

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

Fakulta elektrotechniky
a komunikačních technologií

BAKALÁŘSKÁ PRÁCE



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV TELEKOMUNIKACÍ

DEPARTMENT OF TELECOMMUNICATIONS

SROVNÁNÍ OBJEKTIVNÍCH METRIK PRO HODNOCENÍ KVALITY ZVUKOVÝCH SIGNÁLŮ

COMPARISON OF OBJECTIVE METRICS FOR EVALUATING THE QUALITY OF AUDIO SIGNALS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Adam Landsmann

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Pavel Závíška

BRNO 2021

Bakalářská práce

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

Student: Adam Landsmann

ID: 203737

Ročník: 3

Akademický rok: 2020/21

NÁZEV TÉMATU:

Srovnání objektivních metrik pro hodnocení kvality zvukových signálů

POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s objektivními ukazateli pro hodnocení kvality zvukových signálů a jejich dostupnými implementacemi a s problematikou vyhodnocení kvality rekonstrukce, aby co nejvíce odpovídala lidskému vnímání kvality zvuku. Jednotlivé metody podrobně nastudujte a popište. Dále připravte vhodnou testovací zvukovou databázi a na sérii poslechových testů pro několik typů poškození (clipping, chybějící vzorky, zašumění, kvantizace, apod.) vyhodnoťte míru přesnosti jednotlivých objektivních ukazatelů.

DOPORUČENÁ LITERATURA:

[1] Watanabe, K. Objective perceptual audio quality measurement methods. Broadcast Technology no. 35, NHK STRL.

[2] Campbell, D., Jones, E., Glavin, M. Audio quality assessment techniques—A review, and recent developments. Signal processing, vol. 89, issue 8, 2009, pp. 1489-1500. ISSN: 0165-1684.

Termín zadání: 1.2.2021

Termín odevzdání: 31.5.2021

Vedoucí práce: Ing. Pavel Závíška

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

UPOZORNĚNÍ:

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

ABSTRAKT

Tato bakalářská práce pojednává o problematice objektivního testování kvality zvuku se srovnáním se subjektivním typem hodnocení. V práci jsou popsány jevy v interpretaci zvuku skrz lidský sluchový orgán, možné druhy poškození audio signálu a vybrané druhy, jak objektivních, tak subjektivních metod pro hodnocení kvality zvuku. Od každého typu testování je zahrnuto šest metrik, z nichž pro srovnání jsou vybrány tři. Subjektivní poslechové testy byly provedeny pomocí webového rozhraní v souladu s vybranou metodou. Po nasbírání dostatku dat, byly výsledky statisticky analyzovány pomocí analýzy rozptylu. Konečné srovnání výsledků poslechových testů s hodnotami objektivních metrik bylo provedeno regresní a korelační analýzou.

KLÍČOVÁ SLOVA

Kvalita zvuku, objektivní metody hodnocení, poslechové testy, poškození zvuku, subjektivní metody

ABSTRACT

This bachelor thesis deals with the issue of objective testing of sound quality with a comparison with the subjective type of evaluation. The work describes the phenomena in the interpretation of sound through the human auditory organ, possible types of damage to the audio signal and selected types of both objective and subjective methods for evaluating sound quality. Six metrics are included from each type of testing, three of which are selected for comparison. Subjective listening tests were demonstrated using a web interface in accordance with the selected method. After sufficient data were collected, the results were statistically analyzed using analysis of variance. The final comparison of the results of listening tests with the values of objective metrics was proved by regression and correlation analysis.

KEYWORDS

Překlad klíčových slov (Sound quality, objective evaluation methods, listening tests, sound damage, subjective methods)

LANDSMANN, Adam. *Srovnání objektivních metrik pro hodnocení kvality zvukových signálů*. Brno, 2021, 59 s. Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedoucí práce: Ing. Pavel Závíška

PROHLÁŠENÍ

Prohlašuji, že svou bakalářskou práci na téma „Srovnání objektivních metrik pro hodnocení kvality zvukových signálů“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autora

PODĚKOVÁNÍ

Rád bych poděkoval vedoucímu bakalářské práce panu Ing.Pavlu Záviškovi, za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	9
1 Lidské vnímání zvuku	10
1.1 Zvuk	10
1.2 Psychologická akustika	10
1.2.1 Rozsah sluchu	10
1.2.2 Hlasitost zvuku	11
1.2.3 Výška tónu	12
1.2.4 Doba trvání zvuku	12
1.2.5 Maskování	12
1.2.6 Psychoakustický model lidského sluchu	13
1.2.7 Kognitivní model lidského sluchu	13
2 Poškození audio signálů	14
2.1 Clipping	14
2.1.1 Hard clipping	15
2.1.2 Soft clipping	15
2.2 Zašumění	16
2.2.1 Kvantizační šum	16
2.2.2 Harmonické rušení	17
2.3 Chybějící vzorky	17
3 Způsoby hodnocení audio signálů	19
3.1 Subjektivní metody pro hodnocení	19
3.1.1 ABX	19
3.1.2 MOS	19
3.1.3 MUSHRA	20
3.1.4 ACR	21
3.1.5 DCR	22
3.1.6 CCR	22
3.2 Objektivní metody pro hodnocení	22
3.2.1 SNR	23
3.2.2 PEAQ	23
3.2.3 PESQ	25
3.2.4 PEMO-Q	27
3.2.5 ViSQOL	29
3.2.6 Rnonlin	30

4	Příprava poslechových testů	31
4.1	Zvolená metoda subjektivního testování	31
4.2	Zvuková databáze	31
4.3	Poškození vzorků databáze	32
4.4	Rekonstrukce poškozených vzorků	34
4.4.1	SPADE	34
4.4.2	l1 Relaxation	34
4.4.3	Social Sparsity	34
4.5	Objektivní hodnocení kvality vzorků	35
5	Provedení poslechových testů	36
5.1	Server	36
5.1.1	Operační systém	36
5.1.2	Web Mushra	36
5.1.3	Databázový systém	38
6	Zpracování výsledných hodnot	39
6.1	Analýza dat z poslechových testů	39
6.1.1	Vizualizace a vyloučení odlišných dat	39
6.1.2	Analýza rozptylu	41
6.2	Vyhodnocení porovnání výsledků	44
	Závěr	46
	Literatura	48
	Seznam symbolů, veličin a zkratk	56
A	Přílohy	58
A.1	Výsledky naměřených hodnot	58
B	Obsah přiloženého CD	59

Seznam obrázků

1.1	Lidské sluchové pole [1].	11
3.1	Doporučení načasování testování metodou ACR [20].	22
3.2	Blokové schéma metody PEAQ [28].	24
3.3	Blokové schéma rozdělení modelů periferních uší [8].	26
3.4	Blokové schéma metody PESQ [33].	27
3.5	Blokové schéma metody PEMO-Q.	28
3.6	Schéma percepčního modelu PEMO.	29
3.7	Blokové schéma metody Rnonlin.[46]	30
4.1	Spektrální analýza vzorku kontrabasu.	32
4.2	Spektrální analýza vzorku duduku.	33
4.3	Spektrální analýza vzorku zvonkohry.	33
5.1	Ukázka uživatelského rozhraní aplikace Web Mushra.	37
5.2	Blokové schéma propojení relací databáze.	38
6.1	Histogram vybraného vzorku.	40
6.2	Krabicový graf vybraného vzorku.	40
6.3	Všechny zadané hodnoty respondentem číslo 23.	41
6.4	Graf lineární regresní funkce metody ViSQOL.	44
6.5	Graf lineární regresní funkce metody PEMO-Q.	45
6.6	Graf lineární regresní funkce metody PEAQ.	45
A.1	Přehled výsledných hodnot objektivních metrik	58

Úvod

V bakalářské práci je popsána teorie způsobu testování metrik pro vyhodnocování kvality audio signálů. K této teorii bylo nutné uvést problematiku lidské interpretace zvuku pomocí vědního oboru psychologické akustiky. V této kapitole jsou zmíněny jevy a problémy spojeny s vnímáním zvuku pomocí lidského sluchového orgánu, například vnímání hlasitosti či tónové výšky zvuku, doba trvání zvuku a maskování. Pro pozdější upřesnění byly vysvětleny dva modely lidského sluchu, které využívají objektivní metody hodnocení kvality zvuku.

Další kapitola se věnuje poškození audio signálů. Zmíněnými typy degradace signálu jsou clipping, zašumění a chybějící vzorky. Clipping se však nejčastěji dělí na dva druhy díky v rozdílu ořezu amplitudy, na hard a soft clipping. Poškození zašuměním nahrávky je způsobeno několika faktory. Zde zmíněné jsou kvantizační šum, způsobený odchylkou nakvantovaného signálu od původního z důvodu nízké zvolené bitové hloubky, a šum způsobený harmonickým rušením.

Následně jsou v práci popsány způsoby hodnocení audio signálů. Tato kapitola je rozdělena na dvě části dle dělení typů metod. První z nich jsou subjektivní metody. Část obsahuje šest nejrozšířenějších metrik, jako jsou testování ABX, škála MOS se svými stupnicemi SDG a ODG, metrika se zkratkou MUSHRA a tři typy založené na podobném principu. ACR, DCR, CCR. Druhá část je věnována objektivním metrikám vyhodnocování kvality signálu, jejichž důvodem vzniku byla časová a finanční náročnost metod subjektivních. I v této části je vyjmenováno šest metod.

Po shrnutí celé uvedené teorie byla sestavena databáze zvuků zahrnující celé slyšitelné spektrum. Tyto vzorky byly následně poškozeny a rekonstruovány třemi algoritmy. Jednalo se o metodu SPADE, 11 Relaxaci a algoritmus založený na sociální řídkosti. Pro srovnání objektivních metrik vybrána metoda MUSHRA a její volně dostupné webové rozhraní s názvem webMUSHRA. K realizace poslechových testů byl využit průmyslový server s připojenou databází pro ukládání výsledků. Vybranými objektivními metoda ke srovnání byly VisQOL, PEAQ a PEMO-Q,

Pomocí webového rozhraní bylo nasbíráno hodnocení kvality od 50 respondentů, které bylo nutné statisticky zanalyzovat. Po vizualizaci a vyřazení výsledků vybočujících z normálního rozdělení následovala analýza rozptylu s opakovanými měřeními, které poskytla validaci dat a odpověďmi na zadané hypotézy. V závěru práce jsou srovnány vybrané objektivní metriky s výsledky poslechových testů a na základě regresní a korelační analýzy jsou zde vyvozeny výsledky.

1 Lidské vnímání zvuku

Nejdříve je nutné si objasnit několik pojmů z oblasti vědního oboru psychologické akustiky a jakým způsobem je člověk schopen interpretovat zvukový signál. V oblasti percepcie zvuku totiž existuje mnoho fyzikálních vlastností našeho sluchové aparátu, se kterými poté pracují objektivní metody hodnocení kvality audio signálů. Ty často využívají dvou modelů složených ze znalostí vnímání zvuku. Jedná se o psychoakustický a kognitivní model.

1.1 Zvuk

Zvukem se obecně nazývá mechanické vlnění šířící se v okolním prostředí určitou rychlostí. Kmitání se šíří v každém materiálu rozdílně, a tak rychlost šíření tohoto vlnění úměrně závisí na hustotě či typu materiálu. Tato rychlost je přímo ovlivněna vzdáleností částic prostředí. Pokud jsou tyto částice blíže u sebe, vlnění se bude rychleji přenášet z jedné částice na druhou a rychlost šíření bude větší. Také závisí na rychlosti pohybu částic v látce. Tu ovlivňuje nejčastěji její teplota. Zvuk se tedy šíří pomocí zhušťování a zředování částic v látce a tento jev se nazývá akustický tlak [1].

1.2 Psychologická akustika

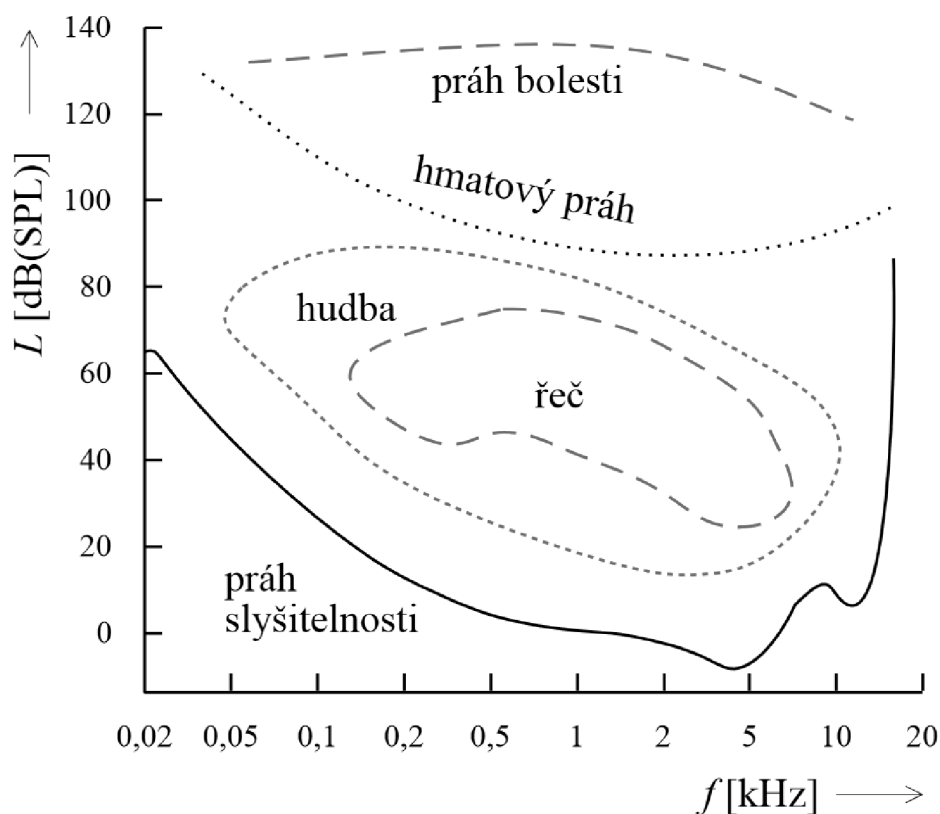
Tento vědní obor se zabývá zákonitostmi a závislostmi mezi zvukovým signálem a jeho lidským vnímáním. Jak z názvu vyplývá, jedná se o spojení dvou věd: psychologie a akustiky. Tyto vědy je však nutné doplnit o znalosti z neurofyzologie, teorie signálů, statistiky, kybernetiky a jiné.

Díky znalosti psychoakustiky jsme schopni komprimovat signály bez ztráty kvality či hodnotit kvalitu audio signálu [2].

1.2.1 Rozsah sluchu

Frekvenční rozsah, který je průměrné lidské ucho schopno vnímat, se pohybuje od 16 Hz až do 20 kHz. Tento rozsah je stanoven uspořádáním a vlastnostmi lidského ucha. Každou frekvenci, z rozsahu tohoto spektra, je lidské ucho schopno vnímat s jinou intenzitou. Tento jev se nazývá sluchové pole a obecně se jedná o dynamický rozsah sluchového orgánu. Jak je možné vidět na obr. 1.1, toto pole je omezené z obou stran sluchovou mezí. Obrázek byl převzat ze skript [1], které využily publikace [3]. Obě tyto meze jsou závislé na frekvenci a typu signálu. Spodní mezí je myšlen práh slyšení, který je minimálním vnímatelným akustickým tlakem pro naše ústrojí. Při

zvyšování intenzity jsme schopni v okolí hladiny 120 dB zvuk vnímat i hmatem. V okolí této hladiny intenzity je možné vlnění nazvat jako nepříjemné na poslech. Po dalším zvýšení hlasitosti se dostáváme k horní mezi. Ta se pohybuje v oblasti 140 dB a je pojmenovaná jako práh bolesti, kde při delším poslechu takto hlasitého signálu dochází jak k dočasné, tak i k trvalé poruše sluchu [4].



Obr. 1.1: Lidské sluchové pole [1].

Tento rozsah se vlivem stárnutí celého aparátu, zmenšuje. Člověk se rodí s konečným počtem sluchových buněk a s přibývajícím věkem tyto buňky ubývají. Jejich regenerace není možná. Ve vyšším věku oslabení sluchu podporuje snižující se pružnost celého vnitřního aparátu kolem bubínku. Také díky prahu slyšitelnosti a jeho prudkému nárůstu, se s věkem ztrácí schopnost vnímat vysoké frekvence [5].

1.2.2 Hlasitost zvuku

Z důvodu, jakým lidské ucho přijímá kmitání zvuku, bylo zvoleno, pro popis akustických veličin, logaritmické měřítko. Proto se pro měření akustických hladin využívá jednotky *bel*, v praxi se však využívá její desetina, tedy decibel. Vnímání hlasitosti

zvuku je ve skutečnosti závislé na kmitočtu a tato závislost kopíruje, dle obr. 1.1, křivku prahu slyšitelnosti [1].

1.2.3 Výška tónu

U lidského sluchovém orgánu není problém pouze u vnímání hlasitosti zvuku. Sluchové ústrojí není schopné vždy prezentovat přijímaný signál takový, jaký opravdu je. Například nejsme schopni přesně určit výšku zvuku neboli tónu. Většina lidí je schopna určit takzvanou relativní výšku tónu. To znamená, že jsou schopni určit, který ze zvukových signálů má vyšší frekvenci, nikoli však o jaké frekvence se jedná. Lidé s absolutním sluchem jsou schopni slyšený tón porovnat s vlastní zvukovou pamětí a tím pádem i tón určit. I pro jev relativní výšky tónu byla definovaná lineární stupnice s jednotou *mel*. Ta přímo vyplývá z Fechner–Weberova zákona. Stupnice udává subjektivní výšku tónu o 1000 Hz a hladině akustického tlaku 40 dB SPL hodnotu 1000 melů [1]. Hlasitost musí být předem definovaná, jelikož je výška tónu závislá na intenzitě zvuku. Například při zvyšování hlasitosti nízkých frekvencí se výška tónu subjektivně zmenšuje. U vysokých frekvencí je tomu naopak [6].

1.2.4 Doba trvání zvuku

Také vnímaná doba trvání zvuku neodpovídá u krátkých ukázek jejich skutečné délce. Jednotkou subjektivní doby trvání zvuku je *dura*. Pokud je doba trvání zvuku kratší než jedna sekunda, bude posluchači připadat tato doba jako delší, než opravdu je. Nejmenší dobou trvání zvuku, kterou je ucho schopno postřehnout, jsou přibližně 2 ms. Tento mezní časový práh nám umožňuje zaznamenat a interpretovat hlásky či hudební nástroje [3].

1.2.5 Maskování

Tento jev nastává v případě, pokud dorazí současně dva zvuky k lidskému uchu a jejich kmitočty od sebe nejsou dostatečně vzdáleny. Vnitřní ucho není schopno přijmout a zpracovat najednou, ale i v krátké časové době více podobných signálů. Možnosti zpracování zvuku ve vnitřním uchu jsou limitované délkou stočeného hlemýždě, ve kterém dochází k přenosu kmitání na nervové impulzy. Jelikož podobné kmitočty dopadají v hlemýždi na víceméně stejná místa, mozek následně vyhodnotí pouze kmitočet z místa s nejvyšším podrážděním [1].

1.2.6 Psychoakustický model lidského sluchu

Pomocí znalosti lidského vnímání sluchu je možné sestavit jeho model. Tento model transformuje vstupní signály časové domény na signál po průchodu bazilární membránou. Díky možné modularitě našel využití v kompresi, ve vkládání vodoznaků, při restauraci audio signálů, a i při hodnocení jeho kvality. Pro každý druh použití se samozřejmě využívá jiný model [8].

Například ve standartu MPEG se pro kódování a kompresi signálu využívá model, který primárně vyhledává hlasité, a tím pádem i maskující, zvuky a ty poté komprimuje. Tím dosáhne menšího datového toku a celkové velikosti nahrávky.

V případě hodnocení kvality audio signálu se využívá psychoakustický model pro objektivní odhad interpretace restaurovaného či poškozeného zvuku [7].

1.2.7 Kognitivní model lidského sluchu

Na rozdíl od psychoakustického modelu, kognitivní model lidského sluchu pracuje s interpretací zvuku přímo lidským mozku. Využívá se při hodnocení kvality zvuku, ve kterém pomáhá určit kvalitu signálu z pohledu lidské psychiky. V metodě PEAQ tento model zpracovává parametry z předchozího psychoakustického modelu. Kognitivní model trénuje vlastní neuronové sítě, které se poté snaží nahradit subjektivní lidské posluchače, se zachováním přesnosti tetování [8].

2 Poškození audio signálů

V této kapitole jsou vysvětleny a popsány různé druhy poškození audio signálů. Jelikož může poškození či degradace audio signálu nastat z mnoha důvodů, byly proto rozděleny do několika skupin. Nejsou zde však sepsány všechny, ale jen nejzásadnější a nejčastější z nich.

2.1 Clipping

Poškození typu clipping je druh nelineárního zkreslení signálu, které se vyskytuje při nesprávném nastavení dynamického rozsahu. Zařízení poté není schopné zpracovat, takto přebuzený signál. Výstupní signál tím pádem bude ořezán o všechny své maximální okamžité výchylky. U harmonického signálu se bude jednat o amplitudy vyšší, než je rozsah systému. V praxi k tomuto poškození dochází jak v analogové i digitální formě [10].

Clipping může nastat hned na několika místech zpracování audio signálu. Jedním z nich je clipping vytvořený zesilovačem, který je nucen vstupní signál zesílit takovým způsobem, pro který nemá dostatečné napájení. Tento stav zesilovače se nazývá saturace. Zesilovač v tomto stavu přetížení není schopen zesílit celý signál rovnoměrně a jeho zesílení se bude postupně zmenšovat do té doby, než se bude rovnat jedné. Poškození však může také nastat i při přehrávání zvukového signálu a systém reprodukcující signál není schopen přebuzený signál zpracovat a přehrát [9]. Výstup saturace se projeví na kvalitě zvuku a způsobí nepříjemné zkreslení signálu či praskání.

V určitých případech je tento druh poškození žádoucí a využívá se záměrně při zpracování signálu z elektrické kytary, kde toto zkreslení sinusového signálu přidává k základnímu signálu jeho harmonické složky a tím obohacuje výsledný signál. Efektové zkreslení elektrické kytary, se tak, dle druhu zapojení pro dosažení výsledného zvuku, dále dělí na několik druhů, jako jsou například *override*, *crunch*, *fuzz* a *distortion*.

Tento druh poškození se nevyskytuje pouze ve zpracování zvukového signálu, ale při zpracování kterýchkoli jiných druhů signálů. Kupříkladu při obrazových signálech, kdy je omezení způsobeno nedostatečným dynamickým rozsahem snímače či jeho nedostatečnou barevnou hloubkou. Dochází tím pádem k limitaci výstupního signálu z důvodu neschopnosti snímače tyto hodnoty zpracovat. To se může projevit jako přeexponované vyšší jasové složky obrazu nebo ve viditelných přechodech mezi odstíny barev [11].

Jelikož se clipping ve zpracování zvuku vyskytuje ve dvou různých podobách, dělíme jej na dva základní typy. **Hard clipping a soft clipping.**

2.1.1 Hard clipping

Při poškození typu hard clipping je signál tvrdě ořezán. Tvrdým ořezáním je v tomto případě myšleno omezení hodnot vstupního signálu, pokud přesáhne prahovou úroveň, plochým oříznutím maxim. Toto omezení lze dle publikace [12] matematicky vyjádřit následovně:

$$y_c[n] = \begin{cases} y[n] & \text{pro } |y[n]| < \theta_c, \\ \theta_c \cdot \text{sgn}(y[n]) & \text{pro } |y[n]| \geq \theta_c, \end{cases} \quad (2.1)$$

kde mezní hodnoty $-\theta_c$ a θ_c určují šířku intervalu, pro který vstupní signál nebude ořezán. V tomto intervalu je přenos vstupního signálu lineární. Jakmile však tento signál přesáhne mezní hodnoty, stane se z něj signál nelineární a bude ploše ořezán.

Výstupní signál z této saturace bude obohacen o liché harmonické složky původního vstupního signálu. Při naprosto přesném a rovném ořezání by mělo být lichých složek teoreticky nekonečné množství.

2.1.2 Soft clipping

Soft clipping na rozdíl od hard clippingu nelimituje výstupní signál tvrdým a plochým oříznutím amplitudových špiček vstupního signálu, ale pouze jej postupně redukuje ve vztahu s mezní úrovní daného systému. Při tomto druhu poškození nedochází ke ztrátě informace a pokud je známa převodní charakteristika systému, je možné tento poškozený signál zrekonstruovat do kvality původního signálu.

Z tohoto důvodu se soft clipping využívá v kytarových zesilovačích, pedálových efektových filtrech a pro ochranu zesilovače před hard clippingem a tím pádem i před rušivým zkreslením výstupního signálu. Zesilovač sleduje hodnoty svého výstupního napětí a pokud se přiblíží k jeho maximu, začne toto napětí komprimovat. Obecná definice převodní charakteristiky soft clippingu lze modelovat libovolnou sigmoidou (například $\arctan(x)$, $\tanh(x)$, $\frac{x}{1+|x|}$ nebo $\frac{x}{\sqrt{1+x^2}}$).

Jedna z možných definicí je následující [13]:

$$y_c[n] = \begin{cases} y[n] & \text{pro } |y[n]| < \theta_c, \\ \text{sgn}(y[n])[(\theta_c - r) \tanh(\frac{|y[n]|-r}{\theta_c-r}) + r] & \text{pro } |y[n]| \geq \theta_c, \end{cases} \quad (2.2)$$

kde θ_c vyjadřuje nastavitelný práh a r je mezní hodnota vymežující horní hranici výstupního signálu, za kterým se tento signál začne komprimovat. Pro výpočet hodnoty r je nutné definovat proměnnou k , která zastupuje tvrdost převodní charakteristiky a nabývá hodnot od 0 do 100. Pomocí ní je možné definovat výpočet jako:

$$r = 0,01 \cdot k \cdot \theta_c, \quad (2.3)$$

z čehož vyplývá, že pokud se bude charakteristika tvrdosti $k = 100$, poté se komprimace výsledného signálu vyrovná poškození zvaném hard clipping, viz část 2.1.1.

2.2 Zašumění

Šum na pozadí signálu obecně znamená přítomnost náhodných rušivých složek v daném signálu. Ty se mohou vyskytovat v nejrůznějších frekvenčních pásmech, či po celém slyšitelném spektru a mimo něj. Existuje několik druhů šumů, kde některé našly využití a jsou tím pádem užitečné. Jedním z nich je takzvaný bílý šum, který má rovnoměrnou výkonovou spektrální hustotu a obsahuje tak všechny frekvence. Využití našel, díky svým charakteristikám, v syntéze a filtraci zvuku, v sirénách a výstražných signálech, v přístrojích pro ochranu soukromí a lepší spánek, ale i pro podkresový signál omezující rušení jiným typem zvuku, například proti hluku od ostatních hostů v restauracích.

Zašumění signálu je často způsobeno špatnými vstupními podmínkami při záznamu signálu, kdy je snímán navíc i okolní hluk a šum. Například rušná ulice za neodtlumenými okny a zdmi. Dalším častým důvodem zašumění signálu je přidání šumu při zpracování signálu zesilovačem, převodníkem, ale i delšími neodstíněnými kabely. Každý analogový přenos zvukového signálu může degradovat signál a přidat šum. Z tohoto důvodu je u každého zařízení psán odstup signálu od šumu. Příkladem tohoto typu šumu může být takzvaný *hiss*. Ten se objevuje při záznamu na analogové magnetické pásky a je způsoben rozměrem magnetických nosičů informací. Dále se objevuje i při převodu signálu z analogového systému na digitální a opačně. Perfektním příkladem je kvantizační šum.

2.2.1 Kvantizační šum

Nejdříve je nutné si vysvětlit důležité pojmy. Vzorkování jako takové, je převod signálu se spojitým časem na signál s diskretním časem. Na takzvaný diskretní signál. Kvantování má za úkol tyto diskretní vzorky vyjádřit pomocí konečné množiny čísel. Při kvantování vznikne kvantizační šum jako odchylka nakvantovaného signálu od původního.

Tato chyba je způsobená nízkým počtem kvantovacích hladin převodníku neboli jeho bitovou hloubkou. Standardně se ve zvukové technice využívá 16, 20 a 24 bitů [14]. Každý přidávaný bit znamená zvýšení odstupu signálu od kvantovacího šumu přibližně o 6 dB.

Tento šum je možné matematicky definovat, při harmonickém signálu, jako odstup signálu od kvantovacího šumu SNR a to následovně [1]:

$$\text{SNR} = 6,02n + 1,76 \text{ [dB]}, \quad (2.4)$$

kde n udává bitovou hloubku převodníku bez znaménkového bitu [14].

Pro zjednodušený výpočet se využívá i odhad založený na velikosti kvantovacího kroku [1]:

$$\text{SNR} = 20 \log(2n - 1) \text{ [dB]}. \quad (2.5)$$

Kvantizační šum se dnes aplikuje do maskovacích pásem při redukci dat pro zjištění neslyšitelnosti frekvenčního maskování. Při prezenci maskování ve zvukové stopě, rozdělené dle částí kmitočtových pásem, nalezneme frekvence, které je možné zkomprimovat či úplně vyřadit. Každé z pásem je individuálně zpracováno na základě informace z psychoakustického modelu daného systému [1].

2.2.2 Harmonické rušení

Harmonické rušení je možné dělit na několik druhů. Typ *hum*, lze definovat jako elektromagnetickou indukci harmonického signálu v zemních smyčkách a rozvodných sítí pro transport signálu. Také se může objevit při indukci v AC transformátorech. Při poslechu tohoto typu poškození je slyšet hukot či hučení. Jelikož se jedná o harmonické rušení, vyskytují se zde násobky základního kmitočtu.

Dalším typem harmonického rušení je *buzz*. V překladu bzukot. Toto rušení vzniká například při prezenci zvuku ze zářivkových světel a tím pádem se dle charakteristiky tohoto zvuku, vyskytuje na vyšších kmitočtech, kde maskuje signál, který je nahráván.

2.3 Chybějící vzorky

Jak již bylo v podkapitole Kvantizační šum 2.2.1 popsáno, vzorkováním je myšlen převod spojitého signálu na diskrétní signál. Při této přeměně byl analogový signál zpracován a byly zaznamenány hodnoty v přesných časových intervalech neboli vzorkovacím kmitočtu. Výsledné vzorky lze poté přenášet digitální soustavou a poté znovu převést na spojitý průběh.

Přenos zvukových signálů skrz digitální rozhraní skrývá mnoho problému a nedostatků. Například při tranferu dat pomocí digitálních zařízení přes synchronizační rozhraní typu ADAT, Word Clock či AES3. Pokud synchronizace neproběhne v pořádku a dva komunikující systémy nemusí očekávat příjem dat, nebo nepracují se stejnými vzorkovacími kmitočty, dojde ke ztrátě vzorků a tím ke degradaci zvukové stopy [14].

I při přenosu zvuku přes Ethernetové či DMX512 rozhraní může dojít ke ztrátě datových paketů z důsledku přerušení spojení, nebo například při přehlcení sběrnice [14].

K tomuto poškození však nemusí docházet pouze při přenosu zvuku, ale i při poškození nosičů, jak digitálních, tak i analogových záznamů. V obou případech je nutné mechanicky poškodit segment s uloženými daty. Například k poškrábání přenosové strany dochází jak u nosiče typu CD, tak i u vinylového nosiče. Z důvodu výpadků a silných lupnutí bude poslech takto poškozeného signálu velice rušivý, až nepříjemný. Restaurace signálu je možná, ale ne pokud vzorků chybí velké množství.

3 Způsoby hodnocení audio signálů

Potřeba hodnotit či testovat jakýkoli zvukový výstup existuje od počátku digitálního zpracování zvuku, a to jak při jejich přenosu nebo pro zjištění kvality restaurace poškozených signálů. Pro tuto potřebu vzniklo nepřehledné množství metod, jak kvalitu měřit či vyhodnotit. Tato kapitola jedná o nejvíce používaných nebo základních z nich.

3.1 Subjektivní metody pro hodnocení

Metody pro hodnocení kvality signálu jde dělit na dvě základní kategorie dle způsobu provádění testování. Jednou z nich je subjektivní kategorie. V té, jak název napovídá, hraje klíčovou roli subjektivní názor. Ten mohou poskytnout pouze testování jedinci neboli posluchači. Poslechové testy musí být pro jejich subjektivitu obsáhlé na subjekty a sesbíraná data. Tudíž zde hraje roli i časová náročnost a následná práce vyhodnocováním dat. Nežádoucí vliv na výsledek může mít jak místo testování a jeho akustika, fyzický či psychický stav jedinců, ale také různorodost celé testované skupiny. Výsledkem je však skutečná subjektivní kvalita audio signálu.

3.1.1 ABX

Test ABX pracuje pouze jako ukazatel jsou-li dva zvukové signály shodné či naopak odlišné při poslechu posluchačem. Testované skupině subjektů jsou předloženy tři nahrávky označené znaky A, B a X. Signál A či B je buď referenčním signálem, nebo naopak signálem rekonstruovaným, přičemž signál X je totožný s jedním z nich. Posluchač má v tomto testu vybrat, které z předložených nahrávek se shodují. Výstupem této metody mohou být pouze dvě z možností. Posluchač vybral správně či nikoliv [15] [16].

3.1.2 MOS

Škála MOS (Mean Opinion Score), normou Mezinárodní komunikační unií [17], je hodnota z předem definované stupnice, která je průměrem všech provedených subjektivních testů na respondentech. Metoda MOS se dnes objevuje napříč všemi druhy hodnocení různých digitálních výstupů, jak v oblasti kvality řeči, zvuku, obrazu či videa. Dále se využívá v mnoha aplikacích, systémech, v laboratorních a monitorovacích prostředích. Stupnice MOS se však nepoužívá pouze v subjektivních testech kvality, ale i jako výstup objektivních algoritmů při měření kvality pomocí automatizovaných metod. „MOS se však často používá bez dostatečného zvážení toho, jak

Tab. 3.1: Tabulka hodnotící škály MOS.

Poškození (<i>Impairment</i>)	SDG	ODG
Neznatelné (<i>Imperceptible</i>)	5,0	0,0
Znatelné, ale ne nepříjemné (<i>Perceptible, but not annoying</i>)	4,0	-1,0
Trochu nepříjemné (<i>Slightly annoying</i>)	3,0	-2,0
Nepříjemné (<i>Annoying</i>)	2,0	-3,0
Velmi nepříjemné (<i>Very annoying</i>)	1,0	-4,0

byla data získána, a inherentních omezení a omezení vyplývajících z návrhu subjektivních testů nebo objektivních metrik.“ Metrika se tím pádem často využívá v nesprávného důvodu či kontextu [18].

Nejpoužívanější pětibodová stupnice MOS se tak dělí na dva druhy hodnocení výsledku: SDG (Subjective Difference Grade) a ODG (Objective Difference Grade). Výstupy z obou druhů stupnic jsou totožné, rozdílem je využití škál buď v subjektivním nebo objektivním testování. Rozdílné jsou i hodnoty použité škály, tak jak je zobrazeno v tab. 3.1, jejíž česká pojmenování byla převzata z literatury [19].

3.1.3 MUSHRA

Metrika se zkratkou MUSHRA (Multi Stimulus test with Hidden Reference and Anchors) je definovaná doporučením ITU-R BS.1534-3 [21] a je vytvořena k hodnocení tzv. „střední kvality zvuku“ výstupu ztrátových kompresních algoritmů, jako například při přenosu zvuku skrz internetové a satelitní rozhraní. Metoda převzala mnoho aspektů z normy [16].

Testování probíhá následovně. Posluchači je předloženo dle doporučení maximálně třináct zvukových nahrávek a aby jejich maximální délka zvukového vzorku nepřesahovala 12 sekund. Toto opatření zabraňuje únavě posluchačů, snížení celkové doby trvání testu, ale také celkové konzistenci testování. Při nutnosti využít delších sekvencí, například při dlouhém a pomalu se pohybujícím se zvuku po trajektorii, je nutné odůvodnit toto rozhodnutí v závěrečném protokolu testu.

Test obsahuje referenční signál pro srovnání, znovu referenční signál, ale v tomto případě již skrytý, pro kvalitnější výstup hodnocení a zamezení chyb respondenta. Dále se v testu nachází alespoň jedna náhodně skrytá kotva. Tyto kotevní signály je možné filtrovat dle nutnosti a míře testované kvality. Neměli by se však využívat ke změně měřítka mezi různými testy. Standartně se využívá kotvy filtrované skrz dolní propust s mezním kmitočtem 3,5 kHz. Pro střední kvalitu se využívá kotva o mezní frekvenci 7 kHz. Šířky pásem kotev odpovídají šířky doporučených pásem různorodých druhů vysílání a přenosu signálů dle doporučení ITU-T G.711, G.712,

G.722 a J.21.

Přesné charakteristiky filtru dolní propusti 3,5 kHz jsou definované následovně:

- Maximální zvlnění pásma = $\pm 0,1$ dB
- Minimální útlum při 4 kHz = 25 dB
- Minimální útlum při 4,5 kHz = 50 dB.

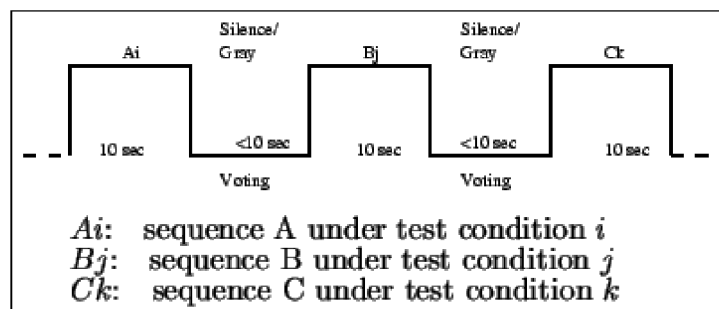
Pomocí těchto kotev je stanoven rozsah stupnic. Metoda využívá stupnice CQS. Ta je definovaná normou ITU-R BT.500 a je navržena pro hodnocení kvality obrazu. Její rozložení se skládá z pěti stupňů, jak je vyobrazeno v tab. 3.2, jejíž překlad je využit z publikace [6]. V praxi je však často využita celá stupnice od 0 do 100. Pokud by například test měl obsahovat 8 zvukových stop a bylo by nutné využít přesnější hodnocení, poté se zvedne celkový počet stop pro respondentův poslech na 11 a referenční signál. Systém by tedy obsahoval jeden označený referenční signál, osm testovaných nahrávek, jednu skrytou referenci, nízkofrekvenční kotvu a kotvu středního pásma [21].

Tab. 3.2: Tabulka hodnotící škály testu MUSHRA

Kvalita (<i>Quality</i>)	Hodnocení
Vynikající (<i>Excellent</i>)	100 – 80
Dobrá (<i>Good</i>)	80 – 60
Přijatelná (<i>Fair</i>)	60 – 40
Špatná (<i>Poor</i>)	40 – 20
Nepřijatelná (<i>Bad</i>)	20 – 0

3.1.4 ACR

ACR (Absolute Category Rating) je metoda subjektivního hodnocení signálu na základně absolutní kvality signálu, definovaná normou [17]. Testování probíhá ve dvou fázích o doporučené délce deset sekund, viz obr. 3.1. Vždy je přehrána ukázka a ihned poté se provádí hodnocení (tzv. „hlasovací čas“). Stupnice této metody se různí dle nutné výstupní přesnosti. Základní a nejpoužívanější škála obsahuje pět hodnot ze škály MOS a to od neznatelné po velmi nepříjemnou, tak jak je vyobrazeno v tab. 3.1. Číslování se shoduje se sloupcem SDG. Pokud je požadována vyšší přesnost, vloží se mezi každý stupeň jeden mezistupeň. Tím vznikne devítibodová škála hodnocení. Existuje i jedenáctibodová stupnice, která je číslována od 0 do 10 a je slovně hodnocena od „nejhorší“ po „nejlepší“.



Obr. 3.1: Doporučení načasování testování metodou ACR [20].

3.1.5 DCR

Na rozdíl od metody ACR se při testování kvality zvuku pomocí metody DCR (Degradation Category Rating) použijí hned dva stimuly, v našem případě zvuky. První z nich je původní referenční nahrávka a druhá je degradována jakýmkoli testovaným zkreslením, šumem, nebo se jedná o signál, který prošel rekonstrukcí. V tomto případě má posluchač možnost porovnat oba signály, a tak zhodnotit velikost degradace. Tím pádem je hodnocení citlivější na menší zkreslení, degradaci, či šum než použití metody ACR. Norma [17] doporučuje použití metody DCR při nižší degradaci signálu na rozdíl od metody ACR, která by se měla dle doporučení využívat na středně a silně degradované vzorky.

3.1.6 CCR

Další subjektivní metodou definovanou Mezinárodní telekomunikační unií [17] je metoda CCR (Comparison Category Rating). Metrika obdobná s předchozí zmíněnou DCR. Test probíhá velice podobně jako je tomu u metody DCR. Porovnávají se dva vzorky, referenční a druhý degradovaný, pouze s tím rozdílem, že jejich pořadí je vybráno náhodně a posluchač tedy neví, který z nich je referenční. Také stupnice byla pro tento styl testování změněn. Posluchač hodnotí podobnost druhého signálu k prvnímu a vše zaznamenává do sedmistupňové škály. Ta nabývá hodnot od -3 do 3 , kde nulový bod vyjadřuje shodnost obou signálů.

3.2 Objektívni metody pro hodnocení

Další kategorií pro hodnocení kvality zvukového signálu je kategorie objektivní. Název vypovídá o nemožnosti ovlivnění výsledku jakýmkoli okolnostmi, protože výsledek bude vždy stejný. Důvodem vzniku byla náročnost jak časová, tak i finanční,

při provádění subjektivních poslechových testů. Metody z této skupiny zvládnou rychle a efektivně zhodnotit kvalitu zvukových zařízení, systémů či sítí. Nicméně se tyto metriky stále snaží co nejvíce se přiblížit výsledku subjektivních metod, a tak zahrnují do svých algoritmů vlastnosti lidského sluchového orgánu a často i psychoakustický model vnímání zvuku.

3.2.1 SNR

V celém znění Signal-to-noise ratio, je v překladu poměr mezi signálem a šumem. SNR – je bezrozměrný poměr výkonu signálu k výkonu šumu obsaženého v záznamu.

Fyzikálně můžeme tento vztah vyjádřit následovně [23]:

$$\text{SNR} = \frac{P_s}{P_n} \quad (3.1)$$

Vzorec lze převést do logaritmického měřítka a tím dosáhnout výstupu v decibelech [23].

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{P_s}{P_n} \quad (3.2)$$

V případě hodnocení kvality rekonstrukce signálu SNR reprezentuje podobnost dvou signálů. Čím vyšší výsledné číslo SNR, tím vyšší je podobnost testovaného signálu k referenčnímu a tím pádem i kvalitnější [24]. Výpočet tohoto odstupe vychází z předšlého vzorce 3.2, avšak signály jsou reprezentovány vektory. Čím vyšší výsledek je, tím podobnější si signály jsou.

Následující vzorec vychází z výzkumné práce [24] a je používán k výpočtu podobnosti signálu v jejich experimentech.

$$\text{SNR}(y, \hat{y}) = 10 \log_{10} \frac{\|y\|^2}{\|y - \hat{y}\|^2}, \quad (3.3)$$

kde y je vektor původního signálu a \hat{y} je vektor rekonstruovaného signálu. Odečtením testovaného vzorku od referenčního a následným vydělením ℓ_2 normy hodnoty referenčního vzorku výsledkem ℓ_2 normy spodního výpočtu a následným vynásobením desetinásobkem logaritmu deseti, dostaneme výsledek testu SNR.

Problémem této metody je, že konečný výsledek udává pouze informace o podobnosti testovaných signálů a není použit žádný z akustických modelů vnímání zvuku.

3.2.2 PEAQ

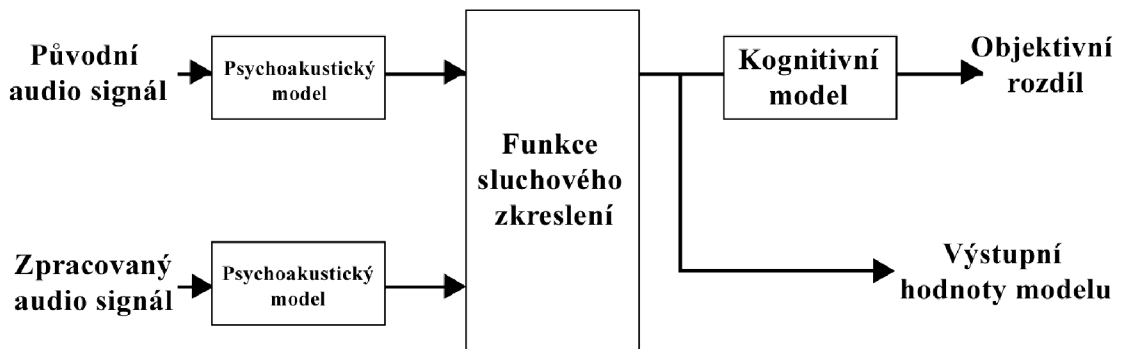
Je zkratkou pro metodu s názvem: Perceptual Evaluation of Audio Quality, která je normou vytvořenou Mezinárodní telekomunikační unií (zkrácené ITU dle Inter-

national Telecommunications Union), s přesným označením ITU-R BS.1387 [25].

Využívá se při vývoji a testování multimediálních zařízení, kodeků, sítí ale také ji lze použít pro objektivní srovnání mezi zařízeními [26]. Při kombinaci s dalšími algoritmy ji lze efektivně využít pro celkové hodnocení systému. Algoritmus využívá psychoakustické principy, které využívá pro zachování úrovně kvality zvuku při snížení bitové rychlosti neboli kompresi. Využití našla zejména v kompresním kódování MPEG a to ve všech svých rozšířeních (MPEG 1, 2 a 3) [27].

Přesnost metody PEAQ při odhadu kvality zařízení nebo systému je pro koncového uživatele důležitá. Špatná kvalita signálu může být pro uživatele nepříjemná, a dokonce i rušivá. Zejména u špičkových zvukových systémů, u kterých bude uživatel zařízení používat k poslechu řeči, hudby a dalších složitých zvuků. Kromě toho lze PEAQ použít k rozlišení různých zařízení z hlediska kvality. K hodnocení kvality těchto zařízení a systémů se tradičně používaly subjektivní poslechové testy, ale takové testy poslechu jsou drahé a časově náročné. Je však třeba poznamenat, že PEAQ byl navržen pouze ke klasifikaci signálů s extrémně malými poruchami [8].

PEAQ lze popsat jako kombinaci dvou částí: první část je psychoakustický model a druhá je kognitivní model, v takovém pořadí jako je uvedeno na obr. 3.2.



Obr. 3.2: Blokové schéma metody PEAQ [28].

Tato metoda hodnotí velikost degradace signálu neboli DCR 3.1.5 (Degradation Category Rating) [29], a ne absolutní kvalitu ACR 3.1.4, tak jako metoda MOS 3.1.2.

Dále se metoda dělí do dvou verzí: „Základní verze“ se využívá v aplikacích, kde je kladen důraz na rychlost a efektivitu, a „Pokročilá verze“ je více přesná, ale je čtyřikrát výpočetně náročnější. Hlavním rozdílem mezi oběma verzemi je, že základní verze má pouze jeden model periferního ucha, který je založený na FFT (Fast Fourier transform – Rychlá Fourierova transformace), zatímco pokročilá verze má tyto modely dva, jeden stejný jako v základní verzi a druhý založený na bance filtrů. Blokové schéma těchto dvou modelů je znázorněno na obr. 3.3. Výstupem

obou verzí jsou tzv. MOV, ze kterých je následně počítán výsledek na stupnici ODG. Toto skóre představuje očekávanou percepční kvalitu degradovaného signálu a může nabývat hodnot od 0 do 4, kde 0 představuje signál s nepostřehnutelným zkreslením a 4 představuje signál s velmi nepříjemným zkreslením. Každá z verzí má vlastní počet bodů MOV. Základní jich produkuje 11 a pokročilá pouze 5. Každý z bodů představuje vstup do neuronové sítě [8].

Vývoj této metody se ubírá na zlepšení percepční výkonnosti, jako například Huberův model nové klasifikace [30] a nový kognitivní model v [31]. Tyto nové práce otevírají možnost přidání nových funkcí a tím se zlepšit přesnost při zachování úrovně složitosti.

3.2.3 PESQ

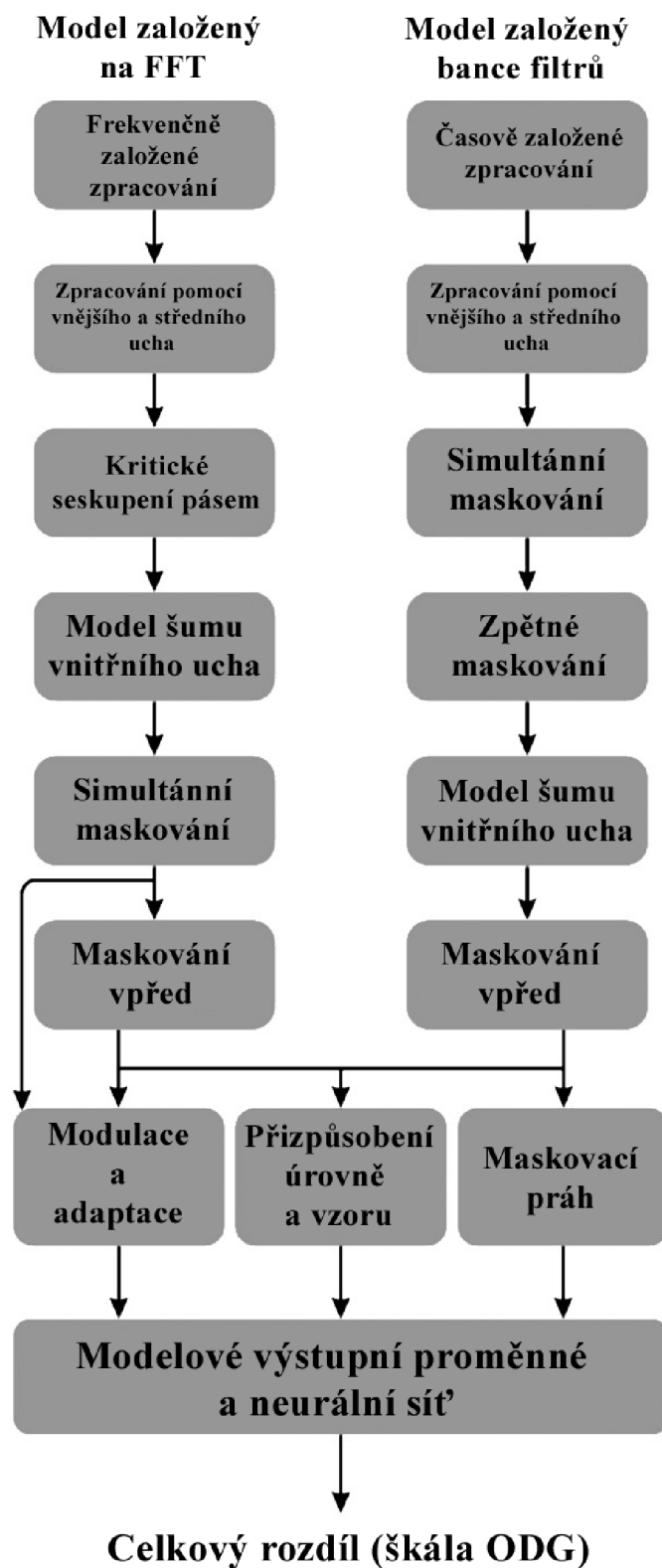
Metoda s názvem PESQ (Perceptual Evaluation of Speech Quality) je popsána normou ITU v roce 2001 a nese název ITU-T P.862 [32]. Vychází z předchozí metody PSQM a využívá algoritmus PAMS (Perceptual Analysis Measurement System) navržený britskou skupinou British Telecom. Důvodem vývoje této metody byl příchod přenosu telefonního signálu přes protokol IP. Tím nastalo několik problémů s kontrolou procházejících dat v paketech neboli skupinách bitů. Ty mohou přicházet v náhodných časových intervalech, a tak není možné je kontrolovat bez ukládání dat do registru a poté následné rekonstrukce. To však způsobuje zpoždění signálu. Žádná z předešlých metod nepracovala s tímto problémem [33].

PESQ tím pádem přinesla řešení pomocí časového vyrovnání signálů, ekvalizací nežádoucích efektů a algoritmem pro průměrování zkreslení, viz obr. 3.4.

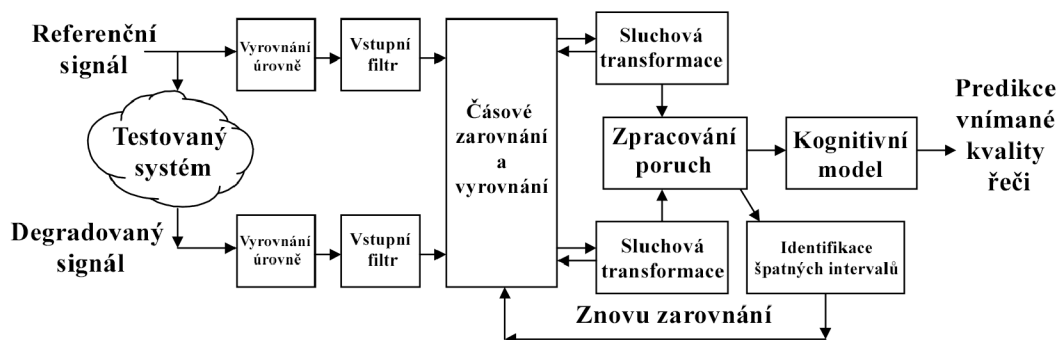
Časového vyrovnání signálů je dosaženo pomocí časové a amplitudové korelace, které jsou zvláště pro každý blok a tím pádem je možné zachytit jejich zpoždění. Následně jsou oba signály zpracovány algoritmem, který má výsledek hodnot v dvou-rozměrném časovo-frekvenčním poli. Toto pole přibližně odpovídá lidskému vnímání daného signálu. Pro odhad hodnotících parametrů je použita tzv. poslechová odchylka, které je rozdílem obou výsledných hodnot ze signálů, jak je vyobrazeno na blokovém schématu. Ve výsledku metody však nejsou zahrnuty poruchy vyskytující se v obousměrné komunikaci, ztrátu hlasitosti v signálu a případně ozvěny.

Tato metoda pracuje na MOS škále [29] pouze s tím rozdílem, že nepracuje v rozsahu 1 až 5, ale v rozsahu $-0,5$ až $4,5$. Pro přepočítání hodnot pro porovnání je nutné se držet doporučení ITU-T P.862.1 [34]. To zajišťuje obousměrný převod mezi škálou MOS SDG a výsledků této metody. Pro přepočítání výsledných hodnot metody na rozsah MOS 3.1.2 využijeme následující vzorec:

$$y = 0,999 + \frac{4,999 - 0,999}{1 + e^{-1,4945x + 4,6607}}, \quad (3.4)$$



Obr. 3.3: Blokové schéma rozdělení modelů periferních uší [8].



Obr. 3.4: Blokové schéma metody PESQ [33].

kde x je výsledek testu PESQ a y je hodnota na škále MOS. Pro opačný převod se využívá tohoto vzorce:

$$x = \frac{4,6607 - \ln\left(\frac{4,999-y}{y-0,999}\right)}{1,4945}. \quad (3.5)$$

3.2.4 PEMO-Q

Metoda navržená v roce 2006 Rainerem Huberem a Birgerem Kollmeirem [35] využívá vlastní percepční model, zkráceně PEMO, který však lze zaměnit za vylepšený sluchový model bez narušení vnitřní struktury celé metriky, jako například model CASP, který využívá základy této metody [36].

Algoritmus celé metody je znázorněn na blokovém schématu na obr. 3.5, který vychází z původní práce [35] spolu s českým překladem z [38]. Vstupní signály jsou ihned na začátku metody časově srovnány a úrovně sjednoceny a poté jsou jednotlivě zpracovány zvoleným percepčním modelem, v našem případě PEMO.

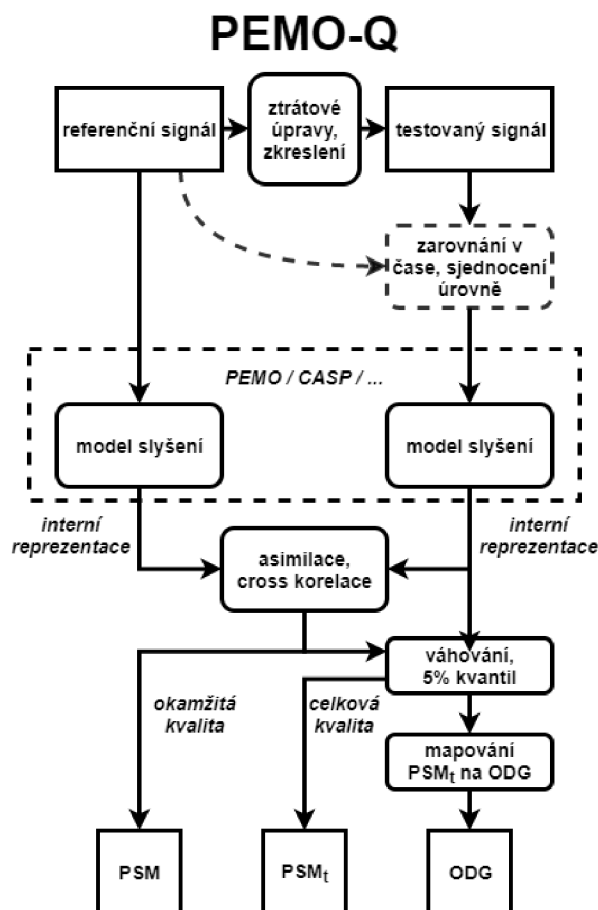
Percepční model PEMO tvoří několik bloků, jak jde vidět na obr. 3.6, který je taktéž převzat z původní práce [35] spolu s českým překladem z [38].

V prvním bloku najdeme 35 gammatónových filtrů čtvrtého řádu o středních kmitočtech od 235 Hz do 14,5 kHz, s odstupem a šířkou pásma odpovídající hodnotě 1 ERB. Tato hodnota je dle publikace [37] definovaná následovně:

$$E = 21,4 \log(4,37F + 1), \quad (3.6)$$

kde E je hodnota ERB a jednotka F jsou frekvence v kHz.

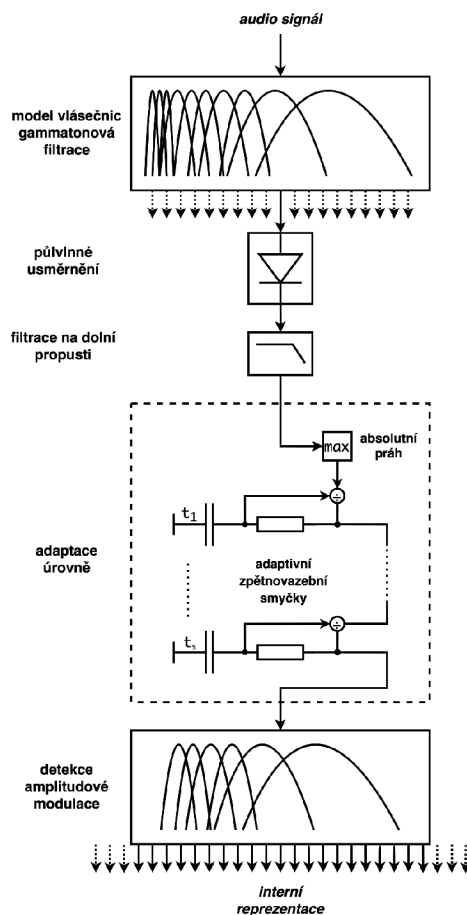
Toto frekvenční rozdělení imituje bazilární membránu a v dalších blocích jsou jednotlivá pásma zpracována zvlášť. Nejdříve probíhá transformace na nervové impulzy pomocí půlvalného usměrnění. Poté jsou signály filtrovány dolní propustí a následně



Obr. 3.5: Blokové schéma metody PEMO-Q.

přivedeny do bloku adaptivní filtrace, ve které je signál limitován pro normalizaci maximální úrovně vstupního signálu. Dále signál prochází několika zpětnovazebními smyčkami zapojenými v kaskádě. Takto normalizovaný signál je v posledním bloku modelu PEMO filtrován osmi kmitočtovými filtry, kde každý z nich má na starost jiné pásmo frekvencí.

Po průchodu signálu percepčním modelem je provedena asimilace a vzájemná korelace signálů přes jejich délku, frekvenční pásma a nezávisle přes všechna pásma z modulační banky. Výsledkem metriky je několik hodnot. První z nich je hodnota PSM. Ta nabývá hodnot od 0 do 1 a reprezentuje okamžitou kvalitu vzorku. Druhou výstupní hodnotou je PSM_t, která představuje celkovou průměrnou hodnotu vnímané kvality nezávislou na druhu vstupního signálu. Pro porovnání výstupu této metody s ostatními jak objektivními, tak i subjektivními metodami, je možné hodnotu PSM_t převést na více rozšířenou škálu ODG ze stupine MOS 3.1.2.



Obr. 3.6: Schéma percepčního modelu PEMO.

3.2.5 ViSQOL

Metoda pro objektivní porovnávání kvality řeči ViSQOL (Virtual Speech Quality Objective Listener) byla vyvinuta výzkumnou skupinou Sigmedia (Signal Processing and Media Applications) [39] a využívá psychoakustický model citlivosti lidského sluchu při porovnávání referenčního signálu s degradovaným. Implementace algoritmu v softwaru MATLAB je dostupná na webových stránkách skupiny Sigmedia. Výstupem metriky je číselná hodnota pětibodové škály MOS.

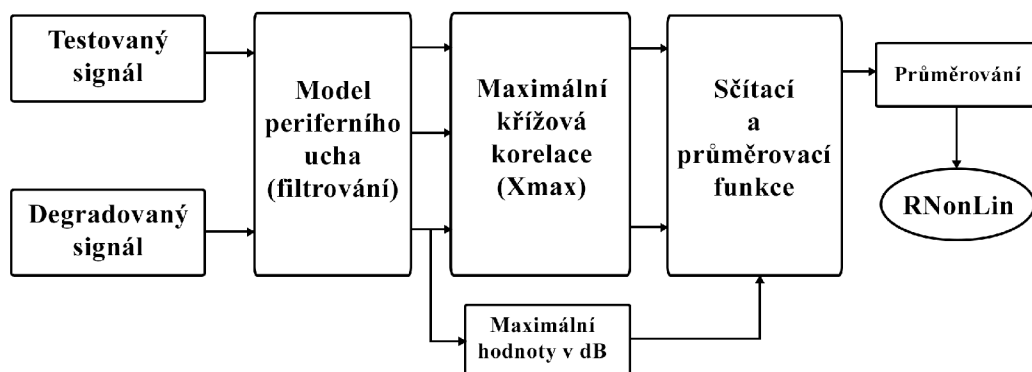
Oba vstupní signály jsou nejdříve vyrovnány a zároveň jsou vyprodukovány spektrogramové reprezentace pro každý signál zvlášť. Poté se signály časově srovnají a začnou se porovnávat jejich spektrogramové reprezentace spolu s mapováním podobnosti s procesní mapou objektivní kvality MOS [40].

Adaptací této metody vznikla specializovaná verze s názvem ViSQOLAudio, která se, jak už název napovídá, specializuje na objektivní hodnocení kvality zvuku. Pro změnu specializace bylo nutné upravit algoritmus následujícími způsoby. Detekce aktivity hlasu byla odstraněna. Počet hodnocených kmitočtových pásem se

zvýšil na víceméně celé pokrytí šířky pásma sluchu, tedy od 50 Hz do 20 kHz. Tato úprava patří mezi nejzásadnější, jelikož základní metoda ViSQOL pracuje pouze s pásmy od 50 Hz do 8000 Hz. Výsledek metody škály MOS byl změněn na jednoduchou stupnici podobnosti, která má rozsah pouze od 0 do 1 (0 – odlišné, 1 – totožné)[41]. I tento algoritmus je dostupný ke stažení ze stránek vývojářů [42].

3.2.6 Rnonlin

Metoda Rnonlin analyzuje rozdíl mezi vstupním testovacím signálem a jeho zkresleným vstupem. Pro zpracování a vyhodnocení stupně degradace signálu využívá model periferního sluchového systému, včetně filtrování vnější a střední části ucha. Tato metrika se využívá pro hodnocení nelineárního zkreslení, což je například clipping nebo kvantizace [43].



Obr. 3.7: Blokové schéma metody Rnonlin.[46]

Blokové schéma této metody (viz obr. 3.7) ukazuje celý proces výpočtu výsledné hodnoty Rnonlin. Nejdříve se oba signály časově vyrovnají, aby bylo odstraněno zpoždění způsobené nelineárním systémem. Následně jsou křivky filtrovány pomocí filtrů 4097 FIR [44]. Poté jsou oba signály znovu filtrovány řadou 40 gammatonových filtrů s šířkou pásma 1-ERBN. Tento druh filtrů se hojně využívá jako model sluchových filtrů a jsou tvořeny sinusovým tónem a rozdělením gamma [45]. Po modelu periferního sluchového systému jsou oba průběhy rozděleny na 30 ms dlouhé části, které jsou poté použity ve výpočtu maximální hodnoty normalizované křížové korelace X_{max} mezi těmito signály. Pro každou část jsou vypočteny hodnoty X_{max} zvlášť a následně jsou tyto výsledky sečteny a zprůměrovány pro jedinou výslednou hodnotu Rnonlin [46].

4 Příprava poslechových testů

Tato kapitola popisuje celkovou práci provedenou před spuštěním testování. Je zde zahrnut výběr metody, složení zvukové databáze, druh poškození a jeho odůvodnění, následná rekonstrukce poškozených signálů a objektivní ohodnocení kvality vzorků.

4.1 Zvolená metoda subjektivního testování

Po zvážení výše vypsanych metrik byla pro realizaci srovnání objektivních metod vybrána testovací metoda MUSHRA (viz část 3.1.3). Tato metrika se jeví jako nejvhodnější z důvodu vyšší přesnosti hodnotící škály.

4.2 Zvuková databáze

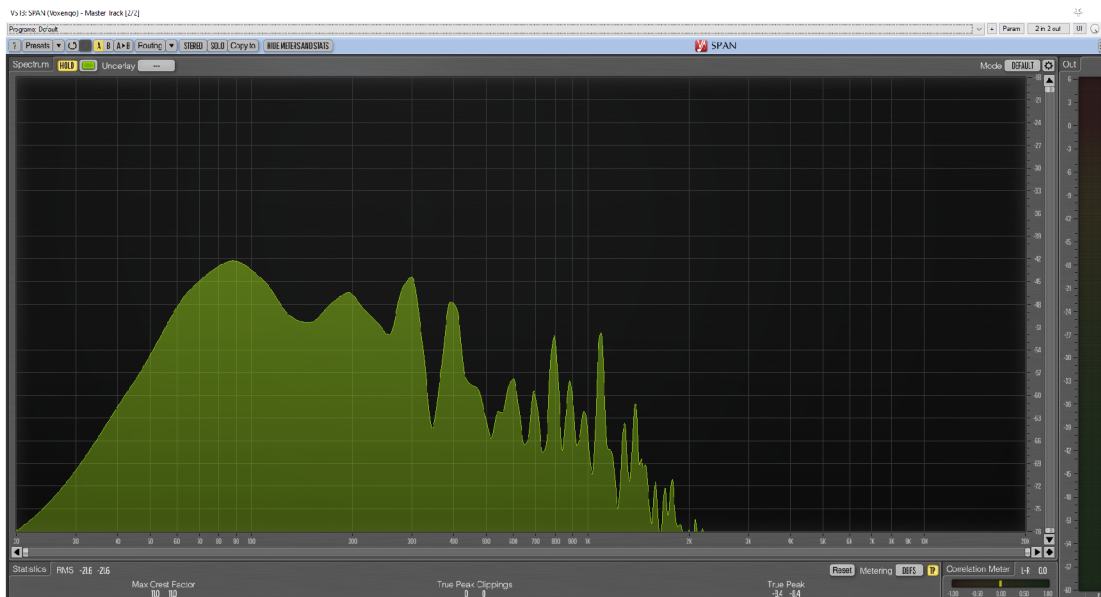
Za účelem testování pomocí poslechových testů, byla vytvořena databáze zvuků. Do ní bylo nutné zahrnout několik faktorů. Prvním z nich byla dostupnost audio nahrávek a možnost jejich využití pro výzkum s ohledem na autorské práva. S ohledem na tento faktor, byla vybrána otevřená databáze zvuků s názvem Freesound ¹ jejíž licence zaštiťuje americká nezisková organizace Creative Commons. Zvukovou databázi tvoří převážně uživatelé a je čistě na nich, jaké práva pro šíření jejich tvorby dají ostatním uživatelům. Na webových stránkách databáze Freesound jsou uvedené čtyři kategorie licencí od Creative Commons. Všechny z nich umožňují využít zvuky pro výzkum.

Dále bylo nutné zahrnout škálu nahrávek obsahující co největší frekvenční pestrosti spolu se zachováním jejich výskytu v reálných případech. Z tohoto důvodu byly při výběru audio stop preferovány akustické nástroje, spolu s celkovými kompozicemi několika nástrojů a v neposlední řadě i lidskou řeč. Spektrální analýza byla provedena ve volně dostupné verzi softwaru Reaper s využitím také volně dostupného přídatného programu pro měření spektra s názvem Span od firmy Voxengo.

Také bylo nutné dodržet doporučená pravidla testování pomocí metody MUSHRA, ve které je uvedenou maximální doporučenou délkou jedné stopy 12 sekund. Všechny přiložené stopy mají tedy délky od 4 do 8 sekund.

Z akustických nástrojů bylo vybráno několik druhů nástrojů. Kontrabas pro jeho pilovitý průběh signálu a omezené zaplnění spektra, které se primárně pohybuje na frekvencích nižších než 1 kHz, viz obr. 4.1, a poté jeho účinnost rychle klesá. Pro ukázkou sinusového průběhu signálu byla zařazena ukáзка dechového hudebního nástroje arménského duduku. Ten má díky dvouplátkovému náústku a provedení,

¹<https://freesound.org>



Obr. 4.1: Spektrální analýza vzorku kontrabas.

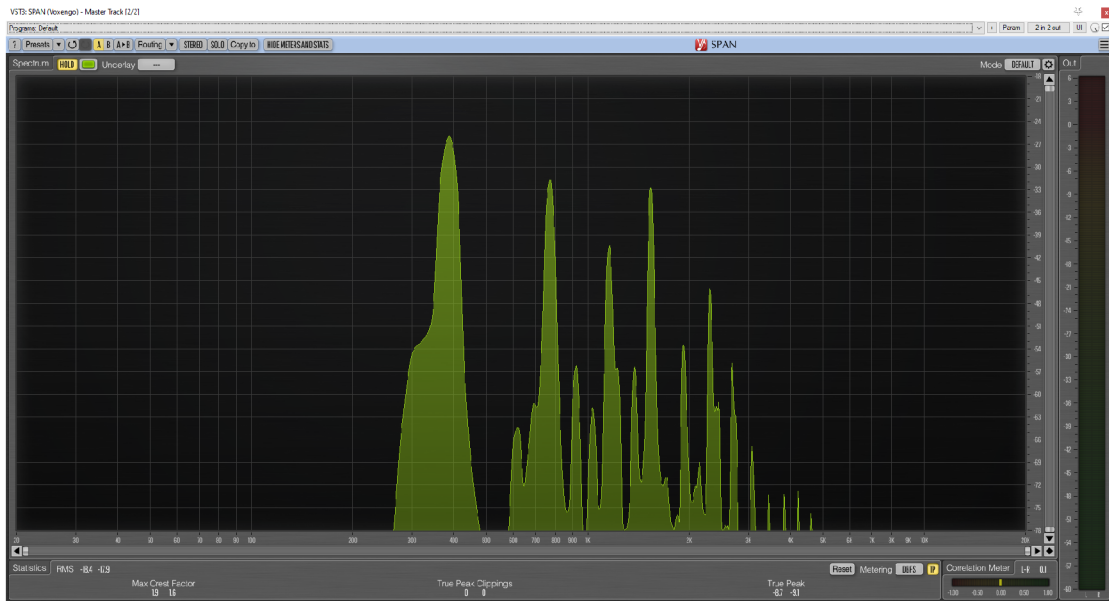
velice jemnou barvu tónu a pohybuje se výhradně ve středních frekvencích, viz obr. 4.2. Z bicích nástrojů byl vybrán zástupce nejvyšších slyšitelných frekvencí spekter, a to zvonkohra, s jejím pronikavým čtvercovým průběhem signálu. Spektrum zvonkohry je vyobrazeno na obr. 4.3.

Mimo nástrojů bylo nutné zahrnout i nenapodobitelnou lidskou řeč, v tomto případě pouze mužského pohlaví. Nutná je i přítomnost kompozic hudebních nástrojů. V tomto případě byla zvolena kompozice s dvěma elektrickými kytarami a kompletní bicí soupravou.

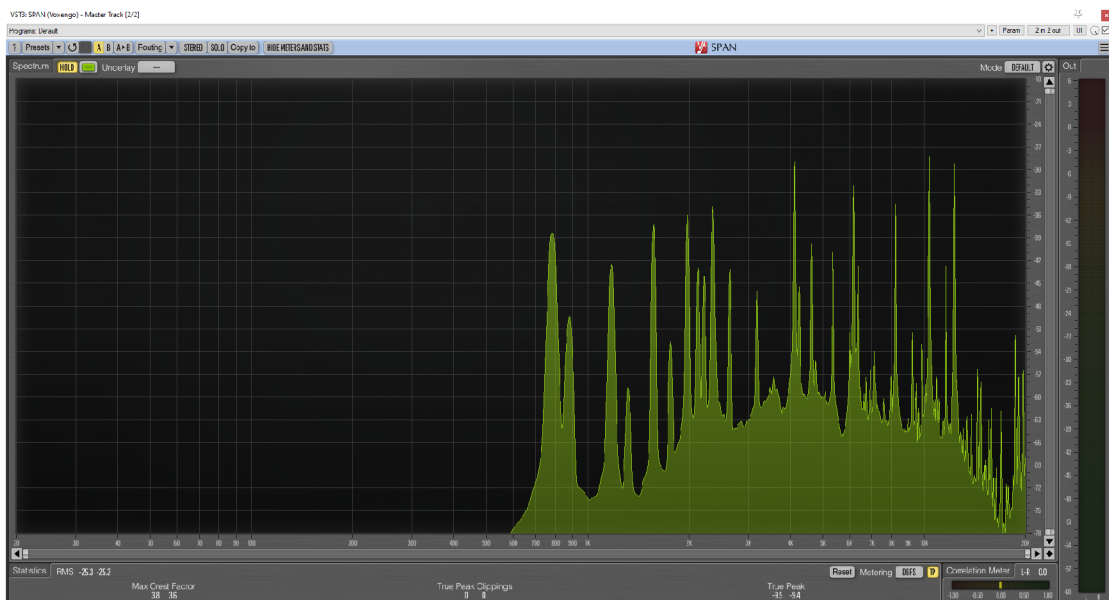
Také výběr formátu zvukového souboru byl velice důležitý. Formát WAV, ve kterém byla databáze vytvořena, je nekompresním formátem, ale pro přenos přes datovou síť má při tak velké databázi obrovských rozměrů. Z tohoto důvodu byl zvolen bezztrátový formát FLAC, který využívá Golombo-Riceova kódování a tím je jeho dekódování výpočetně méně náročné než jiné nekomprimující zvukové kodeky.

4.3 Poškození vzorků databáze

Jako zaměření poslechových testů bylo zvoleno pouze poškození typu Clipping (viz část 2.1). Zúžením testovacího záběru bylo nutné z důvodu celkové délky subjektivního testování, která nesměla přesáhnout stanovený limit dvaceti minut. Při vyšším počtu vzorků se úměrně zvedá délka celého testování, a to má za následek i vyšší únavu sluchového aparátu posluchače a ta má poté za následek nepřesnost celého měření. Všechny ukázky z připravené databáze byly poškozeny v programu Matlab



Obr. 4.2: Spektrální analýza vzorku duduku.



Obr. 4.3: Spektrální analýza vzorku zvonkohry.

pomocí volně dostupného scriptu [47]. Zvolené odstupy zdrojového signálu od zkreslení neboli SDR, byly 3, 5, 10 a 20 dB. Tímto rozdělením bylo dosaženo co největšího pokrytí rozsahu použitých rekonstrukčních metod spolu se zachováním množství vzorků na minimum.

4.4 Rekonstrukce poškozených vzorků

Pro rekonstrukci poškozených signálů bylo použito implementací v Matlabu z práce Pavla Závišky a Pavla Rajmice [47]. Z tohoto balíku metrik byly vybrány tři metody využívající řídkých reprezentací a to: SPADE, l1 Relaxation a Social Sparsity.

4.4.1 SPADE

Rekonstrukční algoritmus SParse Audio DEclipper, zkráceně SPADE, byl vytvořen týmem PANAMA (Parsimony and New Algorithms for Audio & Signal Modeling), který vznikl spojením francouzského Národního výzkumného ústavu v oblasti digitálních věd a technologií Inria a francouzským Národním střediskem pro vědecký výzkum CNRS [48]. Metrika je určena pro rekonstrukci signálu poškozeném clippingem a existuje ve dvou verzích. Verze pracující s modelem analýzy (A-SPADE) a verze s modelem syntézy (S-SPADE), kde obě využívají řídkou regularizaci v časově-frekvenční doméně.

Pro účely této práce byl vybrán algoritmus A-SPADE, který ve výsledcích práce Srđana Kitića překonal pomocí rozdílů SDR metody C-IHT a Social Sparsity. A-SPADE se zařadil na druhé místo po syntézním modelu. Avšak jeho výpočetní a časová náročnost byla několikanásobně menší.

4.4.2 l1 Relaxation

Jak již z názvu vyplývá, algoritmus využívá konvexní vektorové l1 normy k hledání řídkých reprezentací k rekonstrukci clippovaného signálu [49].

V této práci je využit syntetizující Douglas-Rachfordův algoritmus pracující s minimalizací l1 normy.

4.4.3 Social Sparsity

První zmínka o této metodě byla publikována v článku Kaie Siedenburga a jeho týmu, který byl publikován v roce 2014 [50]. Algoritmus je založen na rekonstrukčních metodách využívající řídkých reprezentací. Metrika využívá sociální řídkosti a zvažuje problémy iterativních prahových algoritmů. Způsob jeho provedení umožňuje modelovat závislosti mezi sousedními koeficienty, které sdružuje do skupin,

k čemuž využívá tzv. shrinkage operátory. Těmito operátory jsou: *Lasso* (L), *Windowed Group-Lasso* (WGL), *Empirical Weiner* (EW) a *Persistent Empirical Weiner* (PEW). Využití těchto operátorů výrazně zlepšuje kvalitu rekonstrukce při porovnání s jednoduchým měkkým nastavení prahové úrovně. Autoři použili k vylepšení algoritmu místo báze Gaborův frame založený na Hannově posuvném okně, který měl délky 1024 vzorků. Jako optimalizační algoritmus byl použit relaxovaný (F)ISTA algoritmus.

V této práci je zahrnut pouze operátor PEW, který v publikaci dosáhl nejlepších výsledků ze všech ostatních srovnávaných operátorů. Nastavení prahové úrovně bylo zvoleno $\lambda = 0,0005$, čímž se výpočet algoritmu značně prodloužil.

4.5 Objektívni hodnocení kvality vzorků

Ze všech výše zmíněných metrik pro objektívni hodnocení kvality zvukových signálů byly vybrány tři metody využívající stupnici MOS, která je popsána v části 3.1.2. Tato skutečnost spojuje metriky pro jejich porovnání a vyhodnocení výsledků. Do výběru byly zahrnuty metody PEAQ (viz část 3.2.2), PEMO-Q (viz část 3.2.4) a ViSQOL (viz část 3.2.5). Metriky PEAQ a PEMO-Q byly vyhodnocovány implementacemi v softwaru Matlab. U metody ViSQOL byla použita novější a přesnější verze scriptu v jazyku C++. Tato metrika využívá stupnice MOS-LQS, tudíž bylo nutné výsledky převést na rozsah ODG, který využívají ostatní dvě metody.

Kompletní výpis výsledků je vypsán v příloze A.1.

5 Provedení poslechových testů

Z důvodu aktuálního epidemiologického stavu bylo testování provedeno dálkově přes webový interface. Dle návrhu ze semestrální práce mělo být poslechové testování provedeno skrze online formuláře firmy Google s připojenou databází zvuků. Díky kompresi audio signálu při konverzi dat u jakéhokoli dostupného poskytovatele online přehratelného úložiště, byla pro řešení této práce zvolena doména s vlastním hosting serverem. Při využití online testování bude mít tedy posluchač možnost vyplnit dotazník z jakéhokoli zařízení. Rozdílnost využitých zařízení pro poslech vzorků by mohla být, při běžném druhu testování, brána jako nevýhoda. V tomto případě tato rozdílnost zajistí všestrannou využitelnost metrik napříč rozhraními.

5.1 Server

Za účelem provedení poslechového testování skrz webové rozhraní byl zakoupen rackový server firmy Fujitsu s produktovým názvem PRIMERGY RX200 S7, disponující dvěma procesorovými sloty. V těchto slotech běželi dva šesti jádrové procesory Xeon E5-2630L od firmy Intel s frekvencí 2 GHz. Pro potřeby virtualizace byl alokován celý jejich výpočetní výkon spolu s 4 GB operační paměti a 16 GB prostorem na úložném disku. Jako ovládací systém, neboli hypervisor, serveru bylo zvoleno třetí vydání podnikového hypervisoru ESXi verze 6.0.0 od společnosti VMware. Tento software je poslední podporovanou verzí daným serverem a po několika testovacích provezech se ukázal jako jediný stabilní. Na tomto hypervisoru je možné virtualizovat více operačních systémů, které nejsou vzájemně propojené a tento ovladač jim přiděluje výpočetní výkon.

5.1.1 Operační systém

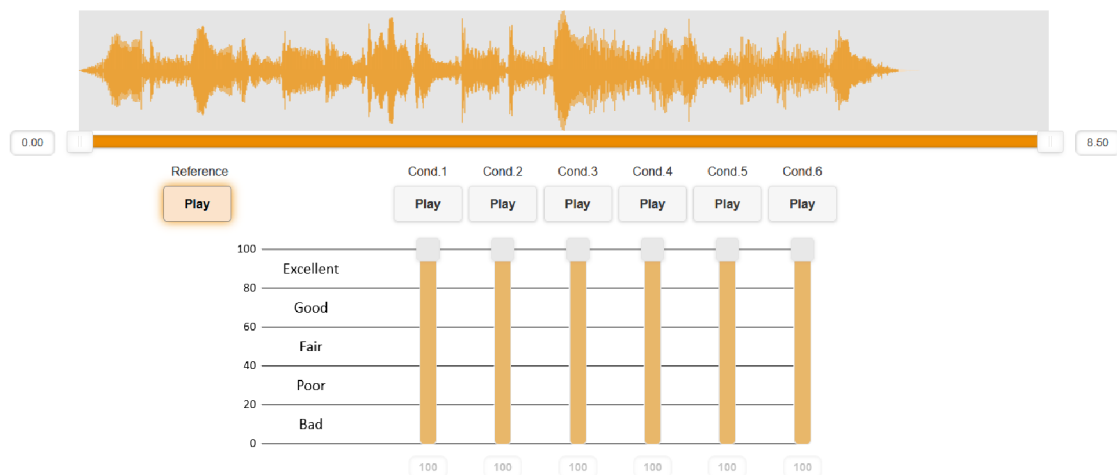
Pro tuto práci byl zvolen svobodný operační systém Linux, jelikož izolační funkce, které Docker využívá, jsou přímo implementované v jeho jádře. Nejlepší vhodný kandidát pro účely práce byl velmi oblíbený systém Ubuntu, který je volně dostupný a šiřitelný. Jeho základem je distribuce Debian. Jedná se o nejdéle vyvíjenou distribuci Linuxu a to od roku 1993. Zvolenou verzí systému Ubuntu byla 20.04 LTS, kde zkratka LTS znamená „Long Term Support“ neboli dlouhodobá podpora systému. Tato verze poskytla plynulý chod celého testování.

5.1.2 Web Mushra

Pro provedení poslechových testů bylo zvoleno volně dostupné webové rozhraní s názvem webMUSHRA z mezinárodních laboratoří Audio Labs [51]. Software, napsaný

v objektově orientovaném jazyce JavaScript, byl vytvořen dle normy ITU-R BS.1534 [21] a je uživatelsky přívětivý. Aplikaci je nutné spouštět pomocí rozhraní Docker, který izoluje aplikace do uzavřených kontejnerů, takzvaná „odlehčená virtualizace“. Kontejner obsahuje jen ty soubory, které nainstalovaná aplikace vyžaduje. Izolací docílí menší provozní náročnosti a své celkové velikosti. Jeho jedinou nevýhodou je absence operačního systému. Musí být tudíž spuštěn jako aplikace v již nainstalovaném systému. Nastavení a rozvržení testu je řešeno soubory typu YAML, což je formát vytvořený pro serializaci strukturovaných dat, je čitelný jak strojem, tak i člověkem a využívá neomezeného vnořování pomocí hierarchie předsezání. Pro editaci souboru byl vybrán open-source textový editor Brackets vytvořený společností Adobe Systems.

Na obr. 5.1 lze vidět rozdělení uživatelského rozhraní splňující požadavky normy [21]. Je zde tlačítko pro přehrání referenčního signálu a škála hodnocení s posuvníkem mezi hodnotami 0 až 100 pro každý stimul.



Obr. 5.1: Ukázka uživatelského rozhraní aplikace Web Mushra.

Díky obsáhlé databázi a zařazení tří rekonstrukčních algoritmů bylo nutné rozdělit testované vzorky po částech splňující požadavky standardu. Ten dle definice musí obsahovat maximálně 12 testovaných vzorků, ve kterých je nutné obsáhnout kotvu a skrytý referenční signál. U každého zvukové ukázky byly vytvořeny čtyři stupně poškození, které byl následně rekonstruovány třemi rekonstrukčními metodami. Spolu s kotevním signálem by tudíž testování přesáhlo maximální počet stimulů a tím degradovalo celý průběh hodnocení kvality signálů.

Dělení úrovní poškození bylo zvoleno v hodnotách odstupů signálu od zkreslení 3, 5, 10 a 20 dB. Na každé stránce testu byly celkově vyobrazeny tři testované stimuly rekonstrukcí v dané úrovni spolu se skrytou referencí, kotevním signálem, který byl zvolen signál s nejhorší úrovní poškození, a druhou kotvou s aktuální úrovní SDR.

Celý test obsahoval úvodní obrazovku s návodem k vyplnění poslechového formuláře, 20 náhodně poskládaných stránek se 6 vzorky a refrenčním signálem a závěrečnou stránku s dotazníkem ohledně posluchače. Dotazník obsahoval kolonku pro vyplnění poslechového zařízení, které ověřovalo kvalitu reprodukce, věk respondenta a jeho pohlaví. Bez těchto údajů nebylo možné dotazník dokončit.

5.1.3 Databázový systém

Výstup tohoto rozhraní je v základu řešen pouze exportem dat do excelové tabulky. To pro tuto práci není nejvhodnější, jelikož se jedná 84 zvukových vzorků z nichž se část opakuje na každé stránce. Při vynásobení zápisů počtem respondentů jsou data nepřehledná a jejich následný přepis by byl velice zdlouhavý. Řešením tohoto problému je vytvoření MySQL databáze s pracovním názvem Mushwrite.

Volně šiřitelný systém MySQL uplatňuje relační databázový systém. Jedná se o nejrozšířenější způsob ukládání dat. Relace neboli tabulky, jsou vzájemně propojeny klíčem. Pro spojení hlavní tabulky s vedlejší, slouží primární a cizí klíče. Primární klíč musí být jedinečný a musí mít nenulovou hodnotu. Tento klíč slouží k identifikaci tabulky, ke které jsou napojené vedlejší tabulky, v nichž využijí cizího klíče. Tento archivační systém spolupracuje se skriptovacím programovacím jazykem PHP, jehož bylo využito ve verzi 7.4.12.

Tato databáze obsahuje tři vzájemně propojené tabulky, viz obr. 5.2, ve kterých se následně jednoduše vyhledává pomocí PHP příkazů. První z tabulek nese název *Test*, která obsahuje primární klíč, uložená data ohledně respondenta a název poslechového testu. Na tuto tabulku je relačně napojena skrz cizí klíč tabulka *Trial*. Do ní je uloženo číslo stránky a její název. Třetí a poslední relace je spojena s předchozí naprosto totožně a nese název *Response*, v níž jsou umístěny názvy stimulů, jeho hodnota a čas vyplnění této hodnoty.



Obr. 5.2: Blokové schéma propojení relací databáze.

6 Zpracování výsledných hodnot

V této kapitole jsou zahrnuty provedené statistické analýzy, vizualizace nasbíraných a vyhodnocených dat a následné porovnání výsledků poslechových testů s hodnotami objektivních metrik.

6.1 Analýza dat z poslechových testů

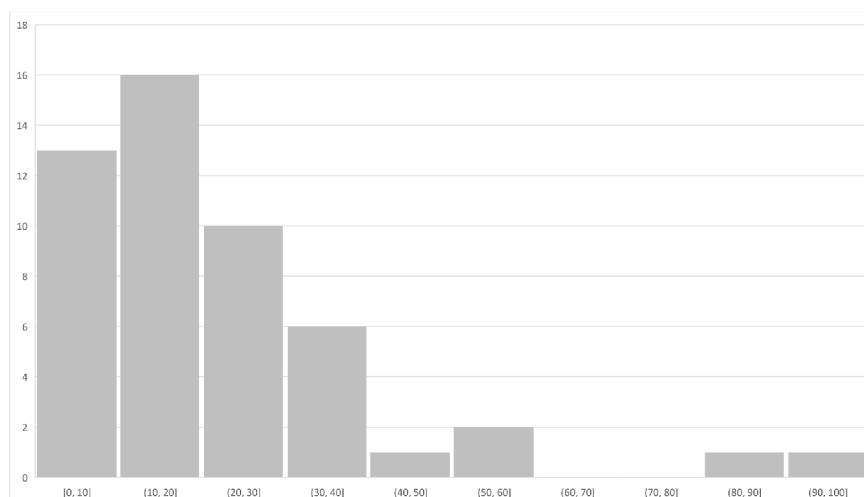
Testování bylo provedeno na skupině o velikosti 50 respondentů. Tato skupina byla primárně obsazena posluchači pracujícími se zvukem, jako například studenty audio inženýrství, zvukaři, střihači, producenty, umělci a zpěváky. Uskupení by tedy mělo být znalé poškození a více kritické ke kvalitě výsledného zvuku. Níže práce pojednává o vizualizaci dat, vyřazení respondentů nesplňující podmínky pro následné vyhodnocení a provedení statistické analýzy rozptylu.

6.1.1 Vizualizace a vyloučení odlišných dat

Základním kamenem celé statistické analýzy je vizualizace naměřených dat. Vizualizace poskytne indikaci šířky a odchylky hodnot. Dle normy [21] je doporučeno využít histogramů, Q-Q grafů či krabicových diagramů.

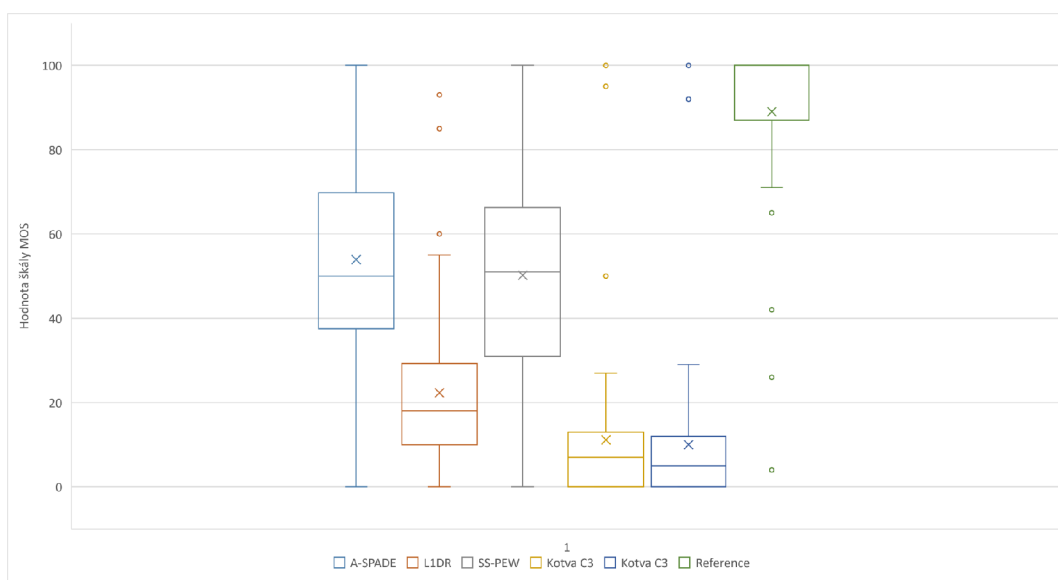
Pomocí vizualizace výsledků v krabicových grafech a histogramech bylo nalezeno dle odlehklých bodů celkem 6 respondentů, kteří dle výsledků nepochopili zadání, nebo nebyli naprosto koncentrovaní. Tuto domněnku lze vyvodit z nevyužití celé hodnotící škály a tím pádem se jejich zadané hodnoty koncentrovaly v malém rozsahu. Při správném rozložení by výsledky měly být rozloženy od 0, reprezentující nejvíce poškozenou a snadno identifikovatelnou kotvu, a 100, kde by při očekávaném výsledku měla být skrytá reference. Dále byly vyřazeny výsledky dvou respondentů, kteří vybočovali z rozptylu ostatních naměřených hodnot. Všechny tyto hodnoty byly v dalším zpracování výsledků vyřazeny.

Pro obsáhlost naměřených hodnot byla vybrána ukázka vzorku elektrické kytary poškozené na úroveň SDR 3 dB a následně rekonstruované algoritmem l1 relaxace (viz 4.4.2), na kterém je celý postup vysvětlen. Na histogramu 6.1 lze vidět náklon složek k levé straně rozsahu, kde mají kumulární vrchol mezi body 10 až 20 a inklinují tak k normálnímu rozdělení. Pouze na hodnotou 50 je vidět nárůst těchto hodnot u kterých by dle Gaussovy funkce byl očekáván další pokles. Největším problémem jsou však na histogramu odlehklé hodnoty, které jsou vyobrazeny nad hodnotami vyššími než 80. Tyto dva výsledky enormně vybočují z normálu ostatních hodnot.



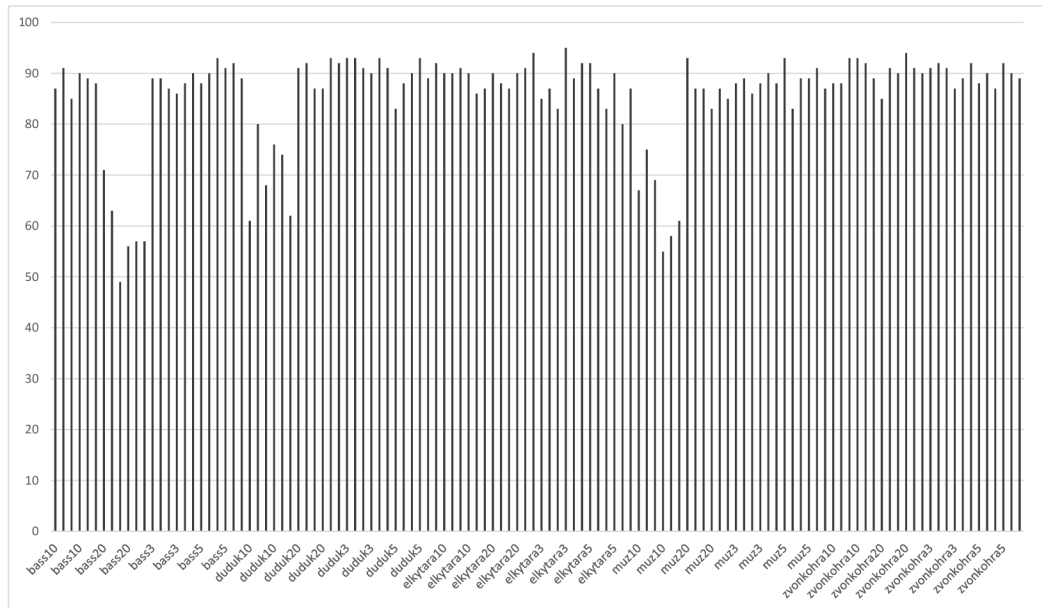
Obr. 6.1: Histogram vybraného vzorku.

Dále bylo nutné data zobrazit v krabicovém grafu 6.2 společně s dalšími rekonstrukcemi, kotvami a skrytou referencí pro celkový náhled na výsledky dané úrovně poškození, které respondenti srovnávaly. V tomto typu grafu je vizualizace dat pomocí jejich kvartilů. Diagram obsahuje minimum a maximum všech dat, horním a dolním kvartilem, mediánem a střední hodnotou označenou křížkem. Od té se odvíjejí standartní odchylky. Body mimo tento boxplot se nazývají odlehlé hodnoty, neboli outliery. V tomto případě nabývají hodnot 60, 85 a 93.



Obr. 6.2: Krabicový graf vybraného vzorku.

Dle zjištění exaktních čísel je možné tyto hodnoty vyfiltrovat v databázi výsledků a dohledat tak, číslo respondenta a jeho celý poslechový test. Při vyhledání výsledku s hodnotou 85 v daném testovacím vzorku byla nalezena shoda s respondentem číslo 23. V grafu 6.3 jsou znázorněny kompletní zadané hodnoty respondentem, ve kterých je viditelné, že se respondent nedržel zadání a nevyužil celého spektra stupnice. Tudíž je nutné celý tento test vyřadit z dalšího zpracování hodnot.



Obr. 6.3: Všechny zadané hodnoty respondentem číslo 23.

6.1.2 Analýza rozptylu

Ke statistické analýze dat byla dle doporučení normy [21], vybrána metoda analýzy rozptylu, neboli ANOVA. Tato metoda matematické statistiky umožňuje určit vliv faktorů na výsledná data. Nejlepším typem metody, pro účely práce, je analýza rozptylu s opakovanými měřeními (*Repeated Measures ANOVA*), která nachází využití při opakujících se měřeních s vícero faktory, či skupinami [52]. Pro její výpočetní náročnost je doporučeno využívat statistických softwarů.

Před započítáním analýzy je nutné stanovit hypotézy, které budou následně vypovídat o zjištěných vztazích mezi daty. Provedením statistické analýzy a zjištěním jejích výsledných hodnot, se zadané hypotézy zamítnou, či potvrdí.

V případě této práce se jedná o čtyři nulové hypotézy:

1. Vnímaná kvalita zvuku testovaných vzorků se mezi sebou nelišila. Myšleno mezi rekonstruovanými signály, kotvami a skrytou referencí.

2. Při měření nezáviselo na respondentovi a neovlivnilo to jeho výsledné hodnoty.
3. Naměřené hodnoty neovlivnil typ základního zvukové signálu.
4. Zvolené odstupy vstupního signálu od zkreslení neměly vliv na poslechové testy.

Pro provedení statistické analýzy byl zvolen software Matlab, který po přidání rozšiřovací sady statistik a strojového učení (*Statistics and Machine Learning Toolbox*) zahrnuje oficiální funkci pro realizaci několika typů analýzy rozptylu. Typ s opakovanými měřeními, využitý v této práci, se v použitém softwaru nazývá **ranova**.

Tato funkce pracuje s importovanou tabulkou v libovolném formátu a vytvořeným modelem opakovaných měření. V matici dat je nutné dodržet pravidel formátování naměřených dat pro analýzu rozptylu. Tabulka musí obsahovat všechny nezávislé proměnné, pro které je nezbytné zjistit jejich vzájemnou ovlivnitelnost. Na prvním sloupci tabulky jsou čísla respondentů, tak jak jim byly přiděleny po vyřazení nepoužitých dat. Tudiž hodnoty 1 až 42. V druhé sloupci jsou čísla dělení zvukových vzorků, dle typu zdrojového signálu. Tento sloupec nabývá hodnot od 1 do 5. Ve třetím sloupci tabulky je vypsáno číselné rozdělení stupně odstavu signálu od zkreslení. Hodnotám byly přiřazeny čísla 1 až 4. V dalších sloupcích jsou již výsledná data, kde první tři z nich jsou hodnoty rekonstruovaných vzorků (A-SPADE, 11 DR, SS PEW), kotva společná pro celý typ signálu, kotva s danou úrovní testované úrovně poškození a skrytý referenční vzorek. Celá tabulka je díky její obsáhlosti přiložena na CD.

Po importování tabulky, přejmenování názvů testů a vytvoření modelu opakovaných měření bylo přistoupeno k realizaci analýzy, na jejímž konci je vypsána tabulka výsledných hodnot, viz tab. 6.1.

V této tabulce jsou vypsány všechny důležité hodnoty pro následující statistické kroky. Nejdůležitější z nich je hodnota p . Tato hodnota se v porovnání meziskupinovým efektem α , určuje, zda se hypotéza zamítne, či ne. V normě [21] je doporučeno zvolit vliv efektu $\alpha = 0,05$. Obecně platí, že pokud je výsledkem hodnota p menší než zvolená hladina α , daná hypotéza se zamítá.

Z tabulky lze vyčíst, že se výsledné hodnoty významně liší a není mezi nimi interakce. První hypotéza se tedy zamítá. Druhou hypotézu nelze zamítnout, jelikož její hodnota převyšuje stanovenou hladinu. Záleželo tedy na respondentovi a jeho vnímání. Třetí a čtvrtá nulová hypotézu se zamítá. Typ zvuku a jeho hodnota SDR měl významný vliv na výsledné hodnoty. Další řádky tabulky ukazují interakci vícero faktorů mezi sebou.

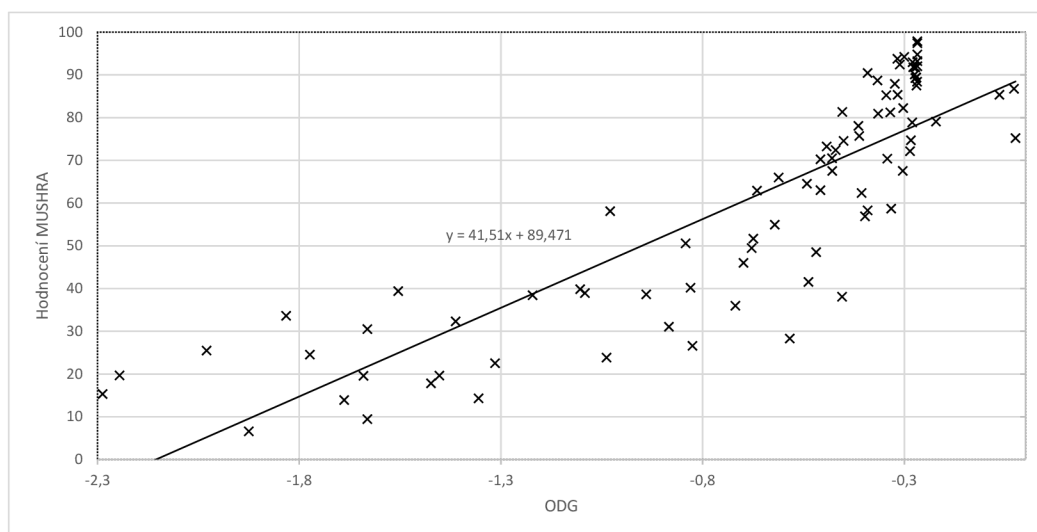
Tab. 6.1: Tabulka výsledných hodnot analýzy rmANOVA.

	SS	Rozdíl	MS	F	<i>p</i> Value	<i>p</i> Value GG	<i>p</i> Value HF	<i>p</i> Value LB
Absolutní člen (Hodnot)	40838,6	5	8167,7	37,225	<0,001	<0,001	<0,001	<0,001
Respondent:Hodnoty	576,1	5	115,2	0,525	0,757	0,712	0,713	0,469
Typ:Hodnoty	5549,7	5	1109,9	5,058	<0,001	<0,001	<0,001	0,025
SDR:Hodnoty	21960,9	5	4392,2	20,018	<0,001	<0,001	<0,001	< 0,001
Respondent:Typ:Hodnoty	1059,4	5	211,9	0,966	0,437	0,423	0,424	0,326
Respondent:SDR:Hodnoty	646,9	5	129,4	0,590	0,708	0,665	0,666	0,443
Typ:SDR:Hodnoty	790,1	5	158,0	0,720	0,608	0,574	0,575	0,396
Respondent:Typ:SDR:Hodnoty	803,1	5	160,6	0,732	0,599	0,566	0,567	0,392
Chyba(Hodnot)	912758,9	4160	219,4					

6.2 Vyhodnocení porovnání výsledků

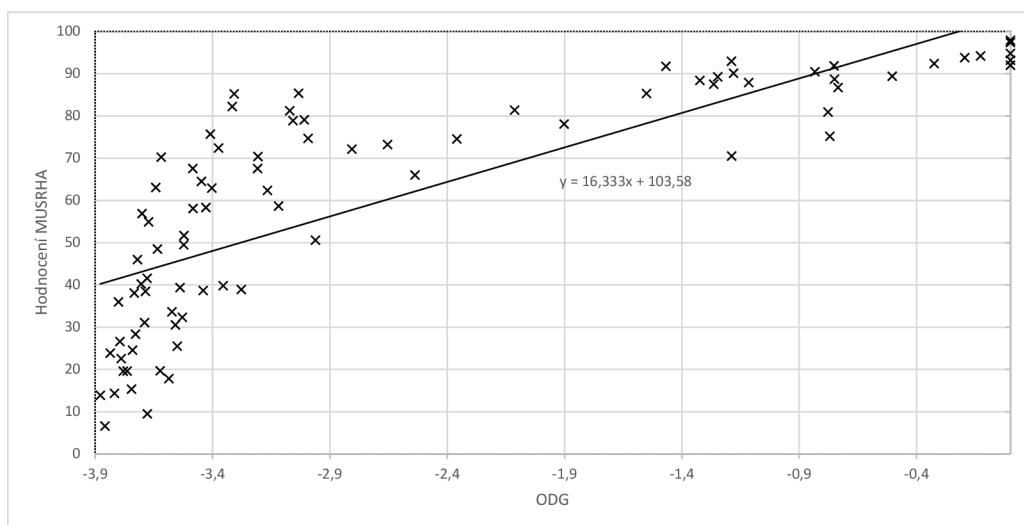
Spojením výsledků předchozí analýzy s daty objektivních metrik dostáváme konečné výsledky práce. Pro porovnání těchto výsledků bylo nutné vypočítat střední hodnoty všech testovaných vzorků v poslechových testech. Do druhého sloupce tabulky byly vypsány kompletní výsledky každé objektivní metody pro hodnocení kvality signálu zvlášť. Pomocí regresní a korelační analýzy je možné vyvodit, která z metod více odpovídá výsledkům poslechových testů.

V grafech lineární regresní funkce lze vidět porovnání hodnot objektivní metriky na škále ODG, která tvoří osu x a hodnocení na škále metody MUSHRA v ose y . Vedle lineární spojnice trendu je i vypsána rovnice grafu. Z těchto grafů lze vyčíst, že nejrovnoměrnější rozložení bodů má metrika ViSQOL, viz obr. 6.4. Oproti tomu metody PEAQ a PEMO-Q měly tendenci hodnotit vícero vzorků hodnocením -4 až $-2,5$. V tomto rozsahu se nachází nejvíce bodů grafu, z něhož lze odvodit využití metod pro jemnější rozdíly kvality od reference. Metrika PEAQ ohodnotila některé referenční signály v rozmezí od 0,191 po 0,209. viz obr. 6.6, což je překvapivý výsledek, který je mimo rozsah vlastní stupnice ODG. Očekávaná hodnota by se měla pohybovat okolo 0, tak jako je tomu u metody PEMO-Q, viz obr. 6.5.

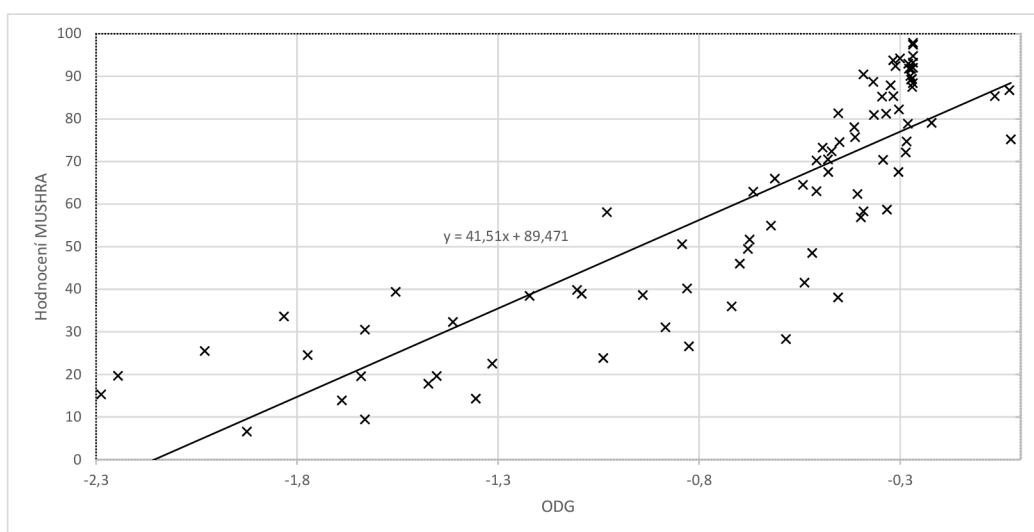


Obr. 6.4: Graf lineární regresní funkce metody ViSQOL.

U korelační analýzy byl vypočten korelační koeficient pro každou metodu zvlášť a zapsán do tab. 6.2. V této tabulce se potvrzují předešlé zjištění spojené s metodou ViSQOL. Její korelační koeficient dosahuje nejvyšší hodnoty. V tomto měření dopadla nejhůře metoda PEAQ.



Obr. 6.5: Graf lineární regresní funkce metody PEMO-Q.



Obr. 6.6: Graf lineární regresní funkce metody PEAQ.

Tab. 6.2: Tabulka korelačních koeficientů objektivních metrik.

Objektivní metoda	Korelační koeficient
PEMO-Q	0,792
Visqol	0,859
PEAQ	0,754

Závěr

V závěru práce je nutné shrnout co bylo v práci provedeno a zjištěno. Na začátku bylo nutné teoreticky popsat vnímání zvuku sluchovým aparátem spolu s jeho psychoakustickými vjemy. Jak lidské ucho reaguje a jakým způsobem člověk interpretuje maskování dvou zvuků, či jeho hlasitost. Po této krátké pasáži bylo možné přistoupit ke dvěma modelům, které berou tyto jevy v potaz a umožňují tak modelovat psychoakustiku sluchového ústrojí. V další části byly popsány tři základní druhy poškození audio signálů spolu s jejich odnožemi. Velký důraz byl vložen do poškození clipping, na jehož rekonstrukci se později celá práce přesunula. Také byly zmíněny druhy šumů spolu se způsoby jejich vzniku a poškození chybějícími vzorky.

Pro následný možný výběr správné subjektivní metriky hodnocení kvality signálu pro poslechové testy, bylo teoreticky popsáno šest subjektivních metod. Z nichž byla, dle jejich znalosti, vybrána jedna, a to s názvem MUSHRA. Tato metoda se díky její nejmodernější verzi spolu s širokou škálou pro přesnost hodnocení, jevila jako nejlepší možnost. Ke srovnání bylo nutné znát i objektivní stranu hodnocení kvality zvukových signálů. Z šesti popsaných metrik byly vybrány ke srovnání tři z nich a to PEAQ, ViSQOL a PEMO-Q. Všechny vybrané metody mají implementaci v programu Matlab a využívají shodnou hodnotící škálu MOS. U algoritmu ViSQOL byl nakonec využita novější verze implementována v jazyku C++.

Po nutném teoretickém úvodu všech potřebných informací bylo možné začít budovat zvukovou databázi a začít vymýšlet poslechové testování. Finální databáze byla omezena pouze na pět různorodých vzorků, tak aby testování nebylo časově náročné a nepřekročilo 20 minut, po kterých účastník ztrácí soustředěnost a jeho sluchový aparát se začíná otupovat. Tento jev by mohl negativně ovlivnit výsledky testů. Frekvenčně pokryly testované zvukové nahrávky celé slyšitelné spektrum a byl kladen důraz na jejich možnost použití s ohledem na autora a licenční společnosti. Vzorky byly poškozeny předem vybraným typem poškození. Zvolené úrovně odstupů signálu od zkreslení byly 3, 5, 10 a 20 dB. Následně byly zvuky rekonstruovány třemi rekonstrukčními algoritmy dodané v Matlab implementacích od vedoucího práce pana inženýra Závišky. Jednalo se o metody SPADE, 11 Relaxace a metoda založená na sociální řídkosti. Celá databáze čítala na 85 zvukových vzorků.

Následně bylo nutné zrealizovat poslechové testy. Z důvodu aktuálního epidemiologického stavu byla zvolena možnost testování skrz webový interface. Pro nepřetržité sledování průběhu a možnosti vyzkoušení si vlastní serverové služby byl zakoupen průmyslový server, na kterém byl virtualizován operační systém s volně dostupným webovým rozhraním s názvem webMUSHRA. Software poskytl předem vytvořenou strukturu testování dodržující normu subjektivní metriky MUSHRA. Jelikož byla zvuková databáze vytvořená ve formátu WAV datově obsáhlá, byla

převedena do bezztrátového formátu FLAC. Tento úkon umožnil načtení webové stránky uživatelem několikanásobně rychleji, díky menšímu objemu dat, který bylo nutné přenést. Testovací software byl napojen na databázový systém MySQL, pro jednodušší třídění výsledků a zajistil případnou zálohu dat při neočekávaném výpadku.

Po nasbírání výsledků od 50 respondentů z řad studentů oboru Audio Inženýrství, profesionální střihačů a hudebních nadšenců, bylo možné přistoupit k validaci a statistické analýze dat. Vizualizace byla provedena pomocí histogramů a následně ověřena na krabicových grafech. Všechny data jsou zařazena na přiloženém CD. Při zjištění nekonvenčních výsledků, jako jsou odlehlé body, nepochopení zadání uživatelem, či případná nesoustředěnost, byly výsledky daného respondenta vyloučeny z dalšího postupu analýzy dat. Celkem bylo vyloučeno 8 respondentů. K provedení statistické analýzy byla dle doporučení normy využita metoda analýzy rozptylu s opakovanými měřeními s názvem rmANOVA. Před provedením byly stanoveny čtyři nulové hypotézy, z nichž pouze jedna nebyla zamítnuta.

Posledním úkonem vykonaným v této práci bylo porovnání výsledků poslecho- vých testů s vybranými objektivními metodami. K tomuto účelu byla zvolena regresní a korelační analýza. Z grafů lineární regresní funkce vyplynul výsledek, která z metrik nejvíce odpovídala výsledkům testu. Nejlépe vycházející metrikou byla ViSQOL. Ta dosáhla nejrovnoměrnější rozložení bodů. Jediným vytknutelným bodem metody je nepokrytí celého rozsahu hodnotící stupnice. Tudíž by se dalo říci, že je metoda připravena pro mnohem více poškozené signály, než byly použity v této práci. Ostatní dvě metody vyšly velice podobně. Pokryly celé své spektrum, ale měly vyšší tendenci se přiklánět k nižšímu hodnocení kvality signálu. Lze tedy odvodit jejich využití pro méně poškozené signály. Následná korelační analýza potvrdila předešlé tvrzení, jelikož korelační koeficient metriky ViSQOL byl nejbližší hodnotě 1. Na druhém místě se s ohledem na tento koeficient umístila metoda PEMO-Q na posledním PEAQ.

Literatura

- [1] SCHIMMEL, Jiří. *Elektroakustika: BELA* [online]. Brno, 2014 [cit. 2020-12-06]. ISBN 978-80-214-4716- 5. Skripta. Vysoké učení v Brně.
Dostupné z URL:
<<https://www.vutbr.cz/studenti/predmety/detail/169924>>.
- [2] *Psychiatrická encyklopedie*. [online]. sv. 5 s. [1367a]
Dostupné z URL:
<<https://ndk.cz/uuid/uuid:41791a50-e34e-11e8-9984-005056825209>>.
- [3] ZWICKER, E. a H. FASTL. *Psychoacoustics: facts and models* 2nd ed. Berlin: Springer, c1999. ISBN 3-540-65063-6.
- [4] SYROVÝ, Václav. *Hudební akustika* [online]. Praha. Akademie múzických umění, 2008. s. [5]. ISBN 978-80-7331-127-8.
Dostupné z URL:
<<https://ndk.cz/uuid/uuid:456f7a00-aec9-11e3-87a3-001018b5eb5c>>.
- [5] MALÝ, Ladislav,. *Stáří a sluch* [online]. Praha. 2011 Senior internet klub [cit. 2016-03-01].
Dostupné z URL:
<<http://sik.vse.cz/ss/stariasluch.pdf>>.
- [6] MELKA, A *Základy experimentální psychoakustiky*. PRAHA, AMU, 2005.328 s. ISBN 9788071963653
- [7] SAYOOD, Khalid. *Introduction to Data Compression: A volume in The Morgan Kaufmann Series in Multimedia Information and Systems*. 5. USA: Morgan Kaufmann, 2018 [cit. 2020-12-09]. ISBN 9780128094747.
Dostupné z URL:
<<https://doi.org/10.1016/B978-0-12-809474-7.00017-3>>.
- [8] CAMPBELL D., Jones E., Glavin M. *Audio quality assessment techniques—A review, and recent developments*. vol. 89 issue 8 2009 pp. 1489-1500, ISSN: 0165-1684.
- [9] VRBA, Kamil, Norbert HERENCŠÁR a Jaroslav KOTON. *Analogová technika: BANA* [online]. Brno, 2011 [cit. 2020-12-06] Skripta. Vysoké učení v Brně.
Dostupné z URL:
<https://moodle-archiv.ro.vutbr.cz/pluginfile.php/444315/mod_resource/content/0/BANA_skripta_Koton_Vrba_2011.pdf>.

- [10] GARY, Davis a Jones RALPH. *Sound Reinforcement Handbook*. 2Rev Ed. USA: Hal Leonard Corporation, 1989. ISBN 9780881889000.
- [11] MCHUGH, Sean. *Camera Histograms: Tones & Contrast*. Cambridge in Colour. Cambridge: Sean McHugh, 2005 [cit. 2020-12-11].
Dostupné z URL:
<<https://www.cambridgeincolour.com/tutorials/histograms1.htm>>.
- [12] HARVILL, Mark a Richard STERN. *Efficient audio declipping using regularized least squares* [online]. Pittsburgh: IEEE, 2015 [cit. 2020-12-06]. ISBN 978-1-4673-6997-8. International Conference on Acoustics, Speech, and Signal Processing.
Dostupné z URL:
<<https://ieeexplore.ieee.org/document/7177964>>.
- [13] DEFRAENE, Bruno, Toon van WATERSCHOOT, Hans JOACHIM FERREAU, Moritz DIEHL a Marc MOONEN *Efficient audio declipping using regularized least squares* [online]. IEEE, 2012 [cit. 2020-12-06]. ISSN: 1558-7924. IEEE Transactions on Audio, Speech, and Language Processing.
Dostupné z URL:
<<https://ieeexplore.ieee.org/document/6255767>>.
- [14] SCHIMMEL, Jiří. *Studiová a hudební elektronika: BSHE* [online]. Brno, 2015 [cit. 2020-12-06]. ISBN 978-80-214-4716-5. Skripta. Vysoké učení v Brně.
Dostupné z URL:
<<https://www.vutbr.cz/studenti/predmety/detail/161237?apid=161237>>.
- [15] HUANG, Yu-Ting a LAWLESS Harry T. *SENSITIVITY OF THE ABX DIS-CRIMINATION TEST*. [online]. Journal of Sensory Studies, Oxford, UK: Blac-kwell Publishing, 9807, 13(2), 229-239 [cit. 2020-11-11]. DOI: 10.1111/j.1745-459X.1998.tb00085.x. ISSN 0887-8250
Dostupné z URL:
<<http://onlinelibrary.wiley.com.ezproxy.lib.vutbr.cz/doi/10.1111/j.1745-459X.1998.tb00085.x/epdf>>.

- [16] Recommendation BS.1116-1 *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. [online]. International Telecommunications Union Radio Communication Assembly, 1997
Dostupné z URL:
<https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-1-199710-S!!PDF-E.pdf>.
- [17] Recommendation P.800.1 *Mean opinion score (MOS) terminology*. [online]. International Telecommunications Union Radio Communication Assembly, 2016
Dostupné z URL:
<<https://www.itu.int/rec/T-REC-P.800.1>>.
- [18] STREIJL, Robert, WINKLER Stefan a HANDS David *Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. Mul-timedia Systems*. [online]. Berlin/Heidelberg: Springer Berlin Heidelberg, 1603,22(2), 213-227 [cit. 2020-11-02]. DOI: 10.1007/s00530-014-0446-1. ISSN 0942-4962
Dostupné z URL:
<<https://www.researchgate.net/publication/298090737_Mean_opinion_score_MOS_revisited_methods_and_applications_limitations_and_alternatives>>.
- [19] ŠPETA, Marek *Programová implementace subjektivních testů zvukové kvality*. [online]. Brno: VUT v Brně. Fakulta elektrotechniky a komunikačních technologií. Ústav Telekomunikací, 2011. 69 s. Vedoucí diplomové práce Ing. Ondřej Rášo.
Dostupné z URL:
<<https://dspace.vutbr.cz/xmlui/bitstream/handle/11012/6139/xspeta00_dip_2011.pdf>>.
- [20] IRISA *Le site web de I'IRISA* [online]. Francie: Beaulieu University Campus - 263 Avenue du Général Leclerc - CS 74205, 35042 RENNES Cedex, 2020 [cit. 2020-11-24].
Dostupné z URL:
<<<http://www.irisa.fr/armor/lesmembres/Mohamed/Thesis/node145.html>>>.

- [21] Recommendation BS.1534-3 *Method for the subjective assessment of intermediate quality levels of coding systems*. [online]. International Telecommunications Union Radio Communication Assembly, 2015
Dostupné z URL:
<<<https://www.itu.int/rec/R-REC-BS.1534-3-201510-I/en>>.
- [22] Recommendation BS.1284-2 *General methods for the subjective assessment of sound quality*. [online]. International Telecommunications Union Radio Communication Assembly, 2019
Dostupné z URL:
<<<https://www.itu.int/rec/R-REC-BS.1284/en>>.
- [23] JOHNSON DON H. *Signal-to-noise ratio*. [online]. Scholarpedia, 2006, 1.12: 2088
Dostupné z URL:
<www.scholarpedia.org/article/Signal-to-noise_ratio>.
- [24] ADLER A., EMIYA V., JAFARI M. *Audio Inpainting*. *IEEE Transactions on Audio, Speech, and Language Processing*. ročník 20, č. 3, March 2012: s.922–932, ISSN 1558-7916, doi:10.1109/TASL.2011.2168211.
- [25] *Method for objective measurements of perceived audio quality*. Recommendation ITU-R BS.1387-1 (PEAQ). , 2001.
- [26] KABAL P. *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*. [online]. Department of Electrical & Computer Engineering, McGill University, 87 s., 2002
Dostupné z URL:
<<http://www.tsp.ece.mcgill.ca>>.
- [27] XIAOPENG H., GUIMING H., XIAOPING Z. *PEAQ-based psychoacoustic-model for perceptual audio coder*. 8th International Conference Advanced Communication Technology, ICACT 2006, Proceedings. 3. 5 pp. - 1823. 10.1109/ICACT.2006.206344.
- [28] WATANABE K. *Objective perceptual audio quality measurement methods*. Broadcast Technology no. 35, NHK STRL.
- [29] *Methods for subjective determination of transmission quality*. Recommendation ITU-T P.800. , 1996.
- [30] HUBER R., KOLLMEIER B. *PEMO-Q-A new method for objective audio quality assessment using a model of auditory perception*. *IEEE Trans. Audio Speech Lang. Process.* 14 (6) (2006) 1902–1911.

- [31] BARBEDO J., LOPES A. *A new cognitive model for objective assessment of audio quality*. J. Audio Eng. Soc. 53 (1/2) (2005) 22–31.
- [32] *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Recommendation ITU-T P.862. , 2001.
- [33] RIX A. W., BEERENDS J. G., HOLLIER M. P., HEKSTRA A. P. *Perceptual evaluation of Speech Quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs*. IEEE Signal Processing Society International Conference on Acoustics, Speech, and Signal Processing (ICASSP), May 2001.
- [34] *Mapping function for transforming P.862 raw result scores to MOS-LQO*. Recommendation ITU-T P.862.1 , 2003.
- [35] RAINER, Huber a Birger KOLLMEIER. *PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception*[online] IEEE Transactions on Audio, Speech, and Language Processing [online]. 14(6), 1902-1911 [cit. 2020-12-09]
Dostupné z URL:
<<https://ieeexplore.ieee.org/document/170988>>
- [36] JEPSEN, Morten Løve, EWERT Stephan D. a DAU, Torsten. *A computational model of human auditory signal processing and perception*[online] Journal of the Acoustical Society of America. 2008 ; Vol. 124, No. 1. pp. 422-438.
Dostupné z URL:
<<https://doi.org/10.1121/1.2924135>>
- [37] GLASBERG, Brian a Brian C.J MOORE. *Derivation of auditory filter shapes from notched-noise data: Hearing Research* [online] 47. 1990 [cit. 2020-12-09]. ISSN ISSN 0378-5955.
Dostupné z URL:
<[https://doi.org/10.1016/0378-5955\(90\)90170-T](https://doi.org/10.1016/0378-5955(90)90170-T)>
- [38] JAN, Novák. *Implementace metody objektivního hodnocení kvality zvuku*[online] Praha, 2018 [cit. 2020-12-09]. Bakalářská práce. České vysoké učení technické v Praze. Vypočetní a informační centrum. Vedoucí práce Rund František.
Dostupné z URL:
<<http://hdl.handle.net/10467/73804>>

- [39] HINES, Andrew, SKOGLUND Jan, KOKARAM Anil, HARTE N. *ViSQOL: The virtual speech quality objective listener*. [online] 2012, International Workshop on Acoustic Signal Enhancement (IWAENC), pp. 1–4.
Dostupné z URL:
<<https://storage.googleapis.com/pub-tools-public-publication-data/pdf/39979.pdf>>
- [40] HINES, Andrew, SKOGLUND Jan, KOKARAM Anil, HARTE Naomi *VISQOL: an objective speech quality model*. [online] EURASIP Journal on Audio, Speech, and Music Processing, 2015, 2015:13
Dostupné z URL:
<<https://asmp-urasipjournals.springeropen.com/track/pdf/10.1186/s13636-015-0054-9>>
- [41] HINES Andrew, GILLEN Eion, KELLY Damien, SKOGLUND Jan, KOKARAM Anil, HARTE Naomi *ViSQOLAudio: An objective audio quality metric for low bitrate codecs* [online] The Journal of the Acoustical Society of America 137, EL449 (2015)
Dostupné z URL:
<<https://asa.scitation.org/doi/10.1121/1.4921674>>.
- [42] ViSQOLAudio: Speech Quality Metric *Sigmedia* [online] Dublin: Signal Processing and Media Applications Group, 2020 [cit. 2020-11-21]
Dostupné z URL:
<<http://www.sigmedia.tv/tools>>.
- [43] TAN C., MOORE CJ B., ZACHAROV N. *Predicting the perceived quality of nonlinearly distorted music and speech signals*. [online] Journal of the Audio Engineering Society. Audio Engineering Society 52(7):699-711, July 2004
Dostupné z URL:
<https://www.researchgate.net/publication/261613967_Predicting_the_perceived_quality_of_nonlinearly_distorted_music_and_speech_signals>.
- [44] GLASBERG B.R., MOORE CJ B. *A model of loudness applicable to time-varying sounds*. [online] Journal of the Audio Engineering Society, 50:331–342, May 2002.
Dostupné z URL:
<https://www.researchgate.net/publication/266395662_A_model_of_loudness_applicable_to_Time-Varying_Sounds>.

- [45] PATTERSON R. D., ALLERHAND M. H. *Time-domain modeling of the peripheral auditory processing: A modular architecture and a software platform*. [online] Journal of the Acoustical Society of America, 98:1890–1894, 1995.
Dostupné z URL:
<<https://doi.org/10.1121/1.414456>>.
- [46] MARTINEZ J., CATALA J. C., FRANCES J., SORIANO J. R. *Comparison of different methods for the subjective sound quality evaluation of compression drivers*. [online] 2009
Dostupné z URL:
<<https://www.researchgate.net/publication/267552440>>.
- [47] ZÁVIŠKA P., RAJMIC P., OZEROV A., RENCKER L. *A Survey and an Extensive Evaluation of Popular Audio Declipping Methods* IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 1, pp. 5–24, Jan. 2021, doi: 10.1109/JSTSP.2020.3042071.
- [48] KITIĆ S., BERTIN N., GRINBONVAL R. *Sparsity and Cosparsity for Audio Declipping: A Flexible Non-convex Approach* [online] LVA/ICA 2015, LNCS 9237s. 243–250, 2015
Dostupné z URL:
<<https://arxiv.org/pdf/1506.01830v2.pdf>>.
- [49] HRBÁČEK R., RAJMIC P., VESELÝ V., ŠPIŘÍK J. *Řídké reprezentace signálů: úvod do problematiky* [online] Elektorevue. Brno(CZ): FEKT VUT Brno, 2011, roč. 2011/50, 13.9., s. 1-10. ISSN 1213-1539
Dostupné z URL:
<<http://www.elektorevue.cz/cz/clanky/zpracovani-signalu/0/ridke-reprezentace-signalu--uvod-do-problematiky/>>.
- [50] SIEDENBURG K., KOWALSKI M., DOFLER M. *Audio declipping with social sparsity* [online] IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, pp. 1577-1581
Dostupné z URL:
<<https://ieeexplore.ieee.org/document/6853863>>.
- [51] SCHOEFFLER M., BARTOSCHEK S., STOTER F., ROESS M. *webMUSHRA — A Comprehensive Framework for Web-based Listening Tests* [online] 2018, Journal of Open Research Software. 6. 10.5334/jors.187.
Dostupné z URL:
<https://www.researchgate.net/publication/322948098_webMUSHRA_-_A_Comprehensive_Framework_for_Web-based_Listening_Tests>.

- [52] KOSELMAN H. J., ALGINA J., KOWALCHUK R. K. *The analysis of repeated measures designs: A review*. 2001, British Journal of Mathematical & Statistical Psychology, 54, (1), 1-20.

Seznam symbolů, veličin a zkratk

ACR	Absolute Category Rating
ADAT	Alesis Digital Audio Tape – Optické rozhraní přenosu zvuku
AES3	digitální zvukový standard Audio Engineering Society
ANOVA	Analysis of variance – Analýza rozptylu
CASP	Computational Auditory Signal-processing and Perception Model – Výpočetní model sluchového zpracování signálu a vnímání
CCR	Comparison Category Rating – Srovnávací kategorie hodnocení
CD	Compact Disk – Kompaktní disk
CQS	Continuous Quality Scale – Kontinuální stupnice kvality
DCR	Degradation Category Rating – Hodnocení kategorie degradace
DMX512	Digital Multiplex 512– Povelový protokol pro digitální přenos
ERB	Equivalent Rectangular Bandwidth – Ekvivalentní obdélníková šířka pásma
FFT	Fast Fourier transform – Rychlá Fourierova transformace
FIR	Finite Impulse Response – Filtr s konečnou impulzní odezvou
FLAC	Free Lossless Audio Codec – Bezztrátový zvukový kodek
IP	Internetový protokol
k	Tvrdość převodní charakteristiky
ℓ_2	ℓ_2 norma, také známa jako Euklidova norma
LTS	Long-term support – Dlouhodobá podpora
MOS	Mean Opinion Score – Průměrné skóre názorů
MPEG	Moving Picture Experts Group – Skupina expertů pro pohyblivý obraz
MUSHRA	MULTiple Stimuli with Hidden Reference and Anchor – Více stimulů se skrytým odkazem a kotvou

ODG	Objective Difference Grade – Hodnocení objektivního rozdílu
PAMS	Perceptual Analysis Measurement System – Systém měření percepční analýzy
PEAQ	Perceptual Evaluation of Audio Quality – Percepční hodnocení kvality zvuku
PEMO	Perception Model – Sluchový model metody PEMO-Q
PEMO-Q	Perception Model Quality Assessment – Metoda objektivního hodnocení zvuku využívající modelu slyšení
PESQ	Perceptual Evaluation of Speech Quality – Hodnocení vnímané kvality řeči
PSM	Perceptual Similarity Measure – Hodnocení vnímané podobnosti
PSQM	Perceptual Speech Quality Measure – Měření kvality vnímání řeči
<i>r</i>	Mezní práh
rmANOVA	Repeated Measures Analysis of variance – Analýza rozptylu s opakovanými měřeními
SDG	Subjective Difference Grade – Hodnocení subjektivního rozdílu
SDR	Signal-to-distortion ratio – Odstup signálu od zkreslení
SNR	Signal-to-noise ratio – Odstup signálu od šumu
SPL	Sound pressure level – Hladina akustického tlaku
ViSQOL	Virtual Speech Quality Objective Listener – Posluchač cílů kvality virtuální řeči
WAV	Waveform audio file format – Křivkový formát zvukového souboru

A Přílohy

A.1 Výsledky naměřených hodnot

Zvuk	Typ rekonstrukce	SDR input	PEMO-Q [ODG]	Visqol [ODG]	PEAQ [ODG]
Kontrabas	aspade	3,000 dB	-3,7198	-0,699	-3,8007
		5,000 dB	-3,6414	-0,5079	-3,7072
		10,000 dB	-3,3153	-0,3031	-3,0566
		20,000 dB	-1,264	-0,27	-0,8503
	ss pew	3,000 dB	-3,672	-0,621	-3,812
		5,000 dB	-3,6181	-0,5082	-3,7189
		10,000 dB	-3,3077	-0,3449	-3,1394
		20,000 dB	-1,1148	-0,3238	-1,4331
	I1 DR	3,000 dB	-3,7886	-1,3145	-3,8387
		5,000 dB	-3,7038	-0,8299	-3,7549
		10,000 dB	-3,4091	-0,4119	-3,0611
		20,000 dB	-1,4682	-0,2731	-0,6731
Muž	aspade	3,000 dB	-3,5287	-1,4123	-3,8574
		5,000 dB	-3,44	-0,9402	-3,7407
		10,000 dB	-3,1182	-0,3325	-2,9278
		20,000 dB	-0,7697	-0,0246	-1,6502
	ss pew	3,000 dB	-3,5579	-1,6313	-3,827
		5,000 dB	-3,3539	-1,1036	-3,6921
		10,000 dB	-3,166	-0,4059	-2,7895
		20,000 dB	-0,734	-0,0284	-1,1362
	I1 DR	3,000 dB	-3,5498	-2,0299	-3,8488
		5,000 dB	-3,5385	-1,5549	-3,7263
		10,000 dB	-3,4026	-0,6655	-2,7572
		20,000 dB	-1,5503	-0,0644	-0,9988
Duduk	aspade	3,000 dB	-3,8006	-0,7192	-3,9047
		5,000 dB	-3,7001	-0,3974	-3,8796
		10,000 dB	-3,0569	-0,2807	-3,3782
		20,000 dB	-1,323	-0,2685	-1,7637
	ss pew	3,000 dB	-3,4838	-0,4788	-3,6335
		5,000 dB	-3,0326	-0,3168	-3,1951
		10,000 dB	-1,2473	-0,2725	-3,907
		20,000 dB	-0,5037	-0,2704	-1,5811
	I1 DR	3,000 dB	-3,7402	-1,7735	-3,9056
		5,000 dB	-3,6858	-1,2213	-3,9031
		10,000 dB	-3,6343	-0,5188	-3,8996
		20,000 dB	-2,8061	-0,2858	-3,7476
Kytara	aspade	3,000 dB	-3,5217	-0,6745	-3,8178
		5,000 dB	-3,3741	-0,4703	-3,7276
		10,000 dB	-3,0705	-0,335	-3,0644
		20,000 dB	-1,1885	-0,2794	-1,0537
	ss pew	3,000 dB	-3,5219	-0,6784	-3,8139
		5,000 dB	-3,4469	-0,5416	-3,7145
		10,000 dB	-3,2062	-0,3423	-3,8277
		20,000 dB	-0,7516	-0,2778	-0,8602
	I1 DR	3,000 dB	-3,7637	-1,4529	-3,8863
		5,000 dB	-3,6896	-0,8835	-3,8446
		10,000 dB	-3,4277	-0,3913	-3,3747
		20,000 dB	-1,1792	-0,2742	-0,6687
Zvonkohra	aspade	3,000 dB	-2,3584	-0,4504	-3,4877
		5,000 dB	-1,9013	-0,4138	-3,0843
		10,000 dB	-0,7773	-0,3653	-1,6563
		20,000 dB	-0,3252	-0,3114	-0,7659
	ss pew	3,000 dB	-2,6543	-0,4925	-3,7558
		5,000 dB	-2,1122	-0,4538	-3,4105
		10,000 dB	-0,7493	-0,3665	-1,6009
		20,000 dB	-0,1269	-0,3003	-0,5141
	I1 DR	3,000 dB	-2,9609	-0,8421	-3,7466
		5,000 dB	-2,5369	-0,6116	-3,5898
		10,000 dB	-0,8331	-0,3913	-1,9866
		20,000 dB	-0,1949	-0,3173	-0,6318

Obr. A.1: Přehled výsledných hodnot objektivních metrik

B Obsah příloženého CD

Na tomto CD jsou uloženy materiály vytvořené v průběhu práce. Je zde kompletní výstup poslechových testů, jak ve formátu SQL, tak i výstup ve formátech CSV. Dále je zde kompletní data ze statistických analýz, vizualizace i výstup rmANOVA ze softwaru Matlab. Poté jsou zde také uloženy výstupní data objektivních metrik, vizualizace a regresní a korelační analýza.

```
/ ..... kořenový adresář příloženého CD
├── vysledky_poslechovych_testu..... výstup dat poslechových testů
│   ├── Mushrwrite.sql
│   ├── response.csv
│   ├── trial.csv
│   ├── test.csv
│   └── default.yaml
├── analyza_vysledku.....analýza výsledků poslechových testů
│   ├── vizualizace.xlsx
│   ├── ANOVAkomplet.xlsx
│   ├── rmANOVA.csv
│   ├── BASS.xlsx
│   ├── DUDUK.xlsx
│   ├── KYTARA.xlsx
│   ├── MUZ.xlsx
│   └── ZVONKOHRA.xlsx
└── objektivni_metody.....výsledky hodnocení objektivními metodami
    ├── Rekosntrukce.xlsx
    └── Obj_metody_komplet.xlsx
```