

# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ  
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS

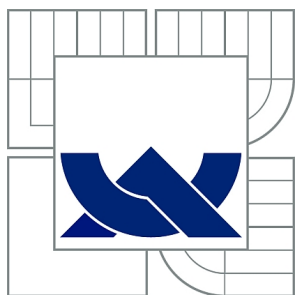
PSYCHOAKUSTICKÝ MODEL LIDSKÉHO SLUCHU

BAKALÁŘSKÁ PRÁCE  
BACHELOR'S THESIS

AUTOR PRÁCE  
AUTHOR

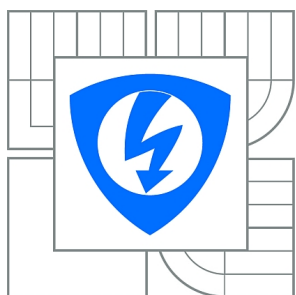
ONDŘEJ LEPA

BRNO 2011



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH  
TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS

## PSYCHOAKUSTICKÝ MODEL LIDSKÉHO SLUCHU

PSYCHOACOUSTIC MODEL OF HUMAN HEARING

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

ONDŘEJ LEPA

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. PETR SYSEL, Ph.D.

BRNO 2011



VYSOKÉ UČENÍ  
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

Ústav telekomunikací

# Bakalářská práce

bakalářský studijní obor  
Teleinformatika

**Student:** Ondřej Lepa

**ID:** 115218

**Ročník:** 3

**Akademický rok:** 2010/2011

**NÁZEV TÉMATU:**

## Psychoakustický model lidského sluchu

### POKYNY PRO VYPRACOVÁNÍ:

Cílem bakalářské práce je seznámit se s psychoakustickým modelem lidského slyšení a prostudovat jeho použití ve standardu MPEG. Jednotlivé části psychoakustického modelu (Fourierovu transformaci, nelineární kmitočtovou transformaci) implementujte jako funkce v prostředí Matlab. Celou implementaci ověřte na zpracování reálných akustických nahrávek.

### DOPORUČENÁ LITERATURA:

[1] Information technology - Generic coding of moving pictures and associated audio information. Standard ISO/IEC13818-3, 1996.

[2] Gold, B., Morgan, N. Speech and Audio Signal Processing. Processing and Perception of Speech and Music. First Edition. New York, Wiley: 2000. ISBN 0-471-35154-7

[3] Zwicker, E., Fastl, H. Psychoacoustics. Facts and Models. Second Updated Edition. New York, Springer: 1999. ISBN 3-540-65063-6

**Termín zadání:** 7.2.2011

**Termín odevzdání:** 2.6.2011

**Vedoucí práce:** Ing. Petr Sysel, Ph.D.

**prof. Ing. Kamil Vrba, CSc.**

*Předseda oborové rady*

### UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## **ABSTRAKT**

Tato práce se zabývá lidským sluchem a psychoakustikou. Přesněji psychoakustickým modelem standardu MPEG podle normy ISO/IEC 13818-3, Information technology - Generic coding of moving pictures and associated audio information Part 3: Audio. Detailně se pak zabývá způsobem, jakým se PAM určuje, jaké metody maskování používá a jaké má možnosti maskování. Dále je realizován jednoduchý výpočet psychoakustických hodnot v prostředí Matlab.

## **KLÍČOVÁ SLOVA**

psychoakustika, PAM, MPEG, Layer, Matlab, ISO, 13818-3, psychoakustický model, lidské sluché ústrojí

## **ABSTRACT**

This thesis deals with human auditory system and psychoacoustics. Deals with psychoacoustic model of standard MPEG following the norm ISO/IEC 13818-3, Information technology - Generic coding of moving pictures and associated audio information Part 3: Audio. Closer look is given to determination of PAM, to methods of masking and what potential of masking it has. Next the simple computation of psychoacoustic values is done in Matlab environment.

## **KEYWORDS**

psychoacoustic, PAM, MPEG, Layer, Matlab, ISO, 13818-3, psychoacoustic model, human auditory system

LEPA, Ondřej *Psychoakustický model lidského sluchu*: bakalářská práce. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací, 2011. 53 s. Vedoucí práce byl Ing. Petr Sysel, Ph.D.

## PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Psychoakustický model lidského sluchu“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení § 152 trestního zákona č. 140/1961 Sb.

Brno .....

.....

(podpis autora)

Děkuji vedoucímu práce,  
panu Ing. Petru Syslovi, Ph.D., za vedení a  
pomoc při opravách faktických  
a stylistických chyb.

# OBSAH

<b>Úvod</b>	<b>11</b>
<b>1 Zvuk</b>	<b>12</b>
1.1 Percepce zvuku . . . . .	12
1.2 Hlasitost zvuku . . . . .	13
1.2.1 Hladina intenzity zvuku . . . . .	14
1.2.2 Jednotka SONE . . . . .	14
1.2.3 Jednotka PHONE . . . . .	14
1.3 Dynamický rozsah . . . . .	15
<b>2 Sluchový orgán</b>	<b>16</b>
2.1 Vnější ucho . . . . .	16
2.2 Střední ucho . . . . .	17
2.3 Vnitřní ucho . . . . .	17
<b>3 Psychoakustika</b>	<b>18</b>
3.1 Práh slyšení . . . . .	18
3.2 Maskování . . . . .	18
3.2.1 Kmitočtové maskování . . . . .	19
3.2.2 Časové maskování . . . . .	20
3.2.3 Kritická pásma . . . . .	21
<b>4 Psychoakustický model pro MPEG</b>	<b>23</b>
4.1 Výpočet spektra . . . . .	25
4.2 Určení SPL . . . . .	26
4.3 Definice absolutního prahu slyšení . . . . .	26
4.4 Určení tónových a šumových složek . . . . .	27
4.4.1 Uložení tónových složek a výpočet jejich SPL . . . . .	28
4.4.2 Uložení šumových složek a výpočet jejich SPL . . . . .	28
4.5 Decimace počtu maskovacích složek . . . . .	29
4.6 Výpočet individuálních maskovacích prahů . . . . .	29
4.7 Výpočet celkového maskovacího prahu . . . . .	31
4.8 Určení minimálního maskovacího prahu . . . . .	31
4.9 Výpočet poměru SMR . . . . .	31
<b>5 Výpočet PAM v prostředí MATLAB</b>	<b>33</b>
5.1 Spuštění programu a výběr souboru . . . . .	33
5.2 Zjištění lokálních maxim . . . . .	34

5.3	Určení tónových a šumových složek . . . . .	35
5.4	Výpočet individuálních maskovacích prahů . . . . .	37
5.5	Výpočet celkového maskovacího prahu . . . . .	37
5.6	Určení Signal-to-Mask poměru a maskovacích úrovní jednotlivých sub- pásem . . . . .	39
<b>6</b>	<b>Závěr</b>	<b>41</b>
	<b>Literatura</b>	<b>43</b>
	<b>Seznam symbolů, veličin a zkratk</b>	<b>45</b>
	<b>Seznam příloh</b>	<b>46</b>
<b>A</b>	<b>Tabulky pro výpočet PAM</b>	<b>47</b>
A.1	Tabulky absolutních prahů . . . . .	47
A.2	Tabulky kritických pásem . . . . .	50
<b>B</b>	<b>Obsah přiloženého CD</b>	<b>52</b>



## SEZNAM OBRÁZKŮ

1.1	Oblast vnímání lidského ucha - převzato z [7]	14
2.1	Model lidského ucha převzato z [18].	16
2.2	Rozložení vláskových buněk převzato z [13].	17
3.1	Práh slyšení a maskovací křivka inspirováno [15]	19
3.2	Principiální zobrazení časového maskování inspirováno [15]	21
4.1	Schéma subpásmového kodéru a dekodéru	24
5.1	Vložení souboru pro výpočet PAM.	34
5.2	Hodnoty určené jako lokální maxima.	35
5.3	Hodnoty určené jako tónové a šumové složky.	36
5.4	Určení individuálních maskovacích prahů.	37
5.5	Určení celkového maskovacího prahu.	38
5.6	Určení maskovacích úrovní pro jednotlivá subpásma.	39
5.7	Určení odstupů signálu od maskovacího prahu.	40

## SEZNAM TABULEK

3.1	Šířka jednotlivých kritických pásem z [15] . . . . .	22
4.1	Hodnoty pro FFT v PAM z [9] . . . . .	26
4.2	Rozsah kmitočtů zkoumaných na výskyt tónových složek z [9] . . . . .	27
4.3	Kritéria pro výběr tónových složek z [9] . . . . .	28
4.4	Hodnoty maskovací funkce z [9] . . . . .	30

# ÚVOD

Při digitálním zpracování zvuku je i v dnešní době vyspělých technologií a nepřehledného množství úložného místa nutné zvukový signál číslicově zpracovávat. To se z větší části děje pomocí komprimace vstupního signálu. Aby bylo ale možné data zjednodušit a vynechat ty části, která nejsou z hlediska zpětného převodu a posluchové kvality potřebná, musí se do komprimace zapojit psychoakustika. Její zapojení spočívá v tom, že zavedený psychoakustický model obsahuje algoritmy, podle kterých se kódér signálu dozví, co může a co nesmí z původního signálu ořezat. Je to tedy soubor postupů a metod, zjištěných dlouhodobým testováním, které matematicky nahrazují lidský sluch a definují schopnosti vnímání. Psychoakustický model zná slabiny lidského sluchu, matematicky je definuje a shrnuje výsledky do tabulek, pro pozdější použití při kódování zvukového signálu.

V této práci se zabývám popisem lidského sluchu, psychoakustikou a způsoby, jakými určuje nepotřebná data. Dále potom standardem MPEG Audio, resp. způsobem, jakým v tomto standardu psychoakustický model 1 označuje a vynechává nepotřebná, pro lidský sluch nadbytečná, zvuková data. V poslední části práce je potom v prostředí Matlab realizován jednoduchý program, který simuluje psychoakustický model MPEG Layer I pro nižší vzorkovací frekvence. Pomocí tohoto programu je realizován výpočet odstupů signálu od maskovací funkce, který je nezbytný pro úspěšné nasazení psychoakustického modelu do kódéru zvukových signálů. Výpočet je realizován na trojici vstupních zvukových vzorků.

Celá práce se opírá o normu ISO/IEC 13818-3, Information technology - Generic coding of moving pictures and associated audio information Part 3: Audio. Z ní jsou čerpány potřebná data a hodnoty pro výpočet a celkovou realizaci samotného programu.

# 1 ZVUK

Zvuk je mechanické vlnění šířící se okolním prostředím určitou rychlostí. Vlnění vzniká zředováním a zhušťováním akustického prostředí a tím je vyvolána právě akustická vlna o určité rychlosti. Tato rychlost závisí především na hustotě prostředí. Čím je prostředí hustější, tím rychlejší přenos zvuku bude. Tato vlastnost souvisí s vlnovým charakterem zvuku. Čím jsou částice prostředí k sobě blíže, tím rychlejší je přenos energie mezi sousedícími částicemi. Naopak čím dále od sebe částice jsou, tím je přenos zvuku horší a pomalejší, např. přenos, resp. nemožnost přenosu, zvuku ve vakuu. Dalším atributem, který ovlivňuje přenos zvuku je teplota okolí. Čím je teplota vyšší, tím je přenos rychlejší. Tento poznatek jde ruku v ruce s dříve zmiňovanou hustotou prostředí, protože čím je teplota látky vyšší, tím rychleji se v ní částice pohybují a také zvukové vlny se díky tomu mohou šířit rychleji. [8]

Zvuk, v číslicové podobě, je posloupnost hodnot signálu. Takovýchto hodnot musí být každou časovou jednotku uloženo dostatečné množství, aby bylo možno následně daný signál opět reprodukovat jako „původní“ zvuk. Tyto hodnoty musejí být také dostatečně jemně rozčleněny mezi nulovou a maximální hodnotu. Takovýto záznam je ovšem velmi objemný - zabírá velké množství paměťového místa a vyžaduje velký datový tok při vysílání. Ke snížení těchto nároků, při zachování kvality, se používají kompresní algoritmy.

## 1.1 Percepce zvuku

Lidské ucho je schopné vnímat zvuky v rozsahu 16 Hz až 20 kHz. Jednotlivé hodnoty se mohou lišit člověk od člověka, tyto jsou však průměrné a určené dlouhodobým měřením. [6] Hodnoty kmitočtů mimo tyto vytyčené hranice nevyvolávají sluchový vjem, ale přesto jsou mnohdy označovány jako zvuk. Například nižší než 16 Hz se nazývají infrazvuk, je možné je vnímat taktálně. Takto se projevuje například kvalitní basový reproduktor. Naopak hodnoty nad 20 kHz jsou označovány jako ultrazvuk. Ten mohou vnímat některá zvířata a kupříkladu netopýři se pomocí něj dokonce orientují v naprosté tmě. Člověk ale tyto hodnoty již není schopen zachytit žádným přirozeným způsobem.

S přibývajícím věkem se u člověka vyskytuje změna sluchové hranice. [12] To se projevuje hlavně oslabeným vnímáním vysokých tónů. Jedná se především o nejvyšší kmitočty, které však pro praktický život nejsou nepostradatelné. Přibližně okolo věku 60ti let už bývá ztráta pozorovatelná i na úrovních okolo 8 kHz. Tato ztráta je pro běžný sluch už velmi významná.

Vnímání výšky tónu, zvuku, pro ucho neprobíhá v celém rozsahu lineárně, ale od hodnoty kmitočtu přibližně 800 Hz je výška vnímána přibližně logaritmicky.

Tato vlastnost jde asi nejlépe vidět na kmitočtech tónu naladěného klavíru, kde jsou jednotlivé oktávy percepčně stejně výškově vzdáleny. Zatímco rozdíl kmitočtů komorního A (440 Hz) a tónu A5 (880 Hz) je 440 Hz, u vysokých tónů A6 (1760 Hz) a A7 (3520 Hz) je rozdíl kmitočtů řádově vyšší, a to 1760 Hz. [19] Z porovnání 440 Hz a 1760 Hz je jasné, že výška není vnímána lineárně. Další poznatek je, že pokud chceme získat percepčně stejně výškově rozdílné tóny, musíme vždy zdvojnásobit kmitočet těch nižších. Rozlišovací schopnost různých tónů je také závislá na hodnotě kmitočtu. Zatímco od tónu znějícího na kmitočtu 1000 Hz dokážeme rozeznat tón lišící se o 2-3 Hz, u vyšších kmitočtů, například 7000 Hz, je již rozlišovací schopnost mnohem horší. Pro rozlišení dvou tónů je v druhé případě potřeba rozdíl alespoň 35 Hz. [14] Zkoumáním se zjistilo, že lidské ucho je nejvíce přizpůsobeno ke vnímání kmitočtů v rozmezí od 500 Hz do 3000 Hz, což odpovídá průměrným hodnotám lidské řeči.

