



# 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

## RESTAUROVÁNÍ CHYBĚJÍCÍCH VZORKŮ ZVUKOVÉHO SIGNÁLU S POUŽITÍM PSYCHOAKUSTICKÉHO MODELU

RESTORATION OF MISSING AUDIO SIGNAL SAMPLES USING A PSYCHOACOUSTIC MODEL

### BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

### AUTOR PRÁCE

AUTHOR

Michal Švento

### VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Ondřej Mokrý

BRNO 2021



# Bakalářská práce

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

**Student:** Michal Švento

**ID:** 212584

**Ročník:** 3

**Akademický rok:** 2020/21

## NÁZEV TÉMATU:

### Restaurování chybějících vzorků zvukového signálu s použitím psychoakustického modelu

#### POKYNY PRO VYPRACOVÁNÍ:

Seznamte se s modelem znehodnocení zvukového signálu při výpadku vzorků. Analyzujte tento problém a seznamte se s vybranými metodami pro jeho řešení s důrazem na metody založené na řídkosti signálů. Zvolené metody doplňte o využití psychoakustického modelu, implementujte v Matlabu a otestujte na simulovaných i reálných nahrávkách. Výstupem budou výsledky objektivního srovnání (SNR, PEMO-Q) metod a také subjektivní hodnocení studenta.

#### DOPORUČENÁ LITERATURA:

[1] RAJMIC, Pavel a Marie DAŇKOVÁ. Úvod do řídkých reprezentací signálů a komprimovaného snímání. Brno: Vysoké učení technické v Brně, 2014. ISBN: 978-80-214-5169- 8.

[2] ZÁVIŠKA, Pavel, Pavel RAJMIC a Jiří SCHIMMEL. Psychoacoustically Motivated Audio Declipping Based on Weighted l1 Minimization. In: 2019 42nd International Conference on Telecommunications and Signal Processing (TSP). IEEE, 2019. Dostupné z: <https://ieeexplore.ieee.org/document/8769109/>

**Termín zadání:** 1.2.2021

**Termín odevzdání:** 31.5.2021

**Vedoucí práce:** Ing. Ondřej Mokrý

**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

Táto bakalárska práca sa zaoberá rekonštrukciou krátkeho úseku poškodeného zvukového signálu. Signál reprezentujeme ako riedky vektor pomocou diskkrétnej Gaborovej transformácie. Na rekonštrukciu sú využité nástroje konvexnej optimalizácie. Je riešená optimalizačná úloha pomocou Douglasovho–Rachfordovho a Chambollovho–Pockovho algoritmu. Do algoritmov je zapojený psychoakustický model, ktorý by mal zlepšiť objektívne parametre zrekonštruovaného signálu. Porovnanie je realizované objektívnou metódou SDR, PEMO-Q a taktiež subjektívne.

## KLÚČOVÉ SLOVÁ

Chambolov–Pockov algoritmus, Douglasov–Rachfordov algoritmus, Gaborova transformácia, konvexná optimalizácia, l1 minimalizácia, proximálne operátory, psychoakustický model

## ABSTRACT

This bachelor thesis deals with the reconstruction of short-time damaged audio signal. The signal is represented by sparse signal representation using discrete Gabor transform. Convex optimization tools are used for the reconstruction. The optimization problem is solved using the Douglas–Rachford and Chambolle–Pock algorithm. Psychoacoustic model is involved in algorithm to obtain better results in objective metrics. The comparison is realised by an objective method SDR, PEMO-Q and also subjectively.

## KEYWORDS

Chambolle–Pock algorithm, Douglas–Rachford algorithm, Gabor transform, convex optimization, l1 minimalization, proximal operators, psychoacoustic model

ŠVENTO, Michal. *Restaurování chybějících vzorků zvukového signálu s použitím psychoakustického modelu*. Brno, 2021, 46 s. Bakalárska práca. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav telekomunikací. Vedúci práce: Ing. Ondřej Mokřý

## VYHLÁSENIE

Vyhlasujem, že svoju bakalársku prácu na tému „Restaurování chybějících vzorků zvukového signálu s použitím psychoakustického modelu“ som vypracoval samostatne pod vedením vedúceho bakalárskej práce, s využitím odbornej literatúry a ďalších informačných zdrojov, ktoré sú všetky citované v práci a uvedené v zozname literatúry na konci práce.

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

Brno .....

.....

podpis autora

## POĎAKOVANIE

Chcel by som poďakovať môjmu vedúcemu pánovi Ing. Ondřejovi Mokrému za užitočné pripomienky, rady a ochotu vždy pomôcť. Rovnako ďakujem svojej rodine za trpezlivosť pri písaní tejto práce.

# Obsah

Úvod	9
<b>1 Riedka reprezentácia signálu</b>	<b>11</b>
1.1 Norma vektora	11
1.2 Diskrétna Gaborova transformácia	11
1.2.1 Parametre DGT	12
1.3 Definovanie problému	12
<b>2 Aparát používaný na rekonštrukciu</b>	<b>15</b>
2.1 Konvexná optimalizácia	15
2.1.1 Proximálne algoritmy	16
2.1.2 Proximálne operátory	16
2.1.3 Douglasov–Rachfordov algoritmus	18
2.1.4 Chambollov–Pockov algoritmus	20
<b>3 Psychoakustický model ľudského sluchu</b>	<b>21</b>
3.1 ATH	21
3.2 Kmitočtové maskovanie	22
3.2.1 Simultánne maskovanie a maskovacia nesymetria	22
3.2.2 Nesimultánne maskovanie	23
3.3 Váhovanie koeficientami	24
3.3.1 Váhové vektory vychádzajúce z ATH	24
3.3.2 Váhové vektory vychádzajúce z krivky rovnakej hlasitosti	24
<b>4 Experimenty</b>	<b>27</b>
4.1 Metódy hodnotenia kvality	27
4.1.1 Double-blind triple-stimulus with hidden reference	28
4.1.2 Pomer signálu k poškodeniu	28
4.1.3 PEMO-Q	29
4.2 Rekonštrukčný algoritmus	29
4.2.1 Banka signálov a ich úprava	29
4.2.2 Vstupný signál do algoritmov	31
4.2.3 Overenie funkčnosti algoritmov	32
4.2.4 Nastavenie algoritmov	34
4.2.5 Výsledky algoritmov	35
4.2.6 Porovnanie syntetizovaných signálov	36
4.2.7 Porovnanie reálnych signálov	38

<b>5 Záver</b>	<b>40</b>
<b>Literatúra</b>	<b>41</b>
<b>Zoznam symbolov, veličín a skratiek</b>	<b>45</b>
<b>A Zoznam príloh</b>	<b>46</b>

# Zoznam obrázkov

1.1	Spektrogram akustickej gitary . . . . .	12
1.2	Hannovo okno . . . . .	13
1.3	Absolútna hodnota $c$ pre signál s modifikáciami . . . . .	14
2.1	Operátory mäkké prahovanie a clip . . . . .	18
3.1	Prah počuteľnosti ľudského sluchu . . . . .	22
3.2	Barkova stupnica . . . . .	23
3.3	Normalizované váhové funkcie ATH . . . . .	25
3.4	Váhové funkcie – A, B, C, Z . . . . .	26
4.1	Stupnica ITU-R . . . . .	28
4.2	Syntetizované signály . . . . .	30
4.3	Fade in a fade out . . . . .	31
4.4	Hannove okná na skrátrenom signáli . . . . .	32
4.5	Umiestenie skráteneho úseku vstupujúceho do algoritmu . . . . .	32
4.6	Umelo poškodený testovací signál . . . . .	33
4.7	Závislosť $\ell_1$ normy na <i>MaxIte</i> , DRA . . . . .	34
4.8	Závislosť $\ell_1$ normy na <i>MaxIte</i> , CPA . . . . .	34
4.9	Spektrogram výseku s rôznymi $w$ . . . . .	35
4.10	DRA syntetizované signály . . . . .	36
4.11	CPA syntetizované signály . . . . .	37
4.12	Signál č.2, DRA, ATH1, diera 50 ms . . . . .	37
4.13	Signál č.3, CPA, ATH3, diera 50 ms . . . . .	38
4.14	DRA reálne signály . . . . .	38
4.15	CPA reálne signály . . . . .	39
4.16	Pokles energie u reálneho signálu . . . . .	39



# Úvod

Spolu s vývojom vytvárania zvukových nahrávok sa objavila potreba opraviť existujúci signál, ktorý je poškodený. Príkladov z bežného života je veľké množstvo, napríklad poškodenie nosiča, výpadok pri prenose signálu či zámerné vynechanie vzoriek pri kompresii. Cieľom pri rekonštrukcii je zachovať vzorky, ktoré sú nepoškodené a na základe informácií a priori známych doplniť chýbajúce úseky.

