

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁRSKA PRÁCA

Brno, 2021

ANDREJ HUDEC



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

**AUDIOKODEKY A TESTY PRO ZHODNOCENÍ JEJICH
POSLECHOVÉ KVALITY**
TITLE OF STUDENT'S THESIS

BAKALÁRSKA PRÁCA
BACHELOR'S THESIS

AUTOR PRÁCE
AUTHOR

Andrej Hudec

VEDÚCI PRÁCE
ADVISOR

doc. Mgr. Pavel Rajmic, PhD.

BRNO 2021

Bakalářská práce

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

Student: Andrej Hudec

ID: 203728

Ročník: 3

Akademický rok: 2020/21

NÁZEV TÉMATU:

Audiokodeky a testy pro zhodnocení jejich poslechové kvality

POKYNY PRO VYPRACOVÁNÍ:

Nastudujte principy, které umožňují neztrátovou i ztrátovou komprimaci audiosignálu. Seznamte se s nejvýznamnějšími zástupci kodeků (např. FLAC, MP3, OGG). Zaměřte se přitom také na dostupnost volného kódu (open source). Dle normy [2] navrhnete poslechové testy, které budou hodnotit kvalitu komprimace při fixním bitovém toku. Vybrané kodeky pak v tomto testu použijte. Testovací prostředí navrhnete a implementujete ve vhodném programovacím jazyku. Provedte statistické vyhodnocení výsledků a učiňte závěry. Neopomeňte srovnání kodeků z hlediska výpočetní náročnosti. V textové části práce se soustředte zejména na detailní popis rozdílů mezi vybranými kodeky.

DOPORUČENÁ LITERATURA:

[1] SPANIAS, A. a kol., Audio Signal Processing and Coding. Wiley, 2007. ISBN 978-0-471-79147-8

[2] ITU-R, Methods for the subjective assessment of small impairments in audio systems, Recommendation ITU-R BS.1116-3, 2015.

Termín zadání: 1.2.2021

Termín odevzdání: 31.5.2021

Vedoucí práce: doc. Mgr. Pavel Rajmic, Ph.D.

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

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

Bakalárska práca sa zaoberá audio kodekmi a testami pre zhodnotenie ich posluchovej kvality. Práca ponúka popis vlastností ľudského sluchu, stručný úvod do problematiky audio kódovania, kompresie zvukových dát a posluchové testy. Bakalárska práca tiež obsahuje teoretický popis kódovacích procesov zvolených audio kodekov, vyhodnotenie merania ich výpočtovej náročnosti a vyhodnotenie posluchovej kvality každého z nich pomocou posluchové testy. Na implementáciu prostredia posluchové testy boli použité programovacie jazyky HTML, CSS, Javascript a Python. Zvukové súbory, ktoré boli predmetom skúmania v posluchové testoch boli získané pomocou knižnice FFmpeg.

KLÚČOVÉ SLOVÁ

kódovanie zvuku, kóder, dekóder, kodek, mp3, ogg, vorbis, opus, aac, posluchový test, stratová kompresia, psychoakustický model

ABSTRACT

The bachelor thesis deals with audio codecs and tests to evaluate their listening quality. The thesis offers a description of the properties of human hearing, a brief introduction to the issues of audio coding, types of audio data compression and listening tests. The bachelor thesis also contains a theoretical description of the coding processes of selected audio codecs, evaluation of measuring their computational complexity and evaluation of listening quality of each of them using listening tests. HTML, CSS, Javascript and Python programming languages were used to implement the listening test environment. The audio files examined in the listening tests were obtained using the FFmpeg library.

KEYWORDS

audio coding, coder, decoder, codec, mp3, ogg, vorbis, opus, aac, listening test, lossy compression, psychoacoustic model

HUDEČ, Andrej. *Audiokodeky a testy pro zhodnocení jejich poslechové kvality*. Brno, 2021, 76 s. Bakalárska práca. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedúci práce: doc. Mgr. Pavel Rajmic, PhD.

VYHLÁSENIE

Vyhlasujem, že svoju bakalársku prácu na tému „Audiokodeky a testy pro zhodnocení jejich poslechové kvality“ som vypracoval samostatne pod vedením vedúceho bakalárskej práce, s využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

Ako autor uvedenej bakalárskej práce ďalej vyhlasujem, že v súvislosti s vytvorením tejto bakalárskej práce som neporušil autorské práva tretích osôb, najmä som nezasiahol nedovoleným spôsobom do cudzích autorských práv osobnostných a/alebo majetkových a som si plne vedomý následkov porušenia ustanovenia § 11 a nasledujúcich autorského zákona Českej republiky č. 121/2000 Sb., o práve autorskom, o právach súvisiacich s právom autorským a o zmene niektorých zákonov (autorský zákon), v znení neskorších predpisov, vrátane možných trestnoprávných dôsledkov vyplývajúcich z ustanovenia časti druhej, hlavy VI. diel 4 Trestného zákonníka Českej republiky č. 40/2009 Sb.

Brno

.....

podpis autora

POĎAKOVANIE

Rád by som poďakoval vedúcemu práce doc. Mgr. Pavlovi Rajmicovi, PhD. za odbornú pomoc, cenné rady, konzultácie a čas, ktorý si práca vyžadovala.

Obsah

Úvod	12
1 Ľudský sluch	13
1.1 Hranice počutelnosti ľudského sluchu	13
1.2 Frekvenčná analýza v ľudskom sluchu	14
1.3 Efekt maskovania	15
1.3.1 Frekvenčné maskovanie	15
1.3.2 Temporálne maskovanie	16
2 Kódovanie audio signálov	17
2.1 Bitová rýchlosť	17
2.2 Bezstratová kompresia	17
2.2.1 Run-length kódovanie	18
2.2.2 Huffmanovo kódovanie	18
2.3 Stratová kompresia	19
2.3.1 Psychoakustický model	19
2.3.2 Joint stereo kódovanie	20
2.4 Knižnica FFmpeg	21
2.5 Metodika výberu audiokodekov	21
2.6 Postupy kódovania vo vybraných audio kodekoch	22
2.6.1 MP3	22
2.6.2 AAC	25
2.6.3 Vorbis	29
2.6.4 Opus	31
2.7 Porovnanie postupov kódovania zvolených kodekov	34
3 Testy pre hodnotenie posluchovej kvality	36
3.1 Test ABX	36
3.2 Test ABC/HR	36
3.2.1 Hodnotenie percepčného rozdielu poskytnutých stôp	37
3.3 Test MUSHRA	37
3.4 Popis parametrov testu podľa ITU-R	38
3.4.1 Výber respondentov	38
3.4.2 Fáza školenia	38
3.4.3 Testovacia metóda	38
3.4.4 Zvukový materiál pre testovanie	39
3.4.5 Reprodukčné zariadenia	39

3.4.6	Fáza vyhodnocovania výsledkov	39
3.4.7	Prezentovanie výsledkov	39
3.5	Stručný popis použitých štatistických metód	39
3.5.1	t-test	39
3.5.2	ANOVA	40
3.6	Výsledná štruktúra testu	40
4	Príprava programového materiálu	42
4.1	Editácia a kódovanie	42
4.2	Spektrogramy kódovaných audio súborov	43
5	Testovacie prostredie	45
5.1	Úvodná stránka	45
5.1.1	Inštruktážne video	45
5.1.2	Odporúčania pre nastavenie audio reťazca	45
5.1.3	Ukážky úrovni znehodnotenia zvuku vplyvom kompresie	46
5.1.4	Vstupný formulár pre zber dát od respondentov	47
5.2	Testovacie prostredie posluchových testov	47
5.2.1	Blok inštrukcií	48
5.2.2	Blok pre prehrávanie a voľbu prehrávanej stopy	48
5.2.3	Blok identifikácie a hodnotenia kvality	49
5.2.4	Výstupné súbory posluchových testov	50
6	Výsledky práce	52
6.1	Normatívna filtrácia respondentov pred štatistickým vyhodnotením	52
6.1.1	Korekcia získaných dát	52
6.1.2	Selekcia respondentov pred štatistickým vyhodnotením analýzou rozptylu	53
6.2	Štatistické vyhodnotenie a prezentácia výsledkov	56
6.2.1	Vyhodnotenie výsledkov – ukážka JAZZ	56
6.2.2	Vyhodnotenie výsledkov – ukážka METAL	61
6.2.3	Vyhodnotenie výsledkov – ukážka SYMFÓNIA	64
6.3	Meranie výpočtovej náročnosti	67
	Záver	69
	Literatúra	71
	Zoznam symbolov, veličín a skratiek	73
	Zoznam príloh	75

Zoznam obrázkov

1.1	Sluchové pole	14
1.2	Príklad frekvenčného maskovania	16
1.3	Príklad temporálneho maskovania	16
2.1	Príklad Run-Length kódovania	18
2.2	Jednoduchý príklad tvorby binárneho stromu pri Huffmanovom kódovaní	19
2.3	Bloková schéma MP3 kódera	22
2.4	Bloková schéma AAC kódera	26
2.5	Bloková schéma Vorbis kódera	29
2.6	Bloková schéma OPUS kódera	31
4.1	Spektrogram pôvodného súboru (ukážka metal)	43
4.2	Ukážky sprektrogramov (metal)	44
5.1	Panely s ukážkami znehodnotenia	46
5.2	Vstupný formulár	47
5.3	Inštrukcie k testom	48
5.4	Blok pre prehrávanie a voľbu stopy	48
5.5	Blok identifikácie a hodnotenia kvality	49
5.6	Výstupný CSV súbor	50
6.1	Graf zastúpenia reprodukčných zariadení respondentov	52
6.2	Rozloženie r -hodnôt kodekov (JAZZ / 96 kbps)	56
6.3	Rozloženie r -hodnôt kodekov (JAZZ / 80 kbps)	58
6.4	Rozloženie r -hodnôt kodekov (METAL / 96 kbps)	61
6.5	Rozloženie r -hodnôt kodekov (METAL / 80 kbps)	62
6.6	Rozloženie r -hodnôt kodekov (SYMFÓNIA / 96 kbps)	64
6.7	Rozloženie r -hodnôt kodekov (SYMFÓNIA / 80 kbps)	65

Zoznam tabuliek

2.1	Kritériá a konečný výber skúmaných kodekov	21
3.1	Preddefinované body poskytnuté respondentovi	37
3.2	Kľúčové parametre definované normou ITU-R BS.1116-3	40
3.3	Zvolené parametre posluchového testu podľa normy ITU-R BS.1116-3	40
3.4	Prehľad obsahu testov	41
6.1	Možné kombinácie hodnotenia respondenta	52
6.2	Štatistické hypotézy t-testu	55
6.3	Výsledok selekcie respondentov	55
6.4	Logika vyhodnocovania p hodnôt v štatistickej analýze	56
6.5	Vplyv faktoru KODEK na výsledky r -hodnôt (JAZZ / 96 kbps) . . .	57
6.6	Signifikantné rozdiely medzi kodekmi (JAZZ / 96 kbps)	57
6.7	Vplyv faktoru KODEK na výsledky r -hodnôt (JAZZ / 80 kbps) . . .	58
6.8	Signifikantné rozdiely medzi kodekmi (JAZZ / 80 kbps)	58
6.9	Vplyv faktorov BITRATE, KODEK a ich interakcie na výsledky r - hodnôt (JAZZ)	59
6.10	Signifikantné rozdiely medzi kodekmi pri zohľadnení faktorov BIT- RATE, KODEK a ich interakcie (JAZZ)	59
6.11	Zostupné zoradenie kodekov podľa dosiahnutých výsledkov (JAZZ) . .	60
6.12	Vplyv faktoru KODEK na výsledky r -hodnôt (METAL / 96 kbps) .	61
6.13	Signifikantné rozdiely medzi kodekmi (METAL / 96 kbps)	61
6.14	Vplyv faktoru KODEK na výsledky r -hodnôt (METAL / 80 kbps) .	62
6.15	Signifikantné rozdiely medzi kodekmi (METAL / 80 kbps)	62
6.16	Vplyv faktorov BITRATE, KODEK a ich interakcie na výsledky r - hodnôt (METAL)	63
6.17	Signifikantné rozdiely medzi kodekmi pri zohľadnení faktorov BIT- RATE, KODEK a ich interakcie (METAL)	63
6.18	Zostupné zoradenie kodekov podľa dosiahnutých výsledkov (METAL)	63
6.19	Vplyv faktoru KODEK na výsledky r -hodnôt (SYMFÓNIA / 96 kbps)	64
6.20	Signifikantné rozdiely medzi kodekmi (SYMFÓNIA / 96 kbps)	64
6.21	Vplyv faktoru KODEK na výsledky r -hodnôt (SYMFÓNIA / 80 kbps)	65
6.22	Signifikantné rozdiely medzi kodekmi (SYMFÓNIA / 80 kbps)	65
6.23	Vplyv faktorov BITRATE, KODEK a ich interakcie na výsledky r - hodnôt (SYMFÓNIA)	66
6.24	Signifikantné rozdiely medzi kodekmi pri zohľadnení faktorov BIT- RATE, KODEK a ich interakcie (SYMFÓNIA)	66
6.25	Zostupné zoradenie kodekov podľa dosiahnutých výsledkov (SYMFÓ- NIA)	66

6.26	Výsledky merania časov kódovania	67
6.27	Výsledky merania časov dekodovania	68

Úvod

Prvopočiatky digitalizácie zvuku sa datujú od 30. rokov 20. storočia, kedy britský vedec Alec Reeves vynášiel spôsob prevodu analógového zvuku do digitálnej formy, známy ako pulzná kódová modulácia (PCM). Za ďalší podstatný mílnik je možné považovať vznik diskkrétnej kosínusovej transformácie (DCT) na začiatku 70. rokov. So začínajúcim rozmachom osobných počítačov a vynálezom kompaktných diskov začiatkom 80. rokov začala vznikať myšlienka znižovania dát v zvukovom zázname. Pôvodne sa prvé koncepty stratovej kompresie zvuku vytvárali ako súčasť štandardov pre kódovanie videa, postupne sa však stratové kódovanie zvuku stalo samostatnou disciplínou zvukového inžinierstva. Globálnym rozšírením internetu a spopularizovaním streamovania dát po internetovej sieti sa začal dôraz na znižovanie objemu dát audio súborov zväčšovať ešte viac. [1]

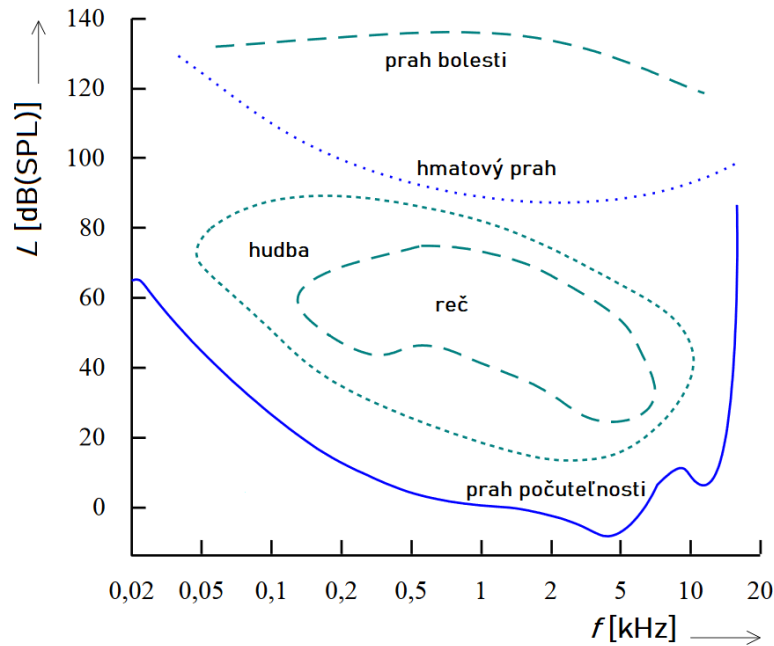
V dnešnej dobe má užívateľ možnosť výberu zo širokej škály audio kodekov využívajúcich stratovú kompresiu dát. V tejto práci bude vytvorený komplexný profil niekoľkých z najznámejších dostupných kodekov, určených na kódovanie hudobných súborov. V prvej kapitole je priblížené, akým spôsobom analyzuje ľudský sluch zvuk, a aké má celý reťazec sluchového orgánu nedostatky. Nasleduje kapitola stručne opisujúca problematiku kódovania audio signálov – sú tu zhrnuté rôzne postupy kódovania spolu s parametrami, na ktorých závisí výsledná kvalita kódovaného audio súboru. Táto kapitola obsahuje tiež metodiku výberu skúmaných kodekov spolu s popisom ich funkcií, a je ukončená stručným zhrnutím rozdielov medzi popisovanými kodekmi. Nasledujúca kapitola sa venuje rozboru rôznych druhov bežne používaných posluchového testov, detailnému rozboru zvoleného typu testu a teoretickému návrhu samotného testu. V ďalšej krátkej kapitole je priblížený postup, ktorým boli získané súbory pre posluchový test. Predposledná kapitola približuje detaily stránky s posluchoвыми testami, pomocou ktorej boli získané dáta pre štatistické vyhodnotenie. Záverečná kapitola je chronologickým opisom štatistickej analýzy, zakončená prezentáciou výsledkov vykonaných posluchového testov a merania výpočtovej náročnosti kodekov.

1 Ľudský sluch

Zvuk ako fyzikálny jav je mechanické vlnenie šíriace sa v látkovom prostredí. Vnímanie zvuku je komplexným procesom, ktorý je potrebné vysvetliť v niekoľkých rovinách – vonkajšie ucho slúži ako prijímač, ktorého primárnou úlohou je nasmerovať zvuk do zvukovodu. Zvukovod vedie tieto mechanické vlny do systému stredného ucha, skladajúceho sa z ušného bubienku a drobných kostičiek, pomocou ktorých sa vibrácie prenášajú na kochleu – špirálovitý orgán naplnený tekutinou. V dolnej časti kochley sa nachádza bazilárna membrána obsahujúca vlásokové bunky, v ktorých dochádza k transformácii mechanickej energie na elektrické impulzy, smerujúce skrz sluchové nervy do mozgu. Na to, aby sme „počuli“, je teda potrebné zvukovú vlnu zachytiť, usmerniť, transformovať, preniesť a analyzovať. [2]

1.1 Hranice počutelnosti ľudského sluchu

Zrejme najzákladnejším limitom sluchu človeka sú hranice počutelnosti popisujúce, v akom frekvenčnom rozsahu dokáže sluchový orgán operovať a koľko energie dokáže prijať a vyhodnotiť bez trvalého poškodenia. Počutelný bude každý zvuk s hladinou akustického tlaku vyššou ako minimálny prah počutelnosti. Ďalším výrazným prahom je tzv. hmatový prah, pri ktorom dochádza ku podráždeniu hmatových teliesok v uchu. Prekročenie tohto prahu avizuje, že prichádzajúci zvuk sa blíži ku tzv. prahu bolesti. Prekročenie prahu bolesti vedie nevyhnutne k bolesti, pri dlhodobom pôsobení stimulu nad touto hranicou môže dôjsť aj k trvalému poškodeniu sluchu. Z grafu 1.1 je možné pozorovať, že ľudský sluch dokáže rozoznať frekvencie v rozsahu 20 Hz–20 kHz, krivka prahu počutelnosti je výrazne frekvenčne závislá. Je dôležité všimnúť si, že zvukový stimul vo frekvenčnom rozmedzí 2–5 kHz potrebuje pre rozoznanie z celého rozsahu najmenšie množstvo energie, a teda je logické, že bude v tejto oblasti ucho najcitlivejšie. Graf 1.1 je potrebné chápať ako zovšeobecňujúci popis, reálne schopnosti ľudského ucha sú individuálne a navyše závislé na veku človeka. [3]



Obr. 1.1: Sluchové pole [3]

