

**UNIVERZITA PARDUBICE  
DOPRAVNÍ FAKULTA JANA PERNERA**

**KATEDRA ELEKTROTECHNIKY, ELEKTRONIKY A  
ZABEZPEČOVACÍ TECHNIKY V DOPRAVĚ**

# **ZVUKOVÝ FORMÁT MP3**

**BAKALÁŘSKÁ PRÁCE**

**AUTOR PRÁCE: Petr Gregar**

**VEDOUCÍ PRÁCE: RNDr. Jan Hula**

**2007**

**UNIVERZITY OF PARDUBICE  
JAN PERNER TRAFFIC FACULTY**

**DEPARTMENT OF ELECTRICAL AND ELECTRONICAL  
ENGINEERING AND SIGNALLING IN TRANSPORT**

**SOUND FORMAT MP3**

**BACHELOR WORK**

**AUTOR: Petr Gregar**

**SUPERVIZOR: RNDr. Jan Hula**

**2007**

## **PROHLÁŠENÍ**

Tuto práci jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích dne 8. 5. 2007

.....  
Petr Grgear

## **SOUHRN**

Bakalářská práce se zabývá popisem kódování digitálně uložených zvukových dat podle standardu MPEG – 1 Layer III (MP3) a strukturou takto vytvořeného souboru.

V úvodní části jsou popsány základní pojmy jako zvuk, vnímání zvuků lidským uchem, způsob záznamu zvuku v číslicové formě či princip reprodukce zvuku. Druhá část se věnuje historii formátu od počátků jeho vývoje až do současnosti. Následuje vlastní popis standardu MP3. Významnou výhodou formátu MP3 jsou i tzv. popisky (ID3 TAGy). ID3 TAGům verze 1 a 2 se věnuje čtvrtá část. Na závěr je nastíněn vliv atributů kódování na kvalitu zvuku.

# **OBSAH**

ÚVOD	8	
1	TEORETICKÝ APARÁT	9
1.1	Zvuk	9
1.2	Lidské ucho a jeho vnímání zvuku	10
1.2.1	Frekvenční rozsah	10
1.2.2	Dynamický rozsah	11
1.2.3	Rozlišování frekvencí	11
1.2.4.	Časové maskování	12
1.2.5	Směrové rozlišení	12
1.3	Záznam zvuku	13
1.4	Pulzně kódová modulace	13
1.5	Waveform audio format	14
1.6	Princip reprodukce zvuku	15
2	HISTORIE FORMÁTU	19
2.1	Vývoj	19
2.2	Zpřístupnění veřejnosti	21
2.3	MP2	22
2.4	MP3	22
2.5	Moving Pictures Expert Group – MPEG	23
2.6	Digitální televizní vysílání	24
2.7	MP3 a licenční poplatky	24
3	MP3	26
3.1	Vrstvy – Layers	26
3.2	Stereo-módy	27
3.2.1	Dual channel	27
3.2.2	Stereo	27
3.2.3	Joint stereo	27
3.3	Rámce	27
3.3.1	Hlavička rámce	28
3.3.2	Datová část	32
3.4	Kódování	32
3.4.1	Algoritmus kódování	32
3.4.2	Kvalita kódování	34
3.4.3	Běžné druhy artefaktů	35
3.4.3.1	Ztráta šířky pásma	35
3.4.3.2	Před – ozvěny	35
3.4.3.3	Double – speak	35
3.4.4	Měření kvality zvuku	35
3.4.4.1	Poslechové testy	36
3.4.4.2	Objektivní základní měření	36
3.4.4.3	Techniky měření založené na psycho-akustickém modelu	36
3.4.5	Vliv šířky pásma na kvalitu	36
3.4.6	Ladění algoritmu pro různé přenosové rychlosti	37
3.4.7	Ukládání	37
3.5	Dekódování	38

4	ID3 TAG	39
4.1	ID3 TAG verze 1	39
4.2	ID3 TAG verze 2	39
4.2.1	Hlavička ID3 TAGu verze 2	40
4.2.2	Extended leader – rozšířená hlavička	41
4.2.3	Padding – „vycpávání“	43
4.2.4	Footer – zápatí	43
4.2.5	Přehled rámce ID3 TAGu	43
4.2.5.1	Příznaky hlavičky rámce ID3 TAGu	44
4.2.5.2	Timestamps – časové značky	46
4.2.5.3	Languages – jazyky	46
4.2.5.4	Odkazy	46
4.2.6	Umístění TAGu	46
4.2.7	Unsynchronizace	46
4.2.8	Metoda sychsafe	47
4.2.9	Typy rámců	47
4.2.9.1	Textové informační rámce	48
4.2.9.2	COMM	49
4.2.9.3	URL rámce	49
4.2.9.4	UFID	49
5	Porovnání vlivu atributů kódování na kvalitu	50
	ZÁVĚR	53
	Seznam použité literatury	55
	Seznam tabulek	56
	Seznam obrázků	57
	Seznam zkratk	58
	Seznam příloh	60

## ÚVOD

Mpeg 1 – Layer 3, obvykle označovaný jako MP3 je rozšířený formát kódování digitálního zvukového záznamu.

Zvuk, v číslicové podobě, je posloupnost hodnot amplitudy signálu. Takovýchto hodnot musí být každou časovou jednotku uloženo dostatečné množství, aby bylo možno následně daný signál opět reprodukovat jako „původní“ zvuk (u Audio-CD je těchto vzorků 44 000 každou sekundu). Tyto hodnoty musejí být také dostatečně jemně (s malým minimálním odstupem 2 hodnot) rozčleněny mezi nulovou a maximální hodnotu (Audio-CD používá 16-ti bitové slovo –  $2^{16}$  úrovní). Takovýto záznam je ovšem velmi objemný - zabírá velké množství paměťového místa a vyžaduje velký bit-rate při vysílání. Ke snížení těchto nároků, při zachování kvality, se používají kompresní algoritmy.

MP3 užívá ztrátový kompresní algoritmus, který je navržen k tomu, aby redukoval velké množství dat nekomprimovaného zvukového záznamu. Takto upravený záznam se musí při reprodukci jevit jako věrná „napodobenina“ původního záznamu. Formát byl vymyšlen skupinou inženýrů společností Philips, CCETT, IRT a Fraunhofer Society. Tyto společnosti pracovaly v rámci projektu EUREKA 147 DAB – výzkumný program zabývající se problémem vysílání digitálního radia. Formát MP3 se stal standardem ISO/IEC v roce 1991. Kompresní algoritmus MP3 odstraňuje ze záznamu zvuky, které nejsou lidským uchem slyšitelné a zbývající data zaznamená efektivním způsobem. K výběru zvuků, které budou odstraněny, užívá psychoakustický model. Komprese zvuku dle formátu MPEG má 3 vrstvy: Layer I, Layer II a Layer III (MP3).

# 1 TEORETICKÝ APARÁT

## 1.1 Zvuk

Zvuk je mechanické chvění v hmotném prostředí, které je schopno vyvolat sluchový vjem. Frekvenční rozsah vnímání lidského sluchu leží v rozmezí 20 Hz až 20 000 Hz. Tyto hodnoty jsou orientační. (Každý jedinec má frekvenční rozsah i krajní hodnoty vnímání zvuku jiné, s rostoucím věkem se snižuje horní mez rozsahu atd.) Mechanické chvění o frekvenci ležící mimo tento interval není člověk schopen sluchem vnímat. Zvuk s nižší frekvencí než 20 Hz nazýváme infrazvuk. Zvuk jehož frekvence leží nad 20 kHz je označován jako ultrazvuk.

Zdroj vlnění označujeme jako zdroj zvuku. Látkové prostředí, kterým se toto vlnění šíří nazýváme vodičem. Vodič zvuku zprostředkuje „akustické spojení“ mezi zdrojem zvuku a přijímačem.

Zdrojem zvuku může být jakékoli chvějící se těleso. Kmitání tělesa může být jak vlastní tak vynucené. O intenzitě vlnění vyvolaném chvěním tělesa v jeho okolí nerozhoduje pouze jeho chvění. Má na ni vliv i to, jak je těleso dobrým či špatným zářičem zvuku. Jakým zářičem těleso je závisí především na jeho tvaru. Membrána reproduktoru je dobrým zářičem proto, že rozechvívá vzduch na jedné straně směrem k posluchači, zatímco vlnění vyvolané zadní stranou je tlumeno uvnitř ozvučnice a tak se k posluchači nedostane. Naopak například struna napjatá mezi body tuhého tělesa je špatným zářičem zvuku. Při chvění struny vzniká přetlak ve směru pohybu struny a podtlak na opačné straně. Struna je tak zdrojem dvou vlnění, které se šíří do okolí s téměř opačnou fází. Vzhledem k tomu, že délka vlny zvuku je, v porovnání s příčnými rozměry struny, velmi velká, se tato dvě vlnění interferencí vzájemně ruší.

Vodičem zvuku je v akustice především vzduch, ale zvuk se šíří i v kapalinách a pevných látkách. Rychlost šíření zvuku je v různých látkách odlišná. Pohybuje se od stovek metrů za sekundu až do několika tisíc metrů za sekundu.

Přijímačem může být sluchový orgán, mikrofon či jiný snímač.

Zvuky můžeme z hlediska periodicity průběhu rozdělit na tóny a hluky. Tón vznikne pravidelným periodickým kmitáním tělesa. Lidské ucho vnímá tóny jako zvuk určité „výšky“. Zdrojem tónů může být hudební nástroj, lidské hlasivky atd. Jako hluk označujeme nepravidelné vlnění. Hluky vznikají nepravidelným kmitáním těles.



Zdrojem hluku je například přeskočení elektrické jiskry, výbuch či srážka těles. Hluků využíváme i v hudbě. Zdrojem hluků je totiž i mnoho hudebních nástrojů jako bicí a podobně.

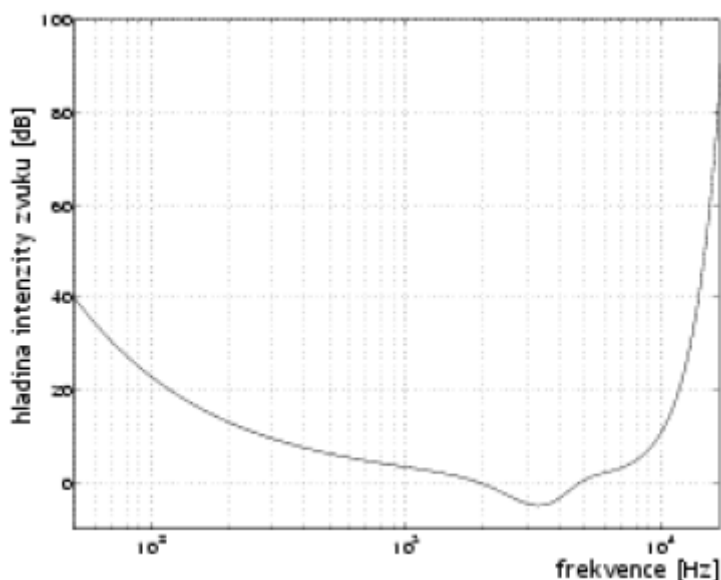
Každý zvuk má svoji fyzikální intenzitu, která se označuje jako hladina intenzity zvuku či akustický tlak. Měří se v decibelech. Decibel je poměrová jednotka a je vztažena k referenční hodnotě. Zde je referenční hodnotu  $2 \cdot 10^{-5}$  Pa (při 1 kHz). Akustický tlak klesá s druhou mocninou vzdálenosti od zdroje zvukového vlnění. Tóny jsou charakterizovány i frekvencí, která určuje výšku tónu. Další charakteristikou je průběh kmitání – ten ovlivňuje zabarvení zvuku. Zapřičiňují to vyšší harmonické, které mohou ležet mimo frekvenční rozsah vnímání zvuku a přesto je lidské ucho vnímá, ale pouze jako onu „barvu“ zvuku.

## 1.2 Lidské ucho a jeho vnímání zvuku

Lidské vnímání zvuku je velice složitý proces, závislý na mnoha faktorech.

### 1.2.1 Frekvenční rozsah

Lidské ucho je přizpůsobeno především vnímání řeči v rozmezí 500 Hz až 3 000 Hz. Nejvyšší informační hodnotu má lidská řeč v rozmezí 500 Hz až 2 000 Hz.



Oblast mezi 2 000 Hz a 4 000 Hz je velice důležitá pro srozumitelnost. Jak již bylo uvedeno výše, lidské ucho je schopno vnímat zvuky od cca 20 Hz až do cca 20 000 Hz (horní hranice výrazně klesá s rostoucím věkem). Omezeného frekvenčního rozsahu je využito při případném komprimování záznamu.

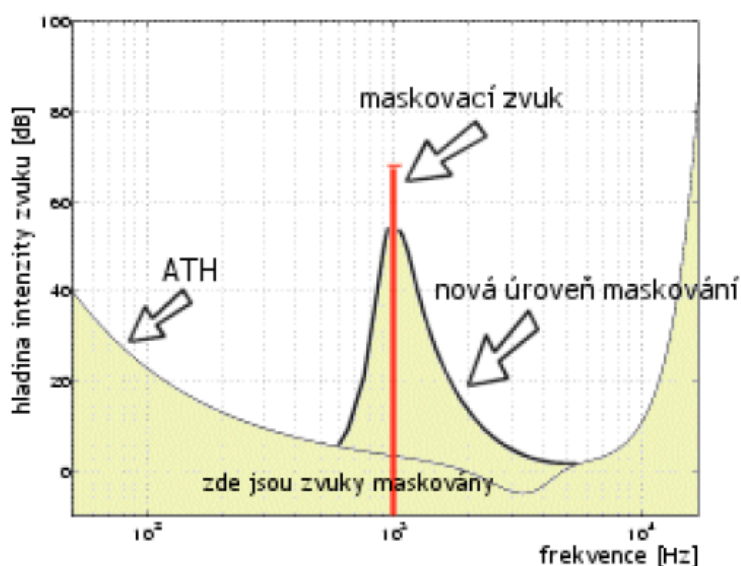
Obr.1 Křivka ATH  
(Zdroj: <http://home.zcu.cz/~mtoman/mp3.pdf>)

## 1.2.2 Dynamický rozsah

Dynamický rozsah je vnímán jako poměr nejhlasitějšího a nejnižšího vnímatelného zvuku. Dynamický rozsah je různý na různých frekvencích a uprostřed slyšitelného pásma je to asi 120 dB. Na okrajích pásma je citlivost (a tím i dynamický rozsah) mnohem menší. Citlivost lidského ucha ukazuje křivka ATH (Absolute Threshold of Hearing – Obr. 1). Tato křivka zobrazuje množství akustické energie potřebné k rozeznání zvuku mladým zdravým člověkem v nehlukném prostředí. Citlivost lidského ucha je nesmírně velká. Ve frekvenční oblasti kde je ucho nejcitlivější 3 500 Hz je zdravé odpočínuté lidské ucho schopno zaznamenat zvuk vyvolaný rozkmitem vzduchových částic o jednu miliontinu milimetru. Bylo-li by lidské ucho ještě o jeden řád citlivější, zaznamenalo by jako šum molekulární pohyb částec vzduchu.

