VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

DIPLOMOVÁ PRÁCE

Brno, 2021

Bc. Tomáš Peloušek

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ
DEPARTMENT OF TELECOMMUNICATIONS

SIMULACE ZKRESLENÍ ZVUKOVÉHO SIGNÁLU V PERCEPČNÍM ZVUKOVÉM KODÉRU
SIMULATION OF AUDIO SIGNAL DISTORTION IN PERCEPTUAL AUDIO ENCODER

DIPLOMOVÁ PRÁCE
MASTER'S THESIS

AUTOR PRÁCE — **Bc. Tomáš Peloušek**
AUTHOR

VEDOUCÍ PRÁCE — **doc. Ing. Jiří Schimmel, Ph.D.**
SUPERVISOR

BRNO 2021

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

Diplomová práce

magisterský navazující studijní program **Audio inženýrství**
specializace Zvuková produkce a nahrávání
Ústav telekomunikací

Student: Bc. Tomáš Peloušek ***ID:*** 186158
Ročník: 2 ***Akademický rok:*** 2020/21

NÁZEV TÉMATU:

Simulace zkreslení zvukového signálu v percepčním zvukovém kodéru

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s principy a metodami ztrátového kódování zvukového signálu využívajícími k redukci počtu bitů psychoakustického modelu, vytvořte jejich rešerši a souhrn dostupných parametrů kódování používaných streamovacími službami. V prostředí Matlab implementujte základní kodér, tj. pouze po smyčku kvantování vzorků, a odpovídající dekodér. Následně vytvořte funkce pro demonstrování vlivu parametrů kodéru na kódovaný signál. Zaměřte se zejména na alokaci bitů, přepínání délky oken kmitočtové transformace a kódování stereofonních zvukových signálů. Pomocí volně dostupných implementací metod objektivního hodnocení kvality zvuku demonstrujte vliv těchto parametrů na kvalitu zvukového signálu. K realizovaným funkcím vytvořte dokumentaci a jednoduchá grafická rozhraní uživatele. Při implementaci můžete vycházet z volně dostupných implementací metod percepčního kódování zvuku.

DOPORUČENÁ LITERATURA:

[1] BOSI, M., GOLDBERG, R. E. Introduction to Digital Audio Coding and Standards. Kluwer Academic Publishers, 2003. ISBN 1-4020-7357-7.

[2] ITU-R BS.1116-1: Methods for the Subjective Assessmen–t of Small Impairments in Audio Systems Including Multichannel Sound System. 2015. International Telecommunication Union.

Termín zadání: 1.2.2021 ***Termín odevzdání:*** 24.5.2021

Vedoucí práce: doc. Ing. Jiří Schimmel, Ph.D.
Konzultant: Jaroslav Musil (Audified, s.r.o.)

doc. Ing. Jiří Schimmel, Ph.D.
předseda rady studijního programu

ABSTRAKT

Tato práce se zabývá problematikou tvorby programu, který simuluje zkreslení vznikající při ztrátovém kódování zvukového signálu, a to v programovacím prostředí MATLAB. V rámci práce byl vytvořen kodér s dynamickou alokací bitů a přepínám délek oken pro váhování, který v závislosti na uživatelském požadavku na velikost datového toku mění výslednou subjektivní kvalitu signálu. Teoretická část představuje rešerši základních principů ztrátového kódování a detailněji popisuje fungování kodéru standardu MPEG 1 vrstva 3. V praktické části je pak popsán princip fungování realizovaného programu a jeho součástí. Dále je provedeno srovnání kvality výstupu programu pro různé úrovně zkreslení s odpovídajícím nastavením u běžně dostupného kodéru MP3 a to pomocí metody PEMO-Q.

KLÍČOVÁ SLOVA

alokace bitů, MP3, MPEG-1 vrstva 3, PEMO-Q, percepční kódování, přepínání délek oken, psychoakustický model

ABSTRACT

This thesis deals with the issue of the creation of a programme that would simulate the distortion that appears during the process of lossy audio coding. As the environment for the creation, the MATLAB programming language has been chosen. An encoder, which changes the subjective signal quality according to customer preferences for the bitrate, has been created as a practical part of this thesis. Its function is based on a dynamic bit allocation technique and includes an optional window switching algorithm. The theoretical background for the creation of the programme consists of an explanation of the main principles of lossy coding with emphasis on MPEG1 layer 3 operating principles. The practical chapter describes how the created programme and its parts work, and it includes results of the run quality testing. The testing was conducted using the objective assessment method PEMO-Q, and consisted of comparing the objective quality of the programme's outputs to the quality of samples on which a regular MP3 encoder with identical settings was used.

KEYWORDS

bit allocation, MP3, MPEG-1 layer 3, PEMO-Q, perceptual encoding, psychoacoustic model, window switching

PELOUŠEK, Tomáš. *Simulace zkreslení zvukového signálu v psychoakustickém ztrátovém kodéru*. Brno, 2020, 53 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: doc. Ing. Jiří Schimmel, Ph.D.

Vysázeno pomocí balíčku thesis verze 4.02; http://latex.feec.vutbr.cz

PROHLÁŠENÍ

Prohlašuji, že svou diplomovou práci na téma „Simulace zkreslení zvukového signálu v psychoakustickém ztrátovém kodéru“ jsem vypracoval samostatně pod vedením vedoucího diplomové 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 diplomové práce dále prohlašuji, že v souvislosti s vytvořením této diplomové práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno ....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu diplomové práce panu doc. Ing. Jiřímu Schimmelovi Ph.D. za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.
Děkuji též všem, kteří mě podporovali, pomáhali mi dešifrovat syntakticky nefunkční věty v indické angličtině, asistovali s překlady a jazykovou stránkou textu, ale taky mi nosili oříšky a vždy mi dodali chuť psát dál.

Obsah

Úvod

Tato práce se zabývá problematikou tvorby programu, který simuluje zkreslení vznikající při ztrátovém kódování. Cílem práce bylo vytvořit kodér, který v závislosti na uživatelském požadavku na velikost datového toku mění výslednou subjektivní kvalitu signálu.

Program byl vyvíjen dle zadání v programovacím prostředí MATLAB. Toto prostředí je pro tento úkol vhodné, neboť obsahuje již spoustu hotových funkcí použitelných pro zpracování zvukových signálů a umožňuje jednoduchou práci s proměnnými v podobě vektorů a matic.

Tuto problematiku diskutují mimo jiné Jayaraman Thiagarajan a Andreas Spanias ve své publikaci „Analysis of the MPEG-1 Layer III (MP3) Algorithm using MATLAB“, Marina Bosi a Richard E. Goldberg v publikaci „Introduction to digital audio coding and standards“. Uvedené publikace tvoří velkou část teoretického východiska této práce.

Práce začíná teoretickou částí, v níž jsou popsány principy, na jejichž základě percepční kódování pracuje. Nejprve je zde představeno obecné blokové schéma hybridního kodéru následované detailním popisem jednotlivých bloků s důrazem na psychoakustický model a popis principů použitých při kvantování signálu. V rámci popisu každého bloku je zařazena kapitola detailněji popisují řešení ve standardu MPEG-1 vrstva 3, na jehož principech kodér realizovaný v rámci praktické části funguje.

Komentář k praktické části práce sestává z popisu realizovaného stereofonního kodéru a srovnání kvality jeho výstupu s výstupem běžně dostupného MP3 kodéru pro odpovídající nastavení. Pro srovnání kvality je použito realizace objektivního hodnocení zvukových signálů PEMO-Q vytvořené v rámci [12]. Kodér obsahuje i jednoduché grafické rozhraní uživatele umožňující demonstraci vlivu vstupních parametrů na kvalitu výstupního signálu a demonstraci principu dynamické alokace bitů.

V závěru je diskutována úspěšnost naplnění zadání práce a možnosti využití programu v rámci práce realizovaného.

1 Percepční kódování zvuku

Percepční kódování zvuku je založeno na principu redukce součástí signálu, které nejsme kvůli limitacím našeho sluchového systému schopni vnímat. Při tomto druhu kódování se kombinuje jak kódování ztrátové, tak bezztrátové. Z technik bezztrátové komprese se většinou využívají Run length coding a Huffmanovo kódování, které přiřazuje různým hodnotám bitové řetězce o různé délce na základě četnosti jejich výskytu. Nejčastěji se vyskytující hodnoty jsou tak kódovány pomocí bitových řetězců o nejkratší délce, a tím je komprimován datový tok. [1]

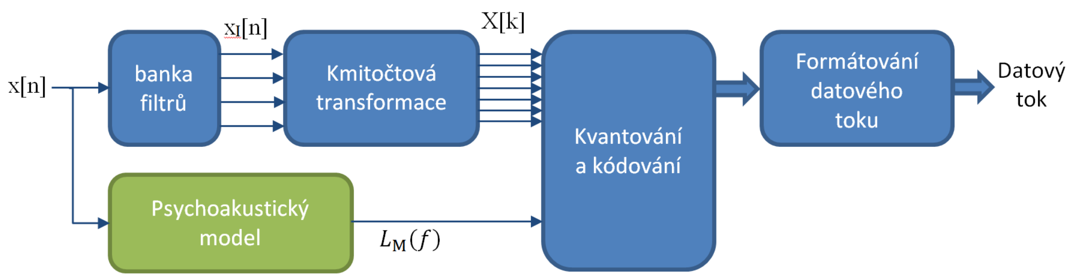
Ztrátové kódování redukuje datový tok frekvenčně závislým kvantováním signálu ve spektrální oblasti na základě informací o relevantnosti získaných z psychoakustického modelu, a požadavku uživatele na velikost datového toku.[2]

1.1 Hybridní kodér

Při percepčním kódování se v současnosti využívá nejčastěji tzv. hybridních kodérů.

1.1.1 Blokové schéma hybridního kodéru

Blokové schéma hybridního kodéru představuje obr. 1.1. Tento typ kodéru využívá



Obr. 1.1: Blokové schéma hybridního kodéru [2]

kombinaci filtrace signálu bankou filtrů s uniformním kmitočtovým rozložením, podvzorkování a následné kmitočtové transformace. Po těchto operacích je nutné provést redukci aliasingu způsobeného přeslechy mezi sousedními kanály. Ta se provádí křížovým váhováním v sousedních pásmech. Souběžně s tím prochází signál psychoakustickým modelem. Psychoakustický model slouží k určení globálního maskovacího prahu L_M v každém časovém rámci a určení odstupu signálu od masky (Signal to Mask Ratio – SMR). Tato informace je předána do bloku kvantování a kódování, kde je signál na základě těchto informací nakvantován a zakódován pomocí Huffmanova kódování. Informace o použitých kódovacích tabulkách pro Huffmanovo kódování a měřítkových koeficientech se společně s nakvantovanými spektrálními koeficienty formátuje do datového toku.[2]

1.1.2 Analyzující banka filtrů

Při zpracování zvukových signálů je výhodné analyzovat zvlášť různá jejich frekvenční pásma. V rámci hybridního kodéru tuto úlohu společně vykonávají dva bloky, a to blok analytické banky filtrů a blok frekvenční transformace.[2] Signál prochází bankou filtrů a je rozložen do pásem se stejnou šířkou. Toto rozdělení neodpovídá vlastnostem lidského sluchu který oproti bance filtrů s rovnoměrným rozložením analyzuje zvukový signál v kritických pásmech, kde šířka pásma je frekvenčně závislá. Toto rozložení můžeme aproximovat pomocí filtrů se zlomko-oktávovým rozložením. Implementace banky zlomko-oktávových filtrů jsou ovšem výpočetně náročné, a tudíž nevhodné pro zpracování signálů v reálném čase.[3] V rámci rodiny standardů MPEG se tak využívá banka s pseudo-kvadraturními zrcadlovými filtry (PQMF), jejíž výhodou je téměř dokonalá rekonstrukce původního signálu, potlačení aliasingu v sousedních pásmech a snadná implementace pomocí algoritmu nízkou výpočetní náročností. Pro standard MPEG-1 vrstva 3 sestává banka z 32 filtrů s lineární fázovou odezvou. [3] Tyto filtry vznikají kosinovou modulací prototypového filtru typu dolní propust jehož impulsní odezva je definována přímo tabulkou v rámci přílohy normy. Zde je definováno pouze prvních 256 vzorků, zbylá část impulsní odezvy vzniká zrcadlením, jak je patrné z obrázku 1.2.[4] Výsledná pásmová propust tak vzniká frekvenčním posunem prototypového filtru o $f_{\mathrm{vz}}/64$. Tato hodnota zároveň představuje šířku propustného pásma filtrů. Průběhy modulových kmitočtových charakteristik prototypové dolní propusti H_0 a pásmové propusti H_1 představuje obrázek 1.3.[1]. H_1 je v pořadí první pásmovou propustí analytické banky vytvořenou pomocí modulace H_0.

Výstup i-tého pásma analytické PQMF banky tak můžeme definovat jako

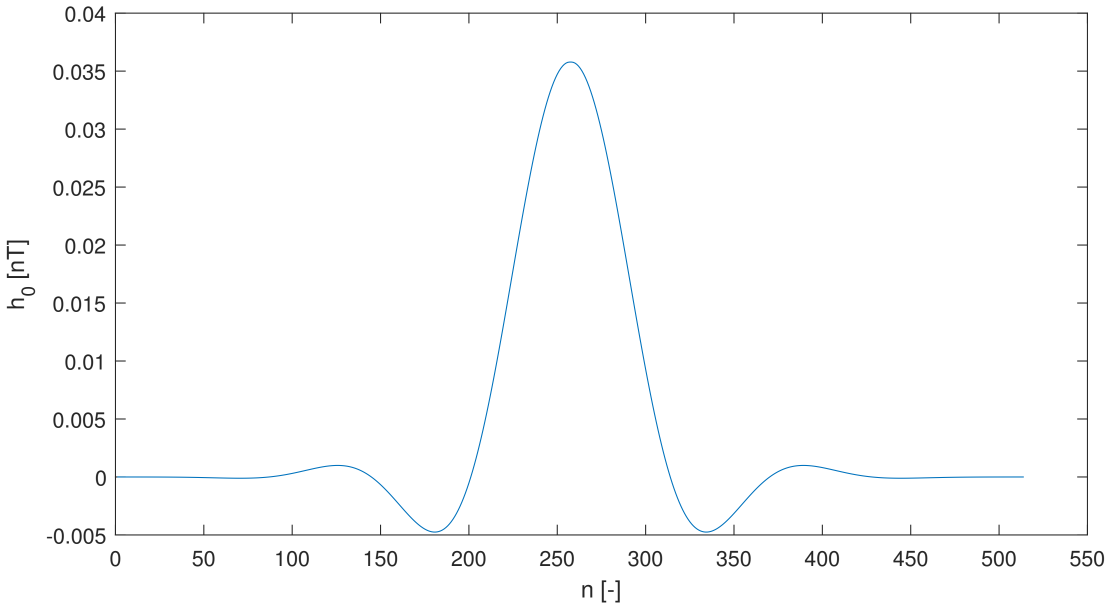
$$s_i = \sum_{k=0}^{63} \sum_{j=0}^{7} \mathrm{M}_{i,k} \left(c\left(k + 64j\right) x\left(k + 64j\right) \right), \tag{1.1}$$

$$\mathrm{M}_{i,k} = \cos\left(\frac{\left(2i + 1\right)\left(k - 16\right)\pi}{64} \right), \tag{1.2}$$

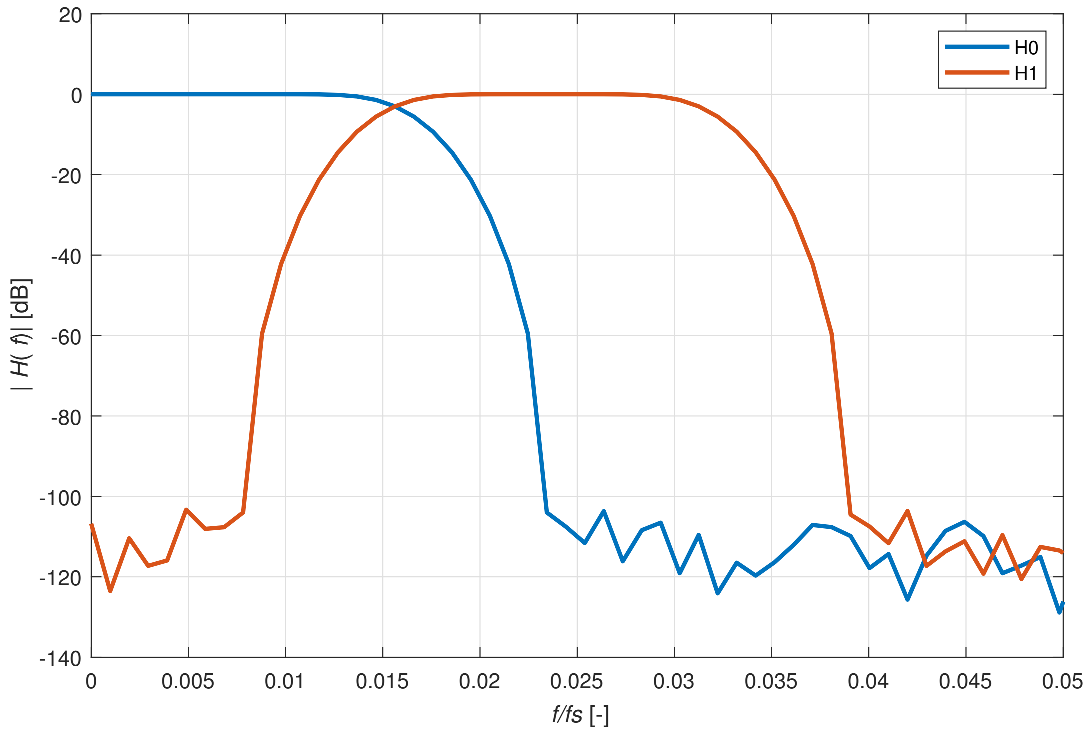
kde $i = 0$ až 31 a $k = 0$ až 63.[3]

1.1.3 Psychoakustický model

Nároky na psychoakustický model se odvíjí od konkrétní aplikace, obecně je však kladen důraz na rychlost a přesnost výpočtu SMR a nízkou výpočetní náročnost.[2]



Obr. 1.2: Impulsní charakteristika prototypové dolní propusti



Obr. 1.3: Modulové kmitočtové charakteristiky H_0 a H_1

Výpočet SMR probíhá v několika krocích.

Kmitočtová transformace a váhování oknem

Časový rámec prochází kmitočtovou transformací a je váhován Hannovým oknem. Fourierova transformace diskrétního signálu je dána rovnicí

$$X\left(e^{-j\omega}\right) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n}, \tag{1.3}$$

kde $\omega = 2\pi f$, f reprezentuje kmitočet signálu v závislosti na vzorkovacím kmitočtu a $x[n]$ představuje jednotlivé vzorky audio signálu. Nejčastěji se dnes k realizaci Fourierovy transformace diskrétního signálu používá algoritmu zvaného FFT (Fast Fourier Transform).[5]

Určení absolutních hladin akustického tlaku

Pro spektrální složky získané kmitočtovou transformací je nutné stanovit odpovídající hodnoty hladiny akustického tlaku, protože lidské vnímání akustického tlaku a s ním související subjektivní hlasitost mají logaritmický charakter. Hladina akustického tlaku je poměrová veličina, která používá jako vztažnou hodnotu akustického tlaku 20 µPa, což odpovídá prahové hodnotě schopné vyvolat sluchový vjem při kmitočtu 1 kHz. Problémem při kódování je ovšem fakt, že nevíme, při jaké hladině hlasitosti bude hudba koncovým uživatelem konzumována. Proto se vychází z předpokládaného dynamického rozsahu, kde 0 dB (SPL) odpovídá hladině kvantovacího šumu viz rovnice (1.4), kde L představuje hladinu číslicového signálu, L_{P} hladinu akustického tlaku v dB (SPL) a B počet bitů signálu.[2]

$$L = 20 \log \frac{1}{2^B} + 20 \log L_{\mathrm{P}} \tag{1.4}$$

Prahová hodnota sluchového vjemu je však značně frekvenčně závislá. Tato závislost se obvykle označuje jako práh slyšitelnosti a pro její aproximaci se využívá vztahu[2]

$$T_{\mathrm{q}}(f) = 3,64\left(\frac{f}{1000}\right)^{-0,8} - 6,5\mathrm{e}^{-0,6\left(\frac{f}{1000}-3,3\right)^2} + 10^{-3}\left(\frac{f}{1000}\right)^4. \tag{1.5}$$

Této vlastnosti lidského sluchu využíváme společně s efektem maskování při tzv. tvarování maskovacího šumu.

Maskování

Maskování je psychoakustický jev, ke kterému dochází při převodu mechanického kmitání na elektrický impuls v prostředí středního ucha a na sluchový vjem v oblasti vnitřního ucha. Tento jev můžeme pozorovat jak v oblasti kmitočtové, tak v oblasti časové, vychází ovšem ze stejného principu. V případě, že do sluchového orgánu dorazí současně dva nebo více akustických signálů, může se stát, že vjem vyvolaný jedním z nich zeslabí nebo zcela potlačí sluchový vjem způsobený ostatními signály.[2]

V časové oblasti se tento jev projevuje jako tzv. časové maskování, kdy krátce před a po přítomnosti maskovacího signálu nejsou vláskové buňky na bazilární membráně schopny reagovat na další mechanické podráždění. Výsledkem je pak tzv. premasking a post-masking.[2]

V kmitočtové oblasti je jev maskování způsoben naopak tím, že je mechanickým podnětem podrážděna vždy ta část bazilární membrány, která odpovídá určité kmitočtové oblasti, jež je širší než kmitočtové pásmo maskovacímu signálu. V případě, že současně s maskujícím signálem je přítomen další signál o podobném kmitočtu, který je ale slabší, nejsou na jeho přítomnost vláskové buňky schopné reagovat.[2]

Pro popis frekvenčního maskování je nutné nejprve stanovit vztah mezi frekvencí f a číslem kritického pásma z danou vztahem (1.6).

$$z(f) = 13 \arctan\left(\frac{0,76f}{1000}\right) + 3,5 \arctan\left(\left(\frac{f}{7500}\right)^2\right) \text{ [bark]} \tag{1.6}$$

Tento vztah často označujeme jako *critical band rate.*[2][3]

Maskovací křivku pak můžeme aproximovat například rovnicí (1.7), která vychází ze Schröderovy aproximace a používá se v psychoakustickém modelu 2 standardu MPEG-1 vrstva 3.[4][2]

$$\begin{split}F(dz) = 15,8111389 + 7,5(1,05dz + 0,474) - 17,5\sqrt{1 + (1,05dz + 0,474)^2} + \\ +8min\left\{0; (1,05\mathrm{dz} - 0,5)^2 - 2(1,05dz - 0,5)\right\},\end{split} \tag{1.7}$$

kde

$$\mathrm{dz} = z\left(f_{\text{maskovaný}}\right) - z\left(f_{\text{maskovací}}\right) \tag{1.8}$$

je vzdálenost kmitočtů maskovaného a maskovacího kmitočtu na barkové stupnici.[2]

Odlišně vnímáme maskování signálem tónového charakteru oproti maskování signálem s vlastnostmi blízkými šumu. Proto existuje v psychoakustickém modelu výpočetní algoritmus, jehož úkolem je tyto charakterově odlišné typy signálu identifikovat a rozlišit.[2]

Identifikace tónových a šumových složek signálu

Existuje několik používaných algoritmů pro určení tónového či šumového charakteru dané maskovací složky signálu. Detailněji o této problematice hovoří [6].

První ze strategií, které tato práce uvádí, je hledání lokálních maxim spektrální hustoty výkonu (PSD). Při použití této techniky je spektrální složka považována za tónovou v případě, kdy má nejméně o 7 dB vyšší PSD než okolní složky.[2]

Další možností je na základě rovnice (1.9) vypočítat spektrální plochost (Spectral Flattness Measure – SFM) a s její pomocí určit index tonality δ z rovnice (1.10). Pro výpočet spektrální plochosti potřebujeme určit aritmetický průměr výkonového spektra A_{m} a geometrický průměr výkonového spektra G_{m}. Při použití této metody považujeme spektrální složku za tónovou pro hodnoty indexu tonality větší než 0,5.[2]

$$SFM = 10 \log \left(\frac{G_{\mathrm{m}}}{A_{\mathrm{m}}} \right) \tag{1.9}$$

$$\delta = min \left(\frac{SFM}{-60}, 1 \right) \tag{1.10}$$

Psychoakustický model standardu MPEG-2 pak srovnává současnou hodnotu spektrální složky s hodnotou predikovanou na základě dat ze dvou předchozích rámců. Nepředvídatelnost c se pak určí jako

$$c(f) = \frac{\sqrt{\Big(R_{\mathrm{j}}(f) \cos(\phi_{\mathrm{j}}(f)) - \tilde{R}_{\mathrm{j}}(f) \cos(\phi_{\mathrm{j}}(f))\Big)^2 + \Big(R_{\mathrm{j}}(f) \sin(\phi_{\mathrm{j}}(f)) - \tilde{R}_{\mathrm{j}}(f) \sin(\phi_{\mathrm{j}}(f))\Big)^2}}{R_{\mathrm{j}}(f) + |\tilde{R}_{\mathrm{j}}(f)|}, \tag{1.11}$$

kde

$$\tilde{R}_{\mathrm{j}}(f) = 2R_{\mathrm{j}-1}(f) - R_{\mathrm{j}-2}(f),$$
$$\phi_{\mathrm{j}}(f) = 2\phi_{\mathrm{j}-1}(f) - \phi_{\mathrm{j}-2}(f).$$

R zde představuje modul komplexního spektra a ϕ jeho fázi. Index j odkazuje na současný rámec a indexy $j-1$ a $j-2$ na rámce předcházející.[3]

Stanovení nepředvídatelnosti pro všechny spektrální složky až do 20 kHz by bylo značně výpočetně náročné. Proto model MPEG-2 počítá nepředvídatelnost pouze pro prvních 206 složek a zbylým složkám přiřazuje konstantní hodnotu 0,4. [3]

Rozdělení do výpočetních pásem a určení jejich energie

V rámci dalších výpočtů jsou spektrální složky seskupeny do výpočetních pásem, jejichž šířka přibližně odpovídá šířce kritického pásma. Na nízkých kmitočtech tak může výpočetní pásmo sestávat pouze z jedné spektrální složky, zatímco na vysokých kmitočtech výpočetní pásmo sestává z mnoha složek.[7] Celkový počet výpočetních pásem je závislý na použité vzorkovací frekvenci a je definován normou.[4]

Energii každého výpočetního pásma stanovíme jako

$$eb\,(b) = \sum_{i=kmin_b}^{kmax_b} R^2(i), \tag{1.12}$$

kde R je modul, a $kmax_b$ horní hranice a $kmin_b$ spodní hranice výpočetního pásma o indexu b.[3] Váhovaná nepravděpodobnost pro každé výpočetní pásmo b je pak dána jako

$$cb\,(b) = \sum_{i=kmin_b}^{kmax_b} R^2(i)c(i). \tag{1.13}$$

Dále je provedena konvoluce energie a váhované nepravděpodobnosti s rovnicí maskovací křivky (1.7)

$$ecb\,(b) = \sum_{i=1}^{b_{\max}} eb\,(i_{\mathrm{b}})\,F\,(bm_{\mathrm{i}}, bm)\,, \tag{1.14}$$

$$ctb\,(b) = \sum_{i=1}^{b_{\max}} cb\,(i_{\mathrm{b}})\,F\,(bm_{\mathrm{i}}, bm)\,, \tag{1.15}$$

kde bm je průměrná hodnota $z(f)$ ve výpočetním pásmub a $b_{\max}$ index nejvyššího výpočetního pásma v závislosti na dané vzorkovací frekvenci.[3]

Určení indexu tonality

Následně je tato hodnota normována dle vztahu (1.19) a index tonality pro dané výpočetní pásmo určíme jako

$$ti\,(b) = -0,299 - 043cbb\,(b)\,, \tag{1.16}$$

kde

$$cbb\,(b) = \log\left(\frac{ctb\,(b)}{ecb\,(b)}\right). \tag{1.17}$$

Index tonality nabývá hodnot v rozsahu 0 až 1, kde pro hodnoty blížící se jedné mluvíme o tonálním charakteru dominantní maskovací složky ve výpočetním pásmu a pro hodnoty blížící se k nule o charakteru šumovém.[3]

1.1.4 Blok kmitočtové transformace

Blok kmitočtové transformace slouží k dalšímu přesnějšímu členění signálu ve spektrální oblasti. Pro efektivní zpracování zvukových signálů je nutné, aby systém uměl signály tónového charakteru zpracovat s co největší přesností ve spektrální oblasti, a naopak signály šumové s co nejlepším rozlišením v rovině časové. Na základě Heisenbergova principu neurčitosti víme, že u dvou konjugovaných veličin platí, že čím přesněji určíme jednu z konjugovaných vlastností, tím méně přesně můžeme určit vlastnost druhou, což je i případ rozlišení v časové a spektrální rovině u kmitočtových transformací.[8] Ve standardu MPEG-1 vrstva 3 se tak využívá různých délek rámce zpracování pro signály s tónovým a šumovým charakterem.