1.2 Frekvenčná analýza v ľudskom sluchu

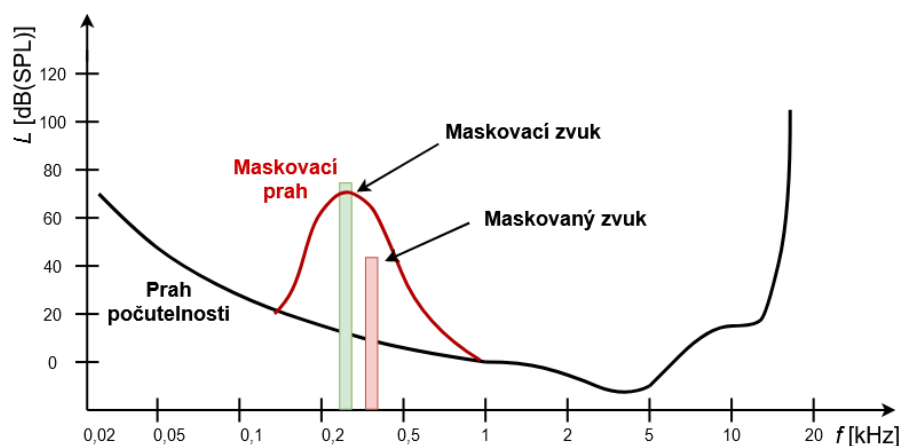
Zvuky ako také je potrebné klasifikovať z hľadiska spektrálneho zloženia. Zjednodušene si môžeme predstaviť jednoduché zvuky o jednej frekvencii (tóny) ako jeden extrém a zvuky skladajúce sa zo všetkých možných frekvencií (šumy) ako opačný extrém. Všetky reálne zvuky môžu byť na základe ich spektrálneho zloženia rozmiestnené niekde na pomedzí tejto škály zvukového charakteru. Bazilárnu membránu kochleárneho slimáka je možné chápať ako sadu za sebou radených pásmových priepustí – vlnenie sa membránou šíri, kým nenarazí na miesto so vzdialenosťou od vstupu membrány odpovedajúcou kmitočtu spracovávanej zvukovej vlny. V tomto bode sa vlna slimákom viac nešíri a jej energia za touto oblasťou strmo klesá. Tieto oblasti dané stredným kmitočtom a šírkou pásma nazývame kritickými pásmami a na celú dĺžku bazilárnej membrány ich pripadá 26. Kritické pásma nemajú konštantnú šírku pásma – šírka pre prvých 5 kritických pásiem je stanovená na 100 Hz, od 500 Hz táto šírka pásma narastá o 20 % v každom ďalšom pásme. Na začiatku, blízko vstupu do bazilárnej membrány, sa detekujú zvuky z hornej hranice frekvenčného rozsahu ľudského sluchu, a čím nižšiu frekvenciu bude zvuková vlna mať, tým hlbšie do tejto membrány prenikne. [4]

1.3 Efekt maskovania

Sluch má okrem obmedzení týkajúcich sa miery prijateľného akustického tlaku a kmitočtového rozsahu aj ďalšie obmedzenia spôsobené činnosťou jednotlivých elementov v sluchovom reťazci. Existujú situácie, pri ktorých informácie vyhodnotenú našim mozgom ako „počuté“ nezodpovedajú reálnym zvukom, ktoré vstupujú do vonkajšieho ucha. Za tento jav môže tzv. maskovací efekt. Príčinu tohto efektu možno hľadať vo fyziologických vlastnostiach stredného ucha. V kapitole 1 je popísané, že spracovávaná zvuková vlna mení energetickú formu vo vlásokových bunkách bazilárnej membrány. Bunky podráždené vplyvom vlniacej sa tekutiny reagujú vždy iba na energeticky najvyšší zvukový podnet, a stimul s rovnakou frekvenciou prichádzajúci súčasne, avšak s nižšou energiou jednoducho ignorujú. Maskujú sa teda najmä signály blízkej frekvencie. [3] Efekt maskovania bol v minulosti objektom skúmania psychoakustických disciplín a dodnes hrá ústrednú úlohu pri audio kódovaní do formátov využívajúcich stratovú kompresiu. Je využitý napr. pre analýzu spektrálnych vlastností signálov v psychoakustických modeloch audio kodekov, ktoré pomocou matematického popisu nedokonalostí ľudského ucha rozlišujú, ktoré informácie je možné z výsledného bitového toku vynechať.

1.3.1 Frekvenčné maskovanie

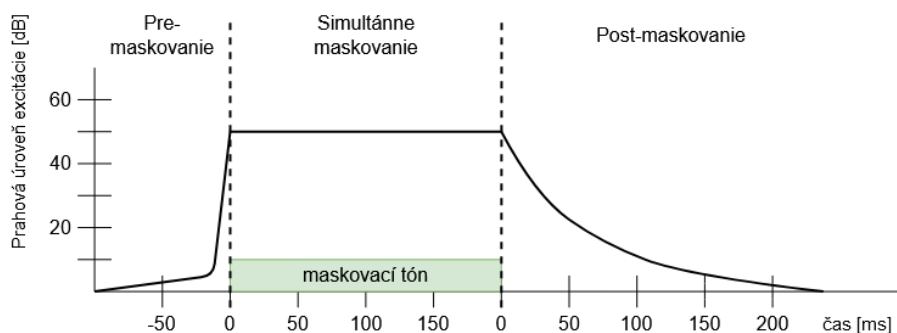
Frekvenčné maskovanie popisuje jav, kedy zo zhľuku zvukov približne rovnakej frekvencie počujeme iba ten, ktorý má najvyššiu hladinu akustického tlaku. V praxi to znamená, že každý zvuk presahujúci absolútny prah počuteľnosti, vytvára svoj vlastný, prídavný maskovací prah, ktorého frekvenčné rozpätie a maskovacia sila závisí predovšetkým od hladiny jeho akustického tlaku. Zvyšné zvuky nachádzajúce sa v blízkosti tohto maskovacieho zvuku budú za predpokladu, že je ich hladina akustického tlaku pod hladinou maskovacieho prahu maskovacieho signálu, nepočuteľné. Vo frekvenčnom maskovaní rozlišujeme tri základné typy interakcií zvukov na základe ich spektrálneho zloženia. Sú nimi tón maskujúci šum, šum maskujúci tón a šum maskujúci hľuk. Vo všeobecnosti platí, že šumové maskovače majú väčšiu maskovaciu silu ako maskovače tonálne. Ďalším dôležitým hľadiskom pre maskovaciu silu signálu je jeho umiestnenie v rámci kritického pásma – čím bližšie bude signál ležať ku stredu tohto pásma, tým väčší maskovací vplyv bude mať. [4]



Obr. 1.2: Príklad frekvenčného maskovania [3]

1.3.2 Temporálne maskovanie

Temporálne (prípadne časové maskovanie) je ďalším dôležitým dôsledkom nedokonalosti ľudského sluchu. Maskovanie totiž neprebíha výhradne len vo frekvenčnej doméne, ale ovplyvňuje aj vnímanie zvukov v čase. Graf 1.3 zobrazuje ilustratívny priebeh nábežnej a zostupnej hrany maskovacieho tónu. Z tohto grafu je evidentné, že maskovanie v čase neprebíha iba po, ale aj niekoľko milisekúnd pred zaznením zvuku. Temporálne maskovanie sa šíri naprieč kritickými pásmami a je popisované trojuholníkovou funkciou, ktorej nábežná hrana má strmosť približne -25 dB a hrana zostupná -10 dB. Trvanie *pre-maskovania* trvá len niekoľko jednotiek milisekúnd, zatiaľ čo *post-maskovanie* môže trvať rádovo stovky milisekúnd. Časové rozpätie závisí samozrejme od hladiny akustického tlaku maskovacieho signálu. [1]



Obr. 1.3: Príklad temporálneho maskovania [4]

2 Kódovanie audio signálov

2.1 Bitová rýchlosť

Bitová rýchlosť (bitrate) popisuje množstvo dátových bitov obsiahnutých v audio súbore za jednotku času. Jej nastavovaním pred kódovaním diktujeme kodeku, koľko bitov môže počas kódovacieho procesu použiť na kódovanie v jednej sekunde. Základnou jednotkou bitovej rýchlosti je bps (bits per second), avšak v praxi sa ďaleko častejšie stretávame s jednotkou kbps (kilobits per second). [5]

$$\textit{bitová rýchlosť} = \textit{vzorkovacia frekvencia} \cdot \textit{počet bitov na vzorku} \cdot \textit{počet kanálov} \quad (2.1)$$

Napríklad zvukové súbory „surového“ zvukového formátu WAV v CD kvalite majú bitovú rýchlosť 1411 kbps ($44100 \times 16 \times 2$). Nakoľko tento parameter priamo súvisí s reguláciou množstva dát vo výstupnom súbore, je jeho nastavenie záležitosťou voľby užívateľa. Rozlišujeme dva základné typy bitovej rýchlosti:

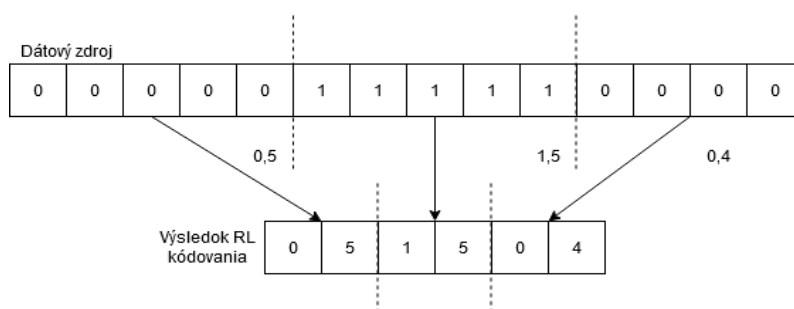
- Konštantná bitová rýchlosť – nastavená fixne pre každú sekundu záznamu
- Variabilná bitová rýchlosť – fluktuujúca priamo úmerne k spektrálnej zložitosti signálov. [5]

2.2 Bezstratová kompresia

Metódy audio kódovania využívajúce bezstratovú kompresiu sú založené na efektívnejšom uložení dát oproti pôvodnému súboru. Ako už z názvu vyplýva, pri týchto metódach nedochádza k strate pôvodných informácií, a fakt, že je vždy možné presne zrekonštruovať pôvodný signál je najpodstatnejšou výhodou tohto prístupu kódovania. Ako ďalšia výhoda môže byť považovaná nižšia výpočtová náročnosť, nakoľko bezstratová kompresia nevyžaduje prítomnosť psychoakustického modelu a blokov zaoberajúcich sa výpočtom vhodnej úpravy spektra výsledného signálu. Na druhej strane toto samotné zjednodušenie zápisu kódovaných informácií nie je ako nástroj ku dostatočnému zníženiu veľkosti dát postačujúci, a kompresné pomery bezstratovej kompresie sú v globálnom merítku vzhľadom na hlavnú motiváciu kompresie audio dát menej efektívne ako postupy stratovej kompresie. Napriek tomu majú formáty bezstratovej kompresie širokú základňu priaznivcov, najmä z rád audiofilov a vlastníkov Hi-Fi systémov. K najznámejším reprezentantom audio formátov využívajúcich bezstratovú kompresiu patria FLAC, AIFF a ALAC. [6]

2.2.1 Run-length kódovanie

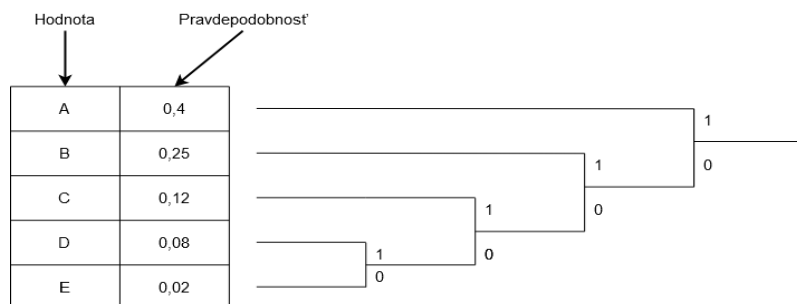
RLK je pravdepodobne najjednoduchším druhom dátovej kompresie. RLK hľadá intervaly znakov rovnakej hodnoty, ktoré následne nekóduje zvlášť, ale zakóduje iba dĺžku intervalu a hodnotu znaku. [7] Jednoduchý príklad princípu tohto druhu kódovania môžete vidieť na obrázku 2.1. Je logické, že efektivita tohto druhu kódovania je výrazne závislá na štruktúre vstupného signálu – za predpokladu, že budú vstupné hodnoty v slede rôznorodé, je účinnosť RLK minimálna. Tento typ entropického kódovania sa používa napr. v MP3.



Obr. 2.1: Príklad Run-Length kódovania [7]

2.2.2 Huffmanovo kódovanie

Huffmanovo kódovanie je založené na entropickom kódovaní a taktiež sa jedná o techniku bezstratovej kompresie dát. Dáta (množina znakov obsiahnutých vo vstupnom slove) sa najskôr štatisticky vyhodnotia a následne zoradia podľa pravdepodobnosti výskytu vo vstupnom dátovom toku. Pre toto rozvrhnutie sa používajú Huffmanove kódové tabuľky. Následne sa vytvorí binárny strom, v ktorom sa dáta zoradia tak, aby bol na najfrekvencovanejšie informácie použitý čo najmenší počet bitov. Množstvo bitov použitých na zakódovanie dát v tomto spôsobe kódovania teda závisí na pravdepodobnosti výskytu danej informácie – čím sporadickejší výskyt je zistený, tým väčší bude počet bitov pre výsledné zakódovanie znaku. [8] Ilustračný príklad Huffmanovho kódu je uvedený na obrázku 2.2. Výsledkom Huffmanovho kódovania sú tzv. prefixové kódy – žiadny symbol kódovaného slova nie je predponou iného symbolu. Huffmanovo kódovanie využíva v závere kódovacieho procesu väčšina audio kodekov.



Obr. 2.2: Jednoduchý príklad tvorby binárneho stromu pri Huffmanovom kódovaní [8]

2.3 Stratová kompresia

Stratová kompresia zvuku vychádza z predpokladu, že originálny zvukový záznam obsahuje informácie, ktoré je možné z bitového toku odstrániť, a tým znížiť veľkosť dát výsledného súboru. Princiipiálne sú odstránené dáta rozdelené na nadbytočné a zbytočné. Za nadbytočné je možné považovať napr. stereofónny záznam, ktorý obsahuje v oboch kanáloch rovnaký signál, ako zbytočné signály si môžeme predstaviť informácie, ktoré nedokážeme, napr. kvôli maskovaciemu efektu, sluchom zachytiť. Audio kodeky využívajúce stratovú kompresiu teda vždy obsahujú percepčné (psychoakustické) modely, ktoré majú za úlohu matematicky popísať vlastnosti ľudského sluchu, a tým riadiť úpravu výstupného spektra tak, aby sa odstránilo čo najviac nepostrehnutelných informácií. Zároveň však musia tieto modely klásť dostatočný dôraz na to, aby signály aj po kompresii pôsobili čo najviac vierohodne vzhľadom k originálnemu záznamu. Stratová kompresia práve vďaka tejto metodike znižovania dát dokáže znížiť veľkosť súboru aj o 80 %. Treba však mať na pamäti, že po zakódovaní sú odstránené dáta navždy stratené a spätná rekonštrukcia skutočného vstupného signálu už nebude možná. Množstvo dát komprimovaného súboru je teda v metódach stratovej kompresie akýmsi kompromisom medzi výslednou posluchovou kvalitou a veľkosťou zakódovaného súboru. [9]

2.3.1 Psychoakustický model

Psychoakustický (niekedy tiež percepčný) model je komponent, využívaný v audio kodekoch so stratovou kompresiou pre matematický popis vlastností ľudského sluchu. Napriek tomu, že sa psychoakustické modely naprieč kodekmi vzájomne líšia, a postupy, ktorými napomáhajú kódovaniu závisia na samotnej implementácii, každý z nich má v podstate niekoľko základných úloh. Pracujú vo frekvenčnej oblasti, ich účelom je predovšetkým operovanie s maskovacími prahmi v menších frekvenčných

blokoch, ktorým hovoríme subpásma. Pre každé subpásma stanovujú psychoakustické modely:

- prah počutelnosti pri absencii zvukového signálu
- celkový maskovací prah
- minimálny maskovací prah (počítaný z celkového maskovacieho prahu)
- pomer maskovacieho signálu k minimálnemu maskovaciemu prahu (pre efektívnejšie riadenie kvantovania).

S vyššie spomenutým súvisí aj klasifikácia spektrálneho zloženia subpásim, kde sa rozlišuje harmonický (tonálny) a šumový (netonálny) charakter. Psychoakustický model taktiež vykonáva decimáciu maskovacích tónov pre lepší odhad maskovacej sily jednotlivých spektrálnych koeficientov v subpásmach. [10]

2.3.2 Joint stereo kódovanie

Ďalšie riešenie pre ušetrenie dát zakódovaného súboru môže byť režim „joint stereo“ kódovania – teda efektívnejšie zoskupovanie a ukladanie dát z jednotlivých kanálov. Takmer všetky hudobné súbory, ktoré máme k dispozícii sú nahraté a vyprodukované stereofónne. V dôsledku konvenčných postupov hudobnej produkcie sa však v každom zvukom súbore bude nachádzať určitý obsah viac-menej v strede tohto priestoru, čo znamená, že bude existovať istá zhoda medzi ľavým a pravým kanálom. To je presne cieľ spájania sterea, ktorého poznáme niekoľko druhov. Spájanie sterea našlo využitie v mnohých, nižšie popisovaných, stratových audio kodekoch. [11]

Stredné/ stranové stereo

Často tiež označované ako M/ S stereo (Middle/ Side). V tomto kódovaní sa niektoré dáta zo vstupných kanálov kódujú do stredného a stranového kanálu. V skratke – stredný kanál obsahuje súčet a stranový kanál rozdiel dvoch vstupných kanálov. Jeho účinnosť je priamo závislá na type vstupného signálu (stúpa s vyššou podobnosťou jednotlivých kanálov), zároveň však môže byť použitý na celý frekvenčný rozsah skladby bez toho, aby zaviedol rušivé zvukové artefakty. [11]

Intenzitné stereo

V intenzitnom stereu sa využíva fakt, že identifikácia smerovosti zvuku ľudským sluchom je obmedzená (resp. s absolútnou presnosťou dokážeme zistiť smer zvu-

kového zdroja iba v určitom frekvenčnom rozmedzí). Zvuky s príliš vysokými, či nízkymi frekvenciami, ktoré sa nachádzajú mimo tohto frekvenčného pásma, môžu byť teoreticky zakódované do jedného kanálu. Intenzitné stereo je zaujímavé kvôli mechanizmu kompresie dát z niekoľkých kanálov do jedného, bohužiaľ je vhodné čisto pre vysokofrekvenčné zložky signálov (na nízkych frekvenciách zavádza zvukové artefakty). Názov odkazuje na spôsob lokalizácie zvukov ľudským sluchom – pri vychýlení zdroja zvuku z mediálnej roviny hlavy musí zvuková vlna prekonať k oboj ušiam rozdielne vzdialenosti, lokalizácia zvuku závisí od rozdielu vnímaných intenzít každým uchom. [11]

2.4 Knižnica FFmpeg

FFmpeg je knižnica obsahujúca open-source zdrojové kódy audio kodekov a taktiež zahŕňa nástroje pre úpravu video a zvukových súborov. V tejto práci poslúži FFmpeg (resp. knižnica pyffmpeg v jazyku Python volajúca FFmpeg písanú v jazyku C) ako nástroj pre kódovanie do zvolených stratových formátov. Knižnica nám umožňuje presné zadefinovanie parametrov, ktoré požadujeme pri kódovaní zvukových súborov do požadovaných formátov. Práca s FFmpeg taktiež uľahčí meranie výpočtovej náročnosti audio kodekov, nakoľko budú všetky kodeky optimalizované cez jeden program. Pre prípadnú podrobnú analýzu ponúka knižnica tiež zdrojové kódy všetkých kodekov. [12]

2.5 Metodika výberu audiokodekov

Tab. 2.1: Kritériá a konečný výber skúmaných kodekov

	AC3	WMA	Vorbis	Opus	SBC	Asao	MP3	MPC	AAC
A	1	0	1	1	1	0	1	0	1
B	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	1	1	1

A–Dostupnosť teoretických špecifikácií

B–Dostupnosť zdrojového open-source kódu

C–Možnosť merania výpočtovej náročnosti

D–Možnosť použitia v posluchovom teste

Po zvolení knižnice ffmpeg ako primárneho nástroja pre získanie kódovaných audio súborov boli pomocou oficiálneho zoznamu knižnice zhromaždené audio kodeky vyhovujúce účelom práce (bežne najviac využívané kodeky pre kódovanie hudby). Následne bolo potrebné zhromaždiť teoretické informácie popisujúce špecifikácie a samotný postup kódovania jednotlivých kodekov – tento parameter sa nakoniec stal aj najväčším filtrom výberu. Kritériá a konečný výber je zobrazený v tabuľke 2.1.