## 1.2.3 Rozlišování frekvencí

Schopnost rozlišení frekvence tónů je u každého člověka jiná. Navíc je frekvenčně závislá. Uprostřed slyšitelného frekvenčního pásma za ideálních podmínek lze rozlišit změnu frekvence o několik centů (cent je 1 / 1200 oktávy). Na nízkých frekvencích je to pod rozdíl 50 Hz, zatímco na nejvyšších kmitočtech



Obr. 2 Frekvenční maskování  
(Zdroj: <http://home.zcu.cz/~mtoman/mp3.pdf>)

dosahuje až 4 000 Hz. Na okrajích pásma je rozlišovací schopnost výrazně omezena. Schopnost odlišit dva frekvenčně blízké tóny ovlivňuje frekvenční maskování. Pokud znějí dva tóny současně, může jeden z nich potlačit slyšitelnost druhého. Tato neschopnost slyšet oba současné tóny se nazývá frekvenční maskování. Maximální úroveň maskovaného signálu je závislá na frekvenční vzdálenosti a úrovni maskujícího signálu. Maskovací schopnost také závisí na

frekvenci maskujícího tónu. Rozlišovací schopnost vnímání tónů s blízkými frekvencemi je ovlivněno šířkou kritického pásma. To má na nejnižších kmitočtech velikost kolem 100 Hz, zatímco na nejvyšších kmitočtech dosahuje až 4 kHz. Tohoto se využívají některé algoritmy pro kompresi zvukových dat. Obrázek 2 ukazuje oblast která se může při kompresi záznamu maskovat (vyřadit).

#### 1.2.4. Časové maskování

Následuje-li po hlasitém tónu velice podobný tón výrazně nižší hlasitosti, je jeho vnímání potlačeno. To je dáno „vlastní“ ochranou sluchového ústrojí. Soustava 3 kůstek vnitřního ucha představuje velice „důmyslnou automatickou regulaci citlivosti sluchu“. V tichém prostředí přenášejí kůstky veškerou akustickou energii z bubínku do oválného okénka (blanka uzavírající vlastní sluchový orgán - hlemýžď). Nejvíce převedené energie je zajištěno jejich vzájemným dokonalým přilehnutím. Při silnějším zvuku (větším akustickém tlaku) se napětí kůstek uvolní, kůstky na sebe přiléhají méně a do vnitřní části ucha se přenáší jen malá část akustické energie a tím je sluchový orgán chráněn před poškozením. Napětí kůstek není řízeno vůlí a je realizováno jemnými svaly. Regulační rozpětí tohoto orgánu je obrovské. Proto je lidské ucho schopno vnímat zvuky s velmi velkým rozsahem intenzit. Porovnáme-li intenzity nejslabších slyšitelných zvuků s prahem bolesti zjistíme že jde o zvuky lišící se svojí intenzitou  $10^{13}$ -krát. Je zřejmé, že bez regulace by se sluchový orgán poškodil při prvním hlasitém zvuku. Po dlouhém pobytu v hlučném prostředí citlivost tedy relativně rychle klesá a následným pobytem v tichém prostředí se opět postupně zvyšuje. Zvyšování citlivosti neprobíhá skokově, ale relativně pomalu – daleko pomaleji než snižování citlivosti. Tato vlastnost lidského sluchu maskuje tiché tóny následující po hlasitému tónu s podobnou frekvencí a může jí využít algoritmus pro kompresi.

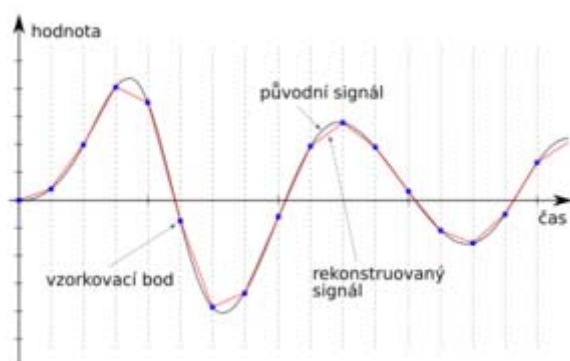
#### 1.2.5 Směrové rozlišení

Lidské ucho není schopno zjistit odkud k němu přicházejí především nízké tóny. Toho lze využít při kompresi záznamu.

### 1.3 Záznam zvuku

Vlnění vzduchu rozechvívá membránu mikrofону, která je jednou z desek kondenzátoru. V závislosti na změně polohy desky se mění kapacita kondenzátoru a tím i napětí mezi deskami (existují i jiné druhy mikrofónů). Toto napětí je zesíleno na patřičnou úroveň, navzorkováno a převedeno A/D převodníkem na binární data. Tyto hodnoty se uloží do paměti.

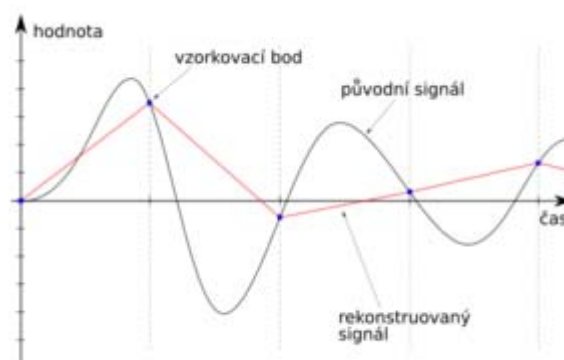
### 1.4 Pulzně kódová modulace



Obr. 3 Vzorkování signálu  
(Zdroj: [http://cs.wikipedia.org/wiki/Pulzně\\_kódová\\_modulace](http://cs.wikipedia.org/wiki/Pulzně_kódová_modulace))

Pulzně kódová modulace (PCM) byla vytvořena roku 1937. Je to modulační metoda, která umožňuje převést časově i amplitudově spojitý signál na posloupnost vzorků, které jsou kvantovány (jejich amplitudy nabývají pouze diskretních hodnot). Amplituda signálu se v pravidelných intervalech odečítá a A/D převodníkem se převádí do binární formy. Přípustné hodnoty amplitud jsou lineárně rozděleny mezi nulovou a maximální hladinou. Počet úrovní je roven  $2^N$ , kde N je počet bitů A/D převodníku. Rozhodujícím parametrem pro přesnost záznamu je tedy jemnost rozlišení úrovní signálu. Převodníky bývají maximálně 24 bitové (CD – 16 bitů). Chybou vzniklou rozdílem mezi určenou a skutečnou hodnotou vzniká kvantizační šum. Další neméně významný parametr ovlivňující kvalitu záznamu je vzorkovací kmitočet. Je to frekvence s jakou se opakuje odečet vzorku a pohybuje se od 8kHz (telefonní linky), přes 44 100 Hz (CD) až po 192 000 Hz (u nových zvukových formátů či při profesionálním zpracování hudby ve studiu). Obr. 4 zobrazuje znehodnocení nízkou frekvencí odečítání. Obr. 3 zobrazuje „dobře“ navzorkovaný signál. Dle Nyquistovy podmínky musí být vzorkovací kmitočet minimálně dvakrát

Pulzně kódová modulace (PCM) byla vytvořena roku 1937. Je to modulační metoda, která umožňuje převést časově i amplitudově spojitý signál na posloupnost vzorků, které jsou kvantovány (jejich amplitudy nabývají pouze diskretních hodnot). Amplituda signálu se v pravidelných intervalech odečítá a A/D převodníkem



Obr. 4 Nevhodné vzorkování signálu  
(Zdroj: [http://cs.wikipedia.org/wiki/Pulzně\\_kódová\\_modulace](http://cs.wikipedia.org/wiki/Pulzně_kódová_modulace))

větší než maximální kmitočet vzorkovaného signálu. Pulzní kódová modulace se používá například na formát Audio CD (red book) . Dva kanály, 16 b, 44 100 Hz. Není-li záznam v tomto formátu nijak upraven (komprese, či parita) zabere 1 minuta:  $44\ 100 * 60 * 2 * 16 / 8 = 10\ 584\ 000$  B. Běžné CD má kapacitu 650 MB (74 minut záznamu dle norem Audio-CD, základní přenosová rychlost 150 kB/s). Omezená kapacita média (či přenosová rychlost) je důvod proč se používají komprese zvukových souborů. Tuto situaci lze řešit většími kapacitami novějších médií jako například: DVD (kapacita 4,5 GB na jednu vrstvu, základní přenosová rychlost 1350 kB/s), novějšími Blu-ray (kapacita 25GB na jednu vrstvu, základní přenosová rychlost 4,5 MB/s) či HD DVD (kapacity 15GB na jednu vrstvu, základní přenosová rychlost 4,45 MB/s).

### 1.5 Waveform audio format

WAV (nebo také WAVE) je zkratka a používaná přípona pro Waveform audio format. Tento zvukový formát vytvořily firmy IBM a Microsoft pro ukládání zvuku na PC. Je to speciální varianta obecnějšího formátu RIFF. Přestože je možné ukládat do WAV souboru zvuk komprimovaně, většinou se používá nekomprimovaný zvuk v pulzně kódové modulaci. Tímto způsobem je uložen zvuk na Audio CD (přidány opravné kódy atd.), což umožňuje snadný převod mezi těmito formáty. PCM je bezztrátový formát a tak se používá WAV nejčastěji při zpracování zvuku. K ukládání zvukových záznamů se dnes používají komprimované soubory. Nejčastěji používané jsou ztrátové formáty (MP3, Vorbis Ogg, WMA, AAC). Toto jsou formáty tzv. ztrátové komprese. Ztrátová komprese sice ušetří více místa na médiu, ale vlastní kompresí se záznam „poškodí“. „Poškození“ spočívá v tom, že ze záznamu se nevratně odstraní některá data. Toto „poškození“ musí být přijatelné, aby měl opětovně reprodukováný zvuk požadovanou kvalitu. Je-li třeba zachovat kvalitu originálního nekomprimovaného záznamu a současně zmenšit velikost souboru, lze použít bezztrátovou kompresi. Formáty bezztrátové komprese jsou například Monkey's Audio, FLAC, Shorten, Apple Lossless či WMA Lossless. Bezztrátové komprese dosahují vyšší kvality, ale nižšího kompresního poměru.

## 1.6 Princip reprodukce zvuku

Jak již bylo uvedeno výše, zvuk je chvění vzduchu. Proto je (pro vnímání informace sluchovou cestou) nutné, převést posloupnost vzorků (viz WAV či PCM) zpět na chvění vzduchu.

Data (posloupnost vzorků) je nutno přivést na vstup D/A převodníku. D/A převodník je součástka, která je schopna převést vstupní binární data na analogový signál na výstupu. Hodnota výstupního signálu však nemůže nabývat nekonečně mnoha hodnot, protože vstupní signál je nespojitý (co se týče amplitudy). Může nabýt jenom určitých diskrétních hodnot. Chybu, která je tímto vytvořena nazýváme kvantizační chyba. Tato chyba se ve výsledném zvuku projeví jako šum. Počet hodnot, které může nabývat výstupní signál, lze vyjádřit pomocí dalšího parametru: kvantizačního kroku (Q), pro nějž platí vztah  $Q = 1 / (2^n - 1)$ . Kde n je počet bitů vstupní informace. Existuje mnoho typů D/A převodníků s rozdílnými parametry i cenami.

Převedený (nyní již analogový) signál musí být patřičně upraven a zesílen. Signál se nejprve upraví dolno-frekvenční propustí, která z něj odstraní nežádoucí vysoké frekvence (neslyšitelné). Poté se signál zesílí operačními zesilovači. Nakonec se signál ještě zesílí výkonovým zesilovačem, aby byl připraven budit kmitací systém reproduktoru (či reproduktorů).

Reproduktor je měnič elektrické energie na akustickou. Vzhledem k jeho malé účinnosti (velká část přiváděné elektrické energie se v něm mění v teplo) je nutno přivést na jeho svorky signál dostatečného výkonu. Hlavní součástí reproduktoru je jeho kmitací systém. Je to pohyblivá cívka obklopena konstantním magnetickým polem permanentního magnetu. Přivedeme-li do této cívky proud, vyvoláme pohyb, který využijeme následovně. Ke kmitací cívce je připevněna membrána reproduktoru. Membrána se pohybuje společně s cívkou a rozechvívá vzduch. Celá pohyblivá část reproduktoru by měla být co nejlehčí, aby byla schopna svým pohybem sledovat průběh přiváděného proudu. Membrána by měla být maximálně tuhá, aby nezkreslovala výsledný zvuk. Proto se na membrány používají různé materiály od papíru až po materiály jako je kevlar, či skelná vlákna. Pohybový systém bývá také mechanicky tlumen, omezí se tak jeho rezonance. Tyto rezonance zkreslují výsledný zvuk a jsou nebezpečné, protože při jejich zesílení může dojít k poškození reproduktoru. Existuje mnoho druhů reproduktorů. Liší se především schopností

reprodukovat určité kmitočtové pásmo. Podle toho dělíme reproduktory na hlubokotónové, středo-tónové a vysoko-tónové. Reproductory se liší svoji konstrukcí, parametry i cenou.

Hlubokotónové reproduktory bývají relativně rozměrné a mají největší povolenou výchylku membrány. Tento druh reproduktorů je svoji konstrukcí uzpůsoben k reprodukci nejnižších frekvencí (od 20 až 50 Hz do několika set Hz). Důležité vlastnosti tohoto typu reproduktorů jsou například: citlivost, impedance, zatížitelnost, rezonanční frekvence reproduktoru, spodní přenášený kmitočet (i rozsah), maximální výchylka membrány, rozměry membrány, mohutnost kmitacího systému, hmotnost pohyblivých částí... Velice podstatný je i průběh citlivosti v závislosti na frekvenci a činitel, který nám ukáže do jaké ozvučnice je reproduktor vhodné umístit.

Středotónové reproduktory mají velice podobnou konstrukci jako hlubokotónové reproduktory. Jsou však většinou menší. Jsou určeny k reprodukci frekvencí od stovek Hz do několika kHz. Tyto reproduktory přenášejí pásmo, ve kterém je lidské ucho nejcitlivější a tak je jejich kvalita velice důležitá. U těchto reproduktorů sledujeme opět některé velice důležité parametry - téměř stejné jako u basových reproduktorů.

Vysokotónové reproduktory se svoji konstrukcí opět příliš neodlišují od výše uvedených typů. Mají však mnohem menší rozměry a jejich membrány jsou vyráběny z jiných materiálů. Vysokotónové reproduktory musejí být chráněny před buzením nízkými frekvencemi, protože při takovém buzení je možné je velice snadno spálit. Vysokotónové reproduktory směřují (nevyzařují zvuk všemi směry jako basové reproduktory, ale pouze v relativně úzkém „paprsku“). Proto se někdy doplňují i prostředky k rozptýlení vyzařovaného zvuku. Mezi důležité parametry patří doporučená dělící frekvence. Ostatní parametry jsou obdobné jako u výše uvedených typů.

K výrobcům kvalitních reproduktorů patří například firmy Scan-speak, Dynaudio, Seas, Audax, či výrobci kvalitních reprosoustav jako například JM-lab nebo B&W. Výrobci reprosoustav však většinou neprodávají měniče samostatně. Výrobou reproduktorů u nás se zabývá firma TVM Acoustic z Valašského Meziříčí.

Vzhledem k velice rozdílným požadavkům na reproduktor, který má produkovat nejnižší (20 Hz) či nejvyšší (20 000 Hz) kmitočty je většinou nutné zapojit více reproduktorů, které jsou dohromady schopny reprodukovat celé slyšitelné

kmitočtové pásmo. Pro jejich správnou spolupráci je nutno rozdělit příchozí signál mezi jednotlivé měniče. Signál, mezi jednotlivé reproduktory, rozděluje zařízení zvané výhybka. Ta může být provedena z pasivních či aktivních součástí. Aktivní provedení vyhybek používá filtry realizované operačními zesilovači a kapacitami či DSP. Tento typ vyhybek pracuje s nižší úrovní signálu. Proto je třeba zesílit výstup z takovéto vyhybky. Naproti tomu pasivní provedení je tvořeno kapacitami, indukčnostmi a rezistory. Taková vyhybka pracuje s výkonovým signálem a tak je možno takovou vyhybku vsadit mezi zesilovač a reproduktor.