Váhování oknem

Signál je před provedením kmitočtové analýzy váhován oknem. Využívají se okna dvou různých délek a to dlouhé okno o délce 36 vzorků pro tonální segmenty a krátké okno o délce 12 vzorků pro netonální segmenty. Pro přechod mezi nimi se používají okna označovaná jako start a stop o délce 36 vzorků.[4]

Průběhy oken můžeme definovat těmito rovnicemi:
dlouhé okno (long)

$$w\left(n\right) = \sin\left(\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right) \quad n = 0, ..., 35 \tag{1.18}$$

krátké okno (short)

$$w\left(n\right) = \sin\left(\frac{\pi}{12}\left(n+\frac{1}{2}\right)\right) \quad n = 0, ..., 11 \tag{1.19}$$

start okno

$$w\left(n\right) = \begin{cases} \sin\left(\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right) & n = 0, ..., 17 \\ 1 & n = 18, ..., 23 \\ \sin\left(\frac{\pi}{12}\left(n-18+\frac{1}{2}\right)\right) & n = 24, ..., 29 \\ 0 & n = 30, ..., 35 \end{cases} \tag{1.20}$$

stop okno

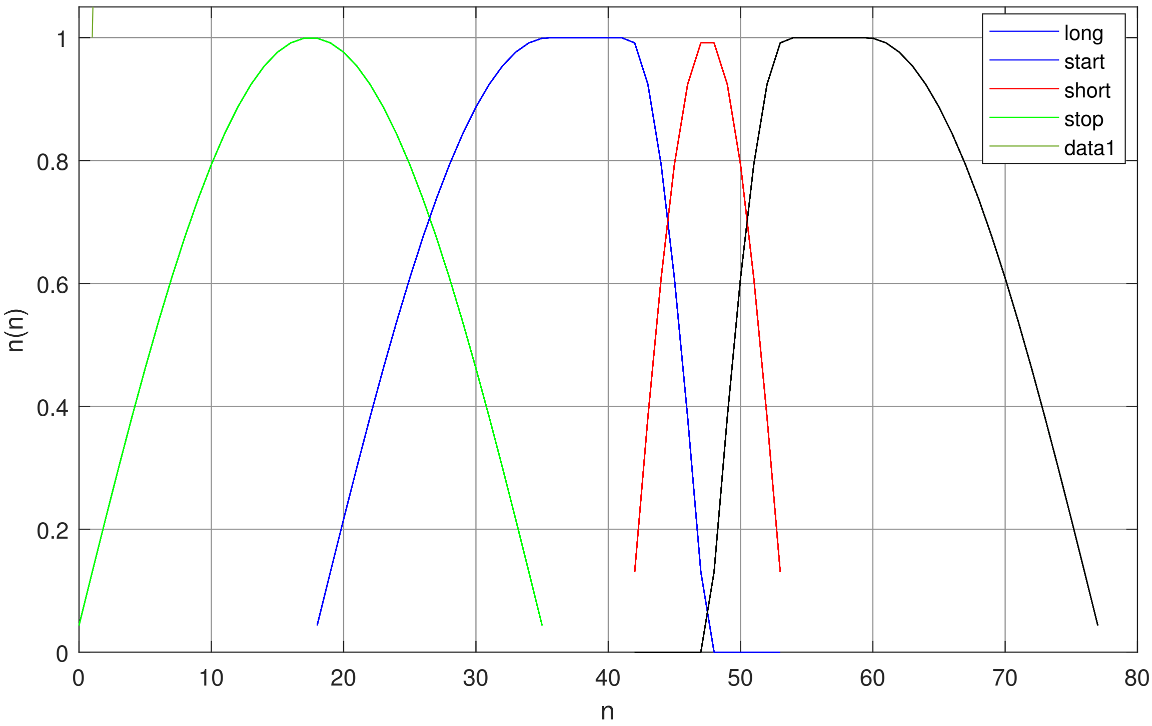
$$w\left(n\right) = \begin{cases} 0 & n = 0, ..., 5 \\ \sin\left(\frac{\pi}{12}\left(n-6+\frac{1}{2}\right)\right) & n = 6, ..., 11 \\ 1 & n = 12, ..., 17 \\ \sin\left(\frac{\pi}{36}\left(n+\frac{1}{2}\right)\right) & n = 18, ..., 35 \end{cases} \tag{1.21}$$

Časový průběh těchto oken představuje obrázek 1.4.

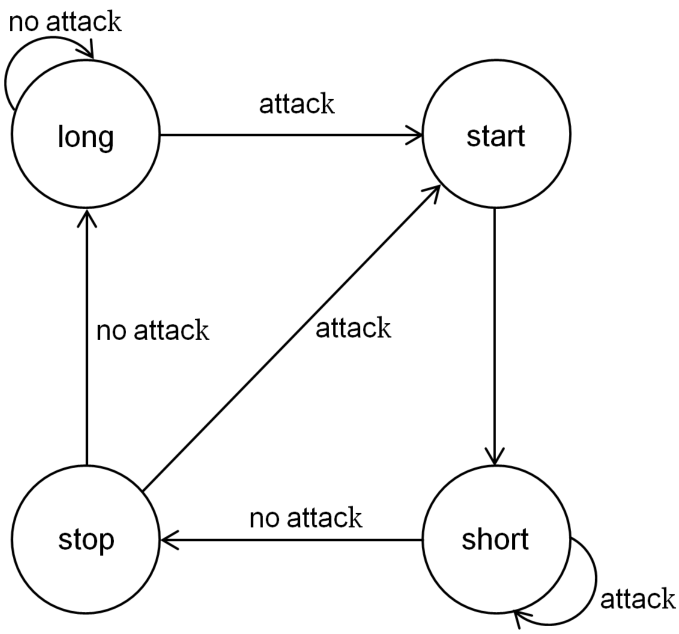
V rámci MPEG-1 vrstva 3 se pro přepínání délky okna kmitočtové zpracování používá prahová hodnota percepční entropie (dále PE). Přepínání oken je řízeno rozhodovacím procesem znázorněným na obrázku 1.5, kde označení *attack* indikuje překročení prahové hodnoty 1800bit/sample při výpočtu percepční entropie daného rámce v psychoakustickém modelu. Označení *no attack* naopak pokles PE pod prahovou úroveň.[3][9]

Modifikovaná diskrétní kosinová transformace

Pro přechod z časové do frekvenční domény se nejčastěji používá DFT realizované pomocí výpočetně efektivní implementace FFT. Při použití DFT je ovšem nutné



Obr. 1.4: Okna používaná v MPEG-1 vrstva 3[2]



Obr. 1.5: Přepínací diagram oken[3][9]

pro omezení vlivu váhování segmentů oknem na pokles energie signálu přičíst vždy k následujícímu segmentu část předešlého. Například při použití Hannova okna se nejčastěji využívá 50% překryvu mezi segmenty. V tomto případě bychom tak fakticky ve frekvenční doméně kódovali dvojnásobek dat, čímž by se fakticky znemožnila jejich úspora.[1] Zde je vhodné využít nativní vlastnosti modifikované diskrétní kosinové transformace (MDCT), jež umožňuje zpracovávat signál s 50% překryvem mezi segmenty bez navýšení zpracovávaného datového toku. Do MDCT tak vstupuje vektor N vstupních vzorků $x_i\left[n\right]$, které jsou transformovány na $N/2$ vzorků $X_i\left[k\right]$ ve frekvenční oblasti podle vztahu

$$X_i\left[k\right] = \sum_{n=0}^{N-1} x_i\left[n\right] \cos\left(\frac{2\pi}{N}\left(n+n_0\right)\left(k+\frac{1}{2}\right)\right) \quad \text{pro } k = 0,...,N/2-1, \tag{1.22}$$

kde

$$n_0 = \left(\frac{n}{2}+1\right)/2 \tag{1.23}$$

pomocí vhodné hodnoty počáteční fáze zajišťuje eliminaci aliasingu.[1]

Výpočetní náročnost realizace MDCT přímo podle definice ovšem stejně jako u DFT roste s N^2. Zde je proto vhodné realizovat MDCT pomocí algoritmu FFT jehož výpočetní náročnost s prodlužující se délkou roste pouze s $N \log_2 N$. Předpis (1.22) tak můžeme upravit na

$$X\left[k\right] = \mathrm{Re}\left\{e^{-\mathrm{j}\frac{2\pi}{N}n_0\left(k+\frac{1}{2}\right)} \sum_{n=0}^{N-1}\left[x\left[n\right]w\left[n\right]e^{-\mathrm{j}\frac{2\pi n}{2N}}\right]e^{-\mathrm{j}\frac{2\pi kn}{N}}\right\}, \tag{1.24}$$

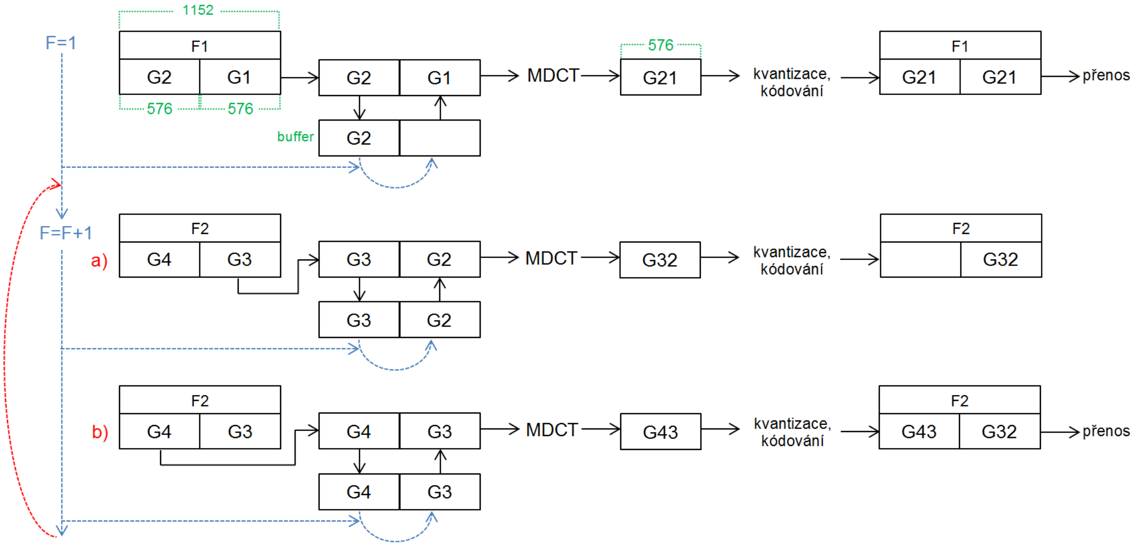
kde $w\left[n\right]$ představuje násobení oknem.

MDCT v MP3

V rámci standardu MPEG-1 vrstva 3 se pracuje s rámci (angl. frame) signálu o délce 1152 vzorků, které se dělí na dvě granule od délce 576 vzorků.[4] Systém řazení granulí a proces jejich vstupu do MDCT představuje schéma 1.6, kde F1 a F2 představují v pořadí první a druhý rámec vstupního signálu, G1 v pořadí první granuli G2 v pořadí druhou atd. Buffer představuje zásobník o délce 1152 vzorků a výstup MDCT označený G21 představuje 576 spektrálních koeficientů reprezentujících granule G1 a G2.

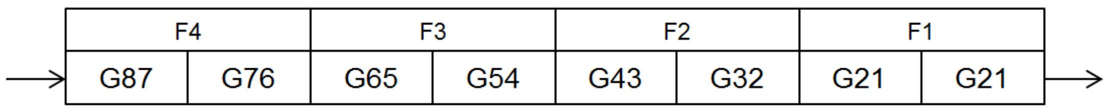
Prvního rámec zpracování představuje výjimku, protože u něj při jeho kódování není možné provést překryv mezi granulemi. Správně by na výstupu pro přenos měl být rámec sestávající z dvou granulí spektrálních koeficientů označených jako G21 a G23, což ovšem není možné, neboť G3 není v ten moment ještě k dispozici pro zpracování. Proto sestává výstup zpracování prvního rámce z dvou za sebe seřazených granulí G21.[3] V rámci zpracování prvního rámce je uložena do bufferu granule G2.

Zpracování všech ostatních rámců signálu má dvě fáze označené ve schématu a) a b), které se cyklicky opakují. Ve fázi a) po načtení F2 (druhého rámce) do



Obr. 1.6: Schéma řazení granulí do MDCT a realizace překryvu

paměti vstupují do MDCT granule G2 a G3, výstup MDCT G23 je po kvantování a kódování uložen do výstupní paměti a do bufferu se uloží G3. Následuje fáze b), kde do MDCT vstupují granule G3 a G4, a výstup MDCT G34 se dále zpracuje a uloží do výstupní paměti. Při pohledu na za sebe seřazené rámce na výstupu (1.7) je již zřetelně patrný překryv mezi granulemi.



Obr. 1.7: Schéma řazení rámců na výstupu kodéru

1.1.5 Blok kvantizace

Redukce datového toku se realizuje v bloku kodéru označeném ve schématu 1.1 jako kvantizace. Zde je datový tok redukován snížením počtu bitů použitých pro zakódování hodnot spektrálních koeficientů, kterými je signál reprezentován. Pro zachování vnímané subjektivní kvality signálu je důležité, aby kvantizační šum, který při kvantování vzniká, nebyl slyšitelný. Toho se dosahuje nastavením úrovně kvantizačního šumu tak, aby ležela pod úrovní globálního maskovacího prahu. Kvantizační šum můžeme definovat jako rozdíl mezi signálem původním a signálem nakvantovaným.[1]

Pro zjednodušený odhad odstupu signálu od šumu (SNR) můžeme použít (1.25) popisující odstup harmonického signálu od šumu, kde n je počet bitů použitého převodníku.[2]

$$SNR = 6{,}02n + 1{,}76 \ \ [\mathrm{dB}] \tag{1.25}$$

Z této rovnice je odvozena nejjednodušší a nejméně výpočetně náročná metoda alokace. Při statické alokaci bitů je možné počet bitů určený pro kódování pásma signálu stanovit jako

$$n\,(i) = SNR\,(i)\,/6{,}02, \tag{1.26}$$

kde $SNR\,(i)$ představuje nejnižší hodnotu odstupu signálu od šumu v daném pásmu. Pro další snížení výpočetní náročnosti bývá tato hodnota často přímo definována tabulkou. Tato metoda se používá například v kodecích LC-ATC a AC2.[2]

Při postupu stanovení počtu bitů pro zakódování daného pásma můžeme vycházet z hodnot percepční entropie (PE), která udává průměrný minimální počet bitů potřebných pro zakódování daného signálu bez toho, aniž by byl signál slyšitelně degradován ve srovnání s původním signálem. Percepční entropii určíme jako

$$PE = -\sum_{b=1}^{z_{\max}} \left\{ cbwidth\,(b) \cdot \log\left(\frac{L_m\,(b)}{eb\,(b)} + 1\right)\right\} \quad [\mathrm{bit/vzorek}]\,, \tag{1.27}$$

kde $cbwidth\,(b)$ odpovídá počtu spektrálních složek ve výpočetním pásmu, L_m a eb jsou hodnoty maskovacího prahu a energie ve výpočetním pásmu.[3]

V rámci psychoakustického modelu MPEG-2 existuje rozhodovací práh 1800 bitů/vzorek, který když hodnota PE pro daný rámec překročí, je tento rámec před kvantováním rozdělen oknem na tři kratší, aby se zaručila přesnější reprezentace signálu v časové doméně.[7][3]

Dynamická alokace bitů v MP3

Psychoakustický model předává bloku kvantování a kódování informace o globálním maskovacím prahu a použité délce okna pro daný rámec signálu. Spektrální koeficienty jsou rozděleny do výpočetních pásem zvaných *scalefactor bands*, kde je každému pásmu přiřazen jeden měřítkový koeficient. Toto rozdělení spektrálních koeficientů, které aproximuje rozdělení do kritických pásem, je definováno v příloze normy [4]. Liší se pro dlouhé a krátké bloky a závisí také na vzorkovací frekvenci vstupního signálu.

Ve standardu MPEG-1 vrstva 3 se využívá nelineární kvantizace, která již ze své podstaty přispívá k tvarování kvantizačního šumu tím, že jsou spektrální složky s větší hodnotou kvantovány s větším kvantovacím krokem. Tím se zvyšuje SNR u menších spektrálních složek.[1]

Kvantizace spektrálních složek je dána rovnicí

$$ix\,(i) = \mathrm{NINT}\left[\left(\frac{|xr\,(i)|}{2^{0{,}25(\mathrm{global\_gain}-210-\mathrm{scale\_factor}(b))}}\right)^{0{,}75} - 0{,}0946\right], \tag{1.28}$$

kde scale_factor (b) a global_gain jsou parametry řídící velikost kvantizačního kroku, $xr\,(i)$ je hodnota spektrální složky na pozici indexu i a NINT je funkce realizující zaokrouhlení na celé číslo.[10]

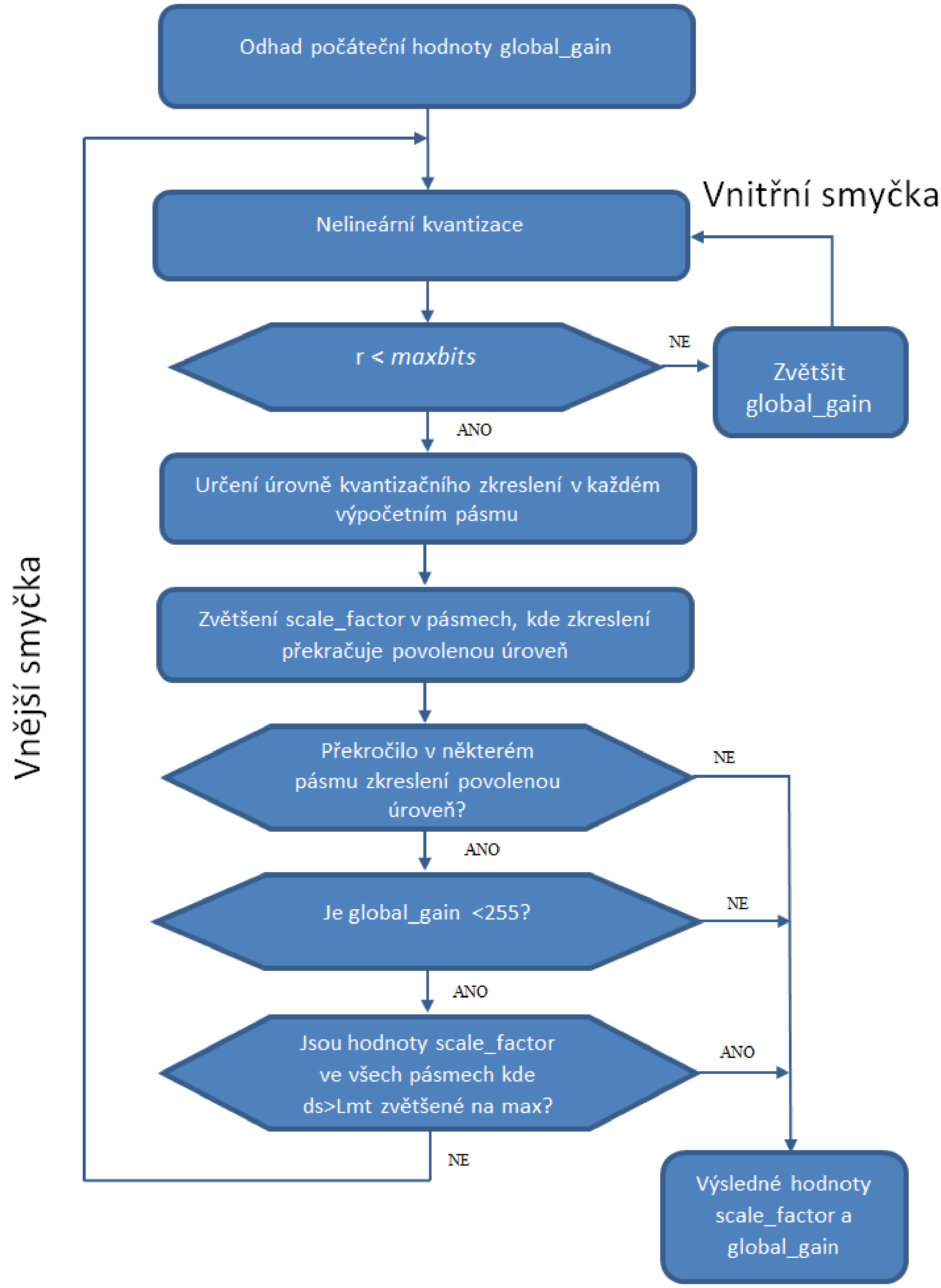
Hodnota parametru global_gain je stejná pro všechny spektrální složky, zatímco parametr scale_factor reprezentuje hodnotu měřítkového koeficientu daného výpočetního pásma. Měřítkový koeficient plní funkci váhy, s jejíž pomocí můžeme v rámci výpočetních pásem regulovat míru kvantizačního šumu.[13]

Pomocí parametru global_gain pak můžeme kontrolovat velikost datového toku. Hledání vhodné kombinace těchto řídících parametrů pro dosažení požadované velikosti datového toku a zároveň zachování úrovně kvantizačního šumu pod úrovní maskovacího prahu se provádí iterativně. Algoritmus sestává ze dvou vnořených smyček viz obr.1.8.[3]

Lmt zde představuje hodnoty globálního maskovacího prahu, *ds* je hodnota zkreslení ve výpočetním pásmu a r počet bitů určených pro kódování rámce. Ve vnější výpočetní smyčce (angl. outer loop nebo distortion loop) srovnává kodér úroveň zkreslení ve výpočetním pásmu s maximální povolenou úrovní pro dané pásmo. Vnitřní smyčka (angl. inner loop nebo rate loop) provádí kvantizaci a kódování spektrálních a měřítkových koeficientů a výpočet celkového počtu bitů potřebných pro zakódování.

Na začátku zpracování každého bloku je nutné nejprve odhadem určit nejnižší hodnotu kvantizačního kroku, při které je výsledný počet bitů využitých pro kódování rámce nižší, než počet bitů, který je rámci přiřazen. Tento odhad se provádí za účelem omezení počtu iterací které musí algoritmus provést, a to pomocí různých numerických metod v závislosti na konkrétní realizaci kodéru. Poté je v rámci vnější smyčky provedena rekvantizace a ve všech výpočetních pásmech realizována kontrola úrovně zkreslení. V případě, že v některém pásmu překročí povolenou úroveň, je v tomto pásmu zkreslení upraveno zvětšením měřítkového koeficientu, čímž se však opět změní počet bitů potřebných pro kvantizaci rámce. Je tudíž opět nutné upravit velikost kvantizačního kroku ve vnitřní smyčce.[13]

Tento proces se opakuje, dokud není dosaženo hodnoty nižší, než je počet bitů rezervovaných pro kódování rámce ($maxbits$), nebo nejsou hodnoty parametru scale_factor v daných pásmech již zvětšeny na svou maximální hodnotu danou normou a zároveň v žádném z výpočetních pásmech nepřekračuje míra zkreslení povolenou úroveň. V případě řešení pro real-time aplikace bývá ještě zahrnuta podmínka dosažení časového limitu určeného pro zpracování bloku za účelem dosažení konstantního zpoždění.[7]



Obr. 1.8: Blokové schéma nelineárního kvantizéru MP3[3]

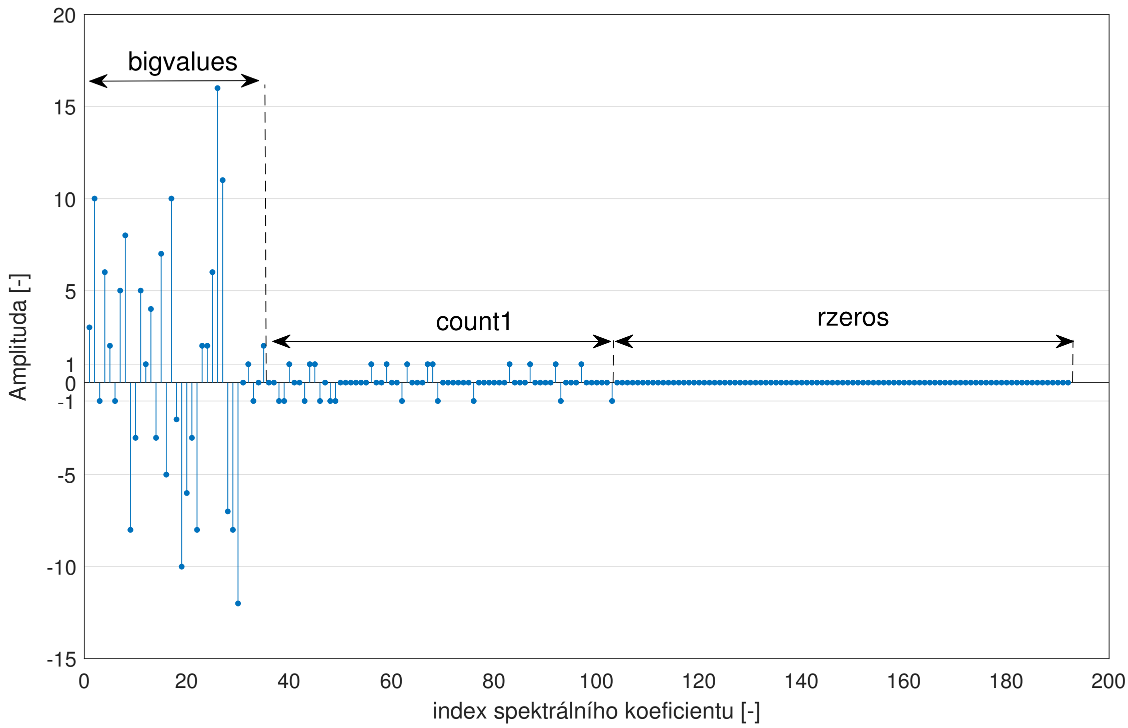
Zkreslení způsobené kvantizací ve výpočetním pásmu se určí dle vztahu

$$scf\_dist\,(b) = \sum_{i=kmin_b}^{kmax_b} \frac{\left(|xr\,(i)| - ix\,(i)^{\frac{4}{3}} \cdot 2^{0{,}25(\mathrm{global\_gain}-210-\mathrm{scale\_factor}(b))}\right)^2}{cbwidth\,(b)} \tag{1.29}$$

Parametr global_gain je kódován na 8 bitů a nabývá rozsahu 0 až 255 zatímco měřítkové koeficienty nabývají různého rozsahu v závislosti na hodnotě parametru scalefac_compress. V případě požadované maximální kvality se používá pro pásma 0-10 (0-5 pro krátké bloky) rozsah 4 bity (0-15) a v pásmech 11-20 (6-11 pro krátké bloky) rozsah 3 bity (0-7).[4]

1.1.6 Kódování v MP3

V rámci MP3 jsou spektrální koeficienty pro kódování rozděleny do tří oblastí zvaných *rzeros*, *count1* a *bigvalues*, které se kódují odlišně. Rozdělení je provedeno na základě kmitočtu, kde na nejvyšších kmitočtech jsou nakvantované hodnoty spektrálních koeficientů výrazně nižší v porovnání s nízkými kmitočty, a je tak možné pro každou oblast použít optimalizovanou sadu Huffmanových tabulek.[3] Příklad rozdělení představuje obr. 1.9, kde jsou zobrazeny nakvantované spektrální koeficienty segmentu váhovaného krátkým oknem.



Obr. 1.9: Rozdělení nakvantovaných spektrálních koeficientů pro kódování

rzeros představuje oblast nejvyšších kmitočtů, kde jsou následkem kvantování hodnoty reprezentovány pouze nulami. Tato oblast se nekóduje, protože její délku je možné zjistit z délky zbylých dvou oblastí a celkové délky rámce.[3]

count1 představuje oblast kde nakvantované spektrální koeficienty nabývají hodnot 1, 0 a -1. Při kódování koeficientů v této oblasti se používají dvě dvě kódovací tabulky a koeficienty jsou kódovány po čtveřicích zvaných *quadruples* na jedno slovo.

Oblast *bigvalues* pak sestává z koeficientů, které nebyly přiřazeny do žádné z předchozích oblastí. Zde se koeficienty kódují po dvojicích pomocí zbylých 30 kódovacích tabulek.[3]

1.1.7 Metody kódování stereofonních signálů

Stereofonní signál reprezentovaný v podobě vektoru vzorků pro každý kanál obsahuje značné množství redundantních informací. Proto byly vyvinuty metody pro kódování stereofonního signálu, které jsou efektivnější než v případě separátního kódování každého kanálu.

Valná většina běžně dostupného hudebního obsahu klade důraz na zvukovou složku nacházející se ve středu stereobáze. Toho využívá například metoda M/S, která pracuje se signálem v podobě kanálu M představujícího součet levého (L) a pravého (R) kanálu a kanálu S, který představuje jejich rozdíl.[1]

$$M = \frac{L+R}{\sqrt{2}} \qquad S = \frac{L-R}{\sqrt{2}} \tag{1.30}$$

Tato transformace je plně reverzibilní, tudíž aplikovatelná na celé spektrum signálu, a běžně se používá i při bezztrátovém kódování. Míra snížení redundance je odvislá od charakteru signálu a maximální účinnosti tato metoda dosahuje v případě, že je signál v obou původních stereofonních kanálech totožný nebo nebo fázově posunutý o π.[1]

V rámci standardu MPEG 1 vrstva 3 se M/S kódování používá v případě, že platí

$$\sum_{k=1}^{N}\left(l_k^2 - r_k^2\right) < 0,8 \sum_{k=1}^{N}\left(l_k^2 + r_k^2\right), \tag{1.31}$$

kde l_k a r_k jsou spektrální koeficienty levého a pravého kanálu získané z FFT v psychoakustickém modelu, a N je délka rámce FFT. V případě použití M/S kódování jsou v psychoakustickém modelu vypočítány upravené hodnoty globálního maskovacího prahu použité pro alokaci bitů.[1]

Další používanou metodou je tzv. *intensity stereo.* Pro určení směru přicházejícího zvuku u vyšších frekvencí (nad 2kHz) využívá náš auditorní systém primárně intenzitní rozdíly mezi signálem přicházejícím do levého a pravého ucha. Pro každé výpočetní pásmo se nejdříve stanoví hodnota energie v levém a pravém kanálu ze spektrálních koeficientů nebo z časového průběhu signálu jako

$$E = \sum_{k_{\min}}^{k_{\max}} |X(k)|^2 = \sum_{n=0}^{N-1} x(n)^2, \tag{1.32}$$

kde $k_{\max}$ $k_{\min}$ nejvyšší a nejnižší index spektrálních koeficientů daného pásma, $x(n)$ jsou vzorky signálu v kritickém pásmu a N je délka signálu. S pomocí energie se následně určí azimut dle sinového či tangentového zákona jako

$$\frac{\sin(\alpha)}{\sin(\alpha_0)} = \frac{E_\mathrm{L} - E_\mathrm{R}}{E_\mathrm{L} + E_\mathrm{R}} \tag{1.33}$$

$$\frac{\tan(\alpha)}{\tan(\alpha_0)} = \frac{E_\mathrm{L} - E_\mathrm{R}}{E_\mathrm{L} + E_\mathrm{R}}, \tag{1.34}$$

kde α je azimut zdroje zvuku, α_0 je azimut pozice reproduktoru (typicky 30°) a E_L a E_R představuje energii v levém a pravém kanále.[2] Přenáší se tak pouze downmix levého a pravého kanálu a informace o azimutu pro dané výpočetní pásmu. Dekodér následně pomoci sinového nebo tangentového zákona určí zesilovací činitele g_L a g_R jako

$$\frac{\sin(\alpha)}{\sin(\alpha_0)} = \frac{g_\mathrm{L} - g_\mathrm{R}}{g_\mathrm{L} + g_\mathrm{R}} \tag{1.35}$$

$$\frac{\tan(\alpha)}{\tan(\alpha_0)} = \frac{g_\mathrm{L} - g_\mathrm{R}}{g_\mathrm{L} + g_\mathrm{R}}, \tag{1.36}$$

pro které platí

$$g_\mathrm{L}^2 + g_\mathrm{R}^2 = 1. \tag{1.37}$$

Poté je váhováním monofonního signálu zesilovacími činiteli vytvořena intenzitní reprezentace původního signálu jako

$$x_\mathrm{L}(t) = g_\mathrm{L} \cdot x_\mathrm{M}(t) \tag{1.38}$$

$$x_\mathrm{R}(t) = g_\mathrm{R} \cdot x_\mathrm{M}(t), \tag{1.39}$$

kde x_M představuje přenášený downmix původního stereofonního signálu.[2]

Při použití této metody dochází k zachování energie, část prostorové informace může být ovšem ztracena. Proto se ve standardu MPEG 1 vrstva 3 používá intensity stereo pouze v případě nižších hodnot datového toku, kde je výhodnější využít takto ušetřené bity pro kódování informace spektrální, která se v tomto případě ukazuje z hlediska subjektivní kvality jako důležitější než informace prostorová.[1]

1.2 Ztrátové kódování v rámci streamovacích služeb

Vzhledem ke zřetelnému trendu nárůstu konzumentů hudby prostřednictvím streamovacích služeb, nabývá tento způsob distribuce hudby na všeobecném významu. Současně s tím se tak zvyšují nároky na výslednou subjektivní kvalitu obsahu a množství dat potřebných pro jeho distribuci. Zde vstupuje do hry ztrátové kódování a jeho využití pro úsporu přenášených dat bez výraznější ztráty subjektivní kvality.

Streamování můžeme definovat jako kontinuální přenos obsahu (nejčastěji audiovizuálního) mezi zdrojovým umístěním souboru (server) a koncovým uživatelem (v našem případě konzument hudby). Rozlišujeme dva základní druhy streamování obsahu, a to přenos v reálném čase (internetová rádia či televize) a model distribuce zvaný „video on demand“. Zde jen nejtypičtějším zástupcem služba YouTube a z placených služeb poskytujících audiovizuální obsah sem můžeme zařadit například Netflix. Na podobném principu fungují také všechny hudební streamovací služby. Základní charakteristikou distribučního modelu video on demand, je možnost koncového uživatele zvolit si obsah z databáze v porovnání s nutností výběru obsahu který je aktuálně vysílán v případě internetové televize či rádia.[14]

Tab. 1.1: Srovnání dostupné kvality streamovacích služeb[1]

služba	**kvalita**				**formát**
	nízká	**standard**	**vysoká**	**velmi vysoká**	
Amazon Music	?	?	320	-	?
Amazon Music HD	-	-	850	3730	?
Apple Music	?	?	256	-	AAC
Deezer	64	128	320	-	MP3
Deezer HiFi	-	-	-	5000	FLAC
SoundCloud Go+	-	-	256	-	AAC
Spotify Free	24	96	160	-	AAC
Spotify Free Web	-	128	-	-	AAC
Spotify Premium	24	96	160	320	AAC
Spotify Premium Web	-	256	-	-	AAC
Tidal HiFi	-	-	1411	2000–9000	FLAC
Tidal Standard	-	320	-	-	AAC
Youtube Music	48	128	256	-	AAC

[1] https://support.spotify.com/cz/article/audio-quality/
https://support.google.com/youtubemusic/answer/9076559?hl=cs
https://support.apple.com/music
https://www.amazon.com/b?ie=UTF8$&$node=14070322011
https://tidal.com/
https://support.deezer.com/hc/en-gb/articles/115003865685-Deezer-Audio-Quality
https://help.soundcloud.com/hc/en-us/articles/360051838074-High-Quality-streaming

V případě služeb nabízejících streamování videa je často používán princip přenosu tzv. multibitrate streamu, který v rámci jednoho datového toku přenáší souběžně data v několika různých variantách kvality. Adaptivní přehrávač na straně koncového uživatele potom v závislosti na rychlosti připojení na server může pro zajištění plynulosti přehrávání přepínat mezi různou velikostí datového toku přehrávaného obsahu.[14]

V rámci služeb pro streamování hudebního obsahu se dostupná kvalita odvíjí od již zmíněné kvality internetového připojení, platformy na které je obsah streamován (webový přehrávač vs. mobilní či desktopová aplikace v případě Spotify) a standardu který si uživatel předplatil. Přehled informací o předplatitelských službách a kvalitě streamovaného obsahu představuje tabulka 1.1 (všechny hodnoty zde uvedené jsou v kb/s).

2 Metody objektivního hodnocení kvality zvukových signálů

Na rozdíl od subjektivních metod hodnocení, které vyžadují účast testovacích subjektů v rámci nákladných a časově náročných poslechových testů, metody objektivního testování využívají dvě základní metody určení kvality zvukového signálu. První využívá porovnání interní sluchové reprezentace referenčního, a testovaného signálu a druhá provádí vyhodnocení šumového (reziduálního) signálu interní reprezentace referenčního signálu a rozdílového signálu.[2] [15]

Mezi metody využívající první zmíněný princip patří například metody Perceptual Audio Quality Measure (PAQM), Perceptual Evaluation of Audio Quality (PEAQ) a metoda PEMO-Q.[2]

2.1 PEMO-Q

Při realizaci hodnocení touto metodou je před vstupem do percepčního modelu nutné signály synchronizovat v čase, aby se zabránilo zkreslení, které může do výpočtu vnášet zpoždění způsobené ztrátovým kódováním. Následně jsou odstraněny rozdíly v RMS úrovni signálů. Přestože časový posun nemá v rámci subjektivního testování žádný vliv, protože posluchač není schopen toto zpoždění postřehnout, do výsledného hodnocení objektivních metod může přinášet značné zkreslení.[15]

Vzájemné zpoždění se určuje pomocí korelace obálek signálů a je kompenzováno zpožděním referenčního signálu o konstantu. Stejně tak je rozdíl v úrovních kompenzován násobením konstantou. Je proto důležité, aby tyto rozdíly byly neměnné v rámci celého testovacího vzorku signálu. V případě že tomu tak není je potřeba teto proces realizovat zvlášť pro každý rámec zpracování signálu. [15]

Třetí krok přípravy signálů pro vstup do percepčního modelu sestává z odstranění částí signálu v nichž výsledná úroveň nepřekročí práh slyšení. Tento krok vychází z předpokladu, že pasáže ticha v nahrávce neovlivňují subjektivní hodnocení testovacích subjektů. Aby se zabránilo odstranění částí, které by mohly obsahovat šum vzniklý kódováním, jsou odstraněny pouze pasáže delší, než je délka jednoho rámce zpracování signálu (až 24 ms u MP3). Aby byla tato délka dodržena, a zároveň aby se mohl projevit efekt časového maskování, nedojde ke kompletnímu odstranění pasáží ticha, pouze k jejich zkrácení na 200ms.[15]

Následuje transformace obou signálů na jejich interní sluchovou reprezentaci pomocí PEMO modelu. Zde je signál filtrován bankou 35 gamatónových filtrů 4-tého řádu kde frekvenční členění odpovídá ERB. [15]

Tab. 2.1: Stupnice hodnocení ukazatele kvality ODG

ODG	**zhoršení**
0,0	neznatelné (*impereceptible*)
-1,0	znatelné, ne nepříjemné (*perceptible, but not annoying*)
-2,0	trochu nepříjemné (*slightly annoying*)
–3,0	nepříjemné (*annoying*)
-4,0	velmi nepříjemné (*very annoying*)

Poté signál prochází řetězcem pěti zpětnovazebních smyček s adaptivními filtry typu dolní propust, které simulují časové maskování. Posledním krokem zpracování je pak detekce amplitudové modulace, jejímž výsledkem je interní sluchová reprezentace signálu.[15]

Výpočet ukazatele PSM_t, která představuje kvalitu testovacího signálu na škále $< 0; 1 >$, kde 1 představuje identitu s referencí, se určí jako "*5% kvantil z váženého průměru vzájemné korelace vnitřních reprezentací reference a testovaného signálu a plovoucího průměru časového průběhu interní reprezentace testovaného signálu*".[12]

Hodnoty PSM_t jsou pak mapovány hyperbolickou funkcí na škálu *Objective Difference Grade* (ODG) jejíž hodnoty přesně odpovídají škále *Subjective Difference Grade* (SDG) používané v rámci subjektivních poslechových testů. Vysvětlení hodnot této stupnice se slovním popisem představuje tabulka 2.1.[15]

3 Realizace kodéru v prostředí MATLAB

V rámci této práce je realizován stereofonní hybridní kodér v prostředí MATLAB s jednoduchým grafickým rozhraním uživatele pro intuitivní testování a demonstraci jak vlivu vstupních parametrů na kvalitu výstupu, tak i grafickou demonstraci některých procesů, které kodér realizuje. Program je optimalizován pro verzi R2020a.

3.1 Programové řešení

Kodér využívá kódů realizujících psychoakustický model a analyzující a syntetizující banku filtrů ze cvičení v předmětu Akustika a zvukové systémy. K němu jsou vytvořeny funkce realizující identifikaci tónových a šumových složek pro přepínání délek okna použitého pro váhování vstupu MDCT, výpočet MDCT, funkce provádějící dynamickou alokaci bitů a kvantizaci podle vzoru MPEG 1 layer 3 a následnou rekvantizaci signálu, IMDCT a zápis do souboru ve formátu LPCM. Blokové schéma kodéru představuje obrázek 3.1 kde $x\left[n\right]$ představuje vzorky vstupního, signálu $x_I\left[n\right]$ vzorky jednoho z I pásem filtrovaných analytickou bankou $X\left[k\right]$ spektrální koeficienty a $L_{\mathrm{M}}(f)$ hodnoty globálního maskovacího prahu.

3.2 Popis funkcí programu

Jádro programu je realizované funkcí *codec*, která volá funkce realizující jednotlivé bloky kodéru. Uživatel může ovlivňovat subjektivní kvalitu výstupního signálu vstupním parametrem *bitrate* a pomocí vstupního parametru *enable_win_switch* může aktivovat nebo deaktivovat přepínání délky oken.

Funkce pracuje s rámci zpracování o délce 576 vzorků nazývanými po vzoru MP3 granule. Datová třída *Granule* definovaná v *Granule.m* slouží k ukládání hodnot různých parametrů a vzorků zpracovávaného signálu, a jejich přenos mezi funkcemi programu. Dále také obsahuje funkce realizující změny formátování matic a vektorů a další pomocné operace.

Vzorky vstupního stereofonního signálu jsou transformovány na signál monofonní a zpracovány funkcí *pazs_c07_model* představující blok psychoakustického modelu. Ta vrací pro rámec vstupního signálu o délce 576 vzorků vektor globálního maskovacího prahu Lm v dB(SPL) a vektor SMR v dB. V rámci psychoakustického modelu je implementován výpočet indexu tonality podle (1.9) a (1.10) pro každý rámec, pomocí něhož se určuje délka použitého okna pro jeho váhování před vstupem do MDCT. Rozhodovací úroveň je zde nastavena na hodnotu parametru $tonIdx = 0,5$.

Z vektoru L_{M} označeného v kódu jako *Lmt* je nutné určit minimální hodnoty pro každé výpočetní pásmo. K tomu funkce slouží *LmttoScfBands*, která vrací minimální

Obr. 3.1: Blokové schéma kodéru

hodnotu *Lmt* v každém pásmu. Rozsahy výpočetních pásem (scalefactor bands) vrací funkce *scalefac_bands_idx* v závislosti vzorkovací frekvenci a délce rámce vstupního signálu. Tyto hodnoty jsou dány normou [4].

Banka pseudo-kvadraturních zrcadlových filtrů ve funkci *PQMFanalysis* provádí filtraci a podvzorkování vstupního signálu. Na výstupu tohoto je po vzoru MP3 signál rozdělený do 32 pásem s 18 vzorky pro každé pásmo.

Následuje blok kmitočtové transformace realizované v podobě modifikované diskrétní kosinové transformace (MDCT) funkcí *MDCT_MP3*. Vstupem této funkce jsou dvě granule signálu označené jako *inputSubband* a *lastInputSubband*, u kterých je provedena frekvenční korekce obrácením znaménka u lichých vzorků sloužící pro eliminaci vlivu frekvenční inverze analytické banky filtrů[3]. Následně jsou obě granule váhovány oknem viz 1.1.4 a je provedena MDCT podle vztahu (1.22). Přepínací logika je realizována funkcí *GetNextWinType* podle schématu 1.5, která je součástí *granule*.

Spektrální koeficienty jsou pomocí funkce *ReshapeToChannels*, která je součástí *Granule*, reorganizovány zpět do podoby vektoru 576 vzorků a to tak, že se MDCT koeficienty ze všech pásem poskládají za sebe. Funkce *PrepareForOuterLoop* resetuje všechny potřebné proměnné a přidělí jim místo v paměti.

Následuje realizace bloku kvantování, kterou provádí podle schématu 1.8 funkce *outer_loop* a *inner_loop*. Nejprve je nutné provést odhad počáteční hodnoty kvantovacího kroku řízené parametrem *global_gain*. Ten provádí funkce *inner_loop* a je v tomto kodéru realizován pomocí metody půlení intervalů.

V rámci funkce *inner_loop* pak funkce *quantizeMP3* realizuje nelineární kvantizaci podle (1.28) a porovnání počtu použitých bitů r (počítá funkce *calculate_bits*) s hodnotou *maxbits*. Protože v rámci kodeku není realizováno Huffmanovo kódování, bylo nutné stanovit hodnotu *maxbits* na základě dostupných informací o formátu datového toku a informací o účinnosti Huffmanova kódování. *Maxbits* se určí jako

$$maxbits = \left[\frac{bitrate \cdot 1000}{f^{\mathrm{vz}}} \cdot 576 - \frac{header + sideinfo}{numGr} \right] \cdot Huffman, \qquad (3.1)$$

kde *header* představuje 32 bitů hlavičky každého rámce MP3, *sideinfo* má rozsah 136 bitů[4], *numGr* je počet granulí v jednom rámci a *Huffman* kompenzuje předpokládanou úsporu dat Huffmanovým kódování (předpokládaná úspora je $1,5:1$).[11]

Funkce *outer_loop* provádí rekvantizaci spektrálních koeficientů a kontrolu kvantizačního zkreslení ve výpočetních pásmech podle vztahu (1.29). Rekvantizaci spektrálních koeficientů provádí funkce *Requantize* podle vztahu inverznímu k (1.28). V rámci kontroly zkreslení v pásmech se iterativně zvětšují hodnoty měřítkových parametrů uložené v proměnné *scalefac_bands*, dokud zkreslení *ds* v daném pásmu neklesne pod hodnotu danou *Lmt*.

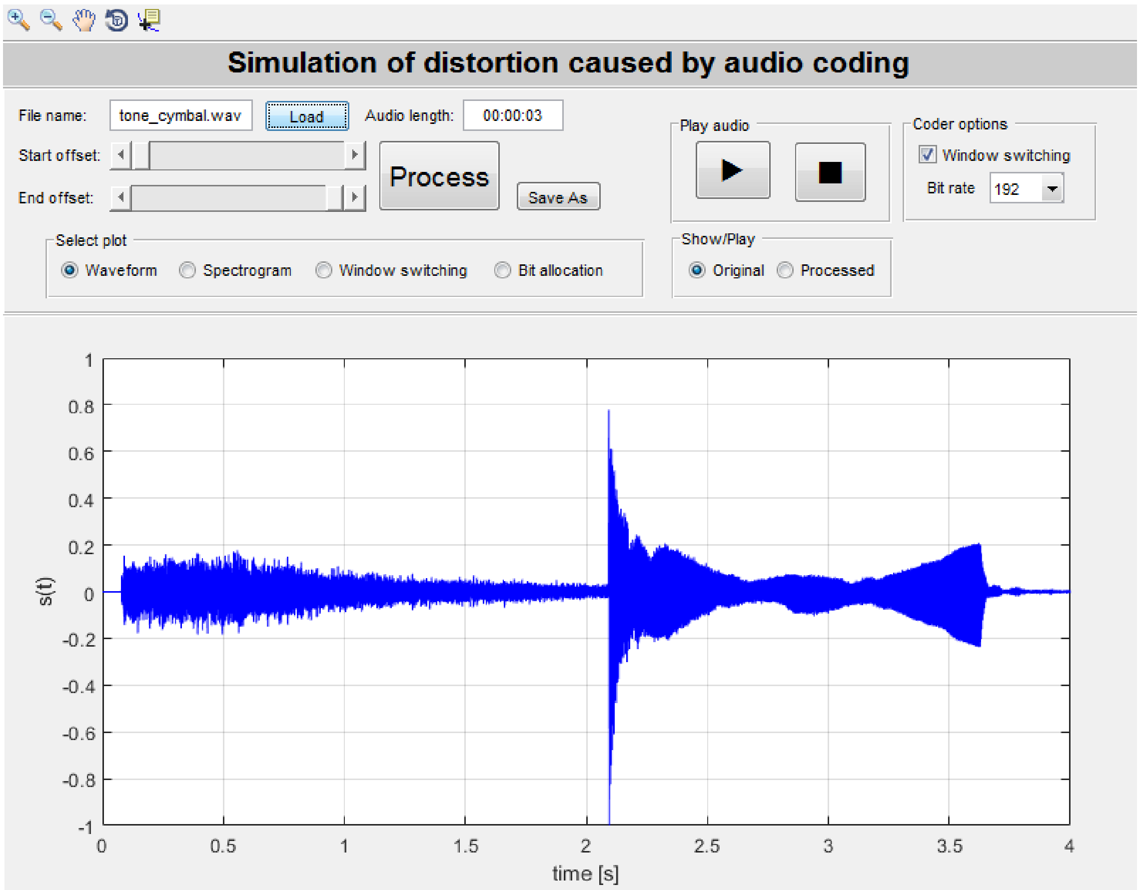
V rámci *outer_loop* je také realizována kontrola podmínek pro ukončení výpočetní smyčky podle 1.8.

V případě krátkých bloků se tento proces realizuje třikrát za sebou pro každých 192 MDCT koeficientů, kterým je přidělena třetina *maxbits*. Celý proces alokace bitů a kvantování se pak provádí pro každý kanál zvlášť. Po ukončení procesu se provede rekvantizace spektrálních koeficientů pomocí výsledných hodnot *global_gain* a *scalefac_bands* a koeficienty se přeformátují pomocí funkce *ReshapeToSubband* do matice pro vstup do IMDCT.

Tu realizuje (funkce *IMDCT_MP3*). Poté je provedena filtrace syntetizující bankou filtrů (*PQMFsynthesis*) jejímž výstupem jsou opět LPCM vzorky signálu.

3.3 Grafické rozhraní uživatele

Grafické rozhraní uživatele (obr. 3.2) se spouští funkcí *MP3GUI.m*. Umožňuje uživa-

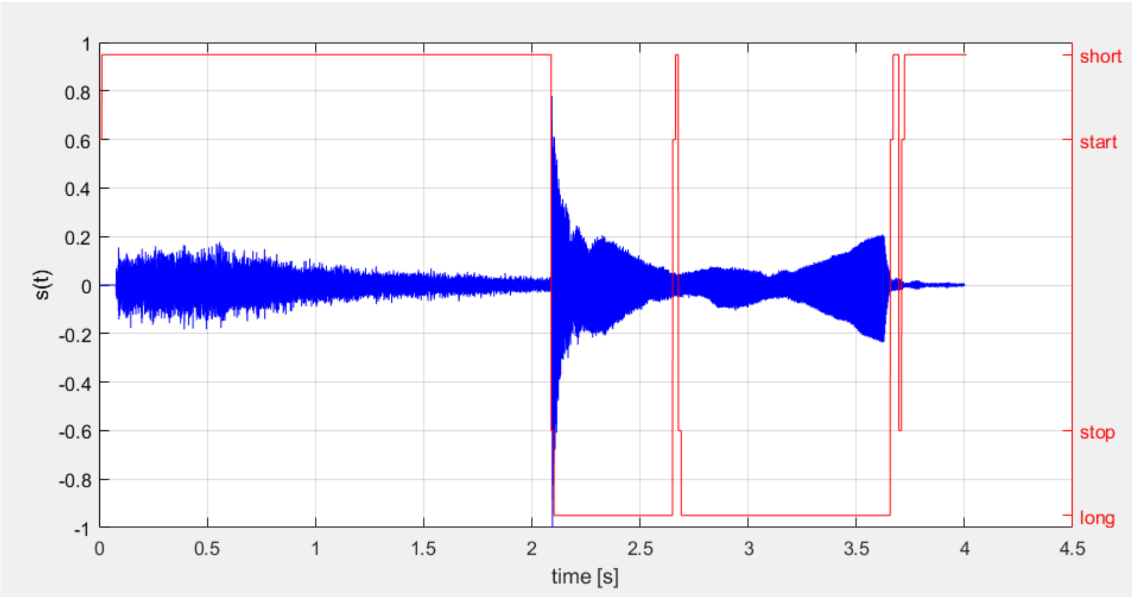


Obr. 3.2: Grafické rozhraní uživatele

teli načíst soubor pro analýzu, vybrat část signálu která se bude zpracovávat, vybrat vstupní parametry kodeku (hodnotu datového toku a aktivaci/deaktivaci přepínání oken), přehrát původní a zkreslený signál a obsahuje čtyři různé režimy grafického rozhraní.

Režimy *Waveform* a *Spectrogram* zobrazují časový průběh signálu a spektrogram. U spektrogramu je při nastavení nižších hodnot datového toku zřetelně vidět úbytek energie na vyšších frekvencích u zkresleného signálu.

Režim zobrazení *Window switching* slouží pro demonstraci přepínání mezi různými délkami oken, které byly použity v průběhu zpracování signálu. Na obrázku 3.3 je výsledek analýzy zvukového souboru tone_ cymbal.wav, který je součástí přílohy.

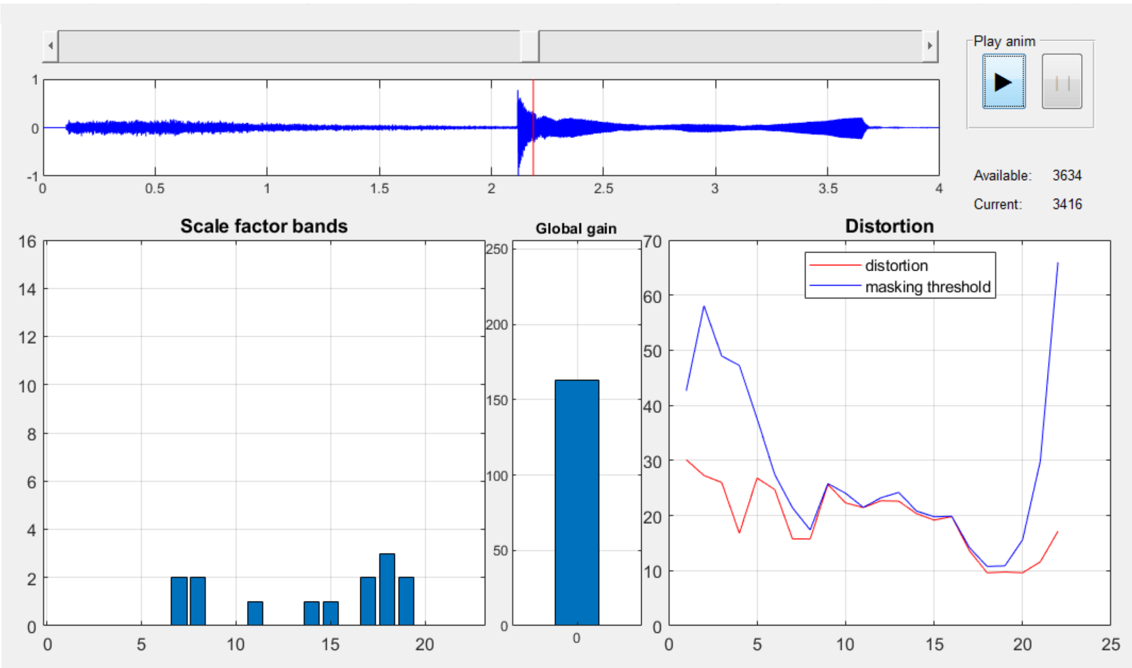


Obr. 3.3: Grafika pro ilustraci přepínání délek oken

Zde je patrné, že v průběhu prvních dvou vteřin, po které v nahrávce zní činel (šumový signál), jsou jsou použita pro váhování krátká okna. Následující vteřinu a půl zní činel společně s hlasitějším syntetizérem, a tak zde převládá tónový charakter a dlouhá okna. Na závěr opět doznívá samotný činel.

Bit allocation umožňuje přehrát animaci demonstrující proces alokace bitů při kódování rámce popsaný v 1.1.5. Konečný stav procesu ukazuje obrázek 3.4.

Pomocí posuvníku uživatel vybere část signálu pro kterou se má zobrazit animace. *Masking threshold* zde představuje *Lmt*, hodnota zkreslení ve výpočetních pásmech je označena *distortion*, *Available* představuje hodnotu *maxbits* a *Current* hodnotu *r*. Animaci a výpočet hodnot pro ni realizují funkce *bitAloc_kodek* a *bitAloc_outer_loop*.



Obr. 3.4: Animace pro ilustraci realizace alokace bitů

4 Ověření účinnosti funkcí kodeku pomocí PEMO-Q

Testování vlivu vstupních parametrů kodeku na kvalitu výstupního signálu byla provedena pomocí veřejně dostupné implementace metody PEMO-Q pro MATLAB realizované v rámci [12].

Pro testování byla byla použita databáze testovacích nahrávek, které byly zvoleny na základě doporučení [17] pro co největší tembrální, dynamickou i rytmickou rozmanitost, různé poměry zastoupení akustických a elektronických nástrojů jakož i signálů s jasným tónovým charakterem či výraznými transienty. Seznam použitých nahrávek ukazuje tabulka 4.1.

Tab. 4.1: Databáze testovacích nahrávek

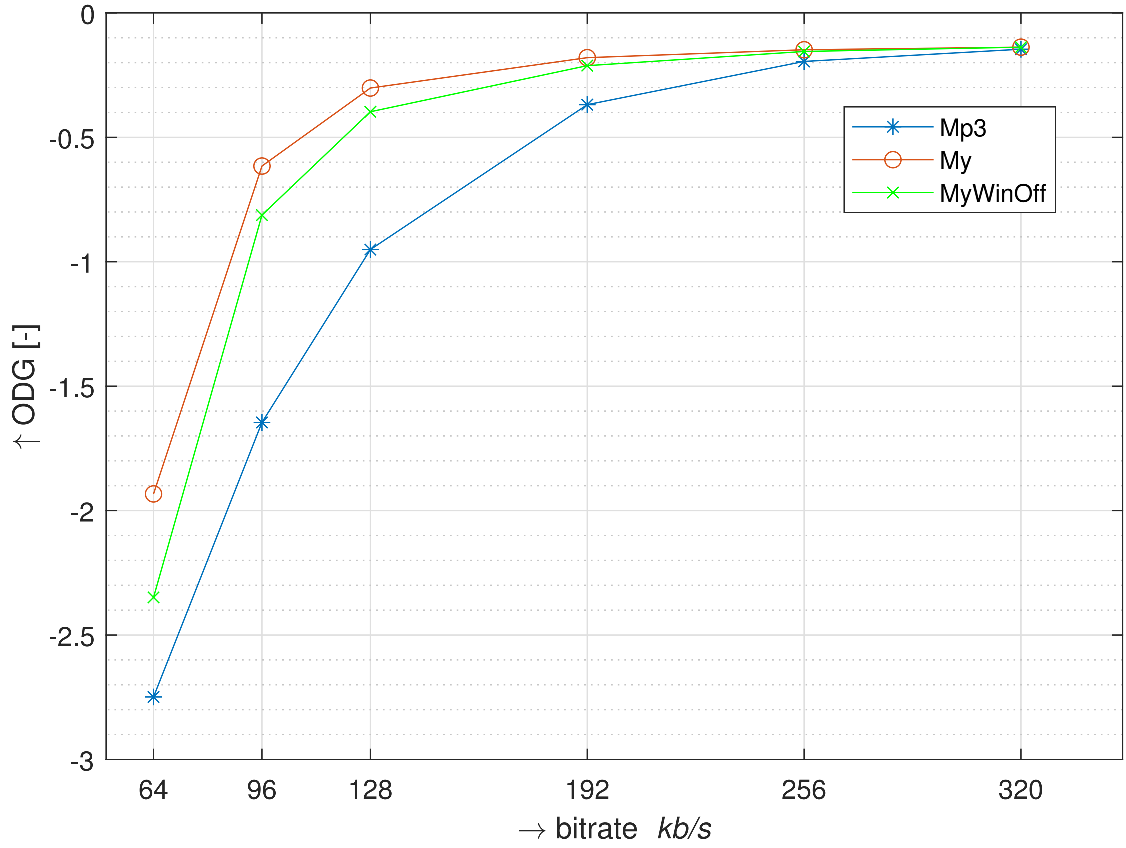
ID	**interpret - skladba**	**žánr**
1	Metallica – Sad But True	metal
2	Skrillex – Scary Monsters And Nice Sprites	electro
3	Leny Andrade – Maiden voyage	vocal jazz
4	The Fred Hersh Trio – Played Twice	instrumental jazz
5	Jamiroquai – Runaway	pop
6	Johnny Frigo – I Love Paris	instrumental jazz
7	The Connecticut Early Music Ensemble – Vivaldi Flute Concerto in D	classical-baroque
8	Westminster Choir – Britten Festival Te Deum	classical-choir
9	Solisti New York – Stravinsky The Royal March	brass band
10	Chris Jones – No Sanctuary Here	bass-heavy rock

Všechny použité testovací nahrávky byly z bezztrátového formátu FLAC převedeny do LPCM formátu .wav pro následné zpracování.

Jako referenčního kódéru bylo užito MP3 enkodéru dostupného v rámci freeware programu Format Factory. Všechny testovací nahrávky byly kódovány jako MP3 s konstantním datovým tokem o šesti různých hodnotách, a to $64, 96, 128, 192, 256$, 320 kb/s. Následně byly nahrávky opět převedeny do formátu .wav který vyžaduje aplikace realizující PEMO-Q analýzu.

Testovací nahrávky byly pro srovnání zpracovány také kodekem realizovaným v rámci této práce při stejných nastaveních datového toku, a to ve dvou variantách - s aktivovaným přepínáním délek okna MDCT a bez. Kompletní výsledky hodnot ODG získaných objektivním testem PEMO-Q představuje tabulka 4. *MP3* zde představuje nahrávky zpracované referenčním kodekem, *My* zpracovaném kodekem vytvořeným v rámci této práce s aktivovaným přepínáním délek okna MDCT a *MyWinOFF* s deaktivovaným přepínáním.

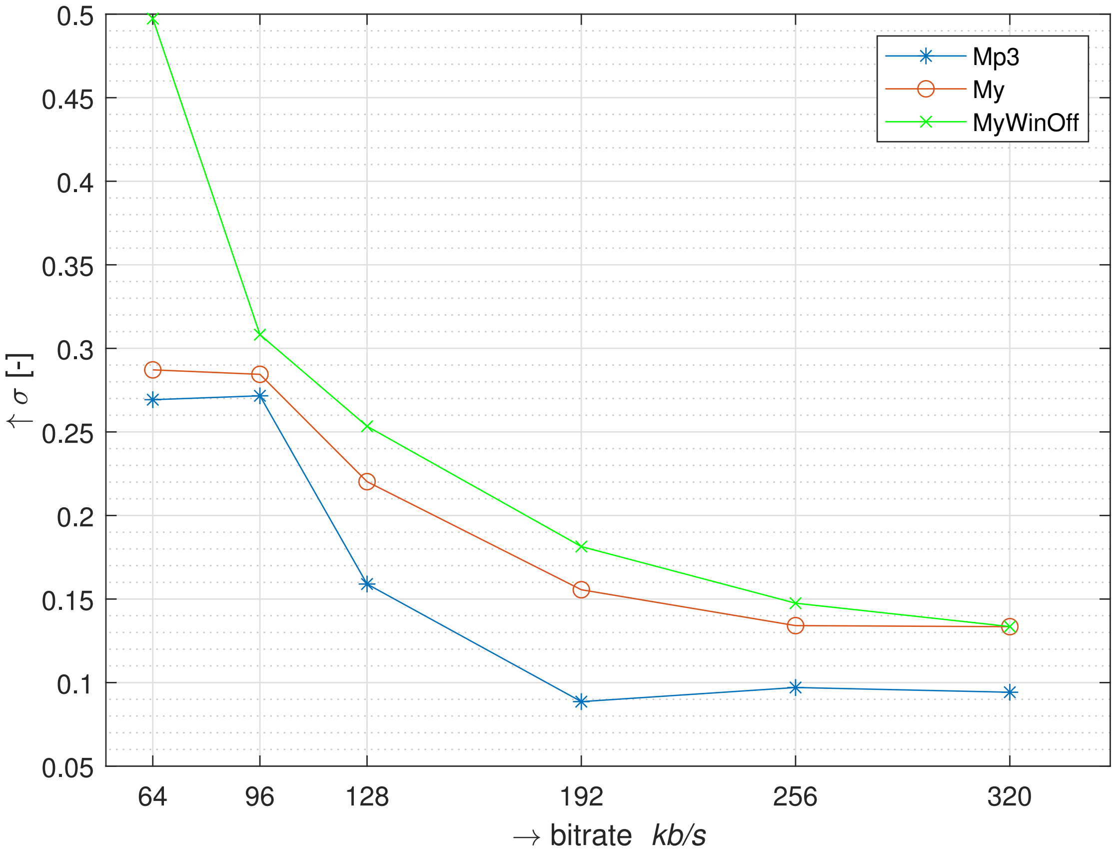
Graf 4.1 pak představuje závislost střední hodnoty ODG ze souboru vzorků na požadované hodnotě datového toku (bitrate) pro referenční kodek a obě varianty nastavení kodeku realizovaného v rámci této práce. Zde je patrné, že při vysokých hodnotách datového toku (320 kb/s) je výstupní kvalita téměř totožná, a to i nezávisle na použití přepínání délek oken kmitočtové transformace. Se snižující se bitrate pak přepínání oken značně zlepšuje výstupní kvalitu. Při přímém porovnání s referenčním kodekem je patrné, že závislost ODG na bitrate ve všech případech kopíruje podobný trend, což ukazuje na úspěšnost této implementace. Zřetelné zhoršení kvality při snížení bitrate na 64 kb/je pak ve všech případech bezpečně rozeznatelné i poslechem.



Obr. 4.1: Srovnání závislosti střední hodnoty ODG na zvolené bitrate

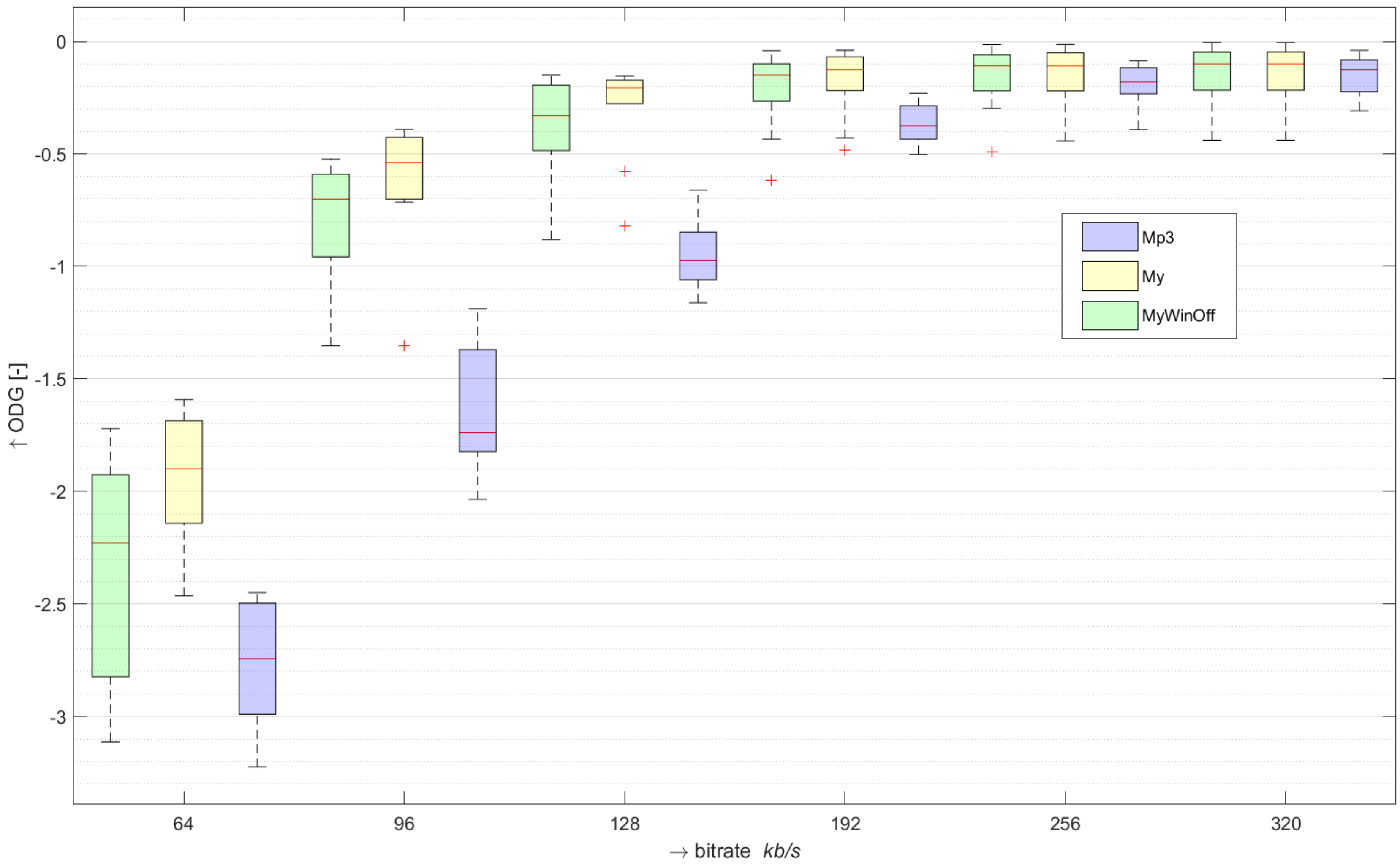
Graf 4.2 pak představuje závislost směrodatné odchylky na bitrate. Zde je naopak patrné, že rozdíly ve výstupní kvalitě jsou závislejší na charakteru vstupního signálu v případě nižšího datového toku. Patrný je zde také vliv deaktivace přepínání oken na výraznější rozptyl výsledné kvality. Zde je opět patrné, že při aktivaci přepínání křivka kopíruje křivku referenční což, poukazuje na správnou funkci této součásti implementace.

Detailnější pohled do statistických výsledků testu pak ukazuje krabicový graf



Obr. 4.2: Srovnání závislosti směrodatné odchylky ODG na zvolené bitrate

4.3, kde za odlehlé hodnoty jsou označeny hodnoty jejichž vzdálenost od 25. nebo 75. percentilu statistického souboru je větší než $1,5$ násobek vzdálenosti mezi 25. a 75. percentilem. Zde je například jasně patrné, že referenční kodek neprodukuje žádné odlehlé hodnoty v porovnání s kodekem vytvořeným v rámci této práce.



Obr. 4.3: Krabicový graf výsledků testu

Tab. 4.2: Výsledky PEMO-Q analýzy databáze testovacích ukázek

MP3										
bitrate\ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
320	-0,08	-0,09	-0,16	-0,27	-0,22	-0,16	-0,04	-0,04	-0,08	-0,31
256	-0,15	-0,19	-0,23	-0,3	-0,23	-0,17	-0,09	-0,09	-0,12	-0,39
192	-0,39	-0,38	-0,48	-0,43	-0,35	-0,37	-0,27	-0,23	-0,29	-0,5
128	-1,05	-1,1	-1,16	-0,85	-0,89	-1,06	-0,79	-0,66	-0,9	-1,05
96	-1,76	-2,04	-1,8	-1,32	-1,59	-1,82	-1,37	-1,19	-1,72	-1,85
64	-3,02	-3,22	-2,52	-2,46	-2,75	-2,99	-2,45	-2,5	-2,74	-2,83
My										
bitrate\ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
320	-0,05	-0,07	-0,13	-0,24	-0,22	-0,15	-0,01	-0,01	-0,07	-0,44
256	-0,05	-0,07	-0,14	-0,28	-0,22	-0,16	-0,01	-0,03	-0,08	-0,44
192	-0,07	-0,08	-0,16	-0,43	-0,22	-0,16	-0,04	-0,07	-0,09	-0,48
128	-0,2	-0,21	-0,24	-0,82	-0,28	-0,2	-0,15	-0,17	-0,17	-0,58
96	-0,45	-0,7	-0,49	-1,35	-0,42	-0,39	-0,58	-0,62	-0,43	-0,72
64	-2,14	-2,46	-1,71	-2,01	-1,94	-1,59	-2,26	-1,69	-1,66	-1,86
MyWinOff										
bitrate\ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
320	-0,05	-0,07	-0,13	-0,24	-0,22	-0,15	-0,01	-0,01	-0,07	-0,44
256	-0,05	-0,07	-0,14	-0,28	-0,22	-0,16	-0,01	-0,03	-0,08	-0,44
192	-0,07	-0,08	-0,16	-0,43	-0,22	-0,16	-0,04	-0,07	-0,09	-0,48
128	-0,2	-0,21	-0,24	-0,82	-0,28	-0,2	-0,15	-0,17	-0,17	-0,58
96	-0,45	-0,7	-0,49	-1,35	-0,42	-0,39	-0,58	-0,62	-0,43	-0,72
64	-2,14	-2,46	-1,71	-2,01	-1,94	-1,59	-2,26	-1,69	-1,66	-1,86

Závěr

V rámci teoretické části této práce je realizována rešerše teoretického základu pro tvorbu programu simulujícího zkreslení vznikajícího percepčním kódováním zvukového signálu. Konkrétně je zde popis základních bloků hybridního kodéru s důrazem na řešení v rámci standardu MPEG-1 vrstva 3.

V praktické části je pak realizován stereofonní hybridní kodér z těchto principů vycházející a k němu příslušící grafické uživatelské rozhraní, které umožňuje demonstraci vlivu vstupních parametrů na kvalitu výstupního signálu, demonstraci přepínání oken použitých pro váhování signálu před vstupem do MDCT a demonstraci principu dynamické alokace bitů.

Pro ověření funkčnosti kodéru bylo provedeno objektivní srovnání kvality jeho výstupu s výstupem běžně dostupného MP3 kodéru pro odpovídající nastavení vstupních parametrů. Toto srovnání bylo realizováno pomocí metody PEMO-Q a bylo provedeno na databázi deseti zvukově a žánrově kontrastních nahrávek pro šest různých nastavení požadovaného datového toku. Zde výsledky ukázaly kladný vliv aktivace funkce přepínání délky oken na výslednou objektivní kvalitu a na snížení rozptylu jejích hodnot v závislosti na charakteru vstupního signálu. Dále se zde ukázal rozdíl objektivní kvality při nižších požadovaných hodnotách datového toku ve srovnání s MP3 kodérem. Tento rozdíl je patrně způsoben nepřesností výpočtu bitů potenciálně potřebných pro zakódování rámce u kodéru realizovaného v rámci této práce. V případě, že by program měl simulovat zkreslení právě MP3 kodérem, je možné hodnoty vstupního parametru představujícího požadovaný datový tok upravit tak, aby se výsledky shodovaly. Zde by bylo ovšem vhodné provést rozsáhlejší objektivní a subjektivní testování.

Oproti zadání práce se nepodařilo realizovat pouze obohacení kodéru o různé druhy stereofonního kódování signálů z důvodu nedostatku dostupných informací o práci s měřítkovými koeficienty a hodnotami povoleného zkreslení v pásmech v režimu joint stereo coding.

Program je vzhledem k intuitivnímu rozhraní a obsahu demonstrativních funkcí možné použít v rámci výuky.

Literatura

[1] BOSI, Marina a Richard E. GOLDBERG. Introduction to digital audio coding and standards. New York: Springer, 2003. ISBN 1-4020-7357-7.

[2] SCHIMMEL, PH.D., doc.Ing. Jiří. Akustika a zvukové systémy. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018.

[3] Jayaraman Thiagarajan; Andreas Spanias, Analysis of the MPEG-1 Layer III (MP3) Algorithm using MATLAB, Morgan & Claypool, 2011, doi: 10.2200/S00382ED1V01Y201110ASE009.

[4] ISO/IEC 11172-3:1993. Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s — Part 3: Audio. 1. ISO, 1993.

[5] SMÉKAL,Z,*Analýza signálů a soustav-BASS*. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2016. ISBN 978-80-214-4716-5.

[6] E. Kurniawati, J. Absar, S. George, C. T. Lau and B. Premkumar, "An investigation into different masking behaviours resulting from estimation of tonality index,"2002 14th International Conference on Digital Signal Processing Proceedings. DSP 2002 (Cat. No.02TH8628), Santorini, Greece, 2002, pp. 1035-1038 vol.2, doi: 10.1109/ICDSP.2002.1028267.

[7] Jacaba S., Joebert. Audio compression using modified discrete cosine transform: the MP3 coding standard. (2001).

[8] Uncertainty principle. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2021-04-22]. Dostupné z: https://en.wikipedia.org/wiki/Uncertainty_principle

[9] NOVÁK, Vladimír. *Percepční kódování zvukových signálů*. Brno, 2011. Bakalářská. Vysoké učení technické v Brně.

[10] WU, Guixing a En-hui YANG. Optimization of MP3 audio encoding by scale factors and global quantization step size. Sep. 13, 2012. United States. US 2012/0232911 A1. Uděleno May 22, 2012. Zapsáno Sep. 13, 2012.

[11] SMITH, Steven W. *The Scientist and Engineer's Guide to Digital Signal Processing*. 2. 1997. ISBN 0-9660176-3-3. Dostupné také z: http://www.dspguide.com/

[12] NOVÁK, Jan. *Implementace metody objektivního hodnocení kvality zvuku.* Praha, 2018. Bakalářská. České vysoké učení technické v Praze.

[13] You, Shingchern & Chen, Woei-Kae. (2008). Efficient quantization algorithm for real-time MP-3 encoders. Multimedia Tools Appl.. 40. 341-359. 10.1007/s11042-008-0210-7.

[14] Streaming media. *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2021-5-4]. Dostupné z: https://en.wikipedia.org/wiki/Streaming_ media# Technologies

[15] HUBER, Rainer a Birger KOLLMEIER. PEMO-Q-A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING.* 2006, **14**(6), 10.

[16] T. Dau, B. Kollmeier, and A. Kohlrausch, "Modeling auditory processing of amplitude modulation: I—modulation detection and masking with narrowband carriers," J. Acoust. Soc. Amer., vol. 102, no. 5, pp. 2892–2905, 1997

[17] *ITU-R BS.1116-3: Methods for the subjective assessment of small impairments in audio systems.* 2015.

Seznam symbolů, veličin a zkratek

DFT	diskrétní Fourrierova transformace
FFT	Fast Fourrier Transform
MPEG	Moving Picture Experts Group
MP3	MPEG-1 vrstva 3
LC-ATC	Low Complexity Adaptive Transform Coding
MDCT	modifikovaná diskrétní kosinová transformace
ODG	Objective Difference Grade
PE	percepční entropie
PQMF	pseudo-kvadraturní zrcadlové filtry
PSD	Power Spectral Density
SDG	Subjective Difference Grade
SNR	Signal to Noise Ratio
SMR	Signal to Mask Ratio
SPL	Sound Pressure Level