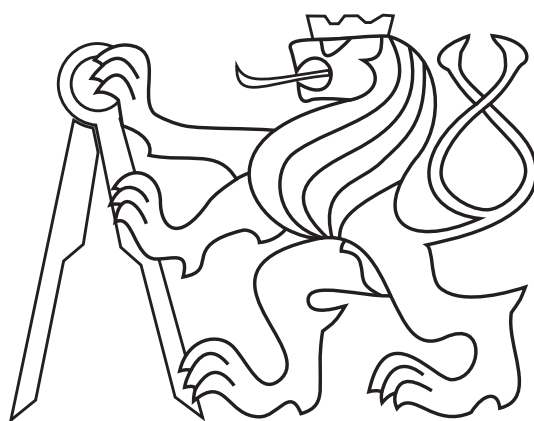


České Učení Vysoké Technické v Praze
Fakulta elektrotechnická
Katedra elektromagnetického pole



Hodnocení kvality zvuku v systémech digitálního rozhlasového vysílání

Audio Quality Assessment of Digital Radio Broadcasting Systems

diplomová práce

Adam Bartyzal

Studijní obor: Rádiová a optická technika

Praha, Květen 2019

Vedoucí práce:

Ing. Karel Ulovec, Ph.D.
Katedra radioelektroniky
Fakulta elektrotechnická
České Učení Vysoké Technické v Praze
Technická 2
160 00 Praha 6
Česká Republika

Copyright © 2019 Adam Bartyzal

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Bartyzal** Jméno: **Adam** Osobní číslo: **425043**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra elektromagnetického pole**
Studijní program: **Elektronika a komunikace**
Studijní obor: **Rádiová a optická technika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Hodnocení kvality zvuku v systémech digitálního rozhlasového vysílání

Název diplomové práce anglicky:

Audio Quality Assessment of Digital Radio Broadcasting Systems

Pokyny pro vypracování:

Uveďte stručný přehled metod pro hodnocení kvality zvukového signálu. Zabývejte se podrobněji metodami s ohledem na možné využití v systémech digitálního rozhlasového vysílání Digital Audio Broadcasting a Digital Radio Mondiale. Připravte, realizujte a vyhodnoťte poslechový test. Zdrojové kódování testovaných vzorků volte s ohledem na použití v uvedených systémech digitálního rozhlasového vysílání. Vzorky připravené pro poslechový test vyhodnoťte také dostupnými objektivními metrikami. Posuďte vhodnost použití objektivních metrik jednak s ohledem na výsledky poslechového testu a jednak s ohledem na výsledky pro různé metody komprese bitového toku použité při zdrojovém kódování. Vybranou objektivní metodu využijte pro vyhodnocení kvality zvuku v systémech digitálního rozhlasového vysílání. Posuďte dosažitelnou kvalitu zvuku v závislosti na rychlosti bitového toku přenášeného zvukového signálu pro různé metody komprese.

Seznam doporučené literatury:

- [1] POCTA, P., BEERENDS, J. G. Subjective and objective assessment of perceived audio quality of current digital audio broadcasting systems and web-casting applications. IEEE Transactions on Broadcasting, 2015.
[2] ULOVEC, K., SMUTNY M. Perceived Audio Quality Analysis in Digital Audio Broadcasting Plus System Based on PEAQ. RADIOENGINEERING, VOL. 27, NO. 1, 2018.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Karel Ulovec, Ph.D., katedra radioelektroniky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **11.02.2019**

Termín odevzdání diplomové práce: _____

Platnost zadání diplomové práce: **20.09.2020**

Ing. Karel Ulovec, Ph.D.
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne.....

.....Podpis Autora

Poděkování

Rád bych tímto poděkoval Ing. Karlu Ulovcovi, Ph.D. a Ing. Františku Rundovi, Ph.D. za cenné rady a konzultace týkající se akademické stránky diplomové práce. Dále bych chtěl vyjádřit dík členům mé rodiny a mým přátelům za podporu psychickou, která byla snad ještě důležitější.

Tato práce byla podpořena grantem studentské grantové soutěže
ČVUT č. SGS17/190/OHK3/3T/13

Abstrakt

Diplomová práce se věnuje problematice hodnocení kvality zvuku ve vztahu k nastavení parametrů vysílání v systémech digitálního rádia. Nejprve jsou stručně popsány systémy DAB/DAB+, DRM(+) a techniky zdrojového kódování zvuku. Další část se věnuje postupům hodnocení poslechové kvality. Napřed je uveden přehled subjektivních poslechových metod a poté jsou vysvětleny principy objektivního testování různými počítačovými algoritmy. Praktická část se soustředí na realizaci poslechových testů MUSHRA a objektivního testování užitím čtyř různých metrik: PEAQ Basic, PEAQ Advanced, PEMO-Q a ViSQOL. Pomocí výsledků subjektivního testování je určena nejlepší metoda, která je následně použita k vyhodnocení kvality přenášeného zvukového signálu v závislosti na bitovém toku a způsobu komprese.

Keywords:

digitální rádio, hodnocení kvality zvuku, MUSHRA, PEAQ, PEMO-Q, ViSQOL.

Abstract

The thesis deals with the issue of sound quality evaluation in relation to the setting of broadcasting parameters in digital radio systems. In the beginning, DAB/DAB+, DRM(+) systems and audio source coding techniques are briefly described. The next part focuses on the procedures of listening quality assessment. First, an overview of subjective listening methods is given, and then the principles of objective testing by various computer algorithms are explained. The practical part focuses on the implementation of listening MUSHRA tests and objective testing using four different metrics: PEAQ Basic, PEAQ Advanced, PEMO-Q and ViSQOL. Using subjective testing results the best method is chosen and used to evaluate the quality of the transmitted audio signal, depending on bitrate and compression method.

Keywords:

digital radio, sound quality assessment, MUSHRA, PEAQ, PEMO-Q, ViSQOL.

Obsah

1	Úvod	1
2	Digitální rádio a metody komprese audiosignálu	3
2.1	Standardy digitálního rozhlasového vysílání	3
2.1.1	Digital Audio Broadcasting	4
2.1.2	Digital Radio Mondiale (+)	7
2.2	Kodeky a profily používané v digitálním rozhlasovém vysílání	8
2.2.1	MPEG-1 Layer II	9
2.2.2	Advanced Audio Coding	10
2.2.3	Opus	14
3	Poslechové testy	17
3.1	Metoda pro nepatrná zhoršení	17
3.2	Obecné metody	19
3.2.1	Párové srovnávání	19
3.2.2	Seřazování do pořadí	19
3.3	Metoda pro střední kvalitu	20
3.3.1	Metodika	20
3.3.2	Výběr subjektů	20
3.3.3	Tréninková fáze	21
3.3.4	Podněty	21
3.3.5	Poslechové podmínky	22
3.3.6	Statistická analýza	23
3.3.7	Prezentace výsledků	23
4	Objektivní metody hodnocení kvality zvuku	25
4.1	Metody užívané v telekomunikační technice	26
4.1.1	PSQM, PESQ a POLQA	26
4.1.2	3SQM	27

4.1.3	ViSQOL	28
4.2	Metody v multimediální technice	29
4.2.1	PEAQ	29
4.2.2	PEMO-Q	31
4.2.3	ViSQOLAudio	32
5	Realizace subjektivních a objektivních testů	33
5.1	Příprava testovaného materiálu	33
5.1.1	Použité kodeky	33
5.1.2	Typy materiálu z hlediska obsahu	35
5.2	Návrh subjektivních testů	36
5.2.1	Důvody pro výběr testu MUSHRA	36
5.2.2	Výběr vzorků	37
5.2.3	Použité vybavení	38
5.2.4	Grafické uživatelské rozhraní	40
5.3	Použité implementace algoritmů pro objektivní testování kvality audiosignálů	43
5.3.1	PEAQ	43
5.3.2	PEMO-Q	44
5.3.3	ViSQOL Audio	45
5.4	Popis ostatních funkcí, skriptů a rozhraní	45
5.4.1	main	45
5.4.2	evaluate	45
5.4.3	showResults	46
5.4.4	makeCell	47
6	Výsledky testů	49
6.1	Výsledky subjektivních testů	49
6.2	Porovnání objektivních a subjektivních výsledků	54
6.3	Výsledky objektivních testů	57
7	Závěr	61
	Literatura	63
A	Digitální přílohy	69
B	Přílohy k subjektivnímu testu	71
B.1	Informovaný souhlas	71
B.2	Doprovodné informace	72
C	Podrobné výsledky objektivního hodnocení	73

Seznam obrázků

2.1	Blokové schéma sestavení DAB/DAB+ signálu.[1]	5
2.2	Naznačení principu rozmístění nosných kmitočtů v systému OFDM [2]	6
2.3	Blokové schéma DRM systému [3]	7
2.4	Blokové schéma MPEG Layer II v systému DAB. [4]	9
2.5	Grafické znázornění vzniku jednotlivých profilů kodeku AAC postupným přidáváním moderních technologií zpracování zvuku. [5]	10
2.6	Blokové schéma AAC kodéru převzaté z normy [6]	11
2.7	Proces spektrální pásmové replikace.[7]	12
2.8	Parametrické stereo [7]	13
2.9	Kodér a dekodér xHE-AAC [8]	14
2.10	Schéma přepínání módů v USAC [8]	14
2.11	Srovnání kodeku Opus s ostatními kompresními metodami podle [9]	15
3.1	Frekvenční toleranční maska studiových sluchátek z [10]	22
4.1	Blokové schéma algoritmu PESQ	27
4.2	Blokové schéma algoritmu 3QSM	27
4.3	Blokové schéma algoritmu ViSQOL[11]	28
4.4	Blokové schéma algoritmu PEAQ	29
4.5	Blokové schéma modelu sluchové cesty v PEMO-Q[12]	31
4.6	Blokové schéma algoritmu PEMO-Q [13]	32
5.1	Blokové schéma zapojení měřicího pracoviště při měření sluchátek Sennhesier HD 200 PRO	39
5.2	Frekvenční charakteristika sluchátek Sennhesier HD 200 PRO změřená na umělém uchu Brüel & Kjær	40
5.3	Grafické rozhraní uvítacího okna	42
5.4	Grafické rozhraní průběhu testu	43
5.5	Uživatelské rozhraní programu OPERA	44
5.6	Grafické rozhraní <code>showResults</code>	46

6.1	Grafické zobrazení výsledků prvního testu	50
6.2	Grafické zobrazení výsledků prvního testu rozdělené do kategorií	51
6.3	Výsledky testu MUSHRA pro nižší bitové rychlosti	52
6.4	Výsledky testu MUSHRA pro nižší bitové rychlosti včetně kategorií	53
6.5	Závislost ODG na SDG včetně oblastí vytyčených směrodatnými odchylkami objektivních a subjektivních testů	56
6.6	Hodnocení algoritmem ViSQOL.	59
C.1	Hodnocení algoritmem PEAQ Basic.	74
C.2	Hodnocení algoritmem PEAQ Advanced	75
C.3	Hodnocení algoritmem PEMO-Q.	76

Seznam tabulek

2.1	Provozní módy DAB/DAB+ [1]	4
2.2	Dostupné bitové rychlosti při použití MPEG Layer II a vzorkovací rychlosti 48 kHz [14]	5
2.3	Dostupné bitové rychlosti při použití MPEG Layer II a vzorkovací rychlosti 24 kHz [14]	5
2.4	Přehled dostupných přenosových rychlostí v DAB/DAB+ pro dostupné stupně ochrany a kódové poměry	6
2.5	Parametry OFDM symbolu.	8
2.6	Parametry OFDM symbolu.	8
3.1	Hodnotící škála podle [15] (Překlady z [16])	18
3.2	Bipolární škála podle [17] (Překlady z [16])	19
3.3	Unipolární sto-dílková škála CQS pro popis kvality se slovními popisy dle [17] a překlady z[16])	20
4.1	Seznam parametrů MOV	30
5.1	Seznam vzorků použitých při objektivním testování kvality	36
5.2	Seznam vzorků použitých v subjektivních testech	37
5.3	Použité bitové rychlosti v testu č. 1	37
5.4	Použité bitové rychlosti v testu č. 2	38
5.5	Parametry sluchátek Sennhesier HD 200 PRO udávané výrobcem [18]	39
5.6	Parametry zvukové karty AxagonHQ ADA-HP udávané výrobcem [19]	41
6.1	Výsledky prvního testu	50
6.2	Výsledky druhého testu	52
6.3	Kvadratické střední chyby ε pro jednotlivé kompresní metody	54
6.4	Průměrná hodnota ε a ε_{radio}	55
6.5	Směrodatné odchylky objektivního hodnocení	55

Úvod

Zkuste to bez drátů, milý Marconi!

-Ladislav Smoljak, Zdeněk Svěrák, *Jára Cimrman ležící, spící*

Letmým pohledem do historie můžeme zjistit, že rozhlasové vysílání je tu s námi už zhruba jedno století. V roce 1922 bylo v Anglii zahájeno první pravidelné rozhlasové vysílání BBC (*British Broadcasting Corporation*). Jen rok na to následoval počátek vysílání i na území Československa. Tehdejší Český rozhlas Radiojournal vysílal monofonně amplitudovou modulací a ta přenášela převážně mluvené slovo skrz éter až do roku 1959, kdy se k ní připojila modulace frekvenční.[20] Postupem času se k rádiím připojily i další technologie jako například televizní vysílání, telekomunikační či datové služby a elektromagnetické spektrum začalo být velmi ceněnou komoditou.

Snaha o rozšíření nabízeného obsahu a zároveň o minimalizaci využitého pásma vedla k myšlence digitalizace rozhlasového vysílání. Digitální rádio, přenášející zvuk ve formě jedniček a nul, umožnilo využít kompresních algoritmů a tím výrazně snížit spektrální náročnost. Tento krok ovšem postavil před poskytovatele vysílání nelehký úkol. Jaké parametry přenosu nastavit, aby posluchač nepostřehl vliv digitální komprese. Jinými slovy jak zachovat dostatečnou kvalitu zdrojového signálu a zároveň docílit co nejefektivnějšího přenosu. Řešení tohoto problému se nalézalo ve výsledcích poslechových testů. Aby byly výsledky testů relevantní, bylo třeba zaručit různorodost vzorků, dostatečné množství subjektů a správnou metodiku. Staré rčení ovšem říká „Čas jsou peníze“ a nejinak tomu je i v případě subjektivních testů.

Rozvoj výpočetní techniky ovšem umožnil nahradit skupiny lidí v poslechových místnostech algoritmy, odhadujícími kvalitu zvuku zakódovaného ztrátovou kompresí. Ta na rozdíl od komprese bezztrátové, nesnižuje pouze entropii dat, nýbrž z původního zdroje odstraňuje irelevanci. Jaká data mohou být označena za irelevantní pro poslech lidským uchem zkoumá ve světě zvukového zpracování obor zvaný psychoakustika. Kodeky využívají objevů této vědní disciplíny, jimiž jsou například spektrální či časové maskování, ve prospěch úspory dat. Obecně se dá říci, že čím modernější kodek je, tím více těchto poznatků využívá. To je sice velice pozitivní z hlediska přenosu signálů skrze kanál, nicméně

1. Úvod

se ukazuje, že algoritmy určené pro objektivní hodnocení audiosignálů si s takovýmto vývojem neumí dobře poradit a výsledky hodnocení dopadají při použití nových efektivnějších kodeků hůře, než kdyby byl hodnotícím prvkem lidský posluchač. To ve výsledku může vést k nesprávnému nastavení parametrů přenosu v digitálním rádiu a tím k neefektivnímu využití spektra.

Ověřením zda k tomuto jevu dochází a v jaké míře se zabývá tato diplomová práce. V následujících kapitolách je uveden vhled do systémů digitálních rádií a v nich používaného zdrojového kódování. Dále je zde přiblížena problematika subjektivního a objektivního hodnocení audiosignálů. Podrobně jsou popsány jednotlivé metody, jenž jsou poté v praktické části využité k testování zvukových vzorků zpracovaných různými kompresními metodami. Dále je v práci popsáno, jaké programové vybavení bylo vyvinuto a využito a v závěru se práce věnuje výsledkům jednotlivých testů společně s jejich souvislostmi. Na jejich základě je vybrána a použita nejvhodnější metoda k vyhodnocování kvality zvuku v systémech digitálního rozhlasu.

Digitální rádio a metody komprese audiosignálu

*Video killed the radio star.
The Buggles, název písně*

Přestože smrt rádia prorokovali „The Buggles“ v historicky prvním videoklipu vysílaném hudební televizí MTV (*Music Television*), analogové rádio je tu s námi dodnes a v následujících letech pravděpodobně bude i nadále. Zdá se že minimálně do roku 2025, dokdy v České Republice platí licence na pásma, ve kterých je klasické FM rádio provozováno. Nicméně digitalizace je všudypřítomný trend a podobně jako se v průběhu poslední dekády přecházelo na digitální vysílání televize DVB-T (*Digital Video Broadcasting – Terrestrial*), čeká nás modernizace i ve vysílání rozhlasových služeb.

2.1 Standardy digitálního rozhlasového vysílání

V současné době existují dva velké paralelně existující standardy. DAB (*Digital Audio Broadcasting*), ke kterému se po úspěchu v severských zemích upínají i tuzemští poskytovatelé rozhlasových služeb, a DRM (*Digital Radio Mondiale*), momentálně zažívající „boom“ v Indii a dalších zemích, kde je rádio provozováno v pásmu středně dlouhých vln. Vedle nich existují i další systémy, jako například Japonský ISDB (*Integrated Services Digital Broadcasting*) standard pro vysílání televize a rádia, nahrazující tamější analogový NTSC-J (*National Television System Committee - Japan*), či IBOC (*In-band on-channel*) užívaný v Americe pod obchodním názvem HD Radio, který vysílá analogový a digitální signál simultánně. Tyto druhy vysílání se ovšem nevztahují k evropské půdě a tudíž se práce v následujících dvou podkapitolách zabývá pouze systémy DAB/DAB+ a DRM(+).

2.1.1 Digital Audio Broadcasting

Standard *EUREKA 147 Digital Audio Broadcasting* (zkráceně DAB) popsaný v normě ETSI EN 300 401[1] je normou vyvíjenou již od osmdesátých let Evropským ústavem pro telekomunikační normy ETSI (*European Telecommunications Standards Institute*) společně s Mezinárodní telekomunikační unií ITU (*International Telecommunication Union*). Jedná se o moderní a univerzální multimediální rozhlasový systém, jehož cílem je nahradit současné FM vysílání. Oproti tomu se pyšní například odolností vůči aditivnímu šumu, kvalitou zvuku srovnatelnou se záznamem na CD, energetickou úsporností a možností využití ve více pásmech. Balík několika rozhlasových stanic společně s doprovodnými datovými službami multiplexuje do jednoho datového toku, který je poté vyslán modulací OFDM (*Orthogonal frequency-division multiplexing*), jejíž hlavní předností je odolnost vůči chybám způsobeným vícecestným šířením a možnost využití jedno-frekvenční sítě.

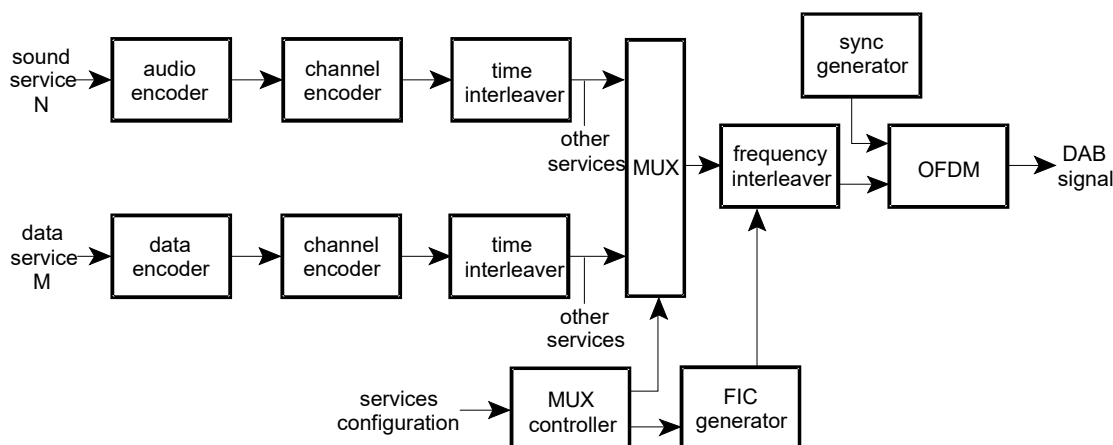
Starší verze systému DAB [21], byla navržena pro přenos různými technologiemi v několika pásmech jak je ukázáno v tabulce 2.1. Definovány byly 4 vysílací módy, které umožňovaly vysílání až do frekvencí kolem 3 GHz. Ve verzi 2.2.1 doporučení ETSI EN 300 401 [1] z ledna 2017 se už s módy II - IV nadále nepočítá a DAB+ tak zůstává ryze terestrickou platformou.

	Mód I	Mód II	Mód III	Mód IV
Použití	Pozemní VHF 168 - 240 MHz	Pozemní L pásmo 1452 - 1492 MHz	Satelitní L pásmo až do 3 GHz	Městská zástavba L pásmo
Počet nosných vln frekvencí	1536	384	192	768
Odstup nosných vln [kHz]	1	4	8	2
Trvání intervalu [μs]	1246	311,5	155,75	623
Délka ochranného intervalu	246,0	61,5	30,75	123

Tabulka 2.1: Provozní módy DAB/DAB+ [1]

Norma [1] definuje pouze vznik signálu na vysílací straně. Řešení na přijímací straně je ponecháno na výrobci koncových zařízení. Na obrázku 2.1 je ukázáno jak je sestavení DAB signálu řešeno.

Z hlediska zdrojového kódování využívá původní DAB kodek MPEG-1 Audio Layer II (více o kodecích v podkapitole 2.2). Je možno použít dva vzorkovací kmitočty: 24 kHz a 48 kHz. Od těch se odvíjí možné bitové rychlosti zdrojového kódování vypsané v tabulkách 2.2 a 2.3. Výstupem kodéru jsou logické rámce (*Logical Frames*) trvající 48 ms respektive 24 ms pro vyšší vzorkovací rychlost. V novější verzi DAB+ metodu MPEG Layer II doplnil modernější kódovací postup HE-AAC v2. Tok nového kodéru je opatřen Reed-Solomonovým ochranným kódem a je vkládán do superrámeců (*Super Frames*), dlouhých 120 ms, které



Obrázek 2.1: Blokové schéma sestavení DAB/DAB+ signálu.[1]

jsou rozděleny do pěti logických rámců DAB. Počet vzorkovacích kmitočtů vyrostl na čtyři. Konkrétně: 16, 24, 32, 48 kHz a možné bitové rychlosti jsou násobky osmi od 8 kb/s až do 192 kb/s.

48 kHz jednokanálový mód (kb/s)	56	64	80	96	112	128	160	192
48 kHz ostatní módy (kb/s)	112	128	160	192	224	256	320	384

Tabulka 2.2: Dostupné bitové rychlosti při použití MPEG Layer II a vzorkovací rychlosti 48 kHz [14]

24 kHz (kb/s)	8	16	24	32	40	48	56	64	80	96	112	128	144	160
---------------	---	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----

Tabulka 2.3: Dostupné bitové rychlosti při použití MPEG Layer II a vzorkovací rychlosti 24 kHz [14]

Po odstranění irelevantních a nadbytečných dat zdrojovým kóděrem, přichází na řadu přidání pečlivě řízené redundantní složky pomocí konvolučního kóděru, který je následován časovým prokladačem. Ten determinovaným způsobem přeskládá tok dat tak, že při zpětném složení se případná shluková chyba, vzniklá při přenosu kanálem, „rozmělní“ a může být opravena konvolučním kódem. Maximální možné přenosové rychlosti multiplexu DAB v závislosti na stupni ochrany jsou v tabulce 2.4.

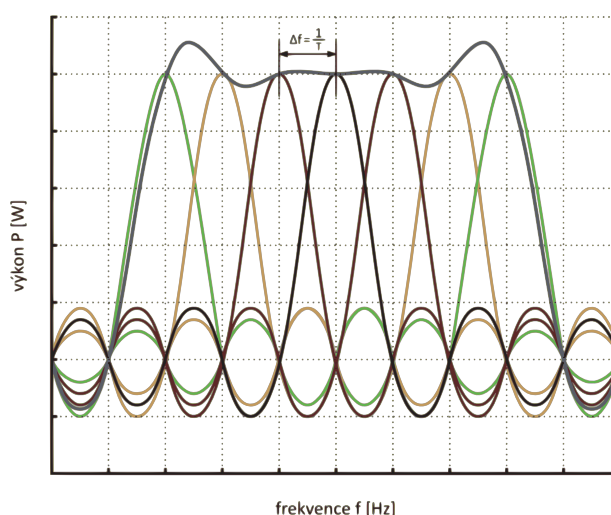
Počet programů v multiplexu DABu je flexibilní, zrovna tak jako nastavení kvality jednotlivých zvukových služeb a jejich doprovodných informací.

Ortogonalně frekvenčně dělený multiplex, neboli OFDM je modulace využívající velkého množství nosných vln. Dostupná šířka pásma je rozdělena do ortogonálních subkanálů. Ty jsou od sebe vzdáleny $\Delta f = \frac{1}{T}$, kde T je doba trvání jednoho symbolu. Signál omezený v čase si lze představit jako signál trvající nekonečně dlouho vynásobený obdélníkovým pulzem. Takovému signálu ve frekvenčním spektru odpovídá výkonová hustota popsaná

Stupeň ochrany	Kódový poměr	Kapacita multiplexu (kb/s)
1-A	1/4	576
2-A	3/8	864
3-A	1/2	1152
4-A	3/4	1728

Tabulka 2.4: Přehled dostupných přenosových rychlostí v DAB/DAB+ pro dostupné stupně ochrany a kódové poměry

funkcí $\frac{\sin x}{x}$. Průchody nulou jednotlivých nosných vln jsou umístěny přesně na sebe, díky čemuž jsou postranní laloky potlačeny, jak naznačuje obrázek 2.2.



Obrázek 2.2: Naznačení principu rozmístění nosných kmitočtů v systému OFDM [2]

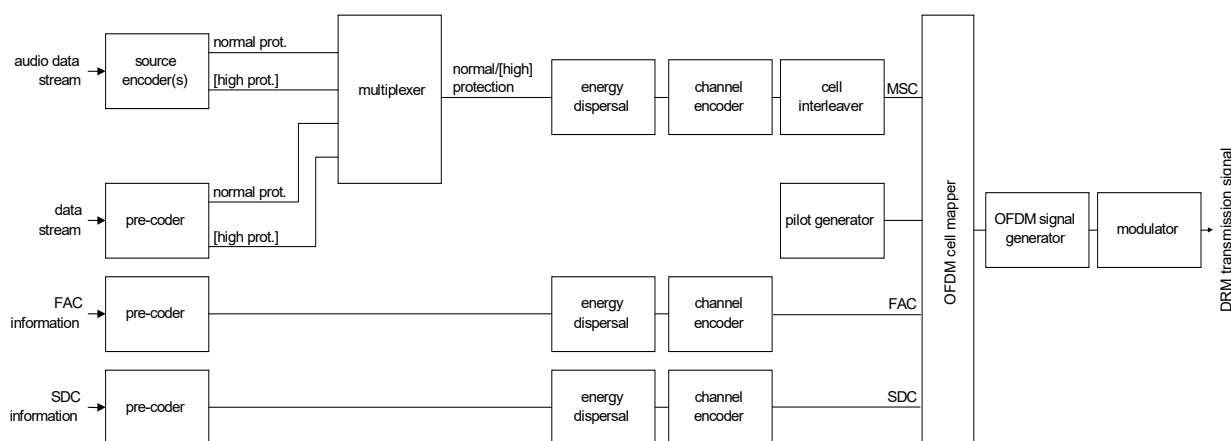
K čemu velké množství nosných frekvencí? Zjednodušeně řečeno, čím více nosných vln, tím delší symbolový čas T . Odražený signál, který do přijímače dorazí s časovým zpožděním může způsobit chybnou detekci symbolu. Dobře se to dá představit na analogové televizi, kde zpožděný signál způsobil chybu známou jako „duch“. V digitálním světě sice „duchové“ neexistují, zato vznikne překryv symbolu zpožděného signálu se symbolem následujícího rámce, tedy jev známý jako mezisymbolová interference ISI (*Inter Symbol Interference*). Prodloužení symbolového času způsobí to, že mírně zpožděný signál dorazí stále „včas“ a nedojde k znehodnocení přijatého symbolu. I přesto ve zlomku vysílacího času vzniká překryv dvou symbolů a proto je do vysílání vkládán ochranný interval GI (*Guard Interval*), neboli čas, kdy je vysílání pozastaveno. Tím na jednu stranu klesá efektivita přenosu, nicméně signál je díky tomu značně odolnější vůči odrazům. Zároveň to umožňuje konstruovat sítě s vysílači na téže frekvenci, čímž dochází k značné úspoře spektra.

Jednotlivé nosné vlny jsou v systému DAB modulovány čtyř stavovým fázovým klíčováním s pamětí neboli DQPSK (Differential Quadrature Phase-Shift Keying).

2.1.2 Digital Radio Mondiale (+)

Na rozdíl od systému DAB, který postupně nahrazuje FM vysílání pouze v pásmu velmi krátkých vln, si DRM brousí zuby i na stanice vysílané amplitudovou modulací ve středních a dlouhých vlnách. ETSI ho definuje normou *ETSI ES 201 980* [3], a k dnešnímu dni existují jeho dvě verze: DRM30 a DRM+. Novější verze standard doplnila o módy pro použití v pásmu VKV a také o modernější druhy zdrojového kódování.

Blokové schéma vysílače DRM je znázorněno na obrázku 2.3. V Multiplexu systému DRM je ve srovnání s digitálním rádiem DAB podstatně menší počet rozhlasových stanic. Typicky jedna, maximálně však čtyři (včetně doplňkových dat). To může být výhodou pro poskytovatele obsahu. Odpadá jim tak problematika složitého domlouvání se s konkurencí a tím je mimo jiné zajištěna větší nezávislost.



Obrázek 2.3: Blokové schéma DRM systému [3]

Pro kanály umístěné ve spektru pod hranicí 30 MHz, nabízí DRM vzhledem k použité modulaci prostor pro datový tok od 8 kb/s po 72 kb/s. Pro kanály nad 30 MHz maximální bitová rychlost povyroste až na 186 kb/s. Nabídka systémů zdrojového kódování je následující:

- MPEG-4 CELP, MPEG-4 XVHC - pro monofonní kódování mluveného slova
- MPEG-4 HE-AAC v2 - pro kódování hudebních stanic
- MPEG xHE-AAC - univerzální kodek pro řeč a hudbu
- MPEG Surround - pro kódování doplňkových stop při vícekanálovém přenosu

Výsledný datový tok se skládá ze třech kanálů:

- MSC (*Main Service Channel*) - kanál obsahující audio a datové služby
- FAC (*Fast Acces Channel*) - informace určené k demodulování signálu

- SDC (*Service Description Channel*) - informace potřebné k dekódování MSC

Každý z bloků má svůj mírně specifický kanálový kódér. Scrambler zajišťuje pseudonáhodný charakter signálu, čímž vyrovnává výkon ve spektru.

Poměr redundance přidané konvolučním kódováním je závislý na zvolené robustnosti. Z hlediska kanálu ve kterém lze DRM využívat definuje standard celkem pět typických prostředí (v tabulce 2.5) a k nim přiděluje možné kombinace nastavení modulace nosné vlny.

Mód	Typické podmínky prostředí
A	AWGN kanál, malé úniky
B	Časově a frekvenčně selektivní kanál s velkým dopplerovským rozprostřením
C	Jako B, ale s větší pravděpodobností příjmu odraženého signálu
D	Jako B, ale s extrémní pravděpodobností příjmu odraženého signálu
E	Časově a frekvenčně selektivní kanál

Tabulka 2.5: Parametry OFDM symbolu.

Konkrétní nastavení jsou vyjmenována v tabulce 2.6, kde časy v ní uvedené jsou vyjádřeny jako násobky elementární časové jednotky $T = 83^{1/3} \mu s$. T_u je čas trvání užitečné doby symbolu a T_g je doba trvání ochranného intervalu.

Seznam Parametrů	Mód robustnosti				
	A	B	C	D	E
T_u [ms]	288 T	256 T	176 T	112 T	27 T
T_g [ms]	32 T	64 T	64 T	88 T	3 T
T_g/T_u	1/9	1/4	4/11	11/14	1/9

Tabulka 2.6: Parametry OFDM symbolu.

OFDM má při terestrickém vysílání řadu výhod a i tento standard dělí kanál na velké množství navzájem se neovlivňujících nosných vln. K jejich modulaci je zde ale na výběr ze třech možností a to mezi čtyř, šestnácti a šedesáti čtyř stavovou kvadraturně amplitudovou modulací (4QAM, 16QAM a 64QAM).

2.2 Kodeky a profily používané v digitálním rozhlasovém vysílání

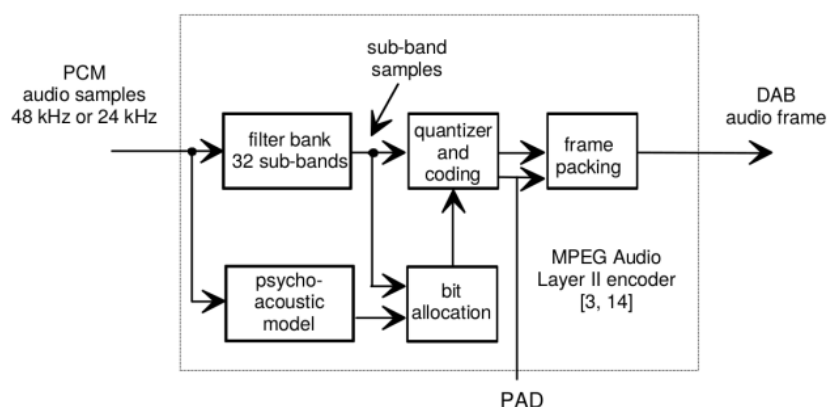
V úvodu je zmíněno, že poznatky z psychoakustiky činí objektivní testování čím dál náročnější. Tato kapitola dává nahlédnout pod pokličku různých kódovacích metod jichž je v

této práci užito. Výběr s jednou výjimkou, konkrétně kodeku Opus, ctí metody používané ve zdrojovém kódování v systémech digitálního rádia definovaných skupinou expertů pro pohyblivý obraz MPEG (*Moving Picture Experts Group*).

2.2.1 MPEG-1 Layer II

Minulý rok tomu bylo již čtvrt století od momentu, kdy byl vydán standard ISO/IEC 111723 [22] zabývající se ztrátovou kompresí multimediálního obsahu. Jeho třetí část pojednávající o kódování audia definuje tři „vrstvy“, jimiž se myslí úroveň integrace psychoakustických poznatků. Standard definuje částečnou dopřednou kompatibilitu jednotlivých vrstev, tedy schopnost alespoň částečně dekodovat tok dat z vyšší vrstvy dekodérem z vrstvy nižší. MPEG Audio Layer III komerčně známá pod označením *mp3* se skrze kapesní přehrávače známe jako „empétrojky“ dostala do povědomí běžné veřejnosti, a i přes své stáří se těší úspěchu do dnešních dní.

V digitálním rádiu či televizním vysílání se ovšem používá její o něco méně komplexní verze MPEG Layer II. Na obrázku 2.4 je naznačen princip kódování.



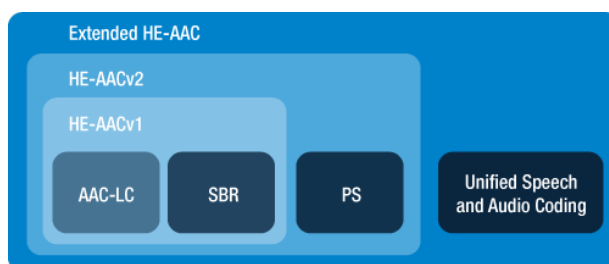
Obrázek 2.4: Blokové schéma MPEG Layer II v systému DAB. [4]

Vstupní pulzně kódovaný signál je rozdělen do dvou cest. V první z nich se nachází banka třiceti dvou filtrů, dělících signál do frekvenčních subpásem, která jsou poté kódována nezávisle. Blok zvaný psychoakustický model využívá principu zvaného spektrální sluchové maskování, neboli schopnosti lidského mozku nevnímat určité kmitočty, zní-li současně s nimi kmitočty jiné. Výstup tohoto bloku říká, jak hrubě kvantizovat (úrovňově rozdělit) jednotlivá subpásma.

Na takto upravené datové toky je poté použito Huffmanovo kódování, které častěji se vyskytujícím datovým slovům přiděluje kratší kódová slova. Tento princip je známý například z Morseovy abecedy. Konečnou fází je „balení“ dat do rámců a opatření hlavičkou obsahující data nezbytná pro dekodér.

2.2.2 Advanced Audio Coding

Standard AAC, přestavený jako součást specifikace MPEG-2, později doplněný o nové profily v MPEG-4 [6] a MPEG-D je rodinou kodeků, která si klade za cíl poskytnout lepší poslechovou kvalitu, než *mp3* při použití stejného bitového toku. Je postaven na modulárním přístupu. Podle potřeby lze využít různých rozšíření základního profilu a tím dosáhnout lepšího kódového zisku, efektivnějšího kódování řeči nebo například nižšího zpoždění. Na obrázku 2.5 je naznačeno jakým způsobem jsou vrstveny profily používané v digitálním rádiu, popsané v následujících podkapitolách.



Obrázek 2.5: Grafické znázornění vzniku jednotlivých profilů kodeku AAC postupným přidáváním moderních technologií zpracování zvuku. [5]

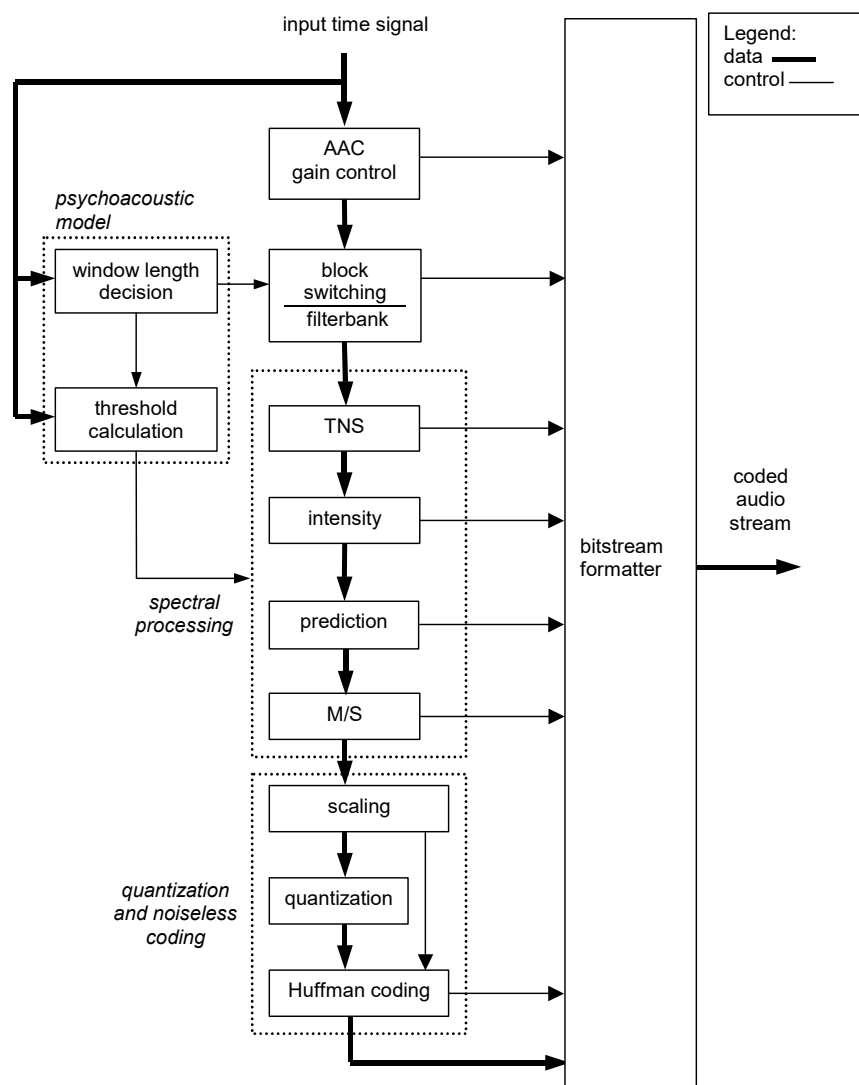
2.2.2.1 Low Complexity

Z funkčního hlediska dosahuje AAC oproti starším kodekům vyššího kódového zisku pomocí několika strategií. Kompletní blokové schéma AAC kodéru převzaté ze standardu ISO 13818-7 [6] se nalézá na obrázku 2.6.

Místo dělení signálu do několika desítek pásem pomocí filtrů v časové doméně, používá AAC LC profil modifikovanou¹ kosinovou transformaci MDCT (*Modified discrete cosine transform*) s proměnlivě dlouhými bloky. Ta je použita k převodu záznamu do 1024 frekvenčních subpásem, které jdou selektivněji kvantizovat výstupem psychoakustického modelu. Dále využívá vyrovnávací paměť pro zpětné adaptivní kódování. Pokud se v paměti kodéru nachází sekvence bitů, která je zrovna na vstupu, uloží se na výstup pouze její adresa. Jedná se o podobný přístup jako známé I, P a B snímky v pohyblivém obraze. V neposlední řadě používá AAC techniky jako jsou například TNS (*Temporal Noise Shaping*), která „posouvá“ kvantizační šum nad slyšitelné pásmo, či JS (*Joint Stereo*), které místo dvou plných kanálů ukládá jejich součet a jejich rozdíl, což je z hlediska entropie dat efektivnější.

Maximální bitová rychlost AAC závisí na vzorkovacím kmitočtu. Pro 44.1 kHz je to 264.6 kb/s, pro 48 kHz 288 kb/s. [6]

¹Modifikace transformace spočívá v úpravě transformačního jádra. Původně kosinový tvar se přibližuje obdélníkovému pulzu

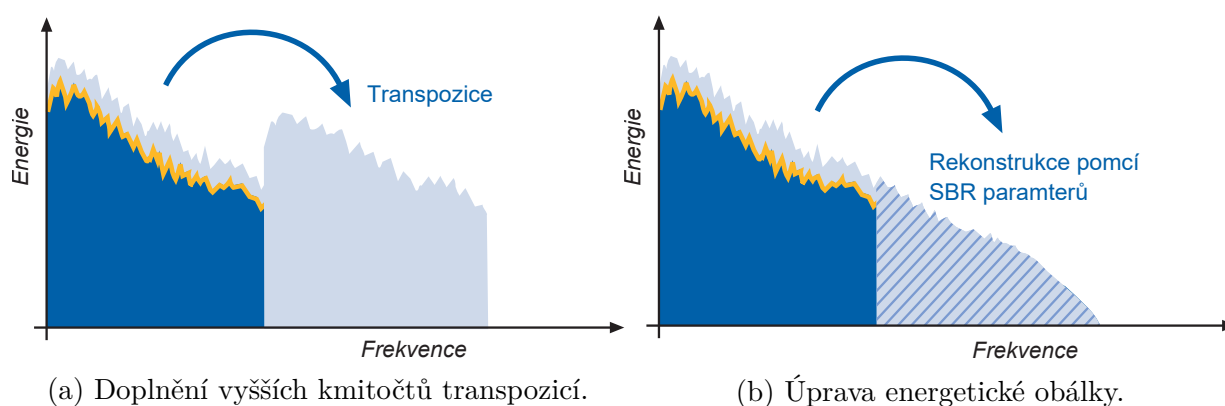


Obrázek 2.6: Blokové schéma AAC kodéru převzaté z normy [6]

2.2.2.2 High Efficiency v1

Vyšší kmitočty zabírají v tradičním kódování audiosignálů poměrně velkou část uložených dat. Ukazuje se ovšem, že z psychoakustického hlediska je exaktní reprodukce posledních jedné až dvou oktáv relativně nedůležitá. Při pouhém potlačení vyšších kmitočtů dolní propustí, jako to dělaly starší kodeky při nedostatečném datovém toku, ovšem označují posluchači vjem jako „neúplný“ či „dutý“ i přes zachování většiny relevantní informace. To společně s faktem, že z fyzikálního hlediska produkují hudební nástroje zvuky na ve spektru se periodicky opakujících kmitočtech, známých jako vyšší harmonické, vedlo k myšlence spektrální pásmové replikace (*Spectral Band Replication*). Obrázek 2.7 naznačuje, jak takový proces probíhá na straně dekodéru. Pomocí transpozice je základní spektrum „překopírováno“ do pásma vyšších kmitočtů a poté je upravena spektrální obálka podle

SBR dat vzniklých při kódování dat vstupního signálu.



Obrázek 2.7: Proces spektrální pásmové replikace.[7]

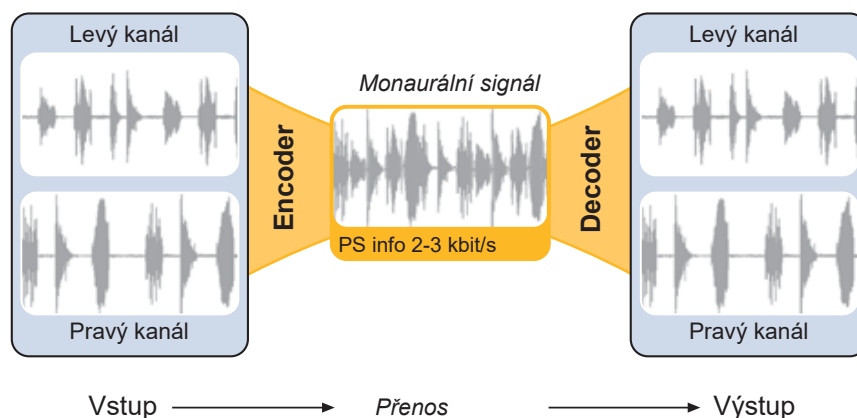
Při opětovném pohledu na obrázek 2.5 vidíme, že kombinace AAC-LC v kombinaci se SBR tvoří profil známý pod jako *High Efficiency AAC v1*. Tento princip obalování původního přístupu modernějšími technologiemi zajišťuje plnou zpětnou a částečnou dopřednou kompatibilitu.

Je nutno podotknout, že HE-AAC v1 není definován pro stejné bitové rychlosti jako jeho předchůdce. Při postupném snižování bitové rychlosti dostává kvalita *Low Complexity* profilu až do přibližně 128 kb/s. Při nižších datových tocích se již začíná být patrná přítomnost kompresních artefaktů. Snížení kvantizačního šumu lze dosáhnout pomocí zúžení pásma a právě tam nastupují výhody SBR.

2.2.2.3 High Efficiency v2

Profil známý jako HE-AAC v2 je dalším evolučním krokem v rodině kodeků AAC. Nad jeho předchozí verzi staví proces známý jako Parametrické Stereo PS (*Parametric Stereo*). Podobně jako u SBR, které namísto původních dat využívá parametrického popisu obálkou, používá parametrické stereo signálovou podobnost. Tentokrát nikoli v kmitočtovém spektru ale v čase. Kodér ze vstupního signálu vezme levý a pravý kanál a smíchá je do jednoho „monoaurálního“ kanálu, jak naznačuje obrázek 2.8. Dále vypočítá z obou původních kanálů tři následující parametry:

- IID *Inter-channel Intensity Difference* - Popisující rozdíl intenzity mezi jednotlivými kanály.
- ICC *Inter-channel Cross-Correlation* - Vyjadřující míru koherence mezi jednotlivými kanály. Je vypočítán korelační funkcí.
- IPD *Inter-channel Phase Difference* - Určující míru „zpoždění“ jednoho signálu vůči druhému.



Obrázek 2.8: Parametrické stereo [7]

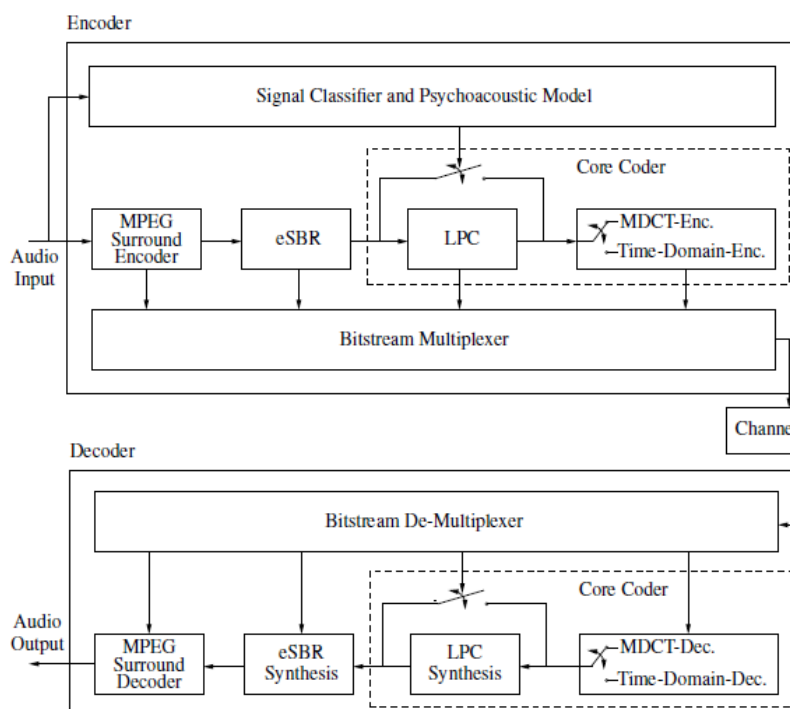
I tento profil se používá až při rychlostech, kdy již kvalita předchozích dvou profilů není dostatečná. 64 kb/s je maximální bitová rychlost při které lze parametrické stereo použít.

2.2.2.4 MPEG-D USAC

Ať už jde o předčítání zpráv, diskuzní pořady či krátké vstupy moderátora mezi hudební reprodukcí, mluvené slovo je v rádiu velmi častým prvkem. Dlouhodobý vývoj přenosu řeči skrze digitální telefonní linku přineslo efektivní kódování mluveného slova zvané LPC (*Linear Predictive Coding*). Vychází z poznatků o rezonanci zvukových vln v dutinách a dekompozici lidské řeči na elementy, které lze popsat databází. Na lidské hlasové ústrojí lze pohlížet jako na soustavu rezonančních trubic, ve kterých jsou zesíleny určité kmitočty, udávající barvu lidského hlasu. LPC analýza pomocí filtrů oddělí tyto kmitočty a vytvoří parametrický popis barvy hlasu. Ve zbytku signálu poté vyhledá řečové elementy a přiřadí jim prvky z databáze, kterou disponuje kodér i dekodér. Na přijímající straně je poté možno zrekonstruovat velmi věrně znějící repliku původního hlasu.

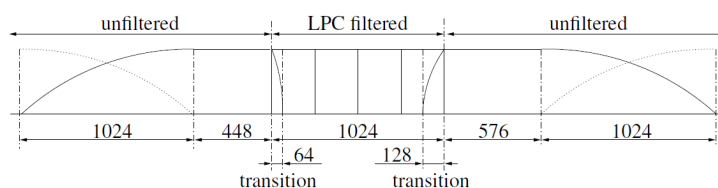
MPEG-D USAC (*Unified Speech and Audio Coding*) vznikl myšlenkou spojit řečové AMR-WB+ (*Extended Adaptive Multi-Rate – Wideband*) a hudební HE-AAC v2 kódování. Standard ISO/IEC 23003-3, známý pod označením MPEG-D, definuje nový přírůstek do rodiny AAC kodeků s názvem xHE-AAC (*eXtended High Efficiency Advanced Audio Coding*) [23], který cílí na bitové rychlosti 32 kb/s a nižší. Při pohledu na obrázek 2.5 se může zdát, že xHE-AAC je jen dalším rozšířením, postaveným nad HE-AAC v2. Ve skutečnosti to tak ale není. USAC využívá stejných technologií jako jeho předchůdce, ale redefinuje většinu bloků aby bylo možné využít LPC analýzu.

Na obrázku 2.9 je naznačeno, jak kodek nakládá se vstupním signálem. Blok MPEG Surround produkuje parametrický popis stereofonního signálu a míchá kanály do jedné stopy. Ta je poté zpracována blokem eSBR (*enhanced Spectral Band Replication*), který SBR použité v HE-AAC v2 doplňuje o efektivnější prediktivní vektorové kódování PVC (*Predictive Vector Coding*). Jednakanálový nízkofrekvenční signál je poté po blocích zpracováván MDCT analýzou obdobnou té v Low Complexity profilu s volitelně doplněnou o



Obrázek 2.9: Kodér a dekodér xHE-AAC [8]

LPC kodér. Volba přemostění řečového kódování lze vynutit natrvalo, či lze vypínat automaticky výstupem psychoakustického modelu rámec po rámci. Skokové přepínání mezi signálem kódovaným by mohlo produkovat kompresní artefakty. xHE-AAC proto definuje komplexní schéma přepínání mezi jednotlivými kódovacími metodami naznačené na obrázku 2.10.



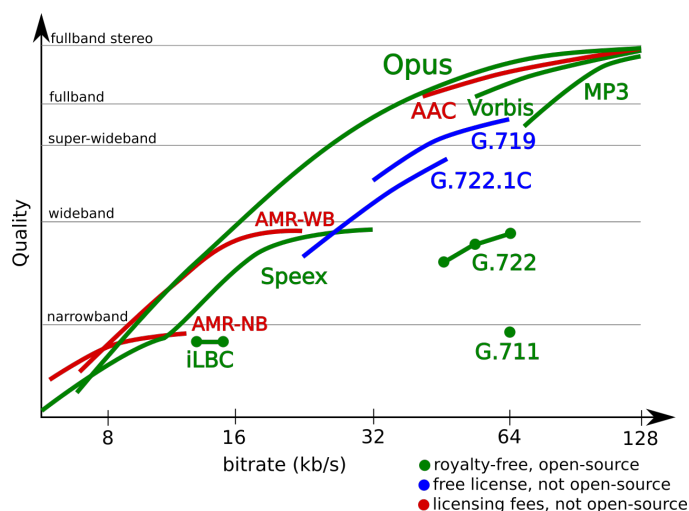
Obrázek 2.10: Schéma přepínání módů v USAC [8]

2.2.3 Opus

Částečně mimo celou koncepci této práce stojí kodek jménem Opus. DAB ani DRM ho nemají ve svých definicích, i když na fóru *HydrogenAudio* [24] jsou k nalezení zmínky tom, že s jistými úpravami lze Opus v DRM používat. To je nicméně velmi okrajová varianta

a není důvodem proč je ve zdejším výčtu uveden. Skutečným důvodem pro jeho volbu byla jeho dostupnost. Je totiž jediným volně dostupným kodekem využívajícím sjednocené kódování řeči a hudby USAC. Srovnání s xHE-AAC by tedy mohlo být zajímavé.

Opus je standardizován pod záštitou IETF (*Internet Engineering Task Force*) jako RFC6716 [25]. Je relativně nový tj. jeho vývoj započal v roce 2007 a první verze spatřila světlo světa 11. Září 2012. V oficiálních materiálech autoři poskytují srovnání s velkým množstvím v současnosti používaných kodeků (obrázek 2.11), kde si vede velmi dobře.



Obrázek 2.11: Srovnání kodeku Opus s ostatními kompresními metodami podle [9]

Z technologického hlediska se velmi podobá xHE-AAC. Tam kde MPEG využívá řečové kódování AMR-WB+, využívá Opus kodek SILK vyvinutý společností Skype Technologies pro internetovou telefonii. Ke kódování zvuku používá CELT (*Constrained Energy Lapped Transform*) otevřený kodek založený, stejně jako AAC, na MDCT analýze [25].

Poslechové testy

Nejkrásnější hudba je lidský smích.

Jan Werich

Do dnešních dní jsou subjektivní poslechové testy nejvěrohodnějším způsobem hodnocení kvality ve zvukové technice. V průběhu let vydala Mezinárodní telekomunikační unie řadu doporučení, podrobně popisujících různé druhy poslechových testů, jejich metodiky, zaměření, podmínky za jakých mohou být prováděny, jakým způsobem je vyhodnocovat atd. Akademická obec se jich drží, jelikož test provedený podle doporučení ITU má oproti testům nestandardizovaným zaručenou výpovědní hodnotu.

Z nepřeberného množství doporučení, jenž ITU-R definuje, se dají vybrat tři normy bezprostředně se týkající subjektivního testování. Seřazeny dle doby vydání jsou to následující:

1. Metoda pro nepatrná zhoršení: ITU-R BS.1116 [15]
2. Obecné metody testování: ITU-R BS.1284 [17]
3. Metoda pro střední kvalitu, známá pod označením MUSHRA: ITU-R BS.1534 [26]

Na následujících řádcích je uveden rámcový popis jednotlivých metod vyjma testu typu MUSHRA, který je ze všech metod popsán nejpodrobněji, jelikož je v praktické části využit pro testování kvality komprimovaných souborů.

3.1 Metoda pro nepatrná zhoršení

Anglicky *Small impairments method* popsaná v [15] cílí na aplikace, kde lze sledovat zhoršení, které může být nepostřehnutelné bez splnění přísných podmínek a bez důkladné statistické analýzy. Své využití nalézá například při vývoji reprodukcí zařízení či návrhu nových kodeků.

Subjekty musí být tak zvaní znalí posluchači (*expert listeners*) neboli lidé, u nichž je předchozími testy prokázáno, že nedisponují žádnou sluchovou vadou. Před samotným testováním musí projít důkladným výcvikem. Taktéž jsou z výsledků vyřazeny všechny výsledky, jenž neodpovídají přísným kritériím, které doporučení definuje.

Při samotném průběhu testování se využívá přístup anglicky zvaný *The double blind triple stimulus with reference*, neboli dvojité slepý test se třemi podněty včetně známé reference. Subjektu jsou prezentovány tři stimuly. Referenční označovaný „A“ a dva neznámé s označením „B“ a „C“, z nichž jeden je stejný jako známá reference a druhý je signál testovaný. Pojmem „nepatrné zhoršení“ se myslí jakákoliv odlišnost od původního signálu. Úkolem subjektu je tuto odlišnost odhalit a ohodnotit na škále v tabulce 3.1. Průměr udělené známky je v anglicky psané literatuře označován jako *MOS* (*Mean Opinion Score*).

Udělení známky jednomu vzorku se říká zkouška. Test se skládá minimálně z pěti takovýchto zkoušek, kde délka podnětu nepřesahuje 25 vteřin. Celé sezení by ovšem nemělo přesáhnout třicet minut, aby případná únava subjektu neovlivnila jeho hodnocení. Po udělení známky se bez prodlení postupuje k další zkoušce.

Samotné vzorky by měly být vybrány tak aby reflektovaly zkoumanou problematiku a zároveň by si měly zachovat dostatečnou rozmanitost. Tj. je-li sledována kvalita mluveného slova, je třeba dbát na rozmanitost barev hlasů, rychlosti předčítání, pohlaví interpreta apod. Kapitola 8 doporučení [15] „Poslechové podmínky“ (*Listening conditions*) definuje, jakým způsobem by měl být posluchačům zvukový materiál prezentován. V nejlepším případě by měl být test proveden v poslechové místnosti s definovanými parametry jakými jsou například: objem a rozměry místnosti, doba dozvuku, referenční reprodukční zařízení atd. Pokud poslechová místnost není k dispozici, lze k testování využít referenčních sluchátek¹. Hluk na pozadí ovšem nesmí překročit hladinu, která by subjekty rušila při poslechu.

Výsledky testů podle tohoto doporučení se prezentují na stupnici v anglosaské literatuře známé jako *SDG* (*Subjective Difference Grade*). Hodnota *SDG* se získá rozdílem udělených známek testovanému a referenčnímu signálu (rovnice 3.1)

$$SDG = \text{Známka}_{\text{testovaný}} - \text{Známka}_{\text{referenční}} \quad (3.1)$$

Zhoršení	Známka (<i>MOS</i>)	<i>SDG</i>
Nepostřehnutelné	5	0
Postřehnutelné, avšak neruší	4	-1
Mírně rušivé	3	-3
Rušivé	2	-4
Velmi rušivé	1	-5

Tabulka 3.1: Hodnotící škála podle [15] (Překlady z [16])

¹Definici referenčních sluchátek se důkladněji věnuje podkapitola 3.3.5.

3.2 Obecné metody

Protože doporučení [15] je v některých ohledech velice striktní, vydala ITU doplňující standard, ve kterém shrnuje další v praxi využívané metody. Definovala pro ně hodnotící škály, parametry prostředí ve kterém testy realizovat a podmínky, které by měli splňovat subjekty testy provádějící.

3.2.1 Párové srovnávání

V této metodě, známé také jako „AB Test“ volí subjekt jeden ze dvou stimulů označených „A“ a „B“. Test může být koncipován tak, že je hledán poslechově přívětivější podnět, či je rozdíl těchto dvou stimulů označen na hodnotící škále. Doporučení rovněž říká, že jeden ze dvou podnětů může, ale nemusí být referenční.

Hodnotících škál je v doporučení popsáno hned několik. Pro případ že je sledováno zhoršení, využívá se stejného známkování jako v tabulce 3.1. Vyskytují-li se při hodnocení vzorky jak horší tak lepší, měla by být použita následující stupnice.

Zhoršení	Známka
Mnohem horší	-3
Horší	-2
Trochu horší	-1
Stejně	0
Trochu lepší	1
Lepší	2
Mnohem lepší	3

Tabulka 3.2: Bipolární škála podle [17] (Překlady z [16])

Zda bude škála využita či nikoli záleží na sledovaných parametrech. Je-li třeba znát kvalitativní rozdíly mezi jednotlivými stimuly, je užití škály doporučeno. Hledá-li se pouze jeden vzorek s žádoucími parametry, stačí známkování typu „lepší/horší“.

Je doporučeno, ale ne vyžadováno, aby subjekty byli z řad znalých posluchačů. Pokud to není možné, měli by minimálně projít tréninkem, aby se seznámili s testovací procedurou. Co se poslechových podmínek týče, odvolává se norma na předchozí doporučení [15].

3.2.2 Seřazování do pořadí

Ve volbě posluchačů a dodržování reprodukčních podmínek se tato metoda neliší od předchozí. Metodika je ovšem jiná. Subjektu je v rámci jedné zkoušky představeno podstatně větší množství podnětů. Typicky pět až devět. Posluchač poté přiřazuje hodnocení subjektivně vnímané kvality na spojitou stupnici *CQS* (*Continuous Quality Scale*), popsané tabulkou 3.3. Numerickou hodnotou od nuly do sta, či vyznačením bodu na úsečce udělí hodnocení každému z prezentovaných stimulů. Nemělo by být dovoleno udělit stejnou známku dvěma stejným podnětům.

Interval	Slovní popis anglicky	Slovní popis česky
80 % - 100 %	Excellent	Vynikající
60 % - 80 %	Good	Dobrá
40 % - 60 %	Fair	Příjemná
20 % - 40 %	Poor	Špatná
0 % - 20 %	Bad	Nepříjemná

Tabulka 3.3: Unipolární sto-dílková škála CQS pro popis kvality se slovními popisy dle [17] a překlady z[16])

3.3 Metoda pro střední kvalitu

Doporučení [26] definuje tzv. Vícestimulový test se skrytou referencí a kotvou MUSHRA (*MUltiple Stimuli with Hidden Reference and Anchor*). Jedná se o metodu navrženou přímo pro hodnocení kvality výstupu ztrátových kodeků. Využití tedy nalézá při hledání optimálních parametrů pro přenos zvuku v internetových službách, rádiích či televizním vysílání.

3.3.1 Metodika

Podobně jako při seřazovacím testu je při provádění testu MUSHRA posluchači při jedné zkoušce představeno větší množství podnětů. Typicky deset, maximálně však patnáct. První z nich je vždy známý, nezkreslený vzorek a ostatní náhodně zamíchané testované stimuly, mezi které je náhodně vložen skrytý referenční vzorek a jedna, či více kotev. Typy vzorků jsou podrobněji popsány v podkapitole 3.3.4. Posluchač hodnotí každý stimul na již předem zmíněné sto-dílkové CQS škále naznačené na obrázku 3.3. Opětovné přehrání kteréhokoliv stimulu by mělo být umožněno.

3.3.2 Výběr subjektů

I přes to že test MUSHRA neslouží k detekci nepatrných zhoršení doporučuje standard vybírat subjekty z řad znalých posluchačů stejně jako doporučení [15]. Minimalizuje se tím možnost chyby, kterou do testu může vnést subjekt vyplňující poslechový test bez předchozích zkušeností. Výběr posluchačů, jejichž hodnocení bude mít vliv na výsledky pokusu by měl být proveden na základě těchto dvou metod:

- Pre-screening

Předtest či forma zaučení, jejíž cílem je objevit, zda posluchač disponuje tzv. normálním sluchem² a zda dokáže hodnotit zvuk kriticky.

- Post-screening

²Co je normální sluch definuje ISO 389

Výběr subjektů na základě výsledků již provedeného testování. Starší verze normy [26] definuje dvě kategorie podmínek na základě kterých je možno vyřadit hodnoty hodnocení konkrétního subjektu a to

- nekonzistence opakovaného hodnocení vzorku se stejnými parametry
- nekonzistence hodnocení v rámci množiny subjektů jako například přílišná extremizace hodnocení.

Nová verze normy [27] dodává konkrétní podmínku z první kategorie a to doporučení vyřadit subjekty, jejichž hodnocení skryté reference je ve více než patnácti procentech nižší než 90%.

3.3.3 Tréninková fáze

Před ostrou fází testu je vhodné vložit tzv. zaučení, neboli fázi pro seznámení s rozhraním a stimuly. Subjektu by měly být představeny všechny možné druhy podnětů s konkrétním popisem. Vysvětlení jak při testu postupovat, jakým způsobem udělit hodnocení apod. Zaučení je možno provést buď formou instruktážního videa, doprovodných instrukcí na přiloženém návodě, či formou testu „nanečisto“.

3.3.4 Podněty

Délka jednotlivých podnětů by neměla překročit dvacet sekund. Obvykle se volí úsek okolo deseti vteřin, potažmo taková doba trvání, aby celý test nepřesáhl délku třiceti minut. Jak už název testu naznačuje každá zkouška obsahuje následující podněty:

- Znamá reference:
Stimul označený slovem „Reference“ obsahující nezkreslenou nahrávku.
- Skrytá reference:
Stimul obsahově shodný se známou referencí, náhodně zamíchaný mezi testované podněty. Slouží k takzvanému *post-screeningu*. Posluchač by měl být schopen takovýto podnět identifikovat a udělit mu vynikající hodnocení (z čehož mimo jiné plyne, že alespoň jedno hodnocení v každé zkoušce by mělo mít hodnotu 100 %). Pokud ho posluchač neodhalí a udělí mu opakovaně nízké kvalitativní hodnocení, měl by být při analýze výsledků vyřazen.
- Skrytá kotva:
Minimálně jeden a nebo více uměle zkraslených vzorků, náhodně vložených mezi ty testované. Může podobně jako skrytá reference sloužit k *post-screeningu* a jeho smyslem je ukotvit kvalitativní škálu „zdola“. Doporučení [26] říká, že alespoň jedna z kotev by měla být odvozena od referenčního vzorku aplikováním dolní propusti s následujícími parametry.

3. POSLECHOVÉ TESTY

$$f_c = 3,5 \text{ kHz}$$

Minimální útlum na 4 kHz = 25 dB

Minimální útlum na 4,5 kHz = 50 dB

Maximální zvlnění v propustném pásmu = $\pm 0,1$ dB

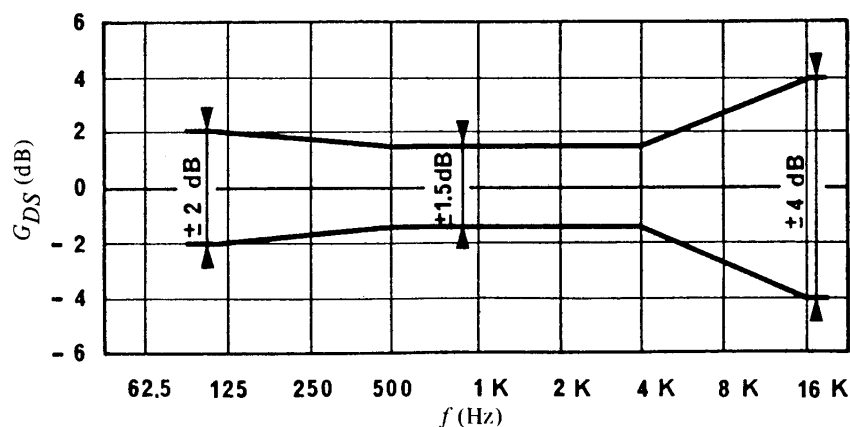
- Testované vzorky:

Stimuly označené písmeny „A“, „B“ a dále na které je aplikováno zkreslení a na nichž je zkoumána problematika. Kvůli vyhodnocení testu je seznam podnětů konzistentní přes všechny zkoušky, tj. i přes to, že každá zkouška obsahuje jinou nahrávku, je na ní pokaždé aplikována stejná paleta zkreslení (pomocí kompresních algoritmů, či jinak přidáných artefaktů).

3.3.5 Poslechové podmínky

Doporučení [26] se odvolává na poslechové podmínky definované v [15] a v [10], tj volbu mezi poslechovou místností osazenou referenční reprosoustavou či sluchátky. Protože v praktické části této práce jsou využita sluchátka, je v tomto textu přiblížena pouze tato možnost. Doporučení [10] říká, že sluchátka musí splňovat následující:

- Časové zpoždění mezi kanály sluchátek nesmí přesáhnout dobu $t_{delay} = 20\mu s$.
- Frekvenční charakteristika musí být plochá tzn. G_{DS} (Odezva sluchátek v difuzním poli) musí ctít masku na obrázku 3.1.



Obrázek 3.1: Frekvenční toleranční maska studiových sluchátek z [10]

Samozřejmostí je, že všichni účastníci testu musí použít stejné poslechové zařízení.

3.3.6 Statistická analýza

Samotné provedení testu by bylo zbytečné bez provedení správné analýzy získaných dat. Podle knihy Základy statistiky [28] se množina subjektů, na kterých je prováděno testování označuje jako „výběr“. Cílem pokusu ovšem není zjistit „výběrový průměr“, nýbrž aproximaci pro celou množinu potenciálních posluchačů. Zavádí se tedy pojem interval spolehlivosti ve kterém s nejvyšší pravděpodobností leží takzvaný „populační průměr“. Jeho kalkulaci zahájíme výpočtem výběrového průměru \bar{u}_{jk} :

$$\bar{u}_{jk} = \frac{1}{N} \sum_{i=1}^N u_{ijk} \quad (3.2)$$

kde, u_{ijk} značí hodnocení uživatele i ve zkoušce j testované metody zhoršení k a N odpovídá množství subjektů. Obdobným způsobem lze spočítat průměry \bar{u}_j či například \bar{u}_k

Doporučení [26] říká, že výsledky mají být prezentovány na 95% intervalu s krajními body:

$$[\bar{u}_{jk} - \delta_{jk}, \bar{u}_{jk} + \delta_{jk}] \quad (3.3)$$

kde δ_{jk} je padesát procent šířky intervalu spolehlivosti daného vztahem:

$$\delta_{jk} = t_{0,05} \frac{\sigma_{jk}}{\sqrt{N}} \quad (3.4)$$

a $t_{0,05}$ je hodnota odpovídající 95% koeficientu spolehlivosti a stupni volnosti $N - 1$, kterou lze najít ve statistických tabulkách. Směrodatná odchylka se poté spočítá podle:

$$\sigma_{jk} = \sqrt{\sum_{i=1}^N \frac{(\bar{u}_{jk} - u_{ijk})^2}{N - 1}} \quad (3.5)$$

Vypočtením těchto hodnot zjistíme, že hodnota populačního průměru se nachází v intervalu spolehlivosti s pravděpodobností 95%.

3.3.7 Prezentace výsledků

Zpracované výsledky by měly být prezentovány takovou formou, aby jim porozuměl jak laik tak expert v oboru. Na přehledném místě v textu by měl čtenář zprvu narazit na grafické zpracování s vyobrazeným intervalem spolehlivosti. To by mělo být doprovázené popisem toho co jednotlivé grafy vyjadřují nicméně není potřeba uvádět veškeré výpočty vedoucí k daným závěrům. Pokud je autor uvést chce, může je přiložit ve formě dodatku.

Opomenuto by tedy nemělo být následující:

- grafická reprezentace dosažených výsledků
- popis množiny subjektů: počet, průměrný věk..

3. POSLECHOVÉ TESTY

- specifikace poslechového materiálu včetně důvodu pro výběr
- charakteristika poslechové místnosti včetně rozměrů, typy měničů, jejich umístění a konfigurace
- popis metodiky testu
- způsob zpracování dat

Objektivní metody hodnocení kvality zvuku

Práce je to, co nikdo nechce dělat.

Tomáš Garrigue Masaryk

Jak už je naznačeno v úvodu, objektivní algoritmy slouží jako alternativa k subjektivním poslechovým testům. Jde o způsob jakým obejít mnohdy náročné testování s velkým počtem posluchačů a tím ušetřit nemalé množství materiálních i časových prostředků.

Snaha o objektivní hodnocení za použití výpočetních metod se dá vysledovat až k počátkům devadesátých let. Existuje několik metod a pohledů na tuto problematiku, nicméně všechny v jádru sdílí obdobný přístup. Konkrétně snahu o modelování lidské sluchové cesty, dále přepočet na vnitřní koeficienty reprezentující nervové vzruchy a v závěru jejich matematické zpracování, často označované jako kognitivní analýza. Výsledkem bývá numerická hodnota odpovídající umístění na hodnotící škále, stejně jako tomu je při subjektivním testování.

V čem se metody liší je jejich zaměření a oblast použití. Telekomunikační sekce ITU-T, která standardizuje systémy jejichž cílem je přenést lidský hlas, se zabývá algoritmy, které sledují převážně srozumitelnost. Naproti tomu hodnocení v systémech definovaných radiokomunikační sekci ITU-R se soustředí spíše na věrohodnost reprodukce originálního signálu v audiovizuálních systémech.

Dále by se metody daly rozdělit podle toho, zda pro svou funkci vyžadují referenci v plné kvalitě či nikoliv na tzv. „Intruzivní“ a „Neintruzivní“. Vyhodnocování kvality řeči bývá častěji řešeno neintruzivně, jelikož algoritmy porovnávají testovaný materiál s vnitřními modely řeči. Hudba, která je zvukově „rozmanitější“ většinou vyžaduje přístup intruzivní.

4.1 Metody užívané v telekomunikační technice

I přes to že se tato práce nezabývá telekomunikační problematikou, je dobré zmínit jaké metody objektivního hodnocení kvality poslechu mluveného slova existují. Ať už proto, že do jisté míry sdílí vnitřní filozofii s těmi z oblasti multimediální techniky, tak proto že vývoj některých algoritmů pro hodnocení kvality hudebních nahrávek vzešel právě z této oblasti.

4.1.1 PSQM, PESQ a POLQA

Jsou tři po sobě jdoucí standardy popisují intruzivní algoritmy určené k hodnocení řeči v telekomunikacích. Společně s evolucí telekomunikačního řetězce se vyvíjely i systémy objektivního hodnocení, nicméně všechny tři staví na modelování sluchové cesty a následné kognitivní analýze.

Nejstarší PSQM *Perceptual Speech Quality Measure* popsany v ITU-T P.861 [29] v roce 1996 se soustředí na oblast telekomunikačního pásma 300 Hz - 3,4 kHz. Signály převádí do frekvenční oblasti, kde pomocí filtrů reprezentujících bazilární membránu vytvoří aproximaci zvuku, který vnímá účastník telefonního rozhovoru. Následovným porovnáváním signálů jsou určeny tři parametry:

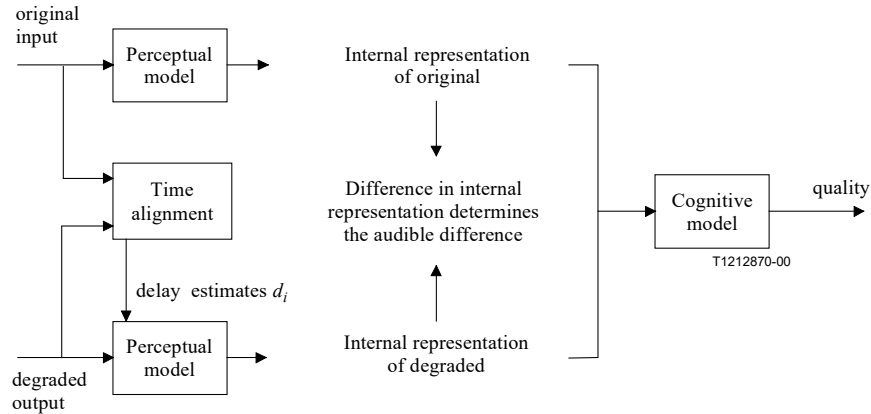
- Časově-frekvenční zvlnění (*Time-Frequency Warping*)
- Frekvenční zvlnění (*Frequency Warping*)
- Úrovnňové zvlnění (*Intensity Warping*)

Jejich váhovaným rozdílem je vypočten parametr označovaný jako *QD* (*Quality Degradation*), který je mapován na škálu *MOS* (*Mean Opinion Score*), která se používá jako výstup subjektivních testů.

Jako hlavní problém při hodnocení systémem PSQM se ukázalo zpoždování řeči způsobené paketovým přístupem v telefonii. Novější standard PESQ (*Perceptual Evaluation of Speech Quality*) dle ITU-T P.862 [30] principiálně popsany obrázkem 4.1 tuto nevýhodu odstranil přidáním bloku *Time alignment*. Ten napřed rozdělí signál na krátké úseky, v telekomunikaci zvané promluvy, korelací obálek odhadne, zda je nutná korekce a pokud ano rozdělí promluvy na rámce dlouhé 64 ms a ty pomocí výsledku korelace po vzorcích seskládá tak, aby testovaný signál odpovídal referenčnímu.

Nejnovější metoda POLQA *Perceptual Objective Listening Quality Assessment* dle ITU-T P.863 z roku 2011 [31] řeší problematiku rozšiřování pásma použitého při přenosu telefonických hovorů. Definuje dva operační módy.

- Úzkopásmový: sloužící pro měření kvality v telefonním pásmu 300 Hz - 3,4 kHz. Pásmový ořez v tomto módu není považován za degradaci signálu a nemá vliv na objektivní hodnocení.
- Super-širokopásmový: ve kterém je šířka pásma rozšířena na 50 Hz - 14 kHz.

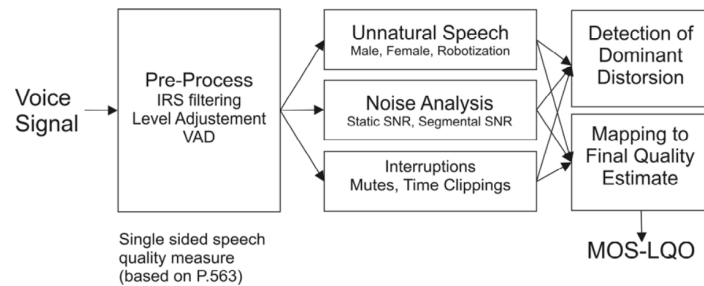


Obrázek 4.1: Blokové schéma algoritmu PESQ

Princip hodnocení je podobný jako u předchozích modelů. Díky většímu výpočetnímu výkonu současných zařízení může POLQA sledovat více parametrů, a provádět složitější matematické operace při kognitivní analýze.

4.1.2 3SQM

ITU-T P.563 *Single Sided Speech Quality Measure* [32] je oproti dvěma předchozím algoritmům metodou intruzivní. Díky tomu lze využít kdekoliv v telekomunikačním řetězci a hodí se tak pro soustavné monitorování sítě. Principiální schéma je na obrázku 4.2.



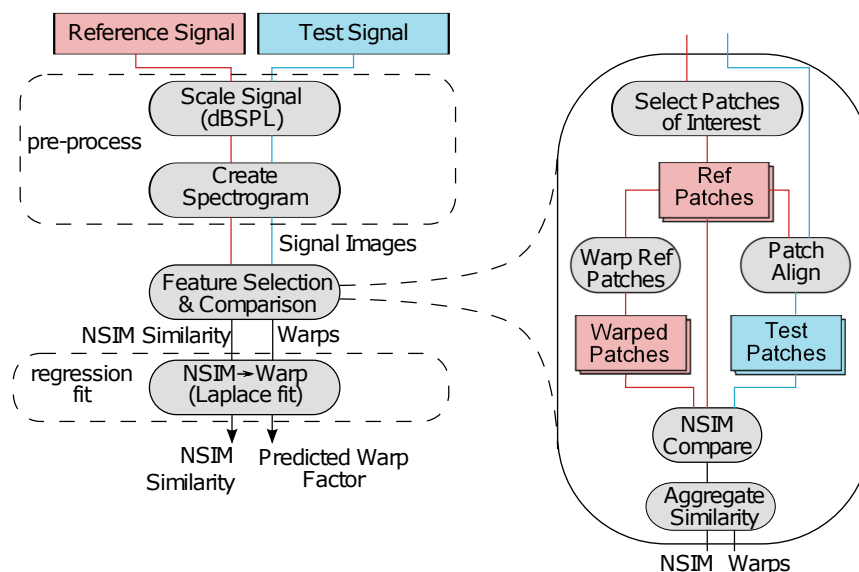
Obrázek 4.2: Blokové schéma algoritmu 3QSM

Signál na vstupu prochází blokem předzpracování. Tím je normován a jsou z něj odstraněny rámce, které s řečí nesouvisí, pomocí tzv. IRS filtru (*Intermediate Reference System*). Nehodí se tak pro detekci parametrů jako je zpoždění, ozvěny či poklesy hlasitosti, které mohou negativně působit na kvalitu telefonních hovorů. Samotná detekce probíhá porovnáváním úrovně obálky signálu.

V signálu se poté sledují parametry jako výška hlasu, variace energie, nepřírozená periodičnost či různé druhy šumu. Na základě těchto parametrů je vytvořen pseudo-referenční signál, pomocí kterého je testovaný signál dále zpracováván intruzivním přístupem podobně jako v metodách PESQ či POLQA.

4.1.3 ViSQOL

The Virtual Speech Quality Objective Listener je nástroj představený ve stejnojmenném článku v roce 2012 [11]. Staví na intruzivním algoritmu zvaném NSIM (*Neurogram Similarity Index Measure*), původně vyvíjeném pro řečovou predikci. Neurogram je analogií ke spektrogramu, kde intenzita barvy představuje sílu vzruchu neurální aktivity. Na obrázku 4.3 je naznačen tok dat algoritmem ViSQOL.



Obrázek 4.3: Blokové schéma algoritmu ViSQOL[11]

V první části označené jako předzpracování dochází k úrovnňovému vyrovnání testového a referenčního signálu. Poté jsou pomocí Fourierovy krátkodobé transformace STFT oba signály převedeny na spektrogramy o třiceti frekvenčních pásmech logaritmicky rozdělených od 250 Hz do 8 kHz.

Část označená jako *Feature Selection & Comparison* slouží k identifikaci vzájemně korespondujících políček¹. Rozměr políčka je 30 rámců x 30 frekvenčních pásem. Algoritmus z obou zpracovávaných signálů vybírá tři takováto políčka nalezením maxima v pásmech 2, 6 a 10, se středními frekvencemi zhruba 250 Hz, 450 Hz a 750 Hz. Poté je na korespondující políčka aplikována metoda výpočtu rozdílu pomocí kvadratické střední chyby² rámec po rámci a pomocí minima této hodnoty jsou políčka zarovnána.

V posledním kroku se na políčka aplikuje NSIM, který vypočítá rozdílové skóre od nuly do jedné, kde nulové hodnocení dostávají signály bez jakékoli podobnosti a jedničku signály identické. Zároveň algoritmus vrací míru časového zkreslení (*Time Warping*).

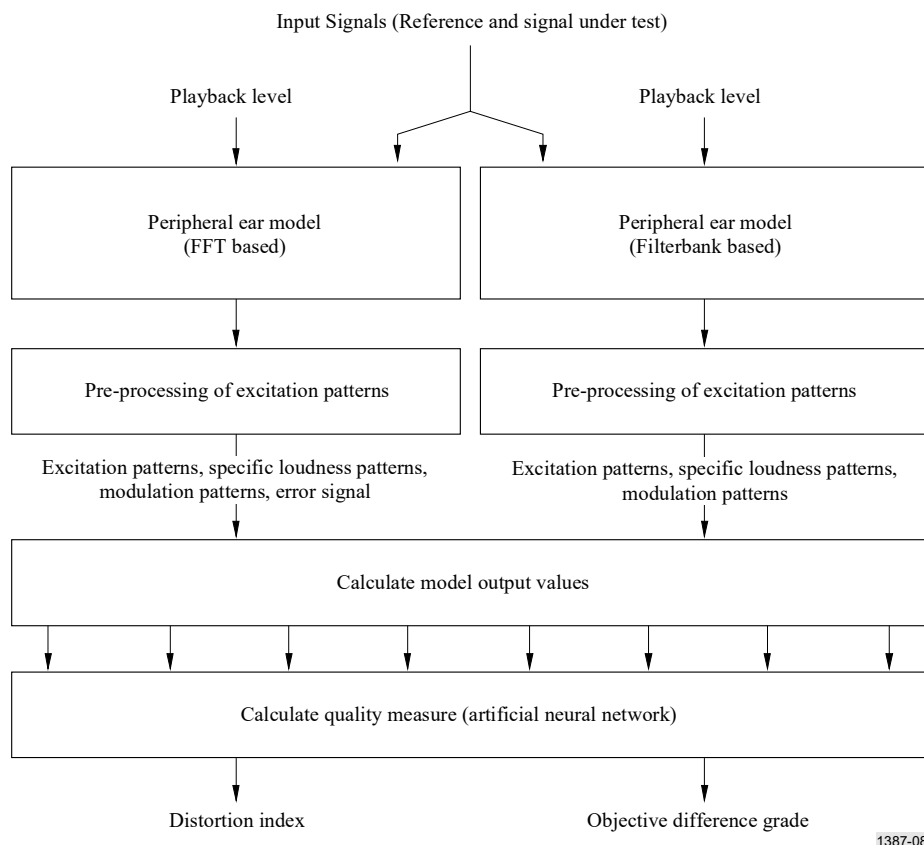
¹Anglicky označována jako *patches*. Žádný dosavadní český překlad nebyl nalezen a z možností ve slovníku slovo políčko nejlépe významově koresponduje.

²RMSE (Root Mean Squared Error)

4.2 Metody v multimediální technice

4.2.1 PEAQ

Perceived Evaluation of Audio Quality zkráceně PEAQ je objektivní metoda hodnocení vnímané kvality zvuku popsaná v ITU-R BS.1387-1 [33]. Princip metody je zobrazen na obrázku 4.4. Jedná se o intruzivní metodu, které je třeba dodat společně se zkresleným signálem signál referenční. Ty jsou převedeny do frekvenční oblasti. Jejich spektrální reprezentace prochází modelem slyšení a následně jsou z nich spočítány *Model Output Values* zkráceně *MOV*. V závěru jsou v bloku zvaném *Artificial Neural Network* provedeny výpočty simulující kognitivní procesy mozku. Výsledkem algoritmu je parametr zvaný *ODG* (*Objective Difference Grade*), což je ekvivalent jednotky *SDG*, představené v podkapitole 3.1



1387-08

Obrázek 4.4: Blokové schéma algoritmu PEAQ

Model PEAQ existuje ve dvou verzích. Základní (*Basic*) a pokročilá (*Advanced*). Ta se od první liší především mírou komplexity výpočtů v sluchové a kognitivní části a použitím druhého modelu sluchové cesty založené na bance filtrů.

Algoritmus PEAQ předpokládá, že signály jsou vzorkovány kmitočtem 48 kHz a zároveň jsou časově a úrovněově vyrovnané. Po načtení je rozdělí do rámců o 2048 vzorcích s

4. OBJEKTIVNÍ METODY HODNOCENÍ KVALITY ZVUKU

Název MOV	Model	Popis
PEAQ Basic		
$BandwidthRef_B$	FFT	Šířka pásma referenčního signálu
$BandwidthTest_B$	FFT	Šířka pásma testovaného signálu
$TotalNMR_B$	FFT	Poměr Šum/Maska
$WinModDiff1_B$	FFT	Váhovaný modulační rozdíl
ADB_B	FFT	Průměrné blokové zkreslení
EHS_B	FFT	Harmonická struktura chyb
$AvgModDiff1_B$	FFT	Průměrný modulační rozdíl 1
$AvgModDiff2_B$	FFT	Průměrný modulační rozdíl 2
$RmsNoiseLoud_B$	FFT	Hlasitost zkreslení
$MFPD_B$	FFT	Maximální vážená pravděpodobnost detekce
$RelDistFrames_B$	FFT	Vzájemně zkreslené rámce
PEAQ Advanced		
$RmsModDiff_A$	Banka Filtrů	
$RmsNoiseLoudAsym_A$	Banka Filtrů	Hlasitost zkreslení
$SegmentalNMR_B$	FFT	Poměr Šum/Maska
EHS_B	FFT	Harmonická struktura chyb
$AvgLinDist_A$	Banka Filtrů	Lineární zkreslení

Tabulka 4.1: Seznam parametrů MOV

vzájemným 50% překryvem. Dále je na rámce aplikováno Hannovo okno a pomocí diskrétní Fourierovy transformace jsou převedeny do frekvenční domény.

Model sluchové cesty se skládá z různých filtrů, reprezentace vnitřního šumu a výpočtu rozprostírání frekvenčních složek pomocí násobení bankou tzv *Spreading functions*. Výstupem těchto matematických operací jsou *MOV*, jejichž seznam je v tabulce 4.1.

Neuronová síť je rovnice 4.1, do které vstupují parametry *MOV* a jejím výstupem je parametr D_I (*Distortion Index*), který je přepočítán na *ODG*.

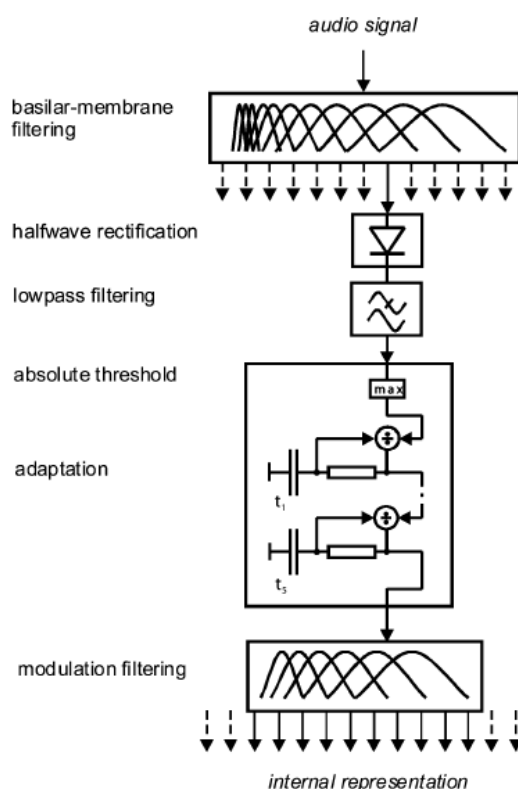
$$D_I = w_{yb} + \sum_{p=0}^{P-1} \left(w_y[p] \text{sig}(w_{wb}[p] + \sum_{q=0}^{Q-1} w_x[q, p] M'_v[q]) \right) \quad (4.1)$$

kde w_y , w_{wb} , w_x jsou váhovací koeficienty dostupné v [34], Q je počet *Model Output Values* značených M' a P je počet uzlů v takzvané „skryté vrstvě“ což je vnitřek orientovaného grafu popisujícího neuronovou síť. Základní model má uzly tři, v pokročilém je jich pět.

Doporučení [33] klade velmi striktní nároky na implementaci tohoto algoritmu. Definuje šestnáct párů hudebních vzorků, u jejichž hodnocení implementací algoritmu by se D_I neměl lišit o $\pm 0.02\%$ od hodnot určených doporučením. Jednou z takových implementací je například program Opera od společnosti Opticom [35].

4.2.2 PEMO-Q

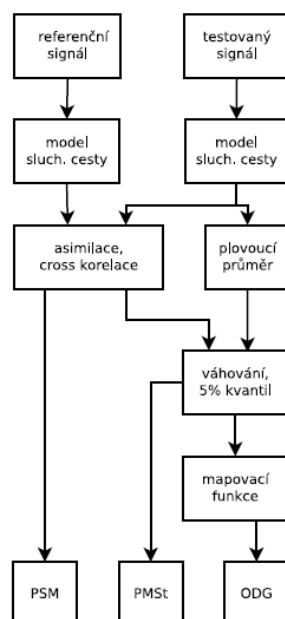
V roce 2006 Rainer Hubner a Birger Kollmeier publikovali článek s názvem *PEMO-Q - A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception* [12]. V tom vysvětlují, že PEMO-Q vychází z algoritmu PEAQ. Rozšiřuje ale jeho přístup o hledání obecných artefaktů jakým je například *pre-echo*³. Dále se pyšní použitím lepšího sluchového modelu, představeného v publikaci [36], který byl v rámci testování a optimalizace porovnáván s daty z šesti velkých subjektivních testů, které provedla ITU a MPEG v letech 1990 až 1995. Blokové schéma tohoto modelu je na obrázku 4.5. Funkce celého algoritmu poté na obrázku 4.6.



Obrázek 4.5: Blokové schéma modelu sluchové cesty v PEMO-Q[12]

Prvním blokem je banka třiceti pěti gammatónových filtrů simulující funkci bazilární membrány. Střední kmitočty jednotlivých filtrů jsou v rozsahu od 235 Hz až po 14,5 kHz. Každá větev je zpracovávána nezávisle. Následné usměrnění a filtrace dolní propustí reprezentuje převod zvuku na nervové impulzy. Zpracovaný signál je přiveden do bloku adaptivní filtrace složeného z pěti zpětnovazebních smyček zapojených do kaskády. Na konci cesty je další banka osmi filtrů modelujících schopnost rozpoznání amplitudově modulovaného zvuku.

³Zvuk vzniklý kompresí, který se ozývá dříve, než v původní nahrávce.



Obrázek 4.6: Blokové schéma algoritmu PEMO-Q [13]

Před kognitivním zpracováním je provedena asimilace signálu. Jedná se o matematický průměr vnitřních tří-dimenzoinálních reprezentací (frekvenční, časové a modulační) testovaného a referenčního signálu tam, kde testovaný nabývá nižších hodnot než signál referenční. Tato operace vychází z poznatku, že „chybějící“ prvky jsou méně rušivé než prvky „přidané“. Poté je provedena korelace pro každý modulační kanál zvlášť přes všechny časové a frekvenční hodnoty. Z kvadrátů výsledků korelačních funkcí je poté vypočítána hodnota *PSM* (*Perceptual Similarity Measure*) a z ní i hodnota *ODG*.

4.2.3 ViSQOLAudio

Evolucí původního algoritmu ViSQOL přestaveného v podkapitole 4.1.3 určeného k hodnocení hudebních signálů je tak zvaný ViSQOLAudio. V článku *ViSQOLAudio: An objective audio quality metric for low bitrate codecs* [37] autoři tvrdí, že k adaptaci původní metody na hodnocení hudebních nahrávek bylo třeba jen drobných změn.

Zde je jejich výčet:

- Systém porovnávání políček referenčního a testovaného signálu je zachován, nicméně všechna políčka jsou nyní označena za „aktivní“ na rozdíl od původních třech signifikantních pro řečové signály.
- Množství frekvenčních pásem při předzpracování bylo rozšířeno na celé slyšitelné pásmo od 50 Hz do 20 kHz.

Realizace subjektivních a objektivních testů

Dvakrát měř, jednou řež.
-české přísloví

K ověření předpokladu, zda objektivní algoritmy hodnotí poslechovou kvalitu nekořektně, k realizaci testů, k přípravě zvukových vzorků a k zobrazení výsledků bylo třeba využít poměrně početného programového vybavení, z jisté části ještě neexistujícího. Vzhledem ke zkušenostem autora s prostředím MATLAB z dob bakalářského studia, bylo pro tyto účely využito právě toto interaktivní programové prostředí a skriptovací programovací jazyk čtvrté generace vyvíjený společností MathWorks [38]. Tato kapitola popisuje jaké existující nástroje byly využity, jak byly propojeny pomocí skriptů napsaných v prostředí MATLAB a jaká uživatelská rozhraní byla vytvořena pro uskutečnění poslechových testů a zobrazování výsledků.

5.1 Příprava testovaného materiálu

Před provedením jakýchkoli testů, ať už poslechových či automatizovaných, bylo potřeba připravit množinu zvukových vzorků spolehlivě reprezentující sledované parametry. To znamenalo výběr implementace kodeků, tvorbu skriptů, které je budou dávkově volat a tím generovat vzorky o různých bitových rychlostech a dalších parametrech.

5.1.1 Použité kodeky

U standardů digitálního rádia je struktura jasně definována na vysílací straně a implementace přijímače je ponechána na výrobci zařízení. Ač to může znít paradoxně, u kodeků je tomu právě naopak. Dekodér je normou důkladně popsán, zatímco kodéry mohou být napsány (či sestaveny) různě. Velké společnosti na poli audiovizuálního zpracování, jako

například Fraunhofer či Apple, si licencují vlastní (proprietární) kodéry a nechávají si za ně zaplatit poměrně velké peníze. Proto se souběžně s nimi snaží open-source komunita vytvářet kodéry vlastní. U dvou vzorků stejného profilu a stejné bitové rychlosti se tak může lišit poslechová kvalita, kvůli čemuž je důležité poznamenat jakým kodérem byl signál zakódován.

Na následujících řádcích jsou vyjmenovány implementace kodérů a dekodérů použitých v této práci.

- o TwoLame:

použitý ke kódování vzorků do MPEG 1 Layer II je adaptací původního freeware kodéru známého jako tooLame. Je volně dostupný z webových stránek [39]. Skript `makeMp2` volá tento kodér napsaný v jazyce C a generuje vzorky s následujícími parametry:

- konstantní bitová rychlost (kb/s): 56, 64, 80, 96, 112, 128, 160, 192, 224, 256, 320, 384
- mód: stereo
- vzorkovací rychlost 44100 Hz

Pro objektivní hodnocení je ovšem nutno mít signál uložený v nezakódované formě. Vzorky tudíž byly převedeny zpět do pulzní kódové modulace pomocí programu FFmpeg [40].

- o Nero AAC Codec

I u rodiny AAC byl v této práci použit freeware nástroj. Nero AAC Codec dostupný z [41] dokáže signál jak kódovat, tak převádět zpět do formy PCM. Dostupné profily jsou: AAC LC, HE-AAC a HE-AAC v2. Skript `makeAac` volá program `NeroAacEnc.exe` a vytváří vzorky s parametry:

- konstantní bitová rychlost : 12 - 256 kb/s pro profil AAC LC, 16 - 128 kb/s pro profil HE-AAC a 12 - 64 kb/s pro profil HE-AAC v2 vždy s krokem 4 kb/s.
- mód: stereo (pro HE-AAC v2 mód mono neexistuje, neboť využívá Parametrického sterea)
- vzorkovací rychlost 44100 Hz

- o Opus

... je od základu vyvíjen jako open-source. Kodek je proto ke stažení přímo na stránkách standardu [9] a není třeba složitě shánět alternativy ve formě komunitních projektů. Podobně jako u dvou předchozích byl v matlabu vytvořen skript pro generování vzorků s následujícími parametry:

- konstantní bitová rychlost : 8 - 128 kb/s s krokem 4 kb/s.

- mód: stereo
- vzorkovací rychlost 44100 Hz
- xHE-AAC

Kódování do xHE-AAC je prozatím problematické neboť jediný současný enkodér poskytuje Fraunhofer Institute za cenu pro komerční využití. Bylo tak třeba porozhlédnout se po zdroji zakódovaných vzorků jinde. Na webu DRM+ [42] je k dispozici ukázkové video v nekomprimované verzi obsahující tři vzorky jak v originálu, tak zakódované pomocí xHE-AAC do dvou bitových rychlostí. Konkrétně do 8 a 12 kb/s. Pomocí programu Avidemux [43] byla z videa extrahována zvuková stopa, ze které byly programem Audacity [44] vystříhány konkrétní zhruba osmi vteřinové vzorky.

5.1.2 Typy materiálu z hlediska obsahu

Doposavad se tato kapitola věnovala tomu, jak hudební vzorky zakódovat a rozkódovat. Neméně důležitým aspektem je ale skutečný obsah nahrávek použitých k testování kvality. Jelikož v éteru rádia existuje mnoho stanic vysílající různý druh obsahu od mluveného slova přes klasickou hudbu až po nezařaditelné alternativní hudební žánry, bylo třeba do palety zpracovávaných vzorků zahrnout různorodý materiál.

Tři vzorky z DRM+ videa představovaly tři kategorie, do kterých jsou nadále rozděleny všechny nahrávky figuruující v této práci. Jedná se o signály:

1. hudební - v indexech a skriptech značené jako *music*
2. řečové - s označením *speech*
3. smíšené - s označením *mixed*

Dalším zdrojem nahrávek byla samotná norma ITU-R BS.1387-1 [33], respektive CD přiložené k softwaru Opera [35] s nahrávkami, které se používají při implementaci algoritmu PEAQ. Jedná se o izolované ukázky různých zvuků či nástrojů jako například klavír, saxofon a zpěv. V rádiu se ovšem takto osamostatněné nástroje vyskytují jen zřídka a proto z nich byl vybrán pouze jeden vzorek, konkrétně úryvek vokální skladby *Tom's Diner* zpěvačky Suzanne Vega, který se ve světě multimediální techniky používá velmi často a může tak fungovat jako určitá reference vůči ostatním testům. Zbytek nahrávek uvedených v tabulce 5.1 připravil Miroslav Smutný při spolupráci na článku *Perceived Audio Quality Analysis in Digital Audio Broadcasting Plus System Based on PEAQ* [45].

Hlavně pro potřeby subjektivního testování byly za pomoci programu Audacity z nahrávek vybrány relevantní desetivteřinové úseky, které byly dále opatřeny pozvolným zesílením v první vteřině a pozvolným útlumem ve vteřině poslední.

Název vzorku	Typ	Popis	Zdroj
'capriccio.wav'	Music	Úryvek klasické hudby	Smutný
'holmes.wav'	Speech	Čtení z knihy Sherlock Holmes	Smutný
'cimrman.wav'	Speech	Záznam hry Opeřený Had	Smutný
'dubstep.wav'	Music	Elektronická hudba	Smutný
'pennylane.wav'	Music	Úryvek písně od The Beatles	Smutný
'refveg.wav'	Music (vocal)	A Capella zpěv	Opera
'frekv3.wav'	Mixed	Řeč s hudebním podkresem	Smutný
'xhedemomusic.wav'	Music	Hra na elektrickou kytaru	DRM
'xhedemospeech.wav'	Speech	Úryvek řeči moderátora BBC	DRM
'xhedemomixed.wav'	Mixed	Řeč s hudebním podkresem	DRM

Tabulka 5.1: Seznam vzorků použitých při objektivním testování kvality

5.2 Návrh subjektivních testů

Jak už je v úvodu naznačeno, objektivní hodnocení nahrávek zakódovaných novými kodeky využívajícími psychoakustické poznatky ve větší míře nekoresponduje s realitou. Na tento fakt již poukazuje akademický článek *Subjective and Objective Assessment of Perceived Audio Quality of Current Digital Audio Broadcasting Systems and Web-Casting Applications*. [46], kde je sledována závislost *ODG* a *SDG* při použití hodnotících algoritmů PEAQ a POLQA. Subjektivní testy provádí párovým srovnáváním podle [15], tedy na velké skupině expertních posluchačů. Pro potřeby této práce se ovšem lépe hodí test s vícenásobným stimulem dle [26]

5.2.1 Důvody pro výběr testu MUSHRA

- Zkoumaná kvalitativní oblast

Protože bylo třeba ověřit zda algoritmy hodnotí věrohodně nejen tam kde je komprese nepostřehnutelná, ale skrze celé kvalitativní spektrum od výborné až po nedostatečnou, je test pro střední kvalitu podle [26] mnohem vhodnější než test pro nepatrná zhoršení použitý v [46]. Největší důraz bylo třeba klást na bitové rychlosti využívané v digitálním rádiu DAB/DAB+ a DRM.

- Posluchači

Subjekty nemusí procházet sérií rozřazovacích vstupních testů. Stačí pouze zaučení přímo v grafickém uživatelském rozhraní samotného testu a doprovodné informace v příloženém návodu (K nahlédnutí v příloze B).

- Podmínky

MUSHRA nemá tak přísné podmínky na provádění testu. K jeho provedení není třeba poslechová místnost s extrémně kvalitním reprodukčním zařízením. Stačí dostatečně

kvalitní sluchátka a úroveň zvukového pozadí taková, aby nerušila posluchače. Další výhodou tohoto testu je, že lze provádět individuálně. Je tedy možné přenášet laptop se zvukovou kartou a sluchátky za subjekty a tím ušetřit čas i náklady.

5.2.2 Výběr vzorků

Jednou z nesporných výhod objektivního testování je absence únavy posluchačů, se kterou je třeba počítat při návrhu subjektivního testu. Doporučení [26] říká, že délka testu by neměla překročit půl hodiny. Přímo naproti tomu jde fakt, že k zachování objektivity je třeba dostatečná diverzita poslechového materiálu. Výběr těch správných vzorků se tak stává poměrně zásadním faktorem pro výpovědní hodnotu navrženého testu.

Jak už je zmíněno v podkapitole 5.1.2. Vzorky byly rozděleny do třech kategorií. Hudební, řečové a smíšené. V druhých dvou zmíněných nedosahuje míra různorodosti vzorků takové výše jako v kategorii hudební. Větší důraz byl tedy kladen právě na ni, a proto je v subjektivním testování zahrnuto několik vzorků reprezentujících různé hudební žánry.

Stimuly musí být, co do metody zkeslení, konzistentní skrze všechny zkoušky testu a nemožnost kódování nových zvukových záznamů do xHE-AAC tak značně zkomplikovala přípravu vzorků. Nakonec byly vytvořeny testy dva. Jeden na vyšších bitových rychlostech, obsahující všechny kodeky vyjma xHE-AAC. A druhý test na nižších datových tocích, kde společně s xHE-AAC vystupovaly zástupci novodobých kódovacích metod.

Zkoušky sestavených testů tedy obsahují vzorky uvedené v tabulce 5.2.

Název vzorku	Typ	Popis	Test
'capriccio.wav'	Music	Úryvek klasické hudby	1
'holmes.wav'	Speech	Čtení z knihy Sherlock Holmes	1
'pennylanew.wav'	Music	Úryvek písně od The Beatles	1
'refveg.wav'	Music (vocal)	A Capella zpěv	1
'xhedemomusic.wav'	Music	Hra na elektrickou kytaru	2
'xhedemospeech.wav'	Speech	Úryvek řeči moderátora BBC	2
'xhedemomixed.wav'	Mixed	Řeč s hudebním podkresem	1,2

Tabulka 5.2: Seznam vzorků použitých v subjektivních testech

bitrate (kpbs)	24	32	48	64	96	128
Opus	•	•				
HE-AAC v2		•	•			
HE-AAC v1			•	•		
LC-AAC				•		
MPEG1 Layer II					•	•

Tabulka 5.3: Použité bitové rychlosti v testu č. 1

bitrate (kpbs)	8	12	24	48	64
xHE-AAC	•	•			
Opus		•	•		
HE-AAC v2			•	•	•

Tabulka 5.4: Použité bitové rychlosti v testu č. 2

Konkrétní bitové rychlosti použité v testech 1 a 2 jsou zobrazené v tabulkách 5.3 a 5.4. Konečná volba proběhla na základě následujících kritérií:

- Bitové rychlosti používané v digitálním rádiu:

Například *Advanced Audio Coding* se v systémech digitálního rádia na území České Republiky [47] nachází ve verzích HE-AAC v1 a HE-AAC v2, tedy v základní verzi obohacené o SBR potažmo o PS. V zahraničí lze nalézt multiplexy v nichž se vysílá AAC LC profilem avšak není jich mnoho. Velký důraz je tak kladen na HE-AAC v2 profil. O něco menší na HE-AAC v1 a Low Complexity je zastoupen v testech jen jednou.

- Oblasti kde se výsledky různých algoritmů rozcházejí v předběžných testech:

Například MPEG Layer II byl v předběžných testech hodnocen různými algoritmy s velkými kvalitativními rozestupy. Vzorke zkomprimované do *mp2* se proto v testu nachází ve dvou bitových rychlostech.

- Výběr stejných bitových rychlostí různých kodeků k porovnání jednotlivých kompresních metod:

Kupříkladu k určení toho, jak moc velký vliv mají rozšíření SBR a PS na výslednou kvalitu.

- Sledování spíše nižších rychlostí pro určení hranice nedostatečné poslechové kvality.

Zeleně jsou zvýrazněny vzorky, ve kterých se testy vzájemně překrývají.

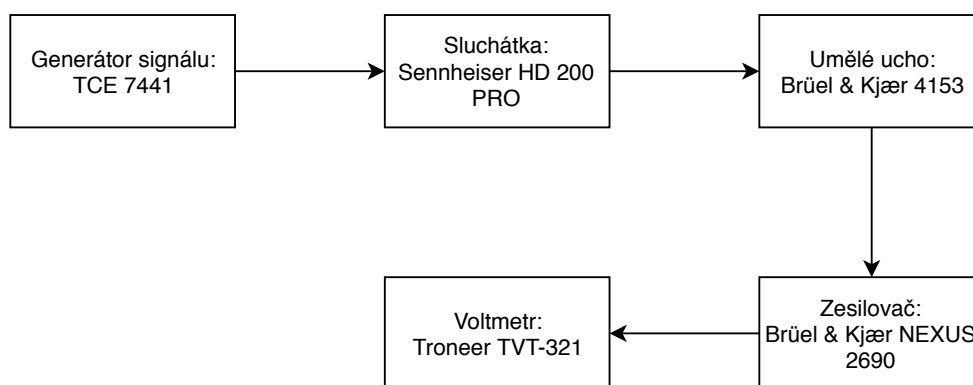
5.2.3 Použité vybavení

Provést subjektivní testy v poslechové místnosti velkou skupinou posluchačů se ukázalo jako časově a hlavně koordinačně velmi náročné. Byla proto zvolena varianta testu přenosného, který lze realizovat kdykoliv a kdekoliv, kde není přítomno zvukové rušení. Testy byly spouštěny přímo v prostředí MATLAB běžícím na čtrnácti palcovém notebooku ThinkPad E431. Jako poslechové zařízení byla použita studiová sluchátka, používaná na katedře Radioelektroniky fakulty Elektrotechnické ČVUT k poslechovým testům v rámci výuky zvukové techniky. Konkrétně šlo o model Sennheisser HD 200 PRO. Parametry udávané výrobcem jsou v tabulce 5.5.

Harmonické zkreslení (THD):	<0,1 %
Jmenovitá impedance:	32 Ohm
Délka kabelu:	2 m
Frekvenční rozsah:	20 - 20 000 Hz
Hmotnost:	184 g
Konstrukce:	uzavřená, dynamická
Akustický tlak:	108 dB (1 Vrms)

Tabulka 5.5: Parametry sluchátek Sennhesier HD 200 PRO udávané výrobcem [18]

Zda frekvenční charakteristika sluchátek odpovídá přísné toleranční masce podle ITU-R BS.708 [10] bylo ověřeno měřením na umělém uchu Brüel & Kjær 4153 [48]. Blokové schéma zapojení se nachází na obrázku 5.1.



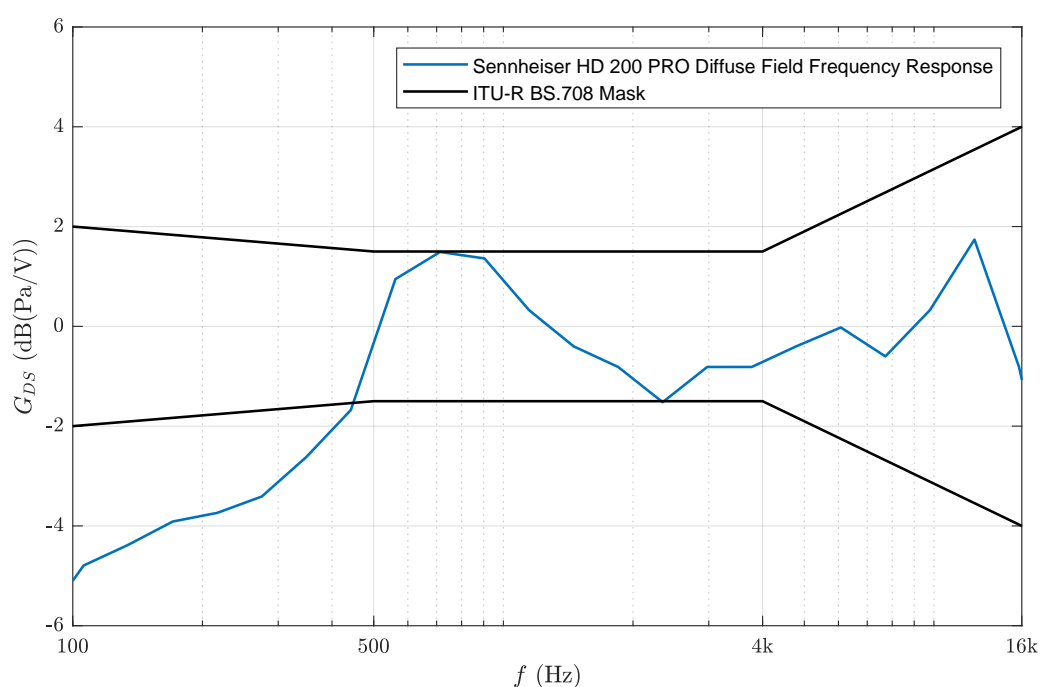
Obrázek 5.1: Blokové schéma zapojení měřicího pracoviště při měření sluchátek Sennhesier HD 200 PRO

Naměřená frekvenční charakteristika je na obrázku 5.2. V nejvyšší části basového pásma a nejnižší oblasti středních kmitočtů nesplňuje přísné hranice masky ITU-R. Sluchátka nicméně byla použita s přihlédnutím k následujícím kritériím.

- Doporučení ITU-R BS.1534 MUSHRA se na kvalitu sluchátek odvolávají skrze normu ITU-R BS.1116 pro měření nepatrných zhoršení. Testování v rámci této práce se však zabývá střední kvalitou.
- Frekvenční charakteristika sluchátek Sennheiser HD 280 PRO, která podle webové stránky rtings.com [49], věnující se podrobnému měření studiových sluchátek, dosahuje velmi podobného zvlnění a vybočuje tak z masky dané doporučením BS.708 stejně jako Sennheiser HD 200 PRO. I přes to jsou sluchátka HD 280 PRO použita v následujících akademických pracích, které taktéž k hodnocení kvality využívají test MUSHRA.
 - *Considering Bluetooth's Subband Codec (SBC) for Wideband Speech and Audio on the Internet* [50] - článek nalezený na webu *Researchgate.com* zabývající se

problematikou kódování širokopásmového zvuku pro přenos internetovým rádiem.

- *Acoustic echo reduction robust against echo-path change with instant echo-power-level adjustment* [51] - článek nalezený na webu *ieeexplore.com* který se zabývá odstraňováním dozvuku v telekomunikacích.



Obrázek 5.2: Frekvenční charakteristika sluchátek Sennhesier HD 200 PRO změřená na umělém uchu Brüel & Kjær

K převodu na analogový signál byla použita externí zvuková karta AxagonHQ ADA-HP, jejíž relevantní technické parametry udávané výrobcem jsou uvedeny v tabulce 5.6. Hladina okolního šumu nebyla při provádění testu měřena, nicméně byl kladen důraz na to, aby ruch okolí nerušil subjekty při poslechu.

5.2.4 Grafické uživatelské rozhraní

Jelikož MUSHRA test se využívá k poslechovým testům více než deset let, existují již před-připravené skripty a programy do kterých stačí naimportovat vzorky a není třeba se zabývat programováním GUI a vyhodnocovacími algoritmy. Většina takových programů je ovšem zatížena licenčními poplatky. Vedle komerčních produktů existují i komunitou psané freeware programy jako jsou WhisPER[52], APE[53] a Scale.

Analogový výstup	2kanálový stereo výstup pro sluchátka nebo aktivní reproduktory
Vzorkovací frekvence	44.1 / 48 / 96 kHz analog
Rozlišení	16 / 24 bit pro přehrávání
Odstup signál / šum	$\geq 95\text{dB}$ (Output SNR) pro DAC
Harmonické zkreslení	$\leq -85\text{dB}$ (Output THD+N)
Sluchátkový zesilovač	20 - 20.000 Hz, výkon 105 mW / 16 ohm, při 1kHz 0.1% THD+N, 16 - 100 Ω

Tabulka 5.6: Parametry zvukové karty AxagonHQ ADA-HP udávané výrobcem [19]

Po zvážení byl vybrán právě program Scale [54] připravený pro použití v jazyce MATLAB. Jeho předností by měla být implementace rozhraní pro provádění všech testů popsanych v kapitole 3. Kromě samotného testu se pyšní i schopností výsledky vyhodnotit.

Po několika pokusech využít Scale se ovšem začaly objevovat problémy. Při větším množství vzorků nahraných do rozhraní MUSHRA prostředí MATLAB hlásilo chyby, vyhodnocování bylo spolehlivé pouze pro test pro nepatrná zhoršení [15] a jednodušší než opravovat existující program se začalo zdát napsat si program od základu.

Pro návrh grafického uživatelského rozhraní se v matlabu nachází komponenta GUIDE [55]. Jedná se o nástroj, ve kterém se pomocí kurzoru myši dají rozmístit objekty z nabídky do prostoru okna, které se poté oživují psaním funkcí zvaných *callbacks*. Ty jsou vyvolány akcemi jako například kliknutím myši, či stiskem klávesy. Objekty jsou definovány souborem vlastností uložených ve speciálním datovém typu *handle* a jdou dle potřeby vyčítat či upravovat obdobně jako standardní datové proměnné v klasických funkcích. Na následujících řádcích je stručně popsáno výsledné rozhraní, ve kterém byly testy MUSHRA realizovány.

- Uvítací okno

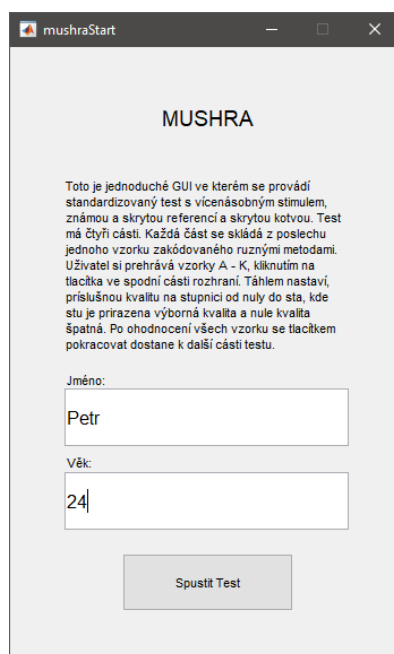
Jednoduché GUI s rámcovým popisem testu, dvěma dialogy pro zadání jména a věku subjektu a tlačítkem pro přechod do další fáze testu. Uvítací okno je k vidění na obrázku 5.3.

- Zaučení

Okno velmi podobné „ostrému“ testování. Na rozdíl od něj zde nejsou vzorky seřazeny náhodně, ale pod každým tlačítkem se nachází zvuková nahrávka odpovídající popisu v pravé části okna. Uživatel si takto může jednotlivé typy stimulů projít a naučit se, co je to „kotva“, „skrytá reference“ a na jaké druhy zkreslení může při testu narazit.

- Okno pro jednu zkoušku testu

Toto okno se přizpůsobuje na základě toho jaký návrhář testu zvolí počet zkoušek a vzorků. Soubor `scenarios.txt` v kořenovém adresáři testu, obsahuje názvy jednot-



Obrázek 5.3: Grafické rozhraní uvítacího okna

livých zkoušek oddělených novým řádkem. Pro představu test popsany v tabulce 5.3 je zapsán takto:

Klasická hudba
Rock
Mluvené slovo
Zpěv(acapella)
Mixed

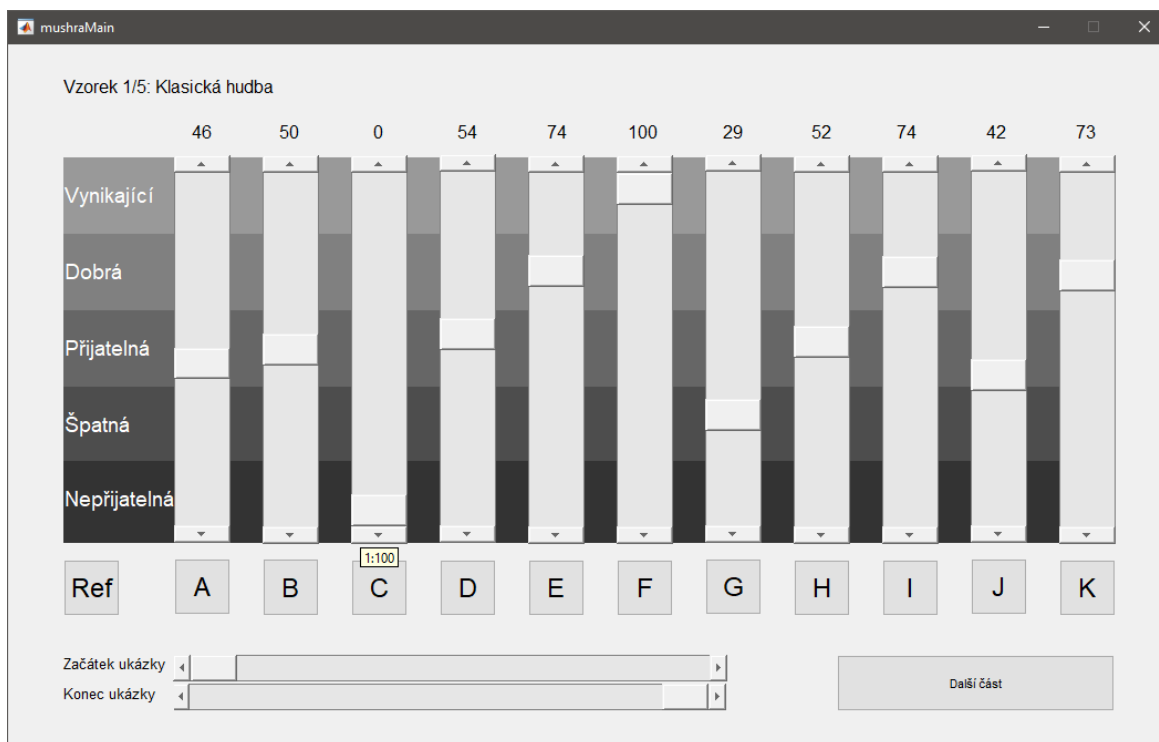
Tomu musí odpovídat počet podsložek uložených do složky **scenarios**. Podle počtu vzorků v podsložkách se aktivuje počet přehrávacích tlačítek hodnotících posuvníků. Maximálně jich ovšem může být 12 včetně dvou referencí a jedné kotvy.

Nad každým posuvníkem se nachází udělené hodnocení v procentech. Pro lepší orientaci jsou posuvníky podkresleny barevně odlišenými oblastmi, které odpovídají kvalitativním úrovním definovaným v doporučení ITU-R BS 1534.

Po vyplnění všech zkoušek je test ukončen dialogovým oknem s nápisem „Konec!“, které volá funkci pro export výsledků do souboru.

O průměrování, post-screening a zobrazení do sloupcových grafů se stará skript **displayResults**, taktéž umístěný v kořenovém adresáři testu.

Posuvníky ve spodní části okna nazvané „Začátek ukázky“ a „Konec ukázky“ slouží k časovému omezení všech přehrávaných vzorků dané zkoušky. Touto funkcí lze využít při rozhodování pouze na základě krátkého úseku vzorku. Například při podezření, že se v ukázce vyskytuje kompresní artefakt.



Obrázek 5.4: Grafické rozhraní průběhu testu

5.3 Použité implementace algoritmů pro objektivní testování kvality audiosignálů

5.3.1 PEAQ

Matlabová implementace PQevalAudio základního modelu PEAQ je volně k dispozici na stránkách univerzity McGill [56]. V předběžných testech byl využit právě tento model, nicméně sami autoři v poznámkách tvrdí, že slouží pro výukové účely a nikoliv pro výzkum na poli zvukového zpracování, jelikož nesplňuje všechna kritéria vytyčená doporučením [33]. Proto se v praxi se základní verzí algoritmu příliš nesetkáváme. Pokročilá verze byla implementována v roce 2003 do programu s názvem Opera¹ německou firmou OPTICOM [35], která se soustředí na vývoj softwaru pro hodnocení kvality zvuku již od roku 1995 a podílela se i na vzniku standardu [33]. Katedra radiotechniky ČVUT má zakoupenou licenci tohoto programu pro akademické využití. K ověření platnosti licence využívá USB klíčenku, kterou je třeba mít připojenou k PC po celou dobu běhu algoritmu.

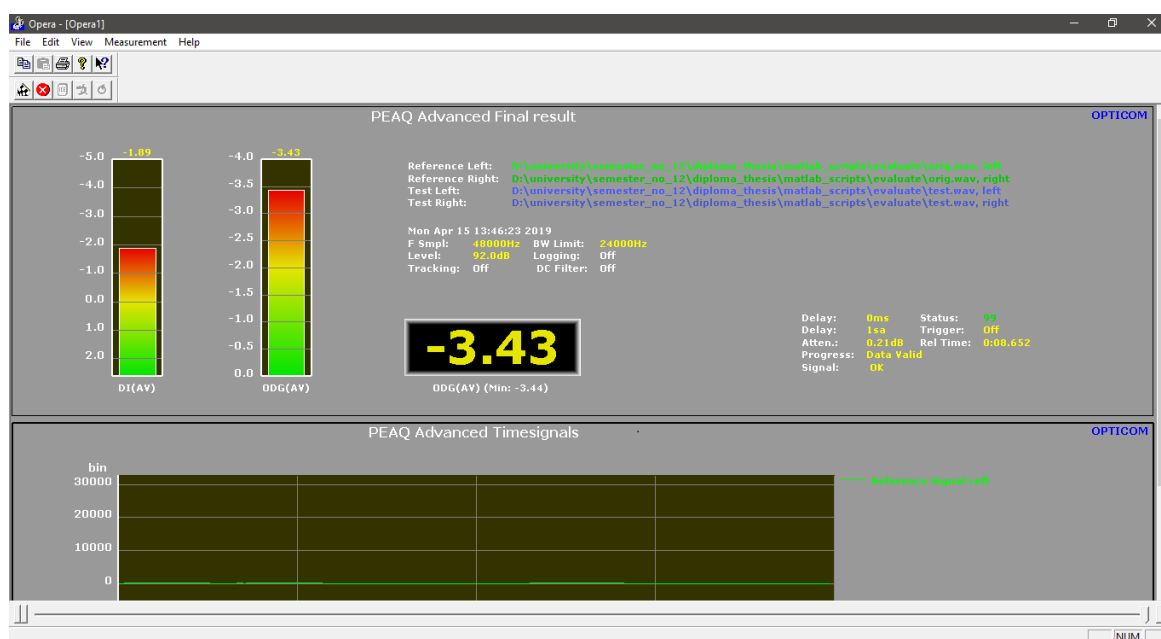
Rozhraní programu (na obrázku 5.5) je poměrně spartánské. V levém horním rohu se nachází tlačítka pro obsluhu. Zbytek prostoru je vyhrazen pro prezentaci výsledků hodnocení. Tlačítkem „Start“ s obrázkem vlajčky se spouští průvodce importem vzorků. Lze

¹Opera obsahuje i základní verzi algoritmu PEAQ.

5. REALIZACE SUBJEKTIVNÍCH A OBJEKTIVNÍCH TESTŮ

použít různé mapování kanálů, kompenzaci zpoždění, a několik filtrů, například odstranění stejnosměrné složky.

Po zadání všech parametrů a cest k referenčnímu a testovanému vzorku přejde program k hodnocení a po dokončení běhu algoritmu zobrazí výsledky jako například: zpoždění, rozdíl hlasitostí vyjádřený v úrovních, délky vzorků a především *ODG* s grafickou reprezentací.



Obrázek 5.5: Uživatelské rozhraní programu OPERA

Takovýto postup je časově náročný a při hodnocení stovek až tisíců vzorků nepoužitelný. Naštěstí lze Opera volat z příkazové řádky a tím je proces automatizovatelný. Volání programu `opera.exe` vrací záznam ve formě tabulky v textovém souboru, ze kterého lze parametry předat jiným programům.

I přes to, že se v akademické sféře používá pokročilý model, čistě ze zvědavosti autora jsou v této práci zahrnuty výstupy obou verzí algoritmu počítané programem Opera.

5.3.2 PEMO-Q

Balík `OpenQual+` vznikl jako přepis implementace algoritmu PEMO-Q z diplomové práce Martina Zalabáka [13] užitím principů objektově orientovaného programování. Zároveň je kód upravený tak, aby poskytoval výsledky blížící se verifikované komerční implementaci PEMO-Q od HörTech gGmbH [57]. Jeho autor Jan Novák ve své bakalářské práci [58] zmiňuje, že balík je optimalizovaný pro MATLAB 2017a a novější. Ve verzi 2018b nicméně vykazoval značně podivné chování často končící chybou, což je důvod proč je v této práci použito právě prostředí 2017a i přes svou téměř dvou roční zastaralost. Smyslem objektově orientovaného přístupu při programování `OpenQual+` byla myšlenka jednoduché rozšíři-

telnosti o další hodnotící algoritmy. V této práci je ovšem využita pouze pro hodnocení metodou PEMO-Q.

Inicializace balíku probíhá skrze metodu `getEvaluationEngine`, která po zadání parametrů vrátí instanci třídy modelu, se kterým se dá následně pracovat jako s obyčejnou funkcí pomocí příkazu `engineInstance.evalObjQuality(y,y1)`, kde `y` je referenční vzorek a `y1` je vzorek testovaný. Ve srovnání s programem Opera je hodnocení poměrně časově náročné.

5.3.3 ViSQOL Audio

Zprovoznění implementace algoritmu ViSQOL bylo ze všech použitých implementací nej-jednodušší. Jedná se o funkci `visqol.m` napsanou v jazyce MATLAB dostupnou z [59] se vstupy pro referenční signál, testovaný signál a parametry. Těmi je konkrétně myšleno, zda bude použit základní model pro hodnocení hlasu či pokročilý ViSQOL-Audio.

Jelikož původně byl algoritmus ViSQOL vyvíjen k hodnocení kvality přenosu řeči v telekomunikacích, neposkytuje výsledky ve formě *ODG*, ale na škále *MOS-LQO* (*Mean Opinion Score - Linear Quadrature Objective*). Je to taktéž jediný výstup funkce `visqol.m`, což může být bráno jako nevýhoda oproti algoritmům PEAQ a PEMO-Q, které disponují větším množstvím výstupních parametrů.

5.4 Popis ostatních funkcí, skriptů a rozhraní

V průběhu této diplomové práce bylo připraveno v rozhraní MATLAB několik skriptů, funkcí a programů, jenž slouží ke generování vzorků, jejich hodnocení a zobrazování výsledků. V následujících podkapitolách je popsána jejich funkce a ovládání.

5.4.1 main

Main je skript propojující většinu funkcí, skriptů a podprogramů, použitých při vyhodnocování kvality zvuku. Jeho kompletní kód je v digitální příloze. Postupně se na základě uživatelského vstupu stará o:

- inicializaci cest
- přípravu zakódovaných souborů
- hodnocení objektivními algoritmy
- zobrazení výsledků

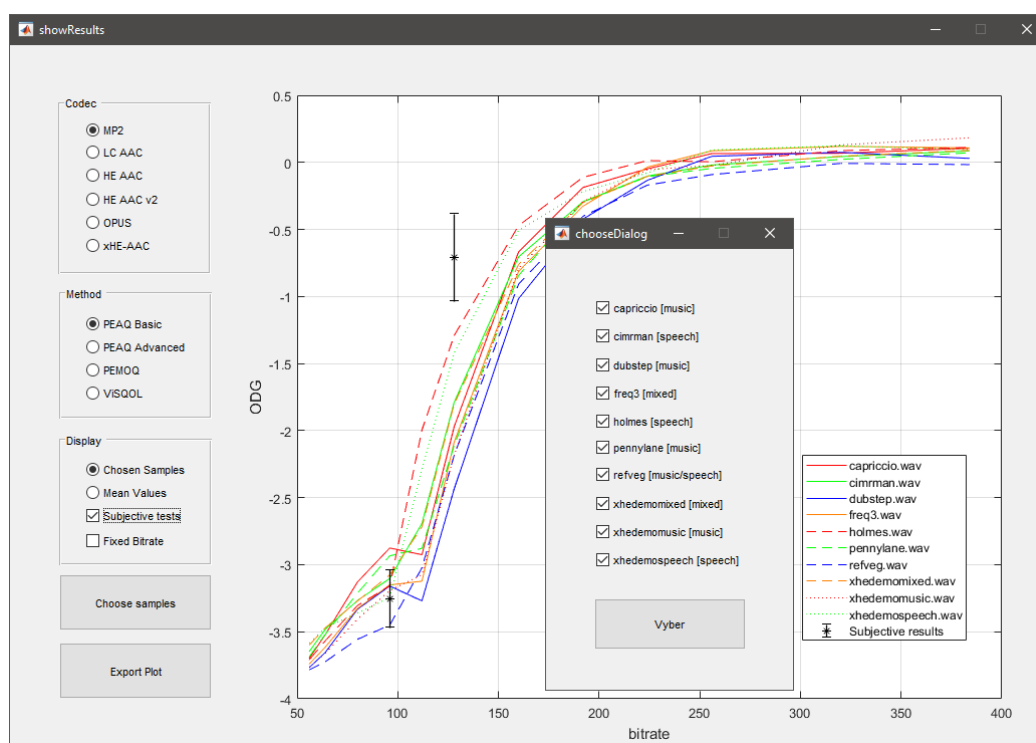
5.4.2 evaluate

Funkce `evaluate` provádí sekvenčně několik operací. Na základě vybraného kodeku přidělí možné bitové rychlosti, ve kterých je hodnocení možné. Následně načte testovaný vzo-

rek, k němu ve složce **original** vyhledá referenci a vyřeší časové zarovnání, bez něhož by hodnocení nebylo korektní. Poté oba vzorky předá hodnoticímu algoritmu, popřípadě provede převzorkování na jiný kmitočet, a výsledné hodnocení uloží do souboru pro další zpracování.

5.4.3 showResults

Grafické rozhraní **showResults** slouží k prezentaci závislostí *ODG* na bitové rychlosti, získaných z výstupů objektivních algoritmů pomocí funkce **evaluate**. K návrhu byl využit systém GUIDE [55]. Okno programu na obrázku 5.6 je rozdělené do dvou částí. Vlevo se nachází ovládací prvky, vpravo je prostor pro zobrazení grafů. Vykreslení grafů reaguje okamžitě na uživatelem volenou kombinaci kodeků (přepínačem *Codec*) a hodnotící metody (přepínačem *Method*). Tlačítkem *Choose Samples* se otevře dialogové okno **ChooseDialog**, ve kterém je možné vybrat si pouze vzorky, jenž uživatele zajímají. Přepínačem *Display* lze vykreslit celou množinu závislostí, či jejich aritmetický průměr společně se směrodatnou odchylkou. Hodnoty na ose *x* lze buďto nechat automaticky se přizpůsobovat obsahu grafu, nebo lze zatrhnacím boxem *Fixed Bitrate* zvolit pevný rozsah v intervalu $\langle 12; 384 \rangle$ kb/s.



Obrázek 5.6: Grafické rozhraní **showResults**

Zatrhnací box *Subjective tests* umožňuje zobrazit výsledky subjektivních testů, pokud jsou k dispozici. Protože rozhraní **showResults** bylo vymyšleno pro vykreslování výsledků hodnocení objektivními algoritmy, je tato funkce pouze doplňková a nectí set vybraných

vzorků tlačítkem *Choose Samples*, který ostatně vůbec nemusí být shodný. Při kliknutí myši na tento box se zobrazí varování o tom, na kterých vzorcích byl subjektivní test proveden. Je ponecháno na uživateli, aby dbal na to jaká data porovnává. Grafy lze exportovat tlačítkem *Export Plot* do formátů pdf, png, jpeg a bmp..

5.4.4 makeCell

Výsledky objektivního hodnocení je do zobrazovače `showResults` třeba importovat. Skript `makeCell` je jednoduchý kód jehož úkolem je kumulace výsledků (závislostí *ODG* na bitových tocích) z funkce `evaluate` do jednoho souboru, se kterým se bude lépe pracovat. V Matlabu pro takové aplikace existuje datový typ zvaný buňka (*cell*), do kterého jdou nořit libovolné datové typy. Buňka vytvořená pomocí funkce `makeCell` má následující kořenovou strukturu:

```
cellName/algorithms/codecs/odgs
```


Výsledky testů

6.1 Výsledky subjektivních testů

Nakonec byly oba testy představené v kapitole 5 provedeny sedmnácti subjekty ve věkovém intervalu od 21 do 62 let. Šest z nich mělo předchozí zkušenosti s poslechovými testy. Poměr mužských posluchačů vůči ženám byl přibližně 2:1. Všem byly poskytnuty průvodní informace, prošli fází zaučení a podepsali informovaný souhlas.

Dvě formy post-screeningu byly aplikovány na množinu získaných hodnocení. První za pomoci skryté reference a druhá pomocí skryté kotvy. Doporučení říká, že subjekty jejichž hodnocení skryté reference je v patnácti procentech případů nižší než devadesát procent, by měli být vyřazeni. To je poměrně přísná podmínka a při jejím aplikování by bylo vyřazeno více než 60 % subjektů. Protože test míří na běžné posluchače rozhlasového vysílání, nikoliv na studiový poslech znalými posluchači, bylo toto kritérium poníženo a vyřazeni byli ti, jejichž průměrné hodnocení vzorků bylo nižší než 90 %. Zároveň nejsou zahrnuta hodnocení posluchačů, jejichž průměrné hodnocení skryté kotvy bylo vyšší než 50 %. Prvním kritériem byly vyřazeni tři posluchači a druhým jeden další. Celkový počet subjektů s relevantním hodnocením byl tedy roven $N = 12$.

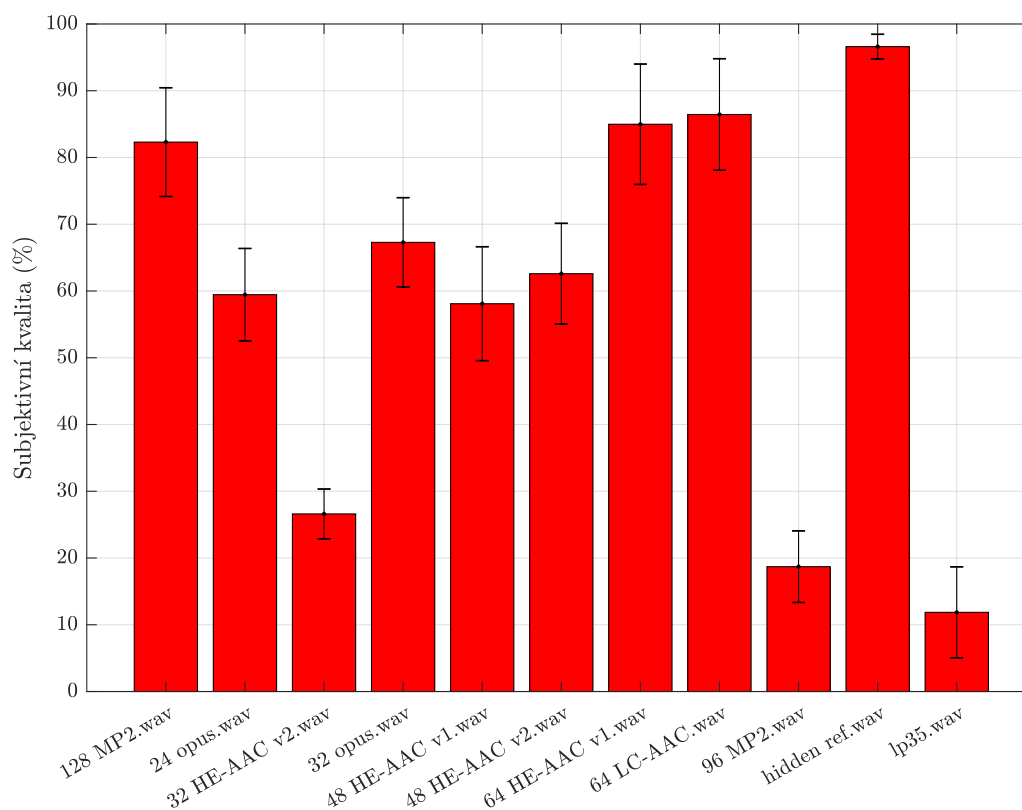
Výsledky poslechových testů jsou vyneseny do tabulek 6.1 a 6.2. Zároveň jsou vyobrazeny pomocí sloupcových grafů, které ukazují průměr subjektivně vnímané poslechové kvality \bar{u}_{jk} . Na obrázcích 6.1 a 6.3 lze vidět průměrné hodnocení přes všechny zkoušky. Grafy 6.2 a 6.4 tyto hodnoty dělí do třech sledovaných kategorií: hudební, řečové a smíšené. Všechny výsledky jsou doplněny chybovými úsečkami vyznačující interval spolehlivosti o velikosti $2\delta_k$ vypočítaný podle vztahu 3.4, ve kterém se s 95% pravděpodobností nachází populační průměr.

Jako ukazatel výpovědní hodnoty dosažených výsledků lze brát shodu v hodnocení vzorků konzistentních pro oba testy (zeleně vyznačených v tabulkách 5.3 a 5.4). Průměrné hodnocení stimulů zakódovaných kodekem HE-AAC v2 o bitové rychlosti 48 kb/s dosáhlo hodnot $\bar{u} = 62,6\%$ s velikostí intervalu spolehlivosti $2\delta = 15\%$ pro první test a $\bar{u} = 61\%$ v intervalu s velikostí $2\delta = 30\%$ pro test druhý. Druhý set vzorků určených pro kontrolu hodnocení mezi dvěma testy byl zkomprimován kodekem Opus s bitovou rychlostí 24 kb/s.

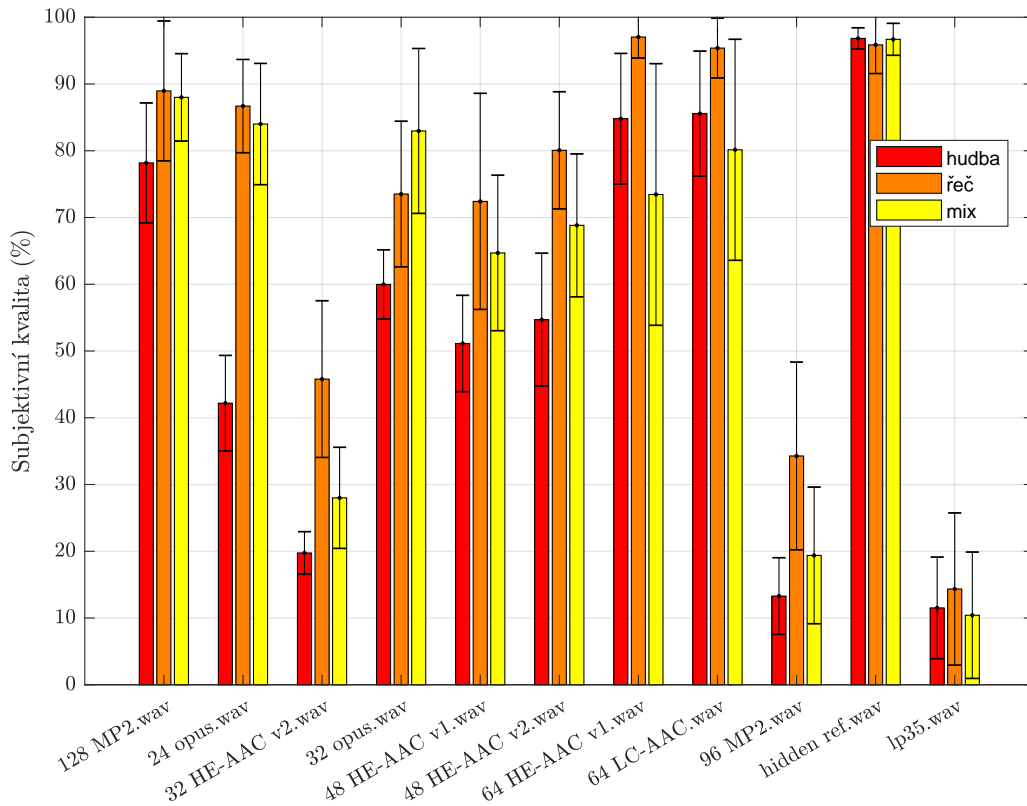
6. VÝSLEDKY TESTŮ

	128 MP2	24 opus	32 HE-AAC v2	32 opus	48 HE-AAC v1	48 HE-AAC v2	64 HE-AAC v1	64 LC-AAC	96 MP2	hidden ref	lp35
\bar{u}_k	82,3	59,4	26,6	67,3	58,1	62,6	85,0	86,4	18,7	96,6	11,9
δ_k	8,1	6,9	3,7	6,7	8,5	7,5	9,0	8,4	5,4	1,9	6,8
$\bar{u}_{jk(music)}$	78,2	42,2	19,7	60,0	51,1	54,7	84,8	85,6	13,3	96,8	11,5
$\delta_{jk(music)}$	9,0	7,2	3,2	5,2	7,2	10,0	9,8	9,4	5,7	1,6	7,6
$\bar{u}_{k(speech)}$	89,0	86,7	45,8	73,5	72,4	80,1	97,0	95,4	34,3	95,9	14,3
$\delta_{k(speech)}$	10,5	7,0	11,7	10,9	16,2	8,8	3,1	4,5	14,1	4,3	11,4
$\bar{u}_{k(mixed)}$	88,0	84,0	28,0	83,0	64,7	68,8	73,4	80,1	19,4	96,7	10,4
$\delta_{k(mixed)}$	6,5	9,1	7,6	12,4	11,6	10,7	19,6	16,6	10,2	2,4	9,5

Tabulka 6.1: Výsledky prvního testu



Obrázek 6.1: Grafické zobrazení výsledků prvního testu



Obrázek 6.2: Grafické zobrazení výsledků prvního testu rozdělené do kategorií

Průměrné hodnocení posluchačů v prvním testu $\bar{u} = 59,4\%$ na intervalu $2\delta = 13,8\%$ se s průměrnou hodnotou získanou druhým testem $\bar{u} = 68,4\%$ na intervalu $2\delta = 12,2\%$ shoduje o něco méně než v prvním případě, se stále však jedná o srovnatelné hodnoty.

Nejlepší hodnocení získala v obou testech skrytá reference a nejhorší skrytá kotva. MPEG Layer II na 96 kb/s dopadl jen o málo lépe než signál omezený dolní propustí o nominálním kmitočtu 3,5 kHz, jelikož při nedostatečné bitové rychlosti jeho psychoakustický model poměrně agresivně kvantizuje vyšší kmitočty a chová se tak jako dolní propust.

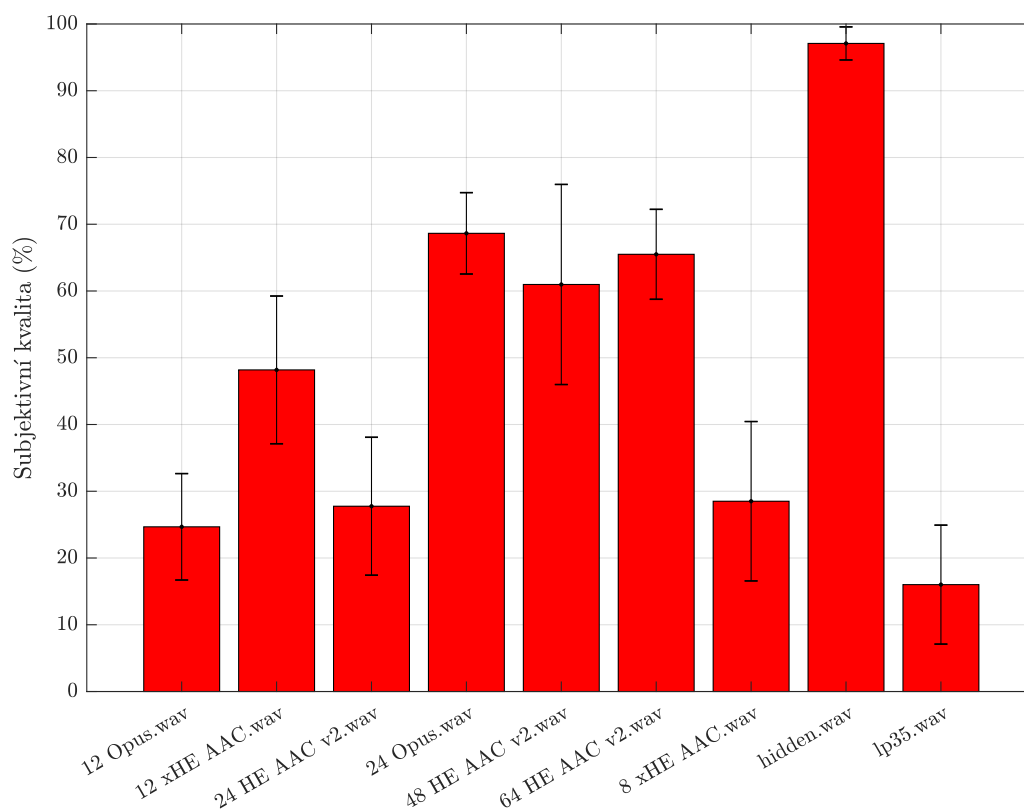
Mezigenerační souboje členů rodiny AAC naznačují, že použití metody *Spectral Band Replication* na rychlosti 64 kb/s nepřináší významné vylepšení subjektivně vnímané kvality. Její růst o pár procentních bodů na hodnotící škále lze ovšem sledovat na rychlosti 48 kb/s při použití parametrického stera. Při poklesu bitové rychlosti na 32/kb už ovšem kvalita padá do oblasti označené jako „špatná“. Proti tomu kodek Opus na stejných rychlostech dosahuje výsledků o zhruba jednu až dvě kvalitativní úrovně výše.

S výjimkou kodeku Opus, u kterého na vyšších rychlostech dopadají nejlépe smíšené vzorky, hodnotí posluchači hudební vzorky hůře než nahrávky z kategorií mluvené slovo a smíšené vzorky. Řeč je ovšem z frekvenčního hlediska méně komplexní než hudba a tak se u ní projevují kompresní artefakty ve srovnání se zakódovanou hudbou méně, i při použití

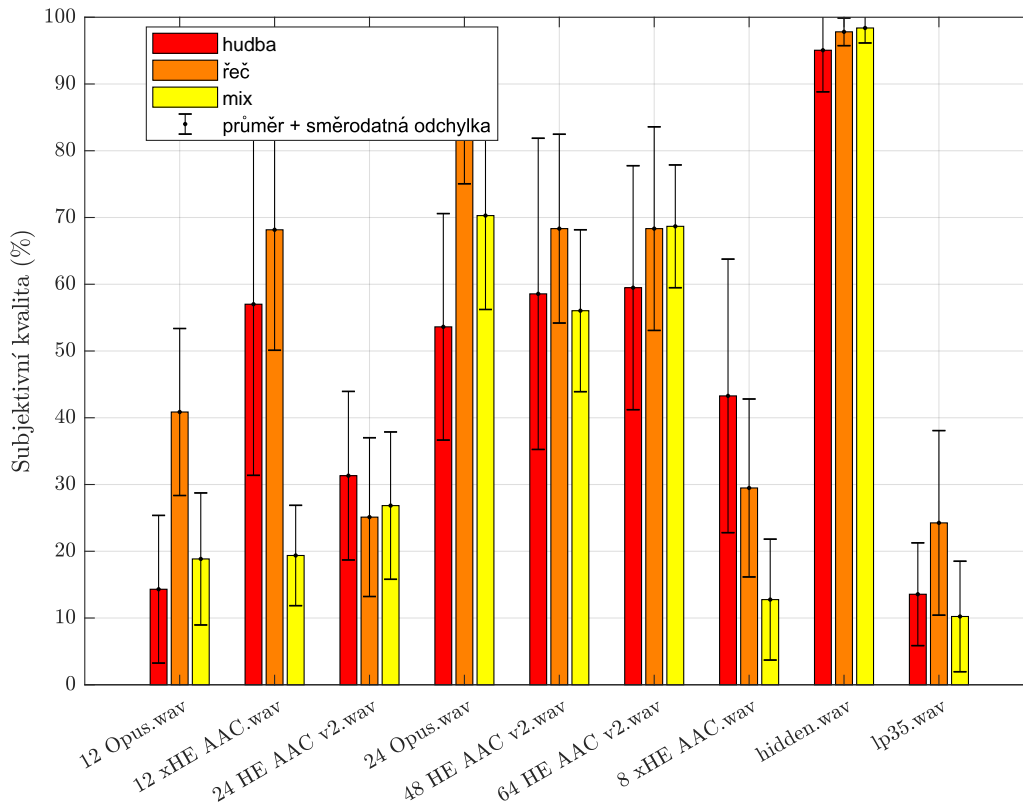
6. VÝSLEDKY TESTŮ

	12 Opus	12 xHE-AAC	24 HE-AAC v2	24 Opus	48 HE-AAC v2	64 HE-AAC v2	8 xHE-AAC	hidden	lp35
\bar{u}_k	24,7	48,2	27,8	68,6	61,0	65,5	28,5	97,1	16,0
δ_k	8,0	11,1	10,3	6,1	15,0	6,7	11,9	2,5	8,9
$\bar{u}_{jk(music)}$	14,3	57,0	31,3	53,6	58,6	59,5	43,3	95,1	13,6
$\delta_{jk(music)}$	11,1	25,6	12,6	17,0	23,3	18,3	20,5	6,2	7,7
$\bar{u}_{k(speech)}$	40,9	68,2	25,1	82,0	68,3	68,3	29,5	97,8	24,3
$\delta_{k(speech)}$	12,5	18,1	11,9	7,0	14,1	15,3	13,3	2,1	13,8
$\bar{u}_{k(mixed)}$	18,9	19,4	26,8	70,3	56,0	68,7	12,8	98,4	10,2
$\delta_{k(mixed)}$	9,9	7,5	11,0	14,1	12,1	9,2	9,1	2,2	8,3

Tabulka 6.2: Výsledky druhého testu



Obrázek 6.3: Výsledky testu MUSHRA pro nižší bitové rychlosti



Obrázek 6.4: Výsledky testu MUSHRA pro nižší bitové rychlosti včetně kategorií

kodeků které nejsou přímo určené ke kompresi mluveného slova. Při započítání překryvů intervalů spolehlivosti dosahují rozdíly v hodnocení jednotlivých kategorií maximálně deseti procent. Celkově se dá pozorovat, že s klesající bitovou rychlostí roste rozptyl získaných hodnot. V testu na nižších bitových rychlostech dosahuje velikost intervalu spolehlivosti hodnoty až $2\delta = 30\%$.

Co druhý test nepotvrdil je výhoda sjednoceného kódování řeči a hudby profilem xHE-AAC. U obou testovaných bitových rychlostí dopadl smíšený vzorek nejhůře ze všech třech kategorií a to s více než desetiprocentní ztrátou oproti průměrné hodnotě. U kodeku Opus dopadají nejlépe téměř výhradně vzorky řečové. Celkově ale xHE-AAC dosahuje vyšší kvality při nízkých rychlostech ve srovnání s ostatními použitými kodeky.

Pokud by tento test měl sloužit jako výběr ideálního kandidáta pro přenos signálu digitálním rádiem, nebyl by zřejmě příliš úspěšný, neboť jen minimum vzorků dosahuje úrovně „Vynikající“ či „Dobrá“. Spíše se výsledky pohybují od hladiny 80 % níže, až po úroveň označovanou jako nedostatečná. Toto „rozprostření“ výsledků přes všechny možné úrovně se však hodí pro porovnávání výsledků získaných objektivními algoritmy. Zvýší se tím výpovědní hodnota grafů v následující podkapitole.

6.2 Porovnání objektivních a subjektivních výsledků

Při určování věrohodnosti objektivních metod jsou brány výsledky subjektivních testů jako referenční ukazatel kvality. Při výpočtech a zobrazení je počítáno s aritmetickým průměrem hodnot subjektivně vnímané kvality přes množinu všech subjektů. Všemi předem zmíněnými algoritmy byl za pomoci skriptů popsaných v kapitole 5 ohodnocen stejný set vzorků jako při poslechových testech. Grafy na obrázku 6.5 zobrazují závislost hodnocení kvality objektivními algoritmy na subjektivně vnímané kvalitě pro jednotlivé hodnotící metody. Po vzoru článku *Objective Assessment of Perceptual Audio Quality Using ViSQOLAudio* [60] byly výsledky všech testů přepočítány na stejnou míru, respektive zobrazeny na stejné škále, aby bylo porovnání přehledné a objektivní. U subjektivních testů byla lineární sto-dílková škála *CQS* jednoduchým vydělením přepočtena na *SDG*. PEAQ a PEMO-Q poskytují výsledky ve formě *ODG* a na tu byly přepočítány i výsledky algoritmu ViSQOL, přepočtem z jeho standardního výstupu *MOS-LQO*. Barevně a za pomoci znaků jsou v nich odlišené metody komprese. Těm je rovněž přidělena eliptická oblast, kde poloosa ve vodorovném směru odpovídá směrodatné odchylce výsledků subjektivních testů a svislá poloosa odpovídá směrodatné odchylce objektivních hodnocení zároveň uvedených v tabulce 6.5.

Jelikož ideální případ závislosti *ODG* na *SDG* je jejich rovnost (čerchovaná modrá přímka) lze velikost chyby spočítat jejich rozdílem. Z těchto rozdílů lze vypočítat střední kvadratickou chybu ε (*Root Mean Squared Error*) podle:

$$\varepsilon = \sqrt{\sum_{j=1}^J \frac{(ODG_j - SDG_j)^2}{J}} \quad (6.1)$$

V tabulce 6.3 je vypočítané ε pro všechny testované metody komprese pomocí všech objektivních metod a v tabulce 6.4 jsou průměrné hodnoty $\bar{\varepsilon}$ a průměru kodeků používaných v digitálním rádiu $\bar{\varepsilon}_{radio}$.

	128 MP2	24 opus	32 HE-AAC v2	32 opus	48 HE-AAC v1	48 HE-AAC v2	64 HE-AAC v1	64 LC-AAC	96 MP2	8 xHE AAC	12 xHE AAC	lp35	uncompressed
$\varepsilon_{\text{peaq-a.}}$	1,70	0,70	0,46	1,34	1,16	1,27	1,49	1,79	0,40	1,62	0,75	0,21	0,15
$\varepsilon_{\text{peaq-b.}}$	1,19	0,99	0,45	1,23	0,96	0,35	1,31	1,77	0,42	1,61	0,63	0,57	0,15
$\varepsilon_{\text{pemoq}}$	0,46	0,48	0,45	1,67	1,04	0,99	0,97	1,30	1,02	1,68	0,73	1,19	0,15
$\varepsilon_{\text{visqol}}$	0,66	1,30	1,03	0,58	0,48	0,35	0,92	1,02	1,64	1,21	0,98	0,99	0,15

Tabulka 6.3: Kvadratické střední chyby ε pro jednotlivé kompresní metody

	$\bar{\varepsilon}$	$\bar{\varepsilon}_{radio}$
PEAQ A.	1,01	1,27
PEAQ B.	0,89	1,04
PEMO-Q	0,87	1,01
ViSQOL	0,87	0,90

Tabulka 6.4: Průměrná hodnota ε a ε_{radio}

	128 MP2	24 opus	32 HE-AAC v2	32 opus	48 HE-AAC v1	48 HE-AAC v2	64 HE-AAC v1	64 LC-AAC	96 MP2	8 xHE AAC	12 xHE AAC	lp35	uncompressed
$\sigma_{\text{peaq-a.}}$	0,51	0,43	0,41	0,43	0,56	0,88	0,66	0,33	0,04	0,05	0,06	0,05	0,00
$\sigma_{\text{peaq-b.}}$	0,36	0,61	0,52	0,61	0,74	0,80	0,69	0,74	0,23	0,27	0,42	0,33	0,00
σ_{pemoq}	0,18	0,31	0,52	0,31	0,54	1,18	0,44	0,29	0,18	0,04	0,04	0,46	0,00
σ_{visqol}	0,11	0,28	0,63	0,28	0,15	0,72	0,28	0,19	0,07	0,46	0,41	0,51	0,00

Tabulka 6.5: Směrodatné odchylky objektivního hodnocení

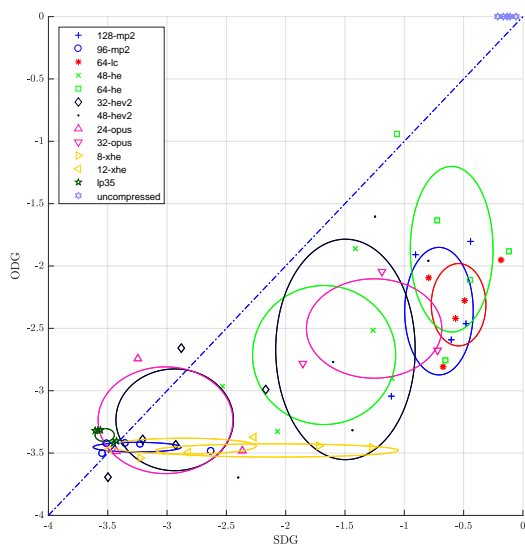
Zjednodušeně řečeno, tabulka 6.3 říká, jak moc se jednotlivé algoritmy „mýlí“ a tabulka 6.5 ukazuje, jak moc náhodné výsledky poskytují.

Bráno podle stáří, hodnocení vzorků zkomprimovaných kodekem MPEG Layer II dosahuje nízkých hodnot rozptylu přes různé druhy nahrávek. Na nižší použité rychlosti 96 kb/s hodnotí s nejmenší chybou oba zástupci algoritmu PEAQ. U PEMO-Q došlo k nadhodnocení o jeden stupeň *ODG* a u algoritmu ViSQOL je tato hodnota ještě vyšší. Jak už bylo řečeno, MPEG Layer II se při nízkých rychlostech chová jako dolní propust, která mění spektrogram (současně tedy i neurogram) jen nepatrně, díky čemuž je „políčková podobnost“ vyšší než u modernějších kompresních metod. PEAQ a na něm založené PEMO-Q zřejmě přikládá větší váhu gammatónovým filtrům modelu ucha na vysokých kmitočtech, a tak hodnotí *mp2* věrněji. Na druhou stranu při vyšší bitové rychlosti je kvalita MPEG Layer II podhodnocena právě algoritmem PEAQ. Nejmenší chybu poskytl PEMO-Q a druhou nejmenší ViSQOL.

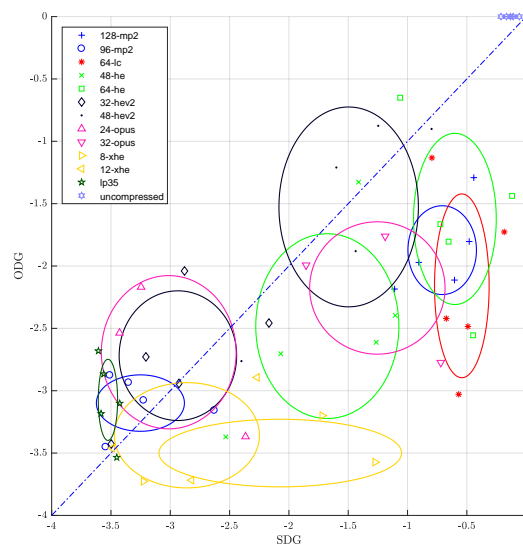
Základní profil AAC LC tvoří v grafech 6.6 zastoupen jen jednou oblastí o bitové rychlosti 64 kb/s, kterou obě verze algoritmu PEAQ podhodnotili o více než jeden stupeň na *ODG* škále. Základní verze má navíc ze všech algoritmů jednoznačně nejvyšší rozptyl. Nejkonzistentnější výsledek a zároveň nejmenší chybu poskytl ViSQOL.

Pokud se k základnímu profilu AAC přidá rozšíření SBR, projevuje se to poklesem kvadratické chyby a současným vzrůstem směrodatné odchylky u všech metod. Tento trend dál pokračuje i s poklesem rychlosti HE-AAC v1 na 48 kb/s, kromě hodnocení algoritmu

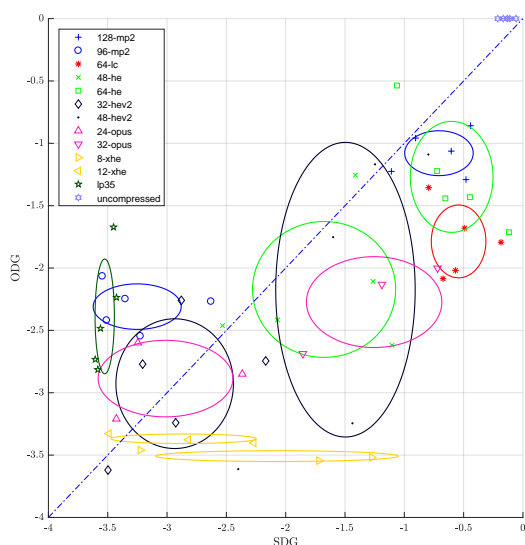
6. VÝSLEDKY TESTŮ



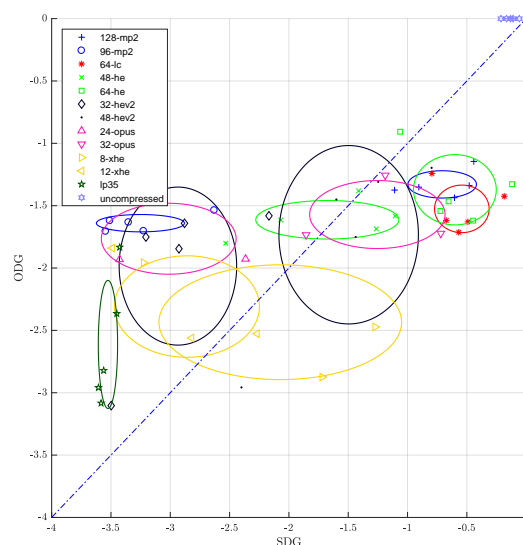
(a) PEAQ Advanced



(b) PEAQ Basic



(c) PEMO-Q



(d) ViSQOL

Obrázek 6.5: Závislost ODG na SDG včetně oblastí vytyčených směrodatnými odchylkami objektivních a subjektivních testů

ViSQOL u kterého rozptyl naopak klesá. Také chybu udává nejmenší a to $\varepsilon = 0,48$.

Dalším generačním krokem k HE-AAC v2 kvadratická chyba roste u algoritmů PEAQ a PEMO-Q, zatímco u hodnocení metodou ViSQOL chyba opět klesá. U tohoto profilu je nejvyšší rozptyl objektivně změřené kvality. Směrodatná odchylka výsledků algoritmu

PEAQ dosahuje téměř 1,2 stupně ODG na obě strany od průměru. PEAQ Advanced poskytuje hodnotu $\sigma = 0,88$ a ViSQOL $\sigma = 0,72$. Oblast s nižší rychlostí 32 kb/s HE-AAC v2 protíná ideální převodní křivku ve třech případech ze čtyř. Jen hodnocení algoritmem ViSQOL je vyšší než u subjektivních testů a to o jeden stupeň ODG.

Kodek Opus v digitálních rádiích nefiguruje a proto jeho hodnocení nemá vliv na jakýkoliv výsledný verdikt. Na nižší rychlosti 24 kb/s ho s nejmenší chybou hodnotil algoritmus PEMO-Q a při vyšším toku 32 kb/s ViSQOL, který dosáhl taktéž nejmenší směrodatné odchylky u obou rychlostí.

Téměř zázračně nízké hodnoty směrodatné odchylky u kodeku xHE-AAC jsou způsobeny spíše tím, že PEAQ ani PEMO-Q si s jeho hodnocením neví příliš rady a tak všem testovaným vzorkům přiřadili hodnotu $ODG \approx 3,5$. Jediný ViSQOL dosahuje hodnocení, které se přibližuje výsledkům subjektivních testů.

Trochu mimo dosavadní výběr stojí dvě poslední oblasti a to skrytá kotva a skrytá reference. Jsou ovšem součástí výstupu subjektivního testování a není důvod nezařadit je do těchto závislostí po bok kompresních metod. Referenční vzorek všechny algoritmy ohodnotily bez jakékoliv ztráty kvality a tak se eliptická oblast zjednodušila na úsečku v pravém horním rohu grafů 6.6. Signál s potlačenými kmitočty od 3,5 kHz výše stejně jako posluchači ohodnotily obě verze algoritmu PEAQ. ViSQOL i PEMO-Q skrytou kotvu nadhodnocují zhruba o jeden stupeň ODG.

Ani jedna z metod neposkytuje ideální výsledky. Lze ale vysledovat trend, že algoritmy PEAQ a PEMO-Q poskytují horší výsledky u modernějších kodeků používaných v digitálních rádiích zatímco ViSQOL má problémy spíše s nadhodnocováním vzorků, které postrádají vyšší harmonické. ViSQOL také v průměru hodnotí s nejmenší kvadratickou chybou, což je ještě výraznější u průměrné hodnoty ε_{radio} , ve které jsou zahrnuty kodeky používané v DRM(+) a DAB/DAB+. Na základě těchto poznatků je označen za nejobjektivnější hodnotící metodu ze všech prezentovaných.

6.3 Výsledky objektivních testů

V zadání této práce je mimo jiné zmíněno, že vybraná objektivní metoda by měla být využita pro vyhodnocení kvality zvuku v systémech digitálního rozhlasového vysílání. Z algoritmizačního hlediska bylo jednodušší vyhodnotit kvalitu všemi metodami, pro všechny kodeky a všechny možné bitové rychlosti a nainportovat je pomocí skriptu `makeCell` do zobrazovacího nástroje `showResults`. Protože v předchozí podkapitole je jako nejvěrohodnější objektivní metoda vybrán ViSQOL, potažmo jeho varianta ViSQOLAudio jsou v tomto textu tedy vybrány pouze závislosti vygenerované algoritmem ViSQOL.

Výsledky z ostatních algoritmů jsou k nahlédnutí v dodatku C, či v samotném nástroji `showResults`, kde si může uživatel výstup libovolně nakonfigurovat a exportovat do souboru.

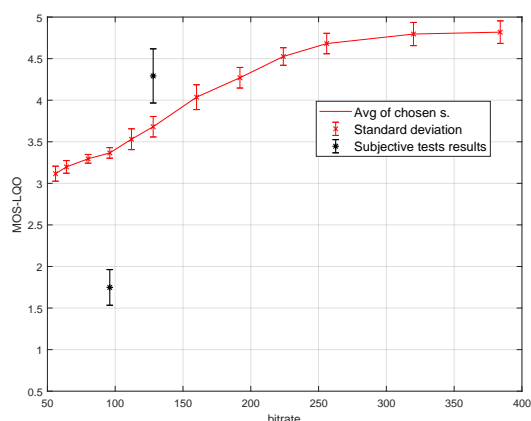
Obrázek 6.6 obsahuje šestici grafů zobrazujících průměrné ODG a směrodatnou odchylku všech vzorků z tabulky 5.1. Do nich jsou rovněž vyneseny výsledky poslechových testů pro

6. VÝSLEDKY TESTŮ

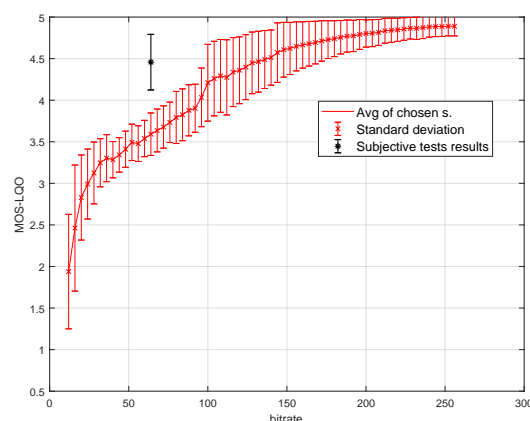
lepší představu toho jak spolu korespondují. Pro DRM+ a DAB+ jsou klíčové grafy 6.6a a 6.6c až 6.6f.

Pro hodnocení *MPEG Layer II* není ViSQOL příliš vhodný avšak to je jediný případ, kdy je radno sáhnout po jiné objektivní metodě. Na velmi nízkých bitových rychlostech stále uděluje hodnocení MOS-LQO vyšší než 3, slovně řečeno lepší než přijatelné. Tento kodek je ovšem v současnosti nahrazován pokročilejšími kompresními metodami a tam si ViSQOL vede podstatně lépe.

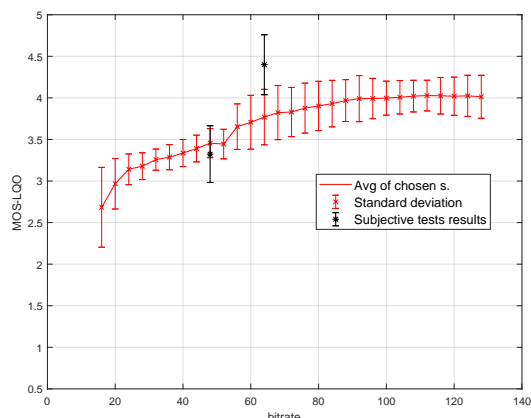
Podobně jako první subjektivní test ukázal, že mezi generacemi HE-AAC v1 a HE-AAC v2 lidské ucho téměř neslyší rozdíl, hodnotí v průměru ViSQOL tyto dva profily velmi podobně. U HE-AAC v2 nicméně podává výsledky se zhruba o třetinu větším rozptylem. Užití rozšíření PS a SBR má své výhody, ale i rizika. S rostoucím bitovým tokem se kódový zisk nabytý použitím metod založených na psychoakustických poznatcích klesá a v určitém momentu již nastoupí výhody nižšího profilu.



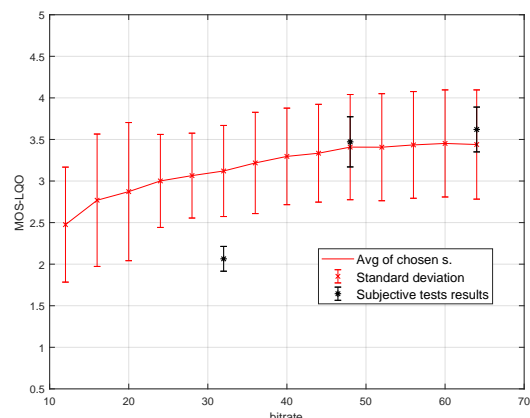
(a) MP2



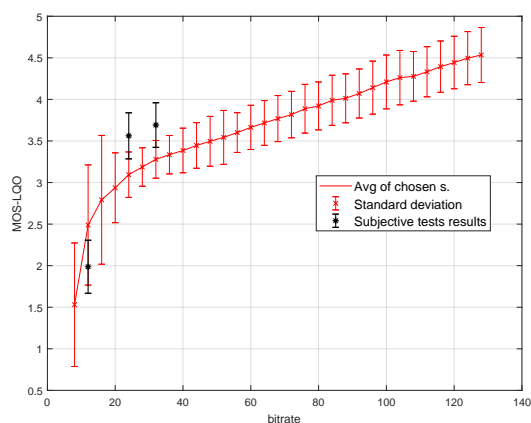
(b) LC-AAC



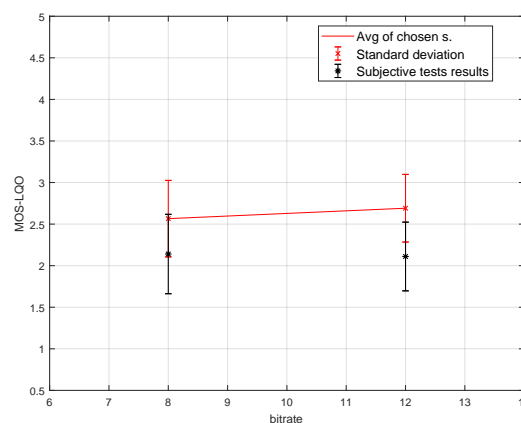
(c) HE-AAC v1



(d) HE-AAC v2



(e) Opus



(f) xHE-AAC

Obrázek 6.6: Hodnocení algoritmem ViSQOL.

I přes to, že k rozhodnutí zvolit právě ViSQOL nevedlo hodnocení vzorků zakódovaných kodekem Opus, je podobnost objektivních a subjektivních výsledků značná.

Ze závislostí na obrázku 6.6 lze říci, že pokud má být kvalita zvuku přenášeného digitálním rádiem hodnocena posluchači jako vynikající nezbyvá než volit z kodeků MPEG Layer II a AAC LC. U prvního ze zmíněných je třeba bitového toku 160 kb/s či více, druhému k překročení spodní hranice vynikající kvality stačí 128 kb/s. Žádná jiná z kompresních metod tohoto kvalitativního prahu nedosahuje.

Pokud je poskytovatel ochotný slevit ze svých požadavků a vysílat kvalitou označovanou jako dobrá (60 % - 80 % na CQS stupnici), dosáhne nejvyšší efektivity volbou profilu HE-AAC v2 na bitové rychlosti 44 kb/s. Kdyby z nějakých důvodů nebylo možné využít rozšíření *Parametric Stereo*, dosáhne stejného výsledku s bitovým tokem 56 kb/s profilů HE-AAC v1 a AAC LC, které na této rychlosti ViSQOL hodnotí stejně. Při použití MPEG Layer II by bylo potřeba dvojnásobného toku 112 kb/s.

Pokud je prioritou úspora prostředků a padne rozhodnutí vysílat pouze přijatelnou kvalitou, lze dle hodnocení algoritmu ViSQOL využít kompresní metodu xHE-AAC. Pro přenos mluveného slova je dostatečnou bitovou rychlostí 8 kb/s. Při přenosu hudebních záznamů je lepší sáhnout po rychlosti 12 kb/s. Jaké kvality xHE-AAC dosahuje na vyšších tocích bohužel kvůli absenci kodéru nebylo zjištěno. V digitálním rádiu DAB, které xHE-AAC neintegruje lze stejné kvality pro vysílání řeči dosáhnout užitím kodeku HE-AAC v2 s tokem 16 kb/s a pro přenos hudebního obsahu nejméně 20 kb/s.

Závěr

Nejjednodušší řešení je obvykle to správné.

-jedna z interpretací filozofického přístupu známého jako Occamova Břitva

Tato práce měla několik dílčích cílů. Zpočátku bylo třeba zmapovat metody pro hodnocení zvukového signálu. Dále bylo třeba realizovat a vyhodnotit poslechový test na množině vzorků reprezentující obsah a kompresní metody používané v systémech *Digital Radio Mondiale* a *Digital Audio Broadcasting*. Stejnou množinu vzorků bylo potřeba oznámkovat dostupnými algoritmy pro objektivní hodnocení kvality zvuku a s ohledem na výsledek subjektivních testů vybrat metodu vhodnou pro objektivní hodnocení signálu používaného v digitálních rádiích. Posledním dílčím úkolem bylo za pomoci vybraného hodnotícího postupu posoudit kvalitu zvukových nahrávek v závislosti na bitovém toku a typu použité komprese.

Jelikož digitální rádio je středobodem této práce, jsou mu společně s popisem zdrojového kódování věnovány první dvě kapitoly. Popsány jsou možnosti a principy systémů DAB/-DAB+ a DRM(+). Dále je uveden stručný vhled do problematiky kompresních standardů MPEG Layer II, MPEG 4 AAC v profilech AAC LC, HE-AAC v1 a HE-AAC v2, MPEG D xHE-AAC a kodeku Opus, který v systémech digitálního rádia nefiguruje, nicméně sdílí s xHE-AAC sjednocené kódování zvuku a řeči a je tak zajímavým doplňkem.

Konzumentem rozhlasového obsahu byl, je a pravděpodobně vždy bude člověk. Lidský vjem je tak stále nejlepším způsobem jak hodnotit kvalitu zvuku. Proto je určování kvality pomocí poslechu lidskými subjekty věnována celá kapitola. Popsány jsou postupy definované mezinárodní telekomunikační unií od přístupů obecných, přes způsob určování nepatrných zhoršení až po metodu pro střední kvalitu s označením MUSHRA.

Ta je využita při subjektivním testování. Pro určení kvality vzorků, rozdělených do kategorií: hudební, řečové s smíšené, byly navrženy dva mírně se překrývající testy. Jeden sledující kvalitu na vyšších bitových rychlostech starších kodeků, a druhý obsahující materiál zakódovaný moderními kodeky, jenž cílí na velmi nízké bitové toky. Bylo pro ně vyvinuto grafické uživatelské rozhraní v programovacím jazyce MATLAB a celkově sedmáct subjektů absolvovalo oba testy.

Pro generování vzorků, jejich objektivní testování a zobrazování získaných výsledků byla vyvinuta skupina znovupoužitelných skriptů a aplikací, taktéž v prostředí MATLAB. Pomocí algoritmů PEAQ Basic, PEAQ Advanced, PEMO-Q a ViSQOL byly ohodnoceny vzorky zakódované všemi již zmíněnými kodeky.

Porovnání subjektivního a objektivního hodnocení odhalilo, že ne vždy lze použít Occamovu břitvu. Žádná z metod totiž není stoprocentně spolehlivá. Nejstarší PEAQ nejlépe hodnotí dnes už historický kodek MPEG Layer II a u novějších kompresních algoritmů klesá věrohodnost jeho hodnocení. PEMO-Q, které algoritmus PEAQ doplňuje o komplexnější model ucha sdílí podobný trend. Nejmodernější ViSQOL, původně vytvořený pro hodnocení řeči v telekomunikacích, si vede lépe u moderních metod komprese. Jako jediný se přibližuje výsledku subjektivního hodnocení vzorků zakódovaných metodou xHE-AAC. Nejstarší MPEG Layer II na nízkém bitovém toku ovšem značně nadhodnocuje. Protože jsou ale starší kodeky vytlačovány novými efektivnějšími kompresními přístupy, jeví se ViSQOL jako nejlepší volba do budoucna.

Zároveň je ale možné, že snížením váhy přikládané výstupům filtrů na vysokých kmitočtech v modelu sluchové cesty u algoritmů PEAQ či PEMO-Q by mohlo vést ke zvýšení věrohodnosti jimi podaných výsledků. Jelikož implementace algoritmů PEAQ Basic, PEMO-Q i ViSQOL byly k dispozici ve formě funkcí pro MATLAB, nabízí se do budoucna možnost jejich úprav či propojení, tak aby podávaly věrohodnější hodnocení pro kompletní škálu kompresních metod.

V této práci je za prozatím nejvěrohodnější metodou zvolen ViSQOL. Pomocí něho je ohodnocen kompletní set vzorků, podrobně popsáný v kapitole 5. Výsledky jsou s komentářem k dispozici v grafické podobě v poslední kapitole této práce, věnované kvalitě přenosu zvuku v digitálních rádiích. Vynikající poslechové kvality lze při zdrojovém kódování dosáhnout pouze použitím starších kodeků na vyšších bitových rychlostech. Vývoj moderních kodeků se soustřeďuje na zvyšování kódového zisku při zachování dobré, či alespoň přijatelné kvality. Té lze dosáhnout už při datovém toku 8 kb/s pro přenos mluveného slova nebo 12 kb/s pro vysílání hudebních záznamů. Protože jsou vysílací mechanismy digitálního rádia podstatně odolnější vůči analogovému rušení, existuje velmi vysoká pravděpodobnost, že se ke koncovému zákazníkovi dostane zvukový signál ve stejné kvalitě, do jaké byl kompresní metodou zakódován. Ta tak ve výsledku mnohdy předčí kvalitu analogového rádia.

Literatura

- [1] European Telecommunications Standards Institute. *ETSI EN 300 401: Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers V2.2.1*. ETSI, 2016, 124 str.
- [2] Modulace s více nosnými kmitočty. Dostupné na: <https://publi.cz/books/234/02.html>
- [3] European Telecommunications Standards Institute. *ETSI ES 201 980: Digital Radio Mondiale (DRM); System Specification ETSI STANDARD*. ETSI, 2017, 196 str.
- [4] Moritzlucaschmid. Audio decoding presentation slides. Aug 2017. Dostupné na: <https://dabtransceiver.wordpress.com/2017/06/30/week-07-audio-decoding/>
- [5] USAC (Unified speech and audio coding). Dostupné na: <http://www.voiceage.com/USAC.html>
- [6] International Telecommunication Union. *13818-7: Information technology — Generic coding of moving pictures and associated audio information — Part 7: Advanced Audio Coding (AAC)*. ISO/IEC, 1990.
- [7] Meltzer, S.; Moser, G. MPEG-4 HE-AAC v2 - audio coding for today's media world. 2005.
- [8] Neuendorf, M.; Gournay, P.; Multrus, M.; a spol. Unified speech and audio coding scheme for high quality at low bitrates. *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, svazek 0, 04 2009: str. 1–4, doi:10.1109/ICASSP.2009.4959505.
- [9] Opus Codec. Dostupné na: <https://opus-codec.org/comparison/>
- [10] International Telecommunication Union. *Determination of the electro-acoustical properties of studio monitor headphones*. ITU-R, 1990, 5 str.

- [11] Hines, A.; Skoglund, J.; Kokaram, A.; a spol. ViSQOL: The Virtual Speech Quality Objective Listener. *Acoustic Echo Noise Control (IWAENC), IEEE Intl. Workshop on*, 01 2012: str. 1–4.
- [12] Huber, R.; Kollmeier, B. PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Transactions on Audio, Speech, and Language Processing*, svazek 14, číslo 6, Nov 2006: str. 1902–1911, ISSN 1558-7916.
- [13] Zalabák, M. *Hodnocení kvality analogového záznamu zvuku s využitím modelu slyšení*. Diplomová práce, České vysoké učení technické v Praze, Praha, 2016.
- [14] European Telecommunications Standards Institute. *ETSI EN 103 466: Digital Audio Broadcasting (DAB); DAB audio coding (MPEG Layer II V 1.1.1)*. ETSI, 2016, 59 str.
- [15] International Telecommunication Union. *Rec. ITU-R BS.1116-1: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. ITU-R, 1994, 26 str.
- [16] Melka, A. *Základy experimentální psychoakustiky*. Praha: Akademie múzických umění v Praze, 2005, ISBN 80-7331-043-0.
- [17] International Telecommunication Union. *Rec. ITU-R BS.1284-2: General methods for the subjective assessment of sound quality*. ITU-R, 2019, 18 str.
- [18] HD 200 PRO. Dostupné na: <https://en-de.sennheiser.com/studio-headphones-noise-reducing-hd-200-pro>
- [19] ADA-HP USB - HQ audio zesilovač. Dostupné na: <https://www.axagon.eu/produkty/ada-hp?ami=181>
- [20] Historie rozhlasu v kostce. Dostupné na: https://www.rozhlas.cz/archiv/historie/_zprava/historie-rozhlasu-v-kostce--1392938
- [21] European Telecommunications Standards Institute. *ETSI EN 300 401: Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers V1.4.1*. ETSI, 2006, 197 str.
- [22] International Organization for Standardization. *Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s*. ISO/IEC, 1993.
- [23] International Telecommunication Union. *ISO/IEC DIS 23003-3: Information technology – MPEG audio technologies – Part 3: Unified speech and audio coding*. ISO/IEC, 2016.
- [24] Hydrogenaudio. Dostupné na: <https://hydrogenaud.io/index.php/topic,109942.0.html>

-
- [25] Valin, J.-M.; Vos, K.; Terriberry, T. Definition of the Opus Audio Codec. RFC 6716, Zář 2012, doi:10.17487/RFC6716. Dostupné na: <https://rfc-editor.org/rfc/rfc6716.txt>
- [26] International Telecommunication Union. *Rec. ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of coding systems*. ITU-R, 2003, 18 str.
- [27] International Telecommunication Union. *Rec. ITU-R BS.1534-3: Method for the subjective assessment of intermediate quality level of coding systems*. ITU-R, 2015, 36 str.
- [28] Řezanková, H.; Loster, T. *Základy statistiky*. Praha: Oeconomica, první vydání, 2013, ISBN 978-80-245-1957-9.
- [29] International Telecommunication Union. *Objective quality measurement of telephone-band (300 - 3400 Hz) speech codecs*. ITU-T, 1996.
- [30] International Telecommunication Union. *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. ITU-T, 2001.
- [31] International Telecommunication Union. *Perceptual objective listening quality assessment POLQA*. ITU-T, 2011.
- [32] International Telecommunication Union. *Single Sided Speech Quality Measure 3SQM*. ITU-T, 2004.
- [33] International Telecommunication Union. *Rec. ITU-R BS.1387-1: Method for objective measurements of perceived audio quality*. ITU-R, 2001, 100 str.
- [34] P. Kabal. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. McGill University, 2003, 96 str.
- [35] Opticom. OPTICOM voice quality PESQ PEXQ speech quality perceptual measurement. 2019. Dostupné na: <https://www.opticom.de/>
- [36] Dau, T.; Kollmeier, B.; Kohlrausch, A. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, svazek 102, číslo 5, 1997: str. 2892–2905.
- [37] Hines, A.; Gillen, E.; Kelly, D.; a spol. ViSQOLAudio: An objective audio quality metric for low bitrate codecs. *The Journal of the Acoustical Society of America*, svazek 137, číslo 6, 2015: str. EL449–EL455, doi:10.1121/1.4921674. Dostupné na: <https://doi.org/10.1121/1.4921674>
- [38] MATLAB and Simulink. Dostupné na: <https://www.mathworks.com/>

- [39] MPEG Audio Layer 2 Encoder. Dostupné na: <http://www.twolame.org/>
- [40] FFmpeg Project. Dostupné na: <https://ffmpeg.org/>
- [41] Nero AAC Codec - Technologies. Dostupné na: <https://web.archive.org/web/20160310025758/http://www.nero.com:80/enu/company/about-nero/nero-aac-codec.php>
- [42] DRM – xHE – AAC demo. Dostupné na: <https://www.drm.org/drm-xhe-aac-demo/>
- [43] Avidemux. Dostupné na: <http://avidemux.sourceforge.net/>
- [44] Audacity. Dostupné na: <https://www.audacityteam.org/>
- [45] Ulovec, K.; Smutny, M. Perceived Audio Quality Analysis in Digital Audio Broadcasting Plus System Based on PEAQ. *Radioengineering*, svazek 27, 04 2018: str. 342–352, doi:10.13164/re.2018.0342.
- [46] Počta, P.; Beerends, J. G. *Subjective and Objective Assessment of Perceived Audio Quality of Current Digital Audio Broadcasting Systems and Web-Casting Applications*. 3, Sep. 2015, 407-415 str.
- [47] Kropáček, R. Stanice DAB. Dostupné na: <http://www.digitalradiodab.cz/rozhlasove-stance.html>
- [48] Artificial Ear/Ear Simulator (IEC 60318-1 coupler). Dostupné na: <https://www.bksv.com/en/products/transducers/ear-simulators/ear-mouth-simulators/4153>
- [49] Vafaei, S. Sennheiser HD 280 Pro Review. May 2018. Dostupné na: <https://www.rtings.com/headphones/reviews/sennheiser/hd-280-pro>
- [50] Hoene, C.; Hyder, M. Considering Bluetooth's Subband Codec (SBC) for Wideband Speech and Audio on the Internet. 05 2019. Dostupné na: <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/49345>
- [51] Fukui, M.; Shimauchi, S.; Hioka, Y.; a spol. Acoustic echo reduction robust against echo-path change with instant echo-power-level adjustment. V *21st European Signal Processing Conference (EUSIPCO 2013)*, Sep. 2013, ISSN 2076-1465, str. 1–5. Dostupné na: <https://ieeexplore.ieee.org/document/6811523>
- [52] Lindau, A. *WhisPER. A MATLAB toolbox for performing quantitative and qualitative listening tests*. 01 2014, 10 str.
- [53] De Man, B.; Reiss, J. APE: Audio Perceptual Evaluation toolbox for MATLAB. *ResearchGate*, 04 2014: str. 18.

-
- [54] Giner, A. V. Scale: A Software Tool for Listening Experiments. *ResearchGate*, 2013: str. 4.
- [55] Create or edit UI file in GUIDE. Dostupné na: <https://www.mathworks.com/help/matlab/ref/guide.html>
- [56] Telecommunications & Signal Processing Laboratory. Dostupné na: <http://www-mmsp.ece.mcgill.ca/Documents/Software/index.html>
- [57] PEMO-Q: HörTech gGmbH. Dostupné na: <https://www.hoertech.de/de/f-e-produkte/pemo-q.html>
- [58] Novák, J. *Implementace metody objektivního hodnocení kvality zvuku*. Bakalářská práce, České vysoké učení technické v Praze, Praha, 2016.
- [59] qxLab. VISQOLAudio. 2019. Dostupné na: <https://qxlab.ucd.ie/index.php/audio-and-music/>
- [60] Sloan, C.; Harte, N.; Kelly, D.; a spol. Objective Assessment of Perceptual Audio Quality Using ViSQOLAudio. *IEEE Transactions on Broadcasting*, svazek 63, číslo 4, Dec 2017: str. 693–705, ISSN 0018-9316, doi:10.1109/TBC.2017.2704421.

Digitální přílohy

Na přiloženém CD jsou následující přílohy v digitální podobě:

- Skripty pro MATLAB:
 - main
 - makeAac
 - makeMp2
 - makeOpus
 - makeWav
 - makeCell
 - mushra
 - showResults
- Vzorky ve zkomprimované a PCM formě
- Text práce v PDF

Přílohy k subjektivnímu testu

B.1 Informovaný souhlas

Informovaný souhlas subjektu

Název projektu: Hodnocení kvality zvuku v systémech digitálního rozhlasového vysílání

Jméno subjektu:

Datum narození:

Vedoucí pokusu: Adam Bartyzal

1. Já, níže podepsaný/podepsaná souhlasím s mou účastí na experimentech. Je mi více než 18 let.
2. Byl(a) jsem podrobně informován(a) o cíli experimentu, o jeho postupech, a o tom, co se ode mě očekává. Experimentátor pověřený prováděním experimentu mě informoval o tom, že s účastí na experimentu nejsou spojena žádná rizika a nemůže dojít k poškození mého zdraví.
3. Porozuměl(a) jsem tomu, že svou účast na experimentu mohu kdykoliv přerušit, nebo odstoupit. Pokud jsem student UK nebo ČVUT, nebude toto odstoupení mít žádný vliv na mé další studium.
4. Při zařazení do experimentu budou moje osobní data uchována s plnou ochranou důvěrnosti dle platných zákonů ČR. Rovněž pro výzkumné a vědecké účely a publikace mohou být moje osobní údaje a výsledky experimentů poskytnuty/publikovány pouze jako anonymizovaná data.
5. Porozuměl(a) jsem tomu, že mé jméno se nebude nikdy vyskytovat v publikacích a zprávách o experimentech. Já pak naopak nebudu proti použití výsledků těchto experimentů.

V Praze dne:

Vlastnoruční podpis dobrovolníka:

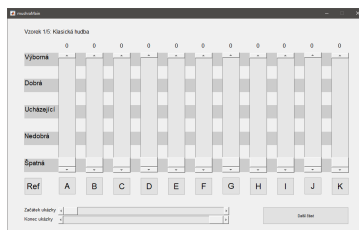
B.2 Doprovodné informace

Průvodní Informace k poslechovému testu

Po dobu následujících přibližně dvaceti pěti minutách se stanete subjektem poslechového testu MUSHRA neboli testu s vícenásobným stimulem, skrytou referencí a kotvou definovaným doporučením ITU-R BS.1534-1.

Na počátku testu (po spuštění skriptu *mushraStart.m*) budete vyzváni k zadání jména a věku. Jak se píše v informovaném souhlasu, tato data budou využita pouze k vyhodnocení experimentu a nebudou nikde publikována.

Tlačítkem *Spustit test* se přejde k první fázi testu, k takzvanému zaučení. To slouží k minimalizaci chyb při provádění ostré fáze testu. Grafické uživatelské rozhraní obsahuje několik důležitých prvků naznačených na následujícím obrázku. Ve fázi zaučení navíc obsahuje vysvětlivky.



Tlačítka označenými velkými tiskacími písmeny se spouští vzorky určené k posouzení. Tábla nad jednotlivými tlačítky slouží k udělení kvalitativního hodnocení na stupnici od nuly do sta příslušným ukázkám. Tato stupnice je podle doporučení ITU-R rozdělena na pět stejně velkých oblastí odpovídajícím subjektivnímu poslechovému vjemu. V každé kapitole testu ukázky vychází z deseti vteřinového referenčního vzorku, který lze spustit tlačítkem *Ref*. Každý je zakódován jinou metodou, tudíž by měl vykazovat jinou poslechovou kvalitu. Navíc jsou mezi ně přimíchány dva umělé vzorky sloužící k otestování věrohodnosti daného subjektu. Skrytá reference, tj. soubor stejný jako známá reference, které byste měli udělit maximální hodnocení. To mimo jiné znamená, že minimálně jedno táhlo by po dokončení hodnocení mělo mít hodnotu 100. Druhý umělý vzorek je tzv. skrytá kotva. Ide o referenční vzorek, na který byla aplikována dolní propust s nominálním kmitočtem 3,5 kHz. Měli byste být schopni takovýto vzorek odhalit a udělit mu horší než uchazející hodnocení.

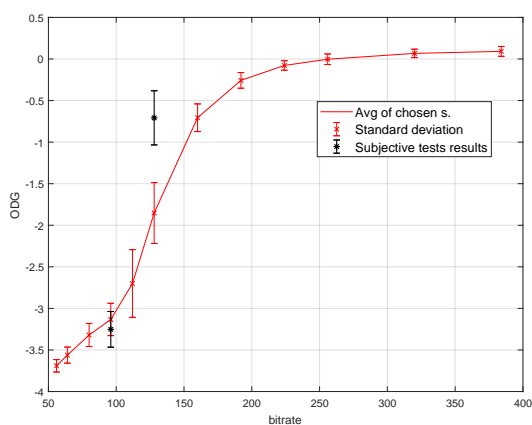
Táhla ve spodní části testu slouží k časovému omezení přehrávaných vzorků. Horní k nastavení začátku, spodní k nastavení konce. Budete-li se chtít rozhodnout kterému ze dvou vám právě posuzovaných vzorků udělíte vyšší kvalitativní ohodnocení, lze těmito táhly vybrat určitou část vzorku a poslouchat pouze ji. Mějte ovšem na paměti, že minimálně jednou byste si měli poslechnout vzorky celé.

Ve fázi zaučení je pořadí vzorků pevně dáno a doprovázeno vysvětlivkami. V ostrém testu budou (vyjma známé reference) ukázky náhodně zamíchány.

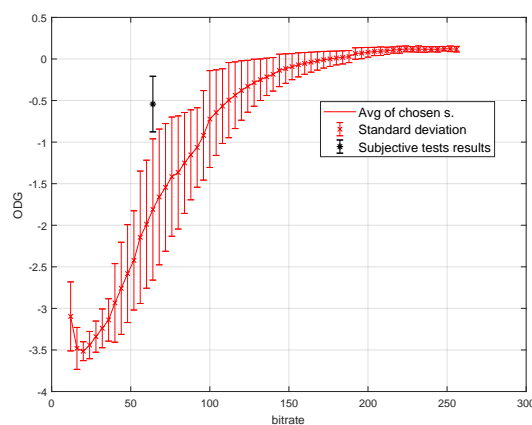
Po udělení hodnocení, tlačítkem *Další část* přejdete k následující kapitole testu. Po dokončení se objeví dialogové okno s nápisem *Konec*.

Podrobné výsledky objektivního hodnocení

Zde jsou uvedeny výsledky objektivního hodnocení ostatními algoritmy.

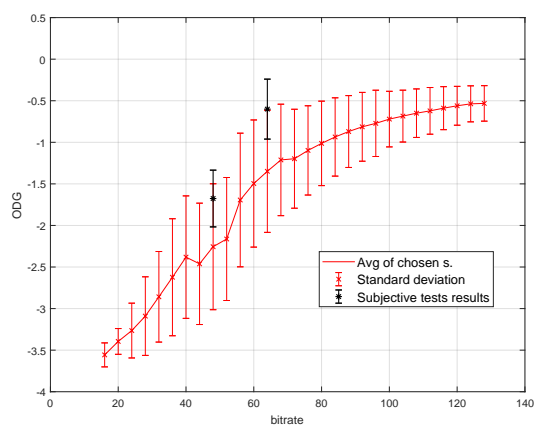


(a) MP2

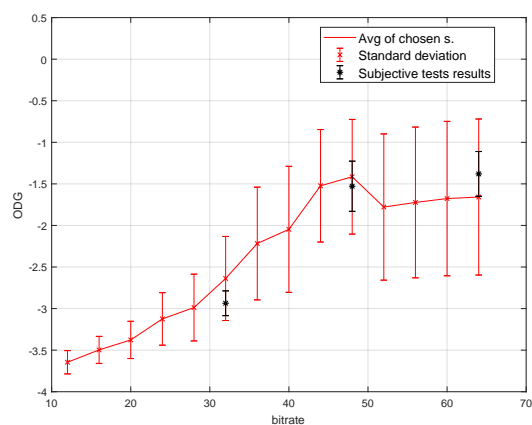


(b) LC-AAC

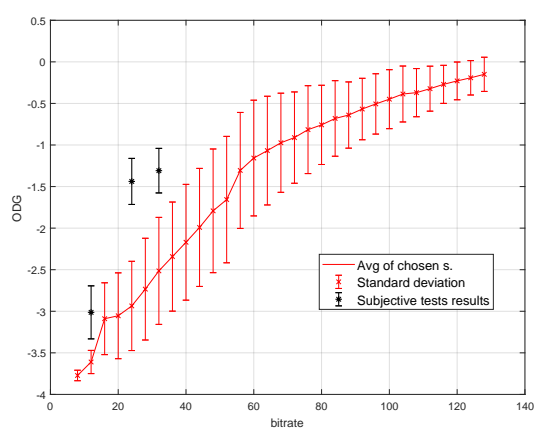
C. PODROBNÉ VÝSLEDKY OBJEKTIVNÍHO HODNOCENÍ



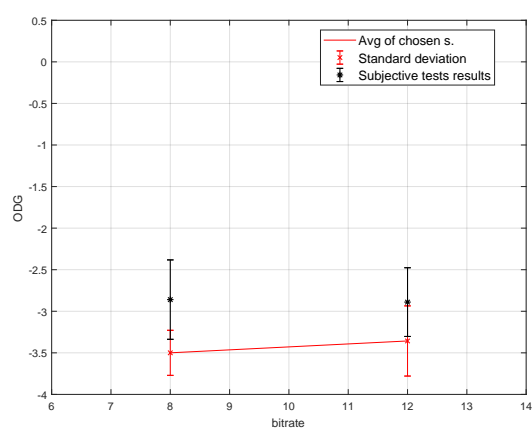
(c) HE-AAC v1



(d) HE-AAC v2

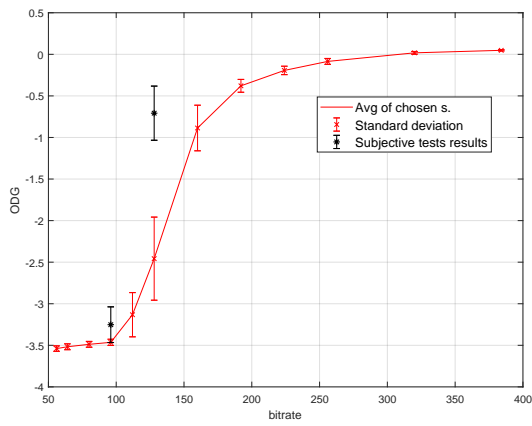


(e) Opus

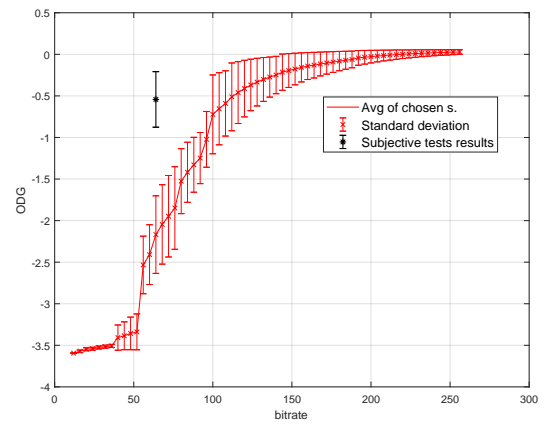


(f) xHE-AAC

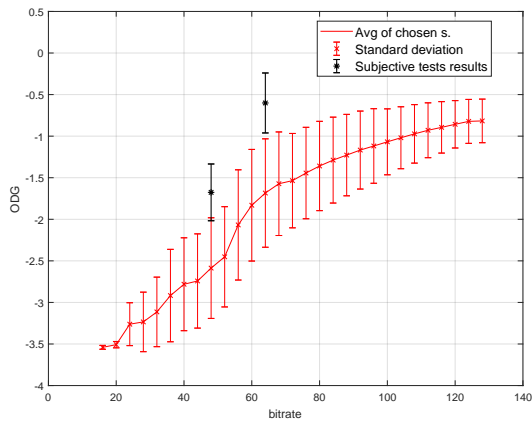
Obrázek C.1: Hodnocení algoritmem PEAQ Basic.



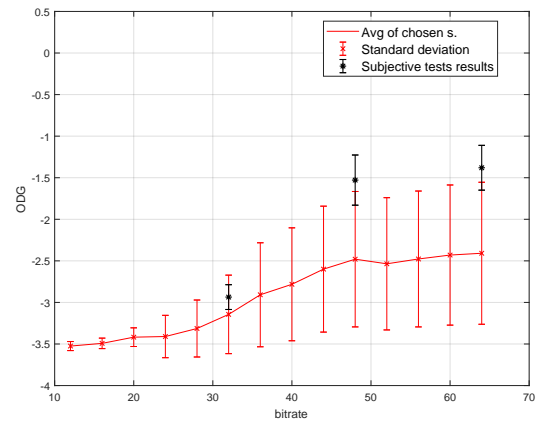
(a) MP2



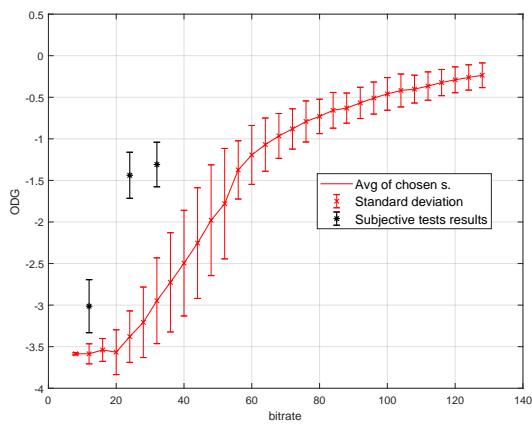
(b) LC-AAC



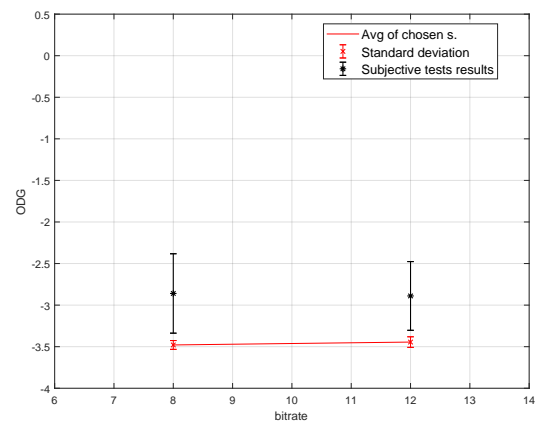
(c) HE-AAC v1



(d) HE-AAC v2



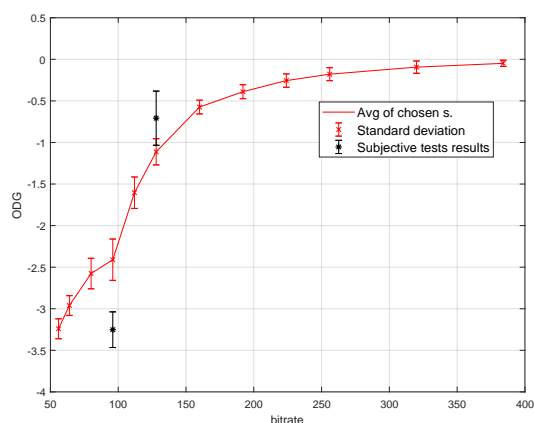
(e) Opus



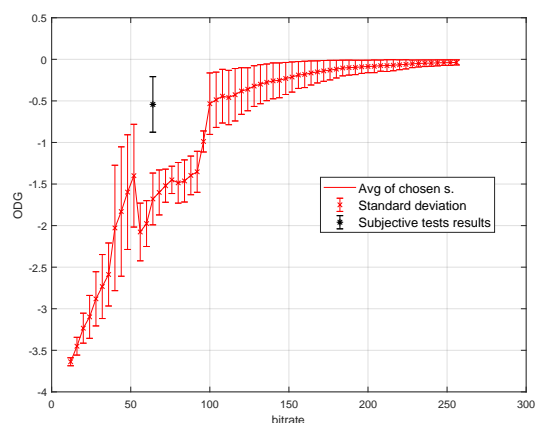
(f) xHE-AAC

Obrázek C.2: Hodnocení algoritmem PEAQ Advanced

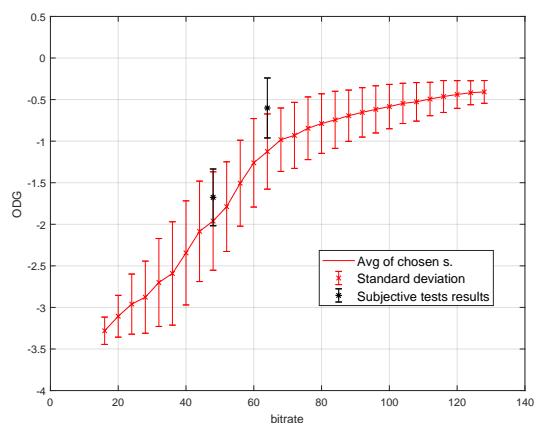
C. PODROBNÉ VÝSLEDKY OBJEKTIVNÍHO HODNOCENÍ



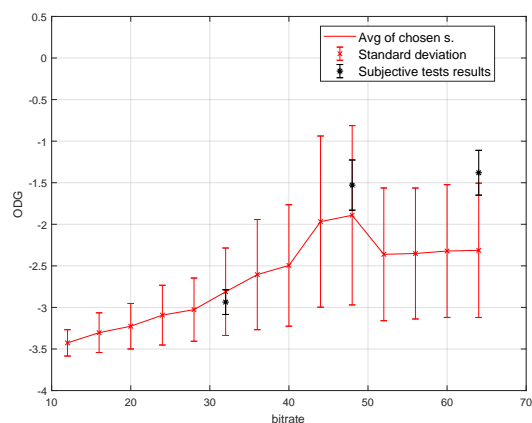
(a) MP2



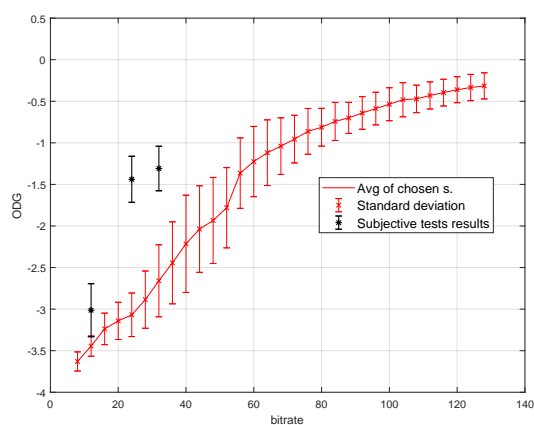
(b) LC-AAC



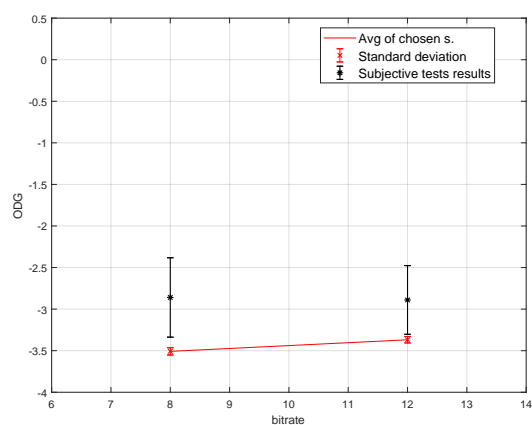
(c) HE-AAC v1



(d) HE-AAC v2



(e) Opus



(f) xHE-AAC

Obrázek C.3: Hodnocení algoritmem PEMO-Q.