Reproduktor vykonává pístový pohyb. V jeden okamžik je na jedné straně membrány je vzduch tlačěn, ale na druhé táhnut za membránou. Kdyby byl reproduktor ve volném prostoru, tato vlastnost by v dostatečné vzdálenosti či v určitých místech vážně znehodnotila zvuk (tzv. akustický zkrat). Zvuk by v extrémním případě neexistoval, protože by se vlnění vzduchu vyvolaná na obou stranách membrány odečetla. Proto se reproduktory (či skupiny reproduktorů) umísťují do ozvučnic. Ozvučnice je jakýsi box, který zabrání šíření vlnění, vyvolaného zadní stranou membrány, do prostoru. Existují i typy ozvučnic, které toto vlnění využijí pro „posílení“ určitého pásma.

Nejjednodušším typem je uzavřená ozvučnice. Tento typ hermeticky odděluje prostor uvnitř boxu (za membránou) od okolí (prostor před membránou). Takové provedení má několik výhod. Chrání reproduktor před utržením, nemění tvar průběhu citlivosti v závislosti na frekvenci, nepoužívá rezonanční obvod (který zkresluje zvuk). Uzavřená ozvučnice má i některé nevýhody. Takovéto ozvučnice jsou poměrně rozměrné a k posluchači vyzařuje jen  $\frac{1}{2}$  (jedna strana) membrány. Energie vyzářená zadní stranou membrány se pohltí uvnitř boxu.

Dalším typem ozvučnice je ozvučnice typu basreflex. Tato ozvučnice používá rezonanční obvod, který zesiluje určité kmitočty. Tento rezonanční obvod je založen na kmitání hmoty vzduchu v basreflexovém nátrubku. Basreflexový nátrubek ústí jednou stranou dovnitř boxu a druhou do okolního prostoru. Tímto způsobem se využije i energie vyzářená zadní stranou membrány. Basreflexová ozvučnice má výhodu především ve větší citlivosti na nízkých frekvencích při menších rozměrech boxu. Taková konstrukce má i řadu nevýhod. Rychlost proudění vzduchu uvnitř nátrubku je omezená, protože při velké rychlosti proudění vzniká na okrajích nátrubku turbulentní proudění, které slyšitelně znehodnotí zvuk. Tento jev lze omezit



zaoblením konců nátrubku. Reproduktor v takovéto ozvučnici není již tolik chráněn před utržením. Basreflexová ozvučnice také více zkresluje zvuk.

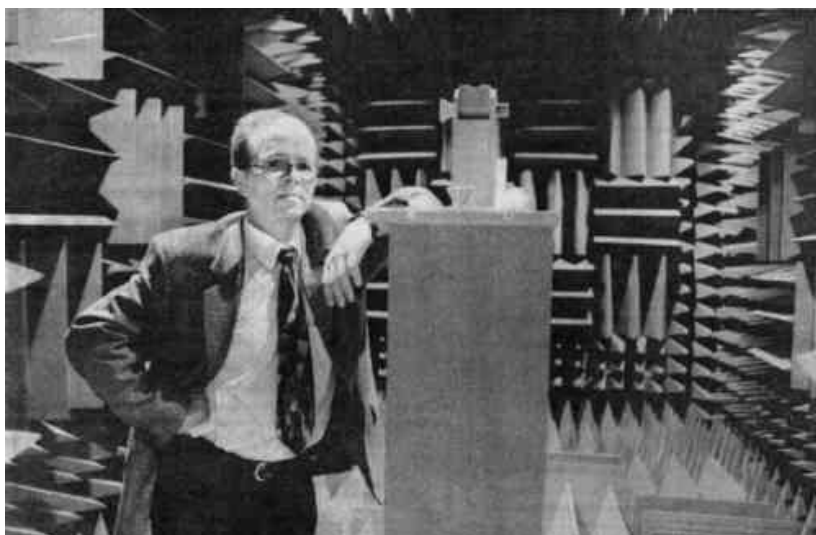
Dále užívaným typem ozvučnice je ozvučnice typu pasivní zářič. Pasivní zářič je reproduktor bez kmitacího systému. Princip funkce je velice podobný basreflexové ozvučnici. Hmotnost membrány pasivního zářiče nahrazuje hmotnost vzduchu v basreflexovém nátrubku. Jelikož je vnitřek boxu hermeticky oddělen od okolí eliminuje tento typ ozvučnice problémy s turbulentním prouděním. Kvalita reprodukce je však závislá na dvou elementech.

Poslední užívanou variantou je ozvučnice typu šnek. Tento typ se používá výhradně pro profesionální ozvučování. Box tvoří jakýsi labyrint ve kterém se zvukové vlny šíří, odrážejí a zesilují až nakonec zesílené vyjdou ven. Tato ozvučnice má největší citlivost na nízkých kmitočtech, ale její návrh je velmi složitý. Takovéto boxy jsou velice těžké a objemné.

## 2 HISTORIE FORMÁTU

### 2.1 Vývoj

Počátkem 80. let začal profesor Dieter Seitzer (Obr. 5) z Erlangen-Nuremberg University v Německu zkoumat problém přenosu komprimované hudby po telefonní



Obr. 5                      Profesor Dieter Seitzer

lince. Sestavil skupinu vědců zabývajících se výzkumem kódování zvuku.

Roku 1979 vyvíjí tým kolem profesora Seitzera první digitální signálový procesor schopný zvukové komprese. Během následujícího vývoje student prof. Seitzera,

Karlheinz Brandenburg (obr. 6), rozšířil základní principy vnímání kódovaného zvuku lidským sluchem (popsal psycho-akustiku). Pod profesorem vedením bylo v tuto dobu vyvinuto mnoho kompresních algoritmů. Karlheinz Brandenburg získal diplom (ing.) z elektrotechniky v roce 1980 a z matematiky 1982.

V roce 1987 výzkumný tým vytvořený mezi Erlangen-Nuremberg University a Fraunhoferským institutem pro integrované obvody začal v rámci projektu evropské unie EUREKA EU147 zkoumat možnosti DAB (digitální vysílání rádia). Tým prof. Heinz Gerhaeuser z Fraunhoferského institutu udělal významný pokrok. S algoritmem LC-ATC (Low Complexity Adaptive Transform Coding) jako základem pro jejich další výzkum. Začalo budování a



Obr.6 Karlheinz Brandenburg

práce na real-time kodeku, který využívá vícenásobný digitální signálový procesor. Hardwarový systém, založený na vícenásobných DSP jednotkách a řadě zvukových

i datových vstupních/výstupních desek. Tento systém byl vybudován od základu týmem vědců mezi nimiž byli Harald Popp a Ernst Eberlein. Do jisté doby existoval LC-ATC jenom jako simulace na počítačích a mohl být testován jen na malém množství zvukového materiálu kvůli vysokým nárokům na výpočetní výkon. Real-time kodek umožnil testovat kompresi v reálných podmínkách a počítat s významnými dodatečnými algoritmickými optimalizacemi.

Roku 1989 dokončil Karlheinz Brandenburg svoji disertační práci. Popsal v ní algoritmus OFC, který má mnoho základních rysů eventuelního MP3 kódovacího zařízení, včetně vysokého frekvenčního rozlišení filtrů, nerovnoměrného kvantování, Hoffmanova kódování atd.. Kódování OFC je považováno za průlom ve své době a za předchůdce MP3. OFC algoritmus skenoval a odstraňoval zvuky pod nebo nad prahem slyšitelnosti lidského ucha. Softwarová část real-time systému OFC byla realizována především Bernhardem Grillem pod vedením profesora Heinze Gerhaeusera. Na základě zapracovaných přídavků Hannoverské University, AT&T, a Thomsonu, tým z Fraunhoferského institutu vylepšil algoritmus OFC a tak zavedl nový zvukový kodek zvaný ASPEC (Adaptive Spectral Perceptual Entropy Coding). ASPEC byl navrhován pro chystaný MPEG zvukový standard. MPEG měl zpočátku 14 návrhů, což přinutilo účastníky aby svoje návrhy sloučily. To vyústilo ve 4 návrhy mezi nimiž byl i ASPEC a MUSICAM. Po formálních testech MPEG zvítězily právě ASPEC a MUSICAM. Tyto algoritmy byly vybrány pro jejich jednoduchost, odolnosti proti chybám a relativně nízkým nárokům na výpočetní výkon při kódování vysoce kvalitního audio záznamu. Ustanoveny byly některé důležité náležitosti jako rychlost vzorkování, struktura rámců, struktura hlaviček či počet vzorků v rámci. Základní myšlenky formátu musicam byly plně zahrnuty do definice formátů ISO MPEG Audio Layer I, Layer II a později i Layer III (MP3).

Pod vedením profesora Mussmanna z Hannoverské univerzity byl vydán standard Layer I (zodpovědný pracovník Leon van de Kerkhof) a Layer II (zodpovědný pracovník Gerhard Stoll). Pracovní skupina (složená z pracovníků: Leon Van de Kerkhof (Nizozemí), Gerhard Stoll (Německo), Leonardo Chiariglione (Itálie), Yves-François Dehery (Francie), Karlheinz Brandenburg (Německo)) si vypůjčila myšlenky algoritmu ASPEC, rozvinula je, obohatila o své nápady a vytvořila Layer III (MP3). Tato vrstva byla navržena k tomu, aby poskytla stejnou kvalitu zvuku při 128 kb/s jako Layer II při 192 kb/s.

Jako důkaz kvality vyrobil a prodal Fraunhoferský institut několik kusů studiového vybavení (19“ racky) založeného na systému ASPEC pro profesionální uživatele jako několik rozhlasových stanic. Jako první bylo toto zařízení použito ke spolehlivému přenosu hudby přes ISDN mezi rozhlasovými stanicemi. Vývoj ASPEC k finálnímu MP3 kodeku zahrnoval některá technická přizpůsobení s dalšími plánovanými kódovacími zvukovými zařízeními plánovaného MPEG – 1, rozšíření o joint stereo kódování, které dovoluje dobře kódovat nejen mono, ale též stereo signál. Všechny algoritmy byly dokončeny roku 1992 jako součást prvního standardu MPEG – 1 (původně pro Video-CD). Tento standard byl publikován roku 1993 jako mezinárodní norma ISO/IEC 11172-3. V jeho zvukové části jsou zahrnuty všechny 3 vrstvy kódování audia (Layer I, Layer II i Layer III). Další práce na audio-části formátu MPEG skončila roku 1994 jako součást druhé skupiny standardů MPEG. Mezinárodní norma MPEG-2 byla vydána roku 1995 jako ISO/IEC 13818-3.

Efektivita komprese kodérů je typicky definována přenosovou rychlostí. Kompresní poměr je závislý na vzorkovací frekvenci a bitové hloubce komprimovaného signálu. Často jsou publikovány kompresní poměry vztažené k formátu Audio-CD (44 100 Hz, 16 bit / kanál, 2 kanály). Karlheinz Brandenburg používal záznam na CD k hodnocení kvality komprimovaného zvuku. Některé konkrétní zvukové výňatky (triangl, harmonika či zvonkohra) byly subjektivně ohodnoceny i profesionály, zabývajícími se zvukem či hudbou.

## 2.2 Zpřístupnění veřejnosti

Referenční softwarová implementace (napsaná v jazyce C, známá jako 11172–5) byla vyvinuta členy výboru ISO MPEG Audio proto, aby poskytovala korektní MPEG Audio soubory (Layer I, Layer II a Layer III). Kodér nepracoval v reálném čase, ale pracoval na množství operačních systémů. To umožnilo simulovat první hardware pro dekódování komprimovaného audia v reálném čase (založené na DSP). Další implementace kodérů formátu MPEG, pracujících v reálném čase, byla vytvořena pro potřebu digitálního vysílání DAB (rádio) a DVB (televize) k domácím přijímačům.

## 2.3 MP2

V říjnu 1993 se na internetu objevily soubory formátu MPEG-1 - Layer II, které byly často přehrávané. Zpočátku byl pro tvorbu MP2 k dispozici pouze Xing encoder (pro kódování) a programy cdda2wav či CDripper (pro extrahování zvukových stop z Audio-CD). Internet Underground Music Archive, který se tak vytvořil, je uznáván jako start on-line hudební revoluce. Internet Underground Music Archive byla první stránka hostující tisíce vysoce kvalitních a autorizovaných nahrávek.

## 2.4 MP3

7. června 1994 uvolnila Fraunhofer Society první softwarový MP3 kódér nazývaný l3enc. Přípona .mp3 byla vybrána (všemi hlasy v hlasování vědců z Fraunhoferského institutu) 14. června 1995. Předtím byla používána přípona .bit. První softwarový MP3 přehrávač Winplay3 (pracující v reálném čase) byl uvolněn 9. září 1995. Nyní již bylo množství uživatelů schopno kódovat a reprodukovat formát MP3 na jejich osobních počítačích. Protože v té době byla kapacita pevných disků poměrně malá, byla tato technologie podstatná pro ukládání a reprodukci neinstrumentálních hudebních děl na osobních počítačích.

V druhé polovině roku 1995 se na internetu začaly šířit soubory MP3. Jejich popularita se výrazně zvýšila s příchodem přehrávače Winamp (firmy Nullsoft uvolněným roku 1997), či přehrávače mpg123 (určeného pro Unix). Další vlna rozšíření přišla s rozmachem peer-to-peer systému sdílení souborů (Napster roku 1999). Takovéto sdílení hudebních souborů bylo možné díky relativně malé velikosti souborů. Hlavní nahrávací společnosti argumentovali (a stále argumentují), že takovýto podíl volně šířené hudby snižuje jejich zisky. To vyústilo až k právnímu napadnutí Napsteru (jehož provoz byl nakonec ukončen) i uživatelů podílejících se na sdílení těchto dat. Programy jako Winamp, Napster či mpg123 zpřístupnily vytváření, sdílení i reprodukci souborů MP3 i běžným uživatelům. Jak se stával formát čím dál tím více uživatelsky oblíbený, začaly se vyrábět přenosné přehrávače. Ty stále zmenšují své rozměry a zvyšují svoji kapacitu. Navzdory popularitě MP3 „obchodníci s on-line hudbou“ často užívají jiné formáty. Tyto formáty jsou tzv. vlastnické. Data jsou zašifrována, aby se zabránilo uživatelům v užívání zakoupené hudby k jiným účelům, než ke kterým je určila nahrávací společnost (hlavně se jedná o sdílení souborů a tím porušování jejich autorských práv). Toto opatření má i své

nežádoucí účinky jako například nemožnost reprodukovat hudbu na jiných zařízeních atd..

## 2.5 Moving Pictures Expert Group - MPEG

Formát MP3 je jeden ze skupiny formátů MPEG. Všechny formáty skupiny MPEG jsou ztrátové a definované jako otevřený standard (specifikace jsou za poplatek přístupné každému). Využívá se zde principů časového a frekvenčního maskování. Standardy MPEG se dělí následovně:

### MPEG – 1

Tento standard obsahuje 3 vrstvy – Layer I, II a III. Layer I se dnes již téměř neužívá. Layer II se užívá pro DAB (digitální vysílání rádia) či Video-CD. Ve spojených státech je Layer II znám jako MUSICAM. Layer III je znám jako MP3. Tyto vrstvy (layer) se liší kvalitou a složitostí implementace.

### MPEG – 2

Používá se k distribuci audio i videosignálu přes satelity či obdobná zařízení. Je také využíván DVD. MPEG – 2 můžeme rozdělit na další části: BC (zpětně kompatibilní s MPEG – 1, přidány další vzorkovací frekvence, možnost kódovat více kanálů) a AAC (označována jako Layer 7, oproti MP3 přidává nové nápady vedoucí ke zmenšení a ke zkvalitnění záznamu). MPEG – 2 AAC: Začátkem roku 1994 prokázaly ověřovací testy značné zlepšení efektivity kódování nových kódovacích algoritmů. Ty ovšem nebyly zpětně kompatibilní se standardem MPEG – 1. Důsledkem toho se zavedl nový standard kódování zvuku MPEG -2 AAC (Advanced Audio Coding). Standard byl dokončen roku 1997 jako IS 13818 – 7.

### MPEG – 3

Původně byl vyvinut pro HDTV (High Definition TV), ale později byl sloučen s MPEG – 2.

### MPEG – 4

Tento standard byl vytvořen se zaměřením na komplexní popis multimediálních objektů. Podporuje přímou interakci s internetem. Obsahuje moduly pro syntézu zvuku a mnohé další pokročilé technologie. Velký důraz byl při vývoji kladen na přijatelnou kvalitu při nízké přenosové rychlosti a škálovatelnost všech částí.

## MPEG – 7

Je standard pro komplexní popis multimediálních objektů – pro snadné vyhledávání. Nemá s audio-kompresí téměř nic společného.

## 2.6 Digitální televizní vysílání

Zkratka DVB označuje Digital Video Broadcasting (vysílání digitálního videa). Existuje několik typů DVB. DVB-T (terrestrial – pozemní vysílání), DVB-S (satellite – satelitní vysílání), DVB-H (handheld – vysílání pro přenosná zařízení), DVB-C (cable – kabelová televize) a DVB-MHP (multimedia home platform – multimediální domácí platforma).

Typ DVB-T používá pro kompresi video-dat (audio + obraz) před vysláním formáty MPEG – 2 (pro kódování zvuku vrstvu Layer II) a výjimečně i MPEG – 4. Na jednom kanále se vysílá tzv. multiplex. To je jakýsi balíček několika televizních i rádiových stanic a doplňkových služeb (EPG (elektronický programový průvodce), teletext...).

## 2.7 MP3 a licenční poplatky

Je všeobecně rozšířeno, že standard MP3 je volně přístupný a k dispozici zdarma pro jakékoli použití. To však není pravda.

V září 1998 rozeslal Fraunhofer Institute dopisy několika vývojářům pracujícím na „volných“ kodérech založených na standardu ISO. V tomto dopise je jasně uvedeno, že každý vývojář či vydavatel kodérů, založených na normě MPEG audio Layer III, je povinen zaplatit licenční poplatek Fraunhoferskému Institutu.

Vývoj standardů probíhá volnou formou. Každý se může připojit k vývojové skupině (dotázáním se na místním MPEG – kontaktním místě) a tak na vývoji pracuje mnoho lidí. Tento druh výzkumu vyžaduje hluboké znalosti a množství času - většinou tolik, že to vyžaduje celodenní zaměstnání. A tak všichni vývojáři pracují pro společnosti, které je za jejich práci platí. Z toho vyplývá, že výsledky práce, kterou tito lidé odvedou, jsou majetkem jejich zaměstnavatele. Schválení formátu ISO obsahuje několik částí. Definuje syntaxi bit-ratu, který musí být správně dekodován dekodérem, nedefinuje však žádná pravidla pro kódování. Navíc poskytuje informativní zdrojový kód kodéru i dekodéru, který je volně dostupný.

Fraunhoferský Institut byl nejvýznamnějším vývojářem MPEG audio Layer III a má svoji práci chráněnou několika patenty. Navíc institut spojil svoje patenty s patenty firmy Thomson. Důvod byl prostý: žádat licenční poplatky za využití této skupiny patentů. Nyní je v této skupině 18 patentů souvisejících s MP3. Některé z patentů mohou být obejity nepoužitím některých rysů standardu, či použitím jiných než specifikovaných algoritmů. Ale existují i patenty, kterým se při implementaci nelze vyhnout a tak je vývojář musí použít (a někdo zaplatit). Existují i patenty jiných firem, které souvisejí s MP3. Tyto firmy mohou v budoucnu požadovat za jejich využití také poplatky, avšak zatím se o to žádná firma mimo Fraunhofer Institute / Thomson nepokusila. Na adrese [www.mp3licensing.com](http://www.mp3licensing.com) je „ceník“ aplikací souvisejících s využíváním standardu MP3 i seznam patentů. Za rok 2005 vygenerovala tato povinnost 100 milionů Euro.



## **3 MP3**

### **3.1 Vrstvy – Layers**

Každý ze standardů obsahuje několik částí, v nichž se popisuje kódování audia, kódování videa, kódování synchronizačních dat a formáty ukládání kódovaných dat. Standard MPEG obsahuje několik vrstev popisujících způsob kódování. Od Layer I k Layer III roste složitost algoritmů, komplexnost a efektivita komprese. To je ovšem vykoupeno většími nároky na výpočetní výkon či na čas potřebný ke kódování a dekódování. Dekodér umožňující dekódovat Layer N umožňuje dekódovat i Layer <N (je zpětně kompatibilní).

Layer I je nejjednodušší schéma založené na algoritmu MUSICAM. Původně bylo určeno pro DCC (Digital Compact Cassette). Dnes se již téměř neužívá. Tato vrstva obsahuje základní mapování vstupního digitálního zvuku do 32 dílčích pásem, pevně dělí data do rámců, kvantování užívá blokové kompenzace a psycho-akustický model určuje adaptivní přidělování bitů.

Layer II je kompromisem mezi kvalitou, rychlostí a kompresním poměrem. Je to vlastně optimalizovaná verze algoritmu MUSICAM. Toto schéma poskytuje doplňkové kódování vzorků a přidělování bitů. Umožňuje rámce různé velikosti. Kvůli relativní jednoduchosti byla tato vrstva vybrána pro DAB.

Vrstva Layer III byla již od počátku vytvářena pro nízké přenosové rychlosti. Je založena na algoritmu ASPEC. Má značně vylepšené kódování. Zvyšuje frekvenční rozlišení. Přidává nestejnorodé kvantování, dále adaptivní segmentaci a entropii kódování kvantovaných hodnot. Přes jeho relativní složitost byl vybrán, protože poskytuje nejlepší poměr kvalita/velikost - to bylo stěžejní pro přenos hudby přes ISDN linky. Chceme-li dosáhnout kvality srovnatelné s kvalitou Audio-CD musíme při použití jednotlivých vrstev volit rozdílný bit-rate. Použijeme-li Layer I je pro dosažení této kvality doporučen bit-rate 384 kb/s. Zvolíme-li Layer II „CD kvality“ je dosaženo mezi 192 až 256 kb/s. Layer III dosahuje takovéto kvality již při 128 kb/s. Z těchto hodnot je jasně patrné, že Layer III je nejvýhodnější. Jeho relativně vysoké nároky na výpočetní výkon, nejsou s rozvojem výpočetní techniky již moc relevantní. Nižší bit-rate je (při stejné kvalitě) výhodnější. Záznam na médiu zabere méně z jeho celkové kapacity (při vysílání nám postačí menší přenosová rychlost přenosové cesty). Následující tabulka znázorňuje „schopnosti“ formátu MPEG – 1 Layer III.

srovnatelná kvalita	mód	Bit-rate	kompresní poměr	max. frekvence
telefon	mono	8 kb/s	1:192	2 500 Hz
AM	mono	32 kb/s	1:48	7 500 Hz
FM	stereo	> 64 kb/s	1:24	12 000 Hz
CD	stereo	> 128 kb/s	1:12	> 16 000 Hz

Tab. 1: „Schopnosti“ formátu MPEG – Layer III

(Zdroj: <http://home.zcu.cz/~mtoman/mp3.htm>, 3.4.2007)

## 3.2 Stereo módy:

### 3.2.1 Dual channel (dual mono)

Bit-rate je rovnoměrně rozdělen mezi oba kanály. Tento mód se uplatní hlavně pro nezávislé kanály. Například komentář ve dvou různých jazycích.

### 3.2.2 Stereo

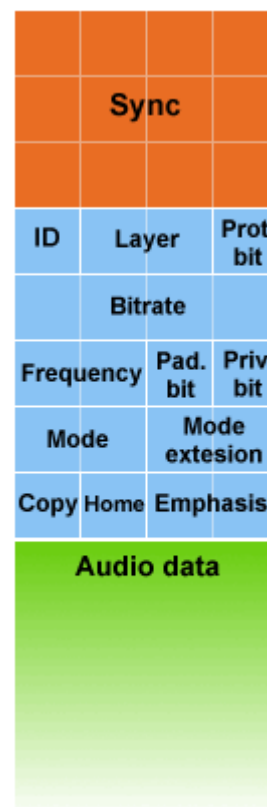
Rozdíl oproti dual channel je pouze v tom, že umožňuje některému kanálu přidělit větší bit-rate. Je-li jeden kanál beze zvuku, je možno (na jeho úkor) přidělit více bitů na kódování druhého kanálu. Tím se zvýší jeho kvalita.

### 3.2.3 Joint stereo

Tento mód využívá podobností v obou kanálech sterea. Enkodér vytvoří rozdílový (side) kanál a součtový (mid) kanál. Tento mód zvýší celkovou kvalitu na úkor částečné ztráty stereofonní informace. Při kódování stereofonního signálu tento režim zajistí znatelně kvalitnější výsledek, než režim dual channel (s rostoucím bit-ratem se rozdíl zmenšuje).

## 3.3 Rámce – Frames

Každý soubor typu MPEG Audio se dělí na jednotlivé rámce. Tyto rámce obsahují vždy hlavičku, část pro detekci chyb, audio-data, pomocná data. Velikost rámců není konstantní a končí hlavičkou dalšího rámce. Rámce jsou na sobě nezávislé – stačí 1 rámeček a dekodér je schopen jej reprodukovat. Rámce jsou základní součástí MP3 souboru.



Obr.7 Rámec MP3

### 3.3.1 Hlavička rámce

Obsahuje 4 bajty. Hlavičkou každá MP3 začíná a je umístěna na začátku každého rámce. ( Microsoft vytvořil formát, začínající jako normální WAV a teprve ve jeho části „data“ začíná první rámeček. To však není MP3 dle ISO/IEC. )

**AAAAAAAA AAABBCCD EEEFFFGH I IJJKLM**

<u>Znak</u>	<u>Délka</u>	<u>Umístění</u>	<u>Popis</u>
-------------	--------------	-----------------	--------------

	[bit]	[ bit – bit ]	
--	-------	---------------	--

<b>A</b>	11	31 – 21	Synchronizace rámce.
----------	----	---------	----------------------

Všechny bity musejí být nastaveny (“1“). Tím se odliší začátek hlavičky od ostatních dat.

<b>B</b>	2	20 – 19	ID verze MPEG audia
----------	---	---------	---------------------

00 – MPEG verze 2.5 (pozdější rozšíření MPEG – 2)

01 – Rezervováno

10 – MPEG verze 2 ( ISO/IEC 13818-3 )

11 – MPEG verze 1 ( ISO/IEC 11172-3 )

Verze MPEG 2.5 byla nedávno přidána ke standardu MPEG 2. Vytvořena byla pro potřebu velmi nízkých bit-ratů, umožňuje nižší vzorkovací frekvence. Jestliže dekodér toto rozšíření nepodporuje, je možno použít pro synchronizaci 12 bitů místo 11.

<b>C</b>	2	18 – 17	ID vrstvy (Layer)
----------	---	---------	-------------------

00 – Rezervováno

01 – Layer III

10 – Layer II

11 – Layer I

<b>D</b>	1	16	Protection bit – určuje zda sou v rámci kontrolní součty.
----------	---	----	---

0 – Chráněno CRC ( 16 bitů CRC následuje hlavičku )

1 – Nechráněno

Znak    Délka    Umístění    Popis

[bit]    [ bit – bit ]

E        4        15 – 12    ID bit-ratu

Bity	V1,L1	V1,L2	V1,L3	V2,L1	V2, L2, L3
0000	vf	vf	vf	vf	vf
0001	32	32	32	32	8
0010	64	48	40	48	16
0011	96	56	48	56	24
0100	128	64	56	64	32
0101	160	80	64	80	40
0110	192	96	80	96	48
0111	224	112	96	112	56
1000	256	128	112	128	64
1001	288	160	128	144	80
1010	320	192	160	160	96
1011	352	224	192	176	112
1100	384	256	224	192	128
1101	416	320	256	224	144
1110	448	384	320	256	160
1111	x	x	x	x	x

Tab. 2: Bit-rate ID

(Zdroj: [http://www.mp3-tech.org/programmer/frame\\_header.html](http://www.mp3-tech.org/programmer/frame_header.html), 23.3.2007)

Všechny hodnoty jsou v kb/s

V1 ( V2 ) – MPEG verze 1 ( 2 či 2.5 )

L1 ( 2 resp. 3 ) – Layer 1 ( Layer 2 resp. Layer 3 )

vf – volný formát. (konstantní bit-rate do 448 kb/s).

Dekodéry nejsou povinny tento „volný“ bit-rate podporovat.

x – nepřípustná hodnota

Soubory MPEG mohou mít tzv. proměnný bit-rate (VBR – variable bit-rate). Každý rámeček může být vytvořen s jiným bit-ratem. Tohoto můžeme využít ve všech vrstvách. Dekodéry podporující vrstvu Layer III musí tuto metodu podporovat (dekodéry nižších vrstev pouze mohou).

Pro vrstvu Layer II existuje nepovolená kombinace bit-ratu a módu stera. Následuje seznam kombinací. ANO označuje povolenou kombinaci, NE označuje kombinaci, která není povolena.

Znak    Délka    Umístění    Popis  
          [bit]    [ bit – bit ]

bit-rate	single channel	stereo	joint stereo	dual channel
free	ANO	ANO	ANO	ANO
32	ANO	NE	NE	NE
48	ANO	NE	NE	NE
56	ANO	NE	NE	NE
64	ANO	ANO	ANO	ANO
80	ANO	NE	NE	NE
96	ANO	ANO	ANO	ANO
112	ANO	ANO	ANO	ANO
128	ANO	ANO	ANO	ANO
160	ANO	ANO	ANO	ANO
192	ANO	ANO	ANO	ANO
224	NE	ANO	ANO	ANO
256	NE	ANO	ANO	ANO
320	NE	ANO	ANO	ANO
384	NE	ANO	ANO	ANO

Tab. 3: Kombinace bit-ratu a režimu sterea

(Zdroj: [http://www.mp3-tech.org/programmer/frame\\_header.html](http://www.mp3-tech.org/programmer/frame_header.html), 23.3.2007)

**F**        2        11 – 10    ID vzorkovací frekvence

bity	MPEG1	MPEG2	MPEG2.5
00	44 100 Hz	22 050 Hz	11 025 Hz
01	48 000 Hz	24 000 Hz	12 000 Hz
10	32 000 Hz	16 000 Hz	8 000 Hz
11	reserv.	reserv.	reserv.

Tab. 4: ID vzorkovací frekvence

(Zdroj: [http://www.mp3-tech.org/programmer/frame\\_header.html](http://www.mp3-tech.org/programmer/frame_header.html), 23.3.2007)

**G**        1        9        Padding bit

0 – rámeček není „vycpán“ (všechny bity využity pro data)

1 – rámeček je „vycpán“ jedním oddílem

„Vycpání“ se užívá pro přesné dosažení žádaného bit-ratu. Příklad: 128 kb/s, 44 100 Hz, Layer II, užívá mnoho 418 bitových a jen málo 417 bitových rámečků k tomu, aby zachovaly 128 kb/s. Pro vrstvu Layer I je místo pro padding 32 bitů. Pro Layer II a III je to 8 bitů.

**H**        1        8        Uživatelský bit - volný.

Má pouze informační význam.

Znak	Délka [bit]	Umístění [ bit – bit ]	Popis
I	2	7 – 6	Režim zpracování kanálů 00 – Stereo 01 – Joint stereo 10 – Dual channel ( 2 mono kanály ) 11 – Jeden kanál ( mono )

J	2	5 – 4	<p>Specifikace režimu zpracování kanálů (při joint stereu).</p> <p>Rozšíření režimu se užívá k připojení/nepřipojení informací, které neovlivňují stereo-efekt. Takto snižuje počet potřebných bitů. Tyto bity jsou dynamicky určovány kóděrem v joint stereo módu. Rozšíření může být v každém rámci jiné stejně jako aktivace či deaktivace režimu joint stereo. Celý frekvenční rozsah MPEG souborů je rozdělen do 32 pásem. Pro Layer I a II tyto bity určují pásma na něž je použito intensity stereo. Pro Layer III tyto bity určí, který se použije typ joint sterea (intensity stereo či m/s stereo).</p>
---	---	-------	---

Hodnota	Layer I & II	Layer III	
		intens. stereo	MS stereo
00	4 pásma	off	off
01	8 pásem	on	off
10	12 pásem	off	on
11	16 pásem	on	on

Tab. 5: ID specifikace zpracování kanálů (joint stereo)

(Zdroj: [http://www.mp3-tech.org/programmer/frame\\_header.html](http://www.mp3-tech.org/programmer/frame_header.html), 23.3.2007)

K	1	3	<p>Copyright</p> <p>0 – Audio není chráněno autorskými právy</p> <p>1 – Audio je chráněno autorskými právy</p>
---	---	---	--

L	1	2	<p>ID originálu</p> <p>0 – Kopie originálního média</p> <p>1 – Originál</p>
---	---	---	---

<u>Znak</u>	<u>Délka</u>	<u>Umístění</u>	<u>Popis</u>
	[bit]	[ bit – bit ]	
<b>M</b>	2	1 – 0	tzv. Emphasis 00 – neplatná kombinace 01 – 50 / 15 ms 10 – Rezervováno 11 – CCIT J.17

Tyto bity jsou zde proto, aby dekodér zjistil, který rámec musí být upraven (Dekodér musí ekvalizérem „srovnat“ zvuk (například po odstranění šumu způsobem Dolby -zvýrazní frekvence kolem 3 200 Hz )). Tato možnost není příliš užívána.

### 3.3.2 Datová část

Upravená data jsou kódována Huffmanovým kódováním. V této části, která následuje hlavičku, mohou být 2 druhy dat. Mohou být přítomna jednak pomocná data pro dekodér, ID3 TAG (verze 2) atp.. Musí však být přítomna data, která obsahují informaci o zvuku.

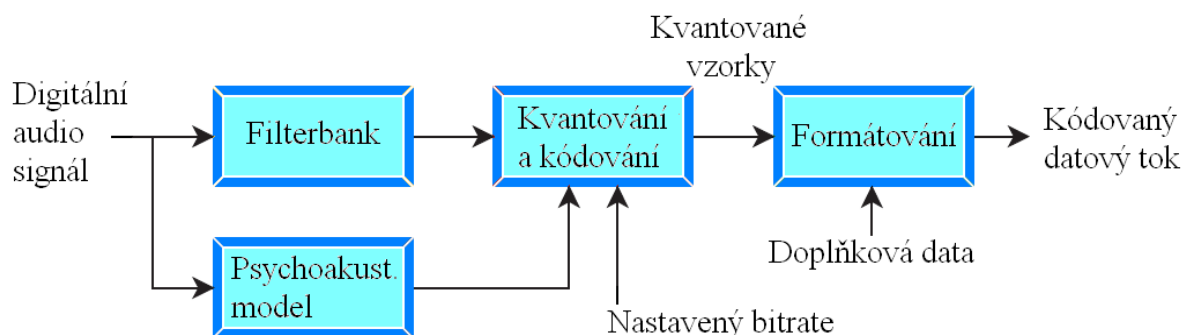
## 3.4 Kódování

Kódování MP3 je kompletně popsáno v normách ISO. Jako pomůcku obsahují tyto normy i popis vzorových kodérů. Tyto kodéry jsou odvozeny z originálních kodérů užívaných pro ověřovací testy. K realizaci kvalitního kodéru je zapotřebí mnoha znalostí a zkušeností. Množství investic nutných k vývoji vysoce kvalitního MPEG kodéru omezilo množství nezávisle vyvinutých implementací kodérů na minimum.

### 3.4.1 Algoritmus kódování

Obrázek 6 znázorňuje schéma typického kodéru MPEG – Layer III. Nejprve je zvuk rozdělen na několik podpásem pomocí filtrů. Souběžně je zpracován vstupní zvuk psychoakustickým filtrem, který určí úroveň maskovacího prahu. Poté jsou data z těchto bloků kvantována a kódována. Tento proces probíhá ve smyčkách, které

hlídají velikost a kvalitu (šum) výsledných dat. Nakonec jsou data zformátována do toku, který vyhovuje normě.



Obr. 8: Schéma postupu kódování

V bloku FILTERBANK je celé kmitočtové spektrum zvuku rozděleno do určitého počtu pásem (32). Ve vrstvě Layer III je použit hybridní typ. Je složen z kaskádního zapojení dvou typů filterbank. První typ je stejný jako u vrstev Layer I a II (tzv. polyphase) a druhý je tzv. Modified Discrete Cosine Transform (MDCT). MDCT typ má daleko vyšší frekvenční rozlišení. Rozdělení každého základního pásma z polyphase filtru na 18 podpásem zvýší potenciál k vyjmutí redundantních dat a tím i efektivity kódování. Dalším přínosem je lepší rozlišení maskovacích úrovní (zmenšení chyb). Paralelně s rozdělením do podpásem je použito tzv. PSYCHOAKUSTICKÉHO MODELU. Na této části se od dob vzniku odvedlo velké množství práce. Modul převede pomocí FFT (Fast Fourier Transformation – Rychlá Fourierova Transformace – tento typ výpočtu používá výpočetní technika k výpočtu fourierovy transformace.) signál z časové do frekvenční oblasti a s touto formou signálu poté pracuje. Výstup tohoto modulu představují hodnoty potřebné pro výpočet maskovacího prahu dané části tzv. SMR. Je-li možné udržet kvantovací šum pod maskovacím prahem každého pásma, pak by výsledek komprese neměl být lidským sluchem rozeznatelný od originálu. Dále se data upravují v modulu KVANTOVÁNÍ A KÓDOVÁNÍ. Tento modul je založen na dvou tzv. iteration loops. Inner loop kvantuje spektrální hodnoty, vypočtené v bloku filterbank a poté je kóduje. Kvantování je provedeno pomocí tzv. power-law kvantizeru. Ten automaticky kvantuje větší hodnoty s nižší přesností. Kvantované hodnoty jsou kódovány Huffmanovým kódováním. To je algoritmus používaný pro bezztrátovou kompresi dat. Převádí znaky vstupních hodnot na bitové řetězce různé délky. Hodnoty které se vyskytují nejčastěji jsou kódovány do bitových řetězců nejmenší délky (nejčastější



hodnota může být kódována do jednoho bitu). Naopak hodnoty, které se vyskytují velmi zřídka jsou kódovány do delších řetězců. Nejjednodušší realizací je vytvoření statistiky četnosti výskytu jednotlivých hodnot a její využití pro následné kódování. Je-li výsledný počet bitů větší než dovoluje bit-rate, zvětší se kvantovací krok a celý proces inner-loop se zopakuje. Outer loop hlídá, zda je zkreslení (kvantování šum) každého pásma je nižší než povolené (maskovací práh vypočtený psycho-akustickým modelem). Je-li zkreslení pásma větší provede se jeho zesílení (zvýšení tzv. scalefactor) a zopakuje se inner loop. Outer loop se ukončí v případě, že každé pásmo má menší kvantovací šum než maskovací práh. Tyto smyčky by měly být nastaveny tak, že skončí (spočítají se) v potřebném čase (při real-time kódování). To je důvod ladění psycho-akustického modelu a obou smyček pro různé bit-raty. Také se někdy zavádějí i další podmínky pro rychlejší ukončení smyček. V modulu FORMÁTOVÁNÍ je nutno uzpůsobit data tak, aby byla kompatibilní s datovým proudem dle MPEG. Datový proud je tvořen rámci, které obsahují hlavičku a data. Data jsou jak pomocná pro dekodér tak vlastní, nesoucí informaci (zakódována Huffmanovým kódováním). V tomto modulu je možno přidat ID3 TAG (viz dále).

Ne všechny kodéry byly vytvořeny stejně. Standardy MPEG nepředepisují způsob provedení kodérů. V extrémním případě by bylo možné kompletně se vyhnout použití psycho-akustického modelu, nepoužít scalefactor (tím i nutnost použití outer loop) a realizovat velmi jednoduchý inner loop. Takovýto kodér by byl velice rychlý, vyhovoval by standardu ISO. Měl by možná s některými zvukovými signály dobrou kvalitu, ale neobstál by při větší pestrosti kódovaných signálů. Zatímco tento kodér by byl snadno realizovatelný, mnohem obtížnější je vyvinout kodér, který nabízí vysokou kvalitu pro různé druhy hudby. Aktuální kodéry MP3 již nabízejí vysokou kvalitu i při nízkých datových tocích.

### 3.4.2 Kvalita kódování

Z výše popsaného procesu kódování plyne, že norma ISO nezaručuje kvalitu zvukového záznamu. Zvuková kvalita se liší v závislosti na různých aspektech. To jak na základních nastaveních, bit-ratu, ale i na sofistikovanosti jednotlivých kodérů. Rozdíly mezi výstupy různých kodérů jsou slyšitelné, i když jsou nastaveny stejné parametry komprese. Následují zvukové artefakty, které při kódování vznikají.

### 3.4.3 Běžné druhy artefaktů

Kodéry jsou nejvíce náchylné ke špatně zvolenému bit-ratu či ke špatně nastaveným parametrům. Takto vzniklé artefakty jsou většinou odlišné od běžného hluku či od zkreslení. Kodéry založené na psycho-akustickém modelu způsobují chyby popisované jako chyba měnící se v čase na jistých frekvencích. Výsledný zvukový záznam může mít následující následky: zkreslený (ne jako harmonické), hlučný pouze v jistém kmitočtovém pásmu, hrubý s problematickou kontrolou chyb (mění se často v čase).

#### 3.4.3.1 Ztráta šířky pásma

Jestliže kodér nedostane blok hudebních dat ke kódování (s požadovanou věrností) v rozmezí bit-ratu dojde k vyčerpání bitů. To může vést k narušení vysokofrekvenčního (hudebně) signálu. To je velice problematické. Šířka pásma se totiž mění (v nejhorším případě) s každým rámcem, což je horší než konstantně nižší šířka pásma.

#### 3.4.3.2 Před – ozvěny

Před-ozvěny jsou velice běžné artefakty u kodérů tohoto typu. Ačkoli je tento název dosti zvláštní, popisuje velice dobře podstatu problému. Signál zvuku se vyskytne dříve než událost, která tento zvuk způsobí.

#### 3.4.3.3 Dvojitý - hlas

Obzvláště při nízkých bit-ratech a nízkých vzorkovacích frekvencích je nevhodné spojení časového rozlišení kodéru a časovou strukturou signálu. Tento jev je nejvíce patrný při reprodukci hlasu a poslechu na kvalitní aparatuře či na sluchátkách. Jeden hlas má pak tendenci znít jako by byl zaznamenán dvakrát a poté překryt.

### 3.4.4 Měření kvality zvuku

Z principu existují tři metody: Poslechové testy, objektivní základní měření či měření na základě psycho-akustického modelu.

#### 3.4.4.1 Poslechové testy

Donedávna byly poslechové testy (správně řízené a se správným vybavením) jedinou dostupnou metodou, pro porovnání výkonů jednotlivých kodérů. Ve spolupráci s řadou „poskytovatelů“ rádiového vysílání a skupinou MPEG vyvinul ITU-R (International Telecommunication Union – Radio) velice propracovaný soubor norem pro provádění poslechových testů. Tyto testy jsou směřovány k tomu, aby kodéry prověřily v těch nejobtížnějších podmínkách (např. se kodér zkouší zvukem materiálů, který je velice obtížné zpracovat - cinknutí skleněných tyčinek o sebe. Poté se zhodnotí výkon pro tento konkrétní materiál. Tato procedura odhalí množství artefaktů (jsou-li přítomny).

#### 3.4.4.2 Objektivní základní měření

Uživatelé se stále snaží o získání přehledu o kvalitě kodéru z parametrů jako odstup signál-šum či šířka pásma dekodovaného signálu. Základním směrem ve vývoji kodérů založených na psycho-akustickém modelu je však zlepšování subjektivní kvality. Sledování pouze těchto parametrů nevyovídá o kvalitě kodéru. Tyto testy mohou být s úspěchem použity při vývoji, ale následně je třeba doladit kodér poslechovými testy.

#### 3.4.4.3 Techniky měření založené na psycho-akustickém modelu

Před více než 15-ti lety bylo zkoumáno využití psycho-akustických modelů k rozpoznání slyšitelnosti zvuků a kvality kodérů. Poslechové testy zatím nemohou nahradit, kvůli nedostatečnému technickému vybavení, ale jsou velice užitečné jako jejich doplněk. ITU-R zpracovala standardy pro měření na základě psycho-akustických modelů a poté vydala doporučení pro systém PEAQ (Perceptual Evaluation of Audio Quality).

#### 3.4.5 Vliv šířky pásma na kvalitu

Tato souvislost neakceptuje lidské vnímání zvuku. Přesto, že zdravé, odpočaté, mladé lidské ucho je schopno slyšet zvuk až o frekvenci 20 000Hz, není

prokazatelně schopno rozlišit nahrávku s pásmem do 20 kHz od té samé s pásmem do 16 kHz. To je důvod proč kodéry omezují frekvence nad 16 000 Hz. Výchyly od ideálně rovné charakteristiky přenosu (v závislosti na frekvenci) v nižších frekvencích jsou již značně slyšitelné. Zvukové artefakty nemají se šířkou pásma také mnoho společného a tak šířka pásma skutečně mnoho o kvalitě kodéru nevypovídá. Jsou-li bity použity pro lepší frekvenční rozsah (směrem nahoru) nebudou již schopny zpětně reprodukovat zvuk o nižších frekvencích. To se může projevit „chudým“ zvukovým projevem. Ještě horší situace nastává, když se situace mění z rámce na rámec (frame). Proto se zavádí omezení šířky pásma. Srovnáním kodéru s omezenou šířkou pásma a kodéru s mnohem vyšší šířkou pásma (při stejném bitratu) ve většině případů zjistíme, že kodér s omezenou šířkou pásma produkuje po dekomprimaci kvalitnější zvuk.

### 3.4.6 Ladění algoritmu pro různé přenosové rychlosti

Jak již bylo vysvětleno výše, dvě iterační smyčky (inner loop a outer loop) se nemusejí dokončit, jestliže špatně navážeme mezi sebou nastavení psychoakustického modelu a přenosovou rychlost k dispozici pro kódovaná data. Proto je nutno nastavit parametry psychoakustického modelu tak, aby se smyčky korektně ukončily. To může vyžadovat taková nastavení, která již zapříčiní slyšitelné rozdíly mezi původním a zakódovaným zvukem. Takováto nastavení jsou naneštěstí nutná a tak se nelze vyhnout slyšitelným rozdílům. Pro každou přenosovou rychlost musí být parametry psychoakustického modelu nastaveny jinak. Tomuto ladění parametrů psychoakustického modelu se věnovalo již značné úsilí.

### 3.4.7 Ukládání

Každý rámec je samostatný prvek. Lze jej vyjmout z řady rámců a přehrát (dekódovat) jej samostatně. Soubor MP3 je tak tvořen hlavičkou a řadou jednotlivých rámců. Každý rámec vrstvy Layer I obsahuje informace o 384 vzorcích, každý rámec vrstev Layer II a Layer III obsahuje informace o 1152 vzorcích. Rámec začíná synchronizačním slovem a končí před následujícím synchronizačním slovem.

Data zakódovaného audia, synchronizace dat a pomocná data mohou být uloženy společně na datovém médiu. Audio se bude mnohem snadněji upravovat, pokud se budou charakteristické body nacházet na adresovatelných místech

v paměti. Přístup na paměť může umožňovat dálkový přístup přes komunikační systém. Takovýto přístup může být zajištěn jinou jednotkou, než samotným dekodérem zvuku. Tato jednotka přijímá pokyny uživatele, čte data (ne audio) z média a posílá uložený bit-rate do dekodéru požadovanou rychlostí. Soubor je rozdělen na rámce a tak je možno posílat požadovanou rychlostí shluky dat do dekodéru.

### 3.5 Dekódování

Dekódování je proces při kterém se z kódovaného proudu dat (kódovaného) opětovně stane signál, který se dá s malou odchylkou považovat za stejný jako



Obr. 9 Schéma dekodéru MPEG Layer III

signál před kódováním. Dekodér je prostředek k provádění dekodování.

Do dekodéru vstupují data komprimovaného datového proudu. V bloku DEKÓDOVÁNÍ se provádí detekce chyb (je-li tato možnost použita při kódování) a proud dat je rozdělen na jednotlivé rámce (frames). V dalším bloku REKONSTRUKCE se obnoví amplitudová verze sady mapovaných vzorků. Poslední blok ZPĚTNÉ MAPOVÁNÍ převede tyto rekonstruované vzorky na PCM. To je již formát, který lze velmi jednoduše zpracovat. Tento postup zobrazuje schématicky obrázek.

## 4 ID3 TAG

Orientace v objemné sbírce zvukových dat je obtížná a uživatel má pouze několik možností, jak jednotlivé soubory odlišit. Přestože nové operační systémy umožňují zadávat dlouhé názvy souborů, není tato volba identifikace souborů právě nejlepší. Z tohoto důvodu byl vytvořen tzv. ID3 TAG - velmi dobrá vlastnost standardu MP3. ID3 TAG slouží k popsání daného souboru. Primárně jsou určeny k popisu audio-souborů, ale jsou dost dobře použitelné i na jiný druh souborů (při správné aplikaci). ID3 TAG je volitelný doplněk MP3 souboru a lze jej doplnit dodatečně, nezávisle na kódování. ID3 TAG má, podle normy, v souboru vždy své pevné místo (bez ohledu na to, zda je soubor uložen na médiu či vysílán). Dekodér není povinen podporovat ID3 TAGy.

### 4.1 ID3 TAG verze 1

Základní verze ID3 TAGu se nachází 128 B před koncem souboru. Tato verze je velice jednoduchá. ID3 TAG není povinný. Následuje jeho struktura:

<u>Pozice</u>	<u>Velikost</u>	<u>Popis</u>
[B - B]	[B]	
0 – 3	3	Hlavička TAGu - řetězec „TAG“
3 – 32	30	Titul
33 – 62	30	Interpret
63 – 92	30	Album
93 – 96	4	Rok vydání
97 – 126	30	Komentář
127 – 128	1	ID stylu hudby



Obr. 10 ID3 TAG v.1

### 4.2 ID3 TAG verze 2

Druhá verze popisku je daleko univerzálnější a dovoluje podrobněji popsat daný soubor. Daní za tyto výhody je větší složitost TAGu. Verze 2 není zpětně kompatibilní s verzí 1 a je umístěna na začátku souboru. Tato verze TAGu je vyvinuta tak, aby bylo možno následně doplňovat dodatečné informace o souboru. ID3 TAG verze 2 je proto navržen jako nádoba, v níž jsou umístěny bloky informací. Tyto bloky jsou nazývány rámce (frames). Každý rámeček začíná hlavičkou, která obsahuje předem

určený (jedinečný) kód – tzv. identifikátor rámce (jedinečný předdefinovaný identifikátor), velikost rámce (potřebnou k přeskočení nevyplněného rámce) a několik příznaků. Struktura ID3 TAGu verze 2 obsahuje: Hlavičku, rozšířenou hlavičku (volitelně), vlastní rámce, výplň (tzv. padding – volitelně) a zápatí (volitelně).

#### 4.2.1 Hlavička ID3 TAGu verze 2

Hlavička má velikost 10 B.

Identifikátor	Obsah	Poznámka
Značka	"ID3"	Velikost 3 B
Verze	\$ xx xx	např.: \$ 04 00 pro ID3v2.4.0
Příznaky	%ABCD0000	využitý je 4. až 7. bit
Velikost TAGu	4 * \$0xxxxxxx	

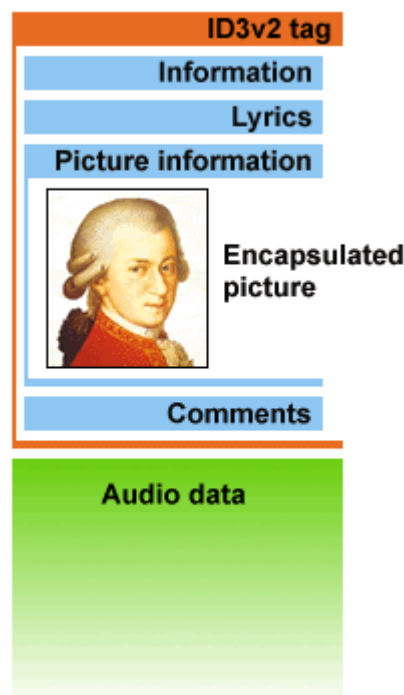
Tab. 6: Hlavička ID3 verze 2.4.0  
(Zdroj: <http://www.id3.org/id3v2.4.0-structure>)

První trojice bajtů je vždy nastavena na „ID3“ a identifikuje začátek hlavičky. Následující dvojice bajtů identifikuje použitou verzi (hlavní – tzv. major version a vedlejší – tzv. revision). Za touto dvojicí jsou umístěny příznaky TAGu. Verze ID3 TAG verze 2.4.0 využívá čtyř příznaků (umístění viz tabulka: Hlavička ID3 verze 2.4.0). Jejich význam je následující:

A - Unsynchronization  
Nastavení určuje, že je použita unsynchronizace (vysvětlení níže).

B - Extended header  
Bit označuje, zda se za hlavičkou nachází rozšířená hlavička. Nastavení bitu signalizuje přítomnost rozšířené hlavičky.

C - Experimental indicator



Obr. 11 ID3 TAG v. 2

Tento bit by měl být nastaven, pokud je TAG v experimentálním stavu vývoje.

- D - Footer present  
Je-li bit nastaven, signalizuje, že na konci TAGu je obsaženo zápatí.

Všechny ostatní bity musejí být vymazány, jinak se při zpracování nerozpoznají příznaky.

Velikost TAGu je popsána čtveřicí bajtů. Nejvýznamnější bit každého bajtu je nastaven vždy na "0" (tzv. systém synchsafe). Tak je využito celkem 28 bitů pro určení velikosti (velikost do 256 MB). Do této velikosti se nezapočítává hlavička ani zápatí (je-li připojeno). Celková velikost = zde uvedená velikost + 20 bajtů (10 B - hlavička a 10 B –zápatí).

ID3 TAG verze 2 můžeme detekovat dle následujícího systému:

\$ 49 44 33 xx xx yy zz zz zz zz

Kde xx označuje verzi, yy představuje příznaky a zz velikost TAGu.

#### 4.2.2 Extended header – „rozšířená hlavička“

Rozšířená hlavička obsahuje dodatečné informace, které mohou lépe nastínit strukturu TAGu. Není však zásadní pro správnou interpretaci informací obsažených v TAGu. Z tohoto důvodu je rozšířená hlavička volitelná.

Identifikátor	Obsah	Poznámka
Velikost	4 * %0xxxxxxx	Velikost celé rozšířené hlavičky
Počet příznakových bajtů	\$ 01	ID3v2.4.0 má jeden př. bajt
Rozšířené příznaky	%0ABC0000	

Tab. 7: Rozšířená hlavička ID3 verze 2.4.0  
(Zdroj: <http://www.id3.org/id3v2.4.0-structure>)

Jednotlivé příznaky jsou závislé na pořadí. Jsou s nimi totiž spojena data, která následují po sobě stejně jako bity příznaků. Data každého příznaku začínají tzv. délkovým bajtem. Ten může obsahovat hodnoty mezi 0 až 127. Následují data, jejichž velikost (počet bajtů) je shodná s číslem v délkovém bajtu. Jestliže příznak nemá připojená data je nastaven bajt označující jejich délku na 0.



- A - Tag update  
Je-li tento příznak nastaven, TAG je updatem TAGu nalezeného dříve v souboru či ve streamu. Tento příznak nemá připojena data – délka dat je \$ 00.
- B - CRC data present  
Nastavení tohoto příznaku ukazuje přítomnost CRC-32 (ISO - 3309) dat v rozšířené hlavičce. Kontrolní součet je počítán ze všech dat mezi hlavičkou a zápatím (tedy včetně paddingu – když je přítomen). Jak vyplývá z popisu je CRC-32 uloženo jako pětibajtové číslo (opět se systémem synchsafe) – délka dat je \$ 05 a vlastní forma 5 \* %0xxxxxxx.
- C - TAG restrictions  
Některé aplikace mohou vyžadovat omezení velikosti TAGu větší, než dle standardu ID3 verze 2 - délka dat je \$ 01. Omezení se řídí dle přiřazeného bajtu následujícím způsobem: % ddeffghh
- dd - TAG size restrictions – omezení velikosti TAGu  
00 - Ne více než 128 rámců a 1 MB (celková velikost)  
01 - Ne více než 64 rámců a 128 KB (celková velikost)  
10 - Ne více než 32 rámců a 40 KB (celková velikost)  
11 - Ne více než 32 rámců a 4 KB (celková velikost)
- e - Text encoding restrictions - omezení kódování textu  
0 - žádná omezení  
1 - string-y jsou kódovány dle ISO 8859-1 či UTF-8
- ff - Text fields size restrictions – omez. velikosti textových polí  
00 - žádná omezení  
01 - žádné string-y nejsou delší než 1024 znaků  
10 - žádné string-y nejsou delší než 128 znaků  
11 - žádné string-y nejsou delší než 30 znaků  
Toto omezení nic neříká o počtu bajtů, které budou použity pro reprezentaci těchto znaků (záleží na kódování). Je-li v textovém poli více řetězců, omezení se týká jejich sumy.
- g - Image encoding restrictions – omezení kódování obrázků  
0 - žádná omezení

- 1 - Obraz kódován jen pomocí PNG či JPEG
- hh - Image size restrictions – omezení velikosti obrázku
  - 00 - žádná omezení
  - 01 - všechny obrázky 256 x 256 pixelů či menší
  - 10 - všechny obrázky 64 x 64 pixelů či menší
  - 11 - všechny obrázky přesně 64 x 64 pixelů (pokud není požadováno jinak)

#### 4.2.3 Padding – „vycpávání“

Volitelně lze vložit „vycpávku“ do místa za posledním rámcem (na konci ID3 TAGu). Je nutno s tím počítat při zadávání velikosti v hlavičce TAGu. Účel této metody spočívá v možnosti dodatečně přidat (či zvětšit) rámce a to bez nutnosti přepisovat celý soubor. Takováto „výplň“ musí mít hodnoty \$ 00. TAG nesmí mít „výplň“ mezi rámci či mezi hlavičkou TAGu a rámcem. Navíc nesmí být tato metoda použita, je-li přítomno na konci TAGu zápatí.

#### 4.2.4 Footer – zápatí

Urychluje nalezení TAGu při hledání od konce souboru. Zápatí je vyžadováno, je-li TAG připojen až za audio data (na konec souboru). Zápatí je přesnou kopií hlavičky, ale nemá identifikátor „ID3“, ale „3DI“.

#### 4.2.5 Přehled rámce ID3 TAGu

Všechny rámce ID3 verze 2 se skládají z hlavičky rámce následovaného jedním či více poli, které obsahují informace. Hlavička má vždy velikost 10 bajtů a její struktura je následující:

<b>Identifikátor</b>	<b>Obsah</b>	<b>Poznámka</b>
ID rámce	\$ xx xx xx xx	
Velikost	4 * %0xxxxxxx	synchsafe
Příznaky	\$ xx xx	

Tab. 8: Rámec ID3 verze 2  
(Zdroj: <http://www.id3.org/id3v2.4.0-structure>)

ID rámce je složený z velkých písmen (A – Z) a číslic (0 – 9). Identifikátory, které začínají písmeny „X“, „Y“ a „Z“ jsou experimentální rámce (k užití pro každého uživatele), bez nutnosti nastavovat experimentální bit v hlavičce TAGu. Po ID rámce následuje údaj o velikosti dat, obsažených v rámci (po šifrování, kompresi i unsynchronizaci). Do této velikosti není započítána hlavička (10 B). Velikost je poté uložena jako 32 bitové číslo (synchsafe). Po velikosti následuje pole příznaků (2 B).

#### 4.2.5.1 Příznaky hlavičky rámce ID3 TAGu

Veškeré nevyužité příznaky musejí být vymazány. První bajt je určen pro stavové příznaky, druhý bajt pro popis formátu. Některé příznaky mají přiřazená data. Jejich pořadí je závislé (data se uloží za hlavičku v pořadí jako mají jejich příznaky).

**% 0ABC 0000      % 0D00 EFGH**

- A - TAG alter preservation  
Tento příznak dekodéru TAGů říká, co dělat s rámcem, jehož identifikátor je neznámý a je jakýmkoli způsobem měněn.
  - 0 - rámec by měl být zachován
  - 1 - rámec by měl být odstraněn
- B - File alter preservation  
Příznak oznamuje dekodéru TAGů, co dělat s rámcem, jehož identifikátor je neznámý a soubor mimo TAGu je měněn.
  - 0 - rámec by měl být zachován
  - 1 - rámec by měl být odstraněn
- C - Read only  
Tento příznak říká dekodéru, je-li obsah tohoto rámce pouze ke čtení. Chceme-li informace v rámci měnit musí být příznak vynulován.
- D - Grouping identity  
Tento příznak signalizuje, zda rámec patří do nějaké skupiny rámců. Je-li nastaven, je k hlavičce přidán ještě identifikační bajt skupiny rámců.
  - 0 - rámec neobsahuje skupinový údaj
  - 1 - rámec obsahuje skupinový údaj

- E - Compression
- Tento příznak ukazuje, je-li rámeček komprimován. Komprimace je provedena metodou zlib. Tato metoda vyžaduje vstupní parametr tzv. „data length indicator“, ten musí být zahrnut v rámci (v případě použití komprese).
- 0 - rámeček není komprimován
- 1 - rámeček je komprimován
- F - Encryption
- Nastavením tohoto příznaku říkáme dekodéru TAGů, že rámeček je šifrován. Šifrování by mělo být provedeno po případné kompresi. Existují různé druhy šifrování, které lze využít.
- G - Unsynchronization
- Příznak signalizuje, jestli je použito na tento rámeček unsynchronizace. Je-li příznak nastaven, všechna data od konce hlavičky do konce rámce jsou unsynchronizována.
- H - Data length indicator
- Tento příznak signalizuje (je-li nastaven), je-li připojen k rámci „data length indicator“. Je to hodnota, které je popsána jako velikost rámce, kdyby byly všechny formátové příznaky nastaveny na nulu. Přiřazená data – synchsafe 32 bitů.

Pořadí rámečků není pevně dané, ale je velice žádoucí, aby byly rámce uspořádány za sebou podle důležitosti informace, které obsahují. TAG musí obsahovat alespoň jeden rámeček. Rámeček musí být minimálně 1 bajt velký – bez hlavičky. Jestliže není řečeno jinak, jsou řetězce kódovány jako ISO 8859-1 (znaky v rozsahu \$20 - \$FF a oddělovač řádku \$ 0A). Takové řetězce jsou v popisech rámce reprezentované jako <text string> či <full text string>. Rámce, které umožňují použít jiné kódování textu, obsahují bajt identifikace kódování. Možná kódování:

- \$ 00 - ISO-8859-1 [ISO-8859-1]. Ukončeno \$ 00.
- \$ 01 - UTF-16 [UTF-16] kódování Unicode [UNICODE] s BOM. Všechny řetězce rámce budou mít stejné pořadí bajtů. Ukončeno \$ 00 00.

- \$ 02 - UTF-16BE [UTF-16] kódování Unicode [UNICODE] bez BOM.  
Ukončeno \$ 00 00.
- \$ 03 - UTF-8 [UTF-8] kódování Unicode [UNICODE]. Ukončeno \$ 00.

#### 4.2.5.2 Timestaps – časové značky

Časové značky jsou odvozeny od ISO 8601. Kompletní popis času vypadá následovně: „YYYY“-„MoMo“-„DD“T„HH“:„MiMi“:„SS“

Tuto přesnost lze snížit následujícími způsoby: YYYY, YYYY-MoMo, YYYY-MoMo-DD, YYYY-MoMo-DDTHH a YYYY-MoMo-DDTHH:MiMi.

#### 4.2.5.3 Languages - jazyky

Pro rozlišení jazyku, jímž jsou napsány informace v rámci, se užívá tři bajtů. Značení jazyků odpovídá normě ISO 639-2 (čeština – cze či ces). Není-li jazyk znám, je používán řetězec „XXX“.

#### 4.2.5.4 Odkazy

Všechny odkazy mohou být relativní. („pic1.png“ či „../pic2.png“)

#### 4.2.6 Umístění TAGu

ID3 TAG verze 2 je umístěn před audio data, aby mohl přehrávač využít informace z něj při streamování (např. vysílání internetových rádií). Je však také možné připojit jej za audio data, nebo udělat kombinaci před a po.

#### 4.2.7 Unsynchronization

Jediný účel unsynchronizace je přizpůsobit ID3 TAG existujícímu softwaru a hardwaru. Unsynchronizace je zbytečná, budou-li soubory zpracovávány výhradně softwarem a hardwarem, který podporuje ID3 TAGy verze 2. Kdykoli je detekována chybná synchronizace (uvnitř TAGu), je vložen nulový bajt. Vkládání pomocných bajtů probíhá následovně: je-li nalezen řetězec - %11111111 111xxxxx měl by být nahrazen tímto řetězcem - %11111111 00000000 111xxxxx. Jsou-li všechny rámece v TAGu unsynchronizovány, měl by být nastaven příznak v hlavičce TAGu. Obsahuje-li TAG alespoň jeden rámeček, který není unsynchronizován, nesmí být tento

příznak nastaven. Je zvyklostí mít celý TAG ošetřen stejně (celý unsynchronizován či nikoli). Jestliže je použito při kódování šifrování či komprese je potřeba použít unsynchronizaci až nakonec. Při dekódování je proces opačný – nejprve se musí provést unsynchronizace (opačná).

#### 4.2.8 Metoda synchsafe

Někdy je nevhodné použít unsynchronizace, protože velikost výsledných dat není předem známa. To je problémem především při práci se „size deskriptor-y“ (např. velikost rámců). Tento problém lze řešit metodou synchsafe. Takto ošetřená čísla nemohou nikdy nabýt takové hodnoty, která by musela být upravena metodou unsynchronization. Systém synchsafe má každý nejvýznamnější bit každého bajtu nastaven na nulu. Je tedy přístupno 7 z 8 bitů každého bajtu (32 bitové číslo má tedy 28 informačních bitů).

#### 4.2.9 Typy rámců:

AENC - Audio encryption	RBUF - Recommended buffer size
APIC - Attached picture	RVAD - Relative volume adjustment
COMM - Comments	RVRB - Reverb
COMR - Commercial frame	SYLT - Synchronized lyric/text
ENCR - Encryption method registration	SYTC - Synchronized tempo codes
EQUA - Equalization	TALB - Album/Movie/Show title
ETCO - Event timing codes	TBPM - BPM (beats per minute)
GEOB - General encapsulated object	TCOM - Composer
GRID - Group identification registration	TCON - Content type
IPLS - Involved people list	TCOP - Copyright message
LINK - Linked information	TDAT - Date
MCDI - Music CD identifier	TDLY - Playlist delay
MLLT MPEG - location lookup table	TENC - Encoded by
OWNE - Ownership frame	TEXT - Lyricist/Text writer
PRIV - Private frame	TFLT - File type
PCNT - Play counter	TIME - Time
POPM - Popularimeter	TIT1 - Content group description
POSS - Position synchronisation frame	TIT2 - Title/songname/cont. description

TIT3 - Subtitle/Description refinement	TRSO - Internet radio station owner
TKEY - Initial key	TSIZ -Size
TLAN - Language(s)	TSRC - ISRC
TLEN - Length	TSSE - Sw/Hw and settings for encod.
TMED - Media type	TYER - Year
TOAL - Original album/movie/show title	TXXX - User defined text info. frame
TOFN - Original filename	UFID - Unique file identifier
TOLY - Original lyricist(s)/text writer(s)	USER - Terms of use
TOPE - Original artist(s)/performer(s)	USLT – Unsynchro. text transcription
TORY - Original release year	WCOM - Commercial information
TOWN - File owner/licensee	WCOP - Copyright/Legal information
TPE1 - Lead performer(s)/Soloist(s)	WOAF - Official audio file webpage
TPE2 -Band/orchestra/accompaniment	WOAR- Official artist webpage
TPE3 -Conductor/performer refinement	WOAS - Official audio source webpage
TPE4 - Interpreted, rmxed, modified by	WORS - Official inet rad.st.homepage
TPOS - Part of a set	WPAY - Payment
TPUB - Publisher	WPUB - Publishers official webpage
TRCK - Track number/Position in set	WXXX - User defined URL link frame
TRDA - Recording dates	
TRSN - Internet radio station name	

#### 4.2.9.1 Textové informační rámce

Rámce, které začínají na písmeno T jsou všechny stejného druhu. Obsahují textové informace všeho druhu. Tyto rámce mají stejnou strukturu, která se dá vyjádřit následovně:

<b>Identifikátor</b>	<b>Obsah</b>	<b>Poznámka</b>
Hlavička rámce	Txxx	TXXX je rezervováno
Kódování textu	\$ xx	viz výše
Vlastní informační text	Kódovaný text	

Tab. 9: Schéma textového typu rámce  
(Zdroj: <http://home.zcu.cz/~mtoman/mp3.pdf>)

#### 4.2.9.2                    COMM

Nejedná se o textový rámec a tak má odlišnou strukturu. Je to komentář.

<b>Identifikátor</b>	<b>Obsah</b>
Hlavička rámce	COMM
Kódování	\$ xx
Jazyk	\$ xx xx xx
Zkrácený popis	kódovaný text zakončený \$ 00
Vlastní popis	kódovaný text

#### 4.2.9.3                    URL rámce

Tab. 10: Schéma rámce typu komentář  
(Zdroj: <http://home.zcu.cz/~mtoman/mp3.pdf>)

Tento typ rámců začíná písmenem W a všechny mají opět stejnou strukturu.

<b>Identifikátor</b>	<b>Obsah</b>	<b>Poznámka</b>
Hlavička rámce	Wxxx	WXXX je rezervováno
Odkaz či adresa	textový řetězec	

Tab. 11: Schéma rámce typu URL  
(Zdroj: <http://home.zcu.cz/~mtoman/mp3.pdf>)

#### 4.2.9.4                    UFID

Tento rámec je tzv. jedinečný identifikátor souboru. Je určený pro orientaci ve velkém počtu souborů. Struktura tohoto rámce je zobrazena v tabulce.

<b>Identifikátor</b>	<b>Obsah</b>	<b>Poznámka</b>
Hlavička rámce	UFID	
Identifikace vlastníka	textový řetězec	ukončený \$ 00
ID souboru	binární kód	až 64 bitů

Tab. 12: Schéma rámce typu UFID  
(Zdroj: <http://home.zcu.cz/~mtoman/mp3.pdf>)



## **5 POROVNÁNÍ VLIVU ATRIBUTŮ KÓDOVÁNÍ NA KVALITU**

Na přiloženém DVD je několik hudebních souborů, zakódovaných kódérem LAME. Pro kódování každého z nich jsem nastavil odlišné parametry (bit-rate, vzorkovací frekvenci a režim stera). Tyto soubory jsem poté poslechem porovnával a hodnotil. Pro porovnávání jsem použil nahrávku skladby Vltava (Bedřich Smetana – Má vlast). Tato skladba je velice pestrá. Obsahuje jak velmi tiché, tak velmi hlasité, značně složité pasáže. Tuto nahrávku mám navíc „naposlouchanou“, což mi velice pomohlo při hodnocení. V následujících řádcích shrnuji své poznatky.

Existuje mnoho kódérů založených na základě původního algoritmu MPEG. Liší se rychlostí, ale hlavně kvalitou výstupních zvukových dat. Nejlepšími kodéry jsou asi takové, které jsou založeny na Fraunhoferově kodeku. Pravděpodobně nejkvalitnější je kódér LAME. Je to otevřený projekt, ale dosáhl velmi vysoké kvality. Navíc je k dispozici zdarma. Tento kodek je odvozen od standardního ISO algoritmu (Fraunhoferova) a je značně inovován. Dnes používá vlastní psycho-akustický model – GPSYCHO. Jeho kód je optimalizován pro využití funkcí moderních procesorů jako například MMX či 3DNow.

Mezi parametry nastavitelné v kodéru patří:

- Bit-rate - kodéry umožňují nastavit požadovaný bit-rate. Nastavuje se dle vlastního uvážení.
- VBR - možnost zapnout funkci variabilního bit-ratu. Tato možnost mnohdy nepřináší výrazné zlepšení při zachování velikosti výsledného souboru.
- Vzorkovací frekvence - nastavení požadované vzorkovací frekvence. Nastavuje se většinou stejná jakou je vytvořen záznam, který chceme kódovat.
- Kvalita - tato volba pravděpodobně ovlivňuje parametry podle nichž pracují iteration loops kodéru. Nastavíme-li vysokou kvalitu, proces kódování trvá déle, ale výsledný zvuk je kvalitnější. Ve všech případech byla nastavena nejvyšší kvalita.
- Režim stera - většinou lze nastavit jakýkoli ze všech režimů zpracování stera (dual channel, stereo, joint stereo či mono).

Nejdůležitějšími parametry jsou bezpochyby vzorkovací frekvence a bit-rate. Vzorkovací frekvence je většinou dána zdrojem zvukových dat (jakým způsobem byla

data nahrána). Změna vzorkovací frekvence (tzv. resampling) je velice náročný proces, který (v reálném čase) plně vyčerpá i moderní DSP. Z tohoto důvodu se vzorkovací frekvence nastaví podle zdrojových dat. Vzorkovací frekvence ovlivní jak vysoké frekvence bude ještě záznam obsahovat (které se zaznamenají). Při nastavení bit-ratu se nemusíme řídit prakticky ničím a tak si můžeme nastavit bit-rate libovolně. To má ovšem vliv na kvalitu následně reprodukováného zvuku. Pro kvalitu srovnatelnou s kvalitou CD se doporučuje použít bit-rate vyšší než 128 kb/s (při stejné vzorkovací frekvenci). Při nižším bit-ratu se musí odstranit mnoho dat a na výsledném zvuku je tento zákrok již znát. V nízkých bit-ratech navíc kodéry ořezávají zvuk nad 16 kHz.

Chceme-li použít nízké přenosové rychlosti je velice výhodné použít režimy sterea. A to buď mono, či joint stereo. Tím se kvalita (při stejném bit-ratu) výrazně zvýší. Při vyšších bit-ratech již není přínos těchto režimů tak znatelný.

Velice diskutabilní volbou je použití variabilního bit-ratu. Tato metoda analyzuje vstupní data a nastavuje bit-rate průběžně dle potřeby. Metoda VBR není doposud (ve většině kodérů) uspokojivě zvládnuta a dosahuje (paradoxně) horších výsledků než konstantní bit-rate (CBR 192kb/s a VBR s průměrným bit-ratem 192kn/s).

Na přiloženém mini-DVD se nachází zvukové ukázky. K jejich kódování jsem použil kodér LAME verze 3.97 (32 bit). Soubory jsou popsány dlouhým jménem dle následujícího systému: Má vlast, Vltava – bit-rate – režim sterea – vzorkovací frekvence. Bit-rate jsem volil mezi 32 a 320 kb/s a vzorkovací frekvenci 44 100 Hz a 20 500 Hz.

Z kvality zvuku, poskytovaného soubory na DVD, je možno usoudit, že výrazný vliv při stejném bit-ratu má jak režim sterea, tak vzorkovací frekvence. Tento vliv je nejvíce zřejmý při nízkém bit-ratu v hlasitější a složitě (hodně nástrojů najednou) části nahrávky.

V následujícím odstavci se pokusím popsat moje subjektivní dojmy z poslechu nahrávek. V popisu zvuku jsem nucen použít mnohá přirovnání, protože pro popis těchto věcí neexistují technické termíny. Doufám, že bude můj popis srozumitelný.

Zvuk MP3 s bit-ratem 32 kb/s je velice podobný jak v režimu stereo tak joint stereo. Je znatelně zkreslený a slitý (nelze odlišit jednotlivé nástroje) s výrazným rušením o vysokých frekvencích (nepříjemný šelest). Režim mono sníží nepříjemné vysoké rušení, ale zvuk je stále velice slitý. Při poloviční vzorkovací frekvenci je zvuk

v režimu joint stereo srovnatelný se zvukem režimu mono s plnou vzorkovací frekvencí. Zvuk souboru s bit-ratem 48 kb/s je velice podobný. Rušení vysokými nepříjemnými zvuky je menší, ale nikterak výrazně. V režimu mono již není rušení tolik patrné. Nastavíme-li bit-rate na hodnotu 64 kb/s lze již v režimu joint stereo slyšet typické rušení jen zřídka. Začínají se objevovat zvuky trianglu. Zvuk je ale ještě pořád dosti slitý (již o poznání méně, než při nižších bit-ratech). Režim mono ještě zlepšil přednes vyšších kmitočtů, rušení již téměř není slyšet. Při bit-ratu 96 kb/s již jsou jasně slyšet údery do trianglu a nepříjemné rušení již nelze významně pozorovat. Zvuk ale zdaleka není stejný jako originál. Je pořád dosti slitý a neklidný. Při 128 kb/s je zvuk již velmi podobný originálu. Rušení se vůbec neobjevuje. Vysoké tóny jsou podány vcelku dobře. Začíná se objevovat i nejspodnější basová linka, která výrazně změkčí zvuk (k lepšímu). Režimy sterea již nelze moc dobře odlišit. Bit-rate 256 kb/s se jeví zvukově jako originální nahrávka. Každé cinknutí trianglu je zcela jasně slyšet. Nástroje hrají každý sám (neslévají se jejich zvuky - jsou slyšet jednotlivé údery na klavír) a znějí tak, jak znějí v původní nahrávce. Zřetelně jsou slyšet i zvuky na pozadí. Spodní tóny jsou prokreslené a měkké (přesto v případě bubnu rázné a rychlé). Nastavíme-li při kódování vysoký bit-rate (nad 128 až 256 kb/s), rozdíly ve zvuku již nelze rozeznat. (s mým vybavením (zdroj: M-audio Revolution 5.1 s o.z. AD826 a AD8260, zesilovač: Harman/Kardon HK675 a repro JM-Lab Focal Electra 926)). Z tohoto důvodu jsem na DVD neumístil všechny kombinace bit-rate, režimu sterea a vzorkovací frekvence.

Formát MP3 není schopen kvalitně komprimovat audiosignál do bit-ratu 64 kb/s. Při větším bit-ratu je již kvalita přijatelná.

## ZÁVĚR

MPEG Layer III (MP3) je pokročilý standard kódování digitálního zvukového záznamu. Využívá se pro výraznou úsporu potřebné úložné kapacity či rychlosti přenášených dat. Ke kódování využívá psycho-akustického modelu. Psycho-akustický model určuje, které zvuky mohou být ze záznamu odstraněny, aniž by zvuk opětovně dekódovaného záznamu byl výrazně odlišný od původního záznamu. Psycho-akustický model zohledňuje lidské vnímání zvuku a využívá nedokonalostí, typických pro lidský sluch. K odstranění redundantních dat se používá frekvenčního maskování, časového maskování, omezení maximální frekvence a neschopnosti lidského sluchu lokalizovat zdroje především hlubokých tónů. Takto upravený záznam je navíc uložen takovým způsobem, který ještě více zmenší jeho výslednou velikost. K tomu se využívá Hoffmanova kódování.

Soubory MP3 mohou být navíc doplněny dodatečnými informacemi. Tyto informace popisují daný soubor a ulehčují orientaci mezi množstvím souborů.

Kódování MP3 je velice náročný proces, vyžadující nemalého výpočetního výkonu. Navíc je vývoj kodéru velice složitý a tak vyžaduje velké množství znalostí, času i prostředků. Proto se téměř žádné originálně vyvinuté kodéry neobjevily. Standard je popsán normou ISO/IEC 11172-3 (MPEG – 1) či ISO/IEC 13818-3 (MPEG – 2), ve které lze najít všechny specifikace i příklady aplikací. Tento standard však není otevřený. Při vývoji tohoto typu kódování bylo zaregistrováno množství patentů. Bez využití alespoň některých není možné kodér či dekodér plně implementovat. Použitím jiných než specifikovaných algoritmů či vypuštěním některých rysů standardu lze některé patenty obejít. Avšak některé z patentů jsou pro implementaci zcela nezbytné a tak je vývojář musí použít a samozřejmě také zaplatit. Většinu těchto patentů vlastní společnost Fraunhoferský institut/Thomson a za jejich použití vyžaduje licenční poplatky. Ceník příslušných sazeb je na stránkách [www.mp3licensing.com](http://www.mp3licensing.com).

Existuje i mnoho jiných kódovacích postupů, které dosahují obdobných výsledků a jejich použití není omezeno (je to volný formát). Formáty MP3, MP3pro, AAC a MP2 jsou všechny ze stejné rodiny (MPEG) a jsou zpoplatněny podobným způsobem (např. zde: [http://www.vialicensing.com/Licensing/MPEG4\\_fees.cfm?product=MPEG-4AAC](http://www.vialicensing.com/Licensing/MPEG4_fees.cfm?product=MPEG-4AAC) je „ceník“ poplatků AAC). Mnoho patentů, souvisejících

s tímto formátem, je vlastněno společnostmi: Fraunhofer institut, Dolby Labs, Sony, Thomson a AT&T. Formáty, které dosahují podobných výsledků, jsou například: Ogg Vorbis či WMA. Při nízkých bit-ratech dosahují modernější kodeky mnohem lepších výsledků. Pokud jde o kvalitu, nejvýhodněji se jeví použít bezztrátové algoritmy kódování. Tyto kodéry jsou schopny zkomprimovat zvuková data takovým způsobem, že po opětovném dekódování nebudou z dat „ořezány“ některé informace. Mezi takovéto kodeky patří například: ALAC, FLAC či Monkey's audio.

Přesto je standard MP3 nejrozšířenějším formátem kódování digitálního zvukového záznamu. Z tohoto důvodu je podporován výrobci elektroniky i vývojáři softwaru na celém světě (přehrávače jak ve formě přístroje tak ve formě softwaru...). V příloze je k dispozici tabulka několika dalších formátů (informací o nich) a tabulka jejich technických detailů.

## **Seznam použité literatury:**

- [1] Informace o formátu MP3  
<<http://www.mp3-tech.org/>>
- [2] Informace o formátu MP3  
<<http://www.all4mp3.com/>>
- [3] Informace o formátu MP3  
<<http://home.zcu.cz/~mtoman/mp3.htm>>
- [4] Informace o formátu MP3  
<<http://le-hacker.org/hacks/mpeg-drafts/11172-3.pdf>>
- [2] Informace o formátu MP3 a poplatcích spojených s jeho aplikací  
<<http://www.mp3licensing.com/>>
- [4] Informace o ID3 TAGu  
<<http://www.id3.org/>>
- [5] Všeobecné informace  
<<http://cs.wikipedia.org/>>  
<<http://en.wikipedia.org/>>

## Seznam tabulek:

Tabulka 1: „Schopností“ formátu MPEG – Layer III	27
Tabulka 2: Bit-rate ID	29
Tabulka 3: Kombinace bit-ratu a režimu sterea	30
Tabulka 4: ID vzorkovací frekvence	30
Tabulka 5: ID specifikace zpracování kanálů (joint stereo)	31
Tabulka 6: Hlavička ID3 verze 2.4.0	40
Tabulka 7: Rozšířená hlavička ID3 verze 2.4.0	41
Tabulka 8: Rámec ID3 verze 2	43
Tabulka 9: Schéma textového typu rámce	48
Tabulka 10: Schéma rámce typu komentář	49
Tabulka 11: Schéma rámce typu URL	49
Tabulka 12: Schéma rámce typu UFID	49
Tabulka 13: Tabulka některých formátů kódování zvuku a stručné info. o nich	Př.1
Tabulka 14: Tabulka některých formátů a jejich technické parametry	Př.2

## Seznam obrázků:

Obr.1: Křivka ATH	10
Obr. 2: Frekvenční maskování	11
Obr. 3: Vzorkování signálu	13
Obr. 4: Nevhodné vzorkování signálu	13
Obr. 5: Profesor Dieter Seitzer	19
Obr. 6: Karlheinz Brandenburg	19
Obr. 7: Rámec MP3	27
Obr. 8: Schéma postupu kódování	33
Obr. 9: Schéma dekodéru MPEG Layer III	38
Obr. 10: ID3 TAG v.1	39
Obr. 11: ID3 TAG v. 2	40



## **Seznam zkratek:**

3DNow	-	multimediální rozšíření funkční sady procesorů firmy AMD
AAC	-	Advanced Audio Coding
ALAC	-	Apple Lossless Audio Codec
AD	-	Analog Devices
ASPEC	-	Adaptive Spectral Perceptual Entropy Coding
ATH	-	Absolute Treshold of Hearing
AT&T	-	American Telephone & Telegraph Company
A/D	-	Analog / Digital
BOM	-	Byte Order Mark
CCETT	-	Centre commun d'études de télévision et télécommunications
CD	-	Compact Disc
CRC	-	Cyclic Rredundancy Check
DAB	-	Digital Audio Broadcasting
DCC	-	Digital Compact Cassette
DSP	-	Digital Signal Processor
DVB	-	Digital Video Broadcasting
DVD	-	Digital Video Disc
D/A	-	Digital / Analog
EPG	-	Electronic Program Guide
FFT	-	Fast Fourier Transformation
FLAC	-	Free Lossless Audio Codec
HDTV	-	High Definition TV
HD DVD	-	High Definitoion Digital Video Disc
IBM	-	International Business Machines Corporation
ID3 TAG	-	popisek souboru MP3
IEC	-	International Electrotechnical Commission
IRT	-	Institut für RundfunkTechnik GmbH
ISDN	-	Integrated Services Digital Network
ISO	-	International Organization for Standardization
ITU-R	-	International Telecommunication Union – Radio
JPEG	-	Joint Photographic Experts Group

LAME	-	Lame Ain't an Mp3 Encoder
LC-ATC	-	Low Complexity - Adaptive Transform Coding
MDCT	-	Modified Discrete Cosine Transform
MMX	-	MultiMedia eXtension
MP3	-	MPEG - 2 Audio Layer III
MPEG	-	Moving Pictures Expert Group
MUSICAM	-	MPEG - 1 Audio Layer II (MP2)
OFC	-	Optimum Coding in Frequency Domain
PCM	-	Pulse Code Modulation
PEAQ	-	Perceptual Evaluation of Audio Quality
PNG	-	Portable Network Graphics
RIFF	-	Resource Interchange File Format
SMR	-	Signal-to-Mask Ratio
UTF	-	Unicode Transformation Format
VBR	-	Variable Bit Rate
WAV	-	Waveform audio format
WMA	-	Windows Media Audio

## **Seznam příloh:**

Příloha 1

Tabulka 13: Tabulka některých formátů kódování zvuku a stručné informace o nich

Příloha 2

Tabulka 14: Tabulka některých formátů a jejich technické parametry

Mini – DVD: Ukázky kvality kódování za daných parametrů

Bakalářská práce ve formátu - .doc

Ukázka části normy 11172 – 3



Kodek	Autor	Datum uvedení	Poslední verze	Cena (USD)		Implementace	Použití	Patentováno
				Autoring	Koncový uživatel			
AAC	ISO/IEC MPEG Audio Committee	1997	ISO/IEC 14496-3	různá	zdarma	OSI: FAAC - iTunes		
ALAC	Apple Computer	2004	QuickTime 7.1.3	?	zdarma	QuickTime, iTunes, Real Player	Beztrátové uchování	?
ATRAC	Sony Corp.	1991	ATRAC3plus	?	?	MiniDisc, Walkman, VAIO, Clie, Playstation3 & Portable, SDDS, SoundForge, RealPlayer, ConnectPlayer	záznam hlasu, profesionální nahrávací studia, beztrátové uchování	Ano
FLAC	Xiph.Org Foundation	2001	1.1.4	zdarma	zdarma	OSI: reference	Beztrátové uchování	Ne
Monkey's Audio	Matthew T. Ashland	2000+	4.01b2	zdarma	zdarma		Beztrátové uchování	?
MP3	ISO/IEC MPEG Audio Committee	1993	ISO/IEC 11172-3	viz práce	?	OSI: LAME - Nero		Ano
Speex	Xiph.Org Foundation	2003	1.1.12	zdarma	zdarma	OSI: reference	Záznam hlasu	Ne
Vorbis	Xiph.Org Foundation	2002	1.1.2	zdarma	zdarma	OSI: reference, aoTuV		Ne
WavPack	Conifer Software	1998	4.31	zdarma	zdarma	OSI: reference		Ne
Windows Media Audio	Microsoft Corp.	1999	9.2	zdarma	zdarma pro uživatele oper. sys. Windows	Windows Media Player		Ano
Musepack	Frank Klemm/MDT	1997	42370	zdarma	zdarma	OSI: reference		Ne

Tab. 13: Tabulka některých formátů kódování zvuku a stručné informace o nich  
(Zdroj: [http://en.wikipedia.org/wiki/Comparison\\_of\\_audio\\_codecs](http://en.wikipedia.org/wiki/Comparison_of_audio_codecs))

Kodek	Typ algoritmu	Vzorkovací frekvence	Bit rate	Bitů na sampl	CBR	VBR	Stereo	Mono
AAC	Ztrátový, hybridní	8 Hz až 192 kHz	8 až 529 kbit/s (stereo)	Libovolně (užívá FP)	Ano	Ano	Dual, Mid/Side, Intensity, Parametric	Ano
ALAC	Bezztrátový	44.1, 48 kHz	Proměnný	?	Ne	Ano	Ano	Ano
FLAC	Bezztrátový	1 Hz až 1048.57 kHz	Proměnný	4, 8, 16, 24(, 32)	Ne	Ano	Ano	Ano
Monkey's Audio	Bezztrátový	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48 kHz	Proměnný	?	Ne	Ano	Ano	Ano
MP3	Ztrátový	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48 kHz	8, 16, 24, 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256, 320 kbit/s	Libovolně (užívá FP)	Ano	Ano	Dual, Mid/Side, Intensity	Ano
RealAudio	Bezztrátový, ztrátový	Různá	Proměnný	Různý	Ano	Ano	Ano	Ano
Speex	?	8, 16, 32 (48) kHz	NB: 2.15 až 24.6 kbit/s - WB: 4 až 44.2 kbit/s	?	Ano	Ano	Intensity	Ano
Vorbis	Ztrátový	1 Hz až 200 kHz	Proměnný	Libovolně (užívá FP)	Ano	Ano	Dual, Lossless, Phase, Point (Intensity)	Ano
WavPack	Bezztrátový, ztrátový, hybridní	1 Hz až 16777.216 kHz	Proměnný v bezztrátovém módu; 196 kbit/s a výše ve ztrátovém módu	Různé v bezztr. módu, min. 2.2 ve ztrát. módu	Ano	Ano	Ano	Ano
Windows Media Audio	Bezztrátový, ztrátový	8, 11.025, 16, 22.05, 32, 44.1, 48, 88.2, 96 kHz	4 až 768kbit/s, proměnný při bezztrátovém kódování	16, 24 v bezztr. Módu, jakýkoli v bezztrátovém (užívá FP)	Ano	Ano	Ano	Ano

Tab. 14: Tabulka některých formátů a jejich technické parametry  
(Zdroj: [http://en.wikipedia.org/wiki/Comparison\\_of\\_audio\\_codecs](http://en.wikipedia.org/wiki/Comparison_of_audio_codecs))



## ÚDAJE PRO KNIHOVNICKOU DATABÁZI

Název práce	Zvukový formát MP3
Autor práce	Petr Gregar
Obor	DI – El.zařízení v dopravě
Rok obhajoby	2007
Vedoucí práce	RNDr. Jan Hula
Anotace	Popis formátu MP3 a ID3 TAGu
Klíčová slova	MP3, ID3 TAG, komprese zvuku



## FORMULÁŘ PRO ZPŘÍSTUPNĚNÍ PRÁCE V ELEKTRONICKÉ FORMĚ – ČESKY

Typ dokumentu	Bakalářská práce		
Autor	Petr Gregar		
E-mail adresa autora	klocar@seznam.cz		
URN			
Název závěrečné práce	Zvukový formát MP3		
Stupeň studia	bakalářské		
Katedra	KATEDRA ELEKTROTECH., ELEKTRONIKY A ZABEZP. TECH. V DOPRAVĚ		
Vedoucí práce	RNDr. Jan Hula		
Klíčová slova	MP3 ID3 TAG komprese zvuku		
Datum obhajoby	27. 6. 2007		
Označení rozsahu zpřístupnění	souhlasím se zveřejněním celé práce	Datum: 30. 5. 2007	Podpis autora:
Abstrakt	Práce se zabývá popisem formátu pro kompresi digitálního záznamu zvuku - MP3.		
Název souboru	BP-Gregar- Format_MP3.doc	Velikost souboru	934 Kb

## FORMULÁŘ PRO ZPŘÍSTUPNĚNÍ PRÁCE V ELEKTRONICKÉ FORMĚ – ANGLICKY

Type of Dokument:	Bachelor work		
Author	Petr Gregar		
Author's E-mail adress	klocar@seznam.cz		
URN			
Title	Sound format MP3		
Degree	bachelor		
Department	DEPARTMENT OF ELECTRICAL AND ELECTRONICAL ENGINEERING AND SIGNALLING IN TRANSPORT		
Advisory committee	RNDr. Jan Hula		
Keywords	MP3 ID3 TAG		
Date of defense	27. 6. 2007		
Availability	unrestricted	Date: 30. 5. 2007	Autor:
Abstract	This essay is dealing with description of format for compression of digital sound recording - MP3		
Filename	BP-Gregar- Format_MP3.doc	Size	934 Kb