie lineárne funkcie, preto spomenuté metódy nazývame **kódovanie lineárnou predikciou** (*linear prediction coding*, **LPC**). Lineárne funkcie síce presne nevyjadrujú skutočnú charakteristiku zvuku (ktoré je vlnením, sinusoidou), môžu ho však dostatočne aproximovať (podobne ako je sínus dobre aproximovaný Taylorovým polynómom zvoleného stupňa), navyše sú nenáročné na výpočet. Zavedenie silnejšieho aparátu by výrazne zvýšilo výpočtovú náročnosť za cenu len malej úspory v dátovom toku.

V každom kroku lineárnej predikcie teda vyberáme vhodnú funkciu (**prediktor**, napr. P) so želaným rádom predikcie (p). Prediktorom odhadneme hodnotu ďalšej vzorky ($P_i := P^p(x_{i-1}, x_{i-2}, \dots, x_{i-p})$) a odpočítame ju od skutočnej hodnoty, aby sme získali reziduál ($r_i := x_i - P_i$). Na spomenutom príklade získame napr. pre prediktor druhého rádu $P^2(x_{i-1}, x_{i-2}) = 2x_{i-1} - x_{i-2}$ ako výsledok hodnotu 18, čo sa zrejme nebude veľmi líšiť od skutočnej hodnoty. Štatisticky sú reziduály (pri voľbe vhodných prediktorov) nízke čísla, s hustým rozdelením okolo nuly – sú teda vhodnými kandidátmi pre entropické kódy.

Vo všeobecnosti možno prediktor P rádu p vyjadriť ako:

$$P^p(x_{i-1}, x_{i-2}, \dots, x_{i-p}) = \sum_{j=1}^p a_j x_{i-j},$$

kde a_1, a_2, \dots, a_p sú koeficienty tohto prediktora. Zvolením vhodného rádu prediktora a jeho koeficientov možno zminimalizovať štandardnú odchýlku reziduálneho signálu (postupnosti reziduálov). Existujú rôzne algebraické optimalizačné metódy riešenia tohto problému⁴⁴. (Na začiatku bloku je, samozrejme, nemožné hneď použiť prediktor rádu p . Prvých p vzoriek preto treba preniesť priamo, resp. možno aplikovať **progressívnu predikciu**⁴⁵, pri ktorej rád predikcie postupne rastie až k číslu p .) Tento postup – keď sa pre spracúvaný blok na základe jeho obsahu

44 V prípade lineárnej predikcie sa používa najčastejšie Levinsonov-Durbinov algoritmus, ktorého časová zložitosť je $O(p^2)$. Nájde optimálne koeficienty prediktora, navyše poskytuje možnosť iteratívneho zvyšovania rádu. Presný postup možno nájsť napr. na <http://www.data-compression.com/speech.html> alebo http://en.wikipedia.org/wiki/Levinson-Durbin_algorithm.

45 Ako ju pre MPEG-4 ALS (p. časť 5.8, str. 128) navrhli MORIYA et al. *Extended Linear Prediction Tools for Lossless Audio Coding*. 2004.

dopredu určia rád a koeficienty prediktora – sa nazýva **dopredne adaptívna predikcia**. Rozhodovanie je na strane kodéra, dekodér tak môže byť konštrukciou a výpočtovou zložitou omnoho jednoduchší.

(Poznámka: V prípade, že sa koeficienty určujú len na základe analýzy už spracovaného signálu – vzoriek, hovoríme o **spätne adaptívnej predikcii**. V tom prípade nie je potrebné prenášať informácie o koeficientoch, keďže sa tieto vypočítajú zhodne v algoritme kodéra i dekodéra. Navyše netreba dopredu poznať hodnoty vzoriek v celom spracúvanom bloku, ale signál možno spracúvať v reálnom čase, čo je mimoriadne výhodné napríklad v prípade kompresie hovoreného slova pri mobilných alebo internetových protokoloch na prenos hlasu. Nevýhodou je rovnako vysoká komplexnosť kodéra i dekodéra.)

Zvolením vhodnej veľkosti bloku a výberom vhodného prediktora (používajú sa aj tabuľky fixne daných prediktorov, aby sa nemuseli prenášať koeficienty, alebo VQ, alebo sa využije istá symetrickosť koeficientov a ich združovanie metódou **LSP** alebo **parcor**⁴⁶) teda možno kompresiu adaptívne prispôbovať potrebám aktuálneho vstupu, za účelom dosiahnutia najvyššej kompresie. Súbežne možno použiť metódu adaptívnosti predikovateľnosti: určí sa **faktor predikovateľnosti** (pomer „spoliehania“ sa na predikciu) m (na začiatku napr. $m=0,75$), pričom nová predikovaná hodnota vzorky bude škálovaná medzi predošlou hodnotou vzorky a pôvodne predikovanou hodnotou vzorky (teda nová predikovaná hodnota vzorky sa vypočíta ako $P_i := m P^p(x_{i-1}, x_{i-2}, \dots, x_{i-p}) + (1-m)x_{i-1}$). V prípade, že by menší faktor predikovateľnosti prispel k menšej chybe, v ďalšom kroku sa m môže zmenšiť (postup pri jeho úprave závisí od konkrétneho kompresného algoritmu).

V praxi sú reálne koeficienty prediktora pre svoj rozsah nevhodné na prenos, preto sa zaokrúhľujú a reziduálny signál sa počíta z takto kvantovaného prediktora.

3.4.3 Hybridné a škálovateľné kódovanie

Na začiatku tejto časti práce (3.4, str. 59) sme si spomenuli nízku účinnosť bezstratových kompresných schém. Pre mnohých používateľov predstavuje veľkosť výsledných dát problém – chcú na svoje prenosné dátové médiá alebo prehrávače uložiť čím viac hudobných záznamov. Zároveň nechcú stratiť výhody bezstratovej kompresie zvuku – možnosť dokonalo rekonštruovať pôvodný záznam.

Prvým, triviálnym riešením je použiť dva rôzne formáty na dosiahnutie oboch cieľov. Tento prístup však v mnohých prípadoch nie je dostatočne pohodlný (dva súbory), alebo nie je vhodný vzhľadom na použitý dátový priestor (z dôvodu duplicity).

46 Bližší popis LSP možno nájsť napr. na <http://www.data-compression.com/speech.html>. Parcor je metóda na zápis tzv. parcorových (z angl. *partial correlation*) alebo reflexných koeficientov – viac možno nájsť napr. na http://en.wikipedia.org/wiki/Levinson-Durbin_algorithm. LPC koeficienty, ich LSP zápis a reflexné koeficienty sú navzájom zameniteľné, každé majú svoje výhody a nevýhody.

Druhou možnosťou je využiť tzv. **hybridné kompresné schémy**. Ich výstupom sú dva súbory. Prvý (menší) súbor poskytuje stratovú, no psychoakusticky dostačujúcu reprodukciu. Zväčša sa používa metóda LPC, teda súbor zahŕňa popis prediktora a kvantovaný (teda približný) reziduálny signál. Druhý (väčší) súbor neslúži na reprodukciu, ale na prípadnú bezstratovú rekonštrukciu pôvodného záznamu. Obsahuje chybovú zložku z kvantovania reziduálneho signálu pri tvorbe prvého súboru, preto z neho možno, v kombinácii s prvým súborom, pôvodný záznam dokonalo zrekonštruovať. Hybridné kódovanie je občas voliteľnou vlastnosťou niektorých bezstratových schém, napr. WavPack (str. 119) a OptimFROG (str. 127).

Tretím riešením je tzv. **škálovateľné kódovanie**, pripomínajúce bezstratové transformačné kódovanie z odseku 3.4.1 (str. 59). Výsledok je bezstratovo uložený v jedinom súbore, no základ netvorí LPC, ale transformačná kódová schéma – tá je svojou konštrukciou vhodnejšia pre úspornú stratovú kompresiu. Zvyšok toku dát tvorí chybová zložka. Jej kódovanie je však akoby viacvrstvé – v toku dát je prítomných viacero hierarchicky usporiadaných tokov. Každý ďalší tok zvyšuje presnosť zápisu chybovej zložky. Použitie je jednoduché – stačí jeden súbor, z ktorého v prípade potreby možno (bez nutnosti dekompresie a opätovnej kompresie) vytvoriť súbor so zvoleným dátovým tokom. Táto možnosť je vhodná najmä pri prenášaní mediálneho obsahu internetom – dátový tok možno prispôbovať vlastnostiam spojenia. Príkladom technológie, ktorá by škálovateľné kódovanie mala ponúknuť, je MPEG-4 SLS (str. 130).

3.4.4 Kompresia stereofonického záznamu

Ako v prípade stratovej kompresie zvuku (rovnomený odsek 3.3.5, str. 57), i pri bezstratových formátoch nesmieme zabudnúť na stereofonický záznam. Medzikanálovú koreláciu možno využiť viacerými spôsobmi, ktoré sa v prípade bezstratovej kompresie zvuku navzájom nevyklučujú, ale dopĺňajú:

- Prvým spôsobom je miesto **osobitného kódovania kanálov** snažiť sa spracúvať ich súčet (združený kanál) a jeden z pôvodných kanálov (doplnkový kanál). Tento postup je zhodný so **spojeným kódovaním kanálov** (ktoré sme bližšie vysvetlili v odseku 3.3.5).
- Podobným spôsobom môžeme využiť podobnosť hodnôt koeficientov prediktorov, ktoré v prípade silnej medzikanálovej korelácie odzrkadľujú rovnaké alebo podobné zvuky, vlnenia v zdroji. Badať podobnosť so **združeným kódovaním kanálov** (tamže).
- Nakoniec možno tento postup využiť aj na reziduály a vytvoriť tak združený a doplnkový reziduálny kanál.

3.5 Zvuk v súčasnom svete elektroniky

Čitateľa sme v predošlých častiach práce (3.3 a 3.4) oboznámili s rozličnými spôsobmi kompresie zvuku. Chýba však celkový pohľad na spracovanie zvuku, na súčasný stav a trendy v práci s ním. V tejto časti práce, pozostávajúcej až z jedenástich odsekov, preberieme technologické aspekty, štandardy a normy v spracovaní zvuku do tej miery, aby sme tento základný prehľad poskytli.

Začneme hardvérom v odseku 3.5.1 *Štruktúra zvukového systému PC* (str. 64). V odsekoch 3.5.2 *Priestorový zvuk a zvukové kanály* (str. 66) a 3.5.3 *Priestorový zvuk v softvéri* (str. 69) sa oboznámime so stereofonickými technikami v praxi. Odseky 3.5.4 *Zvuková syntéza a MIDI* (str. 70) a 3.5.5 *Hudobné sekvencery, moduly* (str. 73) nám poskytnú alternatívne pohľady na prácu s hudbou a jej komponovanie. Odsek 3.5.6 *Ochrana duševného vlastníctva* (str. 74) je pojednaním o autorských právach a ich dôsledkoch a prínosoch do sveta multimédií. V odsekoch 3.5.7 *Úložné médiá* (str. 77) a 3.5.8 *Zvuk a video* (str. 80) sa čitateľ dozvie čosi o minulých, aktuálnych i budúcich trendoch v ukladaní zvukových dát a ich komerčnej využiteľnosti. Krátky odsek 3.5.9 *Rozhrania na prenos zvuku* (str. 82) popíše rôzne spôsoby na prenos digitálneho zvuku. Odsek 3.5.10 *MPEG* (str. 83) prináša prehľad o funkcii tejto organizácie a jej vplyve na svet multimédií. Tematiku uzatvára odsek 3.5.11 *Predpokladaný vývoj* (str. 85).

3.5.1 Štruktúra zvukového systému PC

Napriek názvu tohto odseku sa tu uvedený opis vzťahuje do značnej miery na všetky osobné počítače, nielen na PC. Popíšeme si základné funkčné jednotky zvukového systému a zhrnieme jeho vývoj v priebehu uplynulých rokov.

Jadrom spracovania zvuku na PC je **zvuková karta**. V podstate ide o prídavné zariadenie (nemusí byť v PC prítomné), aj keď je v súčasnosti zvuková karta súčasťou asi každej matičnej dosky na trhu. Na matičnú dosku sa zvyčajne pripája rozhraním PCI, výnimočne cez USB či Firewire, v budúcnosti sa zrejme presadí nové sériové rozhranie PCI-Express. Zvukové karty majú niekoľko analógových vstupov (linkový vstup, mikrofónny vstup) a výstupov (dvojice výstupných reproduktorov), občas je prítomný aj digitálny výstup S/PDIF⁴⁷. Koncovky sú realizované ako samičie **jack** konektory (3,5 mm), okrem mikrofónneho vstupu zvyknú byť typu stereo. Počet výstupných koncoviek je závislý od počtu podporovaných kanálov. V prípade dvojkanálovej zvukovej karty je to jedna (stereo) koncovka, štvorkanálová používa dve, šesťkanálová tri, osemkanálová štyri. (S väčším množstvom kanálov sa v praxi nestretávame.) Zvukové karty rozšírené na trhu zväčša používajú štandardné farebné odlišenie jednotlivých konektorov – mikrofónnemu vstupu je priradená ružová farba,

47 P. odsek 3.5.9 Rozhrania na prenos zvuku, str. 82.

linkovému vstupu modrá, hlavnému výstupu (predné reproduktory) zelená, druhému čierna; digitálny výstup používa oranžovú farbu. S príchodom viackanálového zvuku a potrebou vtesnať na malú plochu karty dostatok vstupných a výstupných koncoviek prišla aj potreba multifunkčných koncoviek – tak sa stáva, že oranžová koncovka je popri digitálnom výstupe použitá i na stredný a basový reproduktor (piaty a šiesty kanál; samozrejme, cez digitálny výstup sa môžu prenášať všetky kanály naraz, nemusí teda vzniknúť potreba používať digitálny i analógový výstup súčasne). Zadným priestorovým reproduktorom (siedmy a ôsmy kanál) sa priradzuje koncovka rozličných farieb (zväčša hnedá). V najnovších zvukových kartách sa dokonca objavuje technológia „jack sensing“⁴⁸, ktorá umožňuje karte zistiť (resp. softvérovo nastaviť) význam jednotlivých koncoviek (vrátane toho, či ide o vstupný alebo výstupný signál).

Na zvukovej karte sa nachádza niekoľko obvodov. V prvom rade je to **kodek**, teda kombinácia A/D a D/A prevodníka⁴⁹. Ďalším (s kodekom historicky najstarším) základným obvodom je operátor frekvenčnej modulácie, generujúci zvukové charakteristiky hudobných nástrojov⁵⁰. Najväčšiu popularitu získali čipy od japonskej firmy Yamaha a OPL⁵¹ sa stalo de-facto štandardom. Až omnoho neskôr, s rastom výpočtovej kapacity a pamäte počítačov, sa na zvukové karty určené pre domácich používateľov dostáva digitálny signálový procesor (**DSP**), ktorý jej umožňuje v reálnom čase spracúvať a upravovať digitálne zvukové dáta (kompresia akustickej hladiny, frekvenčné filtre, umelý dozvuk, efekty atď.). Prvé zvukové karty zvykli mať jeden alebo dva kanály (tvoriace stereofonickú bázu) a neboli plne duplexné (nemohli súčasne zaznamenávať a reprodukovat' zvuk). Neskôr pribudla nielen plná duplexnosť, ale najmä počet nezávislých kanálov, s ktorými mohla zvuková karta pracovať⁵². V súčasnosti je tento počet aspoň 16. Na plošnom spoji zvukovej karty sa nachádzajú ešte vstupné a výstupné zosilňovače a rôzne podporné obvody.

Komunikácia so zvukovou kartou prebieha zo strany softvéru použitím ovládačov zvukovej karty, nainštalovaných v použítom operačnom systéme (komunikujú cez zbernicu s DSP). Možno zachytávať priamy zvuk zo vstupov

48 Akoby „citlivosť na koncovky typu jack“. Vlastnosť väčšiny súčasných kariet, ale i „kodekov“ (p. nasledujúcu poznámku) – napr. Intel HD Audio (kde sa táto črta nazýva „jack retasking“).

49 Na základe kodeku možno vybudovať zvukovú kartu – prevodníky a digitálny mikrokontrolér toku dát sú riešené hardvérovo na matičnej doske (angl. *on-board audio* ako náprotivok prídavných kariet – *add-on card*, tiež nazývaných samostatné - *standalone*), pokročilejšia mixáž zvuku a syntéza nástrojov potom prebieha len softvérovo, na úrovni ovládačov zvukovej karty, zaťažujúc procesor počítača. Takéto riešenia, určené do lacných domácich a kancelárskych počítačov, si vďaka tejto jedinej hardvérovej súčasti vyslúžili tiež prezývku „kodek“. Najrozšírenejším štandardom v tomto segmente je bezpochyby AC'97 od Intelu (najviac 20-bitová hĺbka a 48kHz pre najviac 6 kanálov cez S/PDIF, resp. 96 kHz a 24 bitov pre stereo), v súčasnosti ho začína nahrádzať novší štandard Intel High Definition Audio (32 bitov, 192 kHz, 8 kanálov, podporuje „jack retasking“, čo je označenie pre jack sensing). Pre tento pokrok a postačujúcu kvalitu už nemožno hovoriť o „kodekoch“ ako o nekvalitnom riešení.

50 Viac v odseku 3.5.4 Zvuková syntéza a MIDI, str. 70.

51 Slangová skratka z „FM Operator type L“ – samozrejme, skutočné názvy čipov boli odlišné.

52 Netreba si mýliť s počtom výstupných kanálov, na ktoré bolo možné tieto „vnútorné kanály“ vo zvolenom pomere prenášať.

(linkový vstup, mikrofón), resp. posielat' zvuk priamo na výstup (resp. výstupy). Dáta sa prenášajú vo forme PCM, počet kanálov, vzorkovaciu frekvenciu a rozlíšenie vzorky možno vybrať z rozsahu podporovaného zvukovou kartou. V súčasnosti sú rozšírené najmä stereo (2.0) a 5.1 (ojedinele 7.1 alebo 4.0) karty, podporované vzorkovacie frekvencie sú 44.1 kHz, 48 kHz, začína sa objavovať 96 kHz (ojedinele 192 kHz; samozrejme, karty podporujú aj niektoré nižšie frekvencie, zväčša 32 kHz, 24 kHz, 22,05 kHz, 16 kHz), podporovaná bitová hĺbka vzorky je 8 bitov a 16 bitov, objavujú sa však aj 24-bitové a 32-bitové zvukové karty. Reprodukciu možno rôznymi spôsobmi dotvárať – využitím rozličných nastavení súvisiacich s reprodukciou⁵³. Osobitnou kapitolou riadenia výstupu zvukovej karty je protokol MIDI⁵⁴.

Kvalitnejšie či veľmi kvalitné (tzv. *high end*) rady zvukových kariet niektorých výrobcov sa vyrábajú aj v externom vyhotovení, obyčajne pripojiteľné k počítaču cez rozhranie FireWire alebo USB, nie ako interná PCI karta. Výhodou je menšia náchylnosť na elektromagnetické rušenie, prítomné v skrinke počítača, a tak vyššia vernosť (angl. *high fidelity*, **hi-fi**) reprodukcie zvuku, čo je hlavná požiadavka náročných poslucháčov (tzv. audiofilov).

3.5.2 Priestorový zvuk a zvukové kanály

Aj keď sa slovo „stereofonický“ v súčasnej slovenčine spája s „dvojkanálovým systémom záznamu a reprodukcie zvuku“⁵⁵, pôvodný význam tohto slova je „viackanálový“. Často používaný termín **stereo** pre historicky prvé rozšírené zariadenia pracujúce s dvoma kanálmi (narozdiel od tých predošlých, jednokanálových, **mono**) však tento pôvodný význam vytlačil, preto my tiež pod pojmom *stereo* budeme chápať dvojkanálový zvuk (v primárnom význame: zvuk určený pre ľavé a pravé ucho). Pre **viackanálový zvuk** (viac než dva kanály) sa v súčasnosti skôr ujal termín **priestorový zvuk** (angl. *surround sound*). Pri kompresii takéhoto záznamu sa využívajú metódy podobné tým, ktoré sme uviedli v súvislosti s kompresiou stereofonického záznamu (p. odseky 3.3.5, str. 57 a 3.4.4, str. 63).

Treba však odlišiť počet (diskrétnych) kanálov v zázname od počtu reprodukováných kanálov. Tento počet možno totiž nezávislo meniť – napríklad použitím **maticového dekódovania zvuku** (*matrix decoding*). Vstupný (dekódovaný) signál s istým počtom kanálov sa prevedie do každého výstupného kanálu s intenzitou určenou hodnotami matice. Najjednoduchším príkladom je zmena mono nahrávky na stereo reprodukciu – do každého výstupného kanálu smeruje pôvodný zvuk. Tento postup – zvýšenie počtu reprodukováných kanálov – sa v angličtine označuje **upmix**. Naopak, prevod väčšieho množstva kanálov na menší počet (napr. zmena stereo

53 P. napr. odsek 3.5.3 Priestorový zvuk v softvéri, str. 69.

54 P. odsek 3.5.4 Zvuková syntéza a MIDI, str. 70.

55 Krátky slovník slovenského jazyka. Bratislava: Veda. 2003. Možno overiť na internetovej stránke <http://kssj.juls.savba.sk> [cit. 2006-03-13].

záznamu na mono za účelom reprodukcie na jednom reproduktore) sa označuje ako **downmix**. Matice, ktoré by na spomínané prevody poslúžili, by vyzerali takto:

$$A_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, A_2 = (0,5 \quad 0,5).$$

Vstupné kanály vyjadrujú stĺpce, výstupné kanály riadky matice. Matica A_1 jediný vstupný kanál prevedie do oboch výstupných kanálov s pôvodnou intenzitou, matica A_2 zmieša do jedného výstupného kanála polovičnú úroveň oboch vstupných kanálov (ochrana pred orezávaním). V praxi sa, samozrejme, často používajú omnoho väčšie matice (p. ďalej).

Navyše, je potrebné odlišiť počet reprodukováných kanálov od skutočného počtu reproduktorov. Jeden kanál totiž možno reprodukovať viacerými blízko seba umiestnenými reproduktormi. Jeden z kanálov môže byť vyhradený na reprodukciu hlbokých tónov – veľké rozmery k nemu priradeného reproduktora ho môžu na to predurčovať. Nízkofrekvenčné zložky zo všetkých kanálov sú často presmerované práve do tohto kanála (či už pri zostavovaní záznamu, alebo až / aj tesne pred reprodukciou), využívajúc fakt, že zdroj hlbokých tónov je ťažšie lokalizovateľný (basy „naplnia miestnosť“ bez narušania zážitku z priestorového zvuku).

Za účelom odlišenia všetkých týchto aspektov záznamu zvuku alebo reprodukčnej sústavy sa zaužívalo nasledovné číselné označenie. Prvé číslo indikuje počet (frekvenčne) celorozsahových diskretných kanálov. Nasleduje ho bodka a počet frekvenčne obmedzených kanálov (zväčša jeden kanál **LFE** – *low frequency effects*⁵⁶, teda nízkofrekvenčné efekty). Prvé číslo môže byť rozdelené lomkou na počet kanálov reprodukováných pred poslucháčom a zvyšný počet kanálov. Pri použití maticového dekódovania zvuku sa navyše toto číslo môže upresniť počtom vypočítaných kanálov za chrbtom poslucháča, oddeleným dvojbodkou. (Např. zápis „3/2:3.1“ indikuje 5 celorozsahových kanálov, z toho tri sú pred poslucháčom, tri v jeho okolí – aj keď len dva z týchto troch sú diskretné prítomné v pôvodnom zázname – a v miestnosti sa nachádza aj jeden subwoofer.)

Čitateľ si isto všimol, že s reprodukoványmi kanálmi súvisí aj umiestnenie reproduktorov (reprodukčných sústav), ktoré ich „realizujú“ – koniec koncov, to je zmysel reprodukcie priestorového zvuku. Zaužívalo sa niekoľko štandardných⁵⁷ konfigurácií reproduktorov / kanálov:

- 2.0: Predný ľavý (L) a predný pravý (R) reproduktor – s osou (smerom pohľadu) poslucháča zvyčajne zvierajú uhol 22-30°.

56 Takýto názov používa (najmä) štandard Dolby Digital. Používajú sa aj názvy Low Frequency Enhancement, zriedkavo Low Frequency Extension. Úlohou LFE kanála je reprodukcia veľmi hlbokých tónov. Žiadané je aj dosiahnutie vyššieho akustického tlaku kvôli menšej citlivosti ľudského sluchu na nízke frekvencie (v prípade AC-3 je to +10 dB oproti ostatným kanálom – p. ATSC. Digital Audio Compression Standard (AC-3, E-AC-3). Rev. B. 2005. str. 24). LFE sa realizuje reproduktorom typu subwoofer (p. odsek 2.4.3 *Reprodukcia zvuku*, str. 27).

57 „Štandardy“ v tomto zozname sú zapísané skôr marketingovým označením, ktoré nie je celkom zhodné so značením uvedeným vyššie.

- 4.0: Zväčša k 2.0 pridáva dva zadné (pravý a ľavý) reproduktory – nejednotný štandard. (Známe sú aj 4.1 zostavy, pridávajúce k 4.0 jeden subwoofer.)
- 5.1: K systému 2.0 pridáva ľavý a pravý priestorový kanál (značené ako Sl, Sr, občas L_s, R_s), ktoré zvyčajne zvierajú so smerom pohľadu poslucháča uhol 90-110°, a dvojicu kanálov: predný stredový kanál (C) a basový kanál (LFE). Tento systém je vo sfére domáceho kina de facto štandard.
- 6.1: Oproti 5.1 pridáva do stredu za poslucháčom ďalší kanál (Sb). Využíva ho napr. aj štandard Dolby-Digital EX (p. nižšie), ktorý tento prídavný kanál vypočíta na základe maticového dekódovania.
- 7.1: Tieto systémy zatiaľ existujú len vo svete počítačov. (Podporujú ho v podstate len niektoré zvukové karty, nie spotrebiteľská elektronika; predsa však môžu byť vhodné i na domáce kino – používa sa zdvojenie kanála Sb systému 6.1 do oboch zadných reproduktorov.) Dodáva dva reproduktory za poslucháča, ich os zvierá s pohľadom poslucháča uhol 135-150°.

Je namieste popísať jednotlivé rozšírené formáty priestorového zvuku; také, čo síce boli vyvinuté pre film, no našli uplatnenie aj v domácnostiach. Väčšinu štandardov v tomto odvetví určite vlastní spoločnosť Dolby Laboratories. Medzi prvé formáty priestorového zvuku patrili **Dolby Stereo**, **Dolby Surround** a **Dolby Analog SR** zo 70. a 80. rokov minulého storočia. (Posledný menovaný sa doteraz používa na filmových kotúčoch do kín, zväčša už popri digitálnom zázname zvuku). Používajú maticové dekódovanie na obohatenie stereofonického záznamu o stredový a priestorový kanál (teda spolu štyri reprodukovateľné kanály v dvoch diskretných kanáloch; priestorový kanál má zväčša obmedzenú frekvenčnú charakteristiku). Digitálnej verzii sa dekódovanie týchto štandardov dočkalo v podobe **Dolby Pro Logic** a jeho nasledovníkov, ktoré poskytujú 5.1 (Dolby Pro Logic II), ba až 7.1 (Dolby Pro Logic IIx) zvuk v plnom frekvenčnom rozsahu, maticovo zakódovaný v dvoch diskretných kanáloch.

Dolby Digital (DD), často nazývaný aj **Adaptive Transform Coder 3 (AC-3)** je ďalším dielom tejto spoločnosti. Uviedla ho v r. 1992 a zahrnutím do špecifikácie DVD-Video si získal všeobecnú popularitu. Stratová kompresia s ním spojená je bližšie predstavená v časti 4.7 (str. 104). Používa šesť diskretných kanálov v rozložení 5.1, vzorkovaciu frekvenciu 48 kHz a rozlíšenie vzorky 24 bitov. Dolby Digital EX zvyšuje počet kanálov na 6.1 použitím maticového dekódovania. Novšia verzia **Dolby Digital Plus** (DD+, AC3plus) zvyšuje tento počet na 13.1 a je povinná pre HD DVD prehrávače (nepovinná pre BD), rovnako ako bezstratová kompresia Dolby TrueHD⁵⁸.

Konkurentom je štandard **DTS** firmy Digital Theater Systems z r. 1993, určený predovšetkým do filmov premietaných v kinách. Možno sa s ním stretnúť aj na

⁵⁸ Vychádza z MLP (p. časť 5.10, str. 131), ponúka navyše mnoho metadát, ako napr. vlastná konfigurácia (umiestnenie) reprodukovateľných kanálov, kontrola dynamického rozsahu a normalizácie a podobne.

diskoch DVD-Video, jeho podpora na strane prehrávačov je však voliteľná. Ide o stratovú kompresiu používajúcu kodek **Coherent Acoustics (CAC)**⁵⁹, má vyšší štandardne používaný dátový tok oproti DD (typicky 768 alebo 1536 kbps namiesto 384 alebo 448 kbps), základná verzia ponúka rozlíšenie 24 bit⁶⁰ a 48 kHz pre 5.1 kanálov (LFE kanál do 80, resp. 150 Hz – v prípade DD je to 120 Hz). Ďalším štandardom je formát DTS-ES, ktorý pripomína Dolby Digital EX, no vie pracovať i so skutočným diskretným kanálom (vtedy sa označuje DTS-ES 6.1). DTS NEO:6 je obdobou technológie Dolby Pro Logic, DTS 96/24 rozširuje podporu zvuku s vysokým rozlíšením na všetky kanály. DTS-HD Master Audio ponúka teoreticky neobmedzený počet kanálov a škálovateľnú kompresiu siahajúcu od klasického DTS až po bezstratovú kompresiu. Technológia CAC je povinnou súčasťou štandardov HD DVD a BD, výpočtovo (napriek vyšším typickým dátovým tokom) je menej náročná ako DD, navyše má aj nižšiu latenciu (časové oneskorenie pri kódovaní v reálnom čase).

3.5.3 Priestorový zvuk v softvéri

Doteraz sme spomínali hardvérovú realizáciu na uľahčenie lokalizácie zdroja – viackanálový zvuk⁶¹. Na priestorový dojem zo zvuku však vplýva viac faktorov, ktoré možno simulovať výpočtovou kapacitou (zväčša) DSP čipov na zvukových kartách – v prvom rade môže ísť o akustickú lokalizáciu zdroja ponad skutočné možnosti reproduktorov⁶². Ak uvažujeme o dvoch reproduktoroch, 3D efekty možno rozdeliť do troch skupín:

- Rozšírenie stereofonického dojmu. Možno ho dosiahnuť extrakciou stredového (súčet) a postranného (rozdiel) kanála z pôvodného zdroja a následným posilnením postranného kanála. Jednoduchší spôsob je pripočítať stíšený a fázovo invertovaný signál z každého kanála k druhému kanálu. Ak sa pripočíta signál s oneskorením, vzniká dojem prirodzeného dozvuku.
- Umiestnenie zvuku mimo stereofonickej bázy. Dosiahne sa sofistikovanejšou manipuláciou s fázou a hlasitosťou zvuku na základe jednoduchých psychoakustických poznatkov.
- Umiestnenie zvuku v priestore. Naplno sa využívajú poznatky zo psychoakustiky (najmä HRTF), zahrnujúc fázové posuny a simuláciu dozvuku miestnosti.

59 Tu hovoríme o formáte určenom pre domácich používateľov (p. časť 4.8, str. 106). Verzia určená do kín (apt-X100) má rozlíšenie vzorky 16 bitov, používa delenie na štyri subpásma a upravené ADPCM, podobne ako CAC (stručný popis možno nájsť na <http://www.aptx.com/sitefiles/resources/aptxoverview.pdf>).

60 Avšak rovnako ako DD, i DTS používa typicky len 18 bitov, čo poskytuje dostatočný dynamický rozsah a umožňuje úsporu dátového toku.

61 Programátor si nemusí robiť starosti s obsluhou všetkých výstupných kanálov zvukovej karty používateľa – v prostredí Windows sa napríklad môže spoľahnúť na bezplatnú knižnicu DirectSound, ktorá mu umožňuje zvuk lokalizovať bez ohľadu na druh reprodukcie.

62 P. časť 2.5 *Priestor vo zvuku* (str. 30) a jej odseky, najmä 2.5.3 *Súčasný trendy* (str. 33).

Druhou kategóriou sú jednoduchšie, no o to dôležitejšie faktory vplyvajúce na lokalizáciu zvuku (predovšetkým v interaktívnom prostredí hier), najmä prirodzený dozvuk miestnosti, neskoršie odrazy (echo), útlm (medzi zdrojom zvuku a poslucháčom je prekážka) atď. Najrozšírenejším štandardom v tomto smere je EAX (*Environmental Audio Extensions*), ktorý vyvíja líder vo zvukových kartách pre hráčsku komunitu – Creative Labs. Na trhu je od r. 1998 (s príchodom kariet *Sound Blaster Live!*, tie umožňovali len jednoduchý dozvuk). V súčasnosti (s príchodom kariet *Sound Blaster X-Fi* a piatou generáciou EAX, konkrétne *EAX Advanced HD 5.0*) možno simulovať množstvo detailne konfigurovateľných akustických prostredí s rôznymi spôsobmi šírenia zvuku a celkovým dojmom z neho.

Akoby medzistupňami medzi viacerými diskretnými kanálmi a softvérovým prístupom ku konštrukcii priestorového dojmu zo zvuku sú technológie ako MP3 Surround (p. odsek 4.2.4, str. 91) alebo jej zovšeobecnenie - MPEG Surround⁶³, stále vo vývoji.

3.5.4 Zvuková syntéza a MIDI

V tomto odseku si v stručnosti popíšeme črty protokolu MIDI, najprv si však predstavíme zvukovú syntézu ako takú.

Skôr než sa zvuk úspešne zaznamenával a následne reprodukoval, existovali umelé spôsoby tvorby (syntézy) zvuku (angl. *audio synthesis*) – teda zväčša tónov rôznych nástrojov. Spočiatku na elektronickej úrovni: jednoduchý oscilátor generoval elektrickú vlnu tvaru sinusoidy, ktorá po napojení na reproduktor produkovala tón frekvencie zhodnej s frekvenciou oscilátora. Ako sme si objasnili v časti 2.3 *Charakteristiky zvuku* (str. 21), dodaním harmonických násobkov s vhodnou amplitúdou získavame zvuk podobný danému hudobnému nástroju. Táto metóda sa nazýva **aditívna syntéza**, pre kvalitnejší zvuk však potrebujeme veľké množstvo dodaných harmonických zložiek, teda i generátorov impulzu. Z toho dôvodu je často vhodnejšia **subtraktívna syntéza**. Pri nej sa použitím upraveného oscilátora získa priebeh vlny zhodný so štvorcovou, trojuholníkovou či pílovou vlnou (obsahujúcich množstvo harmonických zložiek). Aplikovaním frekvenčného filtra možno upraviť amplitúdu jednotlivých harmonických frekvencií v signáli (prípadne ich celkom potlačiť), a tak získať želaný signál.

Dokonalejšiu manipuláciu so zvukom ponúka **syntéza frekvenčnou moduláciou**. Frekvenčná modulácia znamená upravovať nosný signál (sinusoida istej frekvencie) modulačným signálom tak, že nosnému signálu rastie či klesá aktuálna frekvencia vo zvolenom rozsahu podľa aktuálnej hodnoty (úrovne amplitúdy) modulačného signálu. V praxi má nosná vlna základnú frekvenciu želaného nástroja,

63 Viac informácií o ňom na stránke <http://www.chiariglione.org/MPEG/technologies/mpd-mps>. Môže používať lubovoľný kompresný formát, napr. MP3, AAC atď.

modulačným signálom je potom nejaký jej harmonický (resp. disharmonický, ak ide o zvony, činely, ...) násobok. Tento postup bol patentovaný v r. 1975 a patent odpredaný firme Yamaha (vypršal v r. 1995, odvtedy ho možno komerčne používať bez poplatkov) a nie je náročné realizovať ho v elektronických obvodoch. V praxi sa však používa zriedka – jeho úlohu zastúpila **fázová modulácia**, omnoho ľahšie realizovateľná v digitálnych čipoch. Pri tej má modulovaná vlna tvar $f(t) = \sin(2\pi f t + a s(t))$, kde $s(t)$ je modulačný signál a nosná vlna frekvencie f je fázovo modulovaná s amplitúdou a . Napriek faktu, že súčasné čipy realizujú fázovú moduláciu, z historických dôvodov sa takto syntetizovaný zvuk označuje ako syntéza frekvenčnou moduláciou (*FM synthesis*).

Ešte dokonalejší zvuk ponúka **syntéza použitím vzoriek** (*wavetable synthesis*⁶⁴). Zvukový čip má prístup k tzv. **vzorkám**⁶⁵, krátkym úryvkom (zväčša jedna-dve vlny) zo záznamu skutočných nástrojov. Tieto sú zaznamenané s istou nominálnou frekvenciou, pričom do požadovaného tónu sú prevzorkované (za pomoci interpolácie pôvodných hodnôt). Tento prístup je náročnejší – požaduje nielen omnoho väčšiu pamäť a výpočtový výkon, ale najmä vytvára potrebu dokonalej nahrávky a spracovania bohatej sady nástrojov. Pri drahších syntézach je zaznamenaných viac tónov rovnakého nástroja, aby sa dosiahla vyššia vernosť syntetizovaných zvukov, príp. sa zaznamenajú pre rovnaký tón aj odlišné nuansy hry (napr. silný alebo slabý úder do klávesu atď.).

Nakoniec, existuje aj metóda **syntézy zvuku na základe fyzikálneho modelovania nástroja** (*physical modelling synthesis*). Zakladá sa na metóde matematického vyjadrenia zvuku (vlnenia), ktorý vzniká v nástroji (napr. bubon ako úder paličky na dvojrozmernú membránu kruhového tvaru s daným polomerom a uchytením, útlmom atď.) Je veľmi náročná vo fáze návrhu aj realizácie (v súčasnosti je už výpočtová sila čipov na tento druh syntézy úplne postačujúca) a používa sa zriedka, podobnosť takto syntetizovaného zvuku so skutočnými nástrojmi je však výnimočne vysoká.

Ako v odseku 2.3.2 *Zvuky hudobných nástrojov* (str. 22) i tu musíme spomenúť javy pri hre na skutočných nástrojoch: nábeh, útlm, držanie, doznenie. Sú realizované zaznamenaním tzv. ADSR⁶⁶ obálok pre každý nástroj – tie určujú intenzitu zvuku v danej fáze zvukového prejavu, realizované sú teda úpravou hlasitosti výstupnej vlny tak, aby ju daná obálka „obklopovala“ (konvolúcia). Tieto obálky možno kvôli lepšej vernosti zvuku navrhnuť aj pre jednotlivé frekvenčné zložky. Pre rozdielnosť zvuku

64 V anglickej terminológii sa technika popísaná v tomto odstavci síce nazýva *sample-based synthesis*, tvorcovia zvukových kariet však z marketingových dôvodov používajú termín *wavetable synthesis*, preto sa prikláňame tiež k termínu „wavetable“. (Skutočná syntéza typu wavetable je zložitejšia, používa sa skôr v profesionálnych syntetizátoroch a nebudeme sa jej venovať).

65 Ide o sled PCM vzoriek. Aby sme odlišili tieto dva významy slova vzorka, budeme tento význam používať v kontexte wavetable.

66 Z anglických ekvivalentov spomenutých fáz zvukového prejavu: attack, decay, sustain, release.

v jednotlivých spomenutých javoch sa v prípade syntézy typu wavetable často osobitne mikrofonicke zaznamenáva každý spomenutý jav.

Na trhu boli rôzne druhy syntetizátorov dávno pred rozšírením osobných počítačov (nehovoriac o zvukových kartách), ktoré sa delili do rôznych kategórií. Lacné neprofesionálne modely ponúkali zväčša jednoduchú aditívnu či subtraktívnu syntézu (v lepších modeloch i FM), obmedzený počet nástrojov a možností. Drahé profesionálne modely podporovali zas okrem syntézy použitím vzoriek i množstvo drobných nastavení ovplyvňujúcich zvuk nástroja (sila úderu, modulácia, panoramatický efekt, vibrácia, dozvuk, legato atď.). Mali nepreberné množstvo rôznych nástrojov (harmonických nástrojov a perkusí), ktoré sa líšili od modelu k modelu (v lepšom prípade od výrobcu k výrobcovi); ďalšie vzorky nástrojov si používateľ mohol dodať do syntetizátora pomocou pamäťových kariet a rôznych proprietárnych rozhraní. Poriadok do tohto sveta priniesol štandard MIDI.

MIDI (*Musical Instrument Digital Interface*) je komunikačný protokol definovaný konzorciom Audio Engineering Society v r. 1983 (verzia 1.0). Jeho úlohou je prenášať informácie o hraných tónoch (nie hudbu samotnú) z jedného zariadenia do druhého – tak môže hranie hudby a jej samotná reprodukcia byť na odlišných miestach, dokonca v odlišnom čase (túto komunikáciu možno zaznamenať – napr. aj v súboroch, o ktorých si povieme nižšie). MIDI však dokáže omnoho viac – zvuková karta môže kontrolovať syntetizátor, ten zas zvukovú kartu alebo iný MIDI nástroj (konštrukčne je najjednoduchšie vyrobiť MIDI klaviatúru, ale existujú aj MIDI gitary, husle, bicie, ...), alebo dokonca pohybovať svetlami na pódium – to všetko vo vzájomnej synchronizácii všetkých napojených zariadení. Komunikácia rozhraním MIDI prebieha prostredníctvom krátkych správ („udalostí“), vyjadrujúcich napr. tón, jeho výšku, hlasitosť, úpravy na kanáloch, perkusie, časový kód na synchronizáciu atď. Jedno zo zariadení potom môže byť použité na reprodukciu (zväčša metódami syntézy opísanými vyššie), resp. záznam. Rýchlosť protokolu je 31 250 bitov za sekundu, čo mierne komplikuje zložitejšie pasáže (veľa udalostí nastane v jednom čase, treba ich však preniesť sekvenčne, čo istú dobu trvá). V súčasnosti existujú štandardy na prenos MIDI správ aj cez USB, ethernet a i.

MIDI 1.0 bolo určené pre profesionálnych hudobníkov, ktorí si naaranžovali svoje nástroje sami. Rozhranie im ponúkalo 128 nástrojov („programov“) a 16 kanálov, 128 rôznych tónov a 128 úrovní hlasitosti. Konfigurovateľnosť bola vysoká, predsa však vznikla potreba jednotného rozloženia nástrojov. V r. 1991 preto vznikol štandard **General MIDI (GM)**, ktorý mapoval na všeobecnú číselnú škálu konkrétne nástroje (napr. 1 – akustický klavír, 14 – xylofón, 72 – klarinet, 128 – výstrel z revolvera) a perkusie (obsadzujúce kanál 10). Navyše, GM špecifikoval minimálne množstvo tónov, ktoré zariadenie muselo byť schopné naraz hrať (polyfónia) – 24. Vytvorila sa tým živná pôda pre skutočné rozšírenie MIDI hudby vo svete PC či herných zariadení. V nasledujúcich rokoch bol dopĺňaný o ďalšie sady nástrojov a nastavenia (často išlo o „štandardy“ veľkých spoločností, nie štandardizačných

komisií). Jedným z rozšírení je aj SP-MIDI (*scalable polyphony*, teda škálovateľná polyfónia), s ktorým sa môžeme stretnúť v súčasných mobilných telefónoch.

Okrem vystúpenia „naživo“ však MIDI umožňoval najmä tvorbu a opakovanú reprodukciu hudobných diel. Keďže MIDI sám osebe je protokol, vznikla potreba ukladania jednotlivých správ do súboru, v ktorom by bol špecifikovaný aj čas, kedy je potrebné tieto správy poslať na MIDI výstup. Definovaný bol štandard SMF (standard MIDI file), používajúci príponu .mid⁶⁷. Komunikácia rozhraním MIDI je dátovo nenáročná, veľkosť súborov je teda extrémne malá (mnohominútové skladby môžu mať niekoľko desiatok kB). Neskôr sa objavil aj štandard XMF (extensible music file), ponúkajúci možnosť integrovať k hudobnému zápisu i vzorky (wavetable) použitých nástrojov.

Tento spôsob reprodukcie hudby (najmä v počítačových hrách a videohrách) dominoval do druhej polovice deväťdesiatych rokov dvadsiateho storočia, keď ho vytlačili postupne moduly (p. nižšie) a skutočné hudobné nahrávky (zväčša uložené v stratových formátoch ako MP3). MIDI bolo (a ostáva) podporované prakticky všetkými zvukovými kartami, v tých lacných (aj najviac rozšírených) však pre jednoduchú syntézu znela takto reprodukovaná hudba nekvalitne⁶⁸, preto si MIDI hudba vyslúžila povesť nekvalitnej alternatívy k digitálnemu záznamu hudby (pričom opak je pravdou – v profesionálnej sfére sa pri tvorbe hudby naďalej používa MIDI). Predsa však ostáva štandardom, ktorý ťažko možno nahradiť čímisi iným. Konzorcium MIDI Manufacturers Association už niekoľko rokov pracuje na koncepte, ktorý by ho doplnil – High Definition MIDI (HD-MIDI).

3.5.5 Hudobné sekvencery, moduly

Termínom **sekvencer** označujeme softvérový produkt určený na komponovanie hudby vo formáte MIDI, v ktorom je hudobný záznam zapisovaný do notovej osnovy. Príbuzný softvérový produkt označovaný anglickým slovom **tracker** realizuje zápis tónov zväčša alfanumericky a slúžia ako editory iných hudobných formátov – modulov.

Modul je označenie súboru, v ktorom je zapísané hudobné dielo vo forme tónov (a ich vlastností, rôznych efektov atď.) a nástrojov reprezentovaných wavetable vzorkami. Vzorky sú uložené vo forme PCM (niektoré novšie moduly poznali jednoduché kompresie vzoriek) pre referenčný tón (v závislosti od formátu). Program prehrávajúci moduly z týchto vzoriek v reálnom čase prevzorkováva zvolené tóny a hlasitosť, mixuje ich (moduly podporujú niekoľko kanálov) a posiela na zvukový

⁶⁷ Spĺňa špecifikáciu RIFF, až na chýbajúcu úvodnú hlavičku. To napravil štandard súboru s príponou .rmi (RMID, zo slov RIFF a MIDI) z dielne Microsoftu, ktorý však nezískal takú širokú podporu.

⁶⁸ Tvorca hudby vo formáte MIDI nemohol ovplyvniť samotný zvukový dojem zo svojej skladby – na drahších modeloch zvukových kariet s dobrou syntézou mohol znieť jeho výtvor ako nahrávka orchestra, na veľmi lacných zas znel celkom synteticky a nekvalitne.

výstup. Podobá sa teda na MIDI súbory, no obsahuje aj samotné nástroje a výstup sa získava výpočtom na procesore počítača. Rozšírili sa najmä na počítačoch tried Amiga a Atari, odkiaľ prenikli aj do sveta PC. Dôvod ich vzniku bol zrejmy: na začiatku deväťdesiatych rokov bola výpočtová sila počítačov a možnosti zvukových kariet dostatočné na spracúvanie vzoriek v reálnom čase. Syntéza frekvenčnou moduláciou neponúkala dostatočnú kvalitu (resp. vernosť) zvuku, slobodu v tvorbe (ľubovoľné wavetable vzorky – aj nezvyčajné zvuky a slová, určite nie prítomné v štandarde General MIDI) a nespĺňala požiadavku na rovnakú reprodukciu (nezávislú od implementácie). Karty so syntézou pomocou vzoriek boli drahé a riešili iba časť problémov. Moduly tak boli akýmsi prechodom medzi MIDI hudbou a skutočnými nahrávkami (na ktoré kapacita vtedajších médií nestačila, ich stratová kompresia bola zas pre výpočtovú náročnosť neuskutočniteľná).

Na trhu sa uchytilo viacero formátov modulov. Celkom prvým bol formát *.mod* (prípona *.mod*) so štyrmi kanálmi, medzi ďalšie rozšírené patrili *.s3m*, *.xm*, *.it* (16 až 32 kanálov, rôzne nastavenia ovplyvňujúce prehrávanie atď.). Pre jednoduchosť tvorby a pekný (i keď trochu syntetický) zvukový prejav za cenu len malej veľkosti súboru boli a sú veľmi obľúbené najmä pre tvorcov zo sveta demoscény, istý čas boli populárne aj medzi tvorcami hier (spomeňme napr. slovenskú hru *Quadrax* od firmy *Cauldron*).

3.5.6 Ochrana duševného vlastníctva

Každé „umelecké dielo ...“, ktoré je výsledkom vlastnej tvorivej duševnej činnosti autora, najmä ... hudobné dielo s textom alebo bez textu a audiovizuálne dielo“⁶⁹ je na území Slovenskej Republiky predmetom autorského práva. Autor má výlučné práva určiť, ako sa s jeho dielom bude nakladať, aký podiel na zisku bude z diela mať atď. V praxi je však situácia iná. Digitálne uloženie autorských diel (či už ide o počítačové programy, filmy, hudbu, ...) spôsobilo, že je neuveriteľne jednoduché rozmnožovať ich, reprodukovat', pozmeňovať, dokonca komerčne predávať (keď autorské dielo bolo odovzdané verejnosti na bezplatné používanie, ktosi ho však pozmení a predáva ako vlastné autorské dielo)⁷⁰. Preto vznikali rôzne iniciatívy, postupy, technológie v snahe obmedziť možnosť voľne kopírovať či reprodukovat' dáta. Narážajú na seba dve požiadavky: slobodné používanie legálne

69 Zákon č. 618/2003 Z. z. o autorskom práve ..., výňatok z § 7.

70 Dôsledok rozšírenia nelegálnych výmenných sietí a digitálnych technológií na spracovanie zvuku možno vidieť na uvedenom príklade: „V roce 1997, kdy se stále ještě vypalovačky vyskytovali v cenových výšinách deseti tisíc korun za kus a internet ve střední Evropě se začínal otevírat průměrnému uživateli, dosahovaly zisky hudebního průmyslu u nás (v ČR – poznámka autora) zhruba tři miliard korun. O pět let později, roku 2002, kdy stojí vypalovačka v průměru patnáct set, internet užívá třicet procent populace, na sídlištích je převaha broadband internetového připojení přes přípojku kabelové televize a na našem trhu je minimálně padesát různých kapesních přehrávačů MP3, jsou zisky českého hudebního průmyslu čtyřikrát nižší, na úrovni sedmi set padesáti milionů.“ (cit. NEUBAUER, Tomáš. Komprimované formáty v praxi. In: *Muzikus* 2004/01.)

zakúpeného diela v rámci možností ustanovených zákonom⁷¹ na jednej strane a snaha⁷² brániť nelegálnemu používaniu diela. Nelegálna je i snaha o odstránenie každej technológie ochrany, presnejšie podľa dikcie zákona: „(1) Neoprávnený zásah do autorského práva je aj: a) odstránenie alebo zmena akejkoľvek elektronickej informácie na správu práv⁷³, b) verejné rozširovanie originálu diela alebo jeho rozmnoženín vrátane ich dovozu, ako aj verejný prenos diela, pri ktorých boli elektronické informácie na správu práv odstránené alebo zmenené bez súhlasu autora. (2) Informácie na správu práv podľa odseku 1 sú údaje, ktoré identifikujú dielo, autora diela alebo nositeľa akéhokoľvek práva k dielu, informácie o termíne a podmienkach použitia diela a akékoľvek číslce alebo kódy, ktoré vyjadrujú takéto informácie, ak je ktorákoľvek z týchto informačných položiek pripojená k rozmnoženine diela alebo ak sa zobrazí v súvislosti s verejným prenosom diela.“⁷⁴

Autori si však zriedkakedy uplatňujú svoje práva sami, často sa spoliehajú na inštitúciu umožňujúcu kolektívnu ochranu ich práv. Poslanie takýchto inštitúcií je jasné, uveďme zdôvodnenie jednej z nich⁷⁵: „Sme tu preto, lebo je nereálne, aby si autor alebo vydavateľ sám vyhľadával používanie svojich diel. Rovnako nie je reálne, aby sa používatelia hudby kontaktovali pri každej skladbe s jej autormi, či vydavateľmi. SOZA na jednej strane používateľom udeľuje licencie na použitie hudby a vyberá za to od nich odmeny a na druhej strane vybrané odmeny rozdeľuje autorom a vydavateľom hudby. Recipročné zmluvy s organizáciami v zahraničí nás oprávňujú zastupovať na území Slovenska práva autorov hudobných diel z celého sveta.“ Na Slovensku pôsobí viac takýchto organizácií, menovite SOZA (Slovenský ochranný zväz autorský pre práva k hudobným dielam – <http://www.soza.sk>) a SLOVGRAM (nezávislá spoločnosť výkonných umelcov – <http://www.slovgram.sk>). Zvyšné tri organizácie (LITA, OZIS, SAPA) sa nevenujú zvukovým a hudobným dielam. Nad dodržiavaním autorských práv má záštitu Ministerstvo kultúry SR.

Okrem práva Európskej Únie, pod ktorej právnym vplyvom stojíme, je hlavným určujúcim faktorom v prípade ochrany autorských práv vývoj na území USA (ovplyvnené globálnosťou sveta a internetovej komunity a v súčasnosti dominantným postavením Spojených Štátov vo svetovom hospodárstve, politike a moderných technológiách). Momentálne znenie zákonov silne ovplyvňuje loby autorských spoločností, veľkých firiem a verejnej mienky (zastupovanej rôznymi organizáciami na ochranu práv spotrebiteľa). Každá táto stránka sleduje najmä svoje záujmy a je náročné vytvoriť „spravodlivý“ kompromis. Navrch v tomto boji majú zväčša autorské spoločnosti, zastupované najmä konzorciom RIAA (Recording Industry Association of America; loby vykonáva často spoločne s organizáciami MPAA –

71 Garantované práva (na citáciu, derivát práce atď.) sú v USA známe pod menom *fair use*.

72 V podstate by malo ísť o snahu štátu, hlavnými bojovníkmi na tomto poli sú však rôzne nahrávacie spoločnosti.

73 P. termín DRM nižšie.

74 Zákon č. 618/2003 Z. z. o autorskom práve ..., § 60.

75 Text zo stránky <http://www.soza.sk> [cit. 2006-03-21].

Motion Picture Association of America – a BSA – Business Software Alliance). Tie sa často vydajú cestou súdnych sporov s jednotlivcami používajúcimi, tvoriacimi či rozširujúcimi „nelegálny“ obsah. Našťastie, aj tí majú občas väčších zástupcov – na pôde USA je to najmä Electronic Frontier Foundation.

Jedna z prvých lastovičiek na poli právnej ochrany diel v digitálnom svete bol americký zákon o domácom zázname zvuku (angl. *Audio Home Recording Act*, AHRA) z r. 1992, ktorý bol reakciou na zariadenia DAT⁷⁶ od Sony. Tá bojovala za „práva používateľov“ – slobodnú možnosť záznamu a možnosť vytvárať kópie originálu pre osobné nekomerčné použitie. RIAA zas stála „na strane umelcov“, v opozícii k možnosti vytvárať identické digitálne kópie za lacný peniaz. Zákon priniesol akýsi kompromis – podiel z ceny každého zapisovateľného média a nahrávacej mechaniky šiel nahrávacím spoločnostiam ako potenciálne „odškodné“. Napriek pochybnej etickej správnosti takéhoto riešenia platia podobné zákony i u nás. Zaujímavosťou je i fakt, že ďalšie zákony (najmä známe *Digital Millennium Copyright Act* – DMCA z r. 1998 a podobné legislatívne úpravy inde vo svete), ktoré stavajú nahrávacie zariadenia nerešpektujúce DRM (p. nižšie) mimo zákon, vo svojej podstate urobili zákony typu AHRA zbytočnými – poplatky sa i napriek tomu naďalej vyberajú.

Rôzne technológie slúžiace účelu ochrany autorských práv vo svete digitálnej hudby (ale aj obrazu, kníh atď.) sa dajú spoločne označiť názvom **Digitálna správa práv** (*Digital Rights Management*, **DRM**). Informácie sú obyčajne súčasťou samotnej dátovej reprezentácie chráneného diela (ako metadáta) a určujú spôsob a rozsah povoleného použitia diela⁷⁷. Najnovšie sa často kombinujú s kryptografickými postupmi – dielo je šifrované, bežný používateľ ho neprečíta, pristupovať k obsahu môže len špecializované zariadenie, resp. softvér (prehrávač). Samozrejme, obe riešenia sa dajú ľahko obísť (zväčša niekto vytvorí program, ktorý túto ochranu prekoná, odstráni atď., tento program sa následne rýchlo šíri po internete), a aj keď je to nelegálne, málokedy sa podarí právne postihnúť vinníka⁷⁸. Tvorcovia nových technológií tak vytvárajú stále zložitejšie schémy, ktoré viac a viac obmedzujú používateľov, často v ich legálnych právach (napr. vytvorenie záložnej kópie produktu atď.) Známý je aj nedávny prípad, keď firma Sony na zvukové CD médiá umiestňovala špeciálny softvér s „agresívnym“ správaním, ktorého úlohou bolo sledovať používateľa a jeho nakladanie s produktom⁷⁹. DRM schémy sú obľúbené najmä v internetových obchodoch s hudbou (iTunes, Napster atď.). Najväčším problémom tejto technológie však v súčasnosti je: „1) ochrana súkromia pred neoprávneným zbieraním osobných údajov a on-line správania sa používateľa

76 P. ďalší odsek 3.5.7 *Úložné médiá*, str. 77.

77 P. tiež druhý odsek citovaného § 60.

78 Ako príklad poslúži známy nórsky programátor „DVD Jon“ Jon Lech Johansen, ktorý prekonal mnoho rôznych DRM ochrán na DVD, v systéme iPod a iTunes, súboroch WMV9 atď.

79 Viac informácií o tejto téme sa nachádza napr. na [cit. 2006-03-21] http://en.wikipedia.org/wiki/2005_Sony_CD_copy_protection_controversy.

prostredníctvom DRM, 2) interoperabilita, ktorá si vyžaduje kompatibilné systémy umožňujúce legálne použitie diela, napr. na rozličných platformách (napr. Windows vs Linux).⁸⁰ Slovensko sa riadi legislatívnym procesom v rámci EÚ, pričom v priebehu tohto roka (2006) sa plánuje schválenie odporúčania členskými štátmi EÚ ohľadom DRM; zároveň sa ráta zo zrušením poplatkov za zapisovateľné médiá v štýle zákona AHRA.

3.5.7 Úložné médiá

V tomto odseku si opíšeme štandardné spôsoby ukladania zvuku z pohľadu používateľa. Vynecháme analógové médiá (rôzne druhy gramofónových platní, analógový záznam na drôte a rozličných druhoch pásov), ktoré hrajú už len okrajovú úlohu, a prejdeme hneď k tým digitálnym.

Digitálna zvuková páska (*Digital Audio Tape*, **DAT**) sa objavila v polovici osemdesiatych rokov 20. storočia ako výsledok práce japonskej spoločnosti Sony a holandskej firmy Philips. Jej úlohou bolo poskytnúť možnosť ukladať zvuk na kompaktné médium (kazetu asi dvakrát menšiu ako v prípade známej a rozšírenej **kompaktnej audiokazety**), keďže CD bolo príliš drahé, rozmerné a v tej dobe neexistovali prepisovateľné CD médiá. DAT umožňuje záznam pri 48, 44,1 a 32 kHz a 16-bitovej hĺbke (stereo) vo formáte PCM. Uchytil sa skôr ako poloprofesionálne riešenie, medzi bežnými používateľmi nemal veľký úspech. Časom sa z neho vyvinul aj štandard na zálohovanie dát vo svete počítačov (DDS – Digital Data Storage) – s podobným osudom. Predajnosť zariadení pracujúcimi s DAT je v súčasnosti nízka, preto Sony posledné kusy expedovala v decembri minulého roka (2005). Formát sa používa naďalej, aj keď už zriedka, stále vytláčaný modernými harddiskovými zariadeniami.

MiniDisc (MD) uviedla firma Sony na začiatku roka 1992 ako ďalšiu snahu o nahradenie kompaktnej audiokazety. Rozmery sa podobá na 3,5-palcovú disketu, v skutočnosti však ide o magneto-optické médium. Poskytuje dostatočnú kapacitu (74 min., 80 min. pri 160-175 MB zvukových dát) použitím stratovej kompresnej schémy (algoritmus ATRAC, zväčša 292 kbps, nezávislé kódovanie oboch kanálov – p. časť 4.6, str. 102) pri nenáročnej cene médií a zariadení. Získal si veľkú popularitu v Japonsku; v Európe a USA však nebol natoľko rozšírený, preto nemá do budúcnosti veľkú perspektívu (aj keď ešte v roku 2005 Sony uvádzala rôzne doplnky a vylepšené verzie tohto formátu, ktorých v súčasnosti už existuje nepreberné množstvo). Prakticky jediný spôsob spojenia MD zariadení s PC je však proprietárny softvérový produkt firmy Sony menom SonicStage.

Digitálna kompaktná kazeta (*Digital Compact Cassette*, **DCC**) vyvinuli firmy Philips a Matsushita (známa skôr svojou produktovou líniou *Panasonic*) na konci

80 Citát [2006-03-21] zo stránok MK SR: <http://www.culture.gov.sk/main/index.php3?id=1546>.

r. 1992 ako konkurenciu MD od Sony. Výhodou bol najmä fakt, že médiami boli obyčajné kompaktné audiokazety, ktoré sa dali kúpiť na trhu za lacný peniaz – nové zariadenia tak dokázali pracovať aj s nimi, aj s DCC. Kvalita bola vysoká (proprietárny algoritmus PASC založený na MP1, 48 kHz, 16 bitov, stereo, 384 kbps), pre jednoduchosť média však dosiahnutá kapacita nekonkurovala iným médiám (45 min.). Ľudia si však už dávno zvykli na pohodlie náhodného prístupu, ktoré im poskytovali CD a MD, preto boli tieto páskové zariadenia trhom prijaté s rozpakmi. Philips upustil od výroby DCC zariadení v novembri 1996.

Kompaktný disk (*Compact Disc*, **CD**, resp. CDDA – *Digital Audio*), hoci vznikol prv než predošlé médiá, uvádzame až tu – na rozdiel od nich je totiž ešte stále hojne používaný a pokladá sa de facto za štandardný spôsob ukladania hudby, pripomínajúc tak gramofónové platne, ktoré tiež sprevádzali ľudstvo dlhú dobu (CD ich v predajnosti za rok predbehlo v r. 1988). CD je spoločným dielom firiem Sony a Philips a uvedené bolo na konci roka 1982. Zvukové dáta sú uložené na optickom médiu s priemerom 120 mm (existuje aj menší štandard – 80 mm) a kapacitou 74 min.⁸¹ (resp. 20 min. pre menší formát; neskôr však boli predstavené disky s vyššími kapacitami ako 80, 90, 99 min.). Na disku je uložená hudba vo forme PCM, čo vedie (pri 44,1 kHz, 16 bitov, stereo) k dátovému toku 1411,2 kbps (176,4 kBps) a pre 74 min. médium veľkosti zvukových dát 783 MB. Vzhľadom na špecifikáciu je frekvenčný rozsah CD niečo vyše 20 kHz a maximálna dynamika nahrávky 96 dB⁸². CD už pár rokov po uvedení získalo veľkú popularitu – bolo ozaj kompaktné, pohodlné, odolné voči škrabancom (vďaka použitiu samoopravných kódov), poskytovalo vysokú kapacitu aj kvalitu, cena výroby pritom nebola vysoká. Čoskoro nato vznikli rôzne deriváty CD (pričom CD teraz často označujeme **Audio CD**, aby sme ho odlišili od nasledujúcich variantov), ako napr. CD-ROM, CD-R, CD-RW (s najrozšírenejšou kapacitou 80 min., teda pribl. 703 MiB⁸³ dát – dôsledok iného kódovania, použitého kvôli vyššej dátovej ochrane), ktoré sú čitateľovi dobre známe. Celkový úspech rodiny CD médií potvrdzuje fakt, že v r. 2004 bolo predaných asi 30 miliárd nosičov⁸⁴. CD v sebe nezahrnovalo žiadnu ochranu pred kopírovaním, s príchodom CD-ROM jednotiek do sveta PC teda bolo jednoduché vytvárať kópie originálnych nosičov. Od r. 2002 začali mnohé spoločnosti vydávať audio CD s integrovanými chybami (od výroby vytvorené nezhody medzi dátami a kontrolnými

81 Priemer kompaktnej audiokazety bol 11,5 cm, CD s takým rozmerom by so zachovaním ostatných parametrov mohlo zaznamenať 60 min. hudby. Existuje rozšírená legenda, ktorá hovorí, že priemer CD bol zvýšený preto, aby sa naňho zmestil záznam Beethovenovej deviatej symfónie (trvajúcej niečo vyše 70 minút). Nie je pravdivá (ako uvádza spoluvorca CD, Kees Immink z firmy Philips, p. IMMINK, K. A. S. The Compact Disc Story). V skutočnosti priemer 12 cm vznikol jednoduchým zaokrúhľením, ktoré len minimálne zväčšilo rozmer, poskytlo však oproti LP platniám dostatočný priestor na umiestnenie bonusových skladieb.

82 Ide o dynamický rozsah hladiny akustického tlaku. Ak berieme do úvahy rôznu citlivosť na rôzne frekvencie (napr. váhovanie ITU-R 468, p. odsek 2.2.2 *Vnímanie hlasitosti*, str. 13), je najmenšia hlasitosť v rozsahu 20 Hz až 20 kHz -86 dB oproti úplnému rozsahu. Psychoakusticky je však táto hodnota vyššia vďaka použitiu rozptyľovania v procese kvantovania.

83 1 MiB = 2¹⁰ bajtov = 1 048 576 bajtov. 703 MiB = 737 MB (1 MB = 1 000 000 bajtov).

84 P. http://en.wikipedia.org/wiki/Compact_disc [cit. 2006-03-22].

sumami, dátové stopy v rozpore so špecifikáciou atď.), o ktorých vedeli, že sú ignorované väčšinou CD prehrávačov, predsa však nie CD-ROM mechanikami na trhu, ktorým to sťažovalo / znemožňovalo kopírovanie. Tieto postupy sa však stretli s veľkou nevhodnosťou konzumentov.

Super Audio CD (SACD) uviedli Sony a Philips v r. 1999 ako vylepšenú verziu CD (po prehratom boji o formát DVD-Audio). Ako médium využíva mierne upravenú špecifikáciu DVD (zväčša jednovrstvové médium kapacity 4,7 GB, dvojvrstvové médium kapacity 8,46 GB, resp. dvojvrstvové médium, kde prvá vrstva je kompatibilná s CD, druhá zas s SACD; rozdiel oproti DVD je najmä v premenlivosti šírky stôp – to je spôsob kódovania dodatočných informácií na ochranu pred kopírovaním), tok dát tvorí jednobitové kódovanie zvuku (2 822,4 kHz, teda 64-krát vyššie než CD) formou PDM⁸⁵ s marketingovým názvom *Direct Stream Digital* (dátový tok je znížený proprietárnym bezstratovým postupom menom *Direct Stream Transfer* založenom na dopredne adaptívnej predikcii⁸⁶). Každé médium povinne obsahuje dvojkanálový záznam, navyše môže obsahovať i viackanálový záznam – 4.0 alebo 5.1 (v SACD označované *multi-channel*, nie *surround sound*). PCM a PDM možno len ťažko priamo porovnávať, v propagačných materiáloch sa uvádza dynamický rozsah 120 dB pri frekvenčnom rozsahu 20 Hz až 20 kHz a najvyššia zaznamenateľná frekvencia 100 kHz. Formát sa preto ujal hlavne medzi audiofilmami (napr. v máji r. 2005 vyšlo 3000 titulov, 40% z nich tvorila klasická hudba) – pre normálnych používateľov je aj kvalita CD veľmi vysoká, preto nemajú potrebu získať „vyššiu vernosť“. Formát však stále hrá svoju rolu a Sony ho neodpísalo – hracia konzola Playstation 3, ktorá vyjde v novembri 2006, ho bude natívne podporovať. V súčasnosti nie je známy prípad nelegálneho kopírovania SACD – obsah je šifrovaný (80-bitová šifra) a pre odlišnosť fyzickej realizácie disku oproti DVD pre SACD neexistujú počítačové mechaniky, len licencované prehrávače, neposkytujúce pôvodný dátový tok.

DVD-Audio (DVD-A) pochádza z dielni firmy Matsushita, ktorých návrh konkuroval návrhu firiem Sony a Philips v boji o priazeň zoskupenia DVD Forum. Na trhu je prakticky od r. 2000. Fyzicky ide o klasické DVD médium, uložený zvuk je zväčša vo forme PCM, 24 bit, 44,1 kHz až 192 kHz, jeden až 6 kanálov. (V prípade 192 kHz je to najviac stereo – pre 4700 MB médium aj to vystačí len na 70 minút záznamu; pre 96 kHz podporuje špecifikácia záznam typu 5.1) Zvuk však možno kódovať bezstratovou kompresnou schémou MLP⁸⁷, čo takmer zdvojnásobuje kapacitu média. Rozšírené sú i „hybridné“ DVD-A, ktoré na druhú stranu média ukladajú hudbu stratovo vo formáte Dolby Digital alebo DTS, aby ju mohli prehrávať i klasické DVD-Video prehrávače, resp. média DualDisc, kde sa na druhej strane nachádza klasické CD s rovnakým záznamom. Na DVD-A sa môžu nachádzať i videosekvencie, interaktívne menu, texty piesní atď., podobne ako na DVD-Video.

85 P. odsek 3.1.2 Kvantovanie, str. 41.

86 P. odsek 3.4.2 Kódovanie lineárnou predikciou, str. 60.

87 P. časť 5.10 Meridian Lossless Packing (MLP), str. 131.

DVD-A má nielen šifrovaný obsah, ale používa i princíp **digitálneho vodoznaku**⁸⁸, jednoznačne identifikujúci zhodu média a obsahu⁸⁹. Značnou nevýhodou je, že DVD-Audio prehrávače kvôli licenčnej politike zoskupenia DVD Forum požadujú priamy analógový šesťkanálový výstup do reprodukčnej sústavy (občas sú vybavené aj stereo výstupom zníženej kvality), všetko pre ochranu pred kopírovaním.

3.5.8 Zvuk a video

Gramofónové platne, CD a rádiá majú silnú konkurenciu v médiách na ukladanie obrazu, televízii, videohráč, interaktívnych dielach atď. Tie totiž rovnako zahŕňajú zvuk, ale aj čosi navyše – obraz, interakciu. Zvuk sa stáva akoby „bočným“ médium a druhoradou, predsa veľmi podstatnou súčasťou obsahu týchto médií. Faktom je, že človeku napr. pri pozieraní filmu viac prekáža nekvalitný zvuk ako nekvalitný obraz. Dokazuje to i nasledujúci prehľad rozličných štandardov na ukladanie filmovej tvorby, ktorý ponúkame na doplnenie predošlého odseku.

Prvou lastovičkou na poli digitálneho ukladania zvuku spolu s obrazom bola technológia **Laserdisc**. Philips a MCA (Music Corporation of America) ho uviedli na americký trh v r. 1978. Ide o extrémne veľký disk (priemer 30 cm), na ktorom je obraz (kompozitný signál) uložený analógovo moduláciou typu PWM. Zvuk môže byť prítomný v analógovej forme (stereo), ale aj digitálnej (PCM, AC-3, DTS, 2 až 6 kanálov – obmedzené len dátovým tokom a dekódovacími zariadeniami na strane používateľa; často obsahoval viac zvukových stôp, napr. komentár k filmu atď.). Formát je na ústupe – v USA sa uchytil medzi videofilmi (ktorí neskôr prešli na DVD-Video, štandard určený pre všetkých), v súčasnosti je tam už menej než milión prehrávačov; v Japonsku mala technológia väčší úspech, v súčasnosti sú tam asi 4 milióny prehrávačov (10% domácností); inde vo svete sa nikdy neujal. Konkurentom je mu omnoho kvalitnejší, lacnejší a pohodlnejší formát DVD-Video.

Video CD (VCD) bol štandard navrhnutý spoločnosťami Sony, Philips, Matsushita a JVC v r. 1993, používa štandardný CD disk a jeho úlohou bolo konkurovať videokazetám. Tomu zodpovedala aj nízka kvalita obrazu (MPEG-1, 1150 kbps, 352×240 pre NTSC, 352×288 pre PAL). Zvuk bol však dosť kvalitný (MP2, 224 kbps, stereo, 44,1 kHz). Tento štandard sa ujal len v Ázii – nielen

88 Digitálny vodoznak (angl. *digital watermark*) je označenie technológií snažiacich sa do digitálnej reprezentácie multimedialného obsahu (zvuk, obraz) vniesť nebadateľné „značky“, ktoré nepočúť resp. nevidiať, no poskytujú (v prípade znalosti postupu ich dekódovania) dodatočné informácie, ako napr. identifikáciu zdroja, skladby, výrobného kódu atď. Jeden z príkladov môže byť aj A©WA od Fraunhoferovho Inštitútu pre grafické spracovanie dát (<http://www.igd.fraunhofer.de/igd-a8/syscop/audio.html>).

89 Jediný rozumný spôsob prekonania ochrany je získanie jedného z dešifrovacích kľúčov – napr. v r. 2004 sa ho podarilo získať rozborom kódu programu WinDVD. Nepodarilo sa však odstrániť vodoznak (teda po napálení dát na DVD-R médium sa prehrávače zaseknú, zistiť rozdiel vo výrobnom kóde medzi zvukovým záznamom a médium). Navyše, novovyrobené médiá možno vybaviť kódom, ktorý sa nedá dekódovať kľúčom, o ktorom sa vie, že je vyzradený. Viac informácií na stránke <http://www.highfidelityreview.com/news/news.asp?newsnumber=14550899>.

neponúkal žiadne skutočné zlepšenie kvality, ale najmä umožňoval lacné kopírovanie filmov a nemal žiadne ochranné prvky, čo spôsobilo jeho odmietnutie zo strany nahrávacích spoločností z USA a Európy. V r. 1998 sa tento formát dočkal vylepšenej verzie, **Super VCD (SVCD)**, s kvalitnejším obrazom a lepším zvukom (128-384 kbps, podpora viac zvukových stôp, najviac 8 kanálov použitím technológie MPEG Multichannel). Vytvorili ho pôvodné spoločnosti v spolupráci s čínskymi vedcami na objednávku čínskej vlády, v snahe oslobodiť sa od licenčných poplatkov za DVD. Tento formát sa ujal v mnohých domácnostiach nielen v Juhovýchodnej Ázii, ale aj v USA a Európe, kde sa používa najmä na amatérske snímky (rodinné oslavy zaznamenané lacnými kamerami atď.).

DVD-Video je asi hlavným dôvodom vzniku štandardu **DVD** (neoficiálne významy: *Digital Versatile Disk*, *Digital Video Disc*). Uviedlo ho v r. 1996 zoskupenie DVD Consortium (potom premenované na DVD Forum), hlavný podiel na technológii týchto 12 cm optických diskov mala firma Matsushita. Obrazovej zložke sa nebudeme venovať, zvuk musí byť vo formátoch PCM, DTS, MP2 alebo AC-3. Zväčša sú zvukové stopy (býva ich viac) kódované vo formáte AC-3, 48 kHz, 16 bit (5.1 a 448 kbps alebo 384 kbps, resp. 2.0 a 192 kbps) a sú súčasťou celého toku dát, chráneného šifrovaním a regionálnymi kódmi⁹⁰. Zvukový výstup z prehrávačov je analógový (zväčša downmixovaný na stereo) aj digitálny (S/PDIF, prenáša pôvodný zvukový tok dát). Pre súčasnú rozšírenosť a popularitu DVD-Video je výrazný útlm v predajnosti nosičov nepravdepodobný. O pár rokov sa mu však vážnym konkurentom stanú disky novej generácie ako HD DVD a BD.

HD DVD je návrhom firiem Toshiba, NEC a Sanyo (s podporou spoločností z IT priemyslu ako Microsoft, HP, Intel), ktorý DVD Forum v r. 2003 ustanovilo ako nástupcu DVD pre obraz s vysokým rozlíšením. Fyzicky používa takmer rovnaké médiá, kapacita jednej vrstvy je však 15 GB (oproti 4,7 GB na DVD). Plánujú sa jednovrstvové a dvojvrstvové médiá (existujú však prototypy trojvrstvových). Výhodou je fakt, že výrobné linky na médiá nepotrebujú takmer žiadne úpravy v technológii výroby. Navyše, čítanie DVD a HD DVD možno kombinovať v jednej optike, dokonca na jednom médiu môžu byť dve vrstvy – DVD a HD DVD, uľahčujúc tak filmovým spoločnostiam i používateľom prechod na nový formát. Medzi podporované formáty zvuku patria PCM, Dolby TrueHD, DTS-HD, DD+, ochranné prvky sú vylepšením technológií prítomných v DVD-Audio. Prvé mechaniky a filmy prišli na trh v marci 2006.

Blu-ray Disc (BD) je spoločným návrhom mnohých firiem, medziiným Sony, Philips, Pioneer (tiež s veľkou podporou firiem z IT branže ako Apple, Dell, no i HP), ktorého špecifikácia bola uverejnená už v r. 2002. Použitie zložitejšej optiky a ochrannej vrstvy hrúbky len 0,1 mm (šesťkrát menej než DVD a HD DVD, preto je ich povrch chránený špeciálnym polymérom menom Durabis) umožňuje vysokú

⁹⁰ Prvú ochranu možno ľahko prekonať pomocou hrubej sily, druhú upraveným softvérom či prehrávačmi.

kapacitu 25 GB na vrstvu, za cenu náročnejšej výroby mechaník i médií. Predpokladá sa používanie jednovrstvových a dvojvrstvových médií (existujú však prototypy so štyrmi či ôsmimi vrstvami, JVC dokonca vyvinulo prototyp kombinovaného DVD/BD média). Podporované zvukové formáty zahŕňujú PCM, Dolby TrueHD, DTS-HD a DD+. Čo sa týka bezpečnostných opatrení, sú podobné štandardu HD DVD. Podpora mnohých nových technológií (napr. Java zabezpečuje interaktivitu, dokonca sa ráta s prístupom na internet) a podpora veľkých spoločností zo sveta IT zaručujú, že BD sa udrží – ba dokonca v dlhodobom horizonte možno aj zvíťazi nad HD DVD, a to i napriek vyššej cene.

Existujú aj rôzne iné alternatívne formáty, no schopnú konkurenciu ponúka len tchajvanský FVD (Forward Versatile Disc – oproti DVD má navyše podporu zvukového formátu WMA), aj to len na území tohto štátu. Otázny je formát VMD (Versatile Multilayer Disc), ktorý má slabú podporu, má vyjsť v treťom kvartáli roku 2006 a zvukové formáty, ktoré bude podporovať, nie sú bližšie známe. Jeho primárnym cieľom je však indický a čínsky trh.

3.5.9 Rozhrania na prenos zvuku

Základným a najviac rozšíreným rozhraním na prenos zvuku je klasický dvojžilový (mono zvuk) alebo trojžilový (stereo) elektrovodivý kábel, zväčša typu jack (3,5 mm), výnimočne RCA (cinch). Okrem týchto analógových však existujú i digitálne rozhrania.

V súčasnosti najrozšírenejším štandardom na prenos audiosignálu digitálnym spôsobom je štandard **S/PDIF** (*Sony / Philips Digital Interface Format*), ktorý je zväčša realizovaný koaxiálnym káblom alebo optickým vláknom (to nie je náchylné na elektromagnetický šum). Podporuje zvuk vo formáte PCM (najviac 2 kanály, 48 kHz, 24 bitov; výnimočne 4 kanály a 16 bitov – nad pôvodnú špecifikáciu) a komprimovaný zvuk (vo formáte MPEG-2, AC-3, DTS, atď. s dátovým tokom nepresahujúcim 3 Mbps). V druhom prípade sa v koncovom zariadení musí nachádzať dekodér, ktorý skomprimovaný zvuk spracuje. Na začiatku vzniklo S/PDIF za účelom prenosu signálu v rámci hi-fi zostáv (najmä drahé CD prehrávače pre audiofilov, aj keď využitie našlo aj v digitálnych páskových záznamových zariadeniach), kde sa DAC umiestňoval často až pred zosilňovač v reproduktore, aby sa zabránilo prípadnému rušeniu signálu elektromagnetickým smogom. Dnes sa s ním stretáme najmä v systémoch „domáceho kina“, na prenos zvykového signálu medzi DVD prehrávačom a reproduktorovou zostavou (resp. zosilňovačom, na ktorý je napojená) opatrenou vhodným dekodérom.

Rozhraním budúcnosti pre konzumný trh sa však má stať **High-Definition Multi-media Interface (HDMI)**. Navrhlo ho konzorcium firiem Sony, Philips, Hitachi, Matsushita, Thomson, Toshiba a Silicon Image na konci roku 2002. Ako

názov napovedá, ide o multimedialne rozhranie, okrem obrazu (vo formáte zhodnom so starším DVI) teda prenáša aj zvuk (celkovo dátový tok pre jednu linku do 5 Gbps). Pre univerzálnosť riešenia sú všetky dáta nekomprimované. HDMI podporuje osem kanálov vo formáte PCM (do 192 kHz a 24 bitov, resp. 2 822,4 kHz a 1 bit). Ochrana pred nelegálnym kopírovaním je garantovaná zložitým proprietárnym návrhom **High-Bandwidth Digital Content Protection (HDCP)** od firmy Intel⁹¹. V prípade prepojenia HD zariadenia (High-Definition, či už ide o obraz alebo zvuk) cez iný kábel než HDMI, (resp. zo zariadením nepodporujúcim HDCP) sa kvalita obrazu aj zvuku znižuje – na 48 kHz, 16 bitov, stereo. Napriek chybám v implementácii DRM ochrany bude toto rozhranie dominovať, keďže podporu mu vyjadrili nielen veľké spoločnosti, ale aj úrady v Európe či USA. Uchytiť sa má najmä v segmente HDTV. (Vysielanie High Definition TV, teda televízie s vysokým rozlíšením – 1280×720, resp. 1920×1080 – je už dnes stále rozšírenejšie, najmä v západných krajinách.)

Treba podotknúť, že HDMI je určené predovšetkým do segmentu konzumnej elektroniky, v počítačovom priemysle však podľa mienky mnohých nestačí – napr. pre rozlíšenie monitora 1600×1200 umožňuje len 85 snímok za sekundu (frekvencia pixelov 165 Mhz pre jednolinkové spojenie). Riešením by mohol byť dvojlinkový kábel (tiež súčasť špecifikácie), implementácia je však náročnejšia. Preto je možné, že sa vo svete PC skôr stretne so štandardom **Unified Display Interface (UDI)**, dátový tok 16 Gbps, prenáša však len obraz, a to vo formáte kompatibilnom s HDMI a DVI, presná špecifikácia UDI má uzrieť svetlo sveta v druhom kvartáli tohto roku). Zaujímavé bude aj sledovať vývoj a prijatie štandardu **DisplayPort**, ktorý má byť použiteľný bez poplatkov (spravuje ho VESA, medzi firmami pracujúcimi na jeho návrhu však pracovali spoločnosti ako ATI, Dell, HP, Philips či Samsung) a je flexibilnejší než všetky doterajšie formáty a rozhrania (s vynikajúcou rýchlosťou do 10,8 Gbps, čo mu umožňuje zahrnúť toky dát predošlých existujúcich štandardov do seba). Prenáša aj zvuk, no presná špecifikácia nie je k dispozícii (podľa predpokladov VESA má byť vypracovaná na konci prvého kvartálu tohto roku).

3.5.10 MPEG

Skratka **MPEG** znamená *Moving Picture Experts Group* (teda skupina expertov na pohyblivé obrázky) a označuje pracovnú skupinu pri organizáciách ISO (International Standards Organization) a IEC (International Electro-Technical Commision; oficiálne označenie skupiny MPEG je ISO/IEC JTC1/SC29 WG11), ktorá sa od r. 1988 pod vedením talianskeho vedca menom Leonardo Chiariglione zaoberá vývojom štandardov na kódovanie a kompresiu videa, zvuku a médií vôbec. Táto skupina, resp. vedci, ktorí za ňou stoja, významnou mierou ovplyvnili vývoj

91 Tieto bezpečnostné opatrenia však majú veľa „osudových chýb“, ktoré ich umožňujú prelomiť, p. <http://apache.dataloss.nl/~fred/www.nunce.org/hdcp/hdcp111901.htm>, <http://www.macfergus.com/niels/dmca/cia.html>.

zvukových formátov spomínaných v našej práci. Tu si spomenieme niekoľko štandardov z ich dielne. Presný popis formátov MP3 a AAC môže čitateľ nájsť v príslušných častiach práce (4.2, str. 89, resp. 4.3, str. 94). (Poznámka: v bežne používanej reči sa pod slovom MPEG myslí aj súbor štandardov, ktorý táto skupina vyvinula a spravuje.)

Najstarší štandard (dokončený roku 1992) má označenie **MPEG-1** a popisuje jednoduché techniky na kompresiu zvuku a obrazu (s nízkym rozlíšením, vhodným na digitálnu telefóniu, vysielané⁹² video atď.) na základe psychoakustickej a psychovizuálnej redundancie zdroja. Časť štandardu zaoberajúca sa zvukom navrhuje tri rôzne stupne komplexnosti kompresie, v závislosti od želanej výpočtovej náročnosti: **MP1**, **MP2** a **MP3** (celým názvom MPEG-1, part 3: audio, layer 1-3). Podporované vzorkovacie frekvencie sú 32 kHz, 44,1 kHz a 48 kHz pri 8-bitovej alebo 16-bitovej hĺbke; počet kanálov bol nanajvýš dva (možnosti mono, dva nezávislé mono kanály a v prípade MP3 aj stereo – nezávislé alebo spojené kódovanie). Dátový tok siahal od 32 kbps po 320 kbps (MP3, pre MP2 je stropom 384 kbps, pre MP1 až 448 kbps).

MPEG-2 (špecifikácia dokončená roku 1994), čo sa týka kompresie zvuku, v časti MPEG-2 part 3 dodáva možnosť kompresie viacerých kanálov (do 5.1 – pri ňom môže dátový tok siahať do 1024 Mbps – MPEG Multichannel; zväčša v kombinácii s MP2) a umožňuje použiť aj menšie vzorkovacie frekvencie (16 kHz, 22,05 kHz, 24 kHz)⁹³ pri dátovom toku 8 kbps až 160 kbps. Vzniká však aj nová vetva kompresie zvuku (part 7) – **AAC** (*Advanced Audio Coding*), nekompatibilná s MPEG-1, no efektívnejšia. Kompresia obrazu sa naďalej zdokonaľuje (za cenu vyššej výpočtovej náročnosti) a je vhodná na obraz s vyšším a vysokým rozlíšením (PAL a HDTV) pri vysokých dátových tokoch.

MPEG-4⁹⁴ nevznikal kvôli dosiahnutiu vyššej kompresie, ale predovšetkým preto, aby poskytol omnoho vyššiu funkcionálnu ako predošlé verzie. Zahrnoval v sebe databázové služby, interaktivitu, mobilné aplikácie atď. To všetko s cieľom poskytnúť multimediálne informácie v najrozličnejšej forme. Kompresia obrazu i zvuku siahala od veľmi nízkeho dátového toku (napr. 2 kbps ako dolný limit na kompresiu hovoreného slova) k vysokým dátovým tokom (v prípade zvuku realizovaných algoritmom AAC). MPEG-4 part 3 rozširuje špecifikáciu AAC medziiným o technológie HE-AAC, AAC-SSR a BSAC. Z formátov na kódovanie zvuku spomeňme ešte štruktúrovaný zvuk (angl. structured audio), formát popisujúci

92 Angl. *streaming*, teda šírenie toku (prúdu) dát (napr. cez internet) a ich prehrávanie počas prijmu signálu. Je alternatívou prehrávania súborov, teda lokálne (úplne) dostupných dát.

93 Proprietárne rozšírenie MPEG-2 navrhnuté Fraunhoferovým Inštitútom pre integrované obvody, tiež označované „MPEG-2.5“, ponúka i vzorkovacie frekvencie 8 kHz, 11,025 kHz, 12 kHz.

94 MPEG-3 mal pôvodne ponúknuť špecifické postupy pre kompresiu obrazu s vysokým rozlíšením (HDTV), pri vtedajšej výpočtovej sile zariadení však nebolo možné využiť mnohé navrhované postupy, preto sa od nich upustilo a výskum sa spojil s ešte prebiehajúcim postupom štandardizácie MPEG-2. Navyše, skratka MP3 sa v ušiach verejnosti spájala so zvukovým formátom, preto MPEG od označenia MPEG-3 celkom upustil.

model syntézy zvuku a inštrukcie na reprodukciu. Trh však oň zatiaľ neprejavil záujem. Zavádza sa podpora **škálovateľného kódovania zvuku** – možnosti meniť (zmenšovať vynechaním časti dát⁹⁵) dátový tok, a tým docieľiť znižovanie výslednej kvality zvuku. Súčasťou štandardu je aj podpora pre vektorové kvantovanie frekvenčnej domény – štandard TwinVQ (nezískal veľkú popularitu kvôli vysokej výpočtovej náročnosti). Kompresia obrazu sa tiež zdokonaľuje. Prvá verzia MPEG-4 bola schválená roku 1998, odvtedy sa však tento štandard dočkal mnohých dodatkov a je ešte v súčasnosti stále obohacovaný o nové technológie a postupy. Obsahuje podporu pre DRM schémy.

Treba podotknúť, že spomínané zvukové štandardy skupiny MPEG sú do značnej miery kompromisom medzi normatívnosťou a informatívnosťou. Štandard definuje formát výstupného toku dát a štandardný (referenčný) dekodér, konštrukcia kodéra je však ponechaná voľne na tvorcu. Dokonca sú prípustné dekodéry spĺňajúce stanovenú odchýlku od referenčného dekodéra – umožňujú implementovať ich aj použitím celočíselnej aritmetiky.

Konzorcium MPEG ešte vyvíja štandardy s označením MPEG-7 (zaoberajúce sa štandardizáciou popisu multimediálneho obsahu) a MPEG-21 (zaoberajúce sa správou a univerzálnym popisom práv pri používaní a šírení multimédií).

3.5.11 Predpokladaný vývoj

V súčasnosti žijeme v digitálnej ére, používanie digitálnych technológií a služieb sa neustále prehlbuje. Človek predvčerajška potreboval „vlastniť“ nahrávku, aby vedel, že je jeho, že s ňou môže manipulovať, že si ju môže vždy vypočuť. Ak by však mal vlastniť všetky nahrávky, ktoré sa mu páčia, vyšlo by ho to draho. Človek včerajška si zvykal na isté kompromisy, nemusel fyzicky „vlastniť“ zvukový záznam, stačilo, ak ho vedel v prípade potreby získať – počnúc piesňami na želanie v rádiách, končiac sieťami peer-to-peer na výmenu digitalizovanej hudby.

Človek dneška je iný. Zvykol si na svet informácií, svet pohodlia. Do popredia už celkom prenikla *služba*. Aj keď nahrávacie spoločnosti reagujú pomaly (rýchlejšie v implementácii stále komplikovanejších DRM schém ako v poskytovaní nových služieb, vhodných v informatizovanom svete), svet sa mení. Ľudia počúvajú satelitné rádiá (v USA najmä v aute⁹⁶), online rádiá a digitálne terestriálne rádiá. Najnovšie sa stávajú účastníkmi *podcast* sietí, sťahujú si z internetu hudbu na svoje prenosné prehrávače, v snahe počúvať ich kedy chcú, kde chcú a ako chcú, a najmä bez reklám. V súčasnosti si médiá určené pre spotrebnú (tzv. „čiernu“, „obyvateľskú“) elektroniku (napr. DVD-Video) môžeme pozrieť už i na počítači a počítačové multimédia

95 P. odsek 3.4.3 *Hybridné a škálovateľné kódovanie*, str. 62. Porovnaj s technológiou *peeling* formátu Ogg Vorbis, časť práce 4.9, str. 108.

96 Dokonca väčšina Američanov počúva hudbu najmä v aute – p. SCHADLER, T. et al. *The Future of Digital Audio*. 2005. Figure 1.

môžeme často spokojne pozerať použitím spotrebnej elektroniky. Spolu s prienikom „multimediálnych centier“ (účelových počítačov) do obývačiek sa stiera hranica medzi domácou zábavou a svetom počítačov, čo vydavateľské spoločnosti spočiatku sledovali s nevôľou. Teraz si však začínajú uvedomovať, že dôležitejší je komerčný úspech medzi používateľmi, popularita a jednoduchosť služby ako neustále obviňovanie používateľov z nelegálnych praktík. Poskytovanie často drahých a nedokonalých nahrávok, predsa však pohodlným spôsobom, prináša úspech, ako vidno v extrémne rýchlo rastúcom obrate on-line obchodov s hudbou. V r. 2008 sa v U.S.A. tržby z digitálnej hudby odhadujú na 4,6 miliárd dolárov, čím prekonajú nielen predaj CD nosičov⁹⁷.

⁹⁷ Tamže, odstavec „The future of digital audio is bright“.

4 PREHLAD STRATOVÝCH ZVUKOVÝCH FORMÁTOV

V nasledujúcich častiach ponúkame prehľad rôznych stratových kompresných algoritmov a formátov. Samozrejme, toto je len výber z mnohých – stratových formátov existuje nepreberné množstvo, mnohé z nich sa však nikdy skutočne neujali (napr. TwinVQ), alebo mali od počiatku špecifické použitie s obmedzeným dosahom a do sveta PC sa ani nemohli dostať (napr. PAC), alebo sú utajené a je ťažké získať o nich relevantné informácie (napr. QDesign). Treba však uviesť, že aj pri väčšine popísaných proprietárnych technológií bolo náročné získať čo i len stručnú špecifikáciu.

Kapitola 4 pozostáva z jedenástich častí. Prvá krátka časť (4.1, str. 88) predstaví čitateľovi formát *ADPCM*, ktorý je dnes vo svete PC už dávno prekonaný, no ešte pred desiatimi rokmi bol pre bežných používateľov jedinou dostupnou technológiou na kompresiu zvuku. Rozsiahla druhá časť (4.2, str. 89) sa zaoberá formátmi *MPEG-1/2 Audio*, najmä *MP3* a *MP2* a ich derivátmi *MP3pro*, *MP3 Surround*. Rozsiahla tretia časť (4.3, str. 94) obsahuje informácie o štandarde nahradzujúcom známu „empétrojku“: *AAC* a jej derivát *HE-AAC*. V ďalších častiach si predstavíme proprietárne formáty *Windows Media Audio (WMA)* (4.4, str. 98), *RealAudio (RA)* (4.5, str. 100), *ATRAC* (4.6, str. 102), *Dolby Digital (DD, AC-3)* (4.7, str. 104) a *Coherent Acoustics (CAC, DTS)* (4.8, str. 106). Nasledujú dva opensource formáty: *Ogg Vorbis* (kapitola 4.9, str. 108) a *Musepack (MPC)* (časť 4.10, str. 111).

Posledná, jedenásta časť (str. 112) je všeobecným úvodom ku kompresii hovoreného slova, ako k špecifickej aplikácii stratovej kompresie. Ako príklad sa spomína opensource formát *Speex* (str. 113).

4.1 ADPCM

Adaptívna diferenciálna modulácia pulzným kódom (ADPCM) je asi najstaršou používanou metódou na stratovú kompresiu zvuku. Vychádza z diferenciálnej PCM (DPCM). V prípade PCM kódujeme každú vzorku jej absolútnou hodnotou, v prípade DPCM sa len vhodným kódovaním vyjadří rozdiel hodnoty aktuálnej vzorky oproti hodnote predošlej vzorky¹. Úlohou ADPCM bolo aj za cenu stratovosti zabezpečiť dostatočnú redukciu dátového toku. Reziđuál (rozdiel hodnoty vzoriek) nie je kódovaný presne, ale na základe aktuálneho kvantovacieho kroku, ktorý sa adaptívne upravuje (pri vysokej dynamike signálu sa kvantovací krok zvyšuje – spätne adaptívne kvantovanie).

ADPCM je natívne podporovaná väčšinou operačných systémov. Algoritmus² dnes už neexistujúcej firmy Interactive Multimedia Association (IMA ADPCM) zo začiatku deväťdesiatych rokov sa stal de facto štandardom. Redukoval zvuk s rozlíšením 16 bitov na 4 bity na vzorku (1 bit znamienko, 3 bity diferenciacia). Opodstatnenie bolo najmä v zázname hovoreného slova – štvornásobná redukcia dát bola pri kapacite vtedajších médií dostatočným pokrokom. V dnešných časoch – aspoň vo svete PC – algoritmus ADPCM už opodstatnenie nemá. Dôvodom je nízka subjektívna kvalita výsledného záznamu a nepriaznivý kompresný pomer.

1 Čitateľ si isto všimol, že DPCM a ADPCM vychádzajú z bezstratovej kompresie zvuku metódou lineárnej predikcie prvého stupňa (p. odsek 3.4.2 *Kódovanie lineárnou predikciou*, str. 60).

2 Možno nájsť napríklad na <http://www.cs.columbia.edu/~hgs/audio/dvi>.

4.2 MPEG-1/2 Audio (MP3)

V tejto časti práce sa oboznámime s formátom MP3, ktorý je v súčasnosti najúspešnejším spôsobom psychoakustickej kompresie zvuku. Oboznámime sa tiež s jeho modifikáciami (MP3pro, MP3surround) a spomenieme i formáty MP2 a MP1. (Pre bližšie informácie týkajúce sa štandardov MPEG čitateľa odporúčame na odsek 3.5.10 MPEG, str. 83.)

MP3 je populárna skratka pre tretiu zvukovú „vrstvu“ štandardu **MPEG-1** (po angl. *MPEG-1 layer 3*), **MP2** a **MP1** označujú druhú a prvú vrstvu (nižšie číslo indikuje nižšiu výpočtovú náročnosť). Na požiadavku MPEGu vyvinúť kvalitný stratový zvukový formát pre trh reagovali viaceré zoskupenia návrhmi algoritmov, pričom najúspešnejšie z nich boli MUSICAM a ASPEC³. Vedci združení v konzorciu MPEG z nich vyvinuli jednotlivé vrstvy (MP1, MP2, MP3), štandardizované v rokoch 1990-1992 (rozšírenia sa dočkali so štandardom MPEG-2 v r. 1994). Skratka MP3 (resp. MP2, MP1) označuje algoritmus, formát, v slangu však najmä súbory (.mp3), ktoré sa stali asi najznámejším spôsobom ukladania a šírenia hudby, skratkou známou aj medzi počítačovými laikmi.

4.2.1 MP1, MP2

MP1 je zjednodušením MP2. Bol určený na dátové toky okolo 192 kbps na kanál. Keďže sa s ním v praxi nestretávame, budeme sa mu venovať len okrajovo.

Formát MP2 (prípona .mp2) bol štandardizovaný na základe návrhu firmy Philips – algoritmu menom MUSICAM z r. 1990 (medzi používateľmi PC sa s ním dalo stretnúť od októbra 1993 vďaka kodeku firmy Xing). Jeho úlohou je kódovať zvuky s vysokým dátovým tokom (128 kbps na kanál) poskytujúc nízku výpočtovú náročnosť a vysokú odolnosť voči chybám pri prenose dát. Veľkosť okna je 1024 (MP1) alebo 2048 vzoriek (MP2), okno sa delí na symetrické „granuly“ rovnakej dĺžky. Algoritmus MP1 a MP2 je založený na subpásmovej kódovacej schéme, celý frekvenčný rozsah jednej granuly sa delí na 32 subpásiem. Veľkosť výsledného rámca je 384 (MP1) alebo 1152 vzoriek (MP2), prenášané subpásma má teda dĺžku 12 alebo 36 vzoriek. Na základe jednoduchého psychoakustického modelu sa vypočíta odstup signálu od maskovacieho signálu a určia sa vhodné kvantovacie kroky (rovnomerne, nie logaritmicke pokrývajúce celý rozsah) pre jednotlivé subpásma, ktoré sa neskôr prenášajú (bez frekvenčnej analýzy – kvantovanie teda prebieha v časovej, nie frekvenčnej doméne). Používa stereofonické kódovanie intenzity a rôzne preddefinované tabuľky pre Huffmanovo kódovanie výstupu. Formát MP2 je pre svoju nízku výpočtovú náročnosť a jednoduchú implementovateľnosť používaný

³ Skratky z „Masking pattern adapted Universal Subband Integrated Coding And Multiplexing“ (MUSICAM) a „Adaptive Spectral Perceptual Entropy Coding“ (ASPEC).

v televíznom i rozhlasovom digitálnom terestriálnom vysielaní (DAB, DVB), stretnúť (aj keď zriedka) sa s ním možno na diskoch DVD-Video, používal sa na diskoch VCD a SVCD.

4.2.2 MP3

Formát MP3 (prípona .mp3) bol štandardizovaný v r. 1992 (aj keď hlavný podiel na jeho rozšírení mal až šervérový kodek *l3enc* od nemeckého Fraunhofer Institut für Integrierte Schaltungen, vydaný 7. júla 1994 – od roku 1995 začal úspešne vytláčať MP2). Bol vyvinutý francúzskym zoskupením Thomson a nemeckým Fraunhofer IIS so zámerom poskytnúť kvalitnú kompresiu pri nízkom dátovom toku (najmä na účely digitálneho vysielania) okolo 64 kbps na kanál. Pri jeho návrhu sa čerpano z najlepších vlastností algoritmov MUSICAM a ASPEC. Veľkosť okna je 1152 vzoriek, to sa rozkladá na dve *granuly*, ktoré sa následne delia na 32 subpásiem (dĺžky 18 vzoriek). Subpásma z oboch granúl sa spoja, aby na ne bola aplikovaná MDCT, poskytujúca presnejšie frekvenčné rozlíšenie, bližšie ľudskému vnímaniu. V subpásme možno MDCT aplikovať nielen raz na 36 vzoriek (dlhý blok), ale aj trikrát na 12 vzoriek (krátky blok). Toto menšie frekvenčné a lepšie časové rozlíšenie sa používa pri výskyte zvukov s veľmi krátkym trvaním (bubny ap.). (Pozn.: existuje i tzv. „zmiešaný blok“, keď prvé dve subpásma používajú 36 vzoriek, ostatné 3×12 vzoriek.) Využíva sa všetkých 576 koeficientov z MDCT. Kodéry (ako v prípade takmer všetkých stratových kompresných schém) iteratívne menia koeficienty a nanovo skúmajú odstup šumu od maskovacieho signálu, snažiac sa ho pre daný dátový tok dostať na minimálnu úroveň.

MP3 oproti MP2 dodáva technológiu spojeného kódovania kanálov (joint stereo, M/S). Nízke vzorkovacie frekvencie (16, 22,05 a 24 kHz), dodané štandardom MPEG-2, používajú len jednu granulu.

MP3 našlo rýchlo svoje uplatnenie a doteraz je najrozšírenejším spôsobom psychoakustickej kompresie zvuku.

4.2.3 MP3pro

Fraunhofer IIS a Thomson uviedli rozšírenie formátu MP3 menom MP3pro v r. 2001. Je založené na technológii švédskej firmy Coding Technologies – **replikácii spektrálneho pásma** (angl. *Spectral band replication*, SBR). Vychádza z dovtedajších obmedzení formátu MP3 pri nízkych dátových tokoch (64 kbps a nižšie). Jediný spôsob, ako efektívne kódovať za takýchto podmienok, bolo orezanie frekvenčného spektra – zo vstupu na strane kodéra sa odstránili vysoké frekvencie (napr. nad 8 kHz) a informácie o vyšších zložkách nebolo treba prenášať, čo značne ušetrilo dátový tok. Dôsledkom bol neprirodzený, zastretý zvuk. Pri snahe preniesť aj

vysokofrekvenčné spektrum by však výsledkom bol veľmi nekvalitný záznam, spôsobený šumom z nízkeho rozsahu kvantovacej stupnice.

Návrhári SBR si uvedomili, že väčšina zvuku prítomná vo vyššom spektrálnom pásme sú len alikvotné tóny, ktoré možno odvodiť z nízkych frekvencií prítomných v zázname, alebo majú charakter šumu. Jednoduché „zdvojenie“ spektra samo osebe prinášalo dobré výsledky. Kodér MP3pro funguje na princípe analýzy vysokofrekvenčných zložiek prítomných v zázname a ich následnom odstránení. Štandardný kodér MP3 zakóduje frekvenčne obmedzený záznam. Do toku dát sa následne pridajú „postranné informácie“ (ignorované prehrávačmi bez podpory MP3pro) o postupe „reštaurovania“ vysokých frekvencií v zázname (nie skutočnej rekonštrukcie ich pôvodných charakteristík). Táto technika prináša úžasné výsledky pre dátové toky okolo 64 kbps („postranná informácia“ tvorí len čosi vyše 4 kbps), pre vyššie dátové toky je SBR, samozrejme, nevýhodou. Opodstatnenie má teda pri on-line vysielaní alebo prenosných prehrávačoch, kde je požiadavkou nízky dátový tok, nie vysoká kvalita.

4.2.4 MP3 Surround

Fraunhofer IIS a Thomson vytvorili v spolupráci s firmou Agere Systems a ich technológiou *Binaural Cue Coding* v r. 2004 jednoduché rozšírenie formátu MP3 o priestorový dojem. Špeciálny kodér vytvára downmix 5.1 zvuku na obyčajný stereo záznam (maticovo kódovaný, podobne ako Dolby Pro Logic), pričom dbá, aby v dôsledku stratovej kompresie nezaničili informácie o dôležitých rozdieloch vo fáze, oneskorení a intenzite frekvenčných zložiek – pre poslucháča natoľko potrebných indícií na dosiahnutie priestorového dojmu zo zvuku⁴. Údaje potrebné na ich rekonštrukciu sa prenášajú ako prídavok k zvyčajnému toku dát – ako „postranné informácie“, ignorované prehrávačmi bez podpory MP3 Surround. Tie potom prehrávajú súbory kódované algoritmom MP3 Surround ako obyčajné stereofonické súbory. Postranný dátový tok tvorí zvyčajne 16 kbps⁵, veľkosťou sú teda nové súbory porovnateľné s pôvodným formátom MP3.

Podporu novému formátu vyslovilo DivXNetworks, tvorca známeho formátu DivX. Faktom však ostáva, že MP3 Surround nie je veľmi rozšírený – priestorový zvukový záznam sa viaže skôr s filmovými záznamami, kde sa používajú iné druhy kódovania (na ktoré si používatelia už zvykli). Prípadná úspora dátového priestoru, ktorú nový formát ponúka, nie je pri súčasných veľkostiach médií dostatočným dôvodom na zmenu. Bližšie informácie možno nájsť napríklad v dokumente http://www.iis.fraunhofer.de/amm/download/flyer/dl.html?f=mp3_surround.pdf.

4 Súčasťou štandardu MPEG-2 je aj postup pri maticovom kódovaní priestorového zvuku, tento sa však pre nízku kvalitu (chýbajú práve spomínané „indície“) a malú podporu na trhu nikdy neujal.

5 Zdroj: <http://www.mp3licensing.com/mp3/mp3surround.html>.

4.2.5 Zhrnutie

Medzi výhody formátu MP3 patrí najmä jeho definícia – štandard zahrnuje popis formátu (toku dát) a dekodéra, implementácia kodéra je však ponechaná na tvorcovi. To spôsobilo, že miesto jedného „štandardného“ kodeku si používateľ môže vybrať z nepreberného množstva rôznych implementácií. Za najkvalitnejší sa vo všeobecnosti považuje algoritmus **LAME**⁶, ktorý je navyše bezplatný. Výhoda sa však môže zmeniť na nevýhodu – mnohí komprimujú zvuk do formátu MP3 používajúc nekvalitné algoritmy, spôsobujúc, že dátový tok, ktorý je s LAME pre danú oblasť postačujúci, môže náhle poskytovať úplne neuspokojivé výsledky.

Medzi nevýhody formátov MP1, MP2 a MP3 (podľa štandardu MPEG-1) patrí obmedzenie dátového toku – pre každý rámec musí tento byť z presne stanovenej množiny hodnôt 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 alebo 320 kbps. MPEG-2 dodáva hodnoty 8, 16, 24, 32, 48 a 144 kbps. Aby sa čo najlepšie využil dátový tok, možno v prípade, že ešte v aktuálnom rámci ostane priestor, uložiť doň dáta z nasledujúceho rámca (najviac 511 bajtov) – táto črta sa nazýva **bitová rezerva** (angl. *bit reservoir*). Variabilný dátový tok, ktorý podporujú mnohé kodéry (a našťastie väčšina MP1/MP2 dekodérov, ba všetky MP3 dekodéry) nie je ničím iným ako vhodným výberom dátového toku pre každý rámec záznamu osobitne, podľa potreby (napr. stanovená hodnota NMR). Obmedzený dátový tok (najmä horná hranica 320 kbps) však neumožňuje úplne flexibilne reagovať na rôznorodé vstupné signály.

Ďalšou nevýhodou je nezabezpečenie podpory pre gapless playback – časové oneskorenie (teda veľkosť medzier pridaných do záznamu) sa do výstupného toku dát nezaznamenáva. Existuje však neoficiálna podpora – LAME tieto informácie dodáva do prídavných značiek súborov .mp3 a závisí od prehrávača, či ich podporuje.

Do tretice treba spomenúť zaťaženosť formátu patentmi. Za licencovanie je zodpovedná firma Thomson⁷, ktorá už viackrát hrozila výrobcom a programátorom komerčných MP3 kodekov v USA žalobami pre neodvážanie poplatkov. Nie je však známy prípad žaloby vývojárov opensource kodekov. Navyše, patentové právo v EÚ, ako je známe, v súčasnosti nepodporuje vymáhanie patentu z programovacích postupov, čo umožňuje používať programy ako LAME bez problémov. Pôvodné patenty od Fraunhofer IIS vypršia v apríli roku 2010, čím sa MP3 stane verejným vlastníctvom.

V prospech MP3 oproti MP2 hovorí najmä vyššia kvalita pri nízkych dátových tokoch. Pre vysoké dátové toky (128 kbps na kanál a viac) však MP2 poskytuje vyššiu kvalitu než MP3 (vďaka neprítomnosti stratovej frekvenčnej dekompozície), čo v prípade DAB a DVB, spolu s menšou výpočtovou náročnosťou a vyššou

⁶ Aspoň pri dátových tokoch 128 kbps a vyššie, p. <http://www.rjamorim.com/test/mp3-128/results.html>. Stránku projektu LAME možno nájsť na <http://lame.sourceforge.net>.

⁷ <http://www.mp3licensing.com>.

odolnosťou voči chybám pri prenose dát, rozhodlo v prospech tohto staršieho formátu. (Navyše, zakomponovanie technológie SBR umožňuje MP2 konkurovať s MP3 aj pri nižších dátových tokoch.)

V skratke možno povedať, že formát MP3 je a ostáva (napriek svojej zastaranosti) najpopulárnejší formát. Veľké množstvo hudby archivovanej ako MP3 na strane používateľov (často z čias, keď bol ešte najlepšou alternatívou), ohromná podpora na trhu (nielen softvérová, ale i prenosné prehrávače, veže, DVD prehrávače, ...) a dobré meno (ktoré sa stalo súčasťou slovníka väčšiny mladých ľudí, bez ohľadu na ich vzťah k výpočtovej technike) z neho robia prakticky univerzálne médium na dátovo nenáročný prenos a archiváciu hudby. Chýbajúca podpora DRM schém imponuje jednoduchým používateľom, rozumná výpočtová náročnosť pomáha tvorcom hardvérových dekodérov. Aj keď je AAC už mnohé roky vybraný ako nástupca MP3, v rozšírenosti ho tak skoro neprekoná – priemerný používateľ si zvykli na osvedčený formát a nepotrebuje ho meniť za nový, ktorý len mierne zmenší veľkosť už i tak malých súborov .mp3.

Oficiálna stránka Fraunhofer IIS venovaná technológii MP3 sa nachádza na <http://www.iis.fraunhofer.de/amm/techinf/layer3>, ďalšie podklady možno nájsť aj na **stránkach** <http://www.mp3licensing.com>, <http://www.digital-audio.net/res/docs/pdf/mpegaud.pdf> a iných.

4.3 MPEG-2/4 Audio (AAC)

V tejto časti práce sa oboznámime s formátom AAC (angl. *Advanced Audio Coding*) a jeho modifikáciami. AAC v súčasnosti ponúka asi najefektívnejšie kódovanie zvuku od silno stratového kódovania (počnúc dátovým tokom 2 kbps) po úplne priesvitný záznam pri dátových tokoch nižších než MP3. (Pre bližšie informácie týkajúce sa štandardov MPEG čitateľa odporúčame na odsek 3.5.10 MPEG, str. 83.)

AAC bol navrhnutý ako nástupca MP3 v štandardoch MPEG-2 a MPEG-4, ako štandard bol schválený v apríli r. 1997. Hlavný podiel na jeho vývoji mala firma Dolby. Ponúka široké spektrum nastavení – vzorkovacia frekvencia od 8 kHz do 96 kHz, počet kanálov 1 až 48 (celý frekvenčný rozsah), navyše do 15 LFE kanálov (frekvenčný rozsah do 120 Hz) a 15 nezávislých dátových kanálov. Bitová hĺbka vzorky 8, 16, 24 bitov, dátový tok na kanál do 256 kbps.

Na okná veľkosti 2048 vzoriek (prekrývanie okien 50%, teda veľkosť rámca je polovičná) sa priamo aplikuje MDCT, v prípade náročných pasáží možno veľkosť okna zmenšiť na 256 vzoriek. Formát AAC je navrhnutý modulárne – používateľ si môže vybrať z viacerých druhov⁸:

- **Profil s nízkou zložitou** (angl. *Low Complexity Profile, LC-AAC*). Najjednoduchší, zároveň najviac podporovaný a používaný, keďže ide o pôvodný formát definovaný v MPEG-2. Pri dátových tokoch pod 80 kbps je na všeobecné záznamy (hudba, spev) vhodnejší HE-AAC, v prípade hovoreného slova ostáva lepším LC-AAC (aj pri poklese dátového toku k 8-12 kbps pre mono).
- **Hlavný profil (MAIN)** dodáva k LC-AAC spätne adaptívnu predikciu spektrálnych zložiek. Funguje ako predikcia v prípade bezstratovej kompresie zvuku, prebieha však vo frekvenčnej (nie časovej doméne). Každá spektrálna zložka je predikovaná prediktorom druhého rádu. Spektrálne zmeny oproti predikovaným sú kódované ako reziduálny signál. Tento profil je výpočtovo omnoho náročnejší. Cieleny bol najmä na nízke dátové toky, kde lepšie výsledky tak či onak ponúka HE-AAC, preto nezískal veľkú podporu.
- **Profil so škálovateľnou vzorkovacou frekvenciou** (angl. *Scalable Sample Rate, AAC-SSR*⁹). Jeho autorom je firma Sony. Pred aplikovaním MDCT sa záznam v okne rozdelí na štyri subpásma použitím banky polyfázových kvadraturných filtrov (MDCT sa teda aplikuje na 512 resp. 64 vzoriek). Prvou výhodou je adaptovateľnosť jednotlivých subpásiem na zložitnosť signálu (napr. zvýšenie frekvenčného rozlíšenia pre menšie frekvencie, zvýšenie časového rozlíšenia pre vyššie frekvencie – v súlade s citlivosťou ľudského sluchu). Druhou, hlavnou výhodou je možnosť jednoduchého zmenšovania kvality (presnejšie frekvenčného

⁸ V MPEG-2 označovaných *profil*, v MPEG-4 *zas typ objektu*. Z pohľadu MPEG-4 sú unikátne len prvé štyri, ostatné sú len špeciálnymi prípadmi predošlých, resp. nachádzajú sa v osobitných špecifikáciách.

⁹ Občas tiež *Sample Rate Scalable, AAC-SRS*.

rozsahu) záznamu postupným odstraňovaním subpásiami od najvyšších frekvencií k nižším. Nevýhodou je ďalšie skreslenie, ktoré sa takýmto spracovaním vnáša do signálu (najmä v okolí deliacich frekvencií medzi subpásiami). V praxi sa nepoužíva.

- **Profil s dlhodobou predikciou** (angl. *Long Term Prediction, AAC-LTP*) k profilu LC dodáva jednoduchú dopredne adaptívnu predikciu v časovej doméne, tzv. dlhodobú predikciu¹⁰. Koeficienty prediktora sa posielajú ako postranná informácia. V praxi sa nepoužíva.
- **Kodér s krátkym oneskorením (AAC-LD, angl. *Low Delay*)**. Stavia na existujúcom AAC-LC, no obmedzuje latenciu (oneskorenie medzi vstupom a výstupom kodéra) na 20 ms (pre porovnanie, štandardne sa oneskorenie pohybuje okolo 100 ms).
- **HE-AAC (AACplus)** – popísaný v nasledujúcom odseku.
- **AAC s aritmetickým kódovaním s presnosťou na bit** (angl. *Bit Sliced Arithmetic Coding AAC, BSAC-AAC*) je založené na použití upraveného aritmetického kódu pri kódovaní čísel vo výstupnom toku dát. Umožňuje stupňovať dátový tok po jednotkách 1 kbps, nie 8 kbps, ako je to v prípade štandardného AAC, čím zlepšuje podporu škálovateľnému kódovaniu zvuku¹¹.
- **TwinVQ a CELP**, určené na kompresiu hovoreného slova.
- Navyše dodáva **parametrické kódovanie** určené pre extrémne nízke dátové toky.

Formát AAC sa zväčša ukladá do kontajnera definovaného ako súčasť MPEG-4 (part 14), ktorý je len vylepšeným kontajnerom .mov QuickTime od Apple¹². Štandardná prípona by mala byť .mp4, používatelia sa však skôr stretnú s príponami používanými službou iTunes – .m4a (súbory s prídavnými informáciami ako audio knihy, albumy s hyperlinkami atď.) a .m4p (klasické audio súbory, chránené DRM schémou). AAC zahŕňa niekoľko stupňov ochrany pred chybami v prenosovom kanáli, počnúc úpravou kódových tabuliek, cez algoritmy na detekciu chyby, končiac samoopravnými kódmi¹³. Formát AAC uložený podľa špecifikácie MPEG-2 (s tým sa možno stretnúť už len zriedka¹⁴) používa príponu .aac a nepodporuje DRM schémy. Používa vlastnú identifikačnú značku, ktorej špecifikácia však nie je verejne dostupná¹⁵.

Ďalšou technológiou použitou v AAC je **percepčná náhrada šumu** (angl. *Perceptual Noise Substitution, PNS*), ktorej úlohou je detegovať v signáli šumové

10 P. časť 4.11 *Kompresia hovoreného slova*, str. 112.

11 Princíp škálovateľného kódovania p. v odseku 3.4.3 *Hybridné a škálovateľné kódovanie*, str. 62.

12 Preto je jeho štruktúra takmer zhodná s týmto formátom, ktorý je lepšie popísaný napr. na stránkach <http://developer.apple.com/documentation/QuickTime/RM/PDF.htm>.

13 Na stránke <http://www.iis.fraunhofer.de/amm/techinf/mpeg4/error.html> možno nájsť prehľad použitých technológií.

14 Šírenie zvuku vo formáte AAC v kontajneri definovanom v MPEG-2 je však hojne využívané v prípade internetových rádii, keďže tento spôsob je pre nich bezplatný. Komerčné používanie kontajneru MPEG-4 so sebou prináša platenie licenčných poplatkov.

15 „AAC uses yet another tagging format, which does not at present have a published spec as of 3/1/2006.“ (Zdroj: <http://www.id3.org/faq.html>.)

zložky a tie nekódovať štandardnými technikami, ale parametricky popísať. Dekodér dostane informácie o frekvenčnom pásme, ktoré šum zaberá, a jeho energii. Výsledkom je teda nedeterministické dekódovanie signálu (generovaný šum má charakter pseudonáhodných čísel). AAC v sebe zahŕňa aj technológiu **tvarovania šumu v časovej doméne** (angl. *Temporal Noise Shaping, TNS*). Ide o spätnu adaptívnu predikciu¹⁶ spektrálnej charakteristiky záznamu v časovom priebehu, ktorá sa prispôsobuje signálu – v prípade bohatého frekvenčného spektra reaguje znížením rozlíšenia vo frekvenčnej doméne, stáva sa však citlivejšou v časovej doméne, redukuje tak problém predozveny. Z TNS údajne ťažia najmä záznamy hovoreného slova – ich kvalita sa má drasticky zvýšiť¹⁷. Presnejší popis možno nájsť napr. v HERRE, Jürgen. *Temporal Noise Shaping ... 1999*.

O spopularizovanie AAC sa postarala najmä firma Apple, ktorá ho od r. 2003 natívne podporuje vo svojich produktoch a službách (iTunes, iPod). Dôvodom je najmä možnosť integrovania DRM schém (FairPlay a Hymn¹⁸). Licencovaním je poverená dcérska spoločnosť konzorcia Dolby – Via Licensing (<http://www.vialicensing.com>).

4.3.1 HE-AAC (AACplus)

Formát s názvom **Vysokoefektívne AAC** (angl. *High Efficiency AAC, HE-AAC*) nazývaný tiež **AACplus**, je jedným z rozšírení pôvodného formátu štandardom MPEG-4. Je určený na nízke dátové toky. K AAC dodáva technológiu SBR (p. odsek 4.2.3 MP3pro, str. 90). Ak sa navyše použije technológia **PS (parametrické stereo**, tiež z dielne Coding Technologies), ide o HE-AAC v2 (AACplus v2). Princíp PS spočíva v hĺbkovej analýze stereofonickej informácie v zázname, pričom rozdiely medzi kanálmi (nielen frekvenčné a amplitúdové, ale podľa potreby už aj fázové) sa následne efektívne popíšu v „postranných informáciách“ pôvodného toku dát (s nízkym dátovým tokom). Pôvodné dva signály sú sčítané a komprimované ako mono záznam. Vďaka tejto technológii podáva HE-AAC pri dátovom toku 48-64 kbps v porovnaní s konkurenciou vynikajúci výkon, rovnako exceluje pri 5.1 zvuku a dátovom toku 160 kbps¹⁹.

4.3.2 Zhrnutie

Medzi výhody formátu AAC oproti MP3 patrí lepšia podpora pre zvuky s vysokým rozlíšením, viackanálový zvuk, presnejšia reprodukcia (priama analýza MDCT bez stratovej dekompozície na subpásma), vyššia kvalita pri rovnakom

16 Porov. termín *dopredne adaptívna predikcia* v odseku 3.4.2 *Kódovanie lineárnou predikciou*, str. 60.

17 Zdroj: <http://www.iis.fraunhofer.de/amm/techinf/aac/index.html>.

18 Obe nakoniec boli prelomené – zásluhu na tom mal najmä Jon Johansen („DVD Jon“).

19 Porov. MELTZER, Stefan – MOSER, Gerald. *MPEG-4 HE-AAC v2 ...*

dátovom toku, lepšia ochrana pred chybami, flexibilnejšie využívanie stereofonickej redundancie (osobitné techniky pre rôzne časti frekvenčného spektra), podpora pre gapless playback (ak je formát uložený v kontajneri MPEG-4) a mnoho iného. Existuje niekoľko rozšírených kodérov, za najlepší sa považuje ten od firmy Apple (iTunes AAC), okrem neho ponúka svoje algoritmy aj napr. Nero (Nero AAC), medzi bezplatnými opensource projektmi dominuje FAAC.

AAC je však pre veľa ľudí celkom neznámy formát – napriek jeho veku, hovorí sa o ňom málo. Zatiaľ. Jeho podpora stále rastie a v súčasnosti ho možno prehrať na veľkom množstve multimedialných mobilných telefónov, niektorých hracích konzolách atď.

Rôzne referencie a technické detaily možno nájsť napr. na stránkach <http://www.audiocoding.com/modules/wiki/?page=AAC> a http://www.chiariglione.org/MPEG/tutorials/papers/icj-mpeg4-si/09-natural_audio_paper/gacoding.html.

4.4 Windows Media Audio (WMA)

Windows Media Audio (WMA) je proprietárnym formátom, ktorého tvorcom je spoločnosť Microsoft. Vznikol ako reakcia na licencovanie formátu MP3.

Pôvodný formát je určený na kompresiu stereofonického záznamu pri vzorkovacej frekvencii do 44,1 kHz a 48 kHz, 16-bitovej hĺbke vzorky a dátových tokoch od 64 kbps do 192 kbps. (Získať presnú špecifikáciu možno za licenčných podmienok uvedených na stránke <http://www.microsoft.com/windows/windowsmedia/licensing/default.aspx>.)

S príchodom série WMA 9 pribudli k pôvodnému WMA ďalšie formáty:

- **Windows Media Audio 9 Professional (WMA Pro)** je určený, ako názov napovedá, na náročnejšie úlohy. Preto vznikol ako nový návrh, nekompatibilný s pôvodným formátom WMA. Ako AAC i WMA Pro podporuje široký segment trhu: Pre náročných používateľov poskytuje pri dátových tokoch až do 1 536 kbps podporu 24-bitovej hĺbky vzorky, vzorkovaciu frekvenciu 96 kHz a najviac 8 kanálov. V prípade dátových tokov 96-128 kbps má dosahovať lepšie výsledky ako pôvodný formát WMA (nazývané tiež **WMA Standard**, **WMA Std**). (Podpora nízkych dátových tokov je spomenutá v predposlednom odstavci tohto odseku.) Obsahuje **kontrolu dynamického rozsahu** (angl. *dynamic range control*) – počas kódovania sa zaznamená priemerná a špičková úroveň amplitúdy záznamu. Počas následnej reprodukcie si používateľ môže vybrať, či chce počuť verne celý dynamický rozsah, alebo ho skresliť kompresiou akustickej hladiny (vtedy ide o „tichý posluš“ – angl. *quiet mode*) úpravou špičiek na 12 dB alebo 6 dB nad priemernou akustickou hladinou záznamu. WMA Standard a WMA Pro podporujú dvojprechodové CBR kódovanie a tri druhy VBR kódovania – s dôrazom na konštantnú kvalitu (jeden prechod), priemerný dátový tok (dva prechody) a špičkový dátový tok (dva prechody).
- **Windows Media Audio 9 Lossless (WMA Lossless)** je bezstratový kompresný formát. Tiež obsahuje podporu pre kontrolu dynamického rozsahu.
- **Windows Media Audio 9 Voice (WMA Voice)** obsahuje algoritmus prispôbený kompresii hovoreného slova, určený na dátové toky 20 kbps a menej (v extrémnych prípadoch 4 kbps pre frekvenčný rozsah do 8 kHz). Napriek špecializácii na ľudský hlas si vie poradiť aj s hudbou – automaticky sa vie prispôbiť zdroju, prepínaním medzi WMA Voice a štandardným WMA algoritmom.

Tabuľku s kompatibilnými prehrávačmi a operačnými systémami v závislosti od použitého formátu možno nájsť v tabuľke na stránke <http://www.microsoft.com/windows/windowsmedia/forpros/codecs/audio.aspx>.

Formát WMA sa zväčša ukladá do kontajneru typu Advanced Systems Format (ASF, .asf), často s osobitou príponou .wma. ASF podporuje rozličné DRM schémy²⁰ a vysielateľnosť (streaming). Oficiálny prehrávač pre WMA je Windows Media Player, DirectShow filter²¹ však možno získať bezplatne zo stránok MS. Neoficiálny popis staršieho, no rozšírenejšieho štandardu ASF 1.0 možno nájsť na stránke <http://avifile.sourceforge.net/asf-1.0.htm>, oficiálny popis novšieho (2004), menej používaného štandardu možno získať na stránkach MS: <http://www.microsoft.com/windows/windowsmedia/forpros/format/asfspec.aspx>. WMA používa proprietárne identifikačné značky.

Prvá verzia (WMA 7) sa objavila v decembri r. 2000, súčasná verzia WMA nesie označenie 9.1. Čoskoro sa dočkáme nových vylepšení – črtá sa obohatenie WMA Pro o kompresiu pri nízkych dátových tokoch (24-96 kbbps pre stereo, 128 kbps-256 kbps pre 5.1 zvuk), určené pre bezdrôtové aplikácie a vysielanie cez internet. Podľa predbežných testov (z decembra 2005) bude mať nová technológia oproti najväčšiemu konkurentovi – HE-AAC v2 – pri dátovom toku 64 kbps navrch. Viac informácií možno nájsť na stránke <http://www.microsoft.com/windows/windowsmedia/forpros/codecs/comparison.aspx>.

WMA má vďaka silnému zázemiu (Microsoft) širokú podporu na trhu prehrávačov a elektronických zariadení. V súčasnosti je dokonca po MP3 druhým najrozšírenejším formátom – dôvodom je natívna podpora v najrozšírenejších operačných systémoch od MS. Rozličné referencie možno nájsť na stránke MS: <http://www.microsoft.com/windows/windowsmedia/forpros/AudioProd.aspx>.

20 Popis najnovšej verzie (2.0) možno nájsť na stránke [cit. 2006-04-12] <http://www.spinnaker.com/crypt/drm/freeme/Technical>.

21 Knižnica určená na prehrávanie daného formátu v prostredí OS Windows.

4.5 RealAudio (RA)

RealAudio (RA, .ra) je (spolu s *RealMedia*, RM, .rm) proprietárnym kontajnerom, ktorého tvorcom je spoločnosť RealNetworks. Vznikal v čase všeobecného úspechu internetu, najmä aby poskytol možnosť vysielania²² zvuku (neskôr s RealVideo i obrazu) dostatočnej kvality pri vtedajších nízkych prenosových kapacitách internetových liniek a prípojok.

RA v sebe skrýva podporu pre rozličné formáty na kompresiu zvuku. Začínal s podporou plejády algoritmov určených na kompresiu hovoreného slova (VSELP v RA 1, surový dátový tok 8 kbps; medzi ďalšie patrí LD-CELP v RA 2 a ACELP.net v RA 4), neskôr používal najmä formáty z cudzej produkcie (AC-3 v RA 3, ATRAC3 v RA 8, LC-AAC a HE-AAC v RA 10). Objavili sa však aj dva formáty z vlastnej dielne – Cookov kodek (v RA 6) a bezstratová kompresia RealAudio Lossless (RA 10). V súčasnosti podporuje dátové toky siahajúce od 12 kbps po 800 kbps. Vhodný spôsob kompresie volí kodér (zväčša RealProducer, určený najmä spoločnostiam zaoberajúcim sa vysielaním a archivácii videa) podľa zadaného dátového toku a podmienok. Vďaka AAC je RA vhodný už nielen na dosiahnutie nízkych dátových tokov ako kedysi – súbežne ponúka možnosť dosiahnuť vysokú kvalitu pri rozumných dátových tokoch 128 kbps a viac, navyše získava podporu pre viackanálový zvuk.

Formát RA používa dve prípony. Staršou z nich je prípona *.ra*. Ide o surový tok dát začínajúci hlavičkou, bez podpory pretáčania (ktoré je možné len vďaka tomu, že použitý kodek musí byť CBR). Neskôr (od verzie 6) RA používa kontajner RealMedia (RM, .rm), podobný kontajneru RIFF. Na prelome rokov 2003/2004 sa RealNetworks snažil do RM zapracovať vlastnú DRM schému (Helix DRM), ktorá sa však nedočkala veľkého rozšírenia. Využíva však DRM schému pre formát AAC. Neoficiálny popis štruktúry súborov možno nájsť na stránke <http://wiki.multimedia.cx/index.php?title=RealMedia>.

Oficiálny prehrávač z dielne RealNetworks – RealPlayer, momentálne vo verzii 10.5 – je k dispozícii pre rozličné platformy. Alternatívou v prostredí Windows môže byť bezplatný DirectShow filter RealAlternative (http://www.free-codecs.com/download/Real_Alternative.htm). Napriek proprietárnosti RA firma RealNetworks iniciovala vznik projektu Helix, otvoreného štandardu na šírenie a prehrávanie multimédií, ktorý získal širokú podporu napr. medzi výrobcami mobilných telefónov – p. stránku <https://helixcommunity.org>.

Prvá verzia RA sa objavila v r. 1995, v súčasnosti je k dispozícii verzia RA 10. Keďže je na trhu od adventu šírenia hudby po internete, ľudia ho poznajú a naďalej sa hojne používa. Ostáva však vo sfére vysielania zvuku, kde mu veľmi zdatne konkurujú MP3 a Vorbis, vo sfére vysielania obrazu ho už takmer úplne nahrádza

²² Pri vysielaní (streaming) sa v súčasnosti RA opiera o proprietárne protokoly RTSP a RDT.

Windows Media Video. V boji na dlhé trate, zdá sa, RA prehráva, technologicky zaostáva a nemá čo ponúknuť. Domovskú stránku RA možno nájsť na <http://www.realnetworks.com/products/codecs/realaudio.html>.

4.6 ATRAC

Adaptive Transform Acoustic Coding (teda adaptívne transformačné akustické kódovanie), v skratke **ATRAC**, je proprietárnym formátom vyvinutým firmou Sony pre potreby zvukových médií formátu MiniDisc.

Prvotná verzia (**ATRAC1**) bola uvedená spolu s MiniDiscom v r. 1992. Používa dva kvadrátne zrkadlové filtre, ktorými delí signál vzorkovaný frekvenciou f na tri subpásma s deliacimi frekvenciami $f/8$ (ak $f=44\,100$ Hz, obsahuje prvé subpásmo frekvencie do 5 512,5 Hz) a $f/4$ (druhé subpásmo obsahuje frekvencie od 5 512,5 Hz do 11 025 Hz; tretie subpásmo frekvencie od 11 025 Hz do 22 050 Hz, teda $f/2$ – Nyquistova frekvencia), na ktoré neskôr aplikuje MDCT. Veľkosť okna je 512 vzoriek, v prípade potreby ho možno (nezávisle pre každé subpásmo) rozdeliť na bloky veľkosti 128, resp. 64 vzoriek (podľa toho, či ide o prvé dve subpásma, alebo o tretie subpásmo). Štandardný dátový tok používaný pre stereofonické záznamy je 292 kbps, zaručujúc tak priesvitnosť reprodukovaného záznamu vo väčšine prípadov.

V r. 1999 Sony uviedol vylepšenú verziu s označením **ATRAC3**. Oproti ATRACu pridáva navyše ďalší kvadrátne zrkadlový filter, ktorým delí prvý interval na polovicu ($f/16$, zvyčajne 2 756,25 Hz). V móde LP2 znižuje dátový tok na 132 kbps, kvalitou však nie je pre ostatné formáty konkurenciou²³. Mód LP4 znižuje dátový tok až na 66 kbps, pričom okrem aplikovania spojeného kódovania kanálov sa komplexnosť signálu redukuje i orezaním frekvencií nad 13,5 kHz. Používatelia sa zároveň dočkali technológie **bitmusic**, umožňujúcej zakúpiť si a stiahnuť nahrávky vo formáte ATRAC. Išlo teda o jednu z prvých legálnych služieb na šírenie hudby prostredníctvom internetu.

V r. 2002 uzrela svetlo sveta nová verzia formátu ATRAC – **ATRAC3plus**. Je nekompatibilná s predošlými verziami. Používa okno veľkosti až 4096 vzoriek, ktoré rozdelí použitím polyfázových kvadrátne filtrov na 16 subpásiem, na ktoré jednotlivo aplikuje MDCT.

V r. 2006 pribudol formát **ATRAC Advanced Lossless**, teda bezstratový kompresný formát. Obsahuje v sebe aj záznam vo formátoch ATRAC3 alebo ATRAC3plus, ktorý z neho možno jednoducho extrahovať, ide teda o hybridný kódér²⁴.

Formát ATRAC používa kontajner s príponou .omg (Open MagicGate – MagicGate je označenie proprietárneho protokolu na bezpečné ukladanie mediálneho obsahu použitím šifrovania a DRM schém; uplatnenie našiel aj vo flashových kartách SD – Secure Digital), občas .oma. Jediný dostupný kódér pre formát ATRAC je produkt firmy Sony menom SonicStage, momentálne (2006-04-18) vo verzii 3.4.

23 Aspoň v prípade dátového toku okolo 128 kbps – p. napr. posluchové testy na stránke <http://www.rjamorim.com/test/multiformat128/results.html>.

24 P. odsek 3.4.3 *Hybridné a škálovateľné kódovanie*, str. 62.

Umožňuje prenos zvuku vo formáte ATRAC, PCM alebo MP3²⁵ do prenosných prehrávačov firmy Sony. Z formátu ATRAC možno skladby previesť do iného formátu (napr. PCM v súbore .wav) len v prípade, že sa používateľ preukáže pôvodinou (napr. identický zvuk na Audio CD), alebo nebola použitá DRM schéma (v prípade nahrávania zvuku cez mikrofón alebo vstup prenosných zariadení). Prehrávanie formátu umožňuje len Winamp (použitím zásuvného modulu firmy Liquid Audio) a RealPlayer (pozná ATRAC1 a ATRAC3, lebo ich sám používa vo svojej verzii RA 8). ATRAC podporuje dátové toky z množiny 48, 64, 96, 132, 160, 192, 256, 320 a 352 kbps.

Formát ATRAC už dávno prehráva s konkurenciou – spolupracujú s ním len prehrávače firmy Sony, jediným kódrom je produkt SonicStage, ktorý je navyše údajne pomalý a nestabilný (p. napr. diskusiu na <http://forum.dbpoweramp.com/printthread.php?t=2377>). Dôvodom je nedostupnosť (utajenosť) jeho podrobnej špecifikácie. Technologicky zaostáva za novými formátmi²⁶, ako dokazujú posluchové testy. Domovská stránka formátu je <http://www.sony.net/Products/ATRAC3>. Ak ho chcete používať, začnite od stránky <http://www.connect.com> – potrebujete však prehliadač Internet Explorer²⁷. Jednoduchú špecifikáciu možno nájsť na stránke http://www.minidisc.org/aes_atrac.html.

25 Teda zo súborov .ogg, .wav, .mp3.

26 Aj keď ATRAC3plus je pri nízkych dátových tokoch (64 kbps a 48 kbps) porovnateľný s WMA Standard, ako ukazujú testy na objednávku Sony: http://www.sony.net/Products/ATRAC3/tech/ITS_test_report.pdf a http://www.sony.net/Products/ATRAC3/tech/TESTfactory_Listening_test.pdf.

27 V prípade zadania presnej adresy, napr. http://musicstore.connect.com/mb_us/IESignup.flow, možno použiť ľubovoľný prehliadač.

4.7 Dolby Digital (DD, AC-3)

Dolby Digital (DD) je spoločný (marketingový) názov pre množinu zvukových formátov z dielne Dolby Laboratories. Sú založené na technológii s označením **Adaptívny transformačný kodér** (angl. *Adaptive Transform Coder 3*), preto je známejší skôr pod menom **AC-3** (oba pojmy budeme považovať za ekvivalentné). Tento formát bol vyvíjaný s cieľom ponúknuť komerčne dostupnú digitálnu reprezentáciu priestorového zvuku²⁸.

AC-3 je určený pre 1 až 5.1 kanálov (novšia verzia DD+, resp. AC3plus zvyšuje toto číslo na 13.1), rozlíšenie vzoriek do 24 bitov, vzorkovaciu frekvenciu do 48 kHz a dátové toky od 32 do 640 kbps²⁹ (DD+ do 6 144 kbps). Veľkosť okna pre MDCT je 512 vzoriek (v závislosti od vstupu možno okno rozdeliť na dve časti po 256 vzoriek). Zvláštnosťou je veľkosť rámca 1536 vzoriek, ktorý vzniká spojením šiestich okien³⁰ – rámec sa teda delí na 6 blokov po 256 vzoriek. Koefficienty z MDCT sa rozdelia na 50 skupín, akýchsi subpásiem veľkosti pribl. šestiny oktávy (čo lepšie odzrkadľuje vlastnosti sluchu než uniformné delenie kvadrátovými filtrami). Koefficienty sa následne kódujú ako reálne čísla zadaním exponentu (len pre koeficient vyjadrujúci najnižšiu frekvenciu – pre vyššie frekvencie sú exponenty zadané ako diferencia oproti predošlej hodnote) a mantisy. Koefficienty prvého bloku v rámci sa kódujú priamo, ostatné môžu byť kódované diferenciou oproti ľubovoľnému z predošlých blokov.

Pre zvolené frekvencie možno použiť združené kódovanie kanálov. Súčasťou kódovania je aj kompresia akustickej hladiny, aplikovaná za účelom zmenšenia dynamiky signálu, a výpočet informácie o hlasitosti dialógov (hovorenej reči) v zázname, umožňujúce normalizáciu akustickej hladiny³¹.

Bezstratová kompresia **Dolby TrueHD** je rozšírením formátu MLP (p. časť 5.10 *Meridian Lossless Packing (MLP)*, str. 131), zvyšujúca počet kanálov na 8 a dátový tok na 18 Mbps.

DD, uvedený v r. 1992, je v súčasnosti hojne používaným formátom. Najväčšie uplatnenie získal pri kompresii priestorového zvuku, najmä na nosičoch DVD-Video. Očakáva sa aj jeho všeobecná rozšírenosť na médiách budúcej generácie – HD-DVD a BD. Pri akceptovateľnom³² dátovom toku (zväčša 192 kbps pre stereo signál,

28 Stručný úvod do histórie a uplatnenia AC-3 možno nájsť v odseku 3.5.2 *Priestorový zvuk a zvukové kanály*, str. 66.

29 Ide o CBR. Presnú vzorkovaciu frekvenciu a dátový tok možno vybrať z obmedzenej množiny hodnôt.

30 Netreba zabúdať na 50% prekrývanie okien, p. odsek 3.3.2 *Fourierova transformácia, kosínusová transformácia*, str. 51.

31 Teda zabezpečujúce to, že priemerná hlasitosť (hladina akustického tlaku) každého prehrávaného záznamu bude rovnaká, odvolávajúc sa na hlasitosť dialógov v stredovom reproduktore (podľa normy medzi -25 a -31 dB oproti úplnému rozsahu).

32 Vzhľadom na určenie – veľkokapacitné médiá (kde aj tak väčšinu obsahu zaberá komprimovaný obraz).

384 kbps alebo 448 kbps pre 5.1) a veľkom množstve nastavení (rôzne postranné informácie, určené pre dekodér) poskytuje vysokú kvalitu a na trhu spotrebnej elektroniky patrí medzi najviac podporovaný formát. Práve všeobecná akceptovanosť a otvorenosť štandardu spôsobuje, že formát DD bude ešte dlho používaný. Presnú technickú dokumentáciu možno nájsť v ATSC. Digital Audio Compression Standard (AC-3, E-AC-3). Rev. B. 2005, alebo v dokumentoch na stránke Dolby (http://www.dolby.com/resources/tech_library/index.cfm).

4.8 Coherent Acoustics (CAC, DTS)

Algoritmus a formát **DTS Coherent Acoustics**, alebo len **DTS** (marketingové označenie), resp. **CAC** (kódové označenie formátu; obe označenia budeme považovať za ekvivalentné) pochádza z dielni firmy Digital Theatre Systems³³.

CAC umožňuje zakódovať až 7.1 kanálov s rozlíšením 24 bitov na vzorku pri vzorkovacej frekvencii 8 až 192 kHz a dátových tokoch siahajúcich od 32 kbps do 6 144 kbps. Pre uľahčenie implementácie však CAC rozdeľuje tok dát na tzv. **jadrový zvuk** (angl. *core audio*) a **rozšírený zvuk** (angl. *extended audio*). Tok dát jadrového zvuku ponúka možnosť prenášať len 5 kanálov s frekvenciami pod 24 kHz a 1 LFE kanál (pod 80, resp. 150 Hz – presnú hodnotu určuje kodér) – dokopy 5.1. Dodatočné informácie, potrebné na rekonštrukciu ostatných kanálov a frekvenčných zložiek (pri vzorkovacej frekvencii vyššej než 48 kHz), prenáša tok dát rozšíreného zvuku.

CAC rozdeľuje okná veľkosti 4096, 2048, 1024, 512 alebo 256 vzoriek (veľkosť je zvolená podľa náročnosti signálu), použitím kvadratúrnych zrkadlových filtrov, na 32 uniformných frekvenčných subpásiem (pri vzorkovacej frekvencii 48 kHz je šírka jedného pásma 750 Hz). Na každé sa jednotlivo hľadajú najvhodnejšie koeficienty lineárnej predikcie³⁴ štvrtého rádu a ak má reziduálny signál po kvantovaní dostatočne nízku variáciu, prenesie sa na výstup spolu s vhodnými koeficientmi, zakódovanými vektorovým kvantovaním³⁵. Ide teda o vylepšenú formu ADPCM kódovania. Ak nie je predikcia dostatočne úspešná, kvantujú sa priamo hodnoty vzoriek. Na vysokofrekvenčné subpásma sa predikcia neaplikuje. LFE kanál sa najprv podrobí decimácii³⁶, potom kvantovaniu na osembitové čísla. Na škálovacie faktory³⁷ v rámci subpásma možno aplikovať jednoduché diferenčné kódovanie (ADPCM). Samotné hodnoty vzoriek (aj reziduálov, ako bolo spomenuté vyššie) sa kódujú spätne adaptívnym kvantovaním, zvyšovaním dátového toku teda možno kompresiu „posúvať bližšie k bezstratovosti“.

Ak je potrebný nízky dátový tok, možno aj hodnoty vzoriek (najmä vysokofrekvenčných subpásiem) kódovať vektorovým kvantovaním. Tiež možno aplikovať spojené kódovanie kanálov, dokonca aj stereofonické kódovanie intenzity³⁸. Súčasťou kódovania je aj kompresia akustickej hladiny, aplikovaná za účelom zmenšenia dynamiky signálu.

DTS, uvedený na trh v r. 1993, používa dodnes mnoho filmových štúdií. Stabilná podpora na trhu a zahrnutie do špecifikácie HD-DVD a BD zaručujú jeho

33 Stručný úvod do histórie a uplatnenia CAC možno nájsť v odseku 3.5.2 *Priestorový zvuk a zvukové kanály*, str. 66.

34 P. odsek 3.4.2 *Kódovanie lineárnou predikciou*, str. 60.

35 Teda čísla sú vybrané z tabuľky (v tomto prípade má 4096 záznamov) a prenáša sa len ukazovateľ na vhodný záznam. P. odsek 3.3.4 *Vektorové kvantovanie (VQ)*, str. 56.

36 P. odsek 3.1.1 *Vzorkovanie*, str. 39.

37 Faktory vyjadrujúce dynamický rozsah (amplitúdu) signálu.

38 P. odsek 3.3.5 *Kompresia stereofonického záznamu*, str. 57.

prežitie. Treba však spomenúť, že bol a asi vždy ostane v tieni DD. Jeho výhodou je otvorenosť štandardu, nízka výpočtová náročnosť, vysoká vernosť (pri použití dostatočne vysokých dátových tokov). Nevýhodou ostáva práve vysoký potrebný dátový tok, ktorý ho robí nepraktickým. Jednoduchý popis formátu možno nájsť na <http://www.dtsonline.com/media/uploads/pdfs/whitepaper.pdf>, presnejšia špecifikácia je dostupná napr. na stránkach ETSI (priama linka: http://webapp.etsi.org/action/PU/20020827/ts_102114v010101p.pdf).

4.9 Ogg Vorbis

Ogg Vorbis je bezpochyby najúspešnejším opensource formátom určeným na stratovú kompresiu zvuku. Pochádza z dielni neziskovej inštitúcie menom **Xiph.org Foundation**³⁹, rovnako ako kontajner **Ogg**, s ktorým a pre ktorý bol **Vorbis** (ako sa tento kompresný algoritmus správne označuje⁴⁰) pôvodne vyvíjaný.

Ogg Vorbis bol od počiatku konštruovaný ako mimoriadne flexibilný formát. Je vhodný pre veľmi nízke i extrémne vysoké dátové toky, podporované rozlíšenie vzorky je 16 alebo 24 bitov⁴¹, vzorkovacia frekvencia od 8 do 192 kHz. Rozloženie kanálov možno tiež ľubovoľne prispôbovať – používateľovi poskytuje až 255 diskretných kanálov⁴². Jednotlivé kanály možno zoskupovať do „podmáp“ (angl. *submaps*), každej možno nakonfigurovať vhodnú spektrálnu charakteristiku, spôsob kódovania atď. Tvorcovia vniesli flexibilitu aj do špecifikácie toku dát, keď zahrnuli možnosť neskoršieho rozšírenia súčasnej generácie kodeku (Vorbis I) o waveletovú dekompozíciu signálu či ambisonický zvuk⁴³ (plánované vo Vorbis II).

Ogg Vorbis používa rámce veľkosti mocnín dvojky, siahajúc od 64 vzoriek po 8 192 vzoriek. MDCT sa (kvôli svojej konštrukcii) aplikuje na okná dvojnásobnej veľkosti. Frekvenčná charakteristika sa vo výstupnom toku dát prenáša najprv ako hrubý odhad (v dokumentácii označovaný ako *floor*), ktorý možno kódovať systémom LSP⁴⁴ (teda je výsledkom LPC, frekvencie sú zapísané v Barkovej mierke), alebo zadaním po častiach lineárnej krivky vyjadrujúcej intenzitu (lineárne rastúcich) frekvenčných zložiek zvuku. Všetky súčasné kodéry používajú druhú možnosť. Floor je v skutočnosti akási spektrálna obálka, určujúca rozsah (a teda kvantovací krok) jednotlivých frekvenčných zložiek. K dátam sa následne pridáva presnejší odhad (v dokumentácii označovaný ako *residue*, teda reziduál; budeme radšej používať anglický variant), ktorý sa kóduje viacprechodovým vektorovým kvantovaním (s rôznou, najlepšie zmenšujúcou sa mierkou). Skutočná amplitúda pre danú frekvenciu je potom súčinom príslušných hodnôt floor a residue.

39 Zakladateľom a hlavným vývojárom je Američan Chris Montgomery. Z dielni Xiph.org pochádzajú aj formáty *Speex* (p. odsek 4.11.1, str. 113), *FLAC* (p. časť 5.4, str. 123) a mnoho iných. Domovská stránka: <http://xiph.org>. Možno na nej nájsť aj projekt Icecast, server na vysielanie multimédií prostredníctvom počítačovej siete, natívne podporujúci Ogg Vorbis.

40 V tejto práci budeme často používať celé spojenie *Ogg Vorbis*, a to z praktických dôvodov – formát Vorbis je často nesprávne označovaný ako „Ogg“, keďže sa zvyčajne ukladá do kontajnera Ogg (s príponou .ogg), zatiaľ čo názov „Vorbis“ je neznámy aj mnohým jeho používateľom. Formát a kontajner sú od seba síce nezávislé (a formát Vorbis možno často nájsť napr. v kontajneri Matroska), v mysliach používateľov sa však spájajú. Aj samotní tvorcovia dbali na ich vzájomnú „súhru“, preto dostáva slovné spojenie Ogg Vorbis svoje opodstatnenie. Na okraj treba uviesť, že Ogg je nielen kontajner, ale najmä názov projektu vývoja opensource multimediálnych riešení od Xiph.org Foundation.

41 Algoritmy podľa špecifikácie Vorbis (vrátane referenčného kodéra) pracujú zväčša v rozlíšení 48 bitov a viac, teda bitová hĺbka vzorky nie je obmedzená samotným formátom, ale kodérom.

42 Referenčný kodér podporuje najviac 5.1.

43 P. odsek 2.5.3 Súčasné trendy, str. 33.

44 P. odsek 3.4.2 Kódovanie lineárnou predikciou, str. 60.

Združenie kanálov prebieha na úrovni residue. Vorbis používa dve metódy: buď sa hodnoty pre jednotlivé frekvenčné zložky a združené kanály prekladajú (angl. *interleave*), čo spôsobuje zoskupenie podobných hodnôt a ľahšie vyjadrenie pomocou VQ, alebo združené kanály použijú jedno residue ako súčtovú intenzitu a druhé na vyjadrenie „uhla“, teda rozdielu fáz medzi kanálmi. (Poznámka: rozdiel intenzít je vyjadrený dostatočne dobre hodnotou floor.)

Vorbis (teda samotný zvukový formát) sám osebe produkuje len „surové dáta“ – sled dátových blokov (tzv. paketov), ktoré majú rozličné dĺžky (špecifikácia nijak neobmedzuje ich dĺžku) a nenesú žiadnu ochranu pred chybami alebo synchronizačné značky. Toto bremeno je na použitom kontajneri. Dekodér musí byť schopný tieto surové dáta dekódovať a správne synchronizovať. Navyše sa naň kladie ťažká požiadavka: nesmie mu prekážať ani „vypchatie“⁴⁵ ani orezanie⁴⁶ paketov. Návrh dekodéra však nie je omnoho výpočtovo náročnejší ako v prípade dekodéra MP3, aj keď použitá pamäť je väčšia. Dôvodom je prenášanie kódových slov (či už ide o entropické kódy alebo tabuľku použitú pri vektorovom kvantovaní) a množstva ďalších informácií určených na presné nakonfigurovanie dekodéra (čo môže byť niekoľko stoviek položiek) v hlavičkách toku dát. Vorbis totiž silno stavia na doprednej adaptívnosti⁴⁷ v každej fáze kódovania a návrhu formátu, zatiaľ čo napr. MP3 používa statický slovník a premenlivé sú len jednoduché preddefinované značky v hlavičke každého rámca. Informácia zapísaná v hlavičkách toku dát formátu Vorbis však pre svoj rozsah nemôže byť súčasťou hlavičky každého paketu, ale po iniciovaní dekodéra informáciami v hlavičkách toku dát môže tento dekódovať tok dát už od ľubovoľného paketu. Táto negatívna vlastnosť je daňou za ohromnú flexibilitu formátu Vorbis. V prípade vysielania (keď je potrebné začať prehrávanie od „ľubovoľného“ miesta) sa hlavičky zväčša prenášajú periodicky s určitým intervalom.

Ogg Vorbis používa tri rôzne hlavičky toku dát. V identifikačnej hlavičke sa nachádzajú základné informácie o vlastnostiach prenášaného zvuku, ako počet kanálov, vzorkovacia frekvencia atď. V komentárovej hlavičke sa prenáša identifikačná značka Vorbis comment⁴⁸. Konfiguračná hlavička obsahuje ďalšie nastavenia a informácie pre dekodér, najmä kódové slovníky pre entropické a VQ kódy.

V dôsledku viacprechodovosti pri kódovaní residue možno odstránením menej podstatných vektorov efektívne znížiť výsledný dátový tok už existujúceho toku dát bez nutnosti opätovného kódovania (samozrejme, za cenu zníženia kvality). Táto črta formátu sa v dokumentácii označuje anglickým slovom (*bitrate*) **peeling** („olúpanie“, „ukrojenie“ dátového toku). Faktom však je, že v súčasnosti (2006-04-25)

45 Angl. *padding*, teda dodanie údajov. P. tiež odsek 3.3.3 *Ortogonálne transformácie v praxi*, str. 52.

46 Pod orezaním sa myslí ukončenie paketu (alebo poškodenie, a teda vnesenie chýb do paketu) ešte pred prenesením všetkých očakávaných dátových polí (napr. informácií zo všetkých prechodov pri kódovaní residue). Dekodér musí byť schopný dekódovať rámec na základe už obdržaných dát.

47 Princíp doprednej adaptívnosti porov. s termínom *dopredne adaptívna predikcia* v odseku 3.4.2 *Kódovanie lineárnou predikciou*, str. 60.

48 p. odsek 3.0.1 *Definícia pojmov*, str. 35.

nejestvuje aplikácia, ktorá by túto črtu formátu efektívne využívala. Kvalita výsledku experimentálneho programu *bpeel* (http://www.vinje.com/ogg_downloads.html) je totiž spochybňovaná.

Otvorenosť formátu Vorbis spôsobila vznik rôznych alternatívnych kodérov. Asi najznámejším sa stal tzv. **AoTuV** kodér (<http://www.geocities.jp/aoyoume/aotuv/>), ktorý svojou kvalitou presvedčil aj „oficiálnych“ vývojárov a jeho aktuálne verzie sú postupne zahŕňané do nových verzií referenčného kodéra. Navyše sa na internete objavujú aj rôzne kodéry s optimalizovanou výpočtovou náročnosťou (napr. na stránke <http://homepage3.nifty.com/blackword/index.htm>). Ak pridáme otvorenosť a bezplatnosť formátu, nie je prekvapujúce, že sa Ogg Vorbis stále častejšie objavuje aj v počítačových hrách (Doom 3, Unreal Tournament 2003 a 2004, Halo a i.) a internetových rádiách (mediálnym aj vysielanie Slovenského rozhlasu).

Formát Ogg Vorbis patrí medzi najrozšírenejšie formáty⁴⁹ na stratovú kompresiu zvuku. Dôvodov je mnoho: je úplne bezplatný, ohromne efektívny, veľmi flexibilný, má širokú podporu (čoraz častejšie už i na prenosných prehrávačoch) a mnoho ďalších výhod. Je však navrhnutý na všeobecný zvukový vstup – v prípade napr. kompresie hovoreného slova sú výhodnejšie iné formáty (napr. Speex, tiež od Xiph.org). Univerzálny referenčný kodér má navyše tendenciu používať veľké rámce (za účelom lepšej dekompozície zvuku a zmenšenia dátového toku), čo spôsobuje vysokú latenciu (oneskorenie), vlastnosť nevhodnú pre kompresiu zvuku (hovoreného slova) v reálnom čase. Definitívna špecifikácia dátového toku Vorbis I bola schválená v máji r. 2000, verzia kodéra 1.0 sa objavila v júli 2002, najnovšia oficiálna verzia (z novembra 2005) nesie označenie 1.1.2. Domovskú stránku organizácie Xiph.org Foundation možno nájsť na <http://xiph.org>, podrobná dokumentácia Ogg Vorbis je dostupná na stránke <http://xiph.org/vorbis/doc>. Filtre na prehrávanie v prostredí Windows možno nájsť na <http://www.illiminable.com/ogg>.

⁴⁹ Medzi formátmi rozšírenými na (zväčša nelegálnych) peer-to-peer sieťach v r. 2005 (Zdroj: http://www.cachelogic.com/research/2005_slide18.php) je v používanosti tretí s podielom 12,3% (za MP3 s 64,89% a WMA s 22,81% podielom).

4.10 Musepack (MPC)

Musepack (v skratke **MPC**, kedysi nazývaný aj MPEGplus, MPEG+, MP+) je frivérovým⁵⁰ stratovým kompresným audioformátom vychádzajúcim z MP2. Za jeho vznikom (v r. 1997) stál Nemecký Andree Buschmann, neskôr sa vedúcim projektu a organizácie, ktorá MPC vyvíja (Musepack Development Team), stal jeho rodák Frank Klemm. Určenie je podobné formátu MP2 – dosiahnuť priesvitnosť záznamu pri dátových tokoch okolo 180 kbps (stereo, 16 bit, 44,1 kHz). Algoritmus však dosahuje obstojné výsledky i pre nižšie dátové toky.

Návrh MPC je do značnej miery zhodný s MP2. Lepším ho však robia črty, o ktoré bol MPC oproti MP2 obohatený, napr. technika náhrady šumu v zázname⁵¹, rozhodovanie o spojenom kódovaní kanálov osobitne pre jednotlivé subpásma, efektívnejšie entropické kódovanie subpásmami a i.. Formát je natívne VBR (od 3 do 1300 kbps), nepodporuje CBR/ABR, preto je na vysielanie nevhodný. Najvyššia vzorkovacia frekvencia je 48 kHz.

MPC používa identifikačnú značku APEv2⁵², ale podporuje i APEv1, ID3v1 a ID3v1.1. Súčasná (2006-04-26) verzia špecifikácie toku dát (StreamVersion7 z 2005-06-28) nepodporuje pretáčanie (seek tables⁵³), podporuje najviac dva kanály a rozlíšenie vzorky 16 bitov. S ďalšou verziou (SV8) by sa to malo zmeniť.

Napriek pôvodnej optimalizácii na vysoké dátové toky dosahuje MPC aj pri štandardných tokoch (okolo 128 kbps) veľmi dobré výsledky. V komunite priaznivcov opensource riešení je obľúbený, aj keď vo všetkom, zdá sa, ho predčí Ogg Vorbis – či už ide o univerzálnosť, kvalitu pri danom dátovom toku, neohrozenie patentmi⁵⁴ a i. Softvérová i hardvérová podpora sa stále zväčšuje, teda nie je dôvod tvrdiť, že MPC je nejakým spôsobom zastaraný (keďže vychádza z technológie starej vyše 15 rokov). Domovskú stránku možno nájsť na <http://www.musepack.net>, aktuálny vývoj sa dá sledovať na stránke Franka Klemma <http://www.personal.uni-jena.de/~pfk/mpp/index2.html>.

50 Aspoň súčasné verzie kodéra. Aktuálny dekodér je opensource.

51 Podobné percepčnej náhrade šumu v AAC, p. časť 4.3 *MPEG-2/4 Audio (AAC)*, str. 94.

52 P. odsek 3.0.1 *Definícia pojmov*, str. 35.

53 Tamže.

54 Nad MPC visí Damoklov meč – najmä rôzne patenty súvisiace s MP2. Cit: „In the past, MPC has been under suspicion of violating multiple patents (MP2, PNS, subband) [1]. According to the developers of MPC, all patented code has been removed and it is now believed to be free of patent encumbrance. However, one PNS patent application is still active [2], and it is not trivial to know if MPC's own "noise substitution techniques" avoid its scope or not. Only a careful examination by a patent lawyer would say whether Musepack is truly patentless or not.“ (Zdroj: <http://en.wikipedia.org/wiki/Musepack>, cit. 2006-04-26.)

4.11 Kompresia hovoreného slova

Aj keď kompresné formáty spomenuté v predošlých častiach kapitoly 4 sú univerzálne a použitie nachádzajú najmä pri kompresii hudby, potreba komprimovať zvuk vyvstala z celkom iných pohnútok. S rapidným rastom množstva účastníckych staníc telefónnych rozvodných sietí v priebehu celého dvadsiateho storočia, neskôr s príchodom mobilných sietí, ba možnosťou vysielat', archivovať, syntetizovať a analyzovať záznam ľudskej reči vznikla potreba nájsť vhodný kompresný postup, ktorý by vedel využiť danosti záznamu ľudského hlasu a signifikantne zmenšiť potrebný dátový tok, a to aj za cenu zníženia kvality (nie zrozumiteľnosti) výsledného dekomprimovaného záznamu. Veľmi nízky dátový tok je výhodný aj pri vojenskom využití komunikačných prostriedkov, kde uľahčuje šifrovanie dát a ochranu pred chybami v prenose.

V praxi sa zvyčajne na záznam a prenos hovoreného slova používa vzorkovacia frekvencia 8 kHz⁵⁵ a rozlíšenie vzorky 8 bitov (často sa použije nelineárne kvantovanie – napr. *A-law* v Európe či *μ -law* v Japonsku a USA). Bez kompresie by bol dátový tok 64 kbps, kompresia ho môže ešte výrazne zmenšiť. Jedným z najstarších spôsobom kompresie hovoreného slova je tzv. **vocoder** (z angl. voice coder – kodér hlasu). Vychádza z konštrukcie ľudských hlasiviek⁵⁶. Na tomto mieste si popíšeme jeho digitálnu verziu. Hlasový trakt vocoder chápe ako časovo premenlivý frekvenčný filter. Vstupom filtra je biely šum (pre nehlasové zložky reči, ako dych, sykavky a pod.) alebo pravidelná pulzná charakteristika (ktorá napodobňuje budenie, teda excitovanie hlasiviek pri vydávaní zvuku daného kmitočtu). Vhodnou analýzou krátkeho časového intervalu hlasového záznamu možno zistiť druh vstupu (šum, hlas) a parametre filtra. Vocoder sa štandardne používajú pri dátovom toku 2,4 kbps, pri ktorom je zvuk síce veľmi neprirodený, ostáva však zrozumiteľný.

Filter, odrážajúci jednotlivé formanty, možno popísať aj na základe LPC⁵⁷ – jej princíp je koniec koncov totožný s filtrom s konečnou impulzovou charakteristikou. Na zvolený rámec (krátky zvukový interval, pri vocoderoch 20 ms, teda 160 vzoriek) sa aplikuje dopredne adaptívna lineárna predikcia s vhodným rádom (zvyčajne 10) a koeficientmi predikcie. Pre nízky počet použitých vzoriek sa táto časť spracovania nazýva **krátkodobá predikcia**. Ostávajúci reziduálny signál je vlastne vstupný (budiaci) signál (zväčša tón istej frekvencie a intenzity), alebo šum (sykavky), alebo kombinácia oboch (napr. hláska „z“). O vhodnom zápise reziduálneho signálu rozhoduje jeho frekvenčná analýza – buď sa priamo popíše tón (resp. či ide o šum) a jeho intenzita, alebo sa použije vektorové kvantovanie. Výsledkom je signifikantné zníženie dátového toku. Vocoder delia zvuk len na šumový/tónový, koeficienty

55 Pri 8 kHz sa hovorí o úzkopásmovom (angl. *narrowband*) zvuku. Frekvenčný rozsah je zvyčajne 300-3 400 Hz. Ak je vzorkovacia frekvencia 16 kHz (rozsah 500-7 000 Hz) alebo viac, hovorí sa o širokopásmovom (angl. *wideband*) zvuku.

56 O vlastnostiach záznamu ľudského hlasu sme si hovorili v odseku 2.3.3 (str. 23).

57 P. odsek 3.4.2 *Kódovanie lineárnou predikciou*, str. 60.

kódujú systémom LSP. Reziđuálny signál vřak veľmi ovplyvňuje výsledok dekódovania; pre jeho nepresný popis znie zvuk pochádzajúci z Vocodera neprirodzene.

Novřie techniky na kompresiu zvuku viacmenej vychádzajú z algoritmu CELP⁵⁸ (angl. *Code Excited Linear Prediction*, lineárna predikcia excitovaná kódom). Oproti Vocoderom sa snaží popísať reziđuálny signál čím presnejřie. Použitý dátový tok dosahuje zväčřa 4,8 kbps, veľkosť rámca je 30 ms (240 vzoriek). Koefficienty krátkodobej predikcie sa kódujú systémom LSP. Dominantná frekvencia budiaceho tónu sa prejaví v periodicite reziđuálneho signálu. Periódu možno vyjadriť vo vzdialenosti vo vzorkách, teda aktuálnu vzorku v doterajřom reziđuálnom signáli možno dobre predikovať práve z hodnoty vzorky vzdialenej jednu periódu. Keďže táto vzdialenosť je omnoho vyšřia než rád krátkodobej predikcie, nazýva sa tento postup **dlhodobá predikcia**. Výsledný reziđuálny signál, ktorý zostane po oboch predikciách, sa nazýva **inovácia**. Kóder ho rozdelí na štyri subrámce a kóduje pomocou VQ. Každé kvantovanie inovačného signálu má vřak silný vplyv na dekódovaný signál, preto (v snahe zmenřiť odstup signálu od řumu, resp. zvyřiť prirodzenosť zvuku aplikovaním psychoakustického modelu) sa vykonáva opakované hľadanie najvhodnejřích reprezentantov inovačného signálu z kódového slovníka vektorov.

Treba podotknúť, že na kompresiu hovoreného slova sa s obľubou (najmä pre svoju výpočtovú jednoduchosť a nízku latenciu) používa aj ADPCM.

4.11.1 Speex

Spomenieme si jedného zástupcu algoritmov na kompresiu hovoreného slova z dielne Xiph.org⁵⁹ – **Speex**. Na rozdiel od väčřiny ostatných algoritmov v tejto triede je Speex opensource. Navyře, jeho ťažisko nie je komunikácia prostredníctvom mobilných sietí, ale prostredníctvom internetu (VoIP, teda tzv. IP telefónia). Dôležitá je teda jeho odolnosť voči strate paketov, nie ich pořkodeniu. Ako Vorbis a FLAC, aj Speex, ak ho je potrebné uložiť do súboru, funguje v tandeme s kontajnerom Ogg (zaužívala sa prípona .spx).

Speex je založený na algoritmoch typu CELP. Má podporovať vysokú kvalitu záznamu aj veľmi nízke dátové toky (nie súčasne). Preto môže tok dát obsahovať nielen nosné úzkopásmové spektrum (vzorkovacia frekvencia 8 kHz), ale aj dodatočné informácie na rekonštrukciu širokopásmového spektra (16 kHz, ba i 32 kHz a 48 kHz). Podobne ako Vorbis, určený na kódovanie všeobecných zvukových vstupov, i „hlasový kodek“ Speex je veľmi flexibilný: poskytuje dátové toky od 2 do 44 kbps (CBR, ABR, VBR) a premenlivú komplexnosť kódéra (desať

58 Bližří popis algoritmov Vocoder a CELP možno nájsť napr. na <http://www.data-compression.com/speech.html>.

59 P. časť 4.9 *Ogg Vorbis*, str. 108.

stupňov – najvyšší je výpočtovo päťkrát náročnejší, ale redukuje odstup signálu od šumu o pribl. 1-2 dB). Ďalšie črty sú stereofonické kódovanie intenzity (ostatné formáty ponúkajú len mono), detekcia hlasovej aktivity (pri VBR sa výrazne zníži dátový tok, keďže sa prenáša len informácia o „prítomnosti ticha“; v prípade vysielania sa neprenášajú dáta vôbec) a i.

Použitý rámec má 20 ms, teda 160 vzoriek. V prípade širokopásmového záznamu sa tento najprv rozdelí kvadrátovým zrkadlovým filtrom na dva úzkopásmové záznamy. Pri dlhodobej predikcii sa využíva nielen vzorka so vzdialenosťou periódy kmitočtu, ale aj jej susedné vzorky. Latencia (časové oneskorenie kodéra) dosahuje 30 ms (pri vzorkovacej frekvencii 8 kHz) alebo 34 ms (16 kHz).

Speex sa stáva stále populárnejším riešením, keď ide o šírenie záznamov hovoreného slova cez internet. Hlavným dôvodom je jeho bezplatnosť a softvérová podpora. Domovskú stránku možno nájsť na <http://www.speex.org>, dokumentáciu na <http://www.speex.org/manual.pdf>. Prvé verzie formátu Speex vznikli v roku 2002, verzia 1.0 prišla na svet v marci 2003. Aktuálna verzia kodeku (1.1.12) je z februára 2006 a zahŕňa v sebe aj alternatívny psychoakustický model algoritmu Vorbis, zlepšujúci tak kvalitu pre vyššie dátové toky (8 kbps a viac)⁶⁰.

60 P. VALIN, Jean-Marc – MONTGOMERY, Christopher. Improved Noise Weighting in CELP Coding of Speech - Applying the Vorbis Psychoacoustic Model to Speex. 2006.

5 PREHLAD BEZSTRATOVÝCH ZVUKOVÝCH FORMÁTOV

V nasledujúcich častiach kapitoly 5 ponúkame prehľad rôznych bezstratových kompresných algoritmov a formátov. Samozrejme, toto je len výber z mnohých – bezstratových formátov existuje nepreberné množstvo, mnohé z nich sa však nikdy skutočne neujali (napr. LTAC¹, Ogg Squish), alebo mali od počiatku špecifické použitie s obmedzeným dosahom a v svete PC sa ani nemohli uplatniť (napr. DST), alebo sú utajené, teda je ťažké získať o nich relevantné informácie a ich podiel na trhu je obmedzený na niekoľko proprietárnych produktov (napr. Apple Lossless – ALAC, alebo WMA Lossless – WMAL, alebo Real Audio Lossless – RAL), alebo sa ujali, no ich projekty časom zanikli (napr. RK Audio – RKAU).

Táto kapitola pozostáva z desiatich častí. Prvá krátka časť (5.1, str. 117) predstaví formát *Shorten*, jeden z najstarších, aj keď v súčasnosti už len zriedka používaný algoritmus. Nasledujú proprietárne a opensource formáty *WavPack* (5.2, str. 119), *Monkey's Audio (MA, APE)* (5.3, str. 121), *FLAC* (5.4, str. 123), *Lossless Audio (LA)* (5.5, str. 125), *True Audio (TTA)* (5.6, str. 126) a *OptimFROG* (5.7, str. 127). V ďalších častiach sú uvedené formáty, ktoré patria (5.8 *ALS*, str. 128) alebo budú patriť (5.9 *SLS*, str. 130) do špecifikácie MPEG-4. Na záver uvádzame, pre jeho rozšírenosť, špecifický formát *Meridian Lossless Packing (MLP)* (5.10, str. 131).

Treba povedať, že bezstratové formáty sa oproti stratovým líšia nielen princípom činnosti, ale aj možnosťou zásahov zo strany používateľa. Kým stratové formáty možno konfigurovať najmä jedným parametrom – želaným dátovým tokom – v prípade bezstratových často možno nastaviť úroveň hľadania optimálnej predikcie či zápisu dát. Činnosť týchto algoritmov sa potom dá deliť na dve základné kategórie:

- **Symetrické kódovanie.** Táto vlastnosť je typická pre spätné adaptívne postupy. Parametrom pri kódovaní možno zvýšiť výpočtovú náročnosť kódovania za účelom zníženia výsledného dátového toku. Výsledkom je však aj proporcionálne zvýšenie náročnosti dekódovania. S týmto prístupom sa môžeme stretnúť v algoritmoch MA a LA.
- **Asymetrické kódovanie.** Táto vlastnosť je typická pre dopredne adaptívne postupy. Parametrom pri kódovaní možno zvýšiť výpočtovú náročnosť kódovania za účelom zníženia výsledného dátového toku, pričom výpočtová náročnosť dekódovania ostáva zachovaná (resp. je len marginálne vyššia). Asymetrické kodéry majú síce oproti symetrickým omnoho rýchlejšie dekódovanie, rýchlosť kódovania však zvykne byť pri rovnakej účinnosti nižšia (dokonca vo všeobecnosti dosahujú omnoho nižšiu účinnosť). S týmto prístupom sa môžeme stretnúť v algoritmoch FLAC, TTA a ALS.

1 P. odsek 3.4.1 *Bezstratové transformačné kódovanie*, str. 59.

Výnimkou z tejto schémy sú napr. formáty WavPack a OptimFROG, ktoré ponúkajú obe možnosti na zefektívnenie kompresie, WavPack je navyše hybridný.

5.1 Shorten

Algoritmus **Shorten** vyvinul v rokoch 1992-1993 Angličan (vtedy študent Cambridgu) menom Tony Robinson, teraz z firmy SoftSound Ltd. Licencia formátu umožňuje bezplatné nekomerčné využitie. Formát pôvodne vznikol ako výsledok projektu na kompresiu hovoreného slova so zvoleným odstupom signálu od šumu, s cieľom jeho archivácie a použitia pri výskume algoritmov na rozpoznávanie reči. Ide teda o hybridný kodér (aj keď v praxi sa ujal len ako bezstratový).

Shorten je založený na lineárnej predikcii, pričom v sebe integruje štyri pevne zvolené predikčné formuly:

$$P^0=0,$$

$$P^1(x_{i-1})=x_{i-1},$$

$$P^2(x_{i-1}, x_{i-2})=2x_{i-1}-x_{i-2},$$

$$P^3(x_{i-1}, x_{i-2}, x_{i-3})=3x_{i-1}-3x_{i-2}+x_{i-3}.$$

Okrem nich ponúka aj adaptívnu predikciu ľubovoľného rádu. Podporuje dva kanály, no nevyužíva medzikanálovú redundanciu. Najvyššia vzorkovacia frekvencia je 44,1 kHz.

Stratová kompresia je založená na kvantovaní reziduálov. Shorten ponúka dva módy – môd s vyrovnaným dátovým tokom medzi rámcami a môd so zvolenou najväčšou energiou skreslenia reziduálneho signálu v rámci. Energia skreslenia však nie vždy zodpovedá odstupom signálu od šumu v dekódovanom signáli – do hry vstupuje i použitá lineárna predikcia. Čím je korelácia zdroja vyššia, tým viac sa prejavujú odchýlky v reziduáloch na výsledných hodnotách vzoriek. Kvantovanie reziduálov a koeficientov navyše prebieha s kvantovacím krokom rovným mocnine dvojky (nižšia výpočtová náročnosť), čo znižuje adaptívnosť riešenia.

Shorten ukladá výsledný tok dát do súboru .shn, ktorý bol neskôr rozšírený napr. o tabuľky na pretáčanie (seek-tables) niekoľkými programátormi (Wayne Steilau, Jason Jordan) zhromaždenými v projekte etree.org². Práve tento projekt zaručil formátu Shorten popularitu, keďže bol na ich stránkach vybraný za hlavný spôsob šírenia zvukových záznamov (neskôr jeho miesto nahradil formát FLAC).

Shorten, po toľkých rokoch od jeho vzniku, nedokáže v súčasnosti úspešnosťou kompresie konkurovať iným formátom, ale jeho nízka výpočtová náročnosť a podpora v rôznych prostrediach – vyplývajúca najmä z dlhého času, čo je na trhu – ho robia akýmsi štandardom na poli bezstratovej kompresie zvuku. Z dlhodobého hľadiska sa však do tohto formátu určite neoplatí investovať. Posledná autorova verzia³ (2.3a)

2 Stránka <http://www.etree.org> vznikla v roku 1998 ako iniciatíva hudobných fanúšikov, ktorí túžili legálne distribuovať nahrávky z koncertov svojich obľúbených skupín (s ich povolením), pričom stratová kompresia bola pre nich neprijateľná.

3 Možno stiahnuť na <http://www.softsound.com/wsh23b32.EXE>.

bola vydaná vo februári 1999. Najnovšia rozšírená verzia⁴ (3.6.0) uzrela svetlo sveta v auguste 2004. Veľmi chudobnú domovskú stránku možno nájsť na <http://www.softsound.com/Shorten.html>, stručnú dokumentáciu verzie 1.22 možno nájsť na <http://svr-www.eng.cam.ac.uk/reports/ajr/TR156>.

⁴ Dá sa získať zo stránky <http://www.etree.org/shnutils/shorten>.

5.2 WavPack

WavPack je opensource hybridný bezstratový kompresný formát, ktorý vyvinul v r. 1998 Američan David Bryant. Algoritmus naprogramoval pôvodne pre vlastné potreby – CD-ROM médiá mali vtedy kapacitu len 650 MB až 700 MB (t.j. asi 66 minút), keď teda chcel spraviť zálohu hudobného CD na CD-ROM, potreboval ho skomprimovať. WavPack stavia najmä na kvalite kompresie v kombinácii s vysokou rýchlosťou kompresie i dekompresie.

WavPack je založený na lineárnej predikcii, používa adaptívny faktor predikovateľnosti. Koeficienty sú predvolené (podobne ako v prípade formátu Shorten), predikcia siaha až do ôsmeho rádu. Predikcia funguje vo viacerých prechodoch, teda namiesto kódovania reziduálu po vykonaní prvej predikcie sa pristupuje k ďalšej predikcii, sériovo. Používajú sa dve až šesť predikčných vrstiev. Podporované vzorkovacie frekvencie siahajú k 192 kHz, rozlíšenie najviac 32 bitov (celé číslo alebo IEEE float 32 pre stratový mód).

Pri kódovaní stereofonického záznamu sa predikcia počíta pre oba kanály nezávislo, neprenáša sa však výsledný reziduálny signál, ale rozdielový kanál spolu so súčtovým, ľavým, alebo pravým kanálom, podľa toho, ktorá alternatíva je výhodnejšia. WavPack umožňuje kódovať najviac 5.1 kanálov.

WavPack ponúka okrem symetrického nastavovania stupňa kompresie aj dodatočnú voľbu pre asymetrické kódovanie – miesto použitia pevne zvolených koeficientov predikčných funkcií sa nájdu optimálne hodnoty pre kódovaný rámec. To značne zvyšuje výpočtovú náročnosť kompresie, ale výsledkom je zmenšenie výsledného dátového toku bez zmeny rýchlosti dekompresie.

Ďalšou zaujímavosťou je výber osobitného, autorom navrhnutého kódovania čísel – ako priznáva, jeho efektívnosť je v porovnaní s Riceovými kódmi nižšia asi o 0,15 bitu na vzorku, výhodou je však ľahšie oddelenie bitov s vyššou hodnotou a bitov s nižšou hodnotou, čo príde vhod pri stratovej kompresii.

WavPack ukladá stratovo (hybridne) aj bezstratovo skomprimované súbory do vlastného súboru .wv, korekčný súbor (pri hybridnom kódovaní) má príponu .wvc. Pri hybridnom kódovaní zadá používateľ žiadaný dátový tok (autor uvádza 192 kbps ako najmenší vhodné číslo, 256 kbps ako štandardnú hodnotu, 320 kbps na dosiahnutie vysokokvalitného záznamu). Používateľ si vyberie, či je prioritou najmenšia výsledná veľkosť oboch súborov, alebo je dôležitejšia subjektívna kvalita stratového súboru pri danom dátovom toku bez ohľadu na veľkosť korekčného súboru. Formát podporuje identifikačné značky (ID3v1, APEv1, APEv2), zhlásenie pri reprodukcii a umožňuje vysielanie.

WavPack má dobrú softvérovú podporu a dočkal sa aj hardvérovej podpory zo strany niekoľkých (zatiaľ len veľmi málo) prenosných prehrávačov, a to najmä vďaka tomu, že pri kódovaní a dekódovaní sa používa len celočíselná aritmetika (ak je vstup celočíselný).

WavPack ponúka bezkonkurenčne rýchlu kompresiu a extrémne rýchlu dekompresiu (ktorej sa vyrovnajú len Shorten a FLAC). Vysoká efektivita kompresie napriek jednoduchému návrhu predstihuje väčšinu konkurencie. Pôvodný autor je i naďalej jeho jediným vývojárom. Najnovšia verzia (4.32) pochádza z apríla 2006, z čoho vidno, že na ňom stále pracuje. Aj keď práca na ňom je na pleciah jedného človeka, vzhľadom na jeho kvality ho možno určite len odporučiť. Domovskú stránku možno nájsť na <http://www.wavpack.com>, veľmi stručnú technickú dokumentáciu k staršej verzii (3.97) na <http://www.wavpack.com/397/technical.htm>.

5.3 Monkey's Audio (MA, APE)

Autorom formátu **Monkey's Audio** (ďalej len **MA**; používa sa aj označenie **APE**, podľa vlastného kontajnera .ape) je Američan Matthew T. Ashland. Jeho vývoj sa začal niekedy pred rokom 2000⁵, pričom od počiatku bol tvorený ako proprietárne riešenie s možnosťou bezplatného využitia pre nekomerčné použitie. Pre MA je prioritou kvalitná kompresia pri zachovaní prijateľnej výpočtovej náročnosti.

MA je symetrický algoritmus založený na spätne adaptívnej lineárnej predikcii. Podporuje stereo (aplikuje združenie kanálov), vzorkovaciu frekvenciu do 48 kHz. Rozlíšenie do 16 bitov. Používa rozsahové kódovanie. Nezvyčajným je použitie neurónových sietí na stanovenie vhodných koeficientov prediktora – nepoužíva totiž nezávislé rámce⁶, ale tok dát je nepretržitý. To výrazne spomaľuje pretáčanie a spôsobuje, že výskyt chyby poškodí zvukový záznam až do konca. Formát však obsahuje kontrolné súčty na detekciu prípadnej chyby v toku dát.

Hlavnou silou MA je vysoká efektivita kompresie. To je zrejme hlavný dôvod, prečo sa tento formát teší popularite – a zrejme ešte dlho bude. V prostredí Windows ponúka príjemné používateľské prostredie a vysokú podporu.⁷

Jeho hlavnou nevýhodou bola donedávna reštriktívna licencia – MA bolo možné voľne používať, no nesmeli sa bez povolenia modifikovať, redistribuovať atď. Každé použitie v programoch s otvorenejšími licenciami (napr. GPL) bolo znemožnené, komerčne sa MA dalo použiť iba so súhlasom autora (ku ktorému bolo prakticky nemožné sa dostať). Prednedávnom (r. 2005) autor uvoľnil kód, no chýba ozajstná príručka, vysvetlenie formátu, princípu. Z toho dôvodu bol všetok vývoj na pleciach jedného človeka, čo prakticky ešte stále obmedzuje vplyv MA na platformu Windows. Našťastie, v uplynulých mesiacoch autor zmenil licenciu, ktorá sa priblížila opensource licenciám, no ešte stále nie je kompatibilná s GPL a inými často používanými licenciami. Jediný spôsob pre fanúšikov tohto formátu je teda naprogramovať si podporu z vlastných síl. V uplynulom čase sa objavilo niekoľko takých projektov⁸.

MA ponúka postačujúcu rýchlosť a vysokú účinnosť kompresie, čo ho robí favoritom medzi súčasnými používateľmi systému Windows. Aj konzorcium MPEG ho vo svojej výzve na návrh bezstratovej kompresie⁹ označuje ako dovtedajší vrchol vývoja („state-of-the-art“). Donedávna sa zdalo, že z dlhodobého hľadiska boj s konkurenciou prehrá, a to z dôvodu nízkej podpory mimo prostredia Windows

5 Autorovi tejto práce sa nepodarilo zistiť presný dátum.

6 V prípade nezávislých rámcov možno vykonať dopredne adaptívnu predikciu, teda nájsť najvhodnejšie prediktory pre daný rámec.

7 Matthew T. Ashland nikdy nerozšíril jeho podporu na iné operačné systémy, urobili to však zaňho iní nadšenci.

8 Ich zoznam môžete nájsť v článku na wikipedii: http://en.wikipedia.org/wiki/Monkey's_Audio [cit. 2006-04-28].

9 P. Call for Proposals on MPEG-4 Lossless Audio Coding. 2002. s. 4.

a reštriktívnej licenčnej politiky. Povest' mu zhoršil aj neflexibilný, a preto nestabilný návrh formátu – novšie verzie kodekov MA síce boli schopné pracovať so súbormi kódovanými staršími verziami, naopak to však nie vždy platilo. V súčasnosti sa mnohé tieto neduhy podarilo odstrániť a MA má veľké množstvo verných používateľov. Najnovšia verzia (4.01 beta 2) pochádza z februára 2006. Domovskú stránku možno nájsť na <http://www.monkeysaudio.com>, zjednodušenú technickú dokumentáciu na <http://www.monkeysaudio.com/theory.html>.

5.4 FLAC

Formát **FLAC** (skratka z Free Lossless Audio Codec) vyvíja skupina programátorov pod vedením Američana Josha Coalsona. Kodek debutoval v decembri roku 2000, v júli 2001 vyšla verzia 1.0 (s ňou i ustálenie špecifikácie toku dát). Autor chcel svetu ponúknuť výpočtovo nenáročný a otvorený kodek, pričom staval na kvalitnom návrhu zaručujúcom spätnú kompatibilitu. Používa vlastný kontajner .flac, aj keď sa často používa alternatívne riešenie, tzv. Ogg FLAC, teda uloženie v kontajneri s príponou .ogg¹⁰.

FLAC je asymetrický¹¹ kodek založený na lineárnej predikcii. Referenčný kodér podporuje vzorkovaciu frekvenciu do 96 kHz (formát samotný nemá obmedzenie vzorkovacej frekvencie), rozlíšenie vzorky do 24 bitov (formát podporuje maximálne 32 bitov). Veľkosť rámca nie je obmedzená, aj keď referenčný kodér používa konštantnú veľkosť (zvolenú vzhľadom na vzorkovaciu frekvenciu zdroja). Pri kódovaní každého rámca sa vyberie najlepšia z týchto alternatív:

- Doslovný prepis, tzn. žiadne kódovanie.
- Konštantný blok, vtedy sa použije obyčajné RLE¹².
- Použije sa jeden z prediktorov z tabuľky – tá obsahuje všetky prediktory algoritmu Shorten¹³, navyše dodáva prediktor štvrtého rádu:
$$P^4(x_{i-1}, x_{i-2}, x_{i-3}, x_{i-4}) = 4x_{i-1} - 6x_{i-2} + 4x_{i-3} - x_{i-4} .$$
- Lubovoľný prediktor, a to až do rádu 32. Na kvantovanie koeficientov sa v každom rámci môže použiť iný kvantovací krok.

Pri kódovaní stereofonického záznamu sa predikcia počíta pre oba kanály nezávislo, neprenáša sa však výsledný reziduálny signál, ale rozdielový kanál spolu so súčtovým, ľavým, alebo pravým kanálom, podľa toho, ktorá alternatíva je výhodnejšia. FLAC podporuje najviac 8 kanálov.

Dekompresia potrebuje len celočíselnú aritmetiku, čo spravilo z formátu FLAC ľahko hardvérovo implementovateľný bezstratový kompresný algoritmus. Keďže prioritou bola nízka výpočtová náročnosť dekompresie, špecifikácia nepodporuje vzorkovaciu frekvenciu nad 96 kHz, aj keď samotný formát ju podporuje (pri kódovaní možno deaktivovať kontrolu súhlasu parametrov so špecifikáciou).

Treba povedať, že úspešnosť kompresie algoritmu FLAC nie je ničím výnimočná. Zato rýchlosť je veľmi dobrá, podpora kontrolných súm

10 Keďže sa projekt FLAC v januári 2003 dostal pod krídla Xiph.org Foundation. P. časť 4.9 *Ogg Vorbis*, str. 108.

11 Stanovený stupeň kompresie má mierny, no takmer zanedbateľný vplyv na rýchlosť dekompresie, p. <http://people.ucsc.edu/~rswilson/flactest>.

12 Z anglického *run length encoding*. Kódujú sa len dve čísla – počet vzoriek v bloku a ich hodnota (rovnaká pre každú vzorku).

13 P. časť 5.1, str. 117.

(FlacFingerprint), identifikačnej značky (FlacMetadata¹⁴), široká softvérová a hardvérová podpora a úplná otvorenosť z neho urobili de-facto priemyselný štandard na poli bezstratovej kompresie. V dôsledku toho mnohé skupiny (medziiným Metallica, They Might Be Giants a iné) využívajú FLAC na šírenie svojej hudby, resp. záznamov z koncertov. Na etree.org tiež získal veľmi rýchlo popularitu, na úkor formátu Shorten. Najnovšia verzia (FLAC 1.1.2) je z februára roku 2005. Domovskú stránku možno nájsť na <http://flac.sourceforge.net>, dokumentáciu na <http://flac.sourceforge.net/format.html>.

¹⁴ Je vnútorne úplne zhodné s Vorbis comment, používaným vo formáte Ogg Vorbis. P. odsek 3.0.1 *Definícia pojmov*, str. 35.

5.5 Lossless Audio (LA)

Formát **Lossless Audio (LA)** pochádza od Novozélandčana menom Michael Bevin, prvá verzia (0.2) sa objavila v auguste 2002. Ide teda o veľmi mladý kodek, ktorý navyše ešte stále nedosiahol verziu 1.0. Snahou autora bolo vytvoriť algoritmus na najefektívnejšiu bezstratovú kompresiu zvuku, pri ktorej nízka časová náročnosť nie je prioritou (presne naopak, ide o jeden z najpomalších algoritmov zo všetkých popisovaných v tejto práci).

LA je symetrický algoritmus založený na lineárnej predikcii. Aplikuje 8 vrstiev predikcie. Používa adaptívny faktor predikcie na každej vrstve, čísla kóduje spätne adaptívnym aritmetickým kódovaním. Používa vlastný kontajner .la.

Osobitosťou LA je riešenie medzikanálovej korelácie na úrovni predikcie – prediktory na svoj odhad používajú vzorky z oboch kanálov¹⁵. Tento prístup, aj keď časovo veľmi náročný, umožňuje dosiahnuť neuveriteľne dobrú kompresiu.

Formát LA ponúka veľmi dobré výsledky kompresie, aj keď rýchlosť je nízka. Je však v rukách jedného nadšenca, ktorý na ňom už dlhší čas nepracuje (najnovšia verzia – 0.4b – je z februára 2004). Domovskú stránku možno nájsť na <http://www.lossless-audio.com>, zjednodušenú technickú dokumentáciu na <http://www.lossless-audio.com/theory.htm>.

¹⁵ Podobne ako OptimFROG, p. časť 5.7, str. 127.

5.6 True Audio (TTA)

Formát **True Audio (TTA)** je projektom softvérovej komunity True Audio software, zhromaždenej okolo dvoch Rusov – Alexandra V. Djourika a Pavla Žilina. Od roku 1999 pracovali na bezstratovej kompresii za účelom zálohovania dát nasnímaných vesmírnymi rádioteleskopmi. V roku 2002 sa rozhodli použiť svoje skúsenosti na konštrukciu symetrického algoritmu na bezstratovú kompresiu zvukového signálu a verejnosti predstavili opensource kodek s úspešnou kompresiou i rýchlosťou, konštruovaný s úmyslom budúcej hardvérovej podpory.

TTA podporuje viackanálový zvuk, rozlíšenie vzorky 8, 16 a 24 bitov. Vzorky sú najprv podrobené delta modulácii (LPC rádu 1) s adaptívnym faktorom predikcie, na zvyškový signál sa v druhom prechode aplikuje LPC so spätne adaptívnymi koeficientmi (autori tento postup nazvali **adaptívne prognostické filtrovanie**). Používa jednoduchú medzikanálovú dekoreláciu (M/S) a mimoriadne dlhé bloky – až okolo jednej sekundy. To spôsobuje problém pri prehrávaní a úpravách skomprimovaného záznamu. Používa Riceovo kódovanie s adaptívnou úpravou dĺžky zápisu.

Formát používa vlastný kontajner .tta, podporuje identifikačné značky ID3v1, ID3v2. Autori úspešne spolupracujú aj s vývojármi kontajneru Matroska na zahrnutie podpory pre TTA do ich kontajnera.

TTA ponúka dobrú kompresiu pri nízkej výpočtovej náročnosti, existuje dokonca jeden DVD prehrávač, ktorý ho podporuje. Softvérová podpora je dostatočná. Slabé rozšírenie na trhu však zatiaľ hovorí proti nemu. No stále podlieha vývoju, najnovšia verzia (3.3) je z mája 2005. Domovskú stránku možno nájsť na <http://www.true-audio.com> alebo <http://tta.sourceforge.net>, popis algoritmu možno nájsť napr. na <http://www.true-audio.com/Codec.theory>.

5.7 OptimFROG

OptimFROG je frivérovým projektom Rumuna menom Florin Ghido, ktorý ho vyvíja od roku 1996. Verejnosti však poskytol až verziu 4.0b v roku 2001. Prioritou je účinnosť kompresie s nastaviteľnou výpočtovou náročnosťou.

OptimFrog je symetrický kodek založený na dopredne adaptívnej lineárnej predikcii. Popri tom ponúka aj dodatočnú možnosť zapnutia asymetrického kódovania. Podporuje dva kanály, hĺbku vzoriek do 32 bitov. Používa vlastný formát s príponou .ofr, podporuje ID3v1 a APEv2. Súčasťou OptimFROGu je i technológia hybridnej kompresie, autorom nazvaná **DualStream**. Korekčné súbory majú príponu .ofc. Odporúčaný dátový tok je okolo 330 kbps a vyššie.

OptimFROG používa vlastný zovšeobecnený koncept stereo dekorelácie, keď predikčná funkcia má dva rády (počet použitých vzoriek pri dekorelácii) – jeden pre svoj kanál, druhý pre združený kanál.

OptimFROG ponúka excelentný kompresný pomer (čím si získal svojich verných zákazníkov), no za cenu prehnanej časovej náročnosti kompresie i dekompresie. Aj keď nie je opensource, dosiahol značnú popularitu. Momentálne má širokú podporu na mnohých platformách. Z dôvodov ukončenia školy a zlého zdravotného stavu autora bol formát istý čas bez vývoja, čo sa, našťastie, zmenilo – najnovšia verzia (4.520b1) je z apríla 2006. V najbližších verziách sa očakáva podpora viackanálového zvuku a ďalšie zlepšenie kompresie. Domovskú stránku formátu možno nájsť na <http://www.losslessaudio.org>.

5.8 MPEG-4 ALS

Táto časť práce pojednáva o štandardizovanom formáte ALS. Stručne je spomenutý jeho predchodca, formát LPAC.

LPAC (Lossless Predictive Audio Compression) vznikol ako projekt Tilmana Liebchena z Berlínskej Technickej univerzity. Išlo o jednoduchý a výpočtovo nenáročný algoritmus využívajúci lineárnu predikciu, s podporou pre viaceré platformy. Keďže neponúkal výnimočné výsledky, nikdy sa skutočne neujal. Autor však zachytil výzvu konzorcia MPEG¹⁶ z júla 2002 na odoslanie návrhu na algoritmus bezstratovej kompresie. Spolu sedem firiem a inštitúcií poslalo svoj návrh, pričom konzorcium nakoniec zvolilo algoritmus LPAC ako referenčný model. Výhodu hralo akademické pozadie, výpočtová nenáročnosť (s dôrazom na veľmi ľahkú hardvérovú implementovateľnosť) a spomedzi konkurencie najlepšie výsledky kompresie. (Treba však zdôrazniť, že neboli porovnateľné s výsledkami nekomerčných algoritmov ako FLAC alebo Monkey's Audio.) Zo spomenutých dôvodov sa algoritmu LPAC ďalej nebudeme venovať, keďže výskum spôsobil jeho prerod v nový štandard – ALS.

ALS (Audio Lossless Coding) bol konzorciom MPEG štandardizovaný v júli 2005. Ide o symetrický kodek založený na lineárnej predikcii. Komplexnosť dekódera je však malá, preto časová náročnosť závisí len od zvoleného rádu predikcie (teoreticky najviac 1023). K jeho črtám patrí podpora ľubovoľného PCM vstupu do rozlíšenia 32 bitov, podpora IEEE float 32, vzorkovacej frekvencie do 192 kHz, možnosť kódovania viacerých kanálov a stôp až do počtu 65 536.

ALS využíva postupné zvyšovanie rádu predikcie na začiatku rámcu (progresívna predikcia) – kompromis medzi úspešnou kompresiou a dátovou nezávislosťou rámcov. Prvú vzorku (nepredikovateľnú) nekóduje, na druhú a tretiu používa osobitnú tabuľku Riceovho kódovania. Na ostatné vzorky do hodnoty rádu predikcie použije rovnakú tabuľku ako v prípade ostatných reziduálov. Pri kompresii zvuku vo formáte IEEE float 32 sa postupuje nasledovne: reálne číslo sa zaokrúhli na celé, napr. 16-bitové číslo, na ktoré sa aplikujú všetky štandardné postupy predikcie. Reziduál je však reálny, tzn. prenáša sa aj exponenciálna zložka.

Kanály sa predikujú nezávislo, alebo sa použije jeden z nich spolu s ich diferenciou, podľa toho, ktorá možnosť je výhodnejšia. (Pozn: kodér v prípade viackanálového zvuku môže kanály združovať do párov.) Prvé dva koeficienty prediktorov sa kvantujú osobitným spôsobom, ostatné použitím Riceových kódov. Pri kódovaní koeficientov prediktorov jedného kanála možno kódovať len rozdiel oproti druhému kanálu, alebo ich nekódovať vôbec a použiť koeficienty z druhého kanála, a tak ušetriť prenos niekoľkých dátových položiek (za cenu zhoršenia distribúcie reziduálneho signálu v tomto kanáli). Pri kódovaní reziduálov sa opätovne využíva korelácia kanálov, a to dvoma spôsobmi: aspoň jeden kanál sa kóduje priamo

16 P. Call for Proposals on MPEG-4 Lossless Audio Coding. 2002.

(tzv. referenčný kanál), druhý sa kóduje priamo, alebo sa využije ďalšia vrstva dopredne adaptívnej lineárnej predikcie, ktorá používa ako vstup reziduály z druhého kanála – predošlý, súčasný a nasledujúci, alebo sa k týmto pridáva ešte ďalšia trojica susedných reziduálov s vhodnou vzdialenosťou v čase (ALS tak podporuje aj dlhodobú predikciu reziduálu¹⁷).

Veľkosť rámca je rovnaká pre celý kódovaný vstup (2048 vzoriek pre vstup so vzorkovacou frekvenciou 48 kHz, 4096 vzoriek pre 96 kHz vstup), v prípade potreby ho však možno iteratívne deliť na bloky polovičnej veľkosti, pri ktorých možno znova rozhodovať o ďalšom delení atď. až do piatej úrovne (veľkosť bloku 32-krát menšia oproti veľkosti rámca) a v každom použiť inú tabuľku kódovania. Popri Riceových kódoch možno použiť aj kódovanie založené na BGMC (Block Gilbert-Moore Codes), ktoré navrhli vývojári z firmy Real Networks. V tomto prípade sa kódované hodnoty rozdeľia na štatisticky zriedkavé, ktoré sú kódované Riceovým kódovaním, a štatisticky dominantné – pri tých sa číslice ich zápisu s menšou hodnotou prenášajú priamo, číslice s väčšou hodnotou sú prenášané upraveným aritmetickým kódovaním. Tento postup je však výpočtovo náročnejší a zvyšuje komplexnosť dekodérov.

Formát ALS je konzorciom MPEG navrhnutý ako nový štandard v bezstratovej kompresii zvuku. Široký tím profesionálov, ktorý za ním stojí, aj priemyselné konzorciá – všetci iste uvítajú otvorený formát z dielni MPEG. Účinnosť kompresie je vynikajúca, rovnako je veľmi dobrá rýchlosť kompresie i dekompresie. Formát má teda dosť kvalít, aby sa stal najlepším algoritmom na bezstratovú kompresiu zvuku. Otáznym je len dopyt trhu – bezstratové formáty sú obľúbené najmä medzi nadšencami, no tí preferujú bezplatnosť. Ak sa ALS napriek licenčným poplatkom ujme, zrejme skôr či neskôr vytlačí v akceptovanosti trhom doterajšieho lídra, FLAC. Referenčný softvér (kodér, dekodér), rovnako aj dokumenty pojednávajúce o ALS možno nájsť na stránke <http://www.nue.tu-berlin.de/forschung/projekte/lossless/mp4als.html>.

17 P. časť 4.11 *Kompresia hovoreného slova*, str. 112.

5.9 MPEG-4 SLS

Táto časť práce pojednáva o vznikajúcom štandardizovanom formáte **SLS** (Scalable to lossless coding), vychádzajúcom z algoritmu AAZ.

AAZ (Advanced Audio Zip) vznikol ako projekt dvoch mladých singapurských výskumníkov (Rongshan Yu a Susanto Rahardja) z Institute of Infocomm Research (súčasť Agency for Science Technology and Research – A*STAR). Za svoj návrh získali v roku 2003 zlatú medailu v Tan Kah-Kee-ho cene pre mladých výskumníkov. V októbri 2002 konzorcium MPEG uverejnilo výzvu na návrh algoritmu umožňujúceho hybridnú stratovú kompresiu s vlastnosťou škálovateľnosti prenosovej šírky až do dosiahnutia bezstratovej kompresie. Dôvod bol jednoduchý: stratové kompresné schémy sú vhodné na nízky dátový tok, bezstratové schémy majú dátový tok veľmi vysoký. Škálovateľná schéma by bola medzistupňom, nástrojom na poskytovanie online služieb budúcnosti, spĺňajúc paradigmu *univerzálneho prístupu k multimédiám* (*Universal Multimedia Access – UMA*), ktorá má byť normou pre vznikajúce štandardy MPEG-7 a MPEG-21. Znamená to jedno: ponúknuť distribúciu ľubovoľných multimediálnych dát kdekoľvek a komukoľvek, bez ohľadu na rýchlosť jeho pripojenia do siete. Je zrejmé, že stratové kompresie sú tu výsostne neúčinné (nutnosť prítomnosti viacerých verzií materiálu pre rôzne šírky pásma na serveri distribútora), etablované bezstratové formáty však neumožňujú škálovateľnosť medzi stratovosťou a bezstratovosťou v reálnom čase. Autori AAZ na výzvu konzorcia MPEG zareagovali a v konkurencii bol ich návrh zvolený za referenčný model nového štandardu SLS.

AAZ je konštruovaný ako rozšírenie toku AAC¹⁸, čo ho robí kompatibilným so súčasnými normami MPEG-4. Základný tok dát je kódovaný štandardným spôsobom (stratovo cez AAC), pričom zložka „redukujúca“ jeho stratovosť je „pribalená“ použitím technológie *fine-granular-scalability* (FGS), ktorá má umožniť dynamicky meniť dátový tok po jednotkách menších než 0,4 kbps. Na dosiahnutie tejto vlastnosti používa AAZ tzv. bitový Golombov kód (bit-plane Golomb code, BPGC). Čím viac dát dekodér obdrží, tým bližšie budú dekodované dáta podobné originálu.

Treba uviesť, že požiadavky kladené na FGS sú pre súčasnú informatickú vedu vysoké. Napriek rokom výskumu v tejto problematike sa nepodarilo nájsť dostatočne efektívne kódovanie umožňujúce takéto škálovanie. V súčasnosti navrhované riešenia FGS sa vyznačujú vysokou komplexnosťou a nízkou efektívnosťou, spôsobujúc zvýšenie výsledného dátového toku v priemere o 20-30% oproti neškálovateľným riešeniam.

Vývoj SLS mešká, zdá sa, že ani dopyt trhu nekorešponduje s očakávaním. V súčasnosti nie je dostupná presná špecifikácia, jeho budúcnosť je neistá. Základný prehľad o formáte možno získať na http://www.itsc.org.sg/synthesis/2004/3_AAZ.pdf alebo <http://www.chiariglione.org/mpeg/technologies/mp04-sls>.

18 P. časť 4.3, str. 94.

5.10 Meridian Lossless Packing (MLP)

Formát **Meridian Lossless Packing (MLP)** vzišiel z dielne spoločnosti Meridian Audio, ktorá vyše 20 rokov vyrába spotrebiteľskú elektroniku vo sfére zvuku. Tento algoritmus bol navrhnutý pre potreby štandardu DVD-Audio, kde slúži na ukladanie bezstratového záznamu zvuku. Dôvod vzniku bol prozaický – uloženie šesťkanálovej nahrávky v plnom rozlíšení bez kompresie (96 kHz, 24 bitov) prekračuje prenosovú rýchlosť média asi o 50% a znižuje jeho kapacitu pod hodinu. MLP prišlo vhod.

MLP používa tri proprietárne technológie: *lossless processing and lossless matrixing*, kedy sa najprv na zvukovom zázname (rámce veľkosti typicky 40 až 160 vzoriek) vykoná bezstratová maticová transformácia za účelom odstránenia medzikanálovej korelácie. Nasleduje *lossless waveform prediction*, teda LPC. MLP má na výber dopredne aj spätne adaptívnu predikciu, oba rádu nanajvyš 8. Predikčné funkcie môžu byť pre jednotlivé kanály rôzne. Na konci procesu je *entropické kódovanie* (kodér ma na výber viac kódovaní, medziiným Riceovo). Prioritou pri vývoji MLP bolo obmedzenie špičkového dátového toku, ktorý nesmel prekročiť požadovanú hodnotu, samotná úspešnosť kompresie bola až na druhom mieste. Z toho dôvodu používa MLP aj vlastný návrh práce s vyrovnávacou pamäťou (buffer), bez ktorej by bola táto úloha príťažká. V nej združuje rámce do paketov (typickej veľkosti 640 až 2560 vzoriek – podľa nastavenia). V prípade neúspechu kompresie nejakého rámca v danom kanáli sa použije jeho zápis v pôvodnej PCM forme.

MLP podporuje najviac 63 kanálov (zvyčajné použitie je 5.1), vzorkovacie frekvencie od 32 kHz do 192 kHz (musí byť rovnaká pre všetky kanály, resp. jedna skupina kanálov môže mať vzorkovaciu frekvenciu dvakrát vyššiu než druhá skupina) a rozlíšenie kanálov 14 až 24 bitov (ľubovoľne pre každý kanál). Navyše umožňuje v dátovom toku zaznamenať prídavné informácie o kanáli, napr. jeho pôvod (mono, stereo zvukový zdroj, hierarchické triedenie kanálov, vlastník záznamu, ...). Pre používateľov so stereo reprodukčným systémom (či už na strane prehrávača alebo reproduktorov) môže výrobca disku do toku dať zahrnúť osobitný stereo záznam. Druhou možnosťou je ešte pri kódovaní definovať maticu na downmix signálu (môže sa v rámci záznamu meniť; zapíše sa do výsledného toku dát) – vtedy sú prvé dva kanály priamo zakódované ako želaný stereofonický záznam, zvyšné kanály (spolu s uvedenou maticou) slúžia na prípadnú rekonštrukciu šesťkanáloveho zvuku.

Technológia MLP bola na trh uvedená ešte v priebehu roka 1998, no jej rozšírenosť je dodnes nízka (korešpondujúc s nízkou popularitou DVD-Audio). Prvé domáce prehrávače na DVD-Audio sa totiž dostali na trh na sklonku roka 2000. Výroba takýchto zvukových záznamov je omnoho nákladnejšia než výroba zvukových CD pre bežných fanúšikov. Vysoké licenčné poplatky a utajenosť špecifikácie hrajú tiež dôležitú rolu. MLP je špecifická technológia (napr. k dispozícii

sú len hardvérové kodéry s cenami v tisícoch dolárov), ktorá by sa nemala porovnávať s inými bezstratovými schémami, predsa však svojím výkonom (rýchlosť a úspešnosť kompresie) nie je schopné konkurovať ostatným formátom spomenutým v tejto práci. Jeho nástupcom pre HD-DVD a BD je Dolby TrueHD – vylepšenie MLP pochádzajúce z dielne Dolby Laboratories. Stránka výrobcu je na http://www.meridian-audio.com/p_mlp_in.htm, bližší technický popis možno nájsť na http://www.meridian-audio.com/w_paper/mlp_jap_new.PDF.

6 POROVNANIE ZVUKOVÝCH FORMÁTOV

V predošlých dvoch kapitolách sme uviedli množstvo rôznych zvukových formátov a kompresných algoritmov. Používateľ si ľahko môže vybrať medzi univerzálnymi stratovými formátmi¹ (signifikantné zmenšenie dátového toku za cenu zníženia objektívnej, nie však nutne subjektívnej kvality), stratovými formátmi na kompresiu hovoreného slova² (radikálne zníženie dátového toku a latencie za cenu mierneho alebo značného zníženia kvality) a bezstratovými formátmi³ (mierne zníženie dátového toku pri zachovaní možnosti presne rekonštruovať pôvodné zvukové dáta). Ich popis obsahuje aj vymenovanie výhod a nevýhod oproti iným formátom. Nie je však jednoduché povedať, ktorý z nich je *najlepší* alebo *najvhodnejší*. Všetko totiž závisí od uhla pohľadu – od požiadaviek konkrétneho zákazníka. Nasledujúci text sa pokúša vymenovať tie najdôležitejšie kritériá (pozn: ich poradie neodráža ich relatívnu dôležitosť). Treba ešte podotknúť, že jednotlivé aspekty sa navzájom ovplyvňujú (napr. jednoduchý návrh uľahčuje hardvérovú podporu atď.).

Prvým kritériom je **stupeň štandardizácie**. Samotný formát môže podávať veľmi dobré výsledky, keď však nie je žiadna garancia stability jeho špecifikácie (či už ide o dekodér alebo tok dát), je často nevhodné investovať doň svoju dôveru. Najväčšiu výhodu tu majú medzinárodné štandardy (MP3, AAC, ALS). Istotu ponúkajú aj štandardizované alebo proprietárne⁴ komerčné riešenia – prvé čerpajú zo svojej dlhoročnej prítomnosti na trhu (DD, DTS...), druhé z marketingovej i odbornej sily firiem, ktoré za nimi stoja (WMA). Mierna neistota vládne na poli ostatných projektov. Opensource formáty ponúkajú širokú základňu programátorov ochotných zdokonaľovať softvér alebo rozširovať jeho podporu pre iné prostredia, tí to však robia len ako koníček vo voľnom čase – závisí od konkrétneho projektu, ako rýchlo sa odstraňujú chyby, ako verne sa dodržiava špecifikácia atď. V prípade projektov jednotlivcov a malých firiem je táto neistota najväčšia a už niekoľkokrát sa stalo, že špecifikácia sa dočkala nečakanej zmeny (MA).

Druhým kritériom je **softvérová podpora**. Tvorcovia (v prípade opensource projektov preberá túto zodpovednosť často komunita priaznivcov) sa musia postarať o dobré softvérové zázemie (novému) formátu. Dôležitá je nielen existencia kodéra a dekodéra pre viac platforiem, ale najmä podpora dekódovania v už jestvujúcich prehrávačoch, zvukových editoroch, tzv. programoch na vypaľovanie médií atď.

Tretím kritériom je **výpočtová zložitosť dekódovania**. Algoritmus musí byť navrhnutý tak, aby nebol problém používať ho pri bežnej práci na súčasných osobných počítačoch. Vysoká záťaž procesora (s pamäťou v súčasnosti nie sú také

1 P. kapitolu 4 *Prehľad stratových zvukových formátov* (str. 87) a jej podčasti.

2 P. časť 4.11 *Kompresia hovoreného slova* (str. 112).

3 P. kapitolu 5 *Prehľad bezstratových zvukových formátov* (str. 115) a jej podčasti.

4 V tomto kontexte myslíme *utajené*.

problémy), dlhé čakanie pri pretáčaní, prerušenie prehrávania pri prechode medzi skladbami (opak tzv. gapless playback) komplikujú každodenné používanie takýchto formátov (napr. MA, LA, OptimFROG). Často sú vhodnejšie formáty spoliehajúce sa na to, že kódovanie sa vykoná len raz, dekodovanie mnohokrát, preto je výhodnejšie stavať na asymetrickom prístupe⁵.

Štvrtým kritériom môže byť **hardvérová podpora**, ktorá sa veľmi odvíja od výpočtovej zložitosti dekodovania. Mnohí používatelia chcú mať svoje piesne vždy so sebou (prenosné prehrávače), a investujú len do formátu, ktorý je podporovaný ich obľúbeným prehrávačom. Podpora v stolových prehrávačoch tiež uľahčí rozhodnutie pre ten-ktorý formát.

Piatym kritériom je **výpočtová zložitosť kódovania**. Musí sa pohybovať v rozumných medziach, ináč formát nebude úspešný natoľko, ako by si možno zaslúžil vzhľadom napr. na úspešnosť kompresie (LA, OptimFROG).

Šiestym kritériom sú rôzne vlastnosti, ktoré môže používateľ zvážiť: Opensource alebo proprietárne riešenie (tento faktor hrá často rozhodujúcu úlohu), podpora pre vysielateľnosť (streaming), detekcia chýb, podpora pre vysoké rozlíšenie zvuku alebo viackanálový zvuk atď. Dôležitá je i rozšírenosť formátu a všeobecné povedomie o ňom (aj keď to by v ideálnom prípade malo byť výsledkom ostatných kritérií, predsa vplýva na ďalší úspech formátu).

Siedmym, pre mnohých najdôležitejším kritériom je **úspešnosť kompresie** (vyjadrený dátovým tokom). Tento termín sa možno ľahko kvantifikuje pre bezstratové formáty (priemer pre rôznorodé vstupy). Pre stratové formáty sa však nezaobídeme bez delenia na kategórie. Štandardným spôsobom porovnávaní sú **posluchové testy** v jednotlivých kategóriách.

Táto kapitola sa delí na dve časti: 6.1 *Porovnanie stratových kompresných schém* (str. 135; v nej bližšie objasníme aj spomenuté posluchové testy) a 6.2 *Porovnanie bezstratových kompresných schém* (str. 139).

⁵ P. kapitolu 5 *Prehľad bezstratových zvukových formátov*, str. 115.

6.1 Porovnanie stratových kompresných schém

Ako sme prv spomínali, je náročné kvantifikovať všetky vlastnosti jednotlivých formátov. Na poli stratových kompresných schém je mnoho etablovaných štandardných konkurentov, majú veľmi dobrú softvérovú podporu a niektoré i dobrú hardvérovú podporu. Výpočtová zložitosť transformačných kódovaní je vo všeobecnosti nízka, dokonca sa dá povedať, že je inherentnou vlastnosťou danej schémy a je nemenná, bez ohľadu na zvolený dátový tok či náročnosť skladby⁶.

Siedme kritérium spomenuté v predošlom texte (úspešnosť kompresie), vzhľadom na stratovosť návrhov, možno len ťažko definovať. V prvom rade si treba uvedomiť, že neexistuje objektívna miera – filozofia stratových formátov je založená na odstraňovaní psychoakustickej irelevancie, teda jediný, kto môže posúdiť ich úspešnosť, je poslucháč. **Posluchový test** je jediný spôsob ako rozhodnúť, ktorý formát má „lepšiu kompresiu“.

Ideálne by bolo vedieť vytvoriť škálu kvalít zvukových záznamov a pre jednotlivé algoritmy zistiť, pri akom dátovom toku dosiahnu zvolenú kvalitu. Takýto prístup by však bol veľmi náročný, preto sa používa iný: zvolia sa kategórie, vyjadrené typickým (resp. priemerným) dátovým tokom. Tieto kategórie odzrkadľujú rôzne spôsoby použitia formátu, napr. 40 kbps pre nekvalitné rozhlasové vysielanie, 64 kbps pre kvalitnejšie rozhlasové vysielanie, 80 a 96 kbps pre poslech na prenosných zariadeniach, 128 kbps pre počúvanie na počítači. Len málokedy sa použijú ďalšie kategórie (160 kbps), keďže pri nich už len zriedkavo badať rozdiel medzi originálom a stratovo skomprimovaným záznamom (tzv. priesvitnosť kódovania). Formáty v jednej kategórii sa potom navzájom porovnávajú podľa subjektívnej kvality.

Posluchové testy možno vykonávať na profesionálnej úrovni: starostlivo sa zvolia zvukové záznamy, vyberú vhodní poslucháči a umiestnia sa do vhodnej miestnosti s profesionálnymi slúchadlami na ušiach. Nasleduje poslech, po ňom rozbor výsledkov. Metódy na ich vyhodnotenie sú navrhnuté expertmi. Príkladom profesionálneho testu môže byť napr. verifikácia úspešnosti kódovacej schémy AAC vypracovaná pre MPEG (<http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w2006.pdf>), alebo posluchový test zvukových formátov pri nízkych dátových tokoch, vypracovaný inštitúciou EBU (http://www.ebu.ch/CMSimages/en/tec_doc_t3296_tcm6-10497.pdf).

Častejšie však iniciujú rôzne posluchové testy „laici“. Zostavia vlastnú zbierku záznamov a umiestnia ju na svoju internetovú lokalitu, aby záujemci otestovali zvolené formáty a odoslali späť výsledky. Testovanie prebieha jedným zo

⁶ Výnimkou je azda len Ogg Vorbis, ktorý zo svojho návrhu podporuje možnú „asymetrickosť“ schémy – dlhý čas kódovania za účelom zmenšenia dátového toku, no bez zmeny času potrebného na dekódovanie – aj keď v súčasnosti pre tento prístup nejestvuje softvérová podpora.

štandardných spôsobov, pochádzajúcich zo sféry profesionálnych posluchovéch testov. Najrozšírenejšia je metóda **ABC/HR**. Pri každom zázname („vzorke“) a testovanom formáte má používateľ možnosť prehrať zvuk A, B alebo originál. Jeden zo zvukov A a B je pôvodný záznam (originál), druhý je zakódovaný. Používateľ si zvolí, ktorý z nich (A/B) podľa jeho mienky bol podrobený stratovej kompresii a aká je jeho kvalita. Tú vyjadruje na stupnici 1 (najhoršia) až 5 (neodlíšiteľné od originálu). K použitým formátom sa často dodá tzv. dolná kotva (reprezentujúca veľmi nekvalitný záznam originálu), v prípade testu pri nízkych dátových tokoch aj horná kotva (reprezentujúca veľmi kvalitnú kompresiu), aby zvolené hodnotenie (1-5) viac reprezentovalo objektívnu mierku, nie relatívne zoradenie formátov. Výsledky sa zakreslia do grafu alebo vyjadria číselne ako priemerná odpoveď a štatistická odchýlka pre ten-ktorý formát.

Pre nízke dátové toky sa často používa systém **MUSHRA**, kde poslucháč jednoducho ohodnotí záznam známkou, pričom si je vedomý, že ide o komprimovaný formát (len nevie, ktorý, keďže ich je v teste viac). Možno sa stretnúť aj s **ABX** testmi, tie však vykonávajú najmä jednotlivci, aby našli dátový tok, kedy zvolený formát dosahuje pre daný súbor skladieb priesvitnosť. Princíp spočíva v prehraní pôvodného záznamu, zakódovanej verzie a neznámeho záznamu X, ktorý je A alebo B. Úlohou poslucháča je zistiť identitu záznamu X.

Na internete možno nájsť niekoľko posluchovéch testov pochádzajúcich od „nadšencov“ (teda nie profesionálne testy). Základný úvod do techniky posluchovéch testov, všeobecné rady a niekoľko starých testov (všetko z rokov 2000 až 2002) možno nájsť na stránke používateľa menom ff123 (<http://ff123.net>). Internetová stránka Brazílčana menom Roberto José de Amorim (<http://www.rjamorim.com/test>) obsahuje testy vykonané v rokoch 2003 a 2004 pre rôzne formáty a dátové toky. Rumun Sebastian Mares (<http://www.maresweb.de/listening-tests>) inicioval posluchový test pre 128 kbps pred niekoľkými mesiacmi (2005-12). Ďalším zdrojom výsledkov (aj keď použitá metodika je otázna) je poloprofesionálna stránka <http://www.soundexpert.info>, na ktorej sú vždy aktuálne výsledky, navyše ich môžu obohatiť návštevníci stránky krátkym posluhom.

Uvádzame tabuľku (tab. 6.1, str. 138) poskytujúcu prehľad rôznych vlastností a črt stratových zvukových formátov. Keď nejaký údaj nebol k dispozícii, alebo je otáznym, uvedený je otáznik. Hlavným zdrojom informácií boli spomenuté internetové stránky a stránka http://en.wikipedia.org/wiki/Comparison_of_audio_codecs, o ktorej aktuálnosť sa starajú mnohí jej návštevníci. Treba uviesť, že zistiť všetky vlastnosti a vykonať „objektívne“ ohodnotenie (stĺpce 3-10) formátov pre daný dátový tok je vo svojej podstate nemožné – posluchové testy sa vykonávajú zriedka a testujú sa len malé množiny formátov, zvukových záznamov a dátových tokov, ktoré majú nejakú relevantnú výpoveď. Preto na mnohých miestach známky chýbajú a aj tie udelené sú

len aproximáciou relatívnej kvality formátu v uplynulom období (keďže mnoho formátov sa neustále vyvíja a zlepšuje – najlepšími príkladmi sú Ogg Vorbis a AAC). Popis jednotlivých položiek tabuľky:

- V prvom stĺpci sú vymenované jednotlivé formáty.
- V druhom stĺpci – **LC** – sa uvádza typ licencie. *OS* znamená opensource, teda je k dispozícii zdrojový kód kodéra a dekodéra. *OSI* znamená, že je táto licencia certifikovaná konzorciom Opensource Initiative (<http://www.opensource.org>), teda zaručene spĺňa ducha opensource hnutia. *ŠT* znamená štandardizovanú (ISO) technológiu, *PR* vyjadruje proprietárnu, teda utajenú technológiu. V zátvorke môže byť uvedené (*L*), čo znamená licencované použitie, tzn. za (komerčné) používanie platformy (zväčša kodéra) sa platí. (Poznámka: MPC má utajený kodér, dekodér je však opensource.)
- Tretí stĺpec – **32** – uvádza známkou priemerné ohodnotenie kvality v porovnaní s ostatnými formátmi v tejto kategórii. 1 znamená vysokú kvalitu, 5 nízku. Použitý dátový tok je 32 kbps (alebo približne 32 kbps v prípade VBR kódovania). Tieto údaje (rovnako aj údaje v stĺpcoch 48, 64, 80, 96, 105, 128, 256) sú len štatistickým odhadom, založeným na mnohých posluchovejch testoch jednotlivcov. Treba podotknúť, že toto hodnotenie nekorešponduje s „objektívnym“ hodnotením kvality známkou 1 až 5, ako sme si ho spomínali vyššie (odstavec o ABC/HR).
- Štvrtý až desiaty stĺpec – **48, 64, 80, 96, 105, 128, 256** – sú totožné s tretím stĺpcom, až na použitý dátový tok (vyjadrený názvom stĺpca).
- Jedenásty stĺpec – **S** – vyjadruje známkou 1-5 dostupnú softvérovú podporu.
- Dvanásty stĺpec – **H** – vyjadruje známkou 1-5 dostupnú hardvérovú podporu (zväčša prenosné prehrávače).
- Trinásty stĺpec – **V** – vyjadruje podporu formátu pre vysielateľnosť (angl. *streaming*): *A* znamená podporu, prázdne pole neprítomnosť podpory.
- Štrnásty stĺpec – **E** – indikuje prítomnosť technológií na detekciu chyby v toku dát: *A* znamená podporu, prázdne pole neprítomnosť podpory.
- Pätnásty stĺpec – **P** – vyjadruje podporu pretáčania (angl. *seek*) v súbore, resp. obnovenia dekódovania v prípade napojenia sa na prebiehajúci tok dát. *A* znamená podporu, prázdne pole neprítomnosť podpory. (Poznámka: Musepack má zvyčajne veľmi riedku tabuľku ukazovateľov (*seek table*), preto je pretáčanie výpočtovo náročné. Ogg Vorbis potrebuje dlhší čas na napojenie sa na vysielaný tok dát, keďže čaká na hlavičky toku dát, bez ktorých nemôže začať s dekompresiou.)
- Šestnásty stĺpec – **OS** – je zoznam operačných systémov, v ktorých má formát softvérovú podporu. Uvedený je v skratkách: *Win* znamená operačné systémy typu MS Windows, *Lin* systémy na báze Linuxu, *Mac* operačný systém Mac OS X. V prípade podpory ďalších, tu nespomenutých platforiem je uvedené +.

(Poznámka: technológie DD a CAC sa nezvyknú priamo porovnávať s konkurenciou, keďže ich použitie je špecifické a sú optimalizované na kódovanie priestorového zvuku. Podobne formát Speex – je určený na kompresiu hovoreného slova.)

	LC	32	48	64	80	96	105	128	256	S	H	V	E	P	OS
MP3	ŠT (L)	3		3	3	2		2	2	1	1	A		A	Win, Lin, Mac, +
MP3pro	ŠT (L)	1	1	1	1	2	2			1	?	A		A	Win, Lin, Mac, +
AAC	ŠT (L)	2	2	2	2	2	2	1	1	1	1	A	A	A	Win, Lin, Mac, +
HE-AAC	ŠT (L)	1		1	2	2				1	1	A	A	A	Win, Lin, Mac, +
WMA	PR (L)	2	3	3	3	2		3	3	1	2	A	A	A	Win, Mac
WMA Pro	PR (L)						1	1		2	5	A	A	A	Win, Mac
RA	PR (L)	3		4			3			1	(3)	A	?	A	Win, Lin, Mac
ATRAC	PR (L)		3	3			4	4	1	4	4	?	?	A	Win, Lin, Mac, +
DD (AC-3)	ŠT (L)														Win, Lin, Mac, +
CAC (DTS)	ŠT (L)														Win, Lin, Mac, +
Ogg Vorbis	OSI	1		2			2	1	1	1	2	A	A	A	Win, Lin, Mac, +
MPC	PR/OS							1	1	2	3	A		A	Win, Lin, Mac, +
Speex	OSI									2	5	A		A	Win, Lin, Mac, +

Tab. 6.1 – Prehľad vlastností vybraných formátov na stratovú kompresiu zvuku

Ťažko určovať víťaza, najmä preto, že každý poslucháč má „iné uši“. Sú ľudia, ktorí nevedia odlíšiť originál od súboru MP3 pri 96 kbps (pri použití kodeku LAME). Nájdu sa však aj takí, pre ktorých je MP3 (LAME) pri 192 kbps ľahko odlišiteľný od originálu. Najlepšie je teda vybrať podľa osobného sluchu, požiadaviek atď. Čo sa týka kvality pri dátových tokoch 96-192 kbps, zrejme treba odporučiť jeden z formátov MP3 (LAME), Ogg Vorbis (AuTuV), AAC (iTunes a NERO) alebo MPC. To je však aktuálny stav – všeobecne sa predpokladá, že AAC má najväčšiu možnosť zlepšovania a doladzovania kódérov, Ogg Vorbis je na tom svojou konštrukciou veľmi podobne. MP3 a MPC sú na tom kvalitou dobre (napriek zastaranosti návrhov samotných formátov – dôvodom je len veľmi kvalitné doladenie kódérov), no ich ďalšie zlepšovanie sa už nepredpokladá. Z pohľadu komerčného využitia má najväčšie šance na úspech formát AAC.

6.2 Porovnanie bezstratových kompresných schém

V prípade bezstratových formátov, ako bolo spomenuté, nie je zložité kvantifikovať úspešnosť kompresie. Netreba žiadne posluchové testy, stačí ochota, počítačové vybavenie a voľný čas. Na internete možno nájsť pár príkladov. Dobré, aj keď nie celkom aktuálne (2004-05) porovnanie rýchlosti kompresie, rýchlosti dekompresie a úspešnosti mnohých formátov možno nájsť zakreslené v grafe na stránke Holanďana menom Hans Heijden (<http://web.inter.nl.net/users/hvdh/lossless/lossless.htm>), podobnú prácu z 2005-02, no vo forme tabuľky, možno nájsť na stránke používateľa s prezývkou Speek (<http://members.home.nl/w.speek/comparison.htm>)

Najjednoduchšie metódy (napr. Shorten) vedú ku kompresii okolo 63% (pomer veľkostí skomprimovaných a pôvodných dát) a nie sú výpočtovo veľmi zložité. Výsledky najúspešnejších kompresných schém sa však vždy pohybujú nad 50% (štatistika pre rôznorodé vstupy). Rozdiely v úspešnosti kompresie sú teda nízke. Na poli bezstratových formátov prevláda jediná metóda – LPC. Každé zlepšenie úspešnosti kompresie sa získava za cenu zvýšenia komplexnosti kodéra (asymetrické riešenie) alebo kodéra i dekodéra (symetrické riešenie). Nárast výpočtovej náročnosti je však neúmerne zmenšeniu dátového toku.

Rýchlosť dekódovania je bezpochyby dôležitejšia než rýchlosť kódovania, ktorá sa vykonáva len raz⁷. Najlepšie kompresné pomery síce dosahujú spätne adaptívne kodéry, tie sú však symetrické, teda ich dekódovanie je rádovo rovnako náročné ako kódovanie. Nie je teda jednoduché zvoliť jasného víťaza medzi bezstratovými formátmi.

Uvádzame tabuľku (tab. 6.2) poskytujúcu prehľad rôznych vlastností a čít bezstratových zvukových formátov. Keď nejaký údaj nebol k dispozícii, alebo je otáznym, uvedený je otáznik. Hlavným zdrojom informácií bola internetová stránka http://wiki.hydrogenaudio.org/index.php?title=Lossless_comparison. Jej autor (Roberto José de Amorim) sa neustále stará o jej aktuálnosť.

- V prvom stĺpci sú vymenované jednotlivé formáty.
- V druhom stĺpci – **LC** – sa uvádza typ licencie. *OS* znamená opensource, teda je k dispozícii zdrojový kód kodéra a dekodéra. *OSI* znamená, že je táto licencia certifikovaná konzorciom Opensource Initiative (<http://www.opensource.org>), teda zaručene spĺňa ducha opensource hnutia. *ŠT* znamená štandardizovanú (ISO) technológiu, *PR* vyjadruje proprietárnu, teda utajenú technológiu. V zátvorke môže byť uvedené bližšie upresnenie – (*L*) znamená licencované použitie, tzn. za komerčné používanie platformy (zväčša kodéra) sa platí, a (*S*) vyjadruje

⁷ Dôvod je prostý – vďaka bezstratovosti riešenia nie je potrebné mať záznam vo viacerých verziách (iné dátové toky). Nutnosť používať viac druhov formátov je zriedkavá, pri bezstratových formátoch ju možno navyše transkódovanie (zmena jedného formátu na iný) vykonať bezstratovo (teda bez tzv. generačnej straty).

- dostupnosť knižnice na implementáciu kodeka v iných produktoch.
- Tretí stĺpec – **CMP** – uvádza známkou priemernú úspešnosť kompresie. 1 znamená vysoká úspešnosť, 5 nízka. Tieto údaje (rovnako aj údaje vo štvrtom a piatom stĺpci) sú len štatistickým odhadom, založeným na mnohých porovnaníach a testoch vykonaných na rôznych materiáloch. Prvé číslo vyjadruje „optimálne“ nastavenie kodéra (väčšinou mód „normal“), znamenajúce kompromis medzi rýchlosťou a úspešnosťou kompresie. V prípade že sú k dispozícii výsledky pre rôzne konfigurácie kodéra, najlepší dosiahnuteľný výsledok (za cenu spomalenia kompresie a/alebo dekompresie) je uvedený ako prvé číslo v zátvorke, druhé číslo v zátvorke je zas výsledok pri zvolení najvyššej rýchlosti kompresie. Treba podotknúť, že rozdiel medzi známkou 1 a 5 nie je v objektívnych číslach veľký – výsledky sa v priemere pohybujú okolo 53-63%.
 - Štvrtý stĺpec – **C** – vyjadruje známkou 1-5 ohodnotenie rýchlosti kompresie pre „optimálne“ nastavenie, v zátvorke sú uvedené rýchlosti pre najsilnejšiu a najrýchlejšiu kompresiu (p. tretí stĺpec).
 - Piaty stĺpec – **D** – vyjadruje známkou 1-5 ohodnotenie rýchlosti dekompresie pre „optimálne“ nastavenie, v zátvorke sú uvedené rýchlosti pre najsilnejšiu a najrýchlejšiu kompresiu (p. tretí stĺpec).
 - Šiesty stĺpec – **S** – vyjadruje známkou 1-5 dostupnú softvérovú podporu.
 - Siedmy stĺpec – **H** – vyjadruje známkou 1-5 dostupnú hardvérovú podporu (zväčša prenosné prehrávače).
 - Ôsmy stĺpec – **V** – vyjadruje podporu formátu pre vysielateľnosť (angl. *streaming*): *A* znamená podporu, prázdne pole neprítomnosť podpory.
 - Deviaty stĺpec – **E** – indikuje prítomnosť technológií na detekciu chyby v toku dát: *A* znamená podporu, prázdne pole neprítomnosť podpory.
 - Desiaty stĺpec – **P** – vyjadruje podporu pretáčania (angl. *seek*) v súbore, resp. obnovenia dekódovania v prípade napojenia sa na prebiehajúci tok dát. *A* znamená dobrú podporu, *S* slabšiu (pomalé pretáčanie z dôvodu dlhej doby dekódovania alebo dlhých rámcov), *N* žiadnu.
 - Jedenásty stĺpec – **OS** – je zoznam operačných systémov, v ktorých má formát softvérovú podporu. Uvedený je v skratkách: *Win* znamená operačné systémy typu MS Windows, *Lin* systémy na báze Linuxu, *Mac* operačný systém Mac OS X. V prípade podpory ďalších, tu nespomenutých platforiem je uvedené +.

(Poznámka: formát ALS bol len nedávno štandardizovaný, SLS je ešte len vo štádiu vývoja. MLP je zas technológia so špecifickým využitím. Z toho dôvodu sa mnohé ich vlastnosti neuvádzajú, alebo uvádzajú v zátvorke.)

	LC	CMP	C	D	S	H	V	E	P	OS
Shorten	OS (N)	5	1	1	2	5			A	Win, Lin, Mac, +
WavPack	OSI	3 (2, 4)	1 (4, 1)	1 (2, 1)	1	3	A	A	A	Win, Lin, Mac
MA	OS (N)	2 (1, 3)	2 (4, 1)	3 (4, 2)	2	5	A	A	S	MS, (Lin?)
FLAC	OSI	3 (3, 5)	2 (4, 1)	1	1	2	A	A	A	Win, Lin, Mac, +
LA	PR	1(1)	5	5	3	5	?		A	Win, Lin
TTA	OS	3	1	2	3	4		A	S	Win, Lin, Mac
OptimFROG	PR (S)	1 (1, 2)	4 (5, 3)	4 (5, 3)	3	5	A	A	A	Win, Lin, Mac
LPAC	PR	3	2 (3, 2)	2	4	5		A	S	Win, Lin, +
ALS	ŠT (L?)	?	?	?	(5)	(5)	A	A	A	?
SLS	ŠT (L?)	?	?	?	(5)	(5)	(A)	?	A	?
MLP	PR (L)	?	?	?	(4)	(4)	?	?	?	?
ALAC	PR (L) OS dekodér	3	2	2	4	3	A		A	Win, Mac
RKAU	PR	3 (3, 4)	4 (5, 3)	4 (5, 4)	4	5			A	Win
WMAL	PR (L)	2	3	3	4	5	A	A	A	Win, Mac
RAL		?	2	3	4	5	A	?	A	Win, Lin, Mac

Tab. 6.2 – Prehľad vlastností vybraných formátov na bezstratovú kompresiu zvuku

Ťažko určovať víťaza, no stanovené kritériá najlepšie spĺňajú WavPack, FLAC a TTA. FLAC má spomedzi nich najrýchlejšiu dekompresiu, najväčšiu rozšírenosť a základňu programátorov, WavPack zas pri nižšej náročnosti kompresie podáva trochu lepšie výsledky. Hlavnou nevýhodou TTA je slabá rozšírenosť tohto formátu. Keďže rozdiely v úspešnosti bezstratových algoritmov nie sú veľké, v rýchlosti a rozšírenosti sa zatiaľ nič nevyrovná formátu FLAC, aj keď mu môže byť neskôr veľkým konkurentom štandardizovaný formát MPEG-4 ALS.

7 ZÁVER

Práca *Úvod do metód spracovania zvuku v súčasnom multimediálnom prostredí* je nielen stručnou internetovou encyklopédiou o zvuku, ale najmä ucelenou učebnicou tejto problematiky, o ktorej nie je jednoduché získať aktuálne informácie v slovenčine. Veríme, že predstavuje prínos v oblasti rozvoja e-learningových aktivít na Slovensku.

V kapitole 2 sme poskytli náhľad do problematiky vlastností zvuku z pohľadu fyziky aj poslucháča.

V kapitole 3 sme sa venovali digitalizácii zvuku, jeho kódovaniu a metódam kompresie dát popisujúcich zvuk – stratovej (psychoakustickej) kompresii i kompresii bezstratovej. Na dôvažok sme poskytli náhľad na aktuálny stav v oblasti práce so zvukom vo sfére osobných počítačov.

Záverečné kapitoly ponúkli prehľad stratových (kapitola 4) i bezstratových (kapitola 5) zvukových formátov a ich porovnanie (kapitola 6).

Práca sa nachádza na internetovej stránke <http://zvuk.atrip.sk>.

8 ZOZNAM POUŽITEJ LITERATÚRY

- ATSC. *Digital Audio Compression Standard (AC-3, E-AC-3) : Document A/52B* [online]. Rev. B. 2005. Washington, D.C. : Advanced Television Systems Committee, Inc., 1994-11-10. Rev. B: 2005-06-14 [cit. 2006-04-19]. Dostupné na internete: <http://www.atsc.org/standards/a_52b.pdf>.
- Call for Proposals on MPEG-4 Lossless Audio Coding* [online]. Klagenfurt, AT : MPEG, 2002-07. Dostupné napr. na <<http://www.labs.nec.co.jp/mpeg/audio/public/w5040.pdf>>.
- GOLOMB, Solomon Wolf. Run-Length Encodings. In *IEEE Transactions on Information Theory*, 1966-09, Vol. 12, pp. 399–401. Dostupné na internete napr. na: <http://urchin.earth.li/~twic/Golombs_Original_Paper/Golomb1966.pdf>.
- HALUŠKA, Ján. *The Mathematical Theory of Tone Systems*. New York : Marcel Dekker, Inc. / Bratislava: Ister Science Ltd, 2004. ISBN 0-8247-4714-3 (Marcel Dekker, Inc.), resp. ISBN 80-88683-28-9 (Ister Science).
- HAMM, Russel O. Tubes vs. Transistors: Is There an Audible Difference? In *Journal of the Audio Engineering Society (AES), 43rd Convention*. New York : AES, 1972-09-14. Dostupné na internete: <<http://www.milbert.com/tstxt.htm>>.
- HERRE, Jürgen. *Temporal Noise Shaping, Quantization and Coding Methods in Perceptual Audio Coding: A Tutorial Introduction* [online]. Príspevok na konferencii AES 17th Conference High Quality Audio Coding, 1999-09-05 [cit. 2006-04-08]. Dostupné napr. na: <[http://www.ee.columbia.edu/~marios/courses/e6820y02/project/papers/Temporal Noise Shaping Quantization and Coding Methods in Perceptual Audio Coding - A Tutorial Introduction.pdf](http://www.ee.columbia.edu/~marios/courses/e6820y02/project/papers/Temporal%20Noise%20Shaping%20Quantization%20and%20Coding%20Methods%20in%20Perceptual%20Audio%20Coding%20-%20A%20Tutorial%20Introduction.pdf)>.
- HUFFMAN, David A. A Method for the Construction of Minimum-Redundancy Codes. In *Proceedings of the IRE*, 1952-09, Vol. 40, pp. 1098-1101. Dostupné na internete napr. na: <http://compression.ru/download/articles/huff/huffman_1952_minimum-redundancy-codes.pdf>.
- IMMINK, Kees A. Schouhamer. The Compact Disc Story. 103rd Convention of AES. NY, USA. 1997. In *Journal of the Audio Engineering Society (AES), vol. 40, no. 5*. New York : AES, 1998-05. Dostupné na internete: <<http://www.exp-math.uni-essen.de/~immink/pdf/cdstory.pdf>>.
- LIEBCHEN, Tilman – PURAT, Marcus – Noll, Peter. *Lossless Transform Coding of Audio Signals*. Berlin : Technical University of Berlin, 1997. (Článok na 102. AES Convention, Mníchov.) Dostupné na internete: <<http://www.nue.tu-berlin.de/wer/liebchen/docs/aes102.pdf>>.
- LIEBCHEN, Tilman. *Realisierung einer verlustlosen Transformationscodierung zur Datenkompression von Mono- und Stereo-Audiosignalen*. Berlin : Technische Universität Berlin, 1998-07-02. Dostupné na internete: <<http://www.nue.tu-berlin.de/wer/liebchen/docs/Diplom.pdf>>.
- MACKAY, David J. C. *Information Theory, Inference, and Learning Algorithms*. Version 7.2. 2005-03-28 [cit. 2006-03-03]. Cambridge, UK : Cambridge University Press. Aktuálna verzia dostupná na internete: <<http://www.inference.phy.cam.ac.uk/itprnn/book.pdf>>.
- MARTIN, G. N. N. *Range Encoding: An Algorithm for Removing Redundancy from a Digitised Message* [online]. 1997-11 [cit 2006-03-06]. (Pôvodný rukopis 1979-03). Dostupné na internete: <<http://www.compressconsult.com/rangecoder/rngcod.pdf.gz>>.

- MELTZER, Stefan – MOSER, Gerald. *MPEG-4 HE-AAC v2 – Audio Coding for Today's Digital Media World*. Ebu Technical Review 2006-01 [online] [cit. 2006-04-08]. Dostupné na internete: <http://www.ebu.ch/en/technical/trev/trev_305-moser.pdf>.
- MORIYA, Takehiro – YANG, Dai Tracy – LIEBCHEN, Tilman. *Extended Linear Prediction Tools for Lossless Audio Coding*. IEEE ICASSP 2004. Montreal, 2004-05 [cit. 2006-03-10]. Dostupné napríklad na internetovej stránke: <http://www.nue.tu-berlin.de/Publikationen/papers/0301008_ICASSP2004_Moriya_etal.pdf>.
- NEUBAUER, Tomáš. *Komprimované formáty v praxi*. In: Muzikus 2004/01. Praha : Nakladateľství Muzikus, 2004. ISSN 1210-1443.
- OLEJÁR, Daniel – STANEK, Martin. *Úvod do teórie kódovania* [online]. 2006-01-27 [cit. 2006-03-03]. Aktuálna verzia dostupná na internete: <<http://www.dcs.fmph.uniba.sk/texty/codebook.ps>>.
- OLSON, Harry F. *Accoustical Engineering*. Cit. In Wikipedia. *Sound* [online článok]. Verzia z 2006-02-09 [cit. 2006-02-13]. Dostupné na internete: <<http://en.wikipedia.org/wiki/Sound>>.
- SCHADLER, Ted et al. *The Future of Digital Audio*. [online] USA : Forrester Research, 2005-03-21 [cit. 2006-03-23]. Dostupné na internete pre registrovaných používateľov: <<http://www.forrester.com/Research/Document/0,7211,36428,00.html>>.
- SIBBALD, Alastair (Sensaura Ltd). *An Introduction to Sound and Hearing* [online]. Middlesex, UK : Sensaura Ltd, 2001 [cit. 2006-02-24]. Dostupné na internetovej stránke: <<http://www.sensaura.co.uk/whitepapers/pdfs/dev005.pdf>>.
- SPORER, Th., BRANDENBURG, Kh., EDLER, B. *The Use of Multirate Filter Banks for Coding of High Quality Digital Audio*. 6th European Signal Processing Conference. Amsterdam, 1992-06 [cit. 2006-04-24]. Dostupné na internete napr. na: <http://www.tde.lth.se/ugradcourses/digICproj/doc/eusipco_corrected.ps>.
- VALIN, Jean-Marc – MONTGOMERY, Christopher. *Improved Noise Weighting in CELP Coding of Speech - Applying the Vorbis Psychoacoustic Model to Speex* [online]. Príspevok na konferencii AES 120th Convention, 2006-05-20(-23) [cit. 2006-04-27]. Možno nájsť na <http://people.xiph.org/~jm/papers/aes120_speex_vorbis.pdf>.
- VLACHÝ, Václav. *Praxe zvukové techniky*. 2. aktualizované vydání. Praha : Nakladateľství Muzikus, 2000. ISBN 80-86253-05-8.
- Zákon č. 618/2003 Z. z. o autorskom práve a právach súvisiacich s autorským právom (autorský zákon). Príloha k zákonu č. 618/2003 Z. 2003* [cit. 2006-03-21]. Dostupné napr. na internetových stránkach ministerstva kultúry: <<http://www.culture.gov.sk/main/file.php?id=1426&file=file2799.html>>.