## 1.2 Hlasitost zvuku

Hlasitost zvuku je subjektivní veličina, která je závislá na intenzitě dopadajícího akustického tlaku. Ten je vyvolaný pohybem zvukové vlny v akustickém prostředí. Změna akustického tlaku je v porovnání s atmosférickým tlakem zanedbatelná, navíc samotná hodnota tlaku nijak nereprezentuje hlasitost. Proto se zavedlo měření pomocí směrodatné odchylky tlaku vzduchu, která je při měření v krátkém časovém intervalu prakticky shodná se změnou akustického tlaku. To se označuje jako akustický tlak měřený v Pascalech a hraje hlavní roli v určování hlasitosti.

Testováním se zjistilo, že člověk vnímá zvuk až od určitého tlaku. Ten byl nazván prahový. Při kmitočtu 1 kHz byl tento práh určen jako  $p_0 = 20\mu Pa$ . Této hodnotě potom říkáme práh slyšitelnosti a byla zvolena jako standard. Přesto se od skutečné hodnoty jednotlivých posluchačů může lišit. Akustický tlak logicky klesá s rostoucí vzdáleností, protože energie akustické vlny se pozvolně vytrácí. [8]

Intenzita zvuku, počítaná podle (1.1), představuje množství zvukové energie, která projde určitou plochou za jednotku času. Intenzita klesá se čtvercem vzdálenosti od zdroje zvuku a celkové množství energie, ve Watech, vyzářené daným zdrojem do okolí se nazývá akustický výkon. Hladina akustického tlaku SPL je logaritmičtým vyjádřením velikosti tlaku vyvolaného měřeným zvukem v závislosti od prahového akustického tlaku  $p_0$ .

$$L_p = 20 \log_{10} \frac{p_1}{p_0} \quad (1.1)$$

Úroveň akustického tlaku je ale problémová z hlediska vnímání zvuku člověkem, neboť na různých kmitočtech se zvuk zdá různě hlasitý, i přes shodnou úroveň SPL.

Z toho důvodu se zavedlo více metod hodnocení hlasitosti.

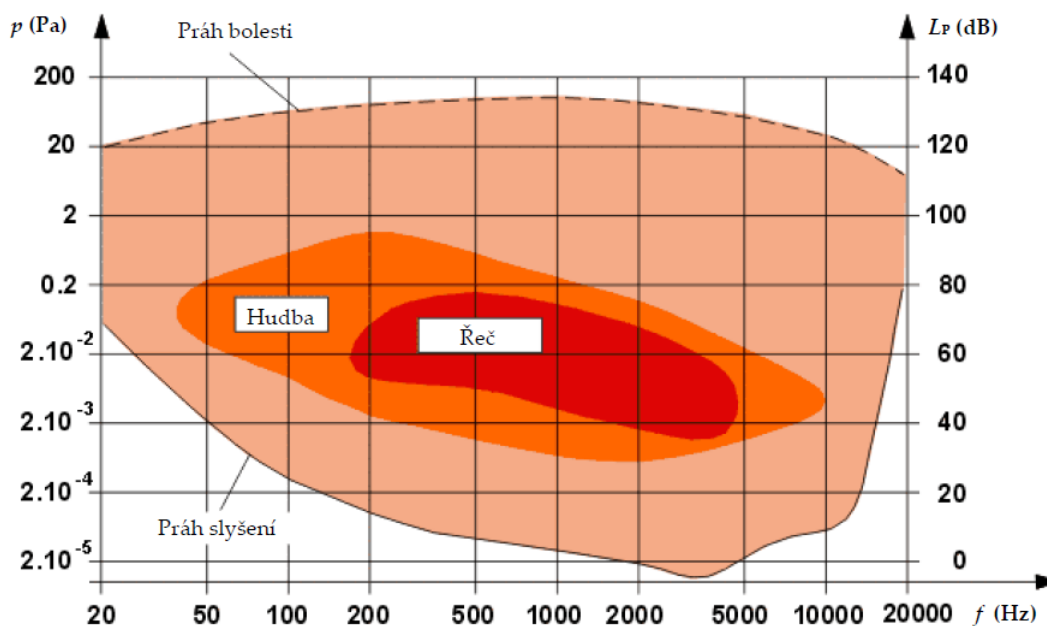
### 1.2.1 Hladina intenzity zvuku

Hladinu intenzity zvuku lze vyjádřit podle (1.2), kde  $I_0$  je prahová intenzita zvuku, měřená dB(SIL).

$$L_i = 10 \log_{10} \frac{I}{I_0} \quad (1.2)$$

### 1.2.2 Jednotka SONE

Další jednotkou zjednodušující měření hlasitosti je SONE. Tentokrát se jedná o jednotku pro subjektivní vnímání hlasitosti. Podle zakladatele S. S. Stevensse představuje 1 sone 40 dB(SPL) při kmitočtu 1 kHz a na rozdíl od dB se jedná o jednotku lineární. Hodnota sonu byla vybrána tak, aby zdvojnásobení sonu mělo pro posluchače efekt zdvojnásobení hlasitosti.



Obr. 1.1: Oblast vnímání lidského ucha - převzato z [7]

### 1.2.3 Jednotka PHONE

Stupnice jednotek PHONE má za úkol eliminovat vliv kmitočtu na vnímanou hlasitost tónů, proto se určila jednotka 1 phone. Ta má hodnotu 1dB(SPL) při kmitočtu 1 kHz.

## 1.3 Dynamický rozsah

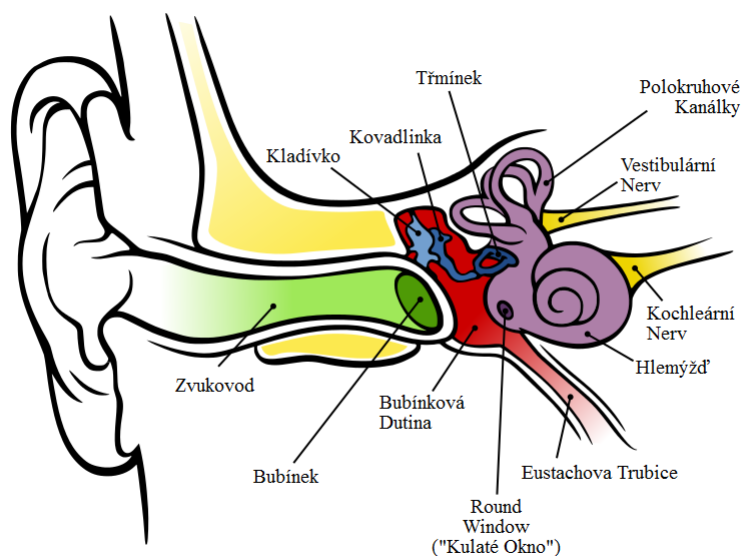
Dynamický rozsah je jeden z dalších rozměrů zvuku, definovaný jako rozdíl mezi nejslabším a nejhlasitějším slyšitelným zvukem. Rozsah dynamiky je stanoven přibližně na 120 dB, přičemž 0 dB patří zvuku nejméně slyšitelnému, resp. zvuku, který již sluch nedokáže vnímat, a 120 dB patří zvuku nejhlasitějšímu. Z bezpečnostního hlediska se v dynamice určují 4 oblasti, a to oblast nižší než 0 dB. Pod tuto hodnotu nebude žádný zvuk slyšitelný. Oblast od 0 dB do 90 dB je považována za oblast slyšitelnosti. Oblast od 90 dB do 120 dB je označována jako oblast nepříjemnosti, kdy zvuky s úrovní hlasitosti v této oblasti jsou nepříjemné, a vrcholná oblast nad 120 dB pojmenovaná jako oblast bolesti, což znamená, že všechny silnější zvuky budou vyvolávat bolest. [14] Při dlouhodobějšímu vystavení oblasti bolesti může dojít k nevratnému poškození zvuku a to buď protržením bubínku, nebo poškozením dále zmíněných vláskových buněk. Dynamický rozsah není ve všech místech slyšitelného spektra stejný. Největší rozsah, právě zmíněných 120 dB, je uprostřed slyšitelného kmitočtového pásma. Směrem k okrajům se dynamický rozsah, tak jako většina veličin týkajících se zvuku a sluchu, kriticky mění.

## 2 SLUCHOVÝ ORGÁN

Ucho je lidský orgán, zajišťující vnímání a zpracování zvukového vzruchu. Skládá se vnějšího, středního a vnitřního ucha.

### 2.1 Vnější ucho

Vnější ucho má dvě části – ušní boltec a zvukovod. Boltec je chrupavčitá část, která má podobný význam jako parabola satelitu. Zvukovod je z části chrupavčitá a z části kostěná trubička o délce asi 3 cm, na jejímž konci se nachází bubínek. Zvukové vlny ze zdroje zvuku jsou tedy soustřeďované boltcem do zvukovodu, jehož nesymetrický tvar způsobuje, že zvukové vlny přicházející z různých směrů jsou mírně odlišné. To je důsledek odrazů od různých povrchů, které pozměňují jeho fázi, amplitudu i frekvenční spektrum. To pomáhá k orientaci v prostoru, protože mozek je potom schopen určit, který zvuk je odražený a který přichází přímo od zdroje. Zvukové vlny potom pokračují dále zvukovodem, což je jednoduchý přibližně válcovitý otvor, vedoucí až k bubínku. V důsledku tvarování zvukovodu mírně zesiluje zvuky o frekvenci od 3 do 12 kHz. [15] Bubínek, vazivová blanka o tloušťce přibližně 0,1 mm, tvoří pomyslný přechod mezi vnějším a středním uchem. Přestože se ve středním uchu také nachází vzduch, nemá žádný přímý spoj s venkovním prostorem. Bubínek pevně odděluje venkovní prostředí a prostředí vnitřního ucha a vlivem přivedeného zvuku se rozkmitá, bubínek rozkmit zesílí a vibrace membrány se dále přenáší už jen pomocí kůstek středního ucha, kladívka, třmínku a kovádlínky.



Obr. 2.1: Model lidského ucha převzato z [18].

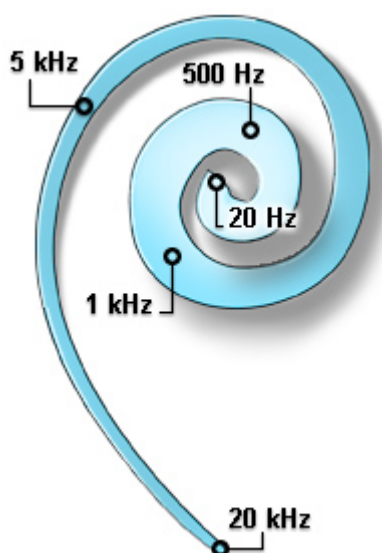


## 2.2 Střední ucho

Střední ucho je systém vzduchem vyplněných dutin a je vystlán sliznicí. Jsou v něm umístěny tři kůstky - kladívko, třmínek a kovadlinka. Ty přenášejí vibrace z bubínku ve formě tlakových vln na kulaté okénko, což je vstupní branou do vnitřního ucha, především do hlemýždě. Střední ucho je stále ještě vyplněno vzduchem. S vnějším prostorem je spojeno pouze skrze Eustachovu trubici přes dýchací cesty, nos a dutinu ústní, kvůli vyrovnávání tlaku na bubínek.

## 2.3 Vnitřní ucho

Vnitřní ucho se nachází v kostěném labyrintu, obklopené části jsou 3 polokruhové kanálky a hlemýžď. Ten začíná kulatým okénkem, které je vstupní branou. Na kulaté okénko přiléhá třmínek ze středního ucha. Vnitřní ucho samotné je tvořeno spirálovitým hlemýžděm, což je vlastně stočená trubička vyplněná endolymfou, směsí tekutin. Vlnění endolymfy, přicházející skrze kulaté okénko, rozechvívá krycí membránu Corttiho orgánu. Ten překrývá vláskové buňky. Corttiho orgán provádí převod mezi kmitáním endolymfy a v závislosti na rychlosti kmitání dochází k rozechvívání v určitém fyzickém prostoru vnitřního ucha, hlemýždě. Podle obrázku 2.2 jsou nejvyšší frekvence detekovány hned za oválným okénkem, kdežto ty nejnižší jsou vnímány až na konci spirály hlemýždě. Pohybem vláskových buněk dochází ke změně jejich elektrického potenciálu, chvění endolymfy tedy přeměňují na elektrické impulzy, které dále zpracovává mozek a tím vzniká sluchový vjem. [13]



Obr. 2.2: Rozložení vláskových buněk převzato z [13].

## 3 PSYCHOAKUSTIKA

Psychoakustika je vědní obor, zabývající se subjektivním vnímáním zvuku. Je považována za spojení mezi lidským vnímáním zvuku a fyzikálními velikostmi zvukového pole. Díky jednotlivým psychoakustickým principům je možné komprimovat zvukové soubory bez zdatelné ztráty poslechové kvality. V podstatě to znamená, že z výchozí vzorek zvukový záznam je zbaven té části slyšitelného spektra, které ucho není schopné vnímat. Ponechány jsou pouze složky potřebné ke správnému vjemu reprodukováného zvuku. Taková komprese je naprosto nezbytná, protože bez ní, by nebylo možné číslicově zpracovávat a uchovávat analogové záznamy v rozumném poměru velikosti ku kvalitě.

### 3.1 Práh slyšení

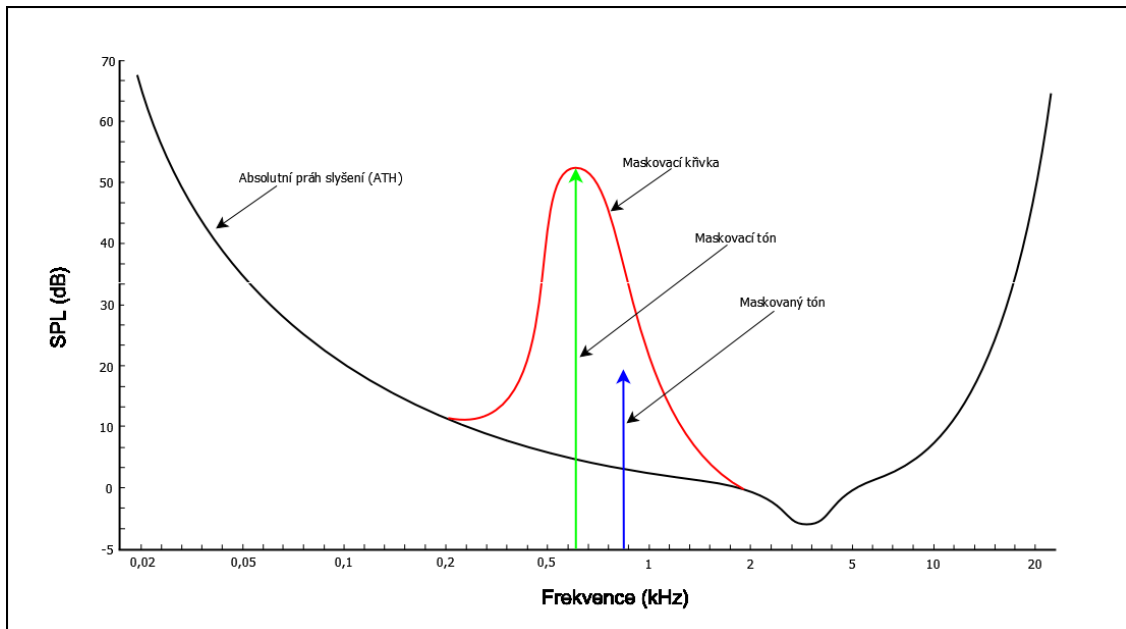
Jak bylo zmíněno výše, sluchový vjem nezávisí pouze na fyzikálních vlastnostech akustické vlny, ale také na vlastnostech sluchového orgánu. Ten má pro různé kmitočty různou citlivost a dynamický rozsah. V závislosti na kmitočtu lze vynést křivku, která pro každý kmitočet prochází hodnotou akustického tlaku, který je lidské ucho schopno vnímat. Tato křivka zobrazená na obrázku 3.1 se nazývá práh slyšitelnosti (threshold in quite). Práh je však závislý na více faktorech. Například vlivem věku se křivka posouvá směrem vzhůru hlavně v oblasti vyšších frekvencí. Tato křivka se dá přibližně aproximovat funkcí podle obrázku rovnice (3.1). Práh slyšitelnosti je velice důležitou součástí psychoakustického kodéru. Veškeré kmitočtové složky, které se nacházejí pod ním jsou totiž neslyšitelné a je možné je ze signálu odebrat bez změny poslechové kvality. [15]

$$A(f) = 3,64 \left( \frac{f}{1000} \right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}\right)-3.3^2} + 10^{-3} \left( \frac{f}{1000} \right) \quad (3.1)$$

### 3.2 Maskování

Maskování je jev, při němž některá část v signálu vjemově překryje jinou, která se tak stává neslyšitelnou. Tohoto jevu se hojně využívá v psychoakustickém kodéru, neboť při přenosu reálných signálů dochází k maskování neustále a ve velké míře ovlivňuje naše vnímání zvuku. Prvek signálu, který způsobí překrytí jiného prvku nazýváme maskující, masker. Prvek překrytý, potom maskovaný, maskee. Maskování dělíme na dva hlavní typy, v závislosti na vzájemné časové pozici maskujícího a maskovaného prvku. Jestliže se maskující a maskovaný prvek vyskytují v signálu současně,

hovoříme o takzvaném souběžném, kmitočtovém, maskování. Jestliže maskující prvek časově předchází nebo následuje za maskovaným prvkem, hovoříme o časovém maskování.



Obr. 3.1: Práh slyšení a maskovací křivka inspirováno [15]

### 3.2.1 Kmitočtové maskování

Kmitočtové maskování je jev, při kterém silnější komponenta v signálu, která se vyskytuje ve stejném okamžiku jako komponenta slabší, tuto slabší překryje a překrytá komponenta přestane být ve zvukovém vzorku slyšitelná. Na obrázku 3.1 vidíme zjednodušený příklad kmitočtového maskování. Signál je tvořen dvěma tóny o blízkých kmitočtech. Tón s nejvyšší amplitudou, v jehož okolí dojde ke změně křivky slyšitelnosti, kompletně překryje druhý tón. Pokud tedy přehrajeme signál tvořený pouze silnějším tónem, z poslechového hlediska se zvuk nezmění, což nám dává velký potenciál k úspoře dat. Rozlišujeme dva druhy maskujících prvků, a to šumové a tónové. Šumový maskující prvek má mnohem lepší potenciál maskovat okolní kmitočtová pásma než tónový. Šumová složka maskuje kmitočtová pásma s maximem  $-5$  dB pod úrovní maskujícího prvku, tónová složka až zhruba  $-20$  dB pod úrovní maskujícího prvku, což snižuje potenciál tónové složky z hlediska psychoakustické komprese. [10] Rozdělení na tóny a šum je možné provádět například pomocí detekce lokálních maxim, kde maximum bude tón, nebo pomocí predikce, kdy tón bude vysoce predikovatelný z po sobě jdoucích kmitočtových spekter.

### 3.2.2 Časové maskování

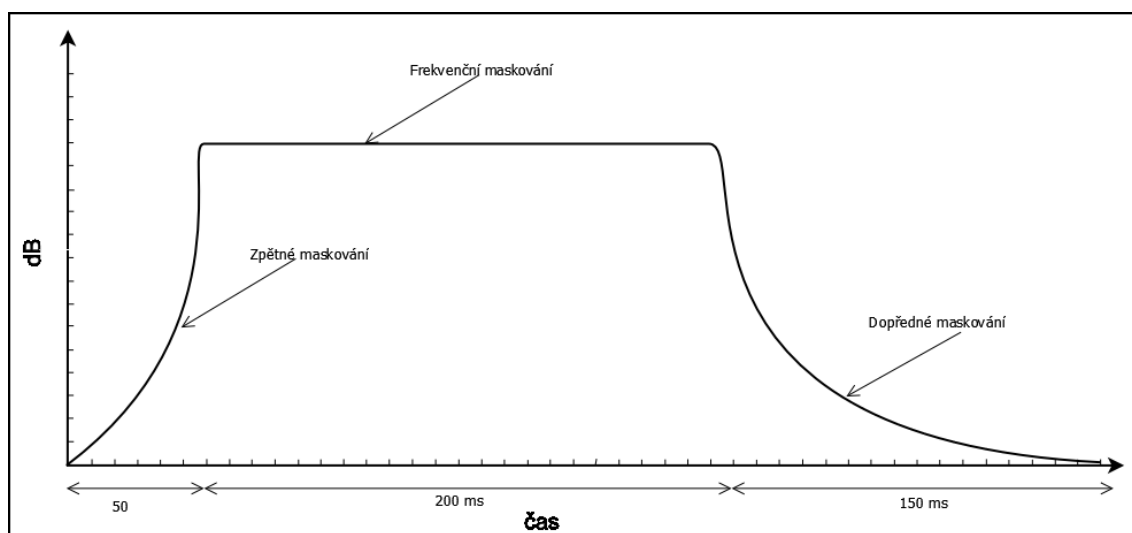
Následuje-li po hlasitém tónu velice podobný tón výrazně nižší hlasitosti, je jeho vnímání potlačeno. To je dáno ochranou sluchového ústrojí. Soustava 3 kůstek vnitřního ucha představuje automatickou regulaci citlivosti sluchu. V tichém prostředí přenášejí kůstky veškerou akustickou energii z bubínku do oválného okénka (blanka uzavírající vlastní sluchový orgán - hlemýžď). Nejvíce převedené energie je zajištěno jejich vzájemným 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 velkým rozsahem intenzit. Porovnáme-li intenzity nejslabších slyšitelných zvuků s prahem bolesti zjistíme, že jde o zvuky lišící se svojí intenzitou  $10^{13}$ -krát. [6] 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. Časové maskování se dělí na dva typy, podle výskytu maskujícího zvuku.

#### Pre-masking

Dopředné maskování, pre-masking, nastává právě tehdy, když maskovaný zvuk předchází zvuku maskujícímu. Tato situace je matoucí, protože se zdá, že zvukový vjem je okamžitý a první zvuk, maskovaný, by měl být již dávno zaznamenaný. Opak je ovšem pravdou a zpětné maskování se obecně vysvětluje tak, že uchu trvá jistou dobu než rozkmitá bubínek na potřebnou výchylku a tím je způsobeno právě zpětné maskování. Pre-masking trvá, oproti post-maskingu, mnohem kratší dobu, řádově desítky milisekund. Dá se použít při kvantování vzorků, kde jsou při nena-dále změně energie dopadajícího zvuku přimíchány nechtěné kvantizační šумы i do částí zvuku s nižší energií.

#### Post-masking

Zpětné maskování, post-masking, nastává, když maskující zvuk přichází do ucha dříve, než zvuk maskovaný. Post-masking spočívá v neschopnosti ucha, rozeznat rychle po sobě znějící zvuky. Vysvětluje se tak, že vláskové buňky jsou rozkmitány



Obr. 3.2: Principiální zobrazení časového maskování inspirováno [15]

maskovaným signálem a působící membrána kulatého okénka stále ještě kmitá. Následující slabší zvuk bude proto opomenut a nezaznamenán. Při vysokých hodnotách akustického tlaku může také nastat případ popsaný v úvodu časového maskování, čili přenosové kůstky jsou ještě příliš roztažené a proto nedokáží slabší zvuk ani přenést. Na obrázku 3.2 je vidět, že dopředné maskování má mnohem delší dobu působení. Protože sluchovému ústrojí trvá poměrně dlouho, než se setrvačností vrátí do původního stavu.

### 3.2.3 Kritická pásma

Experimenty ukázaly, že při frekvenčním maskování se maskovací efekt v okolí maskujícího prvku zužuje směrem k nižším kmitočetům. Bylo zjištěno, že lidské ucho rozděljuje zvuk do 24 takzvaných kritických pásem (critical bands). Uvnitř těchto pásem dochází k nejvýraznějším efektům frekvenčního maskování, neboť lidské ucho v nich nedokáže jasně rozlišit více kmitočetů. Kritická pásma nemají stejný kmitočtový rozsah. Pásma pokrývající nižší kmitočty mají řádově nižší rozsah než pásma pokrývající vysoké frekvence. To souvisí s nelinearitou lidského sluchu. V tabulce 3.1 je naznačeno, jak se šířka jednotlivých kritických pásem po celém vnímatelném kmitočtovém spektru mění. Jestliže na vodorovnou osu vyneseme lineárně všech 24 kritických pásem, potom v tomto měřítku odstraníme kmitočtovou závislost tvaru maskovacího efektu v okolí maskujícího prvku, maskovací efekt bude mít na celém rozsahu stejný tvar. Jedná se o Barkovou stupnici (Bark scale), jejíž jednotkou je 1 Bark. Jelikož celý koncept kritických pásem vychází z fyzických dispozic ucha, ve skutečnosti barková stupnice odpovídá fyzickému rozdělení vláskových buněk

Číslo pásma	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
Šířka [Hz]	100	100	100	100	110	120	140	150	160	190	210	240
Číslo pásma	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.
Šířka [Hz]	280	320	380	450	550	700	900	1100	1300	1800	2500	3500

Tab. 3.1: Šířka jednotlivých kritických pásem z [15]

detekujících jednotlivé kmitočty v uchu. Každý Bark odpovídá stejně dlouhému segmentu hlemýždě. Jedná se tedy o mapování z kmitočtové oblasti do prostorové lokality ucha. Barková stupnice je nelineární stupnicí, pro převod jednotek z Hz do Bark se využívá vzorec (3.2). S použitím Barkové stupnice můžeme jednoduše určit maskovací efekt mezi komponentami signálu nezávisle na jejich kmitočtech, což využijeme v psychoakustickém modelu pro výpočet maskování. [14]

$$KP = 13 \operatorname{atan}\left(0,76 \frac{f}{1000}\right) + 3,5 \operatorname{atan}\left(\left(\frac{f}{1000}\right)^2\right) \quad (3.2)$$

## 4 PSYCHOAKUSTICKÝ MODEL PRO MPEG

Ve standardu MPEG pro kódování zvukových signálů je používán psychoakustický model k vytyčení zvukových složek, které jsou vjemově důležité a naopak, které můžeme vynechat, aniž by byl výsledný záznam nějak slyšitelně postižen. Psychoakustický model hledá hlasité zvuky, které mohou zamaskovat tóny slabší; šумы, které mohou ovlivnit okolní zvukové úrovně; zvuky, které jsou příliš slabé na to, aby je bylo ucho schopné vnímat a měli by být tedy vynechány. Informace z psychoakustického modelu jsou používány k vymezení těch částí spektra, které by měli dostat více bitů a tudíž být kódovány ve větší kvalitě. Také určuje, které části jsou neslyšitelné/nedůležité a měly by tedy dostat méně bitového prostoru.

Místo věnované zpracovanému signálu se určuje již při zpracování. Signál nejprve projde procesem pulzně kódové modulace [3] a dále kvantováním. Tyto metody jsou poměrně známé, proto je zde nebudu dále popisovat. Zajímavější procesy jsou potom váhování a subpásmové kódování.

### Váhování

Zpracování signálu probíhá po malých intervalech, při kterých se předpokládá, že vzorkovaný signál je periodický a jeho perioda je menší, než zvolený interval. Pokud se tak nestane, vznikají chyby zpracování. Tento problém u standardu MPEG řeší použití tzv. váhovacího okénka. [11] Okénko v analyzovaném rámci vybere příslušné vzorky a během zpracování je vyhlazuje přidělováním určité váhy. Okénka se používají jednoduše vynásobením aktuální hodnoty rámce a okna podle (4.1),

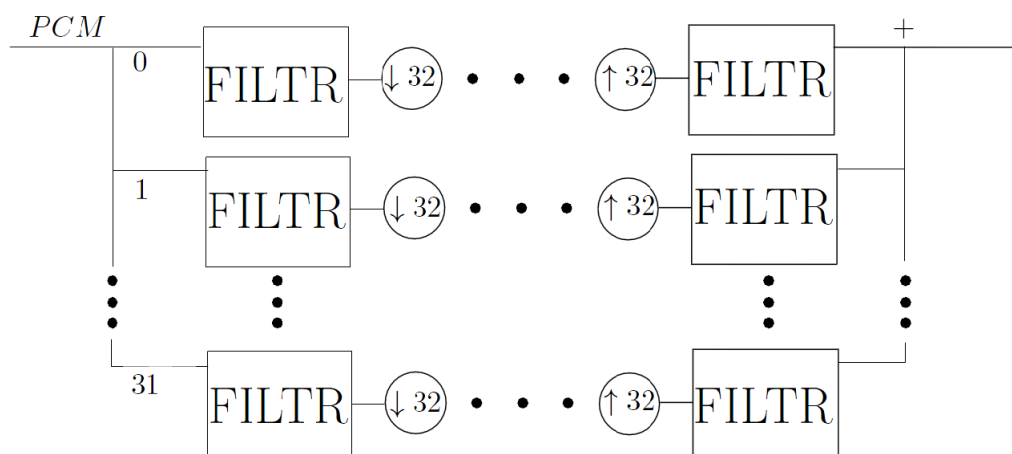
$$y[n] = x[n] \cdot w[n] \quad (4.1)$$

kde  $y(n)$  je výstupní posloupnost,  $x(n)$  je vstupní posloupnost a  $w(n)$  je aplikované okno. Pro standard MP3 se používá tzv. Hannovo okno (4.2), občas pro svůj tvar nazývané kosinové okno. Tato funkce zajišťuje nízký výskyt aliasingu. Výměnou za to je mírně sníženo kmitočtové rozlišení okna. Pokud je okno aplikováno na vzorek signálu ve kmitočtové oblasti, jeho zpětná konverze do časové oblasti nebude postižena zkreslením. Hannovo okno přikládá větší váhu střednímu úseku vzorku, zatímco na okrajích je jeho váha téměř nulová. Díky tomu nedochází ke zkreslení při překryvu jednotlivých vzorků.

$$w(n) = 0,5 \cdot \left(1 - \cos 2\pi \frac{n}{N}\right), 0 \leq n < N \quad (4.2)$$

## Sub-pásmové kódování

Formáty standardu MPEG využívají pro svoje zpracování rozdělení zpracovávaného signálu do několika subpásem. Počet subpásem se liší podle použité vrstvy standardu MPEG. Pro formát MP3 je použito 32 filtrů typu pásmová propust. Ty zajistí, aby se vstupní signál rozdělil do daného počtu stejně širokých pásem, tím se zabrání výskytu aliasingu při překryvu. Typické je použití filtrů v počtu násobků dvou. Sub-pásmové kódování probíhá tak, že vstupní zvukový signál prochází skrze banku filtrů, které ji rozdělí na výše zmíněných 32 subpásem, simultánně prochází skrze psychoakustický model, který určí poměr mezi energií signálu a maskovacím prahem každého subpásma. Tím se určí, jak se rozdělí celkové množství bitů přístupných pro kvantizaci signálů subpásma, aby se minimalizovala slyšitelnost kvantizačního šumu. Poslední blok potom reprezentuje kvantované vzorky subpásma a formátuje je. Vedlejší data, ne nezbytně související se zvukovým signálem, mohou být vložena zde do kódovaného bitstreamu. Dekodér si potom tento bitstream dešifruje, obnoví kvantizovaná subpásmové hodnoty a rekonstruuje zvukový signál ze subpásmových hodnot. [2]



Obr. 4.1: Schéma subpásmového kodéru a dekodéru

V kódování MPEG Audio Layer II je načítáno 1152 zvukových vzorků, které reprezentují jeden rámeček. Pro každý rámeček psychoakustický model vypočítává 32 výstupních hodnot, poměrů signálu k maskovací křivce na subpásmo. Je tedy pouze 32 hodnot k určení, jak rozdělit bity pro 1152 vzorků a to je v podstatě velmi hrubá technika rozdělení. [9]

Různé psychoakustické modely používají různé techniky k rozhodování o těchto 32 hodnotách. Některý model je lepší než druhý, ve smyslu využití správného pokrytí vzorků přidělenými bity. I s použitím opravdu špatného psychoakustického



modelu (například PAM -1) je stále mnohdy možné dosáhnout uspokojivých výsledků. Každý jednotlivý model má svoje silné a slabé stránky, které se poté přenáší i na výsledný formát, ve kterém je ten který model využíván.

Psychoakustický model 1 PAM 1 je stejně jako PAM 2 odvozen ze standardů organizace ISO. Jedná se o jeden z nejjednodušších psychoakustickým modelů, který se ovšem stal velice rozšířeným. Používá se především při kompresi, kde se využívají nižší vzorkovací kmitočty. Především 16 kHz, 22,05 kHz a 24 kHz. [9] Při použití PAM 1 ve formátech MPEG-1 Layer 1 a Layer 2 není nutné provádět žádné principiální změny procesu. Model může být dokonce upraven i pro Layer 3 (formát MP3), což je ale poměrně nevhodné, protože pro Layer 3 je používán PAM 2, který je mnohem vhodnější pro aplikaci vyšších  $f_{vz}$ .

Rozdíl při použití v Layer 1 a Layer 2 je v množství zpracovávaných vzorků. Zatímco při zpracovávání v Layer 1 je počítána bitová alokace pro každý blok 12ti subpásem nebo 384 vstupních PCM vzorků, v Layer 2 je bitový příděl vypočítáván pro 3 bloky, čili 36 subpásem nebo  $(3 \cdot 384 =) 1152$  vstupních PCM vzorků.

Rozdělení bitů pro 32 subpásem se počítá na základě poměru signálu ku maskovacímu prahu všech subpásem. Proto je důležité přesně určit nejvyšší úroveň signálu a minimální maskovací práh pro každé pásmo. Minimální maskovací práh je odvozen z FFT vstupních vzorků PCM signálu.

FFT, vykonávaná paralelně se subpásmovým dělením, kompenzuje nedostatek spektrální selektivity získané subpásmovou bankou filtrů na nízkých frekvencích. Tato technika poskytuje jak dostatečné rozlišení časové složky pro kódování audio signálu, tak uspokojivé spektrální rozlišení pro výpočet maskovacího prahu. Díky tomu mohou být vypočítány hodnoty kmitočtů a úrovně schodovitého zkreslení. To je nezbytné pro výpočet minimálního bitového toku pro ty subpásma, která potřebují nějaké bity pro eliminaci aliasingu v dekodéru.

## 4.1 Výpočet spektra

Převod signálu z časové do kmitočtové oblasti probíhá pomocí Fourierovy Transformace. Standardní váhování, použité na analogový vstupní signál před vstupem do FFT, je za použití Hannova okna ochráněna před zkreslením. Tato ochrana spočívá v rozdílné délce zpracovávaných rámců. To vzhledem k tomu, že Hannovo okno zpracovává pouze hodnoty ve středních částech okna a okrajové hodnoty zůstávají v podstatě nulové, dovoluje překrytí jednotlivých oken pro lepší rozlišení a zároveň ulehčuje zpracování FFT na právě 384 nenulových hodnot.

$$[h!]X_{(k)} = 10 \log_{10} \left| \frac{1}{N} \sum_{l=0}^{N-1} h[k] \cdot s[l] \cdot e^{-jkl \frac{2\pi}{N}} \right|^2 \quad k = 0 \dots N/2 \quad (4.3)$$

	Layer 1	Layer 2
transformační délka N	512 vzorků	1024 vzorků
Velikost okna pro $f_{vz} = 24$ kHz	21,33 ms	42,67 ms
Velikost okna pro $f_{vz} = 22,05$ kHz	23,33 ms	46,44 ms
Velikost okna pro $f_{vz} = 16$ kHz	32 ms	64 ms
Frekvenční rozlišení	$f_{vz}/512$	$f_{vz}/1024$

Tab. 4.1: Hodnoty pro FFT v PAM z [9]

Je nezbytné, aby ještě proběhla normalizace SPL na podporovaných  $96 \text{ dB}_{(\text{SPL})}$ . Společně s FFT se také podle (4.3) počítá hodnota  $X_{(k)}$ , která udává hustotu spektra vstupního signálu  $s_{(l)}$ .

## 4.2 Určení SPL

Pro každé subpásma  $n$  je potřeba určit jeho hodnotu SPL. Ta se spočítá podle (4.4) kde  $X_{(k)}$  je SPL spektrální linky s indexu  $k$  FFT s maximální amplitudou kmitočtového rozsahu obsaženého v subpásma  $n$ . Význam  $scf_{\max}(n)$  se liší podle použité vrstvy formátu MPEG. Ve vrstvě Layer I tato hodnota udává měřítko a ve vrstvě Layer II největší ze tří měřítek subpásma  $n$  v rozmezí jednoho rámce. Ve vzorci se také vyskytuje odpočet  $-10$  dB, což kompenzuje rozdíly mezi hodnotou špičky (peak) a RMS.

$$L_{\text{sb}}(n) = \text{MAX} [X_{(k)}, 20 * \log(scf_{\max}(n) \cdot 32768) - 10] \quad (4.4)$$

Pro výpočet SPL existuje ještě alternativní, efektivnější [9], postup výpočtu  $L_{\text{sb}}(n)$ . Tento postup (4.5) umožňuje kodéru větší potenciální výkon. V této rovnici  $X_{(k)}$  reprezentuje SPL odpovídajícího subpásma  $n$ .

$$L_{\text{sb}}(n) = \text{MAX} [X_{\text{spl}}(n), 20 \cdot \log(scf_{\max}(n) \cdot 32768) - 10] \quad (4.5)$$

hodnota  $X_{(k)}$  je vypočítána podle (4.6)

$$X_{\text{spl}}(n) = 10 \cdot \log_{10} \left( \sum 10^{X_{(k)}/10} \right) \quad (4.6)$$

## 4.3 Definice absolutního prahu slyšení

Práh slyšení, jinak zvaný absolutní práh, je možné vyčíst z tabulek A.1. Pomocí tohoto prahu PAM určí, které složky jsou mimo rozsah slyšitelné části spektra.

Hodnoty jsou předem dané, zprůměrované a normované. Tabulky jsou závislé na použitém vzorkovacím kmitočtu PCM vzorků, protože pro každý vzorkovací kmitočet je potřeba určit nové hodnoty subpásem. Hodnoty jsou uvedené pro každý vzorek ve kmitočtové oblasti, ve které se vypočítává i maskovací práh.

## 4.4 Určení tónových a šumových složek

Vztah jednotlivých tónů maskovacích složen má velký dopad na maskovací křivku. Z toho důvodu se vyplácí rozlišit tónové a netónové, čili šumové, složky. Pro výpočet celkového maskovacího prahu je nezbytné, aby se ze spektra FFT odvodily právě tyto složky. Po celé šířce subpásma se dále analyzuje intenzita netónových složek

Pro určení těchto složek je nejprve nutné definovat lokální maxima spektra. Z nich se potom získají všechny složky, mající tvar podobný funkci sinus, reprezentující tónové části. Následujícím úkolem je vyjádřit intenzitu netónových složek v mezích šířky kritického pásma. Tyto kritická pásma jsou definována v tabulkách kritických pásem A.2. Šířka kritickým pásem se mění v závislosti na pozici v kmitočtovém rozsahu sluchu. V nižších polohách je šířka kritického pásma pouze několik stovek Hz, ale ve vyšších polohách se, díky nelinearitě vnímání zvuku, jejich šířka rapidně zvětšuje a to až ke 4 kHz. [9] To v podstatě znamená, že šířka kritického pásma, je z hlediska výšky tónů stejně velká, avšak ve kmitočtovém vyjádření roste.

K určení, jestli lokální maximum může být tónová složka, je podle tabulky 4.2 zkoumán rozsah  $d_f$  okolo tohoto maxima. Tento rozsah je závislý na poloze maxima v kmitočtové stupnici a na vzorkovací frekvenci  $f_{vz}$ .

Vzorkovací kmitočet 16 kHz	
$d_f = 62,5$ Hz	$0 \text{ kHz} < f < 3,0 \text{ kHz}$
$d_f = 93,75$ Hz	$3,0 \text{ kHz} < f < 6,0 \text{ kHz}$
$d_f = 187,5$ Hz	$6,0 \text{ kHz} < f < 7,5 \text{ kHz}$
Vzorkovací kmitočet 22,05 kHz	
$d_f = 86,133$ Hz	$0 \text{ kHz} < f < 2,756 \text{ kHz}$
$d_f = 129,199$ Hz	$2,756 \text{ kHz} < f < 5,512 \text{ kHz}$
$d_f = 258,398$ Hz	$5,512 \text{ kHz} < f < 10, 336 \text{ kHz}$
Vzorkovací kmitočet 24 kHz	
$d_f = 93,750$ Hz	$0 \text{ kHz} < f < 3,0 \text{ kHz}$
$d_f = 140,63$ Hz	$3,0 \text{ kHz} < f < 6,0 \text{ kHz}$
$d_f = 281,25$ Hz	$6,0 \text{ kHz} < f < 11,250 \text{ kHz}$

Tab. 4.2: Rozsah kmitočtů zkoumaných na výskyt tónových složek z [9]

Všechny spektrální linie, které potom splňují (4.7) a zároveň (4.8) jsou automaticky označeny jako lokální maxima. Takto získané informace o daných vzorcích se dále třídí a ukládají pro následné zpracování a značení.

$$X(k) > X(k - 1) \quad (4.7)$$

$$X(k) \geq X(k + 1) \quad (4.8)$$

#### 4.4.1 Uložení tónových složek a výpočet jejich SPL

Označené a uložené spektrální linie jsou znovu srovnávány s okolními hodnotami. Pokud splňují podmínku (4.9),

$$X(k) - X(k + j) \geq 7dB \quad (4.9)$$

kde se  $j$  vybírá opět podle vzorkovacího kmitočtu a použité vrstvy formátu MPEG z tab. 4.3, jsou definitivně označeny jako tónové a jsou uvedeny identifikující hodnoty, které jednoznačně označují zvolenou složku. Každé takto zvolené tónové složce  $X_{(k)}$  se určí její hodnota  $X_{(tm)}$ , která vyjadřuje její SPL v dB. Dále se určí číslo spektrální linie a označí se tónovou značkou, tzv. tonal-flag. Všem ostatním složkám v mezích zkoumaného kmitočtového pásma je přidělena hodnota  $-\infty$ dB.

Layer I	$f_{vz} = 16$ kHz	$f_{vz} = 22,05; 24$ kHz
$j = -2, 2$	$2 < k < 96$	$2 < k < 64$
$j = -3, -2, 2, 3$	$96 \leq k < 192$	$64 \leq k < 128$
$j = -6, -3, -2, 2, 3, 6$	$192 \leq k < 250$	$128 \leq k < 250$
Layer II	$f_{vz} = 16$ kHz	$f_{vz} = 22,05; 24$ kHz
$j = -4, 4$	$4 < k < 192$	$4 < k < 128$
$j = -6, -4, -2, 2, 4, 6$	$192 \leq k < 384$	$128 \leq k < 256$
$j = -12, -6, -4, -2, 2, 4, 6, 12$	$382 \leq k < 500$	$256 \leq k < 500$

Tab. 4.3: Kritéria pro výběr tónových složek z [9]

#### 4.4.2 Uložení šumových složek a výpočet jejich SPL

Ze zbylých kmitočtových složek jsou po určení tónových součástí určovány ty netónové, resp. šumové, složky. K výpočtu šumových složek ze  $X_{(k)}$  je potřeba určit kritická pásma  $z_{(k)}$  za pomoci tabulek A.2. Při použití  $f_{vz} = 16$  kHz je třeba určit 21 kritických pásem a pro  $f_{vz} = 22,05$  nebo 24 kHz je potřeba 23 kritických pásem. Poté, co byly tonální složky vynulovány, jsou uvnitř každého pásma sečteny

energie zbývající spektrálních čar, aby vytvořily SPL nové, šumové, složky  $X_{(nm)}$  pro příslušné kritické pásmo. Těmto složkám jsou potom, stejně jako tónovým složkám výše, přiřazeny definující hodnoty v podobě hodnoty akustického tlaku  $X_{(nm)}$  v dB a šumové značky, tzv. non-tonal flag. Rozdíl oproti tónové složce je v přiřazení čísla spektrální čáry. Jelikož je šumová složka součtem více jednotlivých kmitočtových složek, je výsledné hodnotě šumové složky přiřazeno číslo složky nejbližší ke geometrickému středu daného kritického pásma.

## 4.5 Decimace počtu maskovacích složek

Decimace je proces snižování počtu maskovacích složek. To probíhá dvěma způsoby. Nejprve je potřeba rozhodnout, zda je maskující složka vůbec ve slyšitelné oblasti, čili nad úrovní absolutního maskovacího prahu ATH, k tomu se používá (4.10), resp. (4.11),

$$X_{tm}(k) \geq LT_g(k) \quad (4.10)$$

$$X_{nm}(k) \geq LT_g(k) \quad (4.11)$$

kde  $X_{tm}(k)$  a  $X_{nm}(k)$  jsou tónové, resp. šumové, složky získané v předchozím kroku a  $LT_q(k)$  je hodnota ATH při kmitočtu o indexu  $k$  zjištěných z A.1.

Druhý krok decimace je vyrušení příliš blízkých maskovacích složek. Složky, které jsou k sobě blíže než 0,5 Bark. Zachová se pouze složka nejsilnější, která bude schopná nejlépe maskovat slabší složky. Slabší složky jsou potom odstraněny ze seznamu tonálních částí, protože jsou již k dalším procesům nepotřebné. K této operaci se používá algoritmus, tzv. sliding window, který funguje podobně jako bubble sort, pouze s tím rozdílem, že data netřídí. Algoritmus je aplikován v oblasti kritických pásem a je nastaven na šířku, výše zmíněných, 0,5 Bark.

## 4.6 Výpočet individuálních maskovacích prahů

Z původních vzorků kmitočtové oblasti, označených  $k$  v množství  $N/2$ , kde je  $N$  určeno z 4.1, je použito pouze menší množství vzorku, označených indexem  $i$ . Teprve tyto prvky jsou použity k výpočtu celkového maskovacího prahu. Jednotlivé složky  $i$  jsou opět uvedeny v tab. A.2

Ve vrstvě Layer I nejsou kmitočtové složky odpovídajícího rozsahu, pokrytého prvními 6ti subpásmi, podvzorkovány. Avšak z dalších složek, které spadají do oblasti následujících 6ti subpásem, je použita pouze každá druhá spektrální čára. Pro následujících 18 subpásem se potom používá pouze každá čtvrtá spektrální čára. Pro Layer II se tyto čáry vybírají odlišně. Pro první tři subpásma pokrývající zkoumané

složky není použito podvzorkování, ale pro následující tři subpásma je již použita pouze každá druhá čára. Z dalších 6ti pásem se využívá pouze každá čtvrtá spektrální čára a z posledních 18ti subpásem se používá pouze každá osmá čára. Počet vzorků  $n$  v každém podvzorkované subpásmové kmitočtové oblasti závisí na použité vrstvě. Ve vrstvě Layer I je jich použito 108 a v Layer II je použito 132 vzorků.

Každé tónové a šumové složce je přiřazena hodnota  $i$  odpovídající nejbližší původní spektrální čáře  $X_{(k)}$ . Hodnota  $i$  je dána z tab. A.2 Individuální maskovací prahy se počítají pro tónovou složku podle (4.12) a pro šumovou podle (4.13),

$$LT_{tm} [z_{(j)}z_{(i)}] = X_{tm} [z_{(j)}] + av_{tm} [z_{(j)}] + vf_{tm} [z_{(j)}z_{(i)}] \quad (4.12)$$

$$LT_{nm} [z_{(j)}z_{(i)}] = X_{nm} [z_{(j)}] + av_{nm} [z_{(j)}] + vf_{nm} [z_{(j)}z_{(i)}] \quad (4.13)$$

kde v Bark a  $LT_{nm}$  jsou jednotlivé maskovací prahy tónové, resp. šumové, složky kritického pásma  $z$  v Bark maskovací složky v kritickém pásmu  $z_m$  v Bark. Výraz  $X_{tm} [z(j)]$  udává SPL maskující tónové složky  $j$  v odpovídajícím kritickém pásmu  $z(j)$ . To samé platí pro výraz  $X_{nm} [z(j)]$  pro šumovou složku.

Zbývající neznámé  $av$  a  $vf$  se nazývají maskovací číslo, resp. maskující funkce maskující složky  $X [z(j)]$ . Maskovací číslo se počítá rozdílně pro tónovou a šumovou složku podle rovnice (4.14), resp. rovnice (4.15).

$$av_{tm} = -1,525 - 0,275 \cdot z_j - 4,5\text{dB} \quad (4.14)$$

$$av_{nm} = -1,525 - 0,275 \cdot z_j - 0,5\text{dB} \quad (4.15)$$

Maskující funkce  $vf$  je závislá na vzdálenosti v  $dz = z(i) - z(j)$  k maskující složce v Bark. Proměnná  $i$  zde reprezentuje číslo spektrální linky na které je vypočítávána maskující funkce a  $j$  určuje tento maskovací člen. Čísla kritickým pásem lze opět nalézt v tab. A.2. Maskovací funkce je stejná pro tónové i šumové složky a je dána podle tab. 4.4.

$vf = 17 (dz + 1) - (0,4 X[z(j)] + 6)$	$-3 \leq dz < -1\text{Bark}$
$vf = dz(0,4 X[z(j)] + 6)$	$-1 \leq dz < 0\text{Bark}$
$vf = -17 dz$	$0 \leq dz < 1\text{Bark}$
$vf = -(dz - 1)(17 - 0,15 X[z(j)]) - 17$	$1 \leq dz < 8\text{Bark}$

Tab. 4.4: Hodnoty maskovací funkce  $z$  [9]

Hodnoty  $X [z(j)]$  udává SPL maskující složky  $j$  v dB a  $vf$  jsou proto počítány v dB. Kvůli vysoké složitosti se upustilo od využívání hodnot  $dz$  větších než 8 a menší

než -3 Bark. To je řešeno tak, že všem hodnotám  $LT_m$ , pro tónové i šumové složky, mimo toto rozmezí je přiřazena hodnota  $-\infty$ dB.

## 4.7 Výpočet celkového maskovacího prahu

Celková maskovací křivka  $LT_g(i)$  je odvozena z vyšších a nižších vrcholů individuálních maskovacích prahů pro každou z  $j$  tónových a šumových maskujících složek a z absolutního prahu slyšení  $LT_q(i)$ , které jsou mimo jiné ukázány v příloze A.1. Výsledný maskovací práh se zjistí součtem energií odpovídajících individuálním maskovacím prahům a absolutnímu prahu slyšení podle (4.16). Celkový počet všech tónových maskujících složek je dán hodnotou  $m$  a počet šumových maskujících složek hodnotou  $n$ . Pro dané  $i$  je vždy možné upravit hodnotu  $j$  tak, aby zahrnovala ty maskující složky, které se nacházejí v rozmezí  $-8$  až  $3$  Bark od  $i$ . Mimo tento rozsah jsou hodnoty  $LT_{tm}$  a  $LT_{nm}$  vždy  $-\infty$ dB.

$$LT_g(i) = 10 \log_{10} \left( 10^{LT_q(i)/10} + \sum_{j=1}^m 10^{LT_{tm}(z(j), z(i))/10} + \sum_{j=1}^n 10^{LT_{nm}(z(j), z(i))/10} \right) \quad (4.16)$$

## 4.8 Určení minimálního maskovacího prahu

Minimální maskovací úroveň  $LT_{\min}(n)$  v subpásmu  $n$  je uvedena v dB podle vztahu (4.17)

$$LT_{\min}(n) = \text{MIN}[LT_g(i)] \quad (4.17)$$

kde  $i$  je číslo subpásma a  $f_{(i)}$  je příslušný kmitočet v tab. A.1. Minimální maskovací práh je potom počítán pro každé subpásma. Tato nově vzniklá úroveň udává minimální hodnotu kmitočtové složky, která bude v případě výskytu slyšitelná.

## 4.9 Výpočet poměru SMR

Zjistit poměr signálu ku maskující složce je potom již jednoduchá operace, kdy se pro jednotlivé subpásma  $n$  odečítá od hodnoty akustického tlaku subpásma  $L_{sb}(n)$  hodnota minimálního maskovacího prahu  $LT_{\min}(n)$ . Čili podle (4.18).

$$SMR_{sb}(n) = L_{sb}(n) - LT_{\min}(n) \quad (4.18)$$

Tato hodnota je uvedena v dB a udává odstup slyšitelného signálu od maskující křivky. Tento odstup je vyjádřen v decibelové míře a udává, jak velký je dynamický rozsah mezi znějícím tónem, resp. znějící kmitočtovou složkou, a nově vypočítanou

maskovací křivkou, vzniklou spojením individuálních maskovacích vlastností jednotlivých maskovacích složek. Tento poměr potom využívá zvukový kodér, který při kódování zvukového signálu používá zjištěný odstup k definování nepotřebných signálních složek.



## 5 VÝPOČET PAM V PROSTŘEDÍ MATLAB

Tato kapitola využívá poznatky předchozích kapitol a zabývá se samotným určením výsledných maskovacích křivek zvukového signálu za použití metod psychoakustického modelu. Ten je simulovaný v prostředí Matlab verze 7, resp. 2009b. Tento software byl zvolen především kvůli přirozené podpoře funkcí, které jsou matematickým základem určení psychoakustického modelu. Hlavně ale proto, že nabízí jednoduchou práci s daty. V prostředí Matlab není potřeba deklarovat proměnné, navíc přirozeným datovým typem je `double`, čili typ s plovoucí desetinou čárkou. Programování v Matlab je velice podobné syntaxi a sémantice z prostředí C, což umožňuje použití již hotových funkcí a tím výrazně usnadňuje práci. Další nespornou výhodou je možnost jednoduše zobrazit data v různých přehledných grafech. Tyto grafy jsou v následující části použity. Prostředí Matlab má ale i své nevýhody, mezi které patří hlavně rychlost. Matlab zapsaný kód nekompile, ale překládá jej až když je žádán, což značně zpomaluje práci výsledného programu. Prostředí je vysoce maticově přizpůsobené, což znamená další zpomalení při provádění potřebných smyček. Psychoakustický model je počítán pro nižší kmitočtová pásma pro Layer I. Tyto vzorkovací kmitočty jsou standardizovány na hodnotách 16 kHz, 22,05 kHz a 24 kHz.

### 5.1 Spuštění programu a výběr souboru

Program nemá z důvodu možného pozdějšího použití ucelený samospustitelný charakter, proto je potřeba spouštět přímo soubor *psycho.m*, který se nachází na příloženém CD ve složce *Model* nebo pokud je jako zdrojová složka nastavena kmenová složka programu, lze výpočet spouštět přímo z prostředí Matlab zadáním příkazu `run psycho`.

Po inicializaci programu se ukáže výběrové okno, ve kterém je možno spustit jeden ze 3 vzorků, které se nachází ve složce *Model/vzorky*. Jedná se o tři vzorky 5.1 lišící se vzorkovací frekvencí. Ta je patrná z názvu souborů, a to 16 kHz, 22,05 kHz a 24 kHz. Obsah vzorků je sinusoidní zvukový signál o nelineárně se zvyšující frekvenci. Délka vzorků je 1 s. Pro jednotlivé vzorky je v programu definován vlastní absolutní práh slyšení, který je třeba pro výpočet celkového maskovacího prahu vzorku. Ve složce se nachází ještě jeden vzorek, a to *sin.wav*, což je původní vzorek se vzorkovací frekvencí  $f_{vz} = 44,1$  kHz. Program samozřejmě funguje i s jakýmkoliv jiným zvukovým souborem ve formátu *\*.wav*. Použití je omezeno pouze použitou vzorkovací frekvencí  $f_{vz}$ , která musí odpovídat frekvencím, pro které je definován práh absolutního slyšení v *absolutni\_prah.m*, čili již udaných 16 kHz, 22,05 kHz a 24 kHz. Nemělo by se ovšem jednat o zvukové formáty, které obsahují nějaká metadata. Tyto

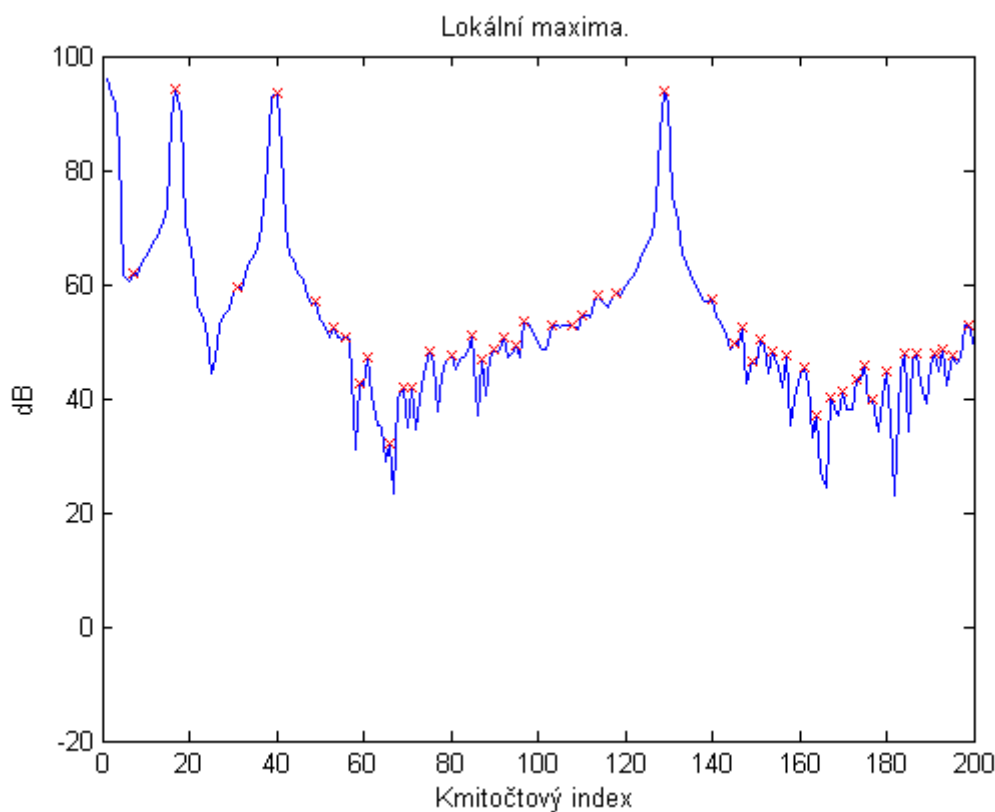


Obr. 5.1: Vložení souboru pro výpočet PAM.

vzorky Matlab odmítne s vypsáním `Index exceeds matrix dimensions`. Program nemá žádné uživatelské prostředí, k dalšímu kroku je třeba pouze odpauzovat výpočet zmáčknutím libovolné klávesy. Dále zobrazené grafy jsou vykresleny pro první vzorek signálu. Zvukové soubory jsou vytvořeny v délce 1000 ms, což znamená, že pro analýzu celého zvukového souboru je potřeba celou proceduru výpočtu PAM absolvovat několikrát. Kolikrát, to záleží na použité  $f_{vz}$  podle tab. 4.1.

## 5.2 Zjištění lokálních maxim

Prvním krokem při načtení souboru je zjištění základních vlastností vzorku, což jsou hodnoty signálu  $x$ . a vzorkovací frekvence  $f_{vz}$ . Ta dále v programu rozhoduje o použitých hodnotách absolutního prahu slyšení a výběru potřebných kritických pásem. Hodnotu  $x$  je potřeba připravit pro výpočet FFT. To probíhá tak, že vstupní vzorek je podle zvolené  $f_{vz}$  otestován na počet vzorků. Ten musí být dělitelný 384. Pokud tak není, je zbylé místo vyplněno nulami. Vzorek potom postupuje přímo do FFT analýzy, která vrátí kmitočtové spektrum. Další operací je aplikace Hannova okna podle (4.2). Následuje určení spektrální hustoty podle (4.3), což je velmi

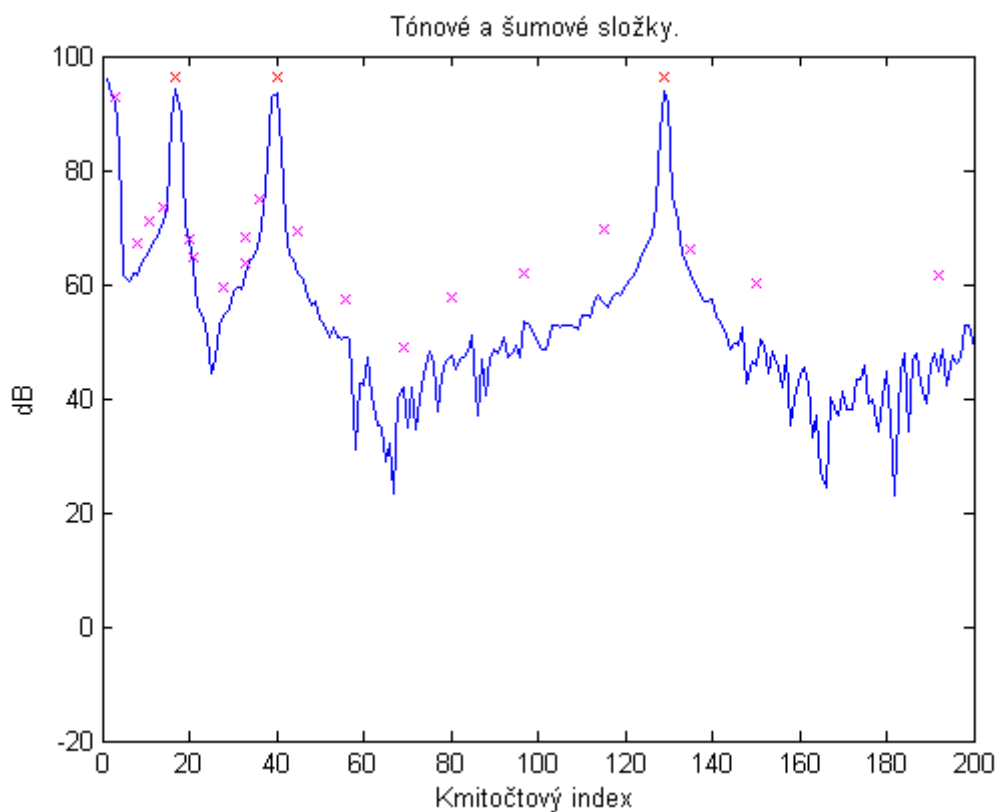


Obr. 5.2: Hodnoty určené jako lokální maxima.

vhodné pro následné zjištění rozložení maxim. Podle normy je ještě nezbytné normalizovat dynamický rozsah vzorku, resp. definovat minimální a maximální hodnoty rozsahu, tak, aby splňoval podmínku celkového rozsahu  $96 \text{ kHz}_{(\text{SPL})}$ . Samotná maxima jsou potom vybírána podle podmínek (4.7) a (4.8), kde  $X_{(k)}$  je hustota spektra vypočítaná dříve. Tím jsou označeny lokální maxima pro daný vzorek a jsou uloženy do databanky a zároveň zobrazeny v grafu 5.2 pro následné určování tónových a šumových složek.

### 5.3 Určení tónových a šumových složek

V dalším kroku programu probíhá určování tónových a šumových složek. Tyto složky se určují z banky lokálních maxim z předchozího kroku. Každé položka z množiny nalezených maxim prochází vylučovacím algoritmem, které porovnává jednotlivé označené položky podle tabulky 4.3. Pokud položka nesplňuje podmínky, je jí přiřazena hodnota  $is\_tonal = 0$ , pokud je splňuje, pokračuje k definici určujících vlastností tak, jak se píše v 4.4.1. Tyto položky jsou okamžitě uloženy to jednorozměrné matice *tonove* jako tónové složky signálu a později jsou vykresleny do grafu 5.3.

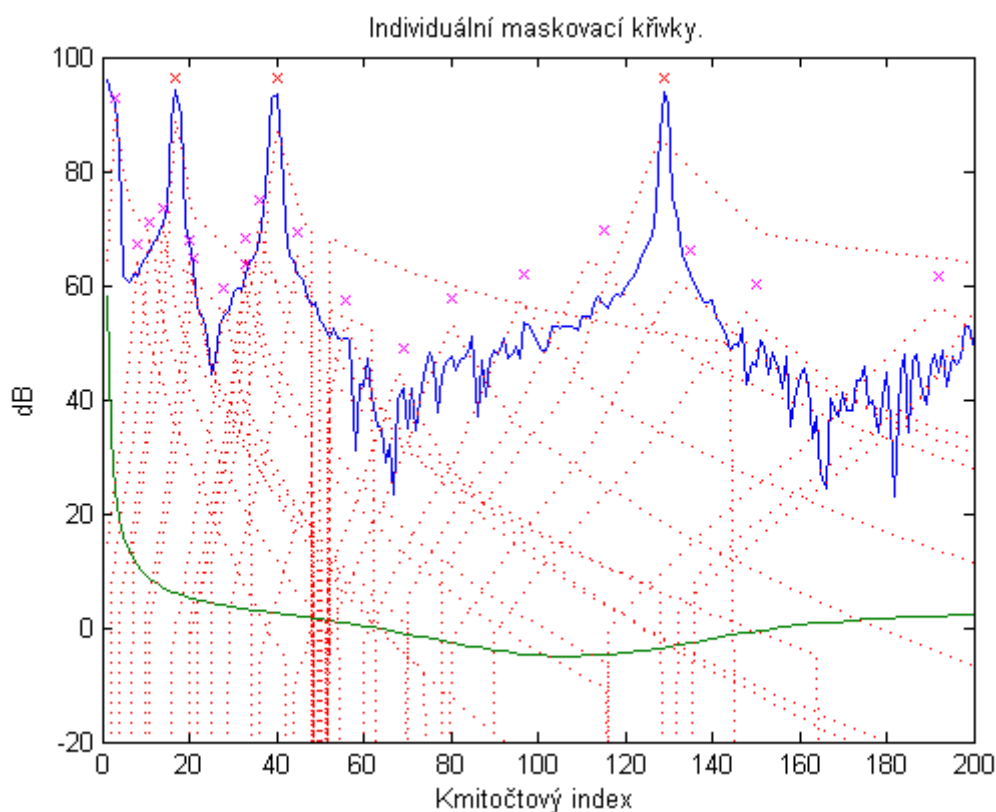


Obr. 5.3: Hodnoty určené jako tónové a šumové složky.

Určení šumových složek, na rozdíl od tónových, probíhá odlišně. Podle postupu v 4.4.2 se nejprve určí hodnota nenulových složek pro dané kritické pásmo, dále se z nich vypočítá jejich geometrický střed a ten se označí jako šumová složka. Hodnoty použitých kritických pásem jsou uvedeny v tabulce A.2. Při nahlédnutí do matice s názvem *sumove* jsou vidět hodnoty zvolených šumových značek závislé na absolutním prahu slyšení pro danou frekvenci a kritickém pásmu, do kterého šumová značka spadá. Hodnoty šumových složek jsou potom vykresleny do grafu 5.3 společně s původním signálem převedeným do kmitočtové oblasti a předem určených tónových složek. V programu se vykreslují nejprve tónové a až potom šumové složky, do dvou různých grafů. Zde jsou ale obě vyznačené složky vykresleny do jednoho grafu kvůli úspoře místa. Červené značky označují tónové složky a fialové reprezentují šumové složky, vypočítané pro jednotlivá kritická pásma. Z grafu je patrné, že šumových složek je mnohem více, než tónových. To je dáno především postupem, jakým jednotlivé složky vznikají. Tónové přímou volbou a šumové se průměrují ze zbytků. Tónových složek je potom tolik, kolik je ve vzorku použito kritických pásem.

## 5.4 Výpočet individuálních maskovacích prahů

Vytyčení individuálních maskovacích prahů probíhá za pomoci tónových a šumových složek. Podle funkcí (4.14), resp. (4.15), je vypočítáno maskovací číslo, které se následně společně s hodnotou maskovacího funkce z tab. 4.4 připočítá k hodnotě tónové či šumové složky. Tím vznikne nová hodnota maskovací funkce LT pro danou šumovou nebo tónovou složku. Tento proces se opakuje pro celé kritické pásmo. Jednotlivé hodnoty LT jsou vyneseny do grafu 5.4. V programu se postupně vykreslují tónové, a jako další krok potom šumové, složky vzhledem k absolutnímu prahu daného použitou  $f_{vz}$ . Zde je pro úsporu místa zobrazen pouze společný graf s již vyznačenými tónovými, a zároveň i šumovými, maskujícími funkcemi.

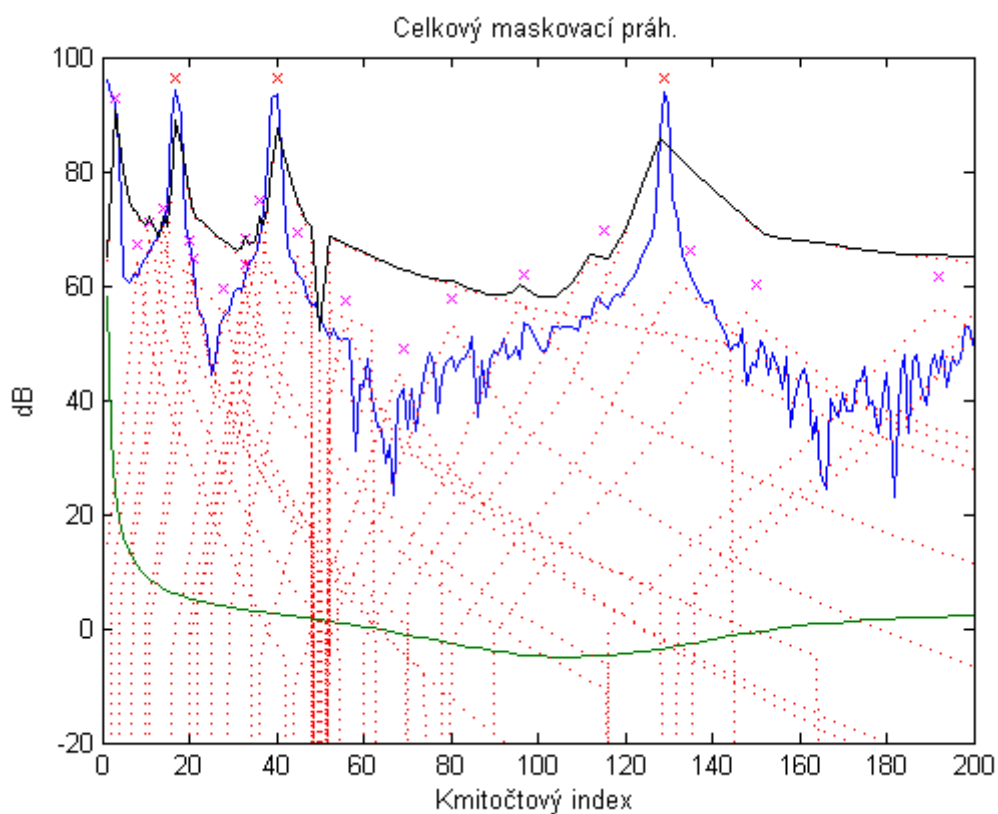


Obr. 5.4: Určení individuálních maskovacích prahů.

## 5.5 Výpočet celkového maskovacího prahu

Nově vzniklá celková křivka maskování je vykreslena černě v grafu 5.5. Jedná se v podstatě o spojení jednotlivých nově vzniklých maskovacích funkcí pro tónové a šumové složky. Celkový práh tyto křivky upravuje vzhledem k ostatním špičkovým

hodnotám a bere v potaz i jejich následné maskující funkce. Při použití jiného vzorku by se mohlo stát, že se vyskytne více tónových složek blízko sebe tak, jak se uvádí v 4.5. Proto je v programu implementována funkce decimace, která je definovaná v souboru *Decimace.m*. Tato funkce vypustí slabší složky v případě, že jsou od sebe jednotlivé složky vzdáleny o méně než  $0,5 \text{ Bark}^1$ , nebo složky, která se nachází pod hranicí ATH. Tato funkce navrácí jednorozměrné matice *Dsumove* a *Dtonalni*, které obsahují hodnoty jednotlivých složek po decimaci. Tyto jsou potom v programu opět vykresleny do grafu. Tento graf zde ale chybí, protože u testovaných vzorků nebyla žádná složka vypuštěna, tudíž není potřeba zobrazovat dvakrát, v podstatě stejný, graf.



Obr. 5.5: Určení celkového maskovacího prahu.

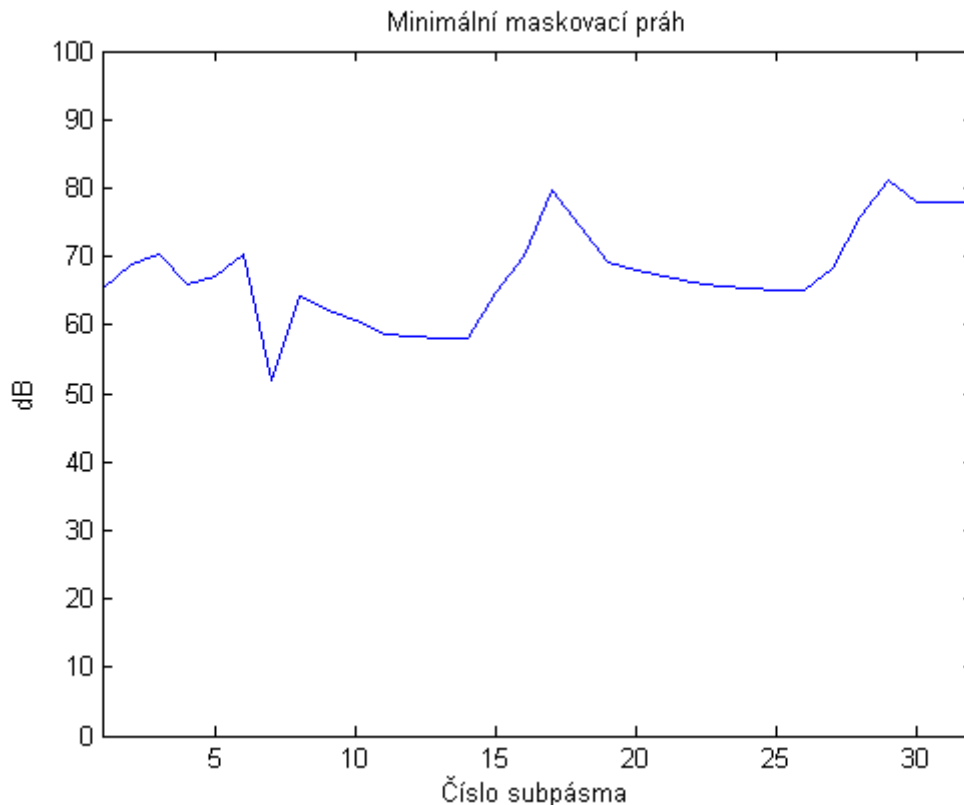
Z grafu 5.5 lze vyčíst, že PAM pro nižší vzorkovací frekvence  $f_{vz}$  z analyzovaného vzorku ořezal více než polovinu nově vzniklých maskujících funkcí. To je dáno jejich pozicí pod hlavní maskovací funkcí tónové složky, která je natolik silná, že překrývá bez problému několik nižších šumových maskovacích funkcí. To v podstatě znamená zbytečné výpočty, které je nutné počítat, ikdyž tyto maskující složky nejsou využity.

<sup>1</sup>Během testů různých vzorků jsem se ale nesetkal s případem, kdy by této funkce bylo opravdu potřeba. Nicméně norma ji vyžaduje, kvůli určení nejlépe maskující složky.

Pokud by v PAM existovala funkce, která by porovnávala jednotlivé tónové a šumové složky ihned, resp. po jejich vyzvednutí ze seznamu lokálních maxim, bylo by možné ulehčit výpočty eliminací na zjevně nepotřebných maskovacích složek šumových.

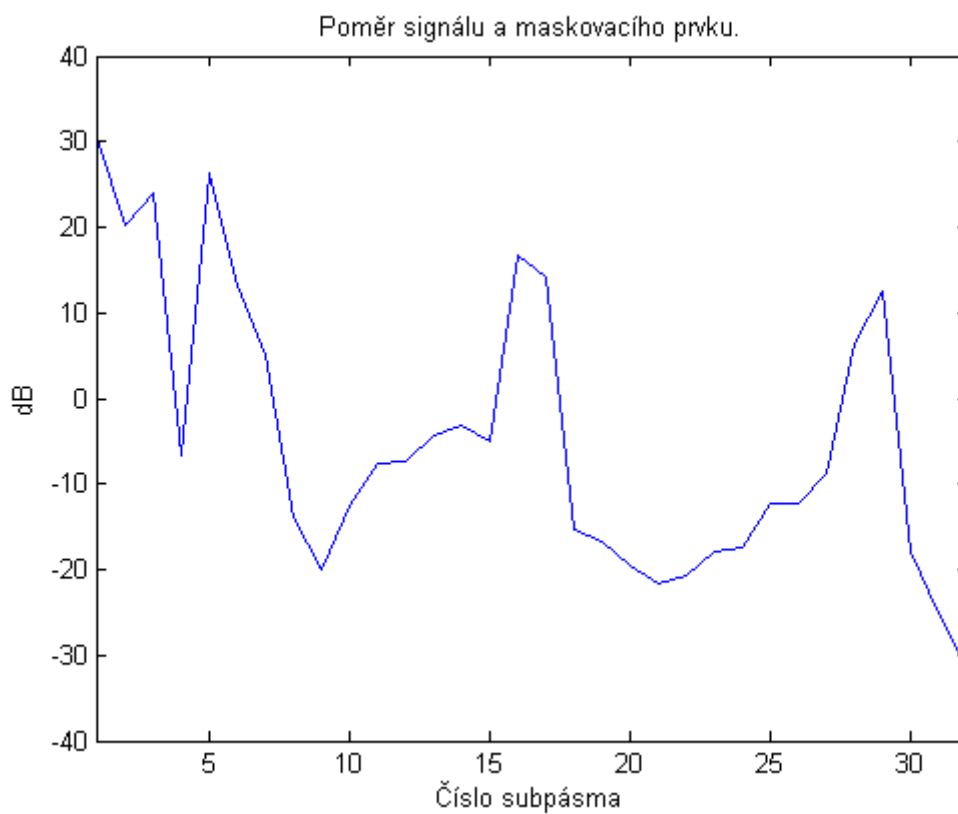
## 5.6 Určení Signal-to-Mask poměru a maskovacích úrovní jednotlivých subpássem

Výsledným krokem výpočtů PAM v prostředí Matlab je výstup v podobě celkových maskovacích úrovní pro jednotlivá subpásma. Ty jsou znázorněny na grafu 5.6. Jelikož je pro lidské ucho nezbytné znát hodnoty maskovacího prahu pro jednotlivá subpásma, jsou tedy tyto hodnoty uvedeny v matici  $LT_{min}$  v hlavním souboru *psycho.m* a mohou být použity jako vstup pro kodér akustického signálu. Udávají minimální hodnoty signálu, který by byl slyšitelný při současném zpracování původního vzorku.



Obr. 5.6: Určení maskovacích úrovní pro jednotlivá subpásma.

Na grafu 5.7 je následně zobrazena ještě poslední potřebná funkce, a to výpočet odstupů signálu od maskujícího členu. Hodnoty zobrazené v záporné části grafu



Obr. 5.7: Určení odstupu signálu od maskovacího prahu.

napovídají, že zkoumaný signál se nachází pod nově vzniklou celkovou maskující křivkou, je potřeba znát pouze hodnotu maskující funkce. To jsou směrodatné informace pro možný kódér signálu.



## 6 ZÁVĚR

Tato práce se zabývala psychoakustickým modelem lidského sluchu a jeho použitím. Jako cíl bylo zadáno seznámení se a prostudování jeho použití ve standardu MPEG a jeho následnou implementaci v prostředí Matlab.

První část zadání, seznámení se s modelem, je uskutečněna v prvních třech kapitolách práce. Seznámení se se základy fyzikálních vlastností zvuku, a hlavně fyziologií lidského sluchu, považuji za nezbytné pro správné pochopení problematiky psychoakustického modelu. Především jsou tyto vědomosti nezbytné k porozumění maskovacích jevů a jejich možného využití při kompresi zvukového signálu.

V části Psychoakustika jsou zavedeny důležité pojmy, především z oblasti nedokonalosti lidského sluchu a možnosti využití těchto nedokonalostí pro úsporu dat při kompresi. Na názorných obrázcích a grafech jsou principiálně zobrazeny základní maskovací funkce psychoakustiky a slouží tak k bližšímu porozumění problematice maskování frekvencí, které jsou, z poslechového hlediska, nepotřebné.

Další část zadání, prostudování použití PAM ve standardu MPEG, je realizováno v kapitole 4, kde se opírám především o normu ISO <sup>1</sup>, ze které celý PAM vychází. Tato část by měla čtenáře seznámit s možnostmi PAM a definuje jeho základní funkce a metody. V této části je definován PAM pro vrstvu Layer I a Layer II, ale v praktické části jsem se zabýval pouze modelem pro Layer I, ve které se nevyužívá časového maskování.

Poslední část zadání, implementace v prostředí Matlab, je realizována pro vstupní vzorky, které se nachází, společně s kódem programu, na příloženém CD. Na těchto vzorcích je možné simulovat výpočet hodnot potřebných pro definování maskujících prahů. Je možné použít i jiných vzorků, které ovšem musí splňovat určitá kritéria, především musí splňovat podmínku vzorkovací frekvence. Postupný průběh výpočtů je provázen názorným zobrazením jednotlivých hodnot. Grafy jsou výstupem přímo z programu Matlab. V Matlabu je realizován kompletní kódér PAM a může být dále implementován pro kódér zvukových signálů.

Celková názornost výpočtu PAM vede k zamyšlení nad funkcí a efektivitou použitého modelu. Je zde vidět, že PAM Layer I pro nižší vzorkovací frekvence je sice velmi efektivní z hlediska úspory dat, je ale velmi neefektivní z hlediska výpočtů. Přidáním několika funkcí by bylo možné eliminovat některé výpočty, které jsou, z hlediska celkového maskovacího prahu, nadbytečné. Z výsledných grafů je také patrné, kolik informací lidské ucho vůbec není schopné zpracovat a tento prostor se tak dá využít k implementaci dalších dat, např. metadata u komerčních zvukových formátů, vodoznaků, či nejrůznější kódovaná data.

---

<sup>1</sup>ISO/IEC 13818-3, Information technology - Generic coding of moving pictures and associated audio information Part 3: Audio

Tato práce čtenářům přináší bližší pohled na možnosti zvukové komprese a celkově na číslicové zpracování zvuku. Na základě znalostí získaných z této práce mohou být testovány další reálné vzorky a samotný model může být rozšířen pro použití s vyšší vzorkovací frekvencí a také další vrstvy, Layer II.

## LITERATURA

- [1] ADAM, P. *Problematika vzorkovania zvuku* [online]. 2006 [cit. 2011-05-19]. Úvod do metód spracovania zvuku v súčasnom multimediálnom prostredí. Dostupné z WWW: <<http://zvuk.atrip.sk/index.php?site=31>>.
- [2] DAVIS, Pan. A tutorial on MPEG/Audio Compression. *IEEE Multimedia Journal*. 1995, 3, s. 1-12.
- [3] *Digital Preservation*. [online]. 2010 [cit. 2011-05-19]. PCM, Pulse Code Modulated Audio. Dostupné z WWW: <<http://www.digitalpreservation.gov/formats/fdd/fdd000016.shtml>>.
- [4] FASTL, H. *Psychoacustics: Facts and Models* Springer, United States, 2nd edition, 1999.
- [5] FOUKAL, J. *Počítače a hudba* [online]. 2003 [cit. 2011-05-19]. MPEG Audio Layer-3. Dostupné z WWW: <<http://www.fi.muni.cz/qruzicka/>>.
- [6] HÁLEK, J., et al *Biofyzika pro bakaláře* Olomouc : Univerzita Palackého, 2002. 213 s.
- [7] HEITEL, T. *Využití psychoakustického modelu a transformace typu wavelet packet pro vodoznačení audio signálů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2010. 85 s, 3 přílohy. Vedoucí diplomové práce Mgr. Pavel Rajmic, Ph.D.
- [8] HENDERSON, T. *The Physics Classroom* [online]. 1998 [cit. 2011-05-19]. The Speed of Sound. Dostupné z WWW: <<http://www.physicsclassroom.com/class/sound/u1112c.cfm>>.
- [9] ISO/IEC 13818-3. *Information technology : - Generic coding of moving pictures and associated audio information - Part 3: Audio*. Genève : ISO/IEC, 1995. 100 s.
- [10] JAROŠ, V. *Experimenty s digitalizovaným zvukem - mp3*. Praha, 2007. 55 s. Bakalářská práce. ČVUT Praha.
- [11] KILGORE, M., WU, J. *MPEG, the MP3 Standard, and Audio Compression* [online]. Stanford, 2003. 12 s. Oborová práce. Stanford University. Dostupné z WWW: <[http://www-ee.stanford.edu/osgood/Sophomore College//Audio Compression and the MP3 Standard.pdf](http://www-ee.stanford.edu/osgood/Sophomore%20College/Audio%20Compression%20and%20the%20MP3%20Standard.pdf)>.

- [12] MALÝ, L. *Stáří a sluch* [[online]. [s.l.], 2005. 3 s. Referát. VŠE Praha. Dostupné z WWW: <<http://sik.vse.cz/ss/stariasluch.pdf>>.
- [13] NETTER, F. H. *Anatomický atlas člověka* 1. vyd. Praha: Grada, 2003. 79 s. ISBN 80-247-0517-6
- [14] SKARNITZL, R. *Akustika řeči* [online]. 2008 [cit. 2011-05-19]. Psychoakustika. Dostupné z WWW: <[http://fu.ff.cuni.cz/vyuka/akustika/3\\_psychoakustika.pdf](http://fu.ff.cuni.cz/vyuka/akustika/3_psychoakustika.pdf)>.
- [15] SVOBODA, R. *Experimenty s psychoakustickou kompresí zvuku*. Praha, 2010. 53 s. Bakalářská práce. ČVUT Praha.
- [16] SVOBODA, V. Základy akustiky (5. díl). *DreamFace.net* [online]. 2004, 1, [cit. 2011-05-19]. Dostupný z WWW: <<http://www.dreamface.net/modules.php?name=Newsfile=articlesid=199>>.
- [17] TŮMA, J. Antialiasing. In *Vzorkování spojitých signálů a jejich obnova* [online]. 2002 [cit. 2011-05-19]. Dostupné z WWW: <<http://homel.vsb.cz/~tum52/Projekty/Antialias.pdf>>.
- [18] Ucho. In *Wikipedia : the free encyclopedia* [online]. St. Petersburg (Florida) : Wikipedia Foundation, 27. 12. 2003, last modified on 1. 5. 2011 [cit. 2011-05-24]. Dostupné z WWW: <<http://cs.wikipedia.org/wiki/Ucho>>..
- [19] WOLFE, J. *Music Acoustics* [online]. 2007 [cit. 2011-05-19]. Note names, MIDI numbers and frequencies. Dostupné z WWW: <<http://www.phys.unsw.edu.au/jw/notes.html>>.

## SEZNAM SYMBOLŮ, VELIČIN A ZKRATEK

- A/D analogově/číslicový převodník - analog/digital convertor
- ATH Absolutní práh slyšení - Absolute Threshold of Hearing
- DAB digitální audio vysílání - Digital Audio Broadcast
- FFT Rychlá Fourierova transformace - Fast Fourier Transformation
- $f_{vz}$  Vzorkovací kmitočet
- Hz Jednotka kmitočtu v SI - Hertz
- IEC Mezinárodní elektrotechnická komise - International Elektro-Technical Commission
- ISO Mezinárodní organizace pro standardy - International Standards Organization
- MPEG Expertní skupina pro pohyblivé obrazy - Moving Picture Experts Group
- PAM PsychoAkustický Model - PsychoAcoustic Model
- RMS efektivní hodnota časově proměnného signálu - Root Mean Squared
- SMR Poměr signálu ku maskovacímu prahu - Signal-to-Mask Ratio
- SNR Odstup signálu od šumu - Signal-noise Ratio
- SPL Hladina akustického tlaku - Sound Pressure Level

## SEZNAM PŘÍLOH

<b>A</b>	<b>Tabulky pro výpočet PAM</b>	<b>47</b>
A.1	Tabulky absolutních prahů . . . . .	47
A.2	Tabulky kritických pásem . . . . .	50
<b>B</b>	<b>Obsah přiloženého CD</b>	<b>52</b>

# A TABULKY PRO VÝPOČET PAM

## A.1 Tabulky absolutních prahů

ISO/IEC 13818-3:1994(E)

©ISO/IEC

**Table D.1a. - Frequencies, critical band rates and absolute threshold**

Table is valid for Layer I at a sampling rate of 16 kHz.

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Thresh. [dB]
1	31,25	0,309	58,23
2	62,50	0,617	33,44
3	93,75	0,925	24,17
4	125,00	1,232	19,20
5	156,25	1,538	16,05
6	187,50	1,842	13,87
7	218,75	2,145	12,26
8	250,00	2,445	11,01
9	281,25	2,742	10,01
10	312,50	3,037	9,20
11	343,75	3,329	8,52
12	375,00	3,618	7,94
13	406,25	3,903	7,44
14	437,50	4,185	7,00
15	468,75	4,463	6,62
16	500,00	4,736	6,28
17	531,25	5,006	5,97
18	562,50	5,272	5,70
19	593,75	5,533	5,44
20	625,00	5,789	5,21
21	656,25	6,041	5,00
22	687,50	6,289	4,80
23	718,75	6,532	4,62
24	750,00	6,770	4,45
25	781,25	7,004	4,29
26	812,50	7,233	4,14
27	843,75	7,457	4,00
28	875,00	7,677	3,86
29	906,25	7,892	3,73
30	937,50	8,103	3,61
31	968,75	8,309	3,49
32	1000,00	8,511	3,37
33	1031,25	8,708	3,26
34	1062,50	8,901	3,15
35	1093,75	9,090	3,04
36	1125,00	9,275	2,93
37	1156,25	9,456	2,83
38	1187,50	9,632	2,73
39	1218,75	9,805	2,63
40	1250,00	9,974	2,53
41	1281,25	10,139	2,42
42	1312,50	10,301	2,32
43	1343,75	10,459	2,22
44	1375,00	10,614	2,12
45	1406,25	10,765	2,02
46	1437,50	10,913	1,92
47	1468,75	11,058	1,81
48	1500,00	11,199	1,71
49	1562,50	11,474	1,49
50	1625,00	11,736	1,27
51	1687,50	11,988	1,04
52	1750,00	12,230	0,80
53	1812,50	12,461	0,55
54	1875,00	12,684	0,29

Index Number i	Frequency [Hz]	Crit. Band Rate [z]	Absolute Thresh. [dB]
55	1937,50	12,898	0,02
56	2000,00	13,104	-0,25
57	2062,50	13,302	-0,54
58	2125,00	13,493	-0,83
59	2187,50	13,678	-1,12
60	2250,00	13,855	-1,43
61	2312,50	14,027	-1,73
62	2375,00	14,193	-2,04
63	2437,50	14,354	-2,34
64	2500,00	14,509	-2,64
65	2562,50	14,660	-2,93
66	2625,00	14,807	-3,22
67	2687,50	14,949	-3,49
68	2750,00	15,087	-3,74
69	2812,50	15,221	-3,98
70	2875,00	15,351	-4,20
71	2937,50	15,478	-4,40
72	3000,00	15,602	-4,57
73	3125,00	15,841	-4,82
74	3250,00	16,069	-4,96
75	3375,00	16,287	-4,98
76	3500,00	16,496	-4,90
77	3625,00	16,697	-4,70
78	3750,00	16,891	-4,39
79	3875,00	17,078	-3,99
80	4000,00	17,259	-3,51
81	4125,00	17,434	-2,99
82	4250,00	17,605	-2,45
83	4375,00	17,770	-1,90
84	4500,00	17,932	-1,37
85	4625,00	18,089	-0,86
86	4750,00	18,242	-0,39
87	4875,00	18,392	0,03
88	5000,00	18,539	0,40
89	5125,00	18,682	0,72
90	5250,00	18,823	1,00
91	5375,00	18,960	1,24
92	5500,00	19,095	1,44
93	5625,00	19,226	1,62
94	5750,00	19,356	1,78
95	5875,00	19,482	1,92
96	6000,00	19,606	2,05
97	6125,00	19,728	2,18
98	6250,00	19,847	2,30
99	6375,00	19,964	2,42
100	6500,00	20,079	2,55
101	6625,00	20,191	2,69
102	6750,00	20,300	2,82
103	6875,00	20,408	2,97
104	7000,00	20,513	3,13
105	7125,00	20,616	3,29
106	7250,00	20,717	3,46
107	7375,00	20,815	3,65
108	7500,00	20,912	3,84

**Table D.1b. - Frequencies, critical band rates and absolute threshold**  
Table is valid for Layer I at a sampling rate of 22,05 kHz.

Index Number i	Frequency [Hz]	Crit.Band Rate [z]	Absolute Thresh. [dB]	Index Number i	Frequency [Hz]	Crit.Band Rate [z]	Absolute Thresh. [dB]
1	43,07	0,425	45,05	55	2670,12	14,909	-3,41
2	86,13	0,850	25,87	56	2756,25	15,100	-3,77
3	129,20	1,273	18,70	57	2842,38	15,284	-4,09
4	172,27	1,694	14,85	58	2928,52	15,460	-4,37
5	215,33	2,112	12,41	59	3014,65	15,631	-4,60
6	258,40	2,525	10,72	60	3100,78	15,796	-4,78
7	301,46	2,934	9,47	61	3186,91	15,955	-4,91
8	344,53	3,337	8,50	62	3273,05	16,110	-4,97
9	387,60	3,733	7,73	63	3359,18	16,260	-4,98
10	430,66	4,124	7,10	64	3445,31	16,406	-4,96
11	473,73	4,507	6,56	65	3531,45	16,547	-4,88
12	516,80	4,882	6,11	66	3617,58	16,685	-4,74
13	559,86	5,249	5,72	67	3703,71	16,820	-4,54
14	602,93	5,608	5,37	68	3789,84	16,951	-4,30
15	646,00	5,959	5,07	69	3875,98	17,079	-4,02
16	689,06	6,301	4,79	70	3962,11	17,205	-3,71
17	732,13	6,634	4,55	71	4048,24	17,327	-3,37
18	775,20	6,959	4,32	72	4134,38	17,447	-3,00
19	818,26	7,274	4,11	73	4306,64	17,680	-2,25
20	861,33	7,581	3,92	74	4478,91	17,905	-1,50
21	904,39	7,879	3,74	75	4651,17	18,121	-0,81
22	947,46	8,169	3,57	76	4823,44	18,331	-0,18
23	990,53	8,450	3,40	77	4995,70	18,534	0,35
24	1033,59	8,723	3,25	78	5167,97	18,731	0,79
25	1076,66	8,987	3,10	79	5340,23	18,922	1,15
26	1119,73	9,244	2,95	80	5512,50	19,108	1,44
27	1162,79	9,493	2,81	81	5684,77	19,289	1,68
28	1205,86	9,734	2,67	82	5857,03	19,464	1,89
29	1248,93	9,968	2,53	83	6029,30	19,635	2,07
30	1291,99	10,195	2,39	84	6201,56	19,801	2,24
31	1335,06	10,416	2,25	85	6373,83	19,963	2,41
32	1378,13	10,629	2,11	86	6546,09	20,120	2,59
33	1421,19	10,836	1,97	87	6718,36	20,273	2,78
34	1464,26	11,037	1,83	88	6890,63	20,421	2,98
35	1507,32	11,232	1,68	89	7062,89	20,565	3,19
36	1550,39	11,421	1,53	90	7235,16	20,705	3,43
37	1593,46	11,605	1,38	91	7407,42	20,840	3,68
38	1636,52	11,783	1,23	92	7579,69	20,972	3,95
39	1679,59	11,957	1,07	93	7751,95	21,099	4,24
40	1722,66	12,125	0,90	94	7924,22	21,222	4,56
41	1765,72	12,289	0,74	95	8096,48	21,342	4,89
42	1808,79	12,448	0,56	96	8268,75	21,457	5,25
43	1851,86	12,603	0,39	97	8441,02	21,569	5,64
44	1894,92	12,753	0,21	98	8613,28	21,677	6,05
45	1937,99	12,900	0,02	99	8785,55	21,781	6,48
46	1981,05	13,042	-0,17	100	8957,81	21,882	6,95
47	2024,12	13,181	-0,36	101	9130,08	21,980	7,44
48	2067,19	13,317	-0,56	102	9302,34	22,074	7,96
49	2110,26	13,450	-0,76	103	9474,61	22,165	8,52
50	2153,32	13,578	-0,96	104	9646,88	22,253	9,10
51	2196,39	13,703	-1,16	105	9819,14	22,338	9,72
52	2239,45	13,826	-1,38	106	9991,41	22,420	10,37
53	2282,51	13,946	-1,61	107	10163,67	22,499	11,06
54	2325,59	14,062	-1,86	108	10335,94	22,576	11,79



**Table D.1c. - Frequencies, critical band rates and absolute threshold**  
Table is valid for Layer I at a sampling rate of 24 kHz.

Index Number i	Frequency [Hz]	Crit.Band Rate [z]	Absolute Thresh. [dB]	Index Number i	Frequency [Hz]	Crit.Band Rate [z]	Absolute Thresh. [dB]
1	46,88	0,463	42,10	55	2906,25	15,415	-4,30
2	93,75	0,925	24,17	56	3000,00	15,602	-4,57
3	140,63	1,385	17,47	57	3093,75	15,783	-4,77
4	187,50	1,842	13,87	58	3187,50	15,956	-4,91
5	234,38	2,295	11,60	59	3281,25	16,124	-4,98
6	281,25	2,742	10,01	60	3375,00	16,287	-4,98
7	328,13	3,184	8,84	61	3468,75	16,445	-4,94
8	375,00	3,618	7,94	62	3562,50	16,598	-4,84
9	421,88	4,045	7,22	63	3656,25	16,746	-4,66
10	468,75	4,463	6,62	64	3750,00	16,891	-4,43
11	515,63	4,872	6,12	65	3843,75	17,032	-4,15
12	562,50	5,272	5,70	66	3937,50	17,169	-3,82
13	609,38	5,661	5,33	67	4031,25	17,303	-3,45
14	656,25	6,041	5,00	68	4125,00	17,434	-3,06
15	703,13	6,411	4,71	69	4218,75	17,563	-2,66
16	750,00	6,770	4,45	70	4312,50	17,688	-2,24
17	796,88	7,119	4,21	71	4406,25	17,811	-1,83
18	843,75	7,457	4,00	72	4500,00	17,932	-1,43
19	890,63	7,785	3,79	73	4687,50	18,166	-0,68
20	937,50	8,103	3,61	74	4875,00	18,392	-0,02
21	984,38	8,410	3,43	75	5062,50	18,611	0,52
22	1031,25	8,708	3,26	76	5250,00	18,823	0,97
23	1078,13	8,996	3,09	77	5437,50	19,028	1,32
24	1125,00	9,275	2,93	78	5625,00	19,226	1,60
25	1171,88	9,544	2,78	79	5812,50	19,419	1,83
26	1218,75	9,805	2,63	80	6000,00	19,606	2,03
27	1265,63	10,057	2,47	81	6187,50	19,788	2,22
28	1312,50	10,301	2,32	82	6375,00	19,964	2,41
29	1359,38	10,537	2,17	83	6562,50	20,135	2,60
30	1406,25	10,765	2,02	84	6750,00	20,300	2,81
31	1453,13	10,986	1,86	85	6937,50	20,461	3,03
32	1500,00	11,199	1,71	86	7125,00	20,616	3,27
33	1546,88	11,406	1,55	87	7312,50	20,766	3,53
34	1593,75	11,606	1,38	88	7500,00	20,912	3,82
35	1640,63	11,800	1,21	89	7687,50	21,052	4,12
36	1687,50	11,988	1,04	90	7875,00	21,188	4,46
37	1734,38	12,170	0,86	91	8062,50	21,318	4,82
38	1781,25	12,347	0,67	92	8250,00	21,445	5,20
39	1828,13	12,518	0,49	93	8437,50	21,567	5,62
40	1875,00	12,684	0,29	94	8625,00	21,684	6,07
41	1921,88	12,845	0,09	95	8812,50	21,797	6,54
42	1968,75	13,002	-0,11	96	9000,00	21,906	7,06
43	2015,63	13,154	-0,32	97	9187,50	22,012	7,60
44	2062,50	13,302	-0,54	98	9375,00	22,113	8,18
45	2109,38	13,446	-0,75	99	9562,50	22,210	8,80
46	2156,25	13,586	-0,97	100	9750,00	22,304	9,46
47	2203,13	13,723	-1,20	101	9937,50	22,395	10,15
48	2250,00	13,855	-1,43	102	10 125,00	22,482	10,89
49	2343,75	14,111	-1,88	103	10 312,50	22,566	11,67
50	2437,50	14,354	-2,34	104	10 500,00	22,646	12,50
51	2531,25	14,585	-2,79	105	10 687,50	22,724	13,37
52	2625,00	14,807	-3,22	106	10 875,00	22,799	14,29
53	2718,75	15,018	-3,62	107	11 062,50	22,871	15,26
54	2812,50	15,221	-3,98	108	11 250,00	22,941	16,28

## A.2 Tabulky kritických pásem

ISO/IEC 13818-3:1994(E)

©ISO/IEC

**Table D.2a. - Critical band boundaries**

This table is valid for Layer I at a sampling rate of 16 kHz.  
The frequencies represent the top end of each critical band.

no	index of Table F&CB	frequency [Hz]	Bark [z]
0	3	93,75	0,925
1	7	218,75	2,145
2	10	312,50	3,037
3	13	406,25	3,903
4	17	531,25	5,006
5	21	656,25	6,041
6	25	781,25	7,004
7	30	937,50	8,103
8	35	1093,75	9,090
9	40	1250,00	9,974
10	47	1468,75	11,058
11	51	1687,50	11,988
12	55	1937,50	12,898
13	61	2312,50	14,027
14	67	2687,50	14,949
15	74	3250,00	16,069
16	79	3875,00	17,078
17	84	4500,00	17,932
18	91	5375,00	18,960
19	99	6375,00	19,964
20	108	7500,00	20,912

**Table D.2b. - Critical band boundaries**

This table is valid for Layer I at a sampling rate of 22,05 kHz.  
The frequencies represent the top end of each critical band.

no	index of Table F&CB	frequency [Hz]	Bark [z]
0	2	86,13	0,850
1	5	215,33	2,112
2	7	301,46	2,934
3	10	430,66	4,124
4	12	516,80	4,882
5	15	646,00	5,959
6	18	775,20	6,959
7	21	904,39	7,879
8	25	1076,66	8,987
9	29	1248,93	9,968
10	34	1464,26	11,037
11	39	1679,59	11,957
12	46	1981,05	13,042
13	51	2325,59	14,062
14	55	2670,12	14,909
15	61	3186,91	15,955
16	68	3789,84	16,951
17	74	4478,91	17,905
18	79	5340,23	18,922
19	85	6373,83	19,963
20	92	7579,69	20,972
21	101	9130,08	21,980
22	108	10 335,94	22,576

**Table D.2c. - Critical band boundaries**

This table is valid for Layer I at a sampling rate of 24 kHz.  
The frequencies represent the top end of each critical band.

no	index of Table F&CB	frequency [Hz]	Bark [z]
0	2	93,75	0,925
1	4	187,50	1,842
2	7	328,13	3,184
3	9	421,88	4,045
4	11	515,63	4,872
5	14	656,25	6,041
6	17	796,88	7,119
7	20	937,50	8,103
8	23	1078,13	8,996
9	27	1265,63	10,057
10	31	1453,13	10,986
11	36	1687,50	11,988
12	42	1968,75	13,002
13	49	2343,75	14,111
14	53	2718,75	15,018
15	58	3187,50	15,956
16	65	3843,75	17,032
17	72	4500,00	17,932
18	77	5437,50	19,028
19	82	6375,00	19,964
20	89	7687,50	21,052
21	97	9187,50	22,012
22	108	11 250,00	22,941

**Table D.2d. - Critical band boundaries**

This table is valid for Layer II at a sampling rate of 16 kHz.  
The frequencies represent the top end of each critical band.

no	index of Table F&CB	frequency [Hz]	Bark [z]
0	6	93,75	0,925
1	13	203,13	1,994
2	20	312,50	3,037
3	27	421,88	4,045
4	34	531,25	5,006
5	42	656,25	6,041
6	49	781,25	7,004
7	54	937,50	8,103
8	59	1093,75	9,090
9	64	1250,00	9,974
10	71	1468,75	11,058
11	75	1687,50	11,988
12	79	1937,50	12,898
13	85	2312,50	14,027
14	91	2687,50	14,949
15	98	3250,00	16,069
16	103	3875,00	17,078
17	108	4500,00	17,932
18	115	5375,00	18,960
19	123	6375,00	19,964
20	132	7500,00	20,912

## B OBSAH PŘILOŽENÉHO CD

