



ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

Fakulta elektrotechnická
katedra radioelektroniky

**Hodnocení kvality analogového záznamu zvuku s využitím
modelu slyšení**

**Analog Audio Record Quality Assessment Using a Model of
Auditory Perception**

diplomová práce

Studijní program: Komunikace, multimédia a elektronika

Studijní obor: Multimediální technika

Vedoucí práce: Ing. František Rund, Ph.D.

Martin Zalabák

České vysoké učení technické v Praze
Fakulta elektrotechnická

katedra radioelektroniky

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: **Martin Zalabák**

Studijní program: Komunikace, multimédia a elektronika
Obor: Multimediální technika

Název tématu: **Hodnocení kvality analogového záznamu zvuku s využitím modelu slyšení**

Pokyny pro vypracování:

Seznamte se s metodami objektivního hodnocení kvality zvukového signálu. Implementujte vybranou metodu využívající model slyšení a ověřte její použitelnost pro hodnocení kvality analogových záznamů zvuku. Zaměřte se na artefakty typické pro analogový záznam, studujte a ověřte možnosti úpravy implementované metody pro hodnocení záznamů u kterých není dostupná reference.

Seznam odborné literatury:

- [1] HARLANDER, N., HUBER, R., AND EWERT, S. D. Sound quality assessment using auditory models. J. Audio Eng. Soc 62, 5 (2014),324-336.
- [2] GODSILL, S., RAYNER, P., AND CAPPE, O. Digital audio restoration. Springer, 2002.

Vedoucí: Ing. František Rund, Ph.D.

Platnost zadání: do konce zimního semestru 2016/2017

L.S.

doc. Mgr. Petr Páta, Ph.D.
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 30. 9. 2015

PODĚKOVÁNÍ

Rád bych poděkoval své rodině, přátelům a blízkým za trpělivost a podporu při psaní této práce.

*Podpořeno grantem Studentské grantové soutěže
ČVUT č. SGS14/204/OHK3/3T/13.*

PROHLÁŠENÍ

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

.....
podpis studenta

V Praze dne

ABSTRAKT

Tato diplomová práce se zabývá objektivním modelem hodnocení kvality PEMO-Q založeným na modelu sluchové cesty a možnostmi jeho použití pro hodnocení analogových záznamů. Kromě analýzy vlivu artefaktů typických pro takové záznamy je otestován princip hodnocení bez přítomnosti reference na bázi vytvoření umělé reference potlačením artefaktů za pomoci Wienerovy fitrace šumu. Jako součást této práce byl model PEMO-Q, včetně alternativního modelu sluchové cesty CASP, implementován v prostředí MATLAB.

Klíčová slova: PEMO-Q, CASP, audio, MATLAB, odstranění šumu

ABSTRACT

This diploma thesis deals with PEMO-Q objective audio quality assessment model based on auditory model and its potential for assessment of analog recordings. The influence of artifacts typical for such recordings is analyzed. Also, the possibility of non-reference assessment using artificial reference made by suppression of artifacts is tested with the help of Wiener noise filtering. As a part of this thesis, a complete PEMO-Q, also with substitute auditory model CASP, is implemented in MATLAB.

Keywords: PEMO-Q, CASP, audio, MATLAB, noise removal

Obsah

Seznam použitých symbolů a zkratek	7
Seznam příloh	8
1 Úvod	11
2 Teoretická část	12
2.1 Metody objektivního hodnocení	12
2.1.1 Model objektivního hodnocení PEMO-Q	13
2.1.2 Model sluchové cesty CASP a jeho možné modifikace	17
2.2 Artefakty v analogových záznamech	20
2.2.1 Metody potlačení šumu	21
2.2.2 Principy potlačení impulzních artefaktů	23
3 Implementace modelů a testů	24
3.1 PEMO-Q	24
3.2 Modifikovaný CASP	27
3.3 Modelované artefakty	28
3.4 Odstranění šumu	30
4 Výsledky	32
4.1 Testy modelovaných artefaktů	32
4.2 Modelový bezreferenční test	32
4.3 Výsledky reálných vzorků	33
5 Závěr	35
6 Zdroje	36

Seznam použitých symbolů a zkratk

ERB	Equivalent Rectangular Bandwidth - ekvivalentní obdélníková šířka pásma [2]
PSM	Perceptual Similiarity Measure - hodnota vnímané podobnosti [8]
SDG	Subjective Difference Grade - hodnocení subjektivního rozdílu [8]
ODG	Objective Difference Grade - hodnocení objektivního rozdílu [8]
FIR	Finite Impulse Response - konečná impulzní odezva [25]
IIR	Infinite Impulse Response - nekonečná impulzní odezva [25]
PEMO-Q	Perception Model - Quality - model vnímání, kvalita [8]
CASP	Computational Auditory Signal-processing and Perception Model - výpočetní model sluchového zpracování signálu a vnímání [15]
DRNL	Dual Resonance Non-Linear filter - dvojný rezonanční nelineární filtr [19]
RMS	Root Mean Square - Odmocnina z průměru kvadrátů
FFT	Fast Fourier Transform - implementace diskrétní Fourierovy transformace

Seznam příloh

Příloha 1: Výstupní závislosti modelů

Příloha 2: CD s MATLAB kódy a výsledky testů s analogovými záznamy

1 Úvod

Problematika hodnocení kvality zvuku je vzhledem ke komplexitě lidského slyšení a jeho popisu záležitostí primárně subjektivních testů. Příprava a realizace takových testů je vzhledem k nutnosti provedení napříč dostatečně velkým vzorkem posluchačů za pomoci vhodné poslechové techniky nákladná na čas i hmotné prostředky [14]. Výhradně subjektivní hodnocení má také negativní aspekt ve skutečnosti, že nelze zaručit konzistentnost a reprodukovatelnost zjištěných výsledků. Tyto skutečnosti jsou motivacemi ke vzniku objektivních metod hodnocení, kde kvalitativní hodnocení je provedeno zpracováním zvukového signálu. To minimalizuje náklady na prostředky spojené s vyhodnocením počítačového algoritmu a může potenciálně umožnit i konzistentnost a reprodukovatelnost zjištěných hodnot. Negativní aspekt tohoto přístupu je obtížnost ověření jestli a do jaké míry daná metoda odpovídá lidskému vnímání zvukové kvality.

Úkolem této práce je seznámit se s metodami objektivního hodnocení kvality zvukového signálu a implementovat vhodnou metodu využívající model slyšení. Dále bude ověřena použitelnost pro hodnocení kvality analogových záznamů vzhledem k artefaktům přítomným v takových záznamech. Pro záznamy, kde není přítomna reference, budou řešeny možnosti přizpůsobení implementovaných metod. Je testován princip vytvoření reference skrze minimalizaci artefaktů na potlačení šumu pomocí Wienerova filtru. Motivací k prozkoumání použitelnosti těchto modelů v této aplikaci je především skutečnost, že i přes jejich možné obecné použití pro vyhodnocení kvality se současná literatura zaměřuje výhradně na aplikaci v oblasti kompresních a přenosových kodeků či digitálního přenosu a nikoliv na hodnocení digitalizovaných analogových nahrávek [6] [21].

V kapitole 2 jsou z teoretického pohledu rozebrány metody objektivního hodnocení kvality a artefakty přítomné v analogových signálech. V následující kapitole bude popsán implementovaný model PEMO-Q a jeho úpravy a také realizované testy. Výsledky těchto testů budou představeny v kapitole 4.

2 Teoretická část

2.1 Metody objektivního hodnocení

Jak bylo naznačeno v úvodu, za nejlepší metodu hodnocení kvality zvuku jsou stále považovány subjektivní testy [21]. Doporučení ITU-R BS.1116 [9] specifikuje systém hodnocení založený na pěti úrovních (znázorněný v tabulce 1), kde hodnota 5 vyjadřuje neslyšitelné zkreslení zvuku a 1 velmi rušivé zkreslení. Při komparativních testech lze použít rozdíl testované a referenční hodnoty (tudíž hodnota 0 vyjadřuje identitu a -4 velmi rušivé zhoršení) též známý jako hodnota SDG [8]. Specifikuje rovněž, že vhodný test vyžaduje nejen dostatečně velký počet posluchačů, ale též musí být posluchači vhodně vyškolení pro hodnocení. Dále jsou popsány minimální požadavky pro poslechové zařízení zahrnující mimo jiné například parametry reproduktorů a akustické vlastnosti poslechové místnosti. Realizace takových testů tedy vyžaduje značné finanční a časové investice.

Tyto obtíže byly motivací pro analýzu vlastností zvuku, které rozhodují o subjektivním měřítku kvality, pro možnost tyto prostředky ušetřit a nahradit hodnocení kvality algoritmem signálového zpracování. Vzhledem k řadě různých perspektiv, jak „kvalitu“ zvuku pojmout, vznikla řada metod pro kvalitativní hodnocení. Ty lze rozdělit do dvou skupin. Jedna skupina jsou metody relativní, intruzivní, které porovnávají testovaný signál vůči referenčnímu „dokonale kvalitnímu“ signálu, a metody absolutní, neintruzivní, které hodnotí jen na základě testovaného signálu. Tato klasifikace má analogii v subjektivních testech, kde dochází rovněž k hodnocení na základě porovnání s referencí či hodnocení zvuku samotného. Dostupné metody neintruzivního hodnocení se zaměřují primárně na řečové zvuky skrze modelování vlastností řeči, pro obecnější hodnocení je tedy na místě využít metody intruzivní [6].

Pro co největší přiblížení se lidskému vnímání kvality využívá řada intruzivních metod modelu sluchové cesty. První takovou standardizovanou metodou je Perceptual Evaluation of Audio Quality neboli PEAQ doporučený v ITU-R BS.1387 [10]. Tato metoda vyhodnocuje testovaný zvuk v modulaci, specifické hlasitosti, pomalých variací zesílení a excitačním vzoru [21]. Samotné kvalitativní hodnocení

Subjektivní zkreslení	Známka
Neslyšitelné	5.0
Slyšitelné, ale nerušivé	4.0
Lehce rušivé	3.0
Rušivé	2.0
Velmi rušivé	1.0

Tabulka 1: Systém hodnocení subjektivního zkreslení zvuku

výstupu modelové sluchové cesty vykonává umělá neuronová síť. Mezi ITU doporučeními lze dále nalézt modely Perceptual Speech Quality Measure (PSQM) [11], kde je místo neuronové sítě použito lineární kombinace různých průměrovaných rozdílů, a navazující Perceptual Evaluation of Speech Quality (PESQ) [12] s mimo jiné vylepšeným psychoakustickým modelem. Tento model je navzdory svému řečovému zaměření běžně používán i pro jiné signály [6]. Posledním ITU standardem v této oblasti je Perceptual Objective Listening Quality Analysis (POLQA) z doporučení ITU-T P.863 přizpůsobený řečovými signály o delší šířce pásma a obohacený o predikci kvality řečových signálů akusticky nahraných s použitím umělé hlavy [13].

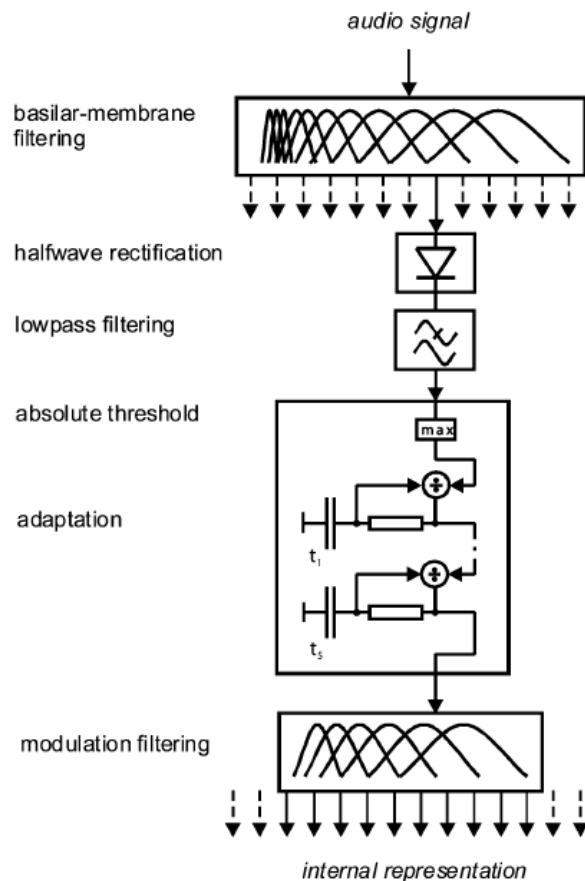
2.1.1 Model objektivního hodnocení PEMO-Q

Metoda využitá v rámci této práce je model PEMO-Q, který sestavili Huber a Kollmeier a představili ve svém článku v roce 2006 [8]. Tato metoda dle svých tvůrců vychází z modelu PEAQ, avšak se snaží adresovat obecné artefakty a typy zkreslení zvuku a nezaměřovat se jen na hodnocení zvuků kódovaných na nízké bitové toky. Dále se snaží poskytnout stálé měřítko, které netrpí na proměnné jevy svázané s adaptací modelu (např. spojené s umělou neuronovou sítí).

Model PEMO-Q je intruzivní, tedy kvalitu vyhodnocuje skrze porovnání s „dokonale kvalitním“ referenčním signálem. Před samotnou analýzou je nutné zajistit, aby referenční i testovaný signál byly úrovní a časově vyrovnané. Navrhovanou metodou vyrovnaní úrovně je vynásobení testovaného signálu skalární hodnotou pro dosažení stejného RMS výkonu. Časovou odchylku lze vyrovnat posunutím jednoho ze signálů. Při neznalosti časové odchylky je vhodné ji odhadnout s využitím vzájemné korelace signálů. V případě, že se tyto odchylky v čase mění, je nezbytné tyto korekce provádět blok po bloku.

Pro převod zvukového signálu na hodnoty odhadovaných nervových stimulů je využito modelu sluchové cesty, znázorněného na obrázku 1. Prvním krokem tohoto modelu je modelování bazilární membrány bankou 35 gammatónových filtrů čtvrtého řádu o středních kmitočtech od 235 do 14500 Hz. Odstupy jednotlivých středních kmitočtů a šířky pásem odpovídají hodnotě 1 ERB [2]. Tím je dosaženo napodobení schopnosti frekvenčního rozlišení lidského sluchu. Výstup každého filtru je dále zpracováván nezávisle.

Signál na výstupu gammatónového filtru je půlvlnně usměrněn (tedy bez zachování záporné půlvlny) a filtrován dolní propustí s mezním kmitočtem 1 kHz. Tento proces zachovává obálku signálu nad tímto kmitočtem a pro nižší kmitočty uchovává informaci o amplitudě a fázi. Tyto dva kroky modelují převod z mechanických kmitů na nervové impulzy. Dále je zvuk limitován úrovní dolní mezí závislou na ma-



Obrázek 1: Blokové schéma modelu sluchové cesty v PEMO-Q [8]

ximální úrovni vstupního signálu (při amplitudě 1 odpovídající úrovni 10^{-5} [6]) a přiveden na vstup bloku adaptivní filtrace.

Tento blok se skládá z pěti zpětnovazebných smyček zapojených do kaskády s dolní propustí ve zpětné vazbě (na obrázku 1 ilustrované pomocí RC článku). Tyto dolní propusti jsou definovány pomocí rozsahu časových koeficientů, konkrétně od 5 do 500 ms. Výstupem zpětné vazby je vstupní signál vydělen. Tato kaskáda má účinek takový, že stacionární signály jsou sníženy přibližně na úroveň 32. odmocniny, což napodobuje logaritmickou kompresi lidského slyšení. Na velké změny však tento proces reaguje přibližně lineárně.

Posledním krokem modelu sluchové cesty je modulační banka filtrů modelující schopnost rozpoznání amplitudové modulace [8] [1]. Filtr s nejnižším kmitočtem je dolní propust druhého řádu nastavená na kmitočet 2.5 Hz. Následují dva filtry s konstantním pásmem 5 Hz a středními kmitočty 5, resp. 10 Hz. Zbytek banky tvoří pět filtrů s konstantní Q hodnotou 2 a překryvy na -3 dB. S takto definovanými parametry odpovídá nejvyšší střední kmitočet hodnotě 129 Hz. Vzhledem k relativně malým kmitočtům a kmitočtovým pásmům oproti vstupním zvukovým signálům a velkým datovým tokům v tomto bodě zpracování (každý vzorek vstupního signálu

je v tomto bodě reprezentován $35 \cdot 8 = 280$ vzorky) je počítána Hilbertovská obálka výstupních signálů, a ta je podvzorkována na nejméně šestinásobek středního resp. mezního kmitočtu filtrů. Z Hilbertovské obálky je brána jen reálná část signálu.

Po průchodu zvukového signálu modelem sluchové cesty (referenční i testovaný zvuk jsou modelem vyhodnoceny nezávisle) a před samotným vyhodnocením je provedena asimilace vnitřní reprezentace testovaného signálu výpočtem

$$\hat{y}_{tfm} = \begin{cases} \frac{y_{tfm} + x_{tfm}}{2} & |y_{tfm}| < |x_{tfm}| \\ y_{tfm}, & |y_{tfm}| \geq |x_{tfm}|. \end{cases} \quad (2.1)$$

Proměnná x představuje hodnotu vnitřní reprezentace referenčního signálu a y hodnotu zkresleného. Indexy t , f a m představují proměnné v čase, frekvenčním kanálu a modulačním kanálu vnitřní reprezentace. Motivací k tomuto kroku je předpoklad, že chybějící či potlačené rysy ve zvuku jsou méně subjektivně invazivní než přidané či zesílené, tudíž je vhodné částečně potlačit rozdíl menších vjemů, než je v referenčním zvuku.

Vyhodnocení objektivní kvality testovaného zvuku je provedeno dvěma metodami. První metoda počítá vzájemnou korelaci přes celou délku signálu a pro všechna frekvenční pásma avšak nezávisle přes všechna pásma modulační banky vztahem

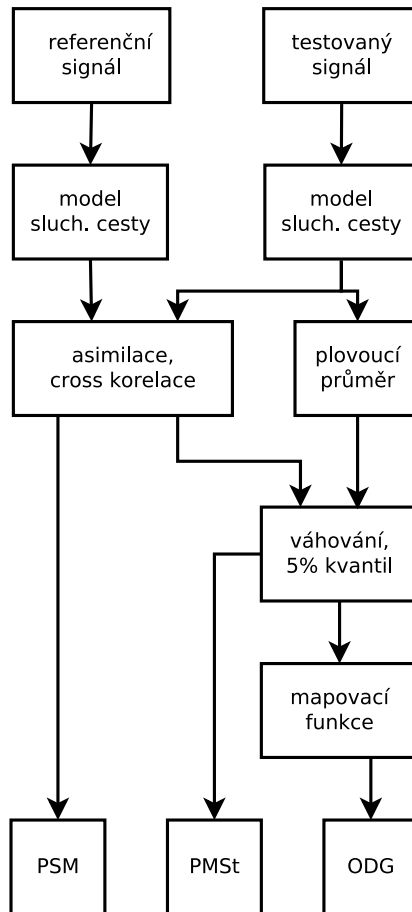
$$r = \frac{\sum_{t,f} (x_{tf} - \bar{x})(y_{tf} - \bar{y})}{\sqrt{\sum_{t,f} (x_{tf} - \bar{x})^2 \sum_{t,f} (y_{tf} - \bar{y})^2}}. \quad (2.2)$$

Hodnoty \bar{x} a \bar{y} představují průměrné hodnoty reprezentací přes čas a frekvenční pásma. Tato hodnota je dále násobena koeficientem w_m spočteným skrze podíl součtu kvadrátů všech hodnot reprezentace zkresleného signálu daného modulačního pásma a součtu kvadrátů hodnot všech modulačních pásem této reprezentace. Výstupem součtu těchto váhovaných vzájemných korelací je finální hodnota PSM . Zjištěná hodnota se má dle tvůrců [8] pohybovat v rozmezí od 0 do 1, kde $PSM = 1$ vyjadřuje identitu, byť teoreticky je prý možné dosáhnout hodnoty až -1.

$$w_m = \frac{\sum_{t,f} y_{tfm}^2}{\sum_{t,f,m} y_{tfm}^2} \quad (2.3)$$

$$PSM = \sum_m w_m r_m \quad (2.4)$$

Druhou výstupní hodnotou modelu je PSM_t , která oproti předcházející hodnotě přidává časovou závislost. Toho je dosaženo počítáním vzájemné korelace (iden-



Obrázek 2: Blokové schéma kompletního modelu PEMO-Q

ticky použitím příslušného vztahu (2.2)) pro 10 ms rámce vnitřních reprezentací. Výsledné hodnoty těchto korelací, jak je nastíněno v obrázku 2, jsou váhovány pohyblivým průměrem časového průběhu vnitřní reprezentace zkresleného signálu ve snaze popsat „okamžitou aktivitu“ ve zvuku. Z takto váhovaných krátkodobých korelací je zvolen 5%-ní kvantil jako výsledná hodnota.

Pro snadné porovnání s *SDG* hodnotami je na výstupu PEMO-Q navíc hodnota PSM_t , která je považována za přesnější odhad [8], mapována na srovnatelnou *ODG* funkcí

$$ODG(x) = \begin{cases} \max\{-4, \frac{a}{x-b} + c\}, & x < x_0 \\ d \cdot x - d, & x \geq x_0. \end{cases} \quad (2.5)$$

Tímto přepočtem je dosaženo hodnoty 0 pro identitu a 4 pro velmi rušivé zkreslení. Konstanty použité při přepočtu jsou $a = -0.22$, $b = 0.98$, $c = -4.13$, $d = 16.4$ a $x_0 = 0.864$.

2.1.2 Model sluchové cesty CASP a jeho možné modifikace

Model sluchové cesty zahrnutý v modelu PEMO-Q je možné nahradit modelem CASP představeném v článku vydaném v roce 2008 [15]. Tento model vychází ze staršího modelu, který definoval Dau *et al.* v roce 1996 [1] a který je v PEMO-Q zahrnut [8], ale v několika jeho částech ho rozšiřuje.

Vzhledem k přidanému nelineárnímu bloku v části bazilární membrány je vhodným prvním krokem zesílit či ztlumit vstupní signál na vhodnou hlasitost. Zvolené hodnoty odpovídají digitálnímu signálu o amplitudě 1 hlasitosti 100 dB. Samotným prvním krokem analýzy je však zahrnutý blok vnějšího a středního ucha přidaný na základě úplné absence ošetření této části slyšení v předchozím modelu. První filtr je definován přenosovou funkcí získanou měřením přenosu mezi sluchátky a bubínkem [20]. Filtr středního ucha byl dále odvozen z dat získaných měřením na zemřelém člověku. [5]. Výstupem těchto filtrů je rychlost třmínku. Frekvenční závislosti těchto filtrů jsou zobrazeny na obrázku 3.

Laboratoř výzkumu slyšení na Univerzitě v Essexu tvrdí, že tento postup a naměřené hodnoty neodpovídají skutečnosti a pro svůj model MAP sestavila alternativu využívající hodnoty přizpůsobené datům živých lidí a počítající výchylku třmínku místo jeho rychlosti [17]. Tím bylo také dosaženo frekvenční nezávislosti v kroku bazilární membrány. Průchod zvuku do středního ucha počítá tento model dvěma rezonančními filtry prvního řádu se zisky 10 a 25 dB, přičemž první filtr má definované pásmo od 1 do 4 kHz a druhý od 2.5 do 7 kHz. Samotná transformace na výchylku třmínku je provedena Butterworthovým filtrem prvního řádu nastaveném na 50 Hz, útlumem pomocí násobné hodnoty $45e - 9$ a Butterworthovou horní propustí s mezním kmitočtem 1 kHz.

Model bazilární membrány byl změněn na DRNL banku [19]. Ta se skládá z lineární a nelineární větve, které se na výstupu sčítají. V lineární části je vstupní signál lineárně zesílen (konstantním koeficientem g) a filtrován kaskádou identických gammatónových filtrů prvního řádu (střední kmitočet f_{clin} , šířka pásma BW_{lin} odpovídající 1 ERB [2]) a dále navíc kaskádou identických dolních propustí s mezním kmitočtem rovným střednímu kmitočtu předcházejících gammatónových filtrů. Hodnota koeficientu g je závislá na použitém pásmu pro daný DRNL filtr banky, což odpovídá frekvenční závislosti bazilární membrány. Počet filtrů a jejich přesné pozice jsou nechány na volbě implementaci.

Nelineární část se skládá z kaskády gammatónových filtrů (střední kmitočet f_{cnlin} , šířka pásma BW_{nlin} odpovídající rovněž 1 ERB), kompresního prvku, opětovné kaskádové gammatónové filtrace a kaskády dolních propustí. Obě kaskády gammatónových filtrů jsou identické a jsou podobně jako v lineární větvi řádu 1,

avšak jejich střední kmitočet je od té odlišný. Mezní kmitočet dolních propustí též odpovídá f_{cnlin} . Kompresní prvek má charakter „zlomené tyče“ a je definován třemi parametry: lineárním ziskem a , ziskem b a exponentem c . Pro nízké úrovně je přenos definován lineární funkcí

$$y[t] = a \cdot x[t], \quad (2.6)$$

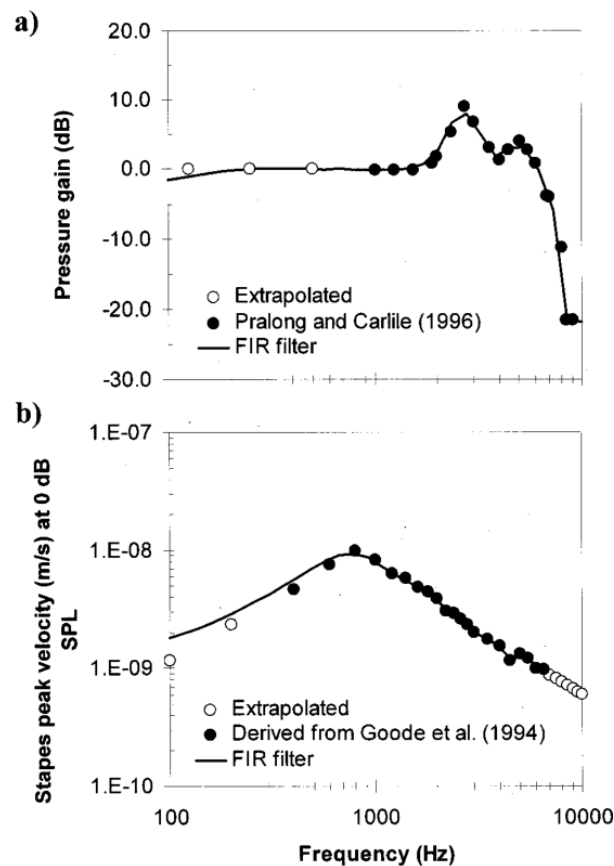
kde $x[t]$ je vstupní a $y[t]$ výstupní vzorek v čase t . Na vyšších úrovních je signál zpracován nelineárně skrze

$$y[t] = b \cdot |x[t]|^c \cdot \text{sign}(x[t]). \quad (2.7)$$

Výstupní hodnota je vždy menší výsledek z těchto dvou. Lze tedy oba vztahy spojit do jednoho výpočtu:

$$y[t] = \text{sign}(x[t]) \cdot \min(a \cdot |x[t]|, b \cdot |x[t]|^c). \quad (2.8)$$

Hodnoty jsou rovněž závislé na frekvenčním pásmu. Typické hodnoty banky uvádí Lopez-Poveda *et al.* ve svém textu [18] z roku 2001 a příklad je uveden na obrázku 4.



Obrázek 3: Charakteristiky filtrů vnějšího a středního ucha použité v CASP [15]

Signal frequency (Hz)	250	500	1000	2000	4000	8000
DRNL linear path						
No. cascaded GT filters	2	2	2	2	2	2
CF _{lin} (Hz)	235	460	945	1895	3900	7450
BW _{lin} (Hz)	115	150	240	390	620	1550
g	1400	800	520	400	270	250
LP _{lin} cutoff (Hz)	CF _{lin}	CF _{lin}	CF _{lin}	CF _{lin}	CF _{lin}	CF _{lin}
No. cascaded LP filters	4	4	4	4	4	4
DRNL nonlinear path						
No. cascaded GT filters	3	3	3	3	3	3
CF _{nl} (Hz)	250	500	1000	2000	4000	8000
BW _{nl} (Hz)	84	103	175	300	560	1100
a	2124	4609	4598	9244	30274	76354
b [(m/s) ^(1-c)]	0.45	0.280	0.130	0.078	0.060	0.035
c	0.25	0.25	0.25	0.25	0.25	0.25
LP _{nl} cutoff (Hz)	CF _{nl}	CF _{nl}	CF _{nl}	CF _{nl}	CF _{nl}	CF _{nl}
No. cascaded LP filters	3	3	3	3	3	3
DRNL filter BF (Hz)	260	508	1002	2006	3978	7720
DRNL filter BW _{3dB} (Hz)	47	70	118	210	415	755

Obrázek 4: Hodnoty vybraných pásem DRNL banky modelující subjekt YO z měření, které provedli Lopez-Poveda *et al.* [18]

Model MAP definici konstant pro DRNL banku zjednodušuje skrze výše uvedené využití výchylky třmínku místo jeho rychlosti. Zisk lineární větve je pro všechna pásma $g = 50$ a parametry kompresního prvku jsou $a = 5 \cdot 10^3$, b takové, aby komprese začala na vstupní hodnotě $10^{-9} \cdot 10^{25/20}$ (tedy 25 dB nad prahem slyšení definovaným úrovní výchylky $1 \mu\text{m}$) a $c = 0.2$. Mezi středními resp. mezními kmitočty lineární a nelineární větve platí závislost

$$f_{clin} = p + q \cdot f_{cnlin}, \quad (2.9)$$

kde $p = 266$ a $q = 0.621$.

Další odlišnosti oproti PEMO-Q tvoří v zesílení před krokem adaptivní kaskády následované druhou mocninou signálu. Zesílení je vyřešené jednoduchým přepočtem se ziskem 50 dB, ačkoliv Harlander *et al.* uvádí, že vhodnější je zisk 38 dB [6]. Dále, vzhledem k nižším hodnotám na úrovni adaptivní smyčky, je vhodné použít nižší dolní mez. Poslední úpravou je přidání dolní propusti prvního řádu před vstupem do modulační banky, což má napodobit nižší citlivost slyšení na rychlé modulace.

2.2 Artefakty v analogových záznamech

Pro průzkum možnosti použití metod objektivního hodnocení kvality na analogových záznamech je vhodné prozkoumat, jaké artefakty se v takových záznamech vyskytují. Godsill *et al.* [3] uvádí několik artefaktů s rozdělením do dvou skupin, a to lokalizované artefakty, které se projevují v signálu na určitých vzorcích (např. škrábance), a artefakty globální, které ovlivňují celý signál. Do skupiny lokalizovaných patří tyto artefakty:

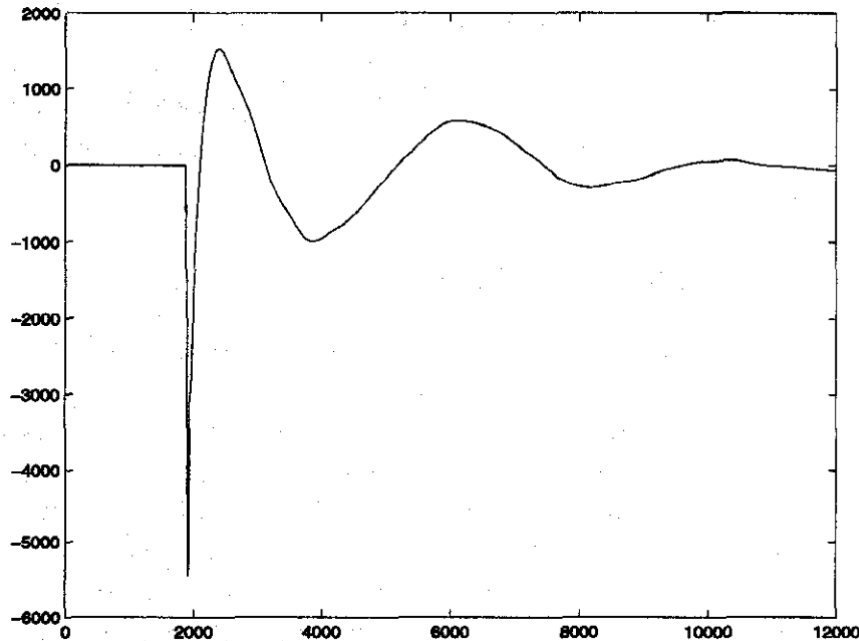
- Krátké poruchy šumového charakteru nahodilé v čase a amplitudě, způsobené např. drobnými škrábanci na vinylových deskách.
- Přechodný nízkofrekvenční impulz: Větší poruchy v médiu, které způsobí nízkofrekvenční resonanci ve snímačovém zařízení na poslech připomínající zvuk úderu.

Artefakty mající globální charakter lze v analogových nahrávkách nalézt tyto:

- Širokopásmový šum: běžně přítomný artefakt projevující se stálým „syčivým“ zvukem. Obvykle se jedná o stacionární šum, avšak ne pravidlem zejména co se týče starších nahrávek. Frekvenční charakteristika šumu může mít mnoho podob.
- Wow a flutter: změny výšky způsobené nestandardním chováním přehrávající mechaniky jako například kolísání rychlosti motoru. Efekt moduluje všechny frekvenční složky. Subjektivní dopad tohoto efektu je velmi závislý na charakteru původního zvuku a parametry tohoto efektu mají mnoho stupňů volnosti (frekvence, frekvenční hloubka, časová závislost) a proto není v rámci této práce z důvodu složitosti analyzován.
- Zkreslení: Obecné nelineární změny jako například ovlivnění amplitudy usměrňovacím jevem či tvrdým limitováním.

Pro analytické důvody je vhodné tyto artefakty modelovat. V témže díle lze zjistit, že krátké poruchy je možné modelovat dvěma způsoby. Prvním je aditivní model, kdy je do signálu na patřičné místo přičten krátký (o délce méně 4 ms) šumový impulz. Druhá metoda je metoda nahrazení, údajně vhodnější pro delší poruchy. V případě delších poruch se v analogových médiích běžně vyskytují resonance, tudíž je v takových situacích přesnější model přechodného nízkofrekvenčního impulzu.

Jak bylo uvedeno výše, nízkofrekvenční impulzy souvisí obvykle s resonancí ve snímačovém zařízení, tedy se bude jednat o kmitání na konkrétním resonančním



Obrázek 5: Příklad přechodného nízkofrekvenčního impulzu z gramofonové nahrávky [4]

kmitočtu (15 až 30 Hz [3]). Vzhledem k přechodnému charakteru a ostrému nástupu lze rovněž předpokládat charakter klesající exponenciály. Tomu odpovídá příklad takového impulzu uvedený ve článku od Godsilla *et al.* na obrázku 5 [4]. Ze samotného příkladu lze odvodit délka řádově jednotek period. Vhodná metoda pro modelování je tedy použít harmonickou funkci začínající v maximu resp. minimu o kmitočtu 15 až 30 Hz s exponenciální obálkou o uvedené délce.

Širokopásmový šum je možné modelovat aditivním bílým šumem. Specifičtější frekvenční charakteristiku lze získat vhodnou filtrací bílého šumu, jako např. váhování vzorků diskretní Fourierovy transformace. Různé formy nelineárního zkreslení lze realizovat aplikací vhodné funkce na vstupní signál. Tvrdý limiter, jako příklad, je možné aplikovat funkcí

$$y[t] = \text{sign}(x[t]) \cdot \min(\text{abs}(x[t]), a), \quad (2.10)$$

kde a je lineární úroveň, kterou má být signál limitován. [25]

2.2.1 Metody potlačení šumu

Pro odstranění šumu ze zvukového záznamu se nabízí řada metod. Mezi ty nejrozšířenější však patří metody založené na zpracování krátkých bloků ve spektrální doméně [3]. Pro tyto metody je nezbytné znát výkonové spektrum šumu (dále známo

jako S_n), které však, za užití předpokladu stacionarity šumu, lze získat z „tichých“ částí signálu.

Zpracování probíhá na krátkých blocích zvukového signálu (pro efektivní výpočet FFT o délce 1024 či 2048 vzorků). Aby nedošlo k frekvenčnímu přetékání a vzniku nespojitostí, je doporučen překryv bloků (např. o 50%) a použití funkce okna (např. Hammingovy nebo Hannovy).

První zde uvedenou metodou je Wienerův filtr [24] definovaný pomocí přenosu tímto vztahem:

$$H(\omega) = \frac{S_x(\omega)}{S_x(\omega) + S_n(\omega)}, \quad (2.11)$$

kde $S_x(\omega)$ je výkonové spektrum bloku signálu bez šumu, tedy požadovaný výstup. Při diskretizaci lze filtr definovat jako funkci

$$f(Y(m)) = \frac{S_x(m)}{S_x(m) + S_n(m)} Y(m) \quad (2.12)$$

s $Y(m)$ jako výstupním spektrem. Při vyjádření výkonového spektra výstupu jako rozdílu výkonového spektra vstupu a šumu a zajištění, že v případě nepřesného odhadu šumu nepůjde hodnota výkonu do záporných čísel (což odporuje definičnímu oboru výkonu)

$$S_x(m) = \begin{cases} |Y(m)|^2 - S_n(m), & |Y(m)|^2 > S_n(m) \\ 0, & |Y(m)|^2 \leq S_n(m) \end{cases} \quad (2.13)$$

vychází finální funkce

$$f(Y(m)) = \begin{cases} \frac{|Y(m)|^2 - S_n(m)}{|Y(m)|^2} Y(m), & |Y(m)|^2 > S_n(m) \\ 0, & |Y(m)|^2 \leq S_n(m). \end{cases} \quad (2.14)$$

Další metoda je spektrální odečítání, které oproti Wienerovu filtru pracuje s amplitudovými spektry. Ekvivalentně lze spektrální odečítání vyjádřit vztahem

$$f(Y(m)) = \begin{cases} \frac{|Y(m)| - S_n(m)^{1/2}}{|Y(m)|} Y(m), & |Y(m)|^2 > S_n(m) \\ 0, & |Y(m)|^2 \leq S_n(m). \end{cases} \quad (2.15)$$

Poslední zmíněnou metodou je výkonové odečítání, kde je výstupní výkon nastaven na výkon vstupu minus očekávaný šumový výkon:

$$f(Y(m)) = \begin{cases} \left(\frac{|Y(m)|^2 - S_n(m)}{|Y(m)|^2} \right)^{1/2} Y(m), & |Y(m)|^2 > S_n(m) \\ 0, & |Y(m)|^2 \leq S_n(m). \end{cases} \quad (2.16)$$

Vzhledem k nepřesnostem v aproximaci šumové hodnoty a krátkodobé nestacionaritě šumu dojde u všech tří metod ke snížení amplitudy. K zesílení efektu potlačení šumu je možné zesílit odhad výkonového spektra šumu, např. vynásobením koeficientem větším 1. Se zvyšováním odhadu spektra však dojde ke zvětšování míry nežádoucího zkreslení redukcí nešumových frekvenčních složek signálu.

2.2.2 Principy potlačení impulzních artefaktů

Pro úplnost jsou zde uvedeny principy možných metod minimalizace dalších představených artefaktů [3]. K odstranění krátkých impulzních poruch je možné použít autoregresní modely. Vzhledem ke skutečnosti, že poruchy samotné mají obvykle výrazně odlišný charakter oproti okolnímu zvuku, lze k detekci těchto poruch použít chybový signál autoregresní syntézy, že jeho úroveň při poruše obvykle prudce vzroste.

Po detekci je možné rovněž využít syntézu autoregresním modelem pro interpolaci signálu v místě poruchy. Literatura uvádí, že je vhodné k aproximaci hodnot AR modelu použít metodu nejmenších čtverců. Je dále uvedeno, že využitím ARMA modelu lze údajně dosáhnout lepších výsledků, avšak se zanedbatelným rozdílem. V případě krátkodobých nízkofrekvenčních impulzů lze dle článku od stejného autora [4] využít stejné metody k detekci i interpolaci, avšak pro lepší výsledky je vhodnější k interpolaci využít Kalmanův filtr.

3 Implementace modelů a testů

V rámci této práce byl v prostředí MATLAB R2012a (s využitím Signal Processing Toolbox) implementován model pro objektivní hodnocení kvality PEMO-Q dle popisu z kapitoly 2.1.1. Dále byl implementován alternativní model sluchové cesty CASP s modifikacemi z modelu MAP a se zohledněním nových poznatků dle popisu z kapitoly 2.1.2. Dále byly vytvořeny modely artefaktů přítomných v analogových nahrávkách a otestována závislost výstupních hodnot PEMO-Q a PEMO-Q s modelem sluchové cesty CASP na míře ovlivnění zvuku artefakty. Vzhledem k rušivosti těchto artefaktů by měly výstupní hodnoty PEMO-Q s rostoucí mírou ovlivnění signálu artefakty klesající tendenci.

Pro využití PEMO-Q bez přítomnosti reference byl vyzkoušen princip vytvoření umělé reference algoritmem potlačení šumu na principu Wienerovy filtrace popsané v kapitole 2.2.1. To bylo provedeno porovnáním hodnot uměle zašuměného signálu využívající originální reference bez šumu a reference vytvořené potlačením šumu. Možnost aplikace této metody byla vyzkoušena na skutečných digitalizovaných záznamech.

3.1 PEMO-Q

Model PEMO-Q popsáný v kapitole 2.1.1 je implementován skrze několik funkcí, konkrétně skrze funkci `pemoq_am` realizující model sluchové cesty pro vložený signál, a funkci `objquality` realizující samotné vyhodnocení hodnot PSM , PSM_t a ODG z dat modelované sluchové cesty. Před využitím modelu sluchové cesty je nezbytné získat strukturu s definicemi hodnot koeficientů filtrů pro daný vzorkovací kmitočet pomocí funkce `pemoq_am_prepare`. Tato separace byla provedena z důvodu zrychlení, že při stálém vzorkovacím kmitočtu je opakovaný výpočet těchto hodnot redundantní. Pro přehlednost je zde implementace popsána přes jednotlivé kroky modelu napříč soubory.

Implementace předpokládá vhodně nastavené úrovně referenčního i testovaného signálu a časovou synchronizaci. Pro asistenci s vhodným nastavením úrovně je k dispozici funkce `preproc_norm`. Výstupem této funkce je násobící koeficient pro testovaný signál vypočtený z poměru mediánů RMS výkonů krátkých rámců signálů (o zvolené délce N). Tyto mediány byly zvoleny vzhledem k charakterům možných artefaktů. Přechodné časové impulzy a šum mohou celkový RMS výkon navýšit. Je však vytvořen předpoklad, že tyto artefakty se při vyhodnocení s krátkými rámci výrazně projeví jen ve statisticky okrajových hodnotách (hlasité impulzy na horním okraji, šum na dolním okraji), tudíž by vliv na mediánovou hodnotu mohl

být minimální.

Prvním krokem modelu sluchové cesty je blok bazilární membrány, tedy banka gammatónových filtrů 4. řádu. Implementace těchto IIR filtrů je převzata z kódu, který napsal Malcolm Slaney [22]. Jediné funkční úpravy tohoto kódu jsou v substituci maximálního středního kmitočtu $f_s/2$ na hodnotu 14.5 kHz a fixním počtu 35 kanálů pro splnění parametrů PEMO-Q. Tyto definice IIR filtrů jsou k nalezení v pomocné funkci `mod_gammafilt` a v příslušných částech `pemoq_am_prepare`. Vzhledem k faktu, že na nízkých kmitočtových pásmech se jeví Slaneyho implementace filtru nestabilně a rezonuje v maximu, je na kmitočtech $f_c < f_s/4$ pracováno se signálem podvzorkovaným na polovinu funkcí `decimate`. Signál je po filtraci pro uniformitu zpětně interpolován funkcí `interp`.

Model vlásečnic je realizován operací usměrnění a low-pass filtrem prvního řádu. K usměrnění je využito vlastnosti interpretu MATLAB, kdy je logická hodnota podmínky „ x je větší než 0” využita jako násobící koeficient:

```
bma=bm.*(bm>0);
```

Pro filtraci je využit IIR filtr prvního řádu spočtený skrze [25]

$$K = \tan \frac{\pi f_c}{f_s} \quad (3.1)$$

$$b_0 = b_1 = \frac{K}{K+1} \quad (3.2)$$

$$a_1 = \frac{K-1}{K+1} \quad (3.3)$$

pro $f_c = 1\text{kHz}$. Krok adaptivní filtrace, včetně aplikace minimální hodnoty $1 \cdot 10^{-5}$ byl realizován již hotovou open-source implementací CASP [16] využívající kompilovaný mex soubor. Motivací k tomu byla nutnost pracovat vzorek po vzorku pro celý signál skrze 35 pásem a pět filtrů v kaskádě. Tento postup je interpretem MATLAB vyhodnocován velmi neefektivně, a to vedlo autory implementace CASP k vytvoření kompilované mex funkce v jazyce C.

První filtr modulační banky byl sestaven pomocí funkce `butter` jako Butterworthův filtr druhého řádu s mezním kmitočtem 2.5 Hz. Vyšší filtry, popsané v kapitole 2.1.1 jsou definovány s příslušnými kmitočty f_c následujícími vztahy: [7]

$$w_0 = \frac{2\pi f_c}{f_s} \quad (3.4)$$

$$e_0 = \exp\left(-\frac{w_0}{Q}/2\right) \quad (3.5)$$

$$b_0 = 1 - e_0 \quad (3.6)$$

$$a_1 = -e_0 \cdot e^{j\omega_0} \quad (3.7)$$

Filtrace sama je provedena pro velký objem dat (8 pásem modulační banky krát 35 pásem frekvenční banky) zvláštní funkcí `mod_packfilter`, která po filtraci signály decimuje na kmitočet $f'_s = \frac{f_s}{n}$, kde n je dolů zaokrouhlený podíl originálního vzorkovacího kmitočtu a desetinásobku středního, v případě dolní propusti mezního kmitočtu filtru modulační banky (s pro každé pásmo banky jinou hodnotou n). Decimační prvek využívá funkci `decimate` s FIR filtrací, že implicitní IIR filtr jeví při tomto použití nestabilitu a mnoho výstupních hodnot roste nad všechny meze. Výstup této banky tvoří cílovou vnitřní reprezentaci.

Po získání hodnot reprezentace je další krok asimilace hodnot testovaného signálu skrze pomocnou funkci `back_assim`. Ta sama (v jednom vyhodnocení pro všechny hodnoty jednoho modulačního pásma) provede operaci odpovídající vztahu 2.1 tímto výpočtem:

```
out=y+(+(y<x) .* ((x-y) ./2));
```

Pro časově nezávislý i závislý výpočet vzájemné korelace je využita pomocná funkce `back_xcorr` se vstupními daty `xtf` (reference) a `ytf` (testovaný zvuk), které mohou být jak krátké rámce, tak celé signály. Vztah 2.2 je realizován následujícím kódem:

```
coeff1=xtf-mean(xtf(:));
coeff2=ytf-mean(ytf(:));
buf1=coeff1.*coeff2;
buf2=coeff1.^2;
buf3=coeff2.^2;
rm=sum(buf1(:))/sqrt(sum(buf2(:))*sum(buf3(:)));
```

Koeficienty pro výpočet PSM ze vztahu 2.3 jsou realizovány sumou kvadrátů jednotlivých modulačních pásem (pro vyhodnocení přes funkci `cellfun` delegovanou na funkci `back_sqsum`). Vyhodnocení koeficientů a následné vynásobení a sumace (pro efektivitu realizované maticovým násobením) je realizováno následujícími dvěma řádky:

```
wm=yssum./sum(yssum);
outvalues.psm=cell2mat(rm)*wm';
```

K výpočtu PSM_t je získána „okamžitá aktivita“ skrze průměry 10 ms dlouhých rámců pomocí funkce `back_iact`. Dále jsou využity funkce vykonávající vzá-

jemnou korelaci (`back_iaq`) a sumu kvadrátů (`back_sqsumst`) s totožnými rámci pro zjištění „okamžité kvality“. Váhování a získání 5% kvantilu je provedeno seřazením hodnot krátkodobé kvality (`buf_iaq`) s přiřazenými hodnotami aktivity stejných rámců (`buf_iact`). Index, u kterého postupná integrace hodnot aktivity přesáhne 5% sumy aktivit, je výsledná hodnota.

```
fivepercent=sum(buf_iact)*0.05;
buf=sortrows([buf_iaq' buf_iact']);
integr=0;
index=1;
while (integr<(fivepercent))
    index=index+1;
    integr=integr+buf(index,2);
end
outvalues.psmt=buf(index,1);
```

Poslední krok je zjištění *ODG* realizované následujícím přepočtem:

```
ODGa=-0.22;ODGb=0.98;ODGc=-4.13;ODGd=16.4;ODGx0=0.864;
if (outvalues.psmt<ODGx0)
    outvalues.odg=max(-4,(ODGa/(outvalues.psmt-ODGb))+ODGc);
else
    outvalues.odg=ODGd*outvalues.psmt - ODGd;
end
```

3.2 Modifikovaný CASP

V rámci této práce byl rovněž implementován alternativní model sluchové cesty vycházející z modelu CASP zahrnující několik úprav převážně z modelu MAP. Pro prvky identické s modelem sluchové cesty obsaženým v PEMO-Q byl využit identický kód, avšak byl přepsán blok bazilární membrány, přidán blok vnějšího a středního ucha a provedeny další úpravy popsané v kapitole 2.1.2.

Blok vnějšího ucha byl vytvořen po vzoru modelu MAP, tedy za pomoci dvou rezonančních filtrů. Koeficienty pro filtry jsou vypočteny pomocnou funkcí `mod_res_filter` následujícím kódem, kde f_c je střední kmitočet, b_w šířka pásma a f_s vzorkovací kmitočet:

```
dt=1/fs;
q=pi*dt*bw;
```

```
tq=tan(q);  
J=1/(1+cot(q));  
K=(2*cos(2*pi*dt*fc))/((1+tq)*cos(q));  
L=(tq-1)/(tq+1);  
rfcoeffs.b=[J 0 -J];  
rfcoeffs.a=[1 -K -L];
```

Koeficienty pro transformační filtry byly získány funkcí butter zahrnutou mezi funkcemi interpretu MATLAB.

DRNL banka modelující bazilární membránu byla rovněž implementována po vzoru modelu MAP. Kmitočty a šířky pásma pro nelineární část byly využity identické jako u modelu sluchové cesty PEMO-Q. Z těch bylo možné vztahem 2.9 vypočítat kmitočty, a při zachování pravidla 1 ERB i šířky pásma, pro lineární část. Řády gammatónových filtrů jsou ve všech případech 1, dolní propusti jsou Butterworthovy filtry (sestavené funkcí butter) a v kaskádách jsou vždy 3 identické filtry. K vyhodnocení kompresního prvku je ještě nezbytné odvodit parametr b . Při znalosti mezní vstupní hodnoty $x = 10^{-9} \cdot 10^{25/20}$, parametru $a = 5 \cdot 10^3$ a $c = 0.2$ lze nalezením řešení rovnice

$$a \cdot x = b \cdot x^c \quad (3.8)$$

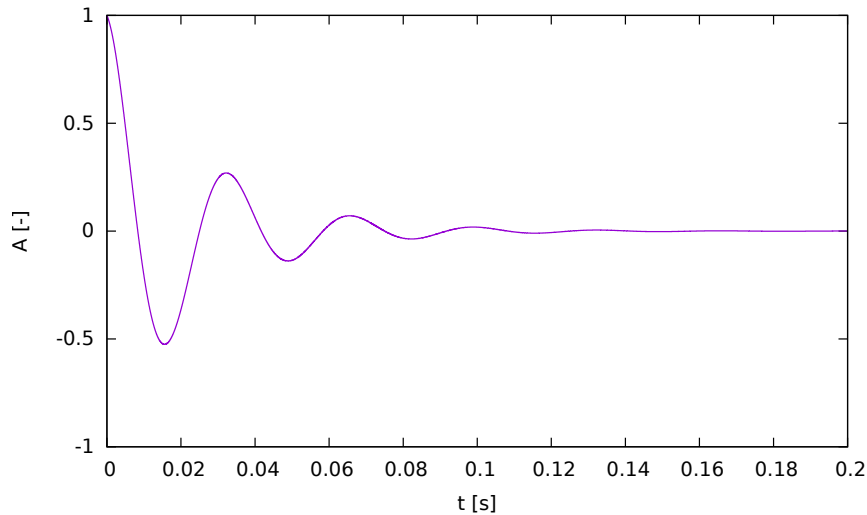
zjistit hodnotu $b = 0.0032$. Zisk lineární části je definován skalárním číslem 50.

Před stupněm adaptivní kaskády byl v porovnání s PEMO-Q přidán expanzní člen, který spočívá v zisku 38 dB a kvadrátu. Minimální hodnota pro adaptivní kaskádu byla snížena na $3 \cdot 10^{-13}$ a přidaná dolní propust o kmitočtu 150 Hz před modulační bankou byla realizována butterworthovým filtrem prvního řádu navrženou funkcí butter.

3.3 Modelované artefakty

Pro analýzu vlivu artefaktů přítomných v analogových záznamech popsaných v kapitole 2.2 na výstupní hodnoty PEMO-Q byly tyto artefakty modelovány pomocí zvláštních funkcí, které do vstupního signálu zanesou zvolenou míru modelovaného artefaktu.

Pro modelování krátkých impulzních poruch je k dispozici funkce `preproc_imp`, která do vstupního signálu aditivně vloží impulzy bílého šumu o zvolené délce a amplitudě odpovídající amplitudě původního signálu. Impulzy jsou rovnoměrně rozloženy po signálu z důvodu reprodukovatelnosti a jejich počet je definován na vstupu zvoleným počtem impulzů na sekundu zvuku.



Obrázek 6: Použitý nízkofrekvenční impulz

Krátké nízkofrekvenční impulzy jsou ve funkci `preproc_trans` modelovány harmonickou funkcí o kmitočtu 30 Hz s exponenciální obálkou o celkové délce 200 ms za pomoci těchto dvou řádků kódu (s vzorkovacím kmitočtem fs):

```
t=0:1/fs:0.2-(1/fs);
imp=exp(-40.*t).*cos(60*pi*t);
```

Impulzy jsou rovněž rovnoměrně rozložené a parametr míry je celkový počet vložených impulzů. Impulz získaný uvedeným vztahem je zobrazen na obrázku 6.

Šum je modelován funkcí `preproc_noise` a přičítá k původnímu zvuku růžový šum. Ten byl zvolen pro charakter blízký lidskému vnímání hlasitosti. Šum je vytvořen pomocí filtrace bílého šumu váhováním ve frekvenční oblasti. Váha je $1/\sqrt{m}$, kde m je index vzorku ve frekvenční oblasti. Váhy jsou vypočteny jen pro unikátní body spektra, body v druhé polovině diskretního spektra odpovídají zrcadlené verzi první poloviny pro zachování reálného charakteru signálu v časové oblasti. Kromě samotného signálu je dalším vstupním parametrem funkce amplituda šumu v dB.

Poslední modelovaný artefakt je nelineární zkreslení v podobě tvrdého limitu za pomoci funkce `preproc_limit`. Limitace samotná je realizována tímto řádkem:

```
y=sign(x).*min(abs(x),threshold);
```

Vstupním parametrem je úroveň limitace v dB relativně k úrovni signálu. Přepočtem do lineární míry je zjištěna mez použitá v uvedeném kódu.

Vliv míry uvedených artefaktů na výstupní hodnoty modelu PEMO-Q byl otestován modelováním artefaktů pomocí výše uvedených funkcí na třech referenč-

ních signálech. Prvním je přes celé slyšitelné spektrum logaritmicky rozmítaný sinusový signál o délce pěti vteřin generovaný funkcí chirp. Druhým je mužská promluva v anglickém jazyce (vzorek č. 50 z testovací kompilace SQAM [23]) a třetím je orchestr (prvních 15 vteřin SQAM vzorku 68).

Čistý referenční vzorek a testovací vzorek s modelovaným artefaktem (s normovaným mediánem výkonů prostřednictvím funkce `preproc_norm`) jsou samostatně vyhodnoceny oběma modely sluchové cesty a jsou vyhodnoceny hodnoty PSM , PSM_t a ODG . Závislosti míry ovlivnění zvuku artefaktem jsou vyneseny do grafů přítomných v příloze 1 a rozebraných v kapitole 4.1.

Pro modelování šumu byly zvoleny úrovně amplitudy šumu od -100 do 50 dB s krokem 10 dB. Krátké šumové poruchy mají stanovenou délku 1 ms a jejich četnost je od 0 do 10 poruch na vteřinu s krokem 0.5 poruch za vteřinu. Limitace probíhá po krocích 5 dB od žádné limitace až po úroveň -90 dB. Krátkých přechodných impulzů je do signálu zaneseno od 0 do 10 . Tento test je k nalezení ve skriptu `arttest.m`.

3.4 Odstranění šumu

K otestování myšlenky vytvoření umělé reference potlačením šumu byla definována funkce `nremove` implementující všechny tři metody uvedené v kapitole 2.2.1. Signál je v této funkci separován na rámce o velikosti 1024 vzorků s překryvem 50% . Z těchto rámců je zjištěna hodnota RMS, přičemž zvolený počet rámců s nejnižší hodnotou je využit k odhadu výkonového spektra šumu. Tento postup vychází z předpokladu, že šum je stacionární aditivní prvek, tudíž v částech signálu s nejnižším výkonem je nejpravděpodobnější výskyt šumu samotného.

Pro potlačení minim způsobených nestacionárním charakterem takto krátkých rámců je výkonové spektrum odhadnuto přes více rámců. Každý frekvenční vzorek výsledného výkonového spektra je vždy ten největší daného indexu mezi vybranými rámci. Počet zohledněných rámců je vhodné zvolit dle množství tichých částí ve zvuku: příliš malý počet může nedostatečně zohlednit krátkodobou nestacionaritu a příliš velký může do odhadu zahrnout i nešumové složky a zvětšit nežádoucí zkreslení zvuku.

Před výpočtem výkonových spekter rámců vybraných pro odhad je aplikováno Hannovo okno pro minimalizaci přetékání. Identické okno je aplikováno též na rámce vstupního signálu při aplikaci samotné filtrace rovněž pro potlačení přetékání a navíc pro odstranění nespojitostí k rekonstrukci typu `overlap-add` po filtraci.

Vytvoření reference je vyzkoušeno dvěma metodami. První, představená ve

skriptu `noreftest.m`, je modelová a využívá uměle zašuměné signály řeči a hudby popsané v předchozí kapitole (harmonický signál je vynechán pro nemožnost použití uvedené metody získání profilu šumu). Výstupní hodnoty PEMO-Q s využitím umělé reference jsou porovnány s těmi získanými se skutečnou referencí.

Druhá metoda je bezreferenční analýza reálných digitalizovaných analogových záznamů poskytnutých vedoucím práce a je vyhodnocena funkcí `subjtest_func` přijímající seznam souborů k analýze v textovém souboru s každým jménem souboru na zvláštním řádku. Odhad spektra šumu není zesílen a počty rámců pro odhad jsou 4 u modelových testů, 5 u analogových záznamů s hudebním obsahem a 40 u záznamů s obsahem jiným.

4 Výsledky

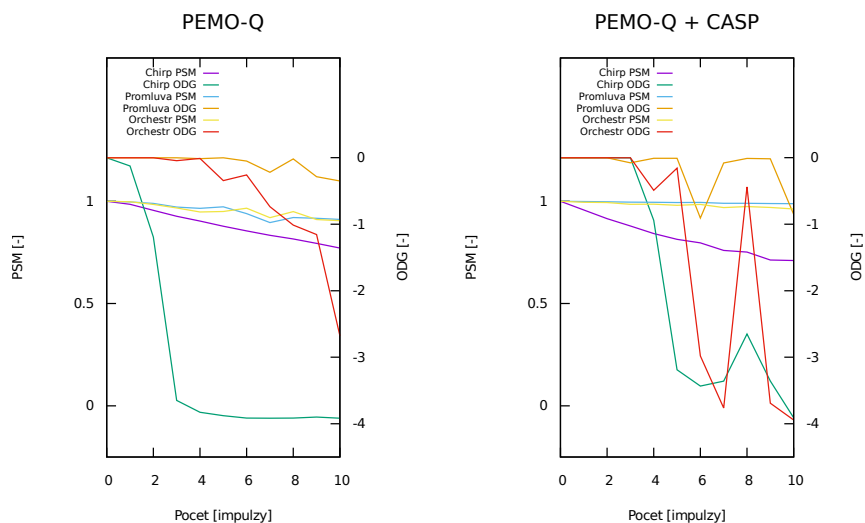
4.1 Testy modelovaných artefaktů

Vyhodnocením uvedených testů byly získány závislosti uvedené v příloze 1 s ukázkou uvedenou na obrázku 7. Zobrazeny jsou jen průběhy hodnot PSM a ODG vzhledem ke skutečnosti, že PSM_t je nosičem identické informace jako ODG . Pro model PEMO-Q byly ve všech případech dosaženy předpokládané výsledky v podobě klesajícího výstupu. V případě krátkých nízkofrekvenčních impulzů se projevila lokální maxima, což však lze přisoudit spojitostí mezi rušivostí impulzu a charakterem zvuku, ve kterém se impulz nachází. Tuto domněnku podporuje fakt, že v rozmítaném harmonickém signálu se žádný takový lokální extrém nevyskytuje.

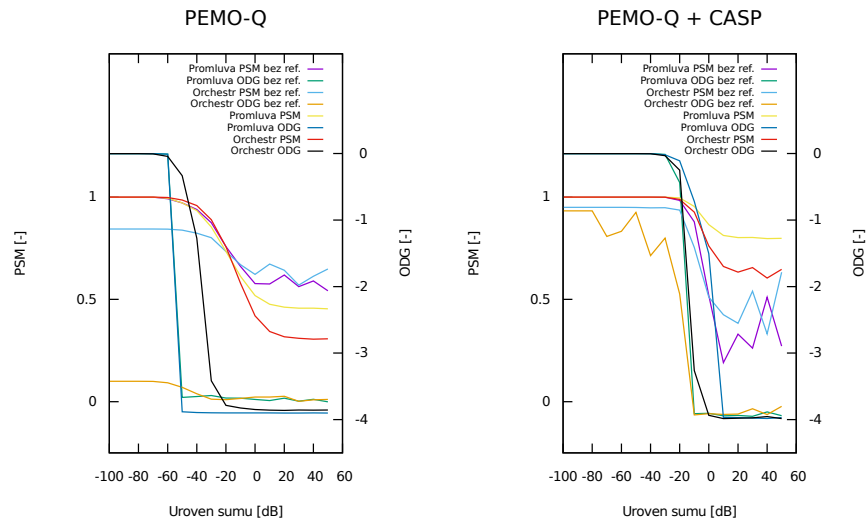
S využitím substitučního modelu sluchové cesty CASP byly splněny předpoklady jen u limitace a při zohlednění výrazně nižší citlivosti i u šumu. Všechny ostatní testy projevily výrazné lokální extrémy. Možné interpretace těchto výsledků jsou špatná citlivost modelu CASP na artefakty krátkého impulzního charakteru, nevhodné parametry modelu pro tyto artefakty či chybná implementace modelu.

4.2 Modelový bezreferenční test

Výsledky testu popsaného v kapitole 3.4 jsou zobrazeny na obrázku 8. Z těchto dat plyne, že korelace hodnot samotného PEMO-Q mezi výsledky získanými s použitím skutečné reference je relativně malá a že bezreferenční metoda podává ve většině případů výsledky lepší, než jaké jsou. Jediná výjimka je časově závislá hodnota



Obrázek 7: Výstupní závislosti hodnot PEMO-Q na počtu n.f. impulzů



Obrázek 8: Výstupní závislosti hodnot PEMO-Q na úrovni šumu s využitím skutečné i umělé reference

u vzorku orchestru, kdy už při velmi nízkých hodnotách šumu je signál vyhodnocen jako velmi rušivě zkreslený. Pravděpodobné vysvětlení tohoto výsledku je špatný odhad šumu způsobený skutečností, že hudební signál obsahuje velmi málo tichých míst.

Poměrně nevýrazné a nejednoznačné výsledky bezreferenčního testu lze vysvětlit možným nedostatečným či nesprávným potlačením šumu Wienerovým filtrem. Vzhledem ke skutečnosti, že profil šumu byl jen hrubě odhadnut, je nejasné do jaké míry jsou tyto výsledky ovlivněny zkreslením signálu filtrací.

Hodnoty získané využitím modelu CASP začínají, opět s výjimkou časové závislé hodnoty u vzorku orchestru, jevit pokles až při velmi vysokých hodnotách šumu. Poté však začínají jevit hodnoty srovnatelné či nižší oproti samotnému PEMO-Q. Hodnoty však v nezanedbatelné míře kolísají, tudíž je nelze považovat za spolehlivý zdroj informací o kvalitě.

4.3 Výsledky reálných vzorků

Získané výsledky pokusné bezreferenční analýzy mají v naprosté většině případů velmi polarizovaný charakter, že mnoho vzorků je hodnoceno jako buď velmi kvalitní s artefakty výrazně pod hranicí slyšitelnosti, anebo velmi rušivě zkreslené. Vzhledem ke skutečnosti, že mnoho negativních výsledků pochází z hudebních vzorků, lze usuzovat, že se často nepodařilo získat vhodný profil šumu.

Tyto fakta, spolu s poměrně nejednoznačnými výsledky předchozího testu, znehodnocují možnost použití bezreferenčního testu v aktuální podobě. Mezi vý-

sledky hodnoty PSM pro samotné PEMO-Q se rovněž objevila řada hodnot „not-a-number“. Pravděpodobné vysvětlení je v nestabilitě adaptivní kaskády, která není ošetřena jako např. hrubým popisem kontroly, který zmínil Harlander *et al.* [6]. Hodnoty jdoucí nad všechny meze vyhodnocené jako nekonečně velké mohou totiž v bloku vzájemné korelace mít uvedený výstup.

Data tohoto testu jsou k dispozici k nahlédnutí v souboru `analog_vysledky.csv` umístěném v příloze 2.

5 Závěr

Z výsledků získaných v kapitole 4.1 lze vyvodit, že implementovaný model objektivního hodnocení PEMO-Q má potenciál pro kvalitativní zhodnocení analogových nahrávek za přítomnosti dostatečně kvalitní reference. Jestli je však vyhodnocená míra rušivosti těchto artefaktů blízká lidskému vnímání není možné určit bez vyhodnocení subjektivních testů se stejnými vzorky.

Substituční model sluchové cesty CASP ve většině případů nepodal žádné přesvědčivé údaje. Je však otázkou, jestli by nebylo možné nalézt pro tyto účely vhodnou úpravu, že parametry tohoto modelu mají mnoho stupňů volnosti a některé jeho bloky mají několik alternativních verzí. V aktuální verzi v příloze této práce však jeho použití k analýze nelze doporučit, kromě výsledků také z důvodu větší složitosti a výpočetní náročnosti v porovnání s modelem sluchové cesty zahrnutým v PEMO-Q.

Co se modifikací přiložených modelů týče, je určitě vhodné dále upravit adaptivní kaskádu modelu PEMO-Q pro zajištění stability, aby výstup modelu nemohl potenciálně obsahovat hodnotu „not-a-number“.

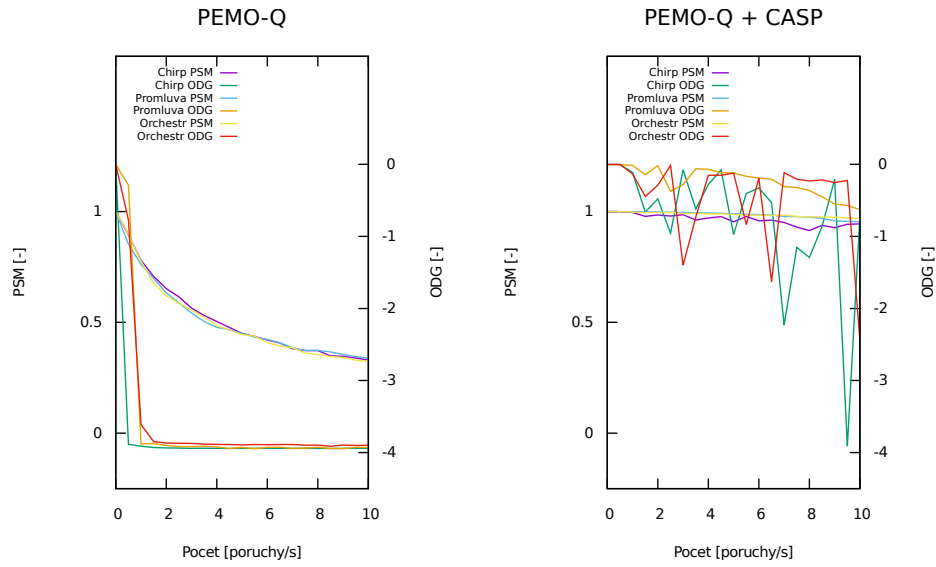
Vzhledem k nepřesným odhadům spektrálních výkonů šumu a neznámému vlivu zkreslení Wienerovým filtrem na výstupy hodnotících modelů i při relativně dobrém odhadu šumového spektra nelze doporučit ani navrženou bezreferenční metodu. Kvůli skutečnosti, že metody rekonstrukce se dají považovat za formu zkreslení snažící se aproximovat neznámý výsledek, je otázkou, jestli metody na tomto principu mají vůbec potenciál být užitečným způsobem bezreferenčního hodnocení kvality.

6 Zdroje

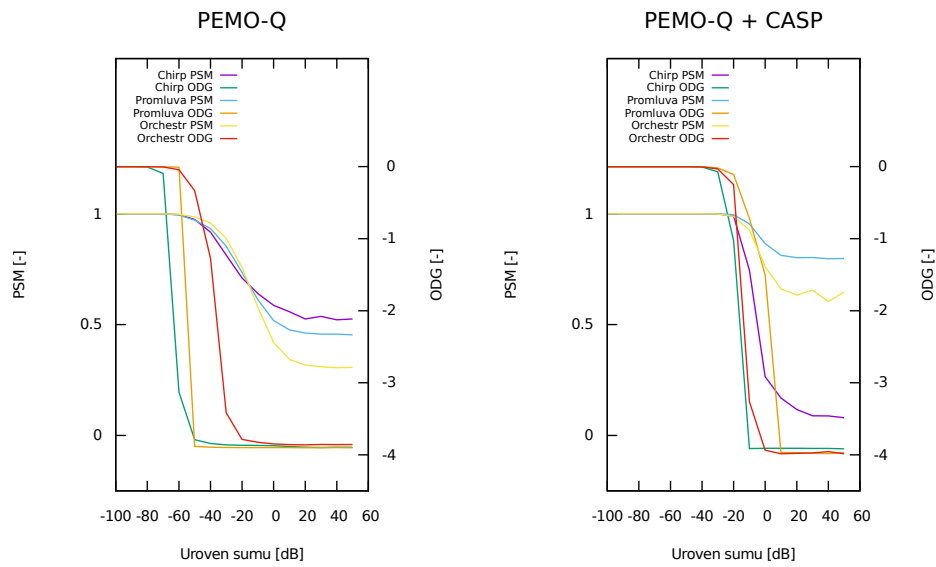
- [1] Torsten Dau. *Modeling auditory processing of amplitude modulation*. BIS Verlag, 1999.
- [2] Brian R Glasberg and Brian CJ Moore. Derivation of auditory filter shapes from notched-noise data. *Hearing research*, 47(1):103–138, 1990.
- [3] Simon Godsill, Peter Rayner, and Olivier Cappé. *Digital audio restoration*. Springer, 2002.
- [4] SJ Godsill and CH Tan. Removal of low frequency transient noise from old recordings using model-based signal separation techniques. *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on*, pages 4–pp, 1997.
- [5] Richard L Goode, Mead Killion, Koshiro Nakamura, and Shinsei Nishihara. New knowledge about the function of the human middle ear: development of an improved analog model. *Otology & Neurotology*, 15(2):145–154, 1994.
- [6] Niklas Harlander, Rainer Huber, and Stephan D. Ewert. Sound quality assessment using auditory models. *J. Audio Eng. Soc*, 62(5):324–336, 2014.
- [7] E. Hogenauer. An economical class of digital filters for decimation and interpolation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(2):155–162, Apr 1981.
- [8] Rainer Huber and Birger Kollmeier. PEMO-Q - a new method for objective audio quality assessment using a model of auditory perception. *IEEE Transactions on Audio, Speech & Language Processing*, 14(6):1902–1911, 2006.
- [9] ITU. *ITU-R Recommendation BS.1116-1: Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, 1997.
- [10] ITU. *ITU-R Recommendation BS.1387-0: Method for Objective Measurements of Perceived Audio Quality*, 1998.
- [11] ITU. *ITU-T Recommendation P.861: Methods for Objective and Subjective Assessment of Quality: Objective Quality Measurement of Telephone-band (300-3400 Hz) Speech Codecs*, 1998.
- [12] ITU. *ITU-T Recommendation P.862: Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, 2001.

-
- [13] ITU. *ITU-T Recommendation P.863: Methods for Objective and Subjective Assessment of Speech Quality: Perceptual Objective Listening Quality Assessment*, 2011.
- [14] ITU. *ITU-R Recommendation BS.1116-3: Methods for the Subjective Assessment of Small Impairments in Audio Systems*, 2015.
- [15] ML Jepsen, Stephan D. Ewert, and Torsten Dau. A computational model of human auditory signal processing and perception. *Journal of the Acoustical Society of America*, 124:422–438, 2008.
- [16] Soendergaard. Peter L. Computational auditory signal processing, 2013.
- [17] Hearing Research Lab. *Matlab Auditory Periphery (MAP), Model technical description*. Essex, 2011.
- [18] Enrique A Lopez-Poveda and Ray Meddis. A human nonlinear cochlear filterbank. *The Journal of the Acoustical Society of America*, 110(6):3107–3118, 2001.
- [19] Ray Meddis, Lowell P O’Mard, and Enrique A Lopez-Poveda. A computational algorithm for computing nonlinear auditory frequency selectivity. *The Journal of the Acoustical Society of America*, 109(6):2852–2861, 2001.
- [20] Danièle Pralong and Simon Carlile. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *The Journal of the Acoustical Society of America*, 100(6):3785–3793, 1996.
- [21] Antony W Rix, John G Beerends, Doh-Suk Kim, Peter Kroon, and Oded Ghizta. Objective assessment of speech and audio quality; technology and applications. *Audio, Speech, and Language Processing, IEEE Transactions on*, 14(6):1890–1901, 2006.
- [22] Malcolm Slaney et al. An efficient implementation of the patterson-holdsworth auditory filter bank. *Apple Computer, Perception Group, Tech. Rep*, 35:8, 1993.
- [23] European Broadcasting Union. EBU SQAM CD - sound quality assessment material recordings for subjective tests, 2008.
- [24] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series*, volume 2. MIT press Cambridge, MA, 1949.
- [25] Udo Zoelzer. *DAFX: Digital Audio Effects*. John Wiley & Sons, Inc., New York, NY, USA, 2002.

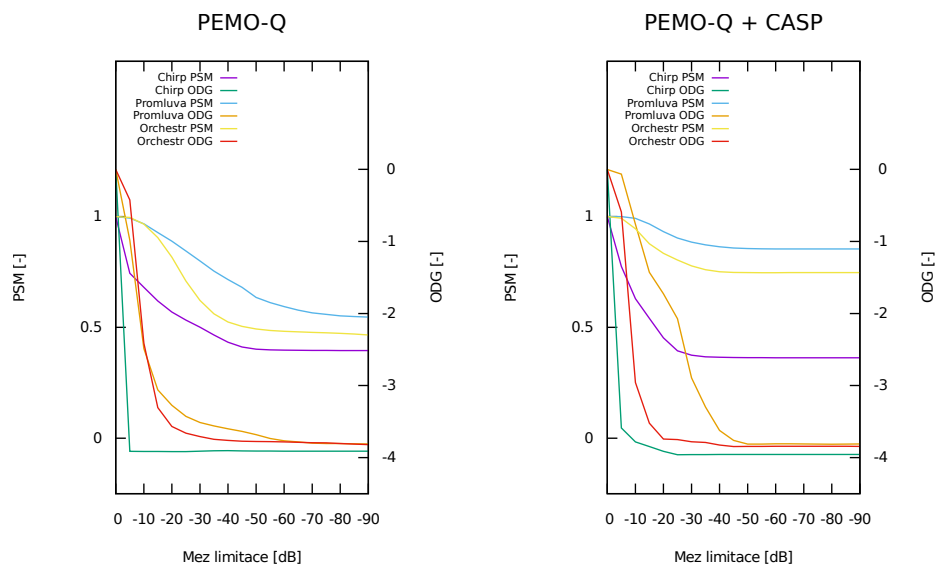
Příloha 1: Výstupní závislosti modelů



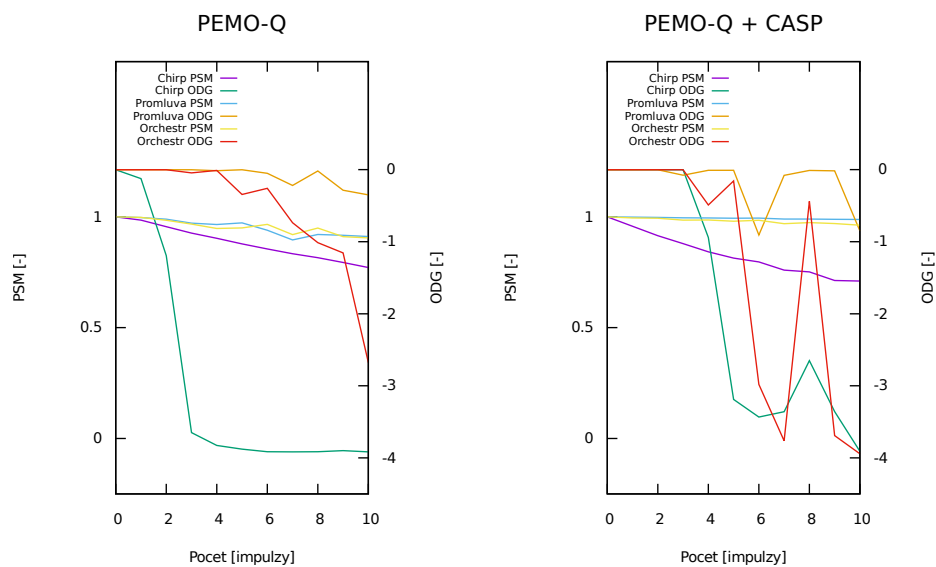
Obrázek 1: Výstupní závislosti krátkých poruch



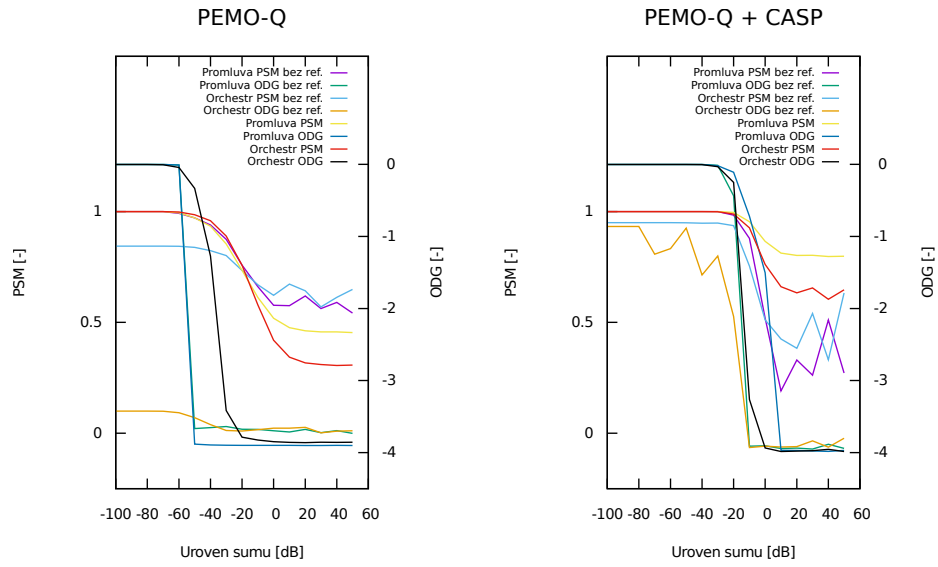
Obrázek 2: Výstupní závislosti šumu



Obrázek 3: Výstupní závislosti limitace



Obrázek 4: Výstupní závislosti přechodného n.f. impulsu



Obrázek 5: Výstupní závislosti šumu v testu s umělou referencí