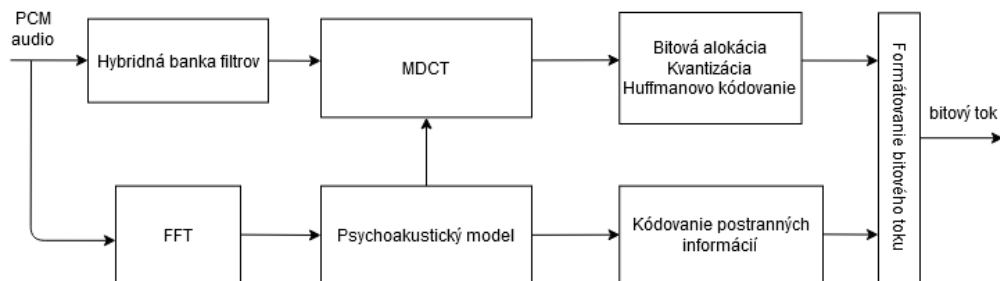
Koncept posluchových testov pripúšťal testovanie maximálne štyroch kodekov. AC3 bol vyradený, nakoľko je využívaný predovšetkým na kódovanie šesť kanálového, priestorového zvuku (nevhodný pre stereofónne hudobné súbory). Keďže zámerom práce bola analýza najznámejších kodekov pre kódovanie hudby, musel rovnako tak menej známy kodek SBC ustúpiť výslednej, notoricky známej a často využíwanej štvorici.

2.6 Postupy kódovania vo vybraných audio kodekoch

Nasledujúca kapitola sa venuje teoretickému rozboru zvolených audio kodekov pomocou popisu funkcií ich blokových diagramov.

2.6.1 MP3

MPEG-1 vrstva 3 (skrátene MP3) je stratový kodek vyvinutý spoločnosťou Moving Picture Expert Group. Prvá verzia bola predstavená v roku 1993. MP3 umožňuje voľbu medzi variabilnou a konštantnou bitovou rýchlosťou v rozmedzí 32–320 kbps a podporuje vzorkovacie frekvencie 32, 44,1 a 48 kHz.



Obr. 2.3: Bloková schéma MP3 kódera [4]

Hybridná banka filtrov

Počiatočný blok, hybridná banka filtrov, prijíma PCM audiodáta a delí ich do dátových blokov, ktorým hovoríme rámce. Dátový rámec obsahuje 1152 navzorkovaných hodnôt, ktoré následne rovnomerne rozdelí na dve kratšie časti, nazývané granule (576 vzoriek). Najdôležitejším úkonom, ktorý táto banka filtrov vykoná je rovnomerné rozdelenie spektra vstupného signálu každého rámca na 32 pásiem zhodnej šírky. Všetky následné výpočty, kvantizácia, a kódovanie sa dejú pre každé toto pásmo individuálne. Jednotlivé filtre obsiahnuté v banke filtrov nemajú dostatočne veľkú strmosť, čo sa prejaví únikom zložiek s frekvenciami blízskymi okraju filtrov do susedného pásma, čím filter zavadza aliasing. Taktiež je dôležité si uvedomiť, že narozdiel od kritických pásiem sluchu majú tieto pásma rovnakú šírku – nízke pásma banky filtrov pokrývajú viac kritických pásiem ľudského ucha. Výstupom tohto bloku sú vzorky usporiadané v spomínaných pásmach. [4]

MDCT

Z týchto dôvodov je priamo za banku filtrov zaradený blok vykonávajúci v jednotlivých pásmach modifikovanú diskretnú kosínusovú transformáciu. Pomocou MDCT sa každé pásmo rozdelí na ďalších 18 jemnejšie rozlíšiteľných subpásiem (celkovo teda už 576). Nemenej dôležitou funkciou MDCT je okienkovanie vstupných dát, umožňujúce meniť dĺžku spracovávanej časti dlhými oknami (1024 vzoriek) a oknami krátkymi (256 vzoriek). Prepínanie týchto okien je riadené psychoakustickým modelom, ktorý sa pomocou prepínania okien snaží predchádzať vzniku zvukových artefaktov. MDCT dokáže taktiež pomocou ortonormálnej transformácie čiastočne znížiť spomínaný aliasing zavedený bankou filtrov. Výstupom MDCT sú spektrálne koeficienty. [4]

Psychoakustický model

Psychoakustický model je možné považovať za hlavný analytický člen celého reťazca, vykonávajúci množstvo operácií potrebných pre prispôsobenie spektra výstupného dátového toku. Prijíma frekvenčné koeficienty, ktoré mu na vstup privádza FFT analýza vstupného signálu. Jeden analytický blok vyhodnocuje 1024 vzoriek (to je dôvod, prečo delí banka filtrov rámec na dve granule, dĺžka analytického okna FFT by nedokázala prekryť vstupný rámec s 1152 vzorkami).

Najzásadnejšou úlohou psychoakustického modelu je výpočet maskovacích prahov pre jednotlivé subpásma vstupného signálu. Pred samotným výpočtom maskovacích prahov je nevyhnutné rozdeliť spektrálne komponenty v subpásmach na tonálne a netonálne, kvôli maskovacej účinnosti. Počíta sa tzv. index tonality, čo je miera pravdepodobnosti opätovného výskytu spektrálneho koeficientu, ktorá určuje, či má komponent charakter tónu alebo šumu. V globále platí, že šumový signál je vysoko nepredvídateľný a tón je naopak vysoko predvídateľný. Po zistení povahy spektra sa pomocou indexu tonality začína výpočet samotných maskovacích prahov. Tieto energetické odhady zahŕňajú aj vplyv temporálneho maskovania, ktorému sa v názvosloví MP3 hovorí funkcia šírenia. Psychoakustický model priamo riadi prepínanie okien v MDCT bloku.

Najčastejším javom, ktorému sa pomocou prepínania okien snaží psychoakustický model vyhnúť sú tzv. pre-echo ozveny. Zvukové artefakty v podobe týchto ozvien nastávajú, keď sa signál s veľmi rýchlym nábehom objaví v blízkosti hranice transformačného bloku bezprostredne po oblasti s veľmi nízkou energiou (tiché pasáže). Prepnutím okna sa docíli lepšie časové rozlíšenie v mieste ostrého tranzientného javu – ozveny sa prejavia v kratšom úseku a nebudú sa šíriť do pasáží, kde má signál nízku energetickú úroveň. Posledným dôležitým krokom je odhad pomeru maskovacieho signálu k minimálnemu maskovaciemu prahu. Úroveň maskovacích prahov sa prevedie do pásiem mierkového faktoru (skupín frekvenčných čiar, ktoré približujú kritické pásma). Vďaka týmto mierkovým pásmam je následne možné odhadnúť hranicu, pod ktorú je možné rozložiť kvantizačný šum. [4]

Alokácia bitov

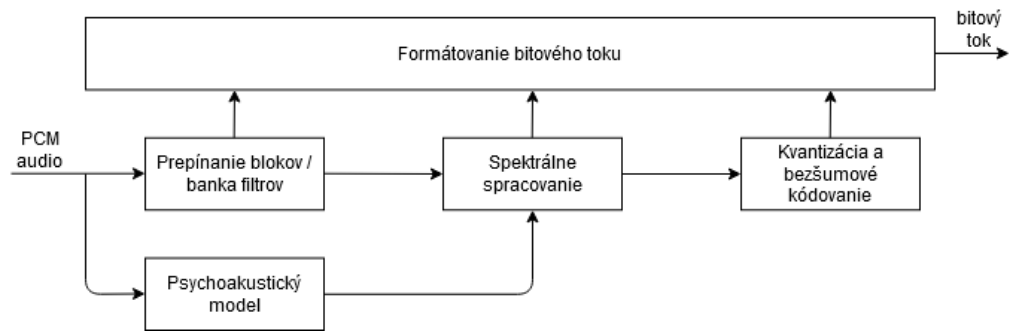
Po predošlých operáciách sa kódovací proces presúva k alokácii bitov. Tá vychádza z dát pochádzajúcich z psychoakustického modelu a jej úlohou je prideliť pre každé subpásma počet bitov potrebných pre potlačenie kvantizačného šumu pod minimálnu hranicu maskovania. Mechanizmus bitovej alokácie je realizovaný pomocou logického člena pozostávajúceho z dvoch slučiek. Vnútoraná slučka upravuje veľkosť kvantizačného kroku dovtedy, kým počet bitov neklesne na dostupnú hodnotu prístupných bitov v bitovej nádrži. Vonkajšia slučka kontroluje výsledné skreslenie v pásmach mierkových faktorov, a vždy, keď zistí nevyhovujúcu mieru skreslenia, zväčší počet pásiem mierkového faktora a opätovne vyzve k činnosti vnútornú slučku. Pre ušetrenie bitov sa konečný výber pásiem mierkového faktora uloží a následne sa zisťuje, či je (z dôvodu ušetrenia dát) možné použiť totožný výber aj pre nasledujúce granule. [4]

Huffmanovo kódovanie

Posledným krokom nasledujúcim po kvantizácii je efektívne zakódovanie jednotlivých pásiem mierkových faktorov pomocou Huffmanovho kódovania. Počet bitov pre jednotlivé nakvantované hodnoty sa pridelí na základe pravdepodobnosti ich výskytu. V tomto algoritme sa kódujú spárované kvantované hodnoty do 32 Huffmanových tabuliek. V MP3 sa frekvenčné čiary zoraďujú do troch kategórií na základe ich amplitúdy. Pre každú túto kategóriu sa potom vymedzia vlastné Huffmanove tabuľky, v prípade poslednej oblasti, pozostávajúcej čisto z núl sa používa Run-length kódovanie, čo ešte väčšmi zefektívni výslednú kompresiu dát. [4]

2.6.2 AAC

Advanced Audio Coding alebo v preklade pokročilé kódovanie zvuku je zvukový formát vyvíjaný od roku 1997, ako priamy nástupca dodnes najrozšírenejšieho strátového formátu MP3. Kodek AAC je na rozdiel od MP3 podstatne univerzálnejší, výkonnejší, má viac možností konfigurácie a všeobecne dosahuje vyššiu kvalitu aj pri nižších bitových rýchlostiach (tvorcovia sa inšpirovali rečovými kodekmi, ktoré kódujú prioritne na čo najnižších bitových rýchlostiach). AAC podporuje 48 kanálov, vzorkovacie frekvencie medzi 8–96 kHz a bitové rýchlosti až po 576 kbps na kanál. Najznámejšou modifikáciou je MPEG-4 (High Efficiency-AAC). Vlastnosti algoritmov AAC sa delia na tri tzv. profily. Hlavný profil pre dosahovanie najvyššej kvality obsahuje oproti iným profilom blok pre riadenie zosilnenia a jeho filtre pre tvarovanie šumu majú vyšší rád. Ďalej to je AAC LC (Low Complexity), neobsahujúci niektoré funkcie hlavného profilu, ako napríklad ovládanie zosilnenia či spektrálnu predikciu koeficientov medzi rámcami, no vďaka nízkej výpočtovej náročnosti postačuje požiadavkám bežných užívateľov – jedná sa o najrozšírenejší AAC profil s ešte vždy veľmi uspokojivou zvukovou kvalitou (tento profil bude podrobený skúmaniu v posluchových testoch). Posledným profilom je AAC SSR (Scalable Sampling Rate), ktorý, ako z názvu vyplýva, poskytuje možnosť frekvenčne škálovateľného vzorkovacieho kmitočtu signálu. [13]



Obr. 2.4: Bloková schéma AAC kódera [13]

Prepínanie blokov/ banka filtrov

Časovo-frekvenčný prevod v AAC zabezpečuje opäť MDCT. MDCT môže na základe charakteru vstupného signálu použiť okno sínusového tvaru alebo okno s tvarom KBD (Kaiser-Bessel Derived). Možnosťou prepínať tvar týchto okien dokáže kodek odhadnúť maskovacie prahy z menšieho počtu spektrálnych koeficientov. Sínusové okno je vhodnejšie pre signály s hustým spektrom, pretože poskytuje užšiu pásmovú selektivitu, čím poskytuje lepšie frekvenčné rozlíšenie. Naopak okno KBD sa používa v prípadoch, keď majú výrazné frekvenčné koeficienty väčšie rozostupy – maskovací prah bude počítaný len so skutočne podstatných zložiek, a tým sa ušetrí bity pre kódovanie. [1] Prekrytie okien je štandardne 50 % – každé okno obsahuje polovicu vzorkov z okna predošlého. Prepínanie blokov riadi psychoakustický model. Dĺžky transformačných okien v AAC sa delia na dlhé (2048 vzoriek) a krátke (256 vzoriek). Podobne ako v MP3 existujú ešte dve prechodné okná, nachádzajúce sa vždy medzi krátkym a dlhým oknom. [13]

Psychoakustický model

Psychoakustický model generuje pomery SMR pre každé subpásmo oneskorených blokov PCM vzoriek, ktoré sú spracovávané v MDCT (aby sa MDCT mohlo včasne prepnúť na inú dĺžku okna). PM vykonáva tiež samotné príkazy pre prepnutie dĺžky okien v MDCT a približný odhad bitov pre dostatočne nízky kvantovací šum v každom subpásme. [13]

Spektrálne spracovanie

TNS

Temporal Noise Shaping, alebo dočasné tvarovanie šumu je prídavný nástroj AAC kóderov, určený pre zlepšenie potláčania zvukových artefaktov a zvýšenie pomeru užitočného signálu k šumu. Príchod tohto konceptu súvisí s faktom, že samotné prepnutie okna v MDCT nemusí byť postačujúce pre elimináciu pre-echo efektov – príliš dynamické signály budú rôznorodé aj v krátkom MDCT okne. Hlavnou úlohou TNS je frekvenčná predikcia obsahu subpásmiem, a tvarovanie kvantizačného šumu tak, aby ostal v rámci celého subpásma pod maskovacím prahom prítomných signálov. TNS navyše prispieva k lepšiemu časovému rozlíšeniu v jednotlivých subpásmach. TNS blok prakticky pozostáva z filtra tvarujúceho šum, obsahujúceho tiež modifikáciu v podobe funkcie lineárne prediktívneho kódovania. Ďalšie vlastnosti tohto filtra závisia priamo od profilu, v ktorom AAC pracuje – vyšší rád filtra je možné očakávať napr. v hlavnom AAC profile, ktorý je prispôsobený na vyššie požiadavky užívateľa. [14]

Spájanie kanálov

Ďalším krokom pri spektrálnom spracovaní je joint stereo kódovanie, ktoré pomáha znižovať výsledný bitový tok, a uplatnenie našlo predovšetkým pri nižších bitových rýchlostiach. AAC má možnosť použiť režim intenzitného a M/S stereo kódovania. Obidva tieto režimy môžu pracovať simultánne a každý z nich môže byť aplikovaný na inú časť spektra zvukového signálu.

M/S stereo kódovanie sa využíva na signály z príslušnej dvojice kanálov (napr. L a R kanál), najmä v oblasti nižších frekvencií. Možnosť kódovania v tomto režime je možné vypnúť pre každý rámec a dokonca aj subpásma osobitne. Zapnutie či vypnutie M/S kódovania sa zaznamenáva v jednom bite, aby dekóder dokázal správne rekonštruovať dáta v bitovom toku.

Intenzitné stereo kódovanie funguje na princípe zdieľania energeticko-časových obálok vysokofrekvenčných signálov medzi dvojicou zvukových kanálov. V AAC pozostáva toto kódovanie z dvoch mechanizmov:

- Intenzitné stereofónne kódovanie – energie dvoch vysokofrekvenčných párových koeficientov v L a R kanáli sú nahradené súčtom ich energií v jednom kanáli
- AAC element pre spájanie kanálu – rozširujúci mechanizmus umožňujúci súčet energií aj medzi kanálovými párami, ktorý môže v prípade viackanálového

zvuku spojiť energie párových koeficientov z akýchkoľvek dvoch kanálov. Navyše je pomocou neho možné vykonanie downmixu viacerých zvukových kanálov do stereu. [14]

Kvantizácia

Kvantizačný proces sa riadi na základe požiadaviek psychoakustického modelu. Rozsah možností bitovej alokácie je samozrejme podmienený požadovanou bitovou rýchlosťou. AAC využíva niekoľko postupov pri bitovej alokácii, najčastejšie však osvedčenú stratégiu dvoch iteračných slučiek. V AAC sa na kvantizáciu pásiem mierkového faktora používajú kvantizátory s nejednotným krokom kvantizácie (je možné ho meniť po 1,5 dB). Toto vedie k lepšiemu zohľadneniu rôznorodosti amplitúd frekvenčných koeficientov a zvýšeniu pomeru užitočného signálu k šumu v pásmach mierkových faktorov. Tvarovanie šumu tu môže prebiehať aj pomocou funkcie zosilnenia pásiem mierkových faktorov. [14]

Bezšumové kódovanie

Vstup do tohto bloku tvorí 1024 spektrálnych koeficientov. Tento blok má možnosť vykonať nasledovné funkcie:

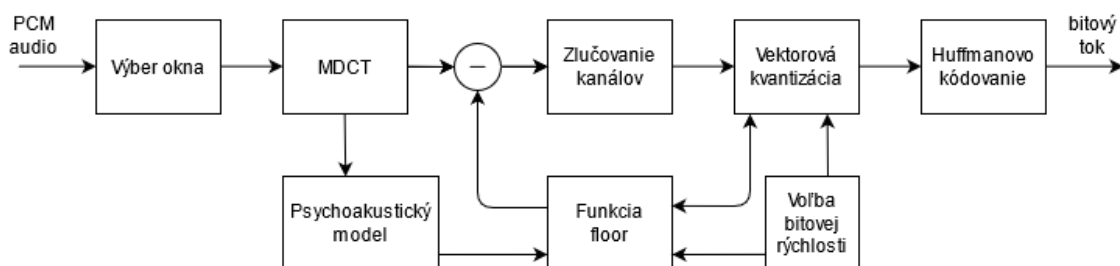
- bezšumová kompresia dynamického rozsahu – jedná sa o vyjadrenie niektorých spektrálnych koeficientov ako rozdiel amplitúdy a pozície iného, susedného koeficientu. Vzhľadom k tomu, že je použitie tohto procesu potrebné následne zakódovať do vedľajších informácií daného rámca, si kóder vždy overí, či bude použitie tohto kompresného prístupu viesť k efektívnejšiemu kódovaniu.
- segmentovanie – v tomto kroku sa zoskupí niekoľko susedných pásiem mierkového faktora do jedného segmentu, aby mohli zdieľať jednu, spoločnú Huffmanovu tabuľku
- zoskupovanie a vkladanie – do súvislých oblastí sa zoskupujú jednotlivé spektrálne koeficienty z krátkych okien. Zníži sa počet bitov na ich zakódovanie a zjednoduší sa réžia kódovania. [14]

Huffmanovo kódovanie

Pre bezstratové uloženie nakvantovaných dát je v AAC k dispozícii 12 Huffmanových tabuliek. Ukladanie do nich prebieha podobne ako v MP3. [14]

2.6.3 Vorbis

Vorbis je stratový zvukový formát od spoločnosti Xiph.org Foundation – počiatky vývoja prototypu nového stratového formátu v tejto organizácii siahajú do roku 1993. Intenzívnejší vývoj však odštartoval až v roku 1998, kedy Fraunhofer Society začala avizovať spoplatnenie vtedy najrozsiahljšieho stratového formátu MP3. Ambíciou Vorbisu, ako novo prichádzajúceho slobodného formátu, bolo tento formát úplne nahradiť. Najznámejším audio kodekom je libvorbis, ktorého prvá verzia bola predstavená v roku 2002. Jeho dátovým formátom je OGG. [15] Kodek Vorbis využíva pokročilejšie kompresné postupy ako jeho predchodcovia (zahŕňa napr. krivky hrubého odhadu spektrálneho zloženia v rámcoch). Podporuje vzorkovacie frekvencie v rozpätí 8–192 kHz s rozlíšením vzoriek 16–24 bitov. Pre kódovanie poskytuje 255 nezávislých kanálov, spolu s možnosťou zoskupovať ich v neskorších fázach kódovania do tzv. podmáp. Všeobecne oproti MP3 dosahuje Vorbis vyšší kompresný pomer s kvalitnejším zvukovým výsledkom. [16]



Obr. 2.5: Blokovaná schéma Vorbis kódera [17]

Výber okna a MDCT

Dĺžka MDCT okna vo Vorbise môže nadobúdať veľkosť 64–8192 vzoriek. Štandardné prekrytie okien pri oknách jednotnej dĺžky je 50 %. Prechod medzi oknami rôznych dĺžok prebieha štandardne pomocou prechodných okien. Hrany týchto okien sú prispôbené tak, aby tvorili 50 % prekrytie s oboma susednými MDCT oknami. [18]

Psychoakustický model

Psychoakustický model má štandardnú úlohu. Z údajov, ktoré prijíma z bloku MDCT poodstraňuje koeficienty, ktoré nemôžu byť ľudským uchom počuté ani pred kódovaním. Výstup z psychoakustického modelu smeruje do bloku, kde sa upravuje pomocou funkcie „floor“. [18]

Funkcia „floor“

Vstupné dáta tohto bloku tvoria psychoakustickým modelom vyselektované spektrálne koeficienty z MDCT. Úlohou tohto bloku je generovať krivky vykonávajúce hrubú aproximáciu spektrálnej obálky vstupného signálu. Vorbis rozlišuje krivky typu „floor 0“ a „floor 1“. Úlohou tejto krivky je aproximácia spektrálnej obálky rámca naprieč frekvenčnými pásmami. Jej výpočet prebieha nasledovne: Prvotná krivka má tvar jednej úsečky medzi počiatočným a koncovým bodom spektrálnej obálky rámca. Následne je táto úsečka rozdelená ďalším bodom v jej strede, ktorý sa priblíži reálnej spektrálnej obálke zvukového signálu. Obidve lineárne oblasti sú opäť symetricky rozdelené a body predelu sa priblížia spektrálnej obálke rámca. Dĺžka procesu delenia novovzniknutých úsečiek (presnosť výslednej krivky) je podmienená bitovou rýchlosťou. Dáta vygenerované funkciou floor sa ďalej kvantizujú pomocou vektorovej kvantizácie. Zároveň sa však odosielaajú do logického člena, ktorý realizuje rozdiel tohto hrubého spektrálneho odhadu od spektrálnej obálky z MDCT bloku. Vzniknú takzvané „zostatky“ – zbytky spektrálnych koeficientov vytvorené odčítaním krivky vzniknutej z floor funkcie od pôvodného spektra rámca z MDCT bloku. Tieto zostatky sú vo výslednom dátovom toku obsiahnuté ako rozdielové vektory. [18]

Zlučovanie kanálov

Tieto zvyšky sú neskôr kódované ako samostatné vektory do jedného kanálu (predpokladajme stereo vstupný signál) v bloku pre zlučovanie kanálov. Existujú tri zlučovacie metódy:

- Lossless stereo – koeficienty sa ukladajú do polárnych súradníc, vo výsledku obsahujú informáciu o amplitúde aj fáze
- Phase stereo – v tomto prípade hrá rolu predpoklad, že ľudský sluch nedokáže rozoznávať malé fázové rozdiely signálov do 4 kHz. Fázy koeficientov sa teda decimujú na 2 alebo 4 možné uhly
- Point stereo – ukladá iba amplitúdu, informácia o fáze nie je zachovaná. [19]

Vektorová kvantizácia

Vstupnými dátami pre vektorovú kvantizáciu sú vzorky z floor funkcie a zvyšky zredukované spájaním kanálov. Všetky tieto dáta sú tu reprezentované ako body

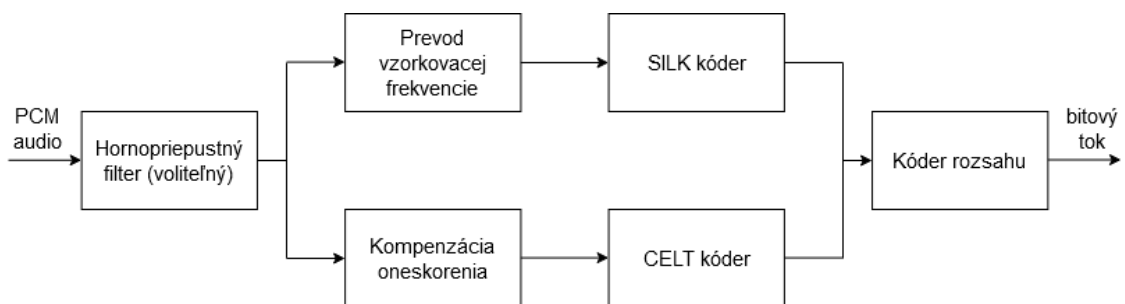
usporiadané v štvorcovej sieti. Na základe rozloženia týchto bodov sa postupne vytvárajú menšie podskupiny, ktoré združujú oblasti husto zoskupených bodov. Každá takáto oblasť ma svoje vlastné ťažisko, pomocou ktorého je možné rozostup týchto bodov a oblastí vyjadriť vektormi. Princípom tohto druhu kódovania je, že počet vektorov potrebný pre vyjadrenie týchto oblastí je nižší, ako pri kódovaní vektorov všetkých bodov jednotlivo. [18]

Huffmanovo kódovanie

Dáta z vektorovej kvantizácie sú v závere kódovacieho procesu zákodované do Huffmanových tabuliek, čím dôjde k finálnej bezstratovej kompresii dát. [18]

2.6.4 Opus

OPUS vyvíjaný spoločnosťou Xiph.org je hybridný audio kodek vhodný pre kódovanie reči aj hudobných súborov predovšetkým v reálnom čase. Prvýkrát bol predstavený v roku 2012 a vďaka svojej univerzálnosti, nízkej algoritmickej latencii a skvelým výkonom sa dnes radí ku špičke v rámci dostupných audio kodekov. Univerzálnosť Opusu je daná vnútornou štruktúrou jeho implementácie – prakticky pozostáva z dvoch kodekov: SILK (kodek spoločnosti Microsoft určený na kódovanie reči) a CELT (kodek pre hudbu vyvinutý Xiph.org). Opus môže pracovať buď v režime SILK (pracuje so vzorkovacími frekvenciami 8, 12, 16 kHz), CELT (pracuje výhradne so 48 kHz) alebo v špeciálnom, hybridnom režime, kedy sa vstupný signál delí medzi tieto dva bloky. Keďže sa práca zameriava na spracovanie hudobných signálov, bude v tejto kapitole popísaná štruktúra a chod dielčieho kodeku CELT. [20]



Obr. 2.6: Bloková schéma OPUS kódera [21]

CELT

OPUS veľmi kvalitne potláča vznik zvukových artefaktov, vďaka pokročilému psychoakustickému modelu. CELT robí percepčný prieskum pomocou spektrálnej obálky, ktorú následne ukladá pomocou kódovania v pásmach, ktoré takmer presne kopírujú kritické pásma ľudského sluchu na Barkovej stupnici. Z každého vstupného rámca určí energiu v jednotlivých frekvenčných pásmach, ktorú neskôr v procese dekódovania rekonštruuje ako spektrálnu obálku. Spektrálne koeficienty jednotlivých pásiem sa získavajú pomocou MDCT. Transformačné okná tohto bloku majú pevné, nízke prekrytie 2,5 ms. Zároveň sa v kóderi aj dekóderi musí použiť preemfázový filter, pretože metóda okien s fixným časovým prekrytím spôsobuje únik spektrálnych zložiek do susedných pásiem. Hlavnou výhodou takýchto okien je zníženie algoritmického oneskorenia. V prípade, že je v rámci identifikovaný výrazný prechodný jav, a je potrebné eliminovať možné pre-echo artefakty, sa zvolí iný prístup – dlhé okno sa symetricky rozdelí na niekoľko menších, ktorých prekrytie je stále 2,5 ms, no vzhľadom na ich dĺžku je toto prekrytie 50 % so susedným oknom, každá vzorka sa teda nachádza v dvoch oknách (metóda s najvyšším možným rozlíšením). [20]

Hrubá kvantizácia energie

Koeficienty získané z MDCT sú v každom subpásme kvantované dvojprechodovým kvantizátorom. Tento typ kvantizátora má nastavené pevné 6 dB rozlíšenie a disponuje možnosťou použiť spektrálnu predikciu medzi rámcami alebo subpásmami (kvôli zjednodušeniu kódovania a potenciálnemu ušetreniu bitov). Po hrubej kvantizácii je zistené, koľko bitov je potrebných pre zakódovanie aktuálneho rámca (môže sa stať, že niektoré pásma nemajú pridelený ani jeden bit). Tieto pásma sú neskôr rekonštruované prídavnou funkciou Opusu – spektrálnym skladaním pásiem za pomoci energie nižších pásiem. Energia tých pásiem, ktoré bity obdržali, môže byť v neskoršom procese ešte prekvantovaná pomocou bloku pre jemnú energetickú kvantizáciu. [20]

Jemná energetická kvantizácia

Keď sa určí pridelenie bitov pre jednotlivé pásma, kóder môže spresniť nakvantovaný energiu v každom pásme. Za predpokladu, že kóderu ostali v závere kódovacieho procesu voľné bity, rozdeľuje ich rovnomerne medzi subpásma pre zvýšenie počtu kvantizačných úrovní, a teda presnejšiu kvantizáciu. [20]

Bitová alokácia

Po kvantizácii spektrálneho zloženia každého subpásma prichádza na rad bitová alokácia. Každému rámcu je na základe hrubej kvantizácie odhadnutý minimálny počet bitov na jeho zakódovanie. Následne sa v kóderi spúšťajú funkcie bitovej alokácie, ktoré majú presne prerozdeliť bity medzi pásma. Zaujímavý je proces preskakovania pásiem – pokiaľ je pásmu pridelených príliš málo bitov, hrozí, že jeho zakódovaním vznikne kvalitatívne veľmi zlá informácia. Z tohto dôvodu sa toto pásmo preskočí, a je neskôr nahradené pomocou techniky spektrálneho skladania, ktorá vygeneruje lepší výsledok. Ušetrené bity sú v danom rámci priradené iným pásmam, v ktorých sa dosiahne ešte vyššia kvalita. Zvyšné bity sú prerozdelené do pásiem tak, aby čo najvernejšie kopírovali odhadnutú spektrálnu obálku (bitové naklonenie alokácie). Celý tento proces bitovej alokácie tvorí takmer konštantný odstup užitočného signálu od šumu a generuje maskovaciu krivku, ktorá však nezahŕňa dva podstatné javy: rozdiel povahy tonálnych a šumových signálov a maskovanie prenikajúce do susedných pásiem. Z toho dôvodu CELT obsahuje funkcie, ktoré tieto nedostatky kompenzujú. Jedná sa o spomínané bitové naklonenie alokácie a zosilnenie vybraných pásiem. Pri zosilňovaní pásiem dochádza ku presunu bitov z pásiem, ktoré majú dostatok bitov na to, aby boli zakódované (nie nahradené pomocou spektrálneho skladania) do pásiem, ktoré požadujú viac bitov. Táto operácia sa spravidla deje v rámcoch, ktoré obsahujú tranzientné javy, a rámcoch, v ktorých boli použité krátke MDCT okná. [20]

Kvantizácia a kódovanie

Bitová rýchlosť sa v Opuse definuje pomocou transportného protokolu. Bitová alokácia sa riadi na základe údajov o zvyšných bitoch, kóder musí mať na začiatku kódovacieho procesu definovanú konečnú bitovú rýchlosť. Opus podporuje akúkoľvek bitovú rýchlosť, ktorá odpovedá celému číslu bitov použitých na jeden rámec. Tento postup má niekoľko výhod: znížia sa výpočtové požiadavky na kóder, nie je potrebné niekoľkostupňové kódovanie, ani bitová nádrž. Takýmto zdynamizovaným procesom sa dosahuje to, že kódované informácie nikdy neprekročia počet bitov, ktoré sú v skorších fázach kódovania pridelené k rámcu. Na kódovanie výsledného bitového toku sa v Opuse používa kóder rozsahu. Na vstup kóderu rozsahu sú privádzané dáta s nakvantovanými hodnotami v subpásmach. Pokiaľ tento kóder detekuje vstupné slovo s príliš dlhým rozsahom, zakóduje určitú časť ako tzv. surové bity, ktoré sú následne priradené na koniec výstupného paketu. Popísaný spôsob ukladania dát prináša výhody ako uľahčenie prepínania medzi dátami zakódovanými

v kóderi rozsahu a surovými bitmi. Navyše sa týmto rozdelením zlepši prevencia proti zavedeným chybám. [20]

Delenie pásiem

Pri vysokých bitových rýchlostiach sa pridelujú niektorým pásmam aj stovky bitov. Preto sa na každé pásmo, ktoré obsahuje viac ako 32 bitov použije M/S stereo kódovanie, čím je dosiahnuté ďalšie zväčšenie kompresného pomeru. [20]

Psychoakustické zlepšenia

Opus vo všeobecnosti zavádza veľmi málo zreteľných zvukových artefaktov, ponúka navyše štyri prídavné psychoakustické metódy pre zníženie rizika počuteľnosti zavedených artefaktov:

- Prefilter a postfiler – tieto filtre majú za úlohu znížiť kvantizačný šum, vzniknutý nízkym prekrytím MDCT okien, ktorý sa prejavuje najmä u vysokotonálnych signálov
- Variabilné časovo-frekvenčné rozlíšenie – prispôbenie rozlíšenia pri oknách, ktoré obsahujú rôznorodé spektrálne signály
- Rozťahovacie rotácie – tieto funkcie majú za úlohu rozptýliť tonálny hluk. Ten vzniká, keď v pásmach, ktoré mali veľa koeficientov zaokrúhlených na nulu, ostane niekoľko koeficientov, ktoré tvoria artefakty tónového charakteru
- Prevencia kolapsu – pásmo, v ktorom bolo príliš veľa koeficientov zaokrúhlených na nulu, môže pri prehrávaní spôsobiť efekt zvukového výpadku. Preto je do tohto pásma pridaný pseudonáhodný šum s výškou minimálnej pásmovej energie, vypočítanej z predošlých dvoch rámcov. [20]

2.7 Porovnanie postupov kódovania zvolených kodekov

V MP3 sa na začiatku kódovacieho procesu delia spracovávané rámce na 32 frekvenčných oblastí so zhodnou šírkou, ktoré sú následne ďalej rozdelené na menšie pásma nazývané pásma mierkového faktora. MP3 má možnosť prepínania dĺžky MDCT okien na krátke (256 vzoriek) a dlhé (1024 vzoriek). Prekrytie okien je 50 % a na

prechod medzi oknami rôznych dĺžok sa používajú prechodné okná. Pre bitovú alokáciu je použitá metóda dvoch vnorených slučiek. Po kvantizácii sa dáta ukladajú do 32 Huffmanových tabuliek.

AAC má možnosť voliť tvar MDCT okna medzi KBD (Kaiser Bessel Derived window) alebo sínusovým. Na rozdiel od MP3 má dĺžka dlhého MDCT okna 2048 vzoriek, pri prepnutí dĺžky okna používa prechodné okná. Oproti MP3 má AAC pokročilejší psychoakustický model a obsahuje mnohé vylepšenia ako napr. TNS a možnosť joint stereo kódovania vo viacerých režimoch. K uloženiu kódovaných dát slúži len 12 Huffmanových tabuliek. Najviac prídavných funkcií má oproti MP3 určite AAC profilu High Efficiency a AAC Main Profile.

Vorbis môže voliť veľkosť okna MDCT v rozmedzí 64–8192 vzoriek, na prechod medzi dĺžkami používa prechodné okná. Spektrum vstupného súboru je pomocou psychoakustického modelu a floor funkcií rozdelené na hrubý odhad spektra, vygenerovaný aproximáciou skrz floor krivky a detaily približujúce presnú, pôvodnú spektrálnu obálku, nazývané zostatky. Podobne ako v AAC sa pred záverečným bezstratovým kódovaním využíva spájanie kanálov. Vorbis využíva dvojstupňové, bezstratové kódovanie – vektorovú kvantizáciu a Huffmanovo kódovanie.

Opus sa skladá z dvoch kodekov, ktoré môžu fungovať jednotlivo alebo v hybridnom režime. Opus (resp. CELT) je jediný zo štvorice kodekov, ktorý pracuje natívne so vzorkovacou frekvenciou 48 kHz. Vstupné rámce delí do frekvenčných pásiem, ktoré takmer presne odpovedajú kritickým pásmam ľudského sluchu. Prekrytie MDCT okien má nastavené fixne na 2,5 ms. Oproti AAC a MP3 má zložitejšiu, no efektívnejšiu bitovú alokáciu. Navyše disponuje pokročilými funkciami ako spektrálne skladanie pásiem, rozťahovacie rotačné funkcie, naklonenie bitovej alokácie a prevencia kolapsu pásiem. Opus ako jediný kodek nepoužíva Huffmanove tabulky.

3 Testy pre hodnotenie posluchovej kvality

Použitie posluchových testov má v tejto téme jednoduchý dôvod – aj keď sa kodeky v mnohom líšia, hlavným požiadavkom, ktorý stojí na vrchole hierarchie ich vlastností, je čo najvernejšia výsledná kvalita zvuku po kompresii dát. Rozlišujeme niekoľko druhov posluchových testov, všetky však majú niektoré spoločné rysy. Návrhár testu musí mať predstavu o obsahu testu, dĺžke testovania, musí respondentom pred testovaním vysvetliť problematiku, spôsob, akým budú hodnotiť, atď. Respondenti by však nemali vedieť o samotnom obsahu príliš veľa technických informácií. Existuje totiž riziko, že respondent môže mať určité znalosti danej problematiky, ktoré budú skresľovať jeho úsudok. Respondenti teda napr. netušia, aké kodeky sa testujú, aké bitové rýchlosti boli pre kódovanie testovacieho materiálu použité – ich úlohou je hodnotiť len referenčnú kvalitu toho, čo je im v danom teste poskytnuté. [22]

3.1 Test ABX

Táto testovacia metóda prišla ako prvý koncept posluchových testov. Respondentovi sú predložené 3 zvukové stopy: stopa A, B a X. Stopy A a B sú vzájomne odlišné a úlohou respondenta je rozhodnúť, či je stopa X identická so stopou A alebo B. Zmyslom tejto metódy je zistiť, či testovaný respondent počuje percepčné rozdiely medzi poskytnutými zvukovými stopami (resp. štatisticky dokázať, že jeho úsudok nie je vecou náhody). Tento druh testu je vhodný napríklad pre výber respondentov k pokročilejším metódam testovania. [23]

3.2 Test ABC/HR

Systém ABC/HR (HR = Hidden Reference, teda skrytá referencia) rozširuje testovanie aj o samotné hodnotenie percepčného rozdielu. Respondentovi sú v každom dielčom teste poskytnuté 3 zvukové stopy s označením A, B a C. Respondent je oboznámený s nasledujúcimi skutočnosťami:

- stopa A je referenčná stopa, ktorej kvalita odpovedá pôvodnému zvukovému súboru pred kompresiou dát
- stopy B a C majú pridelený svoj obsah náhodne. Jedna z nich obsahuje zvukovú stopu identickú so stopou A a druhá z nich je komprimovaná stopa, ktorá je

predmetom jeho hodnotenia. [22]

3.2.1 Hodnotenie percepčného rozdielu poskytnutých stôp

Hlavnou úlohou respondenta je teda správna identifikácia komprimovanej stopy, ktorej zvukovú kvalitu následne hodnotí ako relatívnu ku kvalite stopy A. Respondentovi je zväčša poskytnutá tabuľka 3.1 s popísanými bodmi, odpovedajúcimi stupňu poškodenia. Reálne je v hodnotení poskytnutá respondentovi väčšia voľnosť, hodnotiaci verdikt sa uvádza pomocou číslenej škály, ktorá umožňuje hodnotenie aj medzi vopred definovanými bodmi. [22]

Tab. 3.1: Preddefinované body poskytnuté respondentovi

5–nepostrehnuteľný
4–takmer nepostrehnuteľný
3–postrehnuteľný ale nie príliš rušivý
2–postrehnuteľný a výrazne rušivý
1–znateľný a nepríjemný

3.3 Test MUSHRA

MUSHRA (MUlti Stimulus test with Hidden Reference and Anchor, teda niekoľko stimulový test so skrytou referenciou a „kotvou“), je metóda, ktorá integruje testovanie niekoľkých stimulov do jedného dielčieho testu. Takýto test môže zahŕňať až 15 zvukových stôp – poskytnutú referenciu, skrytú referenciu, kotvu a maximálne 12 testovaných stôp. Základná myšlienka MUSHRA testovania tkvie v zaujímavom psychologickom efekte aplikovanom na respondentovi – za predpokladu, že si respondent uvedomí významný relatívny rozdiel medzi referenčnou stopou a kotvou (kotva je zvuková stopa s evidentným, prehnaným poškodením), vytvorí si subjektívnu stupnicu hodnotenia, do ktorej presnejšie vsádza svoje verdikty pri hodnotení testovaných stôp. Výhodou metódy je nižšia časová náročnosť, za predpokladu, že sa zohľadní dĺžka jednotlivých stimulov (rádovo niekoľko desiatok sekúnd). Testovacia metóda je normalizovaná v odporúčaní ITU-R BS.1534-1. [24]

3.4 Popis parametrov testu podľa ITU-R

ITU-R BS.1116-3 je dokument obsahujúci normatívne odporúčania pre návrh posluchových testov so špecializovaním na hodnotenie drobných percepčných porúch v audio súboroch. Táto norma špecifikuje pre návrhára testu postup prípravy každej fázy. [25]

3.4.1 Výber respondentov

Respondenti pripustení k testovaniu by mali mať aspoň minimálne skúsenosti z oblasti rozboru zvukových dát. Pre každého respondenta sa taktiež odporúča vykonanie audiometrických testov. [25]

3.4.2 Fáza školenia

Prípravná fáza prebieha pred samotným testovaním, a má za úlohu oboznámiť respondentov s testovacím prostredím, klasifikačným procesom, klasifikačnými stupnicami a skúmanou problematikou vôbec. Pomocou názorných ukážok by mali byť respondenti oboznámení so zvukovými artefaktami, ktoré budú v testovaní predmetom ich identifikácie a hodnotenia. Pre lepšie porozumenie môže byť respondentom umožnená vzájomná diskusia v menšom počte ľudí. Pokiaľ sa priebeh školiacej fázy nezanedbá, je pravdepodobné, že vyprodukuje dostatočne poučených respondentov aj z úplných laikov. Počet respondentov pripustených k testu je priamo závislý na technickej príprave testovania, ak sa návrhár testu bude striktne riadiť normou, môže byť postačujúca vzorka len 20 ľudí. [25]

3.4.3 Testovacia metóda

Norma odporúča použitie metódy ABC/HR. Respondent má mať možnosť kontinuálne prepínať medzi stopami, vyžaduje sa presné časové zosúladenie stôp, aby nevznikali rušivé artefakty v podobe praskania pri prepínaní. Stupnica pre hodnotenie by mala obsahovať 50 bodov s 5 záchytnými bodmi. [25]

3.4.4 Zvukový materiál pre testovanie

Obsahový materiál testu by mal byť zvolený tak, aby čo najlepšie spĺňal požiadavky kritického materiálu (plné spektrum, primeraný dynamický rozsah ukážky...), no zároveň príliš nerozptyľoval respondentov. Neodporúčajú sa syntetické nahrávky a príliš atraktívne hudobné vzorky. [25]

3.4.5 Reprodukčné zariadenia

Pre kvalitný posluh sú odporúčané slúchadlá s vyrovnanou kmitočtovou charakteristikou a nízkym časovým oneskorením kanálov. V prípade použitia reproduktorov sú tiež definované parametre ako index smerovosti, nelineárne skreslenie a dynamický rozsah. Pre posluhové testy vykonané na reproduktoroch sa taktiež musí normatívne prispôbiť aj akustika miestnosti. [25]

3.4.6 Fáza vyhodnocovania výsledkov

Vyhodnotenie prebieha pomocou štatistických metód (ANOVA, t-testy, Scheffeho testy alebo Neuman-Keuls testy). [25]

3.4.7 Prezentovanie výsledkov

Výsledky testovania by mali pozostávať z grafov a tabuliek s číselnými štatistickými údajmi. [25]

3.5 Stručný popis použitých štatistických metód

3.5.1 t-test

Táto štatistická metóda umožňuje porovnanie stredných hodnôt jednej (jednovýberový) alebo dvoch odlišných (dvojvýberový alebo tiež párový) skupín. Úlohou týchto testov je predovšetkým prijatie či neprijatie nulovej hypotézy – tvrdenia o neznámych vlastnostiach rozdelenia sledovanej veličiny. Vykonaním t-testov sa v práci

štatisticky overuje, či boli hodnotenia respondentov v teste náhodné alebo systematické. [26]

3.5.2 ANOVA

ANOVA (analýza rozptylu) je pokročilejšou štatistickou metódou – oproti t-testu umožňuje porovnávanie viacerých nezávislých skupín, čo je presne prípad tejto práce, pretože sa porovnávajú hodnotenia respondentov pre štyri rôzne kodeky. Dôležitým druhom analýzy rozptylu pre túto prácu je tzv. RM ANOVA (Repeated Measures, teda s opakovanými meraniami), ktorá vo výsledkoch zohľadní aj koreláciu dát jedného respondenta. [27]

3.6 Výsledná štruktúra testu

Tab. 3.2: Kľúčové parametre definované normou ITU-R BS.1116-3

Norma ITU-R BS.1116-3	
Max dĺžka testovania [min]	20–30
Dĺžka prehrávania stopy [s]	10–25
Dĺžka jedného testu [s]	30–75
Počet testov	10–15

Tab. 3.3: Zvolené parametre posluchového testu podľa normy ITU-R BS.1116-3

Zvolené možnosti	
Max. dĺžka testovania [min]	25–30
Dĺžka prehrávania stopy [s]	5–10
Dĺžka jedného testu [s]	>25
Počet testov	24
Počet kodekov	4
Počet ukážok	3
Počet bitratov	2

Návrh testu vychádza priamo zo spomínanej normy ITU-R BS.1116-3. Parametre testu, ktoré boli zvolené je možné vidieť v tabuľke 3.3. Jediný parameter, ktorý sa nedeje normy je počet testov – toto prekročenie ale bolo zohľadnené v dĺžke prehrávania stôp, vďaka čomu je dodržaná celková dĺžka testovania.

Tab. 3.4: Prehľad obsahu testov

kodek	AAC, MP3, Opus, Vorbis					
ukážka	metal		jazz		symfónia	
bitrate [kbps]	80	96	80	96	80	96
číslo testu	1	2	3	4	5	6

Obsah jednotlivých testov je zobrazený v tabuľke 3.4. Každý kodek bol podrobený 6 testom (3 nahrávky \times 2 bitové rýchlosti, rovnaké pre každý kodek). Týmto konceptom testovania bolo možné vyhodnotiť výslednú kvalitu vzhľadom ku možnostiam kódovacieho algoritmu pri niekoľkých konkrétnych podmienkach.

Test bol koncipovaný ako ABC/HR. V každom teste boli respondentovi ponúknuté 3 zvukové stopy, medzi ktorými mohol kontinuálne prepínať, či už podľa pokynov alebo vlastného uváženia. Respondent mohol test opakovať, kým si nebol istý svojím hodnotením.

4 Príprava programového materiálu

4.1 Editácia a kódovanie

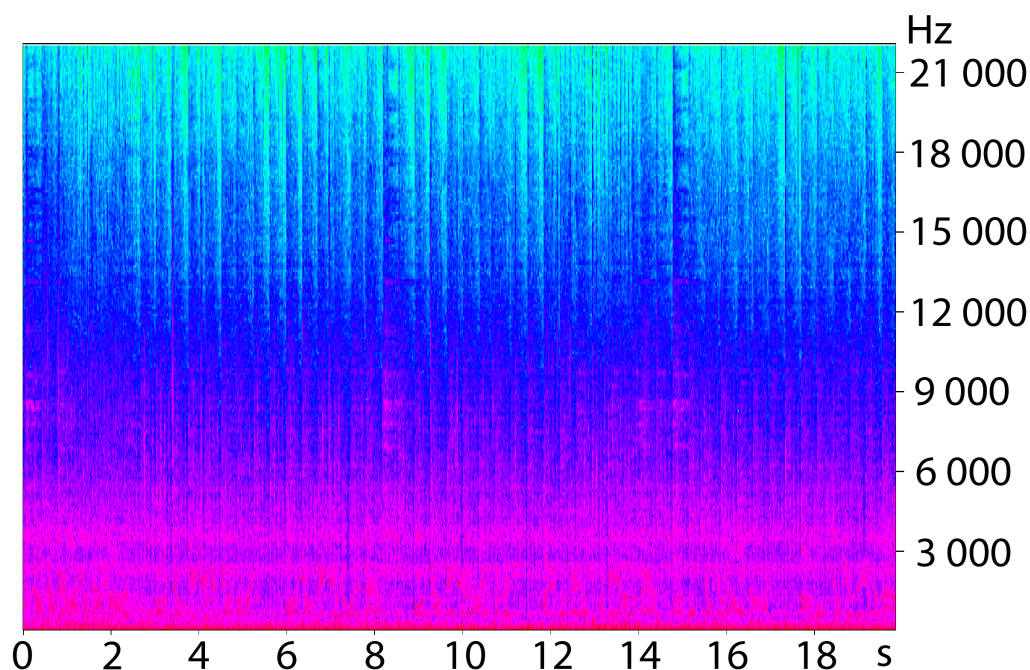
Programový materiál tvoria 3 žánrovo rozličné, profesionálne vyprodukované stereo nahrávky. V dôsledku nízkej dostupnosti hudobných súborov vo formáte WAV, boli uprednostnené súbory v bezstratovom formáte FLAC. Zo všetkých 3 skladieb boli v programe Reaper vystrihnuté 20 sekundové úseky, ktoré boli následne vyrendrované vo formáte WAV (44,1 kHz/16 bit).

Pre kódovanie nastrihaných ukážok bola použitá už spomínaná knižnica ffmpeg. Pre zjednodušenie práce s touto knižnicou bol použitý skript, ktorý automatizoval kódovanie a dekódovanie všetkých súborov naraz. Bolo potrebné vytvoriť zložku, v ktorej sa nachádzali všetky nastrihané ukážky, daný skript a 3 ďalšie zložky pomenované podľa zvolených žánrov, do ktorých sa umiestňovali jednotlivé súbory pripravené pre posluchový test (spätne dekódované ukážky). Skriptu bolo potrebné zdefinovať názvy vstupných zvukových súborov, vzorkovacie frekvencie pre kodeky (44,1 kHz s výnimkou Opusu(48 kHz)), bitové rýchlosti výstupných súborov a výstupné formáty, ktoré požadujeme. Skript fungoval v dvoch krokoch – najskôr pomocou zvolených kodekov z ffmpeg vygeneroval súbory v zvolených formátoch s príslušným názvom (track-bitová rýchlosť-prípona formátu.prípona formátu). Následne vygenerované súbory dekodoval spätne do formátu WAV, umiestnil ich do príslušných zložiek a zmazal súbory vytvorené v prvom medzikroku. Po dokončení procesu bolo v každej zložke 8 súborov vo formáte WAV (4 kodeky \times 2 bitové rýchlosti).

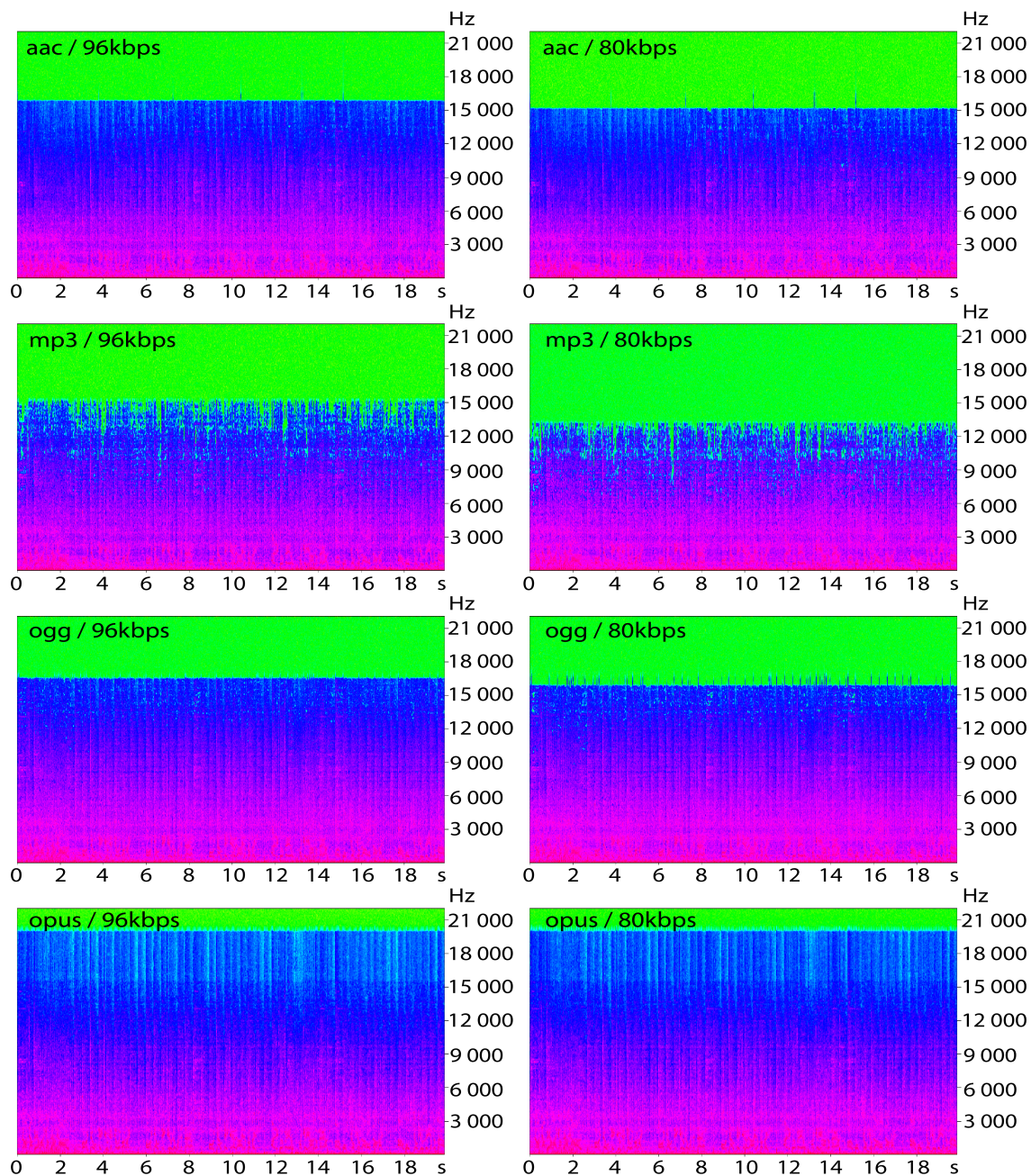
Ukázalo sa, že vygenerované súbory majú pridané nulové vzorky pred aj za zvukovým tokom (v prípade AAC bol rozdiel dĺžok súborov oproti pôvodným niekoľko desiatok ms, mierne zmeny však obsahoval každý dekódovaný súbor). Z tohto dôvodu sa všetkých 8 súborov z každej zložky upravilo na pôvodnú dĺžku pomocou manuálneho zostrihania a zrovnania s časovým priebehom pôvodného, nekomprimovaného súboru. Po tejto úprave boli súbory pripravené pre posluchové testy.

4.2 Spektrogramy kódovaných audio súborov

Spektrogramy sú vygenerované programom SigView, každý z nich obsahuje časovo-spektrálny priebeh oboch kanálov skladby. Dĺžka Hannovho transformačného okna bola nastavená na hodnotu 1024 vzoriek (čo odpovedá približne 23 ms), offset (posun) medzi oknami predstavoval 819 vzoriek, prekrytie medzi oknami teda činilo 205 vzoriek. Spektrogramy slúžia na vizuálnu prezentáciu časovo-frekvenčného priebehu zakódovaných súborov. Na základe porovnávania grafov z obrázku 4.2 a 4.1, bolo možné predbežne vizuálne odhadnúť, aké výsledky dosiahnu kodeky v posluchových testoch. Pri MP3 bola tiež smerodajná informácia, že sa jeho spektrogramy pri jednotlivých bitových rýchlostiach navzájom líšia výraznejšie, ako pri zvyšných kodekoch. Toto zistenie nakoniec viedlo k rozhodnutiu, že zadenovanie záchytných bodov v školiacej fáze posluchových testov bude realizované pomocou kódovania do MP3 s použitím odlišných bitových rýchlostí. Zaujímavé sú spektrogramy Opusu, na ktorých je zreteľne vidieť redšia oblasť medzi 15–20 kHz. Táto oblasť je vytvorená technikou spektrálneho skladania, popisovanou v sekcii 2.6.4.



Obr. 4.1: Spektrogram pôvodného súboru (ukážka metal)



Obr. 4.2: Ukážky sprektrogramov kódovaných súborov (ukážka metal)

5 Testovacie prostredie

Odkaz na stránku: <http://52.59.215.133/>

5.1 Úvodná stránka

5.1.1 Inštruktážne video

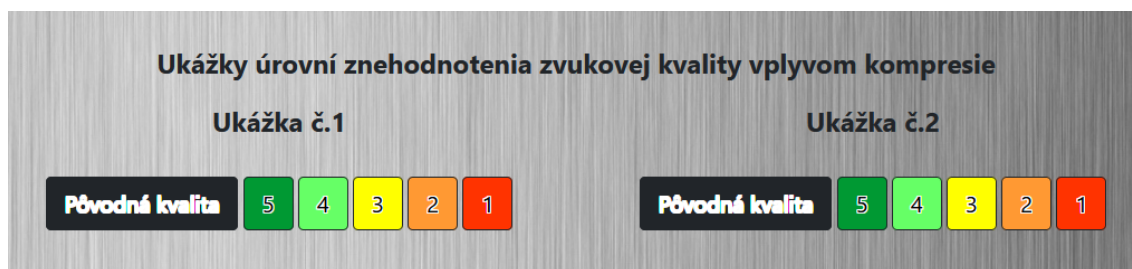
Odkaz na video: <https://www.youtube.com/watch?v=hrBOYxrzKX4>

Pre lepší kontakt a rýchlejší priebeh vstupnej inštruktáže bolo nahraté, zostrihané a uploadované video, ktoré si respondenti mohli spustiť priamo na úvodnej stránke. V tomto videu je respondentom načrtnutá problematika stratového audio kódovania, podrobne popísaný každý element nachádzajúci sa na úvodnej stránke aj v posluchoých testoch. Respondentom je taktiež názorne odprezentovaný postup vyplnenia posluchoých testov, odporúčená krátka prestávka v strede testovania a v závere poďakované za účasť.

5.1.2 Odporúčania pre nastavenie audio reťazca

Keďže zvolený koncept implementácie posluchoých testov neumožňoval všetkým zúčastneným respondentom identické technické podmienky, musela byť v úvode odporúčená čo najoptimálnejšia konfigurácia ich audio reťazcov. Hlasitosť nemohla byť príliš vysoká, zároveň však dostatočná pre zreteľné rozoznanie prítomnosti zvukových artefaktov v kódovaných zvukových stopách. Pre posluh boli uprednostňované slúchadlá.

5.1.3 Ukážky úrovni znehodnotenia zvuku vplyvom kompresie



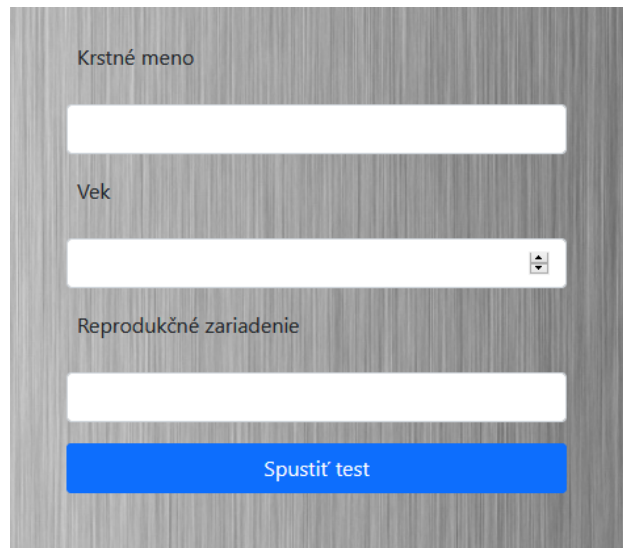
Obr. 5.1: Panely s ukážkami znehodnotenia

Testovaným respondentom sú poskytnuté 2 panely, každý obsahujúci 6 graficky rozdielnych tlačidiel. Tlačidlami respondenti spúšťajú prehrávanie niekoľko sekundovej zvukovej stopy, ktorá podľa číslice na tlačidlo zodpovedá úrovni znehodnotenia z tabuľky 3.1 alebo kvalite pôvodnej kvality ukážky. Respondenti mali za úlohu počúvať tieto úrovne znehodnotenia zostupne. Tabuľka definujúca stupne znehodnotenia je respondentom k dispozícii priamo pod panelmi.

Každý panel obsahuje nahrávku žánrovo blízku ukážkam obsiahnutým v posluchových testoch (metal a jazz). Tieto ukážky boli získané rovnakým postupom, aký sa popisuje v kapitole 4.1. Spektrogramy kódovaných súborov v kapitole 4.2 preukázali, že kodek s predpokladom dosiahnuť najneuspokojivejšie zvukové výsledky je MP3. Stupne znehodnotenia prezentované respondentom v školiacej fáze teda mali nasledujúci obsah:

- Pôvodná kvalita – WAV
- 5 – MP3 / 128 kbps
- 4 – MP3 / 112 kbps
- 3 – MP3 / 96 kbps
- 2 – MP3 / 80 kbps
- 1 – MP3 / 64 kbps.

5.1.4 Vstupný formulár pre zber dát od respondentov



Krstné meno

Vek

Reprodukčné zariadenie

Spustiť test

Obr. 5.2: Vstupný formulár

Formulár obsahuje 3 polia:

- krstné meno respondenta
- vek respondenta
- stručný popis audio reťazca, ktorý bude použitý pre posluh.

Z hľadiska štatistického vyhodnocovania bol najdôležitejším získaným údajom popis reprodukčného zariadenia. Každý respondent mal k dispozícii odlišné technické vybavenie, ktoré je jedným z hlavných faktorov ovplyvňujúcich výsledky. Bolo preto zásadné získať akúkoľvek informáciu približujúcu použité reprodukčné zariadenie.

Ukončenie fázy školenia sa vykonáva odkliknutím tlačidla „Spustiť test“. Úlohou tohto tlačidla sú nasledovné procesy:

- uloženie údajov z formulára do URL adresy ako „query string“
- presmerovanie respondenta do prostredia posluhových testov
- zaznamenanie času, kedy bol test spustený.

5.2 Testovacie prostredie posluhových testov

Po presmerovaní respondentov na stránku s posluhovými testami sa vygeneruje náhodné poradie jednotlivých testov unikátne pre každého jedného respondenta.

5.2.1 Blok inštrukcií

- Stlačením ► spustí prehrávanie stôp
- Prepínaj medzi stopami (ideálne v poradí A-B-A-C) a pozoruj odlišnosti stôp B a C voči referenčnej stope A
- Po identifikácii hľadanej, komprimovanej stopy ohodnot' kvalitu na príslušnom slideri
- Po hodnotení pokračuj k ďalšiemu testu stlačením ►

Obr. 5.3: Inštrukcie k testom

Kvôli veľmi obmedzenej vstupnej inštruktáži bolo potrebné predísť prípadnému zmätočnému konaniu respondentov v posluchovej testoch. Aby boli zúčastnení respondenti rozptýlení počas testov čo najmenej, obsahuje stránka stručné, v bodoch písané inštrukcie k vykonaniu testov.

5.2.2 Blok pre prehrávanie a voľbu prehrávanej stopy



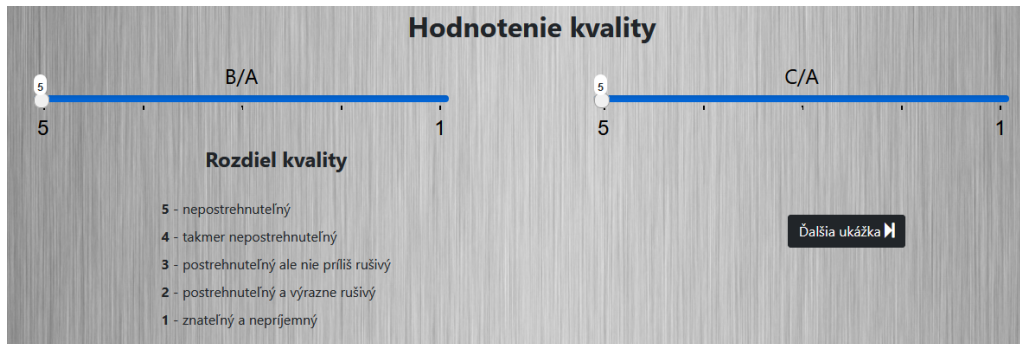
Obr. 5.4: Blok pre prehrávanie a voľbu prehrávanej stopy

Táto sekcia začína počítačom poradie dielčích testov – respondenti majú možnosť pozorovať poradie práve spusteného testu. Tlačidlom „PLAY“ sa spustí prehrávanie prvej stopy (na začiatku každého testu je to defaultne stopa A). Po kliknutí tlačidlo zmení vzhľad na „PAUSE“, aby respondenti vedeli, že prehrávanie stôp bolo úspešne spustené. Manipulácia s týmto tlačidlom nie je respondentom povolená, kým ich internetový prehliadač nenačíta obsah tlačidiel A, B, C.

Práve prehrávanú stopu respondenti nastavujú zakliknutím jedného z tlačidiel A, B a C. Prepnutie stopy spustí obsah zakliknutého tlačidla z časovej pozície, na ktorej bola predošlá stopa zastavená v momente prepnutia (pri prepnutí je nastavený tiež nepatrný crossfade pre ešte plynulejší prechod medzi stopami). Týmto spôsobom je zabezpečený kľúčový predpoklad testov ABC/HR – kontinuálne prehrávanie.

Priradenie obsahu tlačidiel B a C prebieha pri spustení dielčieho testu a je v každom teste náhodné.

5.2.3 Blok identifikácie a hodnotenia kvality



Obr. 5.5: Blok identifikácie a hodnotenia kvality

Stránka ponúka dva posuvné bežce s označením „B/A“ a „C/A“. Tieto bežce slúžia na ohodnotenie relatívnej kvality stôp B a C ku stope A. Číslice na ich okrajoch naznačujú smer stupnice, zvislé čiary, pravidelne rozmiestnené po celej dĺžke dráhy bežca, slúžia ako záchytné body. Záchytné body sú respondentom taktiež k dispozícii priamo pod bežcami. Pri nastavovaní hodnotenia sa navyše aktuálne nastavená hodnota zobrazuje v bubline nad bežcom. Rozsah dráhy bežcov je 50 dielikov (5 záchytných bodov \times 10 medzistupňov).

Respondenti mali za úlohu manipulovať vždy iba jedným z bežcov – stopa, ktorá vyžaduje hodnotenie je v teste len jedna (druhej ostáva hodnotenie 5).

Tlačidlo „Ďalšia ukážka“ realizuje nasledovné úkony:

- uloženie číselných hodnôt jednotlivých bežcov
- uloženie textovej skratky o priradení obsahu stopám B a C
- uloženie aktuálneho počtu sekúnd uplynulých od spustenia testovacieho prostredia
- zvýšenie hodnoty počítadla poradia testov o 1
- zastavenie práve hrajúcej stopy a pretočenie štartovacej pozície nasledujúceho prehrávania na nulu
- nastavenie stopy A ako prvú prehrávanú stopu v nasledujúcom teste
- randomizáciu pridelenia obsahu stôp B a C
- načítanie nových zvukových súborov podľa vygenerovaných parametrov testu.

Po absolvovaní posledného testu sú respondenti presmerovaní na stránku s poďakovaním. Získané údaje z testov sa odosielajú a ukladajú na server po odkliknutí tlačidla „Ďalšia ukážka“ v poslednom teste.

5.2.4 Výstupné súbory posluchových testov

test	codec	genre	bitrate	ratingOriginal	ratingConverted	seconds
0	ogg	jazz	high	5	2	138.82
1	mp3	symp	high	5	1	227.437
2	opus	jazz	low	4.9	5	297.053
3	mp3	jazz	high	5	4	351.849
4	ogg	jazz	low	4.9	5	391.775
5	opus	jazz	high	5	5	407.223
6	mp3	symp	low	5	1	451.479
7	aac	jazz	low	5	4	468.209
8	ogg	metal	low	5	3	482.587
9	ogg	metal	high	5	4	493.406
10	mp3	metal	low	5	1	515.106
11	ogg	symp	high	5	4	522.536
12	opus	metal	low	5	4.9	538.559
13	aac	jazz	high	5	4	556.005
14	aac	metal	high	5	4.4	570.112
15	aac	metal	low	4.7	5	590.775
16	mp3	metal	high	5	3	605.234
17	ogg	symp	low	5	2	624.252
18	opus	symp	low	5	4.2	633.41
19	opus	symp	high	4	5	641.495
20	opus	metal	high	5	4.5	657.4
21	mp3	jazz	low	5	4	676.426
22	aac	symp	low	5	4	711.546
23	aac	symp	high	5	3	720.842

Obr. 5.6: Výstupný CSV súbor

Po dokončení posledného testu stránka vygeneruje dva súbory – CSV, do ktorého sa uložia údaje z jednotlivých testov uložené v príslušných premenných. Taktiež sa vytvorí súbor formátu TXT, do ktorého je vložený údaj vpísaný respondentom do poľa „Reprodukčné zariadenie“ v prvej, školiacej fáze. Oba tieto súbory majú zhodný názov, zložený z údajov zadaných do formulára (krstné meno–vek). Posledným, tretím údajom, ktorý sa v názve týchto súborov vyskytuje je mnohociferné číslo, odpovedajúce počtu uplynulých sekúnd od 1.1.1990, vygenerované v momente ukončenia testu. Táto číselná kombinácia robí názvy oboch vygenerovaných súborov jedinečnými, a predchádza tak ich možnému prepísaniu na serveri v prípade, že by sa objavili respondenti so zhodným menom a vekom.

Na obrázku 5.6 je zobrazený vzhľad obsahu CSV súboru. Hodnoty premenných „test“, „codec“, „genre“ a „bitrate“ sú stránke známe na začiatku každého testu, načítava podľa nich zo zložiek zvukové súbory pre aktuálny test – do tabuľky v CSV sa zapisujú kvôli identifikácii testu.

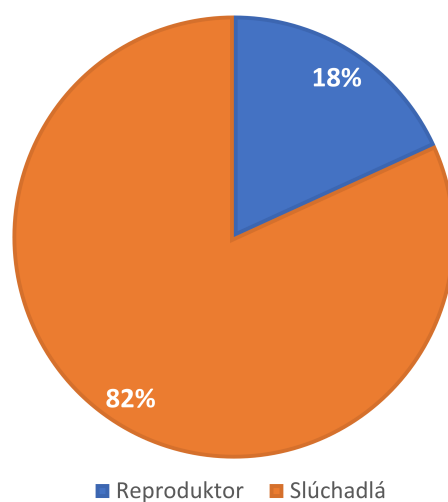
Pri každom teste sa ukladajú údaje o práve vygenerovanom pridelení obsahu tlačidiel B a C, a tiež respondentom nastavené číselné údaje z bežcov B/A a C/A do príslušných premenných. Aby nebolo potrebné manuálne kontrolovať, či respondent identifikoval a hodnotil správnu stopu (zhodu medzi tlačidlom s komprimovanou stopou a odpovedajúcim sliderom) nachádzajú sa v CSV súbore údaje „ratingOriginal“ a „ratingConverted“, ktoré integrujú vyššie spomenuté údaje do dvoch premenných. Údaje v týchto stĺpcoch sú teda zjednodušeným zápisom číselných hodnôt bežcov vzhľadom ku hodnoteným stopám.

Premenná „seconds“ obsahuje čas uplynutý od spustenia testovacieho prostredia po odkliknutie aktuálneho testu. Pomocou tohto údaju bolo možné overiť, či respondent strávil pri teste adekvátny čas.

6 Výsledky práce

6.1 Normatívna filtrácia respondentov pred štatistickým vyhodnotením

Posluchové testy sa zúčastnilo 34 respondentov, zastúpenie reprodukčných zariadení je zobrazené v grafe 6.1.



Obr. 6.1: Graf zastúpenia reprodukčných zariadení respondentov

6.1.1 Korekcia získaných dát

Celý proces „čistenia dát“ je normatívne definovaný, avšak vzhľadom na kvalitu získaných dát bola potrebná prídavná, úvodná korekcia. Z obrázku 5.6 je zrejmé, že správnosť výsledku dielčieho testu respondenta je jednoducho overiteľná porovnaním hodnôt „ratingOriginal“ a „ratingConverted“. Pri tejto prvotnej kontrole hodnotení dielčích testov od všetkých respondentov mohli byť pozorované 3 prípady:

Tab. 6.1: Možné kombinácie hodnotenia respondenta

Možnosť	ratingOriginal	ratingConverted
1	5	< 5
2	< 5	5
3	< 5	< 5

- možnosť 1 – správna identifikácia, ideálny stav
- možnosť 2 – chybná identifikácia, stále však prípustný údaj (má štatistický význam v neskorších fázach vyhodnocovania)
- možnosť 3 – nedodržanie pokynov vyžadujúce korekciu.

Testy, v ktorých respondenti (napriek pokynom v úvodnej fáze) hodnotili obidve stopy, museli byť rozdelené na testy, vyžadujúce korekciu a testy, ktorých výsledky boli úplne vyradené. Test, v ktorom mohla byť vykonaná prídavná korekcia, musel súčasne spĺňať nasledovné podmienky:

- $ratingOriginal \geq 4,5$
- $ratingConverted \leq ratingOriginal - 1$

V prípade, že boli tieto podmienky splnené, sa respondentovo pochybenie „ignorovalo“ a hodnota $ratingOriginal$ sa zaokrúhlila na hodnotu 5. Vo všetkých zvyšných prípadoch boli dielčie testy vyradené. Respondenti, ktorých dáta obsahovali príliš veľa testov určených na vyradenie (viac ako polovicu respondentových testov) boli zo štatistického vyhodnocovania vyradení úplne. V dôsledku tohto kroku boli vyradení 3 respondenti.

6.1.2 Selekcia respondentov pred štatistickým vyhodnotením analýzou rozptylu

Nasledujúce kroky už vychádzali priamo z normatívnych inštrukcií. Bolo potrebné odlišiť respondentov, ktorí preukázateľne spoľahlivo identifikovali hľadanú stopu od respondentov, ktorých hodnotenia boli skôr výsledkom tipovania.

Hodnotenie profilu odbornosti respondenta

Toto prerozdelenie respondentov na dve spomínané skupiny sa realizovalo pomocou jednostranného t-testu. Pred vykonaním t-testu so súborom dát jedného respondenta bolo však potrebné odfiltrovať dáta z testov, ktoré neprispievajú ku vyhodnoteniu skutočného odbornostného profilu respondenta. Táto selekcia dát, vykonaná individuálne pre každého respondenta, mala nasledovný postup:

Prvým krokom bolo získanie rozdielových stupňov hodnotení (r hodnôt), podľa vzorca:

$$r = ratingConverted - ratingOriginal \quad (6.1)$$

Z tabuľky 6.1 vyplýva, že pokiaľ bol respondent úspešný v identifikácii hľadaných stôp, budú jeho r hodnoty (v testoch, v ktorých hodnotil známku nižšou než 5, teda v drvivej väčšine) záporné čísla.

Z týchto r hodnôt sa následne spočítal aritmetický priemer:

$$\bar{x} = \frac{\sum_1^n x}{n} \quad (6.2)$$

a výberová smerodajná odchýlka pre vzorku:

$$s = \sqrt{\frac{1}{n-1} \sum_1^n (x_i - \bar{x})^2} \quad (6.3)$$

Následne sa museli r hodnoty prerozdeliť do dvoch podmnožín:

- r hodnoty testov, ktoré môžu dokázať dostatočnú odbornosť respondenta
- r hodnoty, ktoré by v t-teste spôsobili iba štatistický šum.

Vyhovujúce r hodnoty sa museli nachádzať v intervale, ktorého veľkosť bola spočítaná ako:

$$v = \bar{x} - \Delta s \quad (6.4)$$

kde Δ predstavuje násobok smerodajnej odchýlky (čím vyššie číslo je zvolené, tým je interval relevantných r hodnôt širší – kritérium výberu je benevolentnejšie). Zvyšovanie hodnoty Δ je použité, pokiaľ je pri najstriktnejšej možnosti ($\Delta=1$) v t-testoch neúspešných (teda vyradených) príliš veľa respondentov. Samotná veľkosť intervalu sa definuje ako (\bar{x}, v) .

Je dôležité podotknúť, že táto selekcia dát prebiehala iba na úrovni hodnotenia profilu odbornosti respondenta. Pokiaľ t-test štatisticky preukázal, že hodnotenia respondenta nie sú výsledkom náhodného tipovania (jeho dáta sú pre vyhodnotenie výsledkov posluchové testov relevantné), boli do štatistickej analýzy ANOVA zahrnuté všetky od neho získané dáta (samozrejme s výnimkou dát spomínaných v kapitole 6.1.1).

Vyradovanie respondentov jednovýberovým t-testom

Po individuálnej selekcii testov, v ktorých bola (podľa výrazne nízkeho hodnotenia respondenta) identifikácia správnej stopy príliš jednoduchá, boli relevantné

r hodnoty každého respondenta podrobené jednovýberovému, jednostrannému t-testu. T-testy boli vykonané na hladine významnosti 95 %.

Tab. 6.2: Štatistické hypotézy t-testu

hypotéza	$H_0 : \mu = 0$	$H_1 : \mu < 0$
p hodnota	$< 0,05$	$\geq 0,05$

V tabulke 6.2 sú zobrazené dve možné hypotézy spolu s p hodnotami, ktoré prináležia ich **zamietnutiu**. Úspešný t-test znamenal zamietnutie H_0 v prospech H_1 . Inými slovami – za predpokladu, že vyšla v respondentovom t-teste p hodnota nižšia ako 0,05, bolo možné tvrdiť, že signifikantne často identifikoval hľadanú stopu správne. V opačnom prípade bol respondent vyradený.

T-testy boli vykonávané pri $\Delta=2$ a konečný sumár pred ANOVA vyhodnotením vyzeral nasledovne:

Tab. 6.3: Výsledok selekcie respondentov

Zúčastnení	34
Vyradení kvôli prídavnej korekcii	3
Neúspešní v t-teste	6
Úspešní v t-teste (vhodní pre ANOVA)	25

6.2 Štatistické vyhodnotenie a prezentácia výsledkov

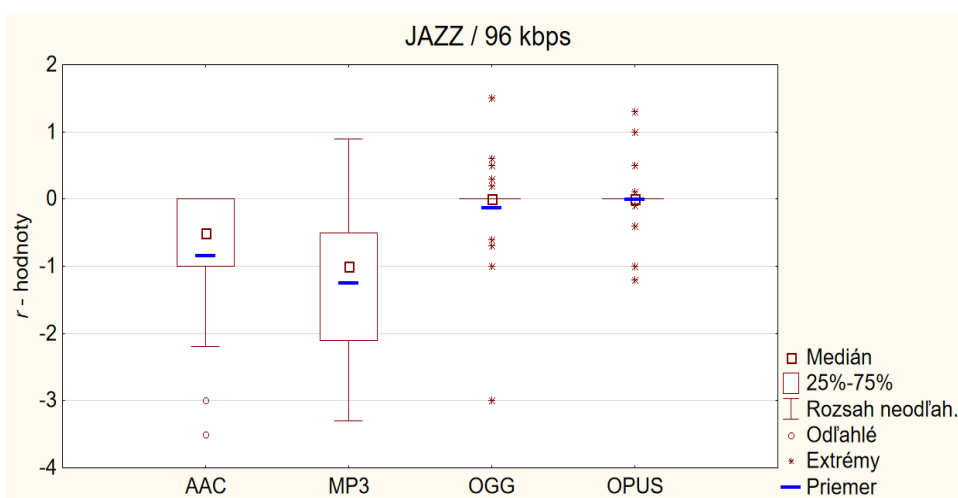
Štatistické vyhodnotenie prebiehalo pomocou analýzy rozptylu s opakovanými meraniami (ANOVA Repeated Measures) v programe STATISTICA. Táto metóda bola vhodnejšia, nakoľko musela byť vo výsledkoch zohľadnená korelácia dát jedného respondenta. Vyhodnocované boli opäť rozdielové r hodnoty respondentov (viď 6.1.2). Každá kapitola obsahuje krabicový graf pre prvotnú vizualizáciu r hodnôt – tieto grafy prezentujú rozdelenie dát pomocou kvartilov (delia oblasť výskytu r hodnôt na štvrtiny). Medián delí oblasť hodnôt na zhodné polovice, priemer zobrazuje strednú hodnotu r hodnôt. Pod grafmi sa vždy nachádzajú výstupné tabuľky analýz, pomocou ktorých je možné štatisticky vyhodnotiť obsah krabicových grafov. Hlavnými údajmi týchto tabuliek sú pre čitateľa p hodnoty. Všetky nasledujúce testy boli vykonané pri hladine významnosti 95 %.

Tab. 6.4: Logika vyhodnocovania p hodnôt v štatistickej analýze

p hodnota	$< 0,05$	$\geq 0,05$
zamietnutie H_0	1	0
príklad p hodnoty	0,002	0,325

6.2.1 Vyhodnotenie výsledkov – ukážka JAZZ

Výsledky kodekov pri bitovej rýchlosti 96 kbps



Obr. 6.2: Rozloženie r -hodnôt kodekov (JAZZ / 96 kbps)

Z grafu 6.2 je zrejmé, že oblasť rozptylu (50 % dát v okolí mediánu) r hodnôt AAC a MP3 je podstatne väčšia ako pri OGG a OPUS, ktorých r hodnoty sú menej variabilné. OGG a OPUS majú okrem veľmi úzkej oblasti rozptylu aj medián blízko nule – je evidentné, že väčšina r hodnôt bola (takmer) nulová. Podľa vertikálneho prieniku krabicových častí je tiež možné predpokladať výskyt signifikantných rozdielov medzi kodekmi (nerovnosť stredných hodnôt súborov ich r hodnôt).

Tab. 6.5: Vplyv faktoru KODEK na výsledky r -hodnôt (JAZZ / 96 kbps)

	SČ	Stupne voľnosti	PČ	F	p
KODEK	23,35772	3	7,78591	11,08797	0,000006
Chyba	46,34478	66	0,70219		

Znenie nulovej hypotézy pred analýzou, ktorej výsledky sú zaznamenané v tabulke 6.5:

- stredné hodnoty r hodnôt sú pre všetky kodeky rovnaké

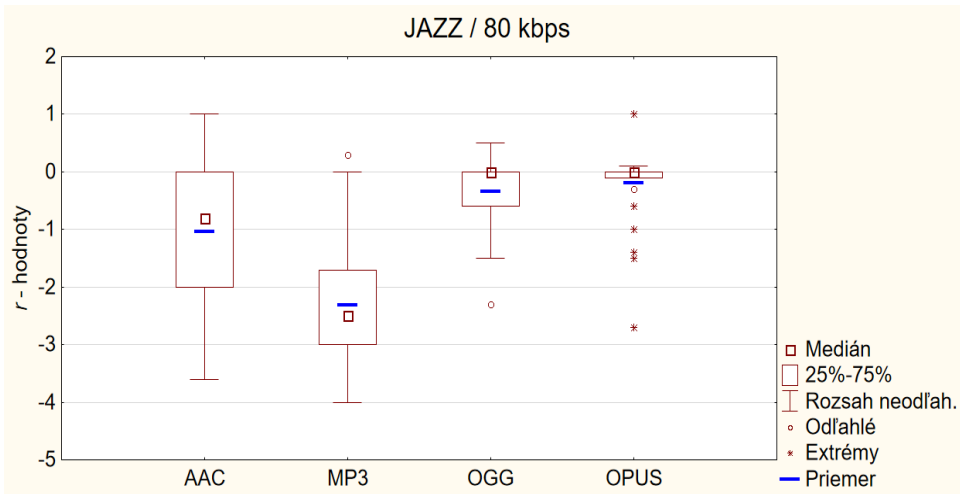
Tabuľka teda na základe logiky z tabuľky 6.4 nastoluje skutočnosť, že faktor „KODEK“ má vplyv na rozdielnosť stredných hodnôt r hodnôt kodekov (hypotéza sa zamietá).

Tab. 6.6: Signifikantné rozdiely medzi kodekmi (JAZZ / 96 kbps)

	AAC	MP3	OGG	OPUS
AAC		0,356251	0,039922	0,006670
MP3	0,356251		0,000380	0,000171
OGG	0,039922	0,000380		0,914902
OPUS	0,006670	0,000171	0,914902	

Na základe výsledkov predošlej tabuľky bolo potrebné vykonať ďalší krok – štatisticky porovnať stredné hodnoty r hodnôt všetkých možných dvojíc kodekov, resp. dokázať, že dvojice nemajú rovnaké stredné hodnoty – rozdiel je signifikantný. V tabulke 6.6 sú dáta usporiadané ako v matici 4×4 . Výskyt signifikantného rozdielu (hrubo zvýraznená p hodnota) je definovaná pozíciou v stĺpci a riadku. Tabuľka dokazuje prítomnosť signifikantných rozdielov presne tak, ako je to možné hrubo, no intuitívne odhadnúť z krabicového grafu 6.2. Z tabuľky teda vyplýva, že neexistuje štatisticky významný rozdiel medzi dvojicou AAC a MP3, ani medzi OGG a OPUS, signifikantný rozdiel však existuje medzi týmito skupinami.

Výsledky kodekov pri bitovej rýchlosti 80 kbps



Obr. 6.3: Rozloženie r -hodnôt kodekov (JAZZ / 80 kbps)

Z grafu 6.3 sú zreteľné posuvy mediánu a priemeru na nižšie hodnoty r hodnôt. Najväčší prepad týchto hodnôt je možné pozorovať u MP3. Navyše má MP3 ako jediný kodek v tomto grafe zúženú krabicovú oblasť (ostatným sa rozšírila), teda respondenti pri MP3 často udávali exaktne nízke hodnotenie.

Tab. 6.7: Vplyv faktoru KODEK na výsledky r -hodnôt (JAZZ / 80 kbps)

	SČ	Stupne voľnosti	PČ	F	p
KODEK	70,46960	3	23,48987	27,14484	0,000000
Chyba	62,30540	72	0,86535		

Stredné hodnoty medzi kodekmi nie sú rovnaké (faktor „KODEK“ má vplyv) – je potrebné porovnať dvojice.

Tab. 6.8: Signifikantné rozdiely medzi kodekmi (JAZZ / 80 kbps)

	AAC	MP3	OGG	OPUS
AAC		0,000183	0,046461	0,010169
MP3	0,000183		0,000150	0,000150
OGG	0,046461	0,000150		0,942890
OPUS	0,010169	0,000150	0,942890	

V ukážke JAZZ sa teda pri bitovej rýchlosti 80 kbps vzájomne líšia všetky možné dvojice okrem OPUS a OGG.

Vplyv faktorov BITRATE a KODEK – ukážka JAZZ

Keďže program STATISTICA neumožňuje v rámci analýzy rozptylu s opakovanými meraniami skúmať vplyv viacerých faktorov na stredné hodnoty r hodnôt, musela byť pre tento účel použitá metóda lineárnych modelov.

Tab. 6.9: Vplyv faktorov BITRATE, KODEK a ich interakcie na výsledky r -hodnôt (JAZZ)

	SČ	Stupne voľnosti	PČ	F	p
BITRATE	8,22397	1	8,22397	14,99942	0,000822
Chyba	12,06228	22	0,54829		
KODEK	77,97147	3	25,99049	25,73807	0,000000
Chyba	66,64728	66	1,00981		
BITRATE*KODEK	5,61321	3	1,87107	3,43932	0,021708
Chyba	35,90554	66	0,54402		

Nulové hypotézy analýzy, ktorej výsledky sú zaznamenané v tabuľke 6.9:

- stredné hodnoty r hodnôt sú pre všetky bitové rýchlosti rovnaké
- stredné hodnoty r hodnôt sú pre všetky kodeky rovnaké
- stredné hodnoty r hodnôt nie sú ovplyvnené kombináciou bitovej rýchlosti a kodeku.

Z tabuľky 6.9 vyplýva, že sa nepotvrdila ani jedna z nulových hypotéz – vplyv na výsledky ukážky JAZZ má bitová rýchlosť, kodek a dokonca aj tretí faktor, ktorým je ich interakcia. Interakcia faktorov znamená, že na výsledky analýzy má výrazný vplyv aj to, akým spôsobom korelujú základné dva faktory.

Tab. 6.10: Signifikantné rozdiely medzi kodekmi pri zohľadnení faktorov BITRATE, KODEK a ich interakcie (JAZZ)

BITRATE		96	96	96	96	80	80	80	80
	KODEK	AAC	MP3	OGG	OPUS	AAC	MP3	OGG	OPUS
96	AAC		0,569858	0,054367	0,006531	0,954766	0,000127	0,320721	0,141415
96	MP3	0,569858		0,000239	0,000132	0,993687	0,000431	0,002375	0,000646
96	OGG	0,054367	0,000239		0,995445	0,002092	0,000126	0,991441	0,999922
96	OPUS	0,006531	0,000132	0,995445		0,000258	0,000126	0,772241	0,945512
80	AAC	0,954766	0,993687	0,002092	0,000258		0,000140	0,025563	0,007391
80	MP3	0,000127	0,000431	0,000126	0,000126	0,000140		0,000126	0,000126
80	OGG	0,320721	0,002375	0,991441	0,772241	0,025563	0,000126		0,999892
80	OPUS	0,141415	0,000646	0,999922	0,945512	0,007391	0,000126	0,999892	

Tabuľka 6.10 zobrazuje prehľad signifikantných rozdielov dvojíc kodekov pri zohľadnení všetkých troch faktorov v ukážke JAZZ. Z tejto tabuľky je tiež možné vyčítať, či sa líši v rámci žánru kvalita kodeku pri použití bitovej rýchlosti 96 a 80 kbps (z tejto tabuľky vyplýva, že sa líši iba MP3). Je dôležité podotknúť, že p hodnoty dvojice kodekov pri rovnakej bitovej rýchlosti v tejto tabuľke sú odlišné od p hodnôt z predošlých analýz ukážky JAZZ práve kvôli zohľadneniu všetkých možných vplyvov (a najmä ich interakcie). Logika interpretácie hodnôt tabuľky je zhodná s predošlými maticovými tabuľkami v kapitole JAZZ, tzn. vychádza z prehľadu tabuľky 6.4.

Tab. 6.11: Zostupné zoradenie kodekov podľa dosiahnutých výsledkov (JAZZ)

	96 kbps	80 kbps
1.	OPUS	OPUS
2.	OGG	OGG
3.	AAC	AAC
4.	MP3	MP3

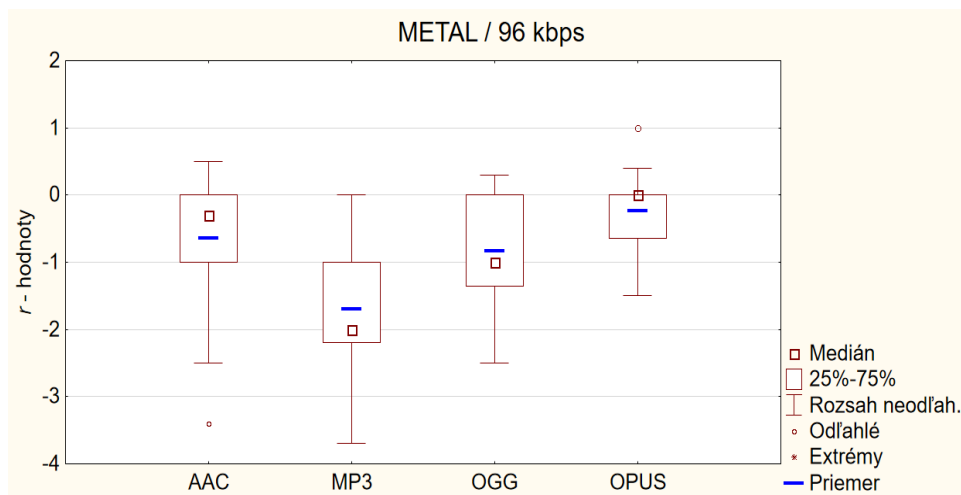
Pri zohľadnení vykonaných štatistických analýz ukážky JAZZ (ktoré však určujú iba zhodu alebo odlišnosť medzi skupinami) a obsahu krabicových grafov (kde je priamo zobrazené rozmiestnenie zozbieraných hodnotení, resp. ich r hodnôt) je možné zoradiť zvukovú kvalitu kodekov tak, ako je to demonštrované v tabuľke 6.11. Pri štatistickom vyhodnocovaní sa mohlo stať, že nebol štatisticky dokázaný signifikantný rozdiel medzi konkrétnou dvojicou kodekov – mohlo sa zdať, že nie je možné určiť ich konkrétnu pozíciu v tabuľke 6.11 (udialo sa napr. pri dvojici AAC a MP3 pri 96 kbps (viď 6.5)). To, akým spôsobom budú AAC a MP3 pridelené na 3. a 4. miesto tabuľky 6.11, bolo možné určiť dvomi spôsobmi:

- Relatívnym porovnaním p hodnôt AAC a MP3 napr. s OGG v tabuľke 6.5. Napriek tomu, že sú MP3 aj AAC štatisticky rozdielne od OGG, má OGG s MP3 podstatne nižšiu p hodnotu (0,000380) ako s AAC (0,039922). To znamená, že AAC vykazuje menší rozdiel (má p hodnotu bližšiu hranici prijatia hypotézy o rovnosti, teda k hodnote 0,05) voči štatisticky preukázateľne kvalitnejšiemu kodeku OGG, preto musí byť radené pred MP3.
- Jednoduchým porovnaním priemerov v grafe pre aktuálnu bitovú rýchlosť – vyššia hodnota znamená vyššie hodnotenia respondentov (vyššiu subjektívnu posluchovú kvalitu).

Týmto spôsobom sú vyhodnotenú tabuľky s konečným verdiktom pri oboch nasledujúcich žánroch.

6.2.2 Vyhodnotenie výsledkov – ukážka METAL

Výsledky kodekov pri bitovej rýchlosti 96 kbps



Obr. 6.4: Rozloženie r -hodnôt kodekov (METAL / 96 kbps)

Graf 6.4 (na rozdiel od grafov ukážky JAZZ) neobsahuje extrémne hodnoty, rovnako tak má len zopár odľahlých hodnôt (r hodnoty sú v rozmedzí krabicových častí a „fúzov“ grafu). Graf taktiež naznačuje, že AAC tentokrát predčí OGG.

Tab. 6.12: Vplyv faktoru KODEK na výsledky r -hodnôt (METAL / 96 kbps)

	SČ	Stupne voľnosti	PČ	F	p
KODEK	23,55000	3	7,85000	15,95627	0,00000
Chyba	32,47000	66	0,49197		

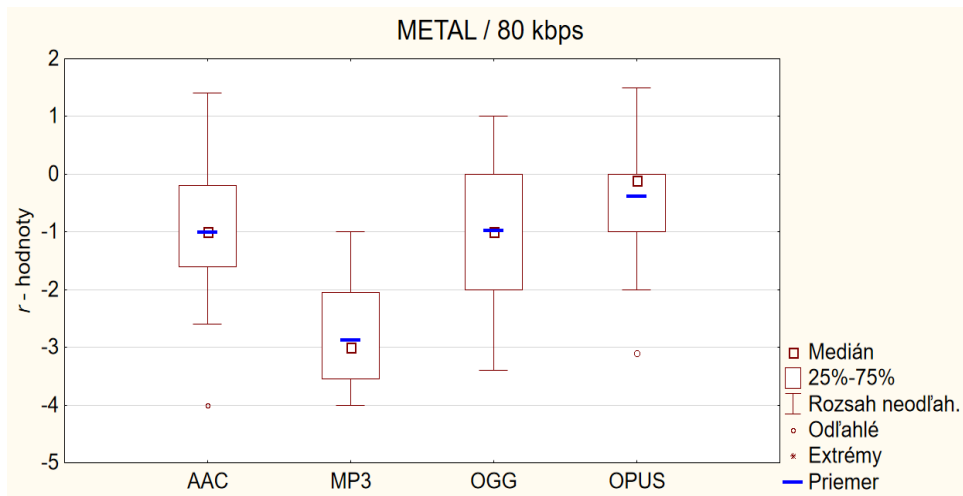
Faktor „KODEK“ má vplyv na stredné hodnoty – je potrebné porovnať dvojice.

Tab. 6.13: Signifikantné rozdiely medzi kodekmi (METAL / 96 kbps)

	AAC	MP3	OGG	OPUS
AAC		0,000181	0,551026	0,309329
MP3	0,000181		0,003504	0,000152
OGG	0,551026	0,003504		0,016158
OPUS	0,309329	0,000152	0,016158	

V ukážke METAL sa pri bitovej rýchlosti 96 kbps líši MP3 od všetkých zvyšných a OGG sa líši od OPUSU.

Výsledky kodekov pri bitovej rýchlosti 80 kbps



Obr. 6.5: Rozloženie r -hodnôt kodekov (METAL / 80 kbps)

Pri znížení bitovej rýchlosti na 80 kbps graf zobrazuje mierne zhoršenie pri AAC, výrazné zhoršenie pri MP3, OGG a OPUS takmer bez zmeny.

Tab. 6.14: Vplyv faktoru KODEK na výsledky r -hodnôt (METAL / 80 kbps)

	SČ	Stupne voľnosti	PČ	F	p
KODEK	52,2072	3	17,4024	15,1176	0,000000
Chyba	58,7078	51	1,1511		

Faktor „KODEK“ má vplyv na stredné hodnoty – je potrebné porovnať dvojice.

Tab. 6.15: Signifikantné rozdiely medzi kodekmi (METAL / 80 kbps)

	AAC	MP3	OGG	OPUS
AAC		0,000177	0,974959	0,828137
MP3	0,000177		0,000227	0,000164
OGG	0,974959	0,000227		0,583646
OPUS	0,828137	0,000164	0,583646	

V ukážke METAL sa pri bitovej rýchlosti 80 kbps líši MP3 od všetkých zvyšných, OPUS, OGG a AAC sa medzi sebou nelíšia.

Vplyv faktorov – ukážka METAL

Tab. 6.16: Vplyv faktorov BITRATE, KODEK a ich interakcie na výsledky *r*-hodnôt (METAL)

	SČ	Stupne volnosti	PČ	F	p
BITRATE	11,9601	1	11,9601	34,48354	0,000018
Chyba	5,8962	17	0,3468		
KODEK	59,0735	3	19,6912	22,67176	0,000000
Chyba	44,2952	51	0,8685		
BITRATE*KODEK	7,0541	3	2,3514	3,09194	0,035045
Chyba	38,7847	51	0,7605		

Faktory „BITRATE“, „KODEK“ aj ich interakcia majú vplyv na stredné hodnoty kodekov – je potrebné porovnať dvojice.

Tab. 6.17: Signifikantné rozdiely medzi kodekmi pri zohľadnení faktorov BITRATE, KODEK a ich interakcie (METAL)

BITRATE		96	96	96	96	80	80	80	80
	KODEK	AAC	MP3	OGG	OPUS	AAC	MP3	OGG	OPUS
96	AAC		0,050909	0,978596	0,983035	0,894869	0,000131	0,612760	0,999997
96	MP3	0,050909		0,370248	0,003875	0,587674	0,000794	0,879921	0,092849
96	OGG	0,978596	0,370248		0,575099	0,999971	0,000131	0,988331	0,996456
96	OPUS	0,983035	0,003875	0,575099		0,359279	0,000131	0,136287	0,937915
80	AAC	0,894869	0,587674	0,999971	0,359279		0,000132	0,999555	0,963807
80	MP3	0,000131	0,000794	0,000131	0,000131	0,000132		0,000135	0,000131
80	OGG	0,612760	0,879921	0,988331	0,136287	0,999555	0,000135		0,767060
80	OPUS	0,999997	0,092849	0,996456	0,937915	0,963807	0,000131	0,767060	

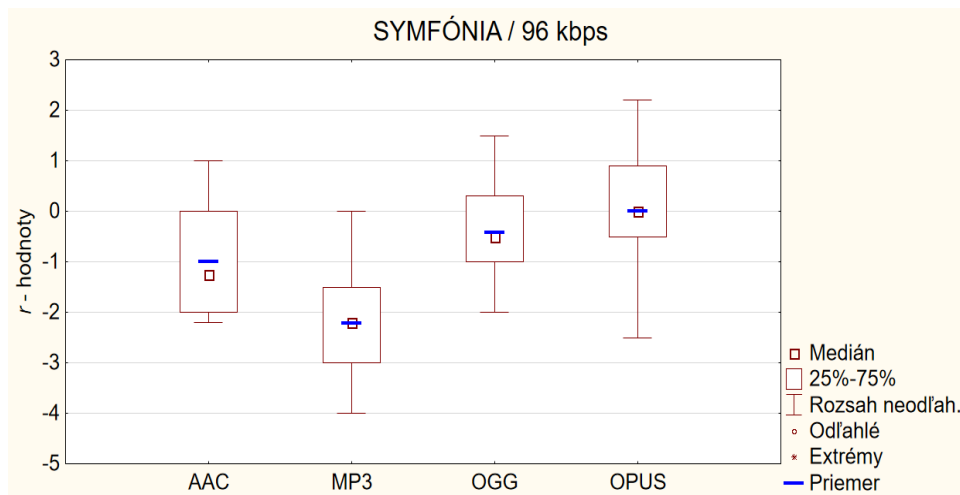
Pri zohľadnení faktorov a ich interakcie je jediným kodekom, ktorý v rámci ukážky METAL vykazuje signifikantné rozdiely medzi bitovými rýchlosťami, iba MP3.

Tab. 6.18: Zostupné zoradenie kodekov podľa dosiahnutých výsledkov (METAL)

	96 kbps	80 kbps
1.	OPUS	OPUS
2.	AAC	AAC
3.	OGG	OGG
4.	MP3	MP3

6.2.3 Vyhodnotenie výsledkov – ukážka SYMFÓNIA

Výsledky kodekov pri bitovej rýchlosti 96 kbps



Obr. 6.6: Rozloženie r -hodnôt kodekov (SYMFÓNIA / 96 kbps)

Graf 6.6 je podobný grafom predošlých ukážok pri 96 kbps. OPUS dominuje, MP3 výrazne zaostáva, OGG sa pri tejto ukážke zdá byť oproti AAC kvalitnejšie.

Tab. 6.19: Vplyv faktoru KODEK na výsledky r -hodnôt (SYMFÓNIA / 96 kbps)

	SČ	Stupne voľnosti	PČ	F	p
KODEK	66,50852	3	22,16951	23,19526	0,000000
Chyba	60,21398	63	0,95578		

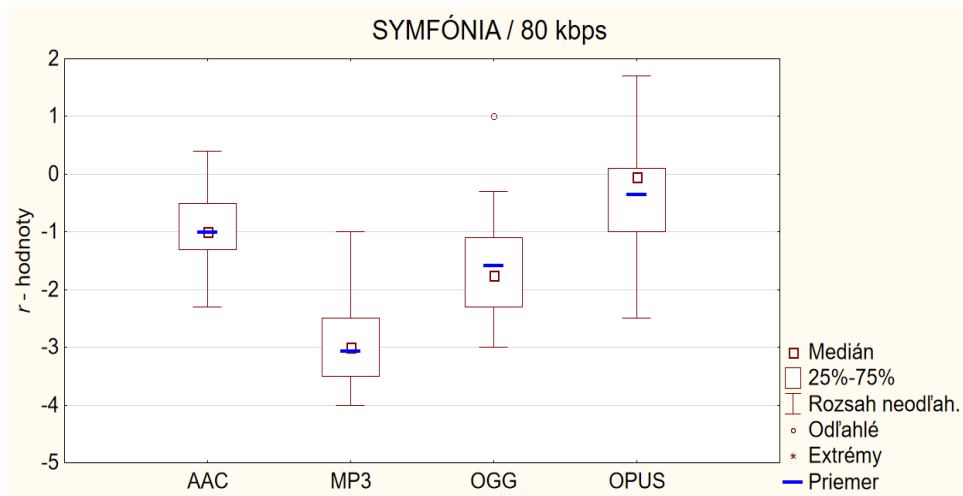
Faktor „KODEK“ má vplyv na stredné hodnoty – je potrebné porovnať dvojice.

Tab. 6.20: Signifikantné rozdiely medzi kodekmi (SYMFÓNIA / 96 kbps)

	AAC	MP3	OGG	OPUS
AAC		0,000396	0,310845	0,003779
MP3	0,000396		0,000154	0,000154
OGG	0,310845	0,000154		0,266908
OPUS	0,003779	0,000154	0,266908	

V ukážke SYMFÓNIA sa pri bitovej rýchlosti 96 kbps líši MP3 od všetkých zvyšných a OPUS sa líši od AAC.

Výsledky kodekov pri bitovej rýchlosti 80 kbps



Obr. 6.7: Rozloženie r -hodnôt kodekov (SYMFÓNIA / 80 kbps)

Graf 6.7 zobrazuje nepatrné zlepšenie pri AAC (užšia krabicová časť od 96 kbps), MP3 sa zhoršilo, OGG vykazuje markantné zhoršenie a OPUS je takmer bez zmeny.

Tab. 6.21: Vplyv faktoru KODEK na výsledky r -hodnôt (SYMFÓNIA / 80 kbps)

	SČ	Stupne voľnosti	PČ	F	p
KODEK	85,2314	3	28,4105	42,3858	0,000000
Chyba	38,2061	57	0,6703		

Faktor „KODEK“ má vplyv na stredné hodnoty – je potrebné porovnať dvojice.

Tab. 6.22: Signifikantné rozdiely medzi kodekmi (SYMFÓNIA / 80 kbps)

	AAC	MP3	OGG	OPUS
AAC		0,000158	0,089674	0,021834
MP3	0,000158		0,000161	0,000158
OGG	0,089674	0,000161		0,000164
OPUS	0,021834	0,000158	0,000164	

V ukážke SYMFÓNIA sa pri bitovej rýchlosti 80 kbps líšia všetky dvojice okrem OGG a AAC.

Vplyv faktorov – ukážka SYMFÓNIA

Tab. 6.23: Vplyv faktorov BITRATE, KODEK a ich interakcie na výsledky *r*-hodnôt (SYMFÓNIA)

	SČ	Stupne voľnosti	PČ	F	p
BITRATE	13,2250	1	13,2250	30,5409	0,000025
Chyba	8,2275	19	0,4330		
KODEK	134,4328	3	44,8109	39,4811	0,000000
Chyba	64,6948	57	1,1350		
BITRATE*KODEK	7,2860	3	2,4287	4,7651	0,004939
Chyba	29,0515	57	0,5097		

Faktory „BITRATE“, „KODEK“ aj ich interakcia majú vplyv na stredné hodnoty kodekov – je potrebné porovnať dvojice.

Tab. 6.24: Signifikantné rozdiely medzi kodekmi pri zohľadnení faktorov BITRATE, KODEK a ich interakcie (SYMFÓNIA)

BITRATE		96	96	96	96	80	80	80	80
	KODEK	AAC	MP3	OGG	OPUS	AAC	MP3	OGG	OPUS
96	AAC		0,000137	0,495399	0,002011	0,999977	0,000131	0,059443	0,059443
96	MP3	0,000137		0,000131	0,000131	0,000155	0,014102	0,143209	0,000131
96	OGG	0,495399	0,000131		0,333856	0,298335	0,000131	0,000233	0,957419
96	OPUS	0,002011	0,000131	0,333856		0,000740	0,000131	0,000131	0,936129
80	AAC	0,999977	0,000155	0,298335	0,000740		0,000131	0,130045	0,024759
80	MP3	0,000131	0,014102	0,000131	0,000131	0,000131		0,000132	0,000131
80	OGG	0,059443	0,143209	0,000233	0,000131	0,130045	0,000132		0,000133
80	OPUS	0,059443	0,000131	0,957419	0,936129	0,024759	0,000131	0,000133	

Pri zohľadnení faktorov a ich interakcie vykazujú v rámci ukážky SYMFÓNIA rozdiely medzi bitovými rýchlosťami MP3 a OGG.

Tab. 6.25: Zostupné zoradenie kodekov podľa dosiahnutých výsledkov (SYMFÓNIA)

	96 kbps	80 kbps
1.	OPUS	OPUS
2.	OGG	AAC
3.	AAC	OGG
4.	MP3	MP3

6.3 Meranie výpočtovej náročnosti

Hlavným údajom vypovedajúcim o výpočtovej náročnosti bol čas potrebný na kódovanie a dekódovanie súborov. Pre toto meranie boli použité ukážky a parametre kódovania totožné s posluchovými testami, pre zaujímavosť bola pridaná bitová rýchlosť 128 kbps. Skript pre toto meranie bol napísaný v programe Python 3.9 a spúšťal sa pomocou cmd. Skript pracoval s ffmpeg a pre každý test mal zadané kodeky, s ktorými pracoval, a počet opakovaní pre kódovanie a dekódovanie každým z nich (v našom prípade 100 opakovaní). V každom jednom meraní mu potom bolo potrebné zadať názov spracovávaného zvukového súboru a požadovanú bitovú rýchlosť. Po spustení skript zakódoval prvým zo zadaných kodekov zvolený zvukový súbor a zaznamenal čas. Po 100 opakovaní vykonal obdobným spôsobom meranie dekódovania. Po vykonaní 100 opakovaní kódovania a dekódovania sa skript presunul ku spracovávaniu zadaného zvukového súboru ďalším kodekom. Po zmeraní a zaznamenaní časov pre všetky 4 kodeky skript vypísal všetky časy na štandardný výstup, zo zaznamenaných časov pre každých 100 opakovaní vypočítal aritmetický priemer a zmazal súbory, ktoré vytvoril z pevného disku.

Konfigurácia počítača, na ktorom bolo toto meranie vykonané vyzerala nasledovne:

- Windows 10 Pro
- CPU Intel Core i7-4720HQ, 2.6 GHz
- RAM 8 GB, 1600 MHz.

Tab. 6.26: Výsledky merania časov kódovania

ukážka	JAZZ			METAL			SYMFÓNIA		
bitrate [kbps]	80	96	128	80	96	128	80	96	128
MP3	8,53	8,67	9,14	8,15	8,63	9,03	8,36	8,25	9,13
AAC	5,81	5,49	5,17	5,91	5,43	6,45	6,89	6,60	5,70
Opus	2,13	2,23	2,46	2,23	2,23	2,29	2,12	1,86	2,02
Vorbis	3,47	3,42	2,93	3,45	3,44	2,85	3,27	3,21	2,61

Tab. 6.27: Výsledky merania časov dekodovania

ukážka	JAZZ			METAL			SYMFÓNIA		
bitrate [kbps]	80	96	128	80	96	128	80	96	128
MP3	0,54	0,54	0,56	0,55	0,56	0,58	0,53	0,54	0,56
AAC	0,40	0,40	0,40	0,41	0,41	0,42	0,38	0,38	0,39
Opus	0,70	0,72	0,78	0,71	0,73	0,79	0,68	0,71	0,77
Vorbis	0,44	0,43	0,45	0,45	0,45	0,46	0,39	0,40	0,41

Informácie v tabuľkách sú udané v percentách reálneho času, podľa vzorca:

$$\frac{\text{zmeraný čas}}{\text{dĺžka nahrávky}} \cdot 100 [\%] \quad (6.5)$$

V tabuľke s hodnotami kódovania vidíme, že najdlhší čas potrebuje pre kódovanie MP3. Dôvodom je fakt, že spomedzi vybraných kodekov sa jedná o najstarší kodek. Zároveň zo sekcie 2.6.1 vyplýva, že proces kódovania v MP3 je podstatne viac mechanický, ako u zvyšných kodekov. Tak, ako sa predpokladalo, najrýchlejšie kódovanie vykazuje Opus, ktorý je primárne navrhnutý pre kódovanie v reálnom čase.

Časy zaznamenané v tabuľke dekodovania sú podstatne nižšie, čo odpovedá predpokladu, že pri dekodovaní prebieha v kodekoch podstatne menej matematických operácií než pri kódovaní. Priemerné časy medzi kodekmi navyše nevykazujú tak odlišné rozdiely, ako v tabuľke kódovania. Pri Opuse je predpokladaná príčina najdlhších hodnôt časov dekodovania spojená s rekonštrukciou energie pásiem pomocou deemfázového a perceptuálneho filtra.

Záver

Primárnym cieľom práce bolo vytvorenie konceptu posluchových testov, pomocou ktorého bude možné vyhodnotiť subjektívnu posluchoвую kvalitu vybraných kodekov používaných pre kódovanie hudobných súborov. Sekundárnym cieľom bola analýza činnosti a popis funkcií zvolených kodekov, a s tým súvisiace vyhodnotenie ich výpočtovej náročnosti. Oba ciele sa podarilo dosiahnuť.

Práca začína popisom vlastností ľudského sluchu nasledovaná stručným úvodom do problematiky audio kódovania spolu s metodikou výberu testovanej štvorice kodekov a spomínaným rozborom každého z nich. Zvolené boli kodeky Opus, Ogg Vorbis, AAC a MP3. Ďalšia kapitola sa zaoberá posluchoвыми testami – popisom rôznych metód, zhrnutie normatívnych odporúčaní pre zvolený koncept testu (ABC/HR) a konečná podoba navrhnutého posluchového testu.

Nasledujúca kapitola prechádza ku praktickej časti práce – je tu popísaný proces vytvorenia programového materiálu a činnosť testovacieho rozhrania. Testovacie prostredie malo podobu dvoch webstránok (inštruktáž a stránka s posluchoвыми testami), na jeho implementáciu boli použité programovacie jazyky HTML, CSC, Python a Javascript. Poslednou kapitolou sú už samotné výsledky práce pozostávajúce zo štatistického vyhodnotenia a výsledkov merania výpočtovej náročnosti.

Na základe výsledkov práce možno za najkvalitnejší kodek považovať Opus, ktorý dosiahol pri posluchoových testoch najlepšie výsledky. Za týmto dominantným výkonom stojí veľmi zložitá implementácia kodeku obsahujúca množstvo prídavných psychoakustických funkcií, a tiež fakt, že sa spomedzi zvolenej štvorice jedná o najnovší kodek. Pri Opuse bola taktiež zameraná najnižšia priemerná doba kódovacieho procesu, v dekódovaní skončil ale na chvoste výberu – vzhľadom na veľmi precízne a zložité dekódovanie bitového toku je tento výsledok však pochopiteľný.

Naopak ako najhorší kodek vyšiel v globále MP3 – dosiahol najhoršie výsledky prakticky v každom posluchovom teste, boli mu namerané najdlhšie časy kódovania, ktoré vyplývajú z veľmi mechanického spôsobu, akým kóduje zvukové súbory. Pri zvážení faktu, že sa jedná o najstarší kodek z výberu, ktorý sa stal štandardom strastového kódovania pred viac ako 30 rokmi, je tento výsledok pochopiteľný. Z hľadiska posluchovej kvality, ktorú ponúka, je teda MP3 v dnešnej dobe skôr prežitkom. Treba však podotknúť, že je stále často využívaný, najmä kvôli natívnej podpore väčšiny multimediálnych prehrávačov.

Ogg Vorbis, ktorého hlavnou ambíciou pri zrode bolo nahradenie MP3, s príslubom lepších zvukových výsledkov pri nižších bitových rýchlostiach, obhájal svoj účel. V posluchoových testoch bol zakaždým výrazne kvalitnejší ako MP3, dosahoval veľmi podobné výsledky ako AAC (v ukážke JAZZ prekonal AAC pri obidvoch bitových rýchlostiach, v ukážke METAL nad ním naopak viedlo AAC a v SYMFÓNII AAC zdolal len pri 96 kbps), na mimoriadne kvalitný Opus však nikdy nestačil. V meraní výpočtovej náročnosti sa zaradil na druhé miesto. Aj keď vzhľadom na výsledky práce sú jeho výkony skôr priemerné, rozhodne sa jedná o kodek s veľmi originálnym prístupom ku kódovaniu, ktorý dodnes nachádza využitie v komerčnej zvukovej sfére.

Profil vyhodnotenia AAC je veľmi blízko Vorbisu – v posluchoových testoch sa umiestňoval na stredných priečkach, v meraní výpočtovej náročnosti obsadil tretiu priečku – oba tieto výsledky sú však vzhľadom na použitý profil nízkej zložitosti (Low Complexity profile) pochopiteľné, AAC pracujúce v hlavnom profile by istotne dosiahlo priaznivejšie výsledky. S Vorbisom ho okrem výsledkov zvukovej kvality spája tiež (naplnený) cieľ prekonať MP3. Historicky bola ale jeho motivácia od Vorbisu dosť odlišná – kým Vorbis, pochádzajúci od konkurenčnej spoločnosti, mal za úlohu vyslovene nahradiť MP3 ako nový štandard, úlohou AAC bolo nahradenie MP3 skôr akýmsi evolučným krokom v napredovaní línie kodekov rodiny MPEG.

Literatúra

- [1] Andreas Spanias, Ted Painter, Venkatraman Atti. *Audio Signal Processing and Coding*. 1. Wiley, Hoboken, New Jersey, 2007. ISBN 978-0-471-79147-8.
- [2] Teri A. Hamill, Lloyd L. Price. *The Hearing Sciences*. 2. Plural Publishing Inc., San Diego, USA , 2014. ISBN 9781597565400.
- [3] Jiří Schimmel. *Elektroakustika*. 1. Brno (Česká Republika): Vysoké učení technické v Brně Fakulta elektrotechniky a komunikačních technologií Ústav telekomunikací, 2016. ISBN 978-80-214-4716-5.
- [4] Jayaraman J. Thiagarajan, Andreas Spanias. *Analysis of the MPEG-1 Layer III (MP3) algorithm Using MATLAB*. 1. Morgan and Claypool, Arizona State Universit., 2012. ISBN 978-1-60845-801-1.
- [5] Mark Harris. *CBR vs VBR Encoding*. citované 10.12.2020 <<https://www.lifewire.com/difference-between-cbr-and-vbr-encoding-2438423>>
- [6] Robin Whittle. *Lossless Compression of Audio*. citované 10.12.2020 <<https://www.firstpr.com.au/audiocomp/lossless/>>
- [7] A. H. Hussein, S. Sh. Mahmud, R. J. Mohammed. *Compression Using Proposed Enhanced Run Length Encoding Algorithm* Ibn AL- Haitham Journal For Pure and Applied Science, [S.l.], v. 24, n. 1, may 2017. ISSN 2521-3407.
- [8] David Salomon. *A Concise Introduction to Data Compression*. Springer-Verlag, London, UK, 2008. ISBN 978-1-84800-072-8.
- [9] Kim Bongjun and Zafar Rafii. *Lossy Audio Compression Identification*. European Signal Processing Conference (EUSIPCO) (2018): 2459-2463.
- [10] Jürgen Herre, Sascha Dick. *Psychoacoustic Models for Perceptual Audio Coding - A Tutorial Review*. Erlangen, Germany, Applied Sciences 9, 2019. ISSN: 2076-3417
- [11] Jürgen Herre. *From Joint Stereo to Spatial Audio Coding - Recent Progress and Standardization*. Naples, Italy, Proc of the 7th Int.Confereneon Digital Audio Effects, 2004.
- [12] *FFmpeg*. citované 18.5.2021 <<https://www.ffmpeg.org/>>

- [13] INTERNATIONAL STANDARD ISO/IEC13818-7 *Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC)*. 2006
- [14] Marina Bosi and Richard E. Goldberg *Introduction to Digital Audio Coding and Standards*. Kluwer Academic Publishers, 2003, ISBN: 1-4020-7357-7
- [15] *Vorbis*. citované 11.12.2020 <<https://sk.wikipedia.org/wiki/Vorbis>>
- [16] Pavol Adam. *Úvod do metód spracovania zvuku v súčasnom multimediálnom prostredí - Časť 4.9: OGG Vorbis*. citované 11.12.2020 <http://zvuk.atrip.sk/index.php?site=4_9>
- [17] *Vorbis*. citované 29.4.2021 <<https://entropia.de/images/c/cc/GPN12-Introduction-to-Vorbis.pdf>>
- [18] Xiph.Org Foundation *Vorbis I specification*. July 4, 2020. <https://xiph.org/vorbis/doc/Vorbis_I_spec.html>
- [19] Xiph.Org Foundation *Ogg Vorbis stereo-specific channel coupling discussion*. <<https://xiph.org/vorbis/doc/stereo.html>>
- [20] Jean-Marc Valin, Gregory Maxwell, Timothy B. Terriberry, Koen Vos. *High-Quality, Low-Delay Music Coding in the Opus Codec*. 135th Audio Engineering Society Convention, New York, USA, 2013
- [21] *Opus (audio codec)*. citované 8.12.2020 <[https://commons.wikimedia.org/wiki/Opus_\(audio_codec\)](https://commons.wikimedia.org/wiki/Opus_(audio_codec))>
- [22] Roberto Amorim. *Listening Test Conduction Handbook*. PITA, 2005
- [23] Jon Boley and Michael Lester *Statistical Analysis of ABX Results Using Signal Detection Theory*. New York, USA, 2009.
- [24] RECOMMENDATION ITU-R BS.1534-1 *Method for the subjective assessment of intermediate quality level of coding systems*. 2001-2003
- [25] RECOMMENDATION ITU-R BS.1116-3 *Methods for the subjective assessment of small impairments in audio systems*. 2015
- [26] *T-test*. citované 18.5.2021 <<https://cit.vfu.cz/stat/FVHE/Teorie/Predn3/ttest.htm>>
- [27] *Analýza rozptylu (ANOVA)*. citované 18.5.2021 <http://user.mendelu.cz/drapela/Statisticke_metody/Prezentace/zakladni/Anova.pdf>

Zoznam symbolov, veličín a skratiek

AAC	Advanced Audio Coding
ANOVA	Analysis of Variance
CD	Compact Disc
CELT	Constrained Energy Lapped Transform
dB	Decibel
FFT	Fast Fourier Transform
FLAC	Free Lossless Audio Codec
Hi-Fi	High Fidelity
Hz	Hertz
ITU-R	International Telecommunication Union Radiocommunication
ITU-R BS	International Telecommunication Union Radiocommunication Broadcast Service
KBD	Kaiser-Bessel Derived
kHz	kilohertz
M/S	Middle / Side
MDCT	Modified Discrete Cosine Transform
MPC	Musepack
MPEG	Moving Picture Experts Group
MUSHRA	Multiple Stimuli with Hidden Reference and Anchor
PCM	Pulse Code Modulation
RLC	Run Length Coding
S/M	Signal to Mask Ratio
SBC	Smart Bitrate Control
TNS	Temporal Noise Shaping

WAV Waveform Audio File Format

WMA Windows Media Audio

Zoznam príloh

A Obsah prílohy

76

A Obsah prílohy

/	Obsah aplikácie posluchových testov
├── static	
│ ├── audio.js	funkcionalita stránky posluchových testov
│ ├── audio_intro.js	funkcionalita inštruktážnej stránky
│ ├── javascript.js	funkcionalita tlačidiel
│ ├── silver.jpg	pozadie stránok
│ ├── slider.css	formátovanie vzhľadu bežcov
│ └── stylesheet.css	formátovanie vzhľadu stránok
├── templates	zdrojové kódy vzhľadu stránok
│ ├── finished.html	vzhľad stránky s poďakovaním
│ ├── intro.html	vzhľad inštruktážnej stránky
│ └── rozhranie.html	vzhľad stránky s posluchovými testami
├── app.py	zápis dát na server
├── requirements.txt	zoznam knižníc
└── zozbierané dáta	anonymizované dáta od všetkých 34 respondentov