Veľkým pokrokom v oblasti spracovania signálov bolo zistenie, že reálne signály sú po vhodnej transformácii približne riedke. Tieto vlastnosti boli pre rekonštrukciu signálu, konkrétne pre doplnenie chýbajúcich vzoriek, využité v článku [1].

Algoritmov používaných pri rekonštrukcii poškodeného signálu je veľké množstvo. K nim patria aj algoritmy postavené na konvexnej optimalizácii, akými sú napríklad: ADMM (Alternating Directions Method of Multipliers) [2], Douglasov–Rachfordov algoritmus (DRA) [3]. Algoritmom nevyužívajúcim konvexnú optimalizáciu je napríklad OMP (Orthogonal Matching Pursuit), ktorý bol spomínaný a využitý v článku [1] a v oblasti rekonštrukcie audiosignálu bol prelomový. Z oblasti nekonvexného prístupu ešte poznáme algoritmus SPADE (the SParse Audio DEclipper) [4], ktorý rieši príbuznú úlohu declipping. Najnovšie algoritmy zastupuje napríklad algoritmus SPAIN (SParse Audio INpainter) [5]. My v tejto práci využijeme konvexný prístup.

V prvej kapitole je popísaný krátky úvod do noriem vektorov a definovaná riedkosť vektora. Malý úsek venujeme časovo-frekvenčnej analýze a syntéze, na ktorú sme zvolili diskretnú Gaborovu transformáciu. V závere kapitoly zdefiniujeme optimalizačnú úlohu, ktorú budeme riešiť.

V druhej kapitole rozoberieme riešenie optimalizačnej úlohy. Riešime ju konvexnou optimalizáciou pomocou  $\ell_1$  minimalizácie. Realizujeme analyzujúci aj syntetizujúci model, ktorý zahŕňa využitie proximálnych operátorov, konkrétne projekcie na konvexnú množinu a mäkkého prahovania. Na konci kapitoly sú popísané algoritmy: Douglasov–Rachfordov algoritmus (syntetizujúci model), Chambollov–Pockov algoritmus (analyzujúci model).

Tretia kapitola je krátkym zhrnutím poznatkov o ľudskom sluchu a jeho psychoakustickom modeli. V závere sú uvedené postupy, akými chceme rekonštrukciu vylepšiť s využitím poznatkov z tejto kapitoly.

Posledná kapitola je rozdelená na dve časti. Prvá časť predstavuje používané subjektívne aj objektívne metódy na hodnotenie kvality zvukového signálu. V druhej časti je popísaný postup rekonštrukcie a diskusia o rekonštruovaných signáloch na základe metód z prvej časti tejto kapitoly a taktiež subjektívneho vnemu poslucháča.

Cieľom tejto práce je objasniť problematiku riedkej reprezentácie signálu využívaného na rekonštrukciu poškodeného signálu. Taktiež chceme poukázať na základy konvexnej optimalizácie, ktorá sa využíva k minimalizácii vektora. V praktickej časti chceme vytvoriť funkčný algoritmus v prostredí Matlab, ktorý nahradí poškodený signál na krátkom úseku čo najpodobnejšími vzorkami s využitím konvexnej optimalizácie.

# 1 Riedka reprezentácia signálu

V tejto kapitole budú ozrejmene základné pojmy pri práci s vektormi a vysvetlené použitie riedkej reprezentácie signálu. Na obrázkoch popíšeme dôvody použitia riedkej reprezentácie pre rekonštrukciu. Na záver kapitoly odvodíme tvar optimalizačnej úlohy.

## 1.1 Norma vektora

Norma vektora je číslo, ktoré vyjadruje jeho veľkosť. Normu  $\ell_p$  definujeme ako:

$$\|\mathbf{c}\|_p := \left( \sum_{i=0}^N |c_i|^p \right)^{1/p} \quad \text{pre } 1 \leq p < \infty, \quad (1.1)$$

kde  $\mathbf{c}$  je (reálny alebo komplexný) vektor dĺžky  $N$ ,  $\mathbf{c} = [c_1, c_2, \dots, c_N]$ . Najznámejšou je norma euklidovská ( $\ell_2$ -norma), ale pre našu prácu sú dôležité  $\ell_1$ -norma a  $\ell_0$ -norma, ktorá podľa (1.1) normou nie je, pretože nespĺňa vlastnosť normy  $\|\alpha \cdot \mathbf{c}\| = |\alpha| \cdot \|\mathbf{c}\|$ , kde  $\alpha$  je konštanta a  $\mathbf{c}$  je vektor. Pre zjednodušenie označenia ju však budeme tiež nazývať normou.

Pod hodnotou  $\ell_0$ -normy myslíme počet nenulových zložiek vektora,  $\ell_1$ -norma je súčtom absolútnych hodnôt prvkov vektora ( $\|\mathbf{c}\|_1 = \sum_i |c_i|$ ) a pod euklidovskou normou ( $\ell_2$ , toto označenie sa však nepoužíva) myslíme  $\|\mathbf{c}\|_2 = \sqrt{\sum_i |c_i|^2}$  [2].

**Definícia 1** Vektor  $\mathbf{c} \in \mathbb{C}^N$  nazývame  $k$ -riedkym ak platí:

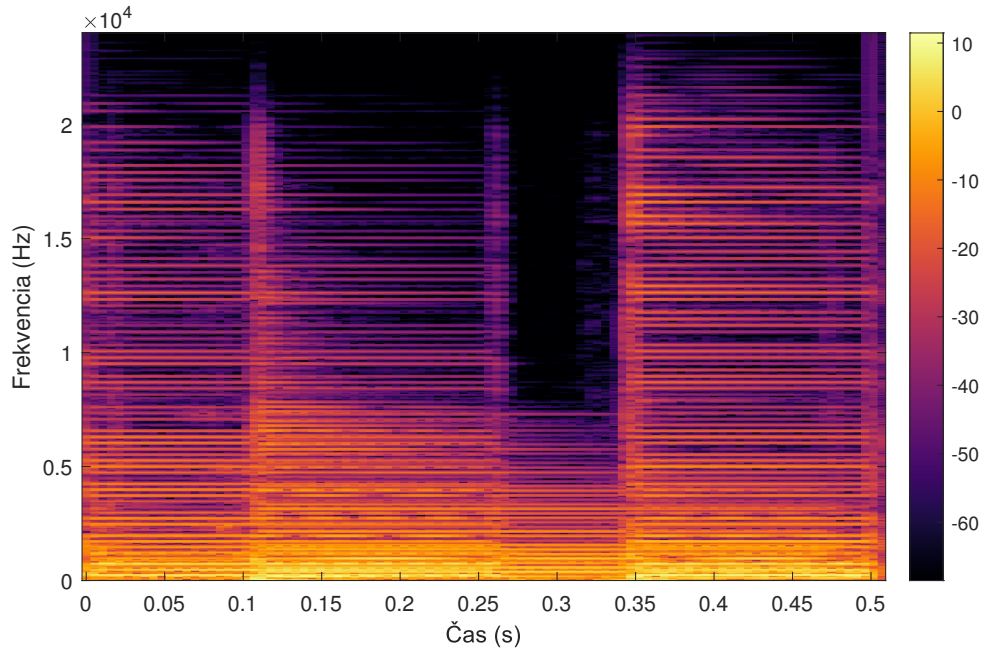
$$\|\mathbf{c}\|_0 \leq k. \quad (1.2)$$

To znamená, že vektor  $\mathbf{c}$  nazývame  $k$ -riedkym, pokiaľ obsahuje najviac  $k$  nenulových zložiek.

## 1.2 Diskrétna Gaborova transformácia

Jedným zo základných parametrov zvukových signálov je frekvencia. Pomocou Fourierovej transformácie dokážeme analyzovať signál a popísať modullovo-frekvenčnú charakteristiku. Fourierova transformácia nesie veľké množstvo informácií, ale nevie tieto informácie (Fourierove koeficienty) prideliť do časovej oblasti, teda vytvorí časovo-frekvenčnú analýzu. K tomu nám slúži diskrétna Gaborova transformácia (DGT), tiež známa ako krátkodobá Fourierova transformácia (STFT).

Výpočet DGT prebieha, zjednodušene povedané, ako posun zvoleného okna, v ktorom sa vypočíta rýchla Fourierova transformácia (FFT). Z toho vyplýva vlastnosť DGT, že počet vypočítaných koeficientov môže byť vyšší ako počet vzoriek signálu, na ktorom ju počítame. Výsledkom je matica koeficientov s veľkosťou  $m \times n$ ,



Obr. 1.1: Spektrogram akustickej gitary.

kde  $m$  je dĺžka FFT a  $n$  je počet posunov okna. Mohli by sme povedať, že  $n$  je podiel počtu vzoriek signálu  $\mathbf{y}$  vzhľadom k dĺžke časového posunu  $a$  zaokrúhlený smerom nahor. Vykreslením absolútnych hodnôt komplexných koeficientov do grafu získavame spektrogram, viď príklad 1.1.

### 1.2.1 Parametre DGT

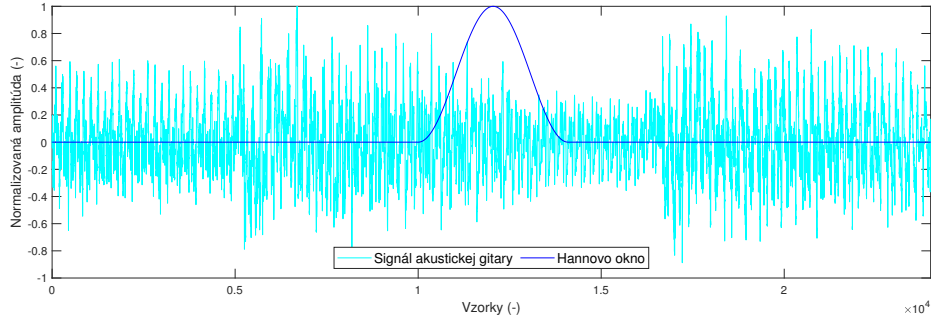
Základné vstupné parametre DGT sú:

- funkcia okna  $g$  a jeho dĺžka  $w$ . Typ použitého okna  $g$ , napr. Hannove, Hammingove, Blackmanove, Gaussove a ďalšie,
- dĺžka časového posunu  $a$ ,
- dĺžka FFT pre každé okno  $m$ .

V obrázku 1.2 je príklad, ako môže vyzeráť okno Gaborovej transformácie.

## 1.3 Definovanie problému

Máme signál  $\mathbf{x} = \{x_1, \dots, x_N\}$ , ktorý je vzorkovaný vzorkovacou frekvenciou  $f_{vz}$ , kde  $N$  predstavuje index poslednej vzorky nášho signálu, teda aj celkový počet vzoriek. Zo signálu  $\mathbf{x}$  vyberieme krátky úsek s fixným počtom vzoriek, v ktorom simulujeme výpadok alebo poškodenie signálu. Signál  $\mathbf{x}$  je nám a priori známy a je



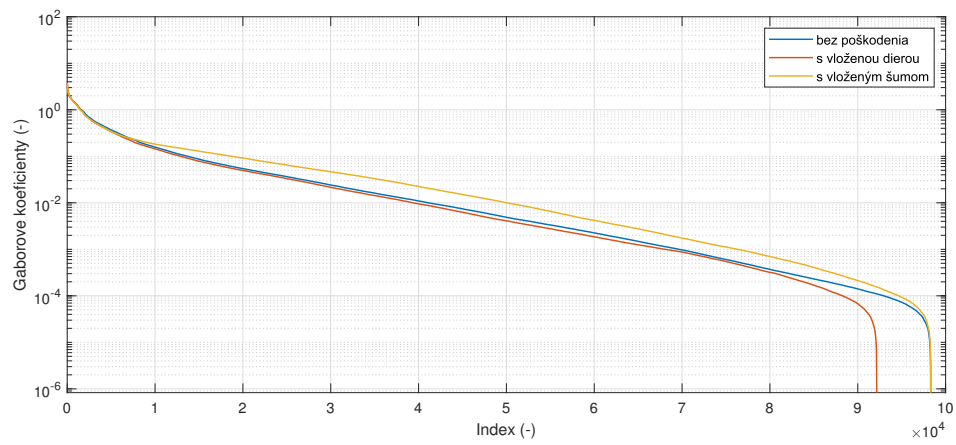
Obr. 1.2: Hannovo okno s dĺžkou  $w = 4096$ . Toto okno sa posúva doprava (v smere času) o hodnotu  $a$ , v oblasti okna vypočíta FFT s dĺžkou  $m$ .

nepoškodený. Nech pozorovaný (poškodený) signál  $\mathbf{y}$  má dĺžku  $N$ , môžeme ho popísať rovnicou:

$$y_n = \begin{cases} x_n & \text{pre } n \in R, \\ 0 & \text{pre } n \in M. \end{cases} \quad (1.3)$$

Index  $n$  popisuje  $n$ -tú vzorku signálu u signálov  $\mathbf{x}$  a  $\mathbf{y}$ . Podľa rovnice (1.3) definujeme dve množiny  $R$  a  $M$ , pre ktoré platí  $R \cup M = \{1, \dots, N\}$ . Signál rozdeľujeme na množinu hodnoverných (reliable) vzoriek a neznámych (missing) vzoriek. Signál  $\mathbf{x}$  obsahuje nepoškodené vzorky. Množina  $M$  obsahuje vzorky, o ktorých nevieme skoro žiadne informácie. Vzorky množiny  $M$  môžu chýbať, avšak môže signál obsahovať lupnutie, a by tieto vzorky boli omnoho väčšie ako vzorky z množiny  $R$ . Keďže je však poškodenie signálu značné, tieto vzorky nemožeme použiť. Preto, pre zjednodušenie, všetky prvky označujeme súhrnne za neznáme. Nastavujeme ich v základe ako nulové.

Našou úlohou je, na základe predpokladu, že tento signál je riedky, dopočítať chýbajúce vzorky. Tento predpoklad by nám mal zaručiť, že doplnený signál bude podobný predtým nepoškodeným vzorkám. Na obrázku 1.3 vidíme porovnanie absolútnych hodnôt koeficientov toho istého signálu, ktorý sme následne upravili vložením diery a šumu. Tieto koeficienty sme zoradili zostupne podľa absolútnej hodnoty, to znamená od najvýznamnejších (rádovo  $10^0$ ) po málo významné (blížiace sa nule, rádovo  $10^{-5}$ ). Zdanlivo najmenej významných koeficientov (podľa absolútnej hodnoty) má signál s vloženou dierou. Posledná časť zoradených koeficientov je skutočne nulová, pretože sme v časovej oblasti vložili do signálu nuly. Analýzou tohoto úseku dostaneme taktiež nulové koeficienty. Signál so šumom je menej riedky, pretože šum pokrýva celú frekvenčnú oblasť rovnomerne, a to sa prejavilo vyššími koeficientmi. Nepoškodený signál je redší ako signál s vloženým šumom.



Obr. 1.3: Absolútna hodnota Gaborových koeficientov zoradených od najvýznamnejšieho pre nepoškodený signál s dĺžkou 500 ms, toho istého signálu s vloženou dierou (50 ms) a signálu s vloženým šumom(50 ms).

## 2 Aparát používaný na rekonštrukciu

V posledných dekádach sa pokrok v spracovaní digitálnych signálov zaoberal riedkou reprezentáciou signálu. Reprezentácia signálu sa dá definovať pomocou sústavy rovníc:

$$\mathbf{y} = \mathbf{D}\mathbf{c}. \quad (2.1)$$

V sústave rovníc vystupuje matica  $\mathbf{D}$ , ktorá je často označovaná ako slovník (Dictionary) a  $k$ -riedky vektor  $\mathbf{c}$ , ktorý predstavuje vektor Gaborových koeficientov signálu  $\mathbf{y}$ . Parameter  $k$ , podľa (1.2), určuje najvyšší možný počet nenulových prvkov vektora  $\mathbf{c}$ . Gaborove koeficienty reálneho signálu ale nebývajú doslova riedke. Namiesto nulových koeficientov sú Gaborove koeficienty veľmi malé, blížiac sa k nule. Na obrázku 1.3 vidíme zoradené absolútne hodnoty koeficientov DGT na príklade reálneho signálu akustickej gitary (bez poškodenia). Os koeficientov je v logaritmickej mierke. Môžeme povedať, že len necelých 2000 koeficientov je v absolútnej hodnote väčších ako 1, čo predstavuje 1,6 % zo všetkých. Koeficienty väčšie ako  $10^{-2}$  a menšie ako 1 pokrývajú 40 % koeficientov a necelých 60 % sú koeficienty menšie ako  $10^{-2}$ . Na tomto vybranom signáli môžeme potvrdiť predpoklad relatívnej riedkosti zvukových signálov.

Súčinom v (2.1) je výsledný signál  $\mathbf{y}$ . Počet riešení pre hľadaný vektor  $\mathbf{c}$  v tejto sústave bude nekonečný. Preto je potrebné klásť požiadavky na slovníkovú maticu  $\mathbf{D}$  a riedkosť vektora  $\mathbf{c}$ . Všetky  $\mathbf{c}$ , ktoré spĺňajú  $\mathbf{D}\mathbf{c}(R) = \mathbf{y}(R)$  nazývame prípustné riešenia [2].

Našou požiadavkou je, aby vektor  $\mathbf{c}$  bol čo najredší. Preto definujeme optimalizačnú úlohu:

$$\arg \min_{\mathbf{c}} \|\mathbf{c}\|_0 \text{ vzhľadom ku } \mathbf{D}\mathbf{c}(R) = \mathbf{y}(R). \quad (P0)$$

Riešením (P0) sú riedke koeficienty. Rekonštruovaný signál získame aplikáciou  $\mathbf{D}$ , čiže inverznej Gaborovej transformácie. Zápisom  $\mathbf{D}\mathbf{c}(R)$  myslíme, že sa najprv vypočíta syntéza  $\mathbf{D}\mathbf{c}$ , a potom sa vezmú vzorky na pozíciách  $R$ .

### 2.1 Konvexná optimalizácia

Riešenie problému (P0) je výpočtovo zložitý, tzv. nedeterministicky polynomiálne ťažké (NP-hard). Je preto potrebné ustúpiť od požiadavky presného riešenia a hľadať riešenie aproximatívne, teda približné. K tomu nám slúži konvexná optimalizácia.

Keďže norma  $\ell_0$  nie je konvexná funkcia, nemôžeme ju použiť na riešenie metódy konvexnej optimalizácie, a preto musíme preformulovať našu minimalizačnú úlohu.

Norma  $\ell_p$ , podľa rovnice (1.1), je konvexná pre  $p \geq 1$  [2], preto riešime úlohu mierne pozmenenú, ale vo výsledku veľmi podobnú s podstatne nižšími nárokmi na výpočet:

$$\arg \min_{\mathbf{c}} \|\mathbf{c}\|_1 \text{ vzhľadom ku } \mathbf{Dc}(R) = \mathbf{y}(R). \quad (\text{P1})$$

### 2.1.1 Proximálne algoritmy

Jedným z možných prístupov konvexnej optimalizácie sú proximálne (iteračné) algoritmy založené na proximálnych operátoroch. Budeme hľadať riešenie konvexnej úlohy bez obmedzujúcej podmienky:

$$\arg \min_{\mathbf{x}} g(\mathbf{x}) + h(\mathbf{x}). \quad (2.2)$$

Úloha (P1) ale požiadavku neobmedzujúcej podmienky v danom tvare nespĺňa, preto ju formálne zapíšeme:

$$\arg \min_{\mathbf{c}} \|\mathbf{c}\|_1 \text{ vzhľadom ku } \mathbf{c} \in \Gamma, \quad (2.3)$$

v ktorej množina  $\Gamma$  je konvexná množina prípustných riešení:

$$\Gamma = \{\mathbf{c} \mid \mathbf{Dc}(R) = \mathbf{y}(R)\}. \quad (2.4)$$

Aby bola úloha (2.3) bez obmedzujúcej podmienky, dokážeme ju previesť pomocou indikátorovej funkcie:

$$\iota_{\Gamma}(\mathbf{c}) = \begin{cases} 0 & \text{pre } \mathbf{c} \in \Gamma, \\ +\infty & \text{pre } \mathbf{c} \notin \Gamma. \end{cases} \quad (2.5)$$

Výsledný tvar bez obmedzujúcej podmienky bude:

$$\arg \min_{\mathbf{c}} \|\mathbf{c}\|_1 + \iota_{\Gamma}(\mathbf{c}). \quad (2.6)$$

### 2.1.2 Proximálne operátory

Projekčný operátor na konvexnú množinu nachádza bod, v tejto konvexnej množine ležiaci najbližšie východnému bodu. Proximálny operátor konvexnej funkcie je prirodzeným rozšírením pojmu projekčného operátora na konvexnú množinu, pretože už nehľadá najbližší bod, ale všeobecne blízky bod [2]. Proximálny operátor definujeme podľa [3] nasledovne:



**Definícia 2** *Nech  $f : \mathbb{R}^N \rightarrow \mathbb{R}$  je zdola polospojité konvexná funkcia s neprázdny m definičným oborom. Pre každé  $\mathbf{x} \in \mathbb{R}^N$  má mininimalizačný problém*

$$\min_{\mathbf{y} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + f(\mathbf{y}) \quad (2.7)$$

*jednoznačné riešenie, ktoré budeme označovať  $\text{prox}_f \mathbf{x}$ . Operátor  $\text{prox}_f : \mathbb{R}^N \rightarrow \mathbb{R}^N$  budeme nazývať proximálny operátor funkcie  $f$ .*

Pre naše potreby sú dôležité tri operátory: projekcia na množinu, mäkké prahovanie a clipping, ktoré v nasledujúcich podkapitolách popíšeme.

### Projekcia na konvexnú množinu

Projekcia  $\text{proj}_\Psi \mathbf{x}$  pre  $\mathbf{x} \in \mathbb{R}^N$  na neprázdnu uzavretú konvexnú množinu  $\Psi \subset \mathbb{R}^N$  je riešením úlohy:

$$\min_{\mathbf{y} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2 + \iota_\Psi(\mathbf{y}), \quad (2.8)$$

teda jedná sa o proximálny operátor indikátorovej funkcie príslušnej množiny.

### Mäkké prahovanie

Druhým operátom, ktorý je pre nás dôležitý, pretože je proximálnym operátorom  $\ell_1$ -normy, vyskytujúcej sa v úlohe (2.6), je mäkké prahovanie. Jeho princíp je jednoduchý. Ak je hodnota  $|c| \leq \gamma$ , vtedy je funkčnou hodnotou 0. Ak je  $|c| > \gamma$ , funkčná hodnota je buď zmenšená o  $\gamma$  pri kladných  $c$ , alebo zväčšená o  $\gamma$  pri záporných  $c$ . Toto platí pre reálne hodnoty. Rozšírením pre komplexné hodnoty je, že zostáva argument a znižuje alebo nuluje sa modul. Správanie operátora popíšeme rovnicou nasledovne:

$$\text{soft}_\gamma(\mathbf{c}) = \text{sgn}(\mathbf{c}) \odot \max(|\mathbf{c}| - \gamma, 0). \quad (2.9)$$

Symbol  $\odot$  predstavuje násobenie po zložkách a  $\text{sgn}$  funkciu signum, ktorá extrahuje znamienko z reálnej časti komplexného čísla.

### Operátor Clipping

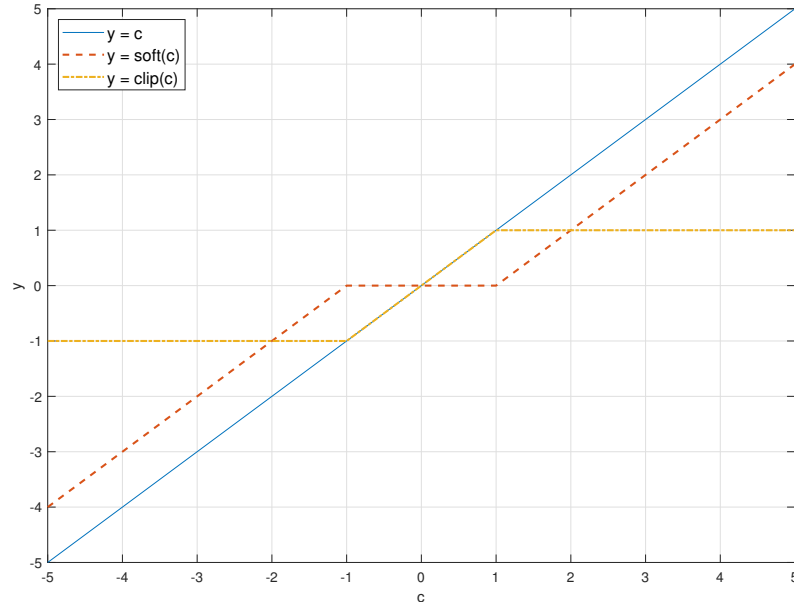
Operátor clipping bude využitý v Chambolle-Pockovom algoritme v kapitole 2.1.4. Definujeme ho nasledovne:

$$\text{clip}_\gamma(\mathbf{c}) = \text{sgn}(\mathbf{c}) \odot \min(\gamma, |\mathbf{c}|), \quad (2.10)$$

čo je explicitné vyjadrenie z rovnice  $\text{clip}_\gamma(\mathbf{c}) = \mathbf{c} - \text{soft}_\gamma(\mathbf{c})$ . Toto vyjadrenie je dokázané napríklad v práci [6]. Operátor funguje ako hard limiter. Hodnoty, ktoré sú

v absolútnej hodnote nižšie ako úroveň  $\gamma$  necháva nezmenené. Hodnoty  $\text{clip}_\gamma(c \leq -\gamma)$  sa rovnajú  $-\gamma$ , naopak  $\text{clip}_\gamma(c \geq \gamma) = \gamma$ .

Pre lepšiu predstavu je správanie oboch operátorov (soft a clip) zobrazené v obrázku 2.1.



Obr. 2.1: Graf proximálnych operátorov mäkké prahovanie a clip,  $\gamma = 1$ .

### 2.1.3 Douglasov–Rachfordov algoritmus

Jednou z možností riešenia optimalizačnej úlohy (P1) je Douglasov–Rachfordov algoritmus (DRA). Jedná sa o syntetizujúci model  $\ell_1$  relaxácie [6]. Keďže vychádzame z rovnice (2.1), kde je operátorom  $\mathbf{D}$ , čo je syntetizujúci operátor (počíta syntézu koeficientov), hovoríme o syntetizujúcom modeli. Z toho taktiež nepriamo vyplýva, že DRA je postavené predovšetkým na výpočte s časovo-frekvenčnými koeficientami signálu.

Všeobecne ho môžeme popísať algoritmom 1:

---

**Algoritmus 1:** Douglasov–Rachfordov algoritmus

---

**Vstup:** Operátory  $\mathbf{D}$  a  $\mathbf{A} = \mathbf{D}^*$ , množina  $\Gamma$ , poškodený signál na pozíciách  $\mathbf{y}(R)$

Nastav počiatočný bod  $\mathbf{c}^{(0)} \in \mathbb{C}^P$

Nastav parametre  $\lambda = 1, \gamma > 0$

**for**  $i = 0, 1 \dots \text{MaxIte}$  **do**

$\tilde{\mathbf{c}}^{(i)} = \text{proj}_\Gamma(\mathbf{c}^{(i)})$

$\mathbf{c}^{(i+1)} = \mathbf{c}^{(i)} + \lambda(\text{soft}_\gamma(\mathbf{2}\tilde{\mathbf{c}}^{(i)} - \mathbf{c}^{(i)}) - \tilde{\mathbf{c}}^{(i)})$

**end**

**return**  $\mathbf{D}(\text{proj}_\Gamma(\mathbf{c}^{(i+1)}))$

---

Algoritmus si teraz rozoberieme krok po kroku. Potrebujeme projekciu na množinu  $\Gamma$ , teda na množinu prípustných koeficientov reálneho signálu. Projekcia  $\mathbf{c}$  na množinu  $\Gamma$  je proximálny operátor indikátorovej funkcie. Projekcia na množinu  $\Gamma$  je definovaná v rovnici (2.11):

$$\text{proj}_\Gamma(\mathbf{c}) = \mathbf{c} - \mathbf{D}^*(\mathbf{D}\mathbf{c} - \text{proj}(\mathbf{D}\mathbf{c})), \quad (2.11)$$

kde  $\mathbf{D}^* : \mathbb{C}^P \rightarrow \mathbb{R}^N$  reprezentuje analýzu signálu, čiže prechod z časovej oblasti do frekvenčnej. V našom prípade používame DGT popísanú v sekcii 1.2. Operátor  $\mathbf{D} : \mathbb{R}^N \rightarrow \mathbb{C}^P$  označuje opačnú operáciu, syntézu signálu, pomocou inverznej diskkrétnej Gaborovej transformácie (IDGT). Pre správny priebeh projekcie je dôležité, aby parametre  $\mathbf{D}$  a  $\mathbf{D}^*$  boli nastavené rovnako [7].

Operátor  $\text{proj}$  definujeme po zložkách ako:

$$(\text{proj}(\mathbf{D}\mathbf{c}))_n = \begin{cases} y_n & \text{pre } n \in R \\ (\mathbf{D}\mathbf{c})_n & \text{pre } n \in M. \end{cases} \quad (2.12)$$

Znamená to, že pri každej iterácii prepíšeme pôvodné nepoškodené vzorky na ich miesta, pretože ich nechceme meniť. Vzorky v oblasti diery ponechávame nezmenené, pretože očakávame, že s príslušným počtom iterácii sa nám v tomto mieste budú vzorky približovať k optimu úlohy.

Výstupom algoritmu 1 bude vektor koeficientov  $\mathbf{c}$ . Rekonštrukčná úloha vyžaduje ako výstup signál. Aby sme zaručili, že sú nepoškodené vzorky nezmenené oproti stavu pred algoritmom, ešte raz vypočítame projekciu na množinu  $\Gamma$  z koeficientov  $\mathbf{c}$ . Následne prevedieme výpočet IDGT pre koeficienty po poslednej projekcii a získame opravený signál. Toto všetko ale platí len za predpokladu, že slovník  $\mathbf{D}$  tvorí tzv. Parsevallovsy tesný frame [8].

Parameter, ktorý sa podpíše na úspešnosti algoritmu je *MaxIte*, čiže počet iterácií (vykonaní) algoritmu. Zastavíme ho po fixnom počte iterácii alebo vtedy, keď sa riešenie ( $\ell_2$  norma časovo-frekvenčných koeficientov) zmení menej ako zvolená úroveň.

## 2.1.4 Chambollov–Pockov algoritmus

Opačným prístupom oproti syntetizujúcemu modelu je model analyzujúci. Jeho základom je riešenie problému:

$$\arg \min_{\mathbf{y}} \|\mathbf{A}\mathbf{y}\|_1 \text{ vzhľadom ku } \mathbf{y} \in \Gamma. \quad (\text{P2})$$

Matica  $\mathbf{A}$  je analyzujúcim operátorom a počíta analýzu signálu  $\mathbf{y}$ , ktorý je vstupom do algoritmu. Možnou realizáciou je Chambollov–Pockov algoritmus (CPA), ktorý je rovnako ako DRA proximálnym algoritmom (opakovane využíva proximálne operátory) [6]. Pribeh algoritmu je:

---

**Algoritmus 2:** Chambollov–Pockov algoritmus

---

**Vstup:** Operátory  $\mathbf{A}$  a  $\mathbf{D} = \mathbf{A}^*$ , množina  $\Gamma$ , poškodený signál na pozíciách  $\mathbf{y}(R)$

nastav parametre  $\sigma, \tau > 0$  splňujúce  $\sigma\tau\|\mathbf{A}\|^2 < 1$

nastav hodnotu primárnej premennej  $\mathbf{p}^{(0)} \in \mathbb{C}^N$  a duálnej premennej

$\mathbf{q}^{(0)} \in \mathbb{C}^P$

nastav hodnotu výstupnej premennej  $\mathbf{x}^{(0)} = \mathbf{p}^{(0)}$

**for**  $i = 0, 1, \dots$  *MaxIte* **do**

$\mathbf{q}^{(i+1)} = \text{clip}_1(\mathbf{q}^{(i)} + \sigma\mathbf{A}\mathbf{p}^{(i)})$   
 $\tilde{\mathbf{x}}^{(i+1)} = \text{proj}(\mathbf{x}^{(i)} - \tau\mathbf{D}\mathbf{q}^{(i+1)})$   
 $\mathbf{p}^{(i+1)} = 2\tilde{\mathbf{x}}^{(i+1)} - \mathbf{x}^{(i)}$

**end**

**return**  $\tilde{\mathbf{x}}^{(i+1)}$

---

Operátory  $\mathbf{A}$  a  $\mathbf{D}$  plnia rovnakú funkciu ako v DRA 1 a sú znova realizované pomocou DGT a IDGT. Zápisom  $\|\mathbf{A}\|^2$  myslíme operátorovú normu. Zaručenie podmienky  $\sigma\tau\|\mathbf{A}\|^2 < 1$  stojí na fakte, že v prípade Parsevalovho framu je operátorová norma  $\mathbf{A} = 1$  [6]. Preto sa podmienka zjednoduší na tvar  $\sigma\tau < 1$ . Ukončenie algoritmu je postavené na rovnakých možnostiach ako v DRA, teda fixný počet iterácií alebo minimálna zmena  $\ell_2$  normy signálu medzi iteráciami.

## 3 Psychoakustický model ľudského sluchu

Skúmanie vlastností ľudského sluchu, predovšetkým jeho modulovej kmitočtovej charakteristiky, siahajú do prvej polovice 20. storočia. Vedomosť o nedokonalostiach ucha poslúžila predovšetkým pri návrhu kódokov pre televízne vysielanie (MPEG-1, MPEG-2, H.264 a mnohé ďalšie) [9]. Zvyšujúce sa nároky na kvalitu už z hľadiska objemu dát nebolo možné prenášať s ohľadom na čas potrebný k prenosu. Kompresiou dát dnes môžeme súbor zmenšiť a s využitím psychoakustického modelu ucha niektoré dáta po analýze zanedbať, pretože náš zmyslový orgán by ich nepočul ani pred kompresiou.

V úlohe rekonštrukcie riešime ale problém odlišný. Algoritmy na rekonštrukciu hľadajú riešenie, ktoré konverguje k optimu úlohy, ale nezahŕňa aspekt, či je zvuk poslucháčovi príjemný. Psychoakustický model by mal v riešení zmierniť vplyvy rušivých elementov, ktoré pri algoritme môžu vzniknúť, v ideálnom prípade ich odstrániť.

Kapitola obsahuje základy ľudského vnímania sluchom a vysvetlenie zapojenia psychoakustického modelu do úlohy doplnenia chýbajúcich vzoriek zvukového signálu.

### 3.1 ATH

V 30. rokoch 20. storočia Fletcher zrealizoval meranie na širšej vzorke poslucháčov, kde skúmal modulovo frekvenčnú závislosť ucha [10]. Výsledkom pozorovania je nelineárna funkcia závislosti úrovne akustického tlaku, pri ktorom človek dokáže rozoznať čistý tón v bezšumovom prostredí na frekvencii. Krivku môžeme približne popísať rovnicou:

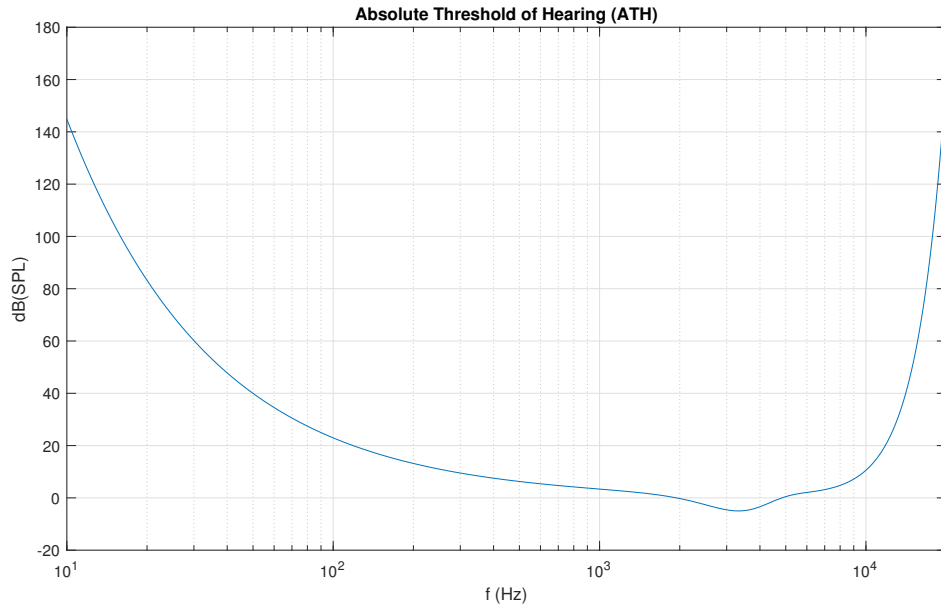
$$T_q(f) = 3,64(F)^{-0.8} - 6,5e^{-0,6(F-3,3)^2} + 10^{-3}(F)^4, \quad (3.1)$$

kde  $f$  je frekvencia harmonického tónu v Hz (pre zjednodušenie značenia je použitá substitúcia  $F = \frac{f}{1000}$ ) a  $T_q$  je hladina akustického tlaku **Sound Pressure Level** (SPL) v decibeloch (dB). Hladina akustického tlaku je definovaná ako:

$$\text{SPL} = 20 \cdot \log_{10} \frac{p}{p_0}, \quad (3.2)$$

kde  $p$  je akustický tlak a  $p_0 = 2 \cdot 10^{-5}$  Pa je referenčná hodnota akustického tlaku (odpovedá 0 dB) [10].

Graf 3.1 je zobrazením priebehu (3.1). Lokálne minimum v rozmedzí 3 až 4 kHz je oblasť, kde je sluch najcitlivejší. Je dôležité poznamenať, že krivka odpovedá približne 20 ročnému poslucháčovi so zdravým sluchom. S rastúcim vekom citlivosť klesá najmä na vyšších kmitočtoch od oblasti minima (4 kHz).



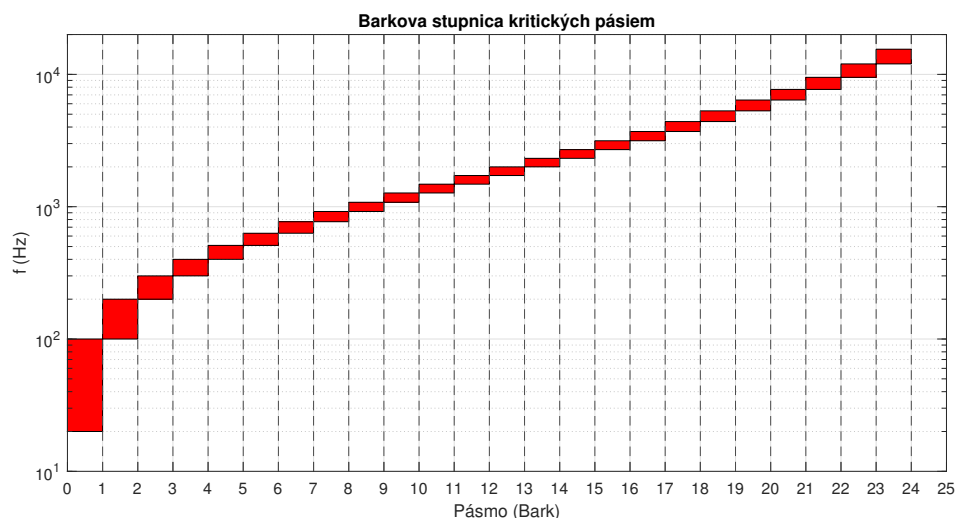
Obr. 3.1: Prah počuteľnosti ľudského sluchu.

## 3.2 Kmitočtové maskovanie

Transformácia akustického signálu v ľudskom uchu na elektrický vnem prebieha na bazilárnej membráne. Bazilárnu membránu si môžeme predstaviť ako priamku s premennou frekvenciou. Znejúci tón v závislosti od amplitúdy a frekvencie sa prejaví ako najväčšia výchylka na membráne. Vnem tónu sa prejaví na membráne aj v oblasti vyššieho a nižšieho kmitočtu s nižšou úrovňou podráždenia. Tento jav spôsobí, že silnejší tón zamaskuje vnemy pod úrovňou týchto amplitúd na blízkych kmitočtoch, a preto nebudú rozpoznateľné. Zwicker zistil, že sluch má 24 kritických frekvenčných pásiem [11]. V grafe 3.2 môžeme vidieť rozloženie týchto pásiem. Vnemy, ktoré sa nachádzajú v jednom kritickom pásme sa vzájomne ovplyvňujú, avšak vnemy v rôznych pasmách na seba vplyv nemajú.

### 3.2.1 Simultánne maskovanie a maskovacia nesymetria

Najväčší vplyv v maskovaní má simultánne maskovanie, teda zamaskovanie tónu alebo šumu iným tónom alebo šumom znejúcich v tom istom momente. V knihe [10] je popísaná maskovacia asymetria, ktorá hovorí o tom, že nezáleží len na úrovni oboch zvukov, ale taktiež na charaktere zvuku. Na popis rozdielu zvukov využijeme SMR (dB) – odstup signálu od najvyššej úrovne (thresholdu) maskovaného signálu, kedy je signál ešte nepočuteľný. Hodnota SMR je rozdielom úrovne signálu, ktorý maskuje (maskér) a signálu, ktorý je zamaskovaný (maskovaný).



Obr. 3.2: Rozloženie frekvencií podľa Barkovej stupnice do príslušného pásma. Pásmu číslo jedna odpovedá plocha medzi 0–1. Podobne je to aj s nasledujúcimi pásmami.

SMR môžeme analyticky popísať:

$$\text{SMR} = S - M, \quad (3.3)$$

kde  $S$  značí úroveň maskéra v dB(SPL) a  $M$  úroveň maskovaného v dB(SPL).

V prípade, keď je tón maskovaný šumom (Noise Masking Tone – NMT) je SMR na hodnote približne 4 dB. Keď je šum maskovaný tónom (Tone Masking Noise – TMN), je SMR podstatne vyššie ako v prípade NMT, a to na úrovni približne 24 dB. Rozdielny odstup SMR pri rôznych typoch signálu označujeme ako maskovacia asymetria.

### 3.2.2 Nesimultánne maskovanie

Vplyv na maskovanie v jednom momente má nielen intenzita v tomto momente, ale aj prítomná maskovacia krivka, a to pred i po doznení zvuku. Maskovacia úroveň pred zaznením nie je taká výrazná a úroveň, kedy môže ovplyvniť vnem je od 1–2 ms. Významnejšia je fáza po zaznení tónu, ktorá môže trvať 50–300 ms [10]. Dôvodom postmaskingu je, že miesto výchylky na bazilárnej membráne po doznení zvuku potrebuje nejaký čas, pokiaľ sa ustáli do stavu pred prebudením. V našom modeli ale túto vlastnosť sluchového aparátu nezahrnieme, pretože zapojenie do algoritmu by bolo zložité a nie je zrejmé, aký prístup zvoliť v oblasti chýbajúceho úseku.

### 3.3 Váhovanie koeficientami

Operácia váhovania je záležitosťou kmitočtovej oblasti audiosignálu. Každé Gabo-rovo okno násobíme po zložkách súborom konštánt – vektorom váh, ktoré budú rešpektovať určité psychoakustické vlastnosti ľudského sluchu s cieľom vytvorenia čo najhodnovernejšej reprezentácie v časovej oblasti. Táto operácia prebieha v operátoroch mäkkého prahovania a clippingu, popísaných detailnejšie v kapitole 2. Cieľom je ponechať v optimalizácii priestor tým zložkám, ktoré nebudú mať vplyv na subjektívny vnem rekonštruovaného signálu.

Nedá sa povedať, ktorý prístup (výber váhového vektora) bude úspešný. V nasledujúcich podkapitolách bude vysvetlený širší výber váh, ktoré budú následne testované a vyhodnotené.

#### 3.3.1 Váhové vektory vychádzajúce z ATH

Krivka ATH popisuje citlivé miesta ľudského sluchu, preto pri výbere váh budeme hľadať inverzné funkcie ku krivke ATH. Zaručia nám kontrolu frekvenčných oblastí zreteľných pre ľudský sluch. V článku [7] sú popísané váhové vektory testované na úlohe declippingu, preto ich otestujeme na úlohe doplnenia chýbajúceho úseku, ktorá je declippingu veľmi podobná. Hodnoty váhových vektorov sú vypočítané nasledovne:

$$\mathbf{w}_{ATH1} = (\mathbf{t} - \min \mathbf{t} + 1)^{-1}, \quad (3.4a)$$

$$\mathbf{w}_{ATH2} = -\mathbf{t} + \tau, \quad (3.4b)$$

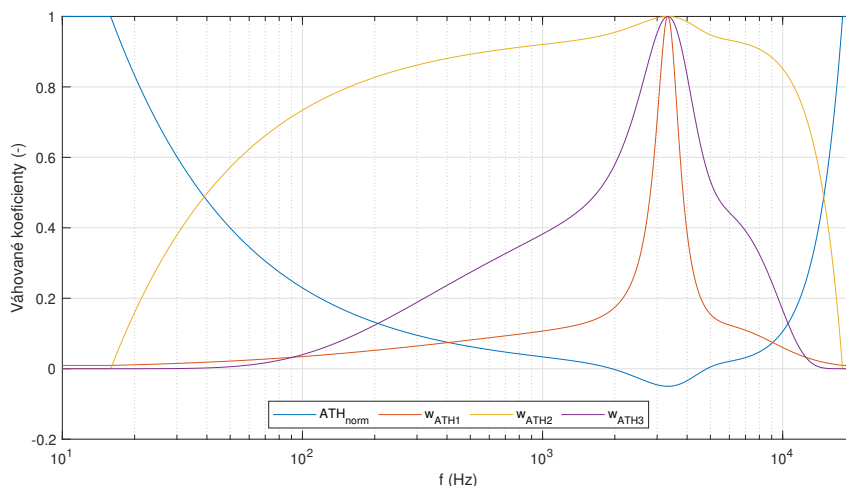
$$\mathbf{w}_{ATH3} = 2 \cdot 10^{-5} \cdot 10^{(-\mathbf{t} + \tau)/20}, \quad (3.4c)$$

kde  $\mathbf{w}$  je váhový vektor,  $\mathbf{t}$  je vektor hodnôt ATH podľa rovnice (3.1), ktorý rešpektuje počet frekvenčných kanálov DGT. Konštanta  $\tau$  je najvyššia hodnota ATH v dB(SPL). Koeficienty váh je potrebné normalizovať, aby mali rovnaké maximum a bol zachovaný charakter každej krivky v rozsahu 0–1. Všetky koeficienty vektora  $\mathbf{w}$  podelíme najvyššou hodnotou prvku vektora  $\mathbf{w}$ . Normalizáciou dosiahneme to, že po váhovaní nebude žiadny harmonický element zosilnený a dôjde len k zoslabeniu určitých častí spektra. Krivky po normalizácii vidíme v grafe 3.3.

#### 3.3.2 Váhové vektory vychádzajúce z krivky rovnakej hlasitosti

V roku 1933 Flechter a Munson publikovali výskum, v ktorom popisujú krivky rovnakej hlasitosti [12]. Zistovali priebeh akustického tlaku v oblasti frekvencií počuteľného pásma. Jednotka hlasitosti (funkcia závislosti akustického tlaku na frekvencii) získala pomenovanie Fon (Ph).





Obr. 3.3: Normalizované váhové funkcie vychádzajúce z predpisu ATH.

Krivky rovnakej hlasitosti boli inšpiráciou pre váhové filtre používané vo zvukomeroch, ktoré sú definované v medzinárodných normách (napr. [13]). My do merania zapojíme filtre typu A, B, C a Z. Ich predpis je inverznou funkciou ku krivkám hlasitosti, A–40 Ph, B–70 Ph, C–100 Ph. Funkcia Z je rovná jednej, teda ponecháva celé kmitočtové pásmo bez zmeny. Predpisy váhovacích funkcií sú nasledovné:

$$\mathbf{w}_A(f) = k(f) \cdot \frac{f^2}{\sqrt{(f^2 + 107 \cdot 7^2)(f^2 + 737 \cdot 9^2)}}, \quad (3.5a)$$

$$\mathbf{w}_B(f) = k(f) \cdot \frac{f}{\sqrt{(f^2 + 3950)}}, \quad (3.5b)$$

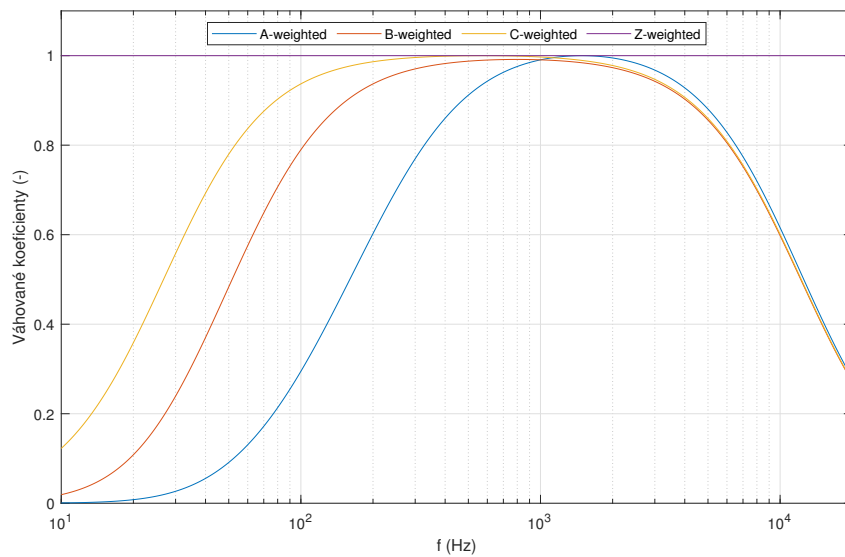
$$\mathbf{w}_C(f) = k(f), \quad (3.5c)$$

$$\mathbf{w}_Z(f) = 1. \quad (3.5d)$$

Pre zjednodušenie zápisu je použitý frekvenčne závislý koeficient  $k(f)$ , ktorý má hodnotu:

$$k(f) = \frac{12194f^2}{(f^2 + 720)(f^2 + 12194^2)}. \quad (3.6)$$

Hodnoty koeficientov v a (3.5) a (3.6) sú prebraté z normy IS 15575 [14], kde je ich výber podrobne popísaný. Pre túto prácu ale podrobný popis koeficientov nie je dôležitý, pretože nás zaujímajú len hodnoty pre výpočet krivky. Funkčné hodnoty týchto predpisov sú v rozmedzí 0 až 1, ale nemajú najvyššiu hodnotu 1, preto pred použitím týchto filtrov váhové koeficienty normalizujeme, aby boli všetky váhové koeficienty v rovnakom rozsahu. V grafe 3.4 vidíme priebeh filtrov po normalizácii.



Obr. 3.4: Normalizované váhové funkcie A, B, C, Z.

## 4 Experimenty

V rámci tejto kapitoly bude popísané testovanie rekonštrukcie poškodených nahrávok. V prvej časti rozoberieme techniky posúdenia kvality opravenej nahrávky. Druhá časť bude venovaná spôsobu realizácie algoritmu a porovnaniu výsledkov získaných z Matlabu. Keďže táto úloha je simulovaná a vieme, ako vyzeral nepoškodený signál, môžeme vypočítať aj  $\ell_1$  normu jeho Gaborovských koeficientov z analýzy. Normu vieme vypočítať aj pre Gaborovské koeficienty z analýzy zrekonštruovaného signálu po skončení algoritmu (DRA alebo CPA). Z pohľadu optimalizačnej úlohy je pre nás najlepšie to riešenie, ktoré má najmenšiu  $\ell_1$  normu, pretože sa jedná o minimalizáciu.

### 4.1 Metódy hodnotenia kvality

Základné delenie hodnotiacich metód je delenie na subjektívne a objektívne. Najsilnejšou doménou subjektívnych meraní je ich posudzovateľ, ktorým je človek a nie počítač. Výhodou je aj to, že poslucháčom opravenej hudby bude taktiež človek. Výsledky sú ale podmienené testovaním na širokej škále poslucháčov a aj zvukových vzoriek, aby mali nejakú výpovednú hodnotu, navyše podliehajú pomerne prísny normám. Preto sa snažíme nahradiť tieto merania objektívnymi metódami, ktoré dokážu merania zásadne urýchliť. Výsledok objektívnych meraní s najlepšimi objektívnymi parametrami po subjektívnom zhodnotení nemusí byť najlepší a vzorka s horšími objektívnymi parametrami môže byť na počutie príjemnejšia. Preto ako cestu zvolíme zapojenie oboch typov hodnotenia. Objektívnymi meraniami získame užší výber, ktorý môžeme zhodnotiť pomocou vlastného vnemu a nebude pre nás časovo náročný. Medzi subjektívne metódy patrí napríklad:

- **S**imple **R**ating without reference (SR) [15, kap. 4],
- **A**bsolute **C**ategory **R**ating (ACR) [16],
- **M**ultiple **S**timuli with **H**idden **R**eference and **A**nchor (MUSHRA) [17],
- **D**ouble-blind **T**riple-**S**timulus with hidden **R**eference (DTSWH)[15, kap. 3].

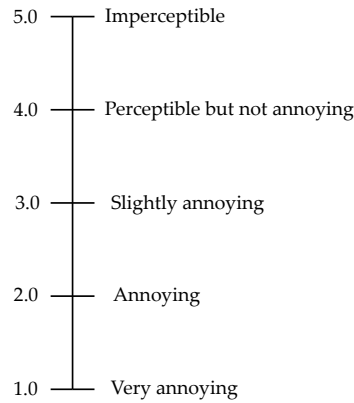
Medzi objektívne metódy patrí:

- **S**ignal-to-distortion ratio (SDR),
- **P**erceptual **E**valuation of **A**udio **Q**uality (PEAQ) [18],
- **P**erception **M**odel – **Q**uality **A**ssesment (PEMO-Q) [19],
- Rnonlin [20],
- **P**erceptual **O**bjective **L**istening **Q**uality **A**ssesment (POLQA) [21],
- **T**he **V**irtual **S**peech **Q**uality **O**bjective **L**istener (ViSQOL) [22].

Vybrané metódy, ktoré budú v tejto práci použité, v nasledujúcich podkapitolách bližšie popíšeme.

### 4.1.1 Double-blind triple-stimulus with hidden reference

Hodnotiacemu poslucháčovi sú prehrávané 3 zvukové signály A, B, C, ktoré môže ľubovoľne prepínať. Signál A je nespracovaný referenčný signál, signály B a C sú spracované požadovaným systémom. Poslucháč hodnotí kvalitu signálov B a C vzhľadom k signálu A na základe stupnice 4.1 [23].



Obr. 4.1: Stupnica ITU-R.

Kvalita je posúdená takzvaným subjective difference grade (SDG) nasledovne:

$$\text{SDG} = \text{grade}_{\text{test}} - \text{grade}_{\text{ref}}, \quad (4.1)$$

kde  $\text{grade}_{\text{test}}$  označuje kvalitu testovaného signálu podľa stupnice 4.1 a  $\text{grade}_{\text{ref}}$  je kvalita referenčného signálu.

### 4.1.2 Pomer signálu k poškodeniu

Z pohľadu rekonštrukčnej úlohy chceme vedieť, ako ďaleko je doplnený signál od pôvodného, teda ako sa líšia od seba na úrovni vzoriek v časovo-amplitúdovej oblasti. Pre toto porovnanie použijeme metódu signal-to-distortion ratio (SDR), v preklade pomer signálu k poškodeniu. SDR definujeme podľa [1] nasledujúcou rovnicou:

$$\text{SDR}(\mathbf{y}(M), \hat{\mathbf{y}}(M)) = 10 \cdot \log_{10} \frac{\|\mathbf{y}(M)\|_2^2}{\|\mathbf{y}(M) - \hat{\mathbf{y}}(M)\|_2^2} \quad [\text{dB}], \quad (4.2)$$

kde  $\mathbf{y}(M)$  reprezentuje nepoškodený signál v miestach, kde sme ho neskôr poškodili. Signál  $\hat{\mathbf{y}}(M)$  reprezentuje zrekonštruovaný signál na pozíciách rekonštrukcie (množiny  $M$ ). Vypočet SDR na mieste nepoškodených vzoriek nie je potrebný, pretože pred syntézou signálu sme nahradili v zrekonštruovanom signáli na pozíciách hodnoverných vzoriek tieto vzorky z pôvodného signálu. Rozdiel  $\mathbf{y}(R) - \hat{\mathbf{y}}(R)$  všetkých nepoškodených úsekov by bol nulový, pretože signály sú rovnaké a nemá zmysel SDR na pozícii hodnoverných vzoriek počítať (SDR by bolo nekonečne veľké). Čím vyššie bude SDR, tým lepšie rekonštrukcia dopadla [24].

### 4.1.3 PEMO-Q

V oblasti objektívnych hodnotení sa jedná o novú metódu predstavenú na začiatku 21. storočia [19]. Táto hodnotiacia technika bola vyvinutá najmä na posúdenie skomprimovaného menejbitového signálu v porovnaní s pôvodným signálom. Využíva sa taktiež pri úlohe, ktorú riešime, teda pri hodnotení rekonštrukcie nekomprimovaného signálu.

Kvalita je posúdená objective difference grade (ODG) s obdobnou škálou ako SDG pri subjektívnom hodnotení. Názvy úrovní sú rovnaké ako pri 4.1, no rozdielne je posunutie numerických hodnôt stupnice. Hodnoteniu Imperceptible prislúcha číselná hodnota 0 a Very annoying hodnota  $-4$ .

## 4.2 Rekonštrukčný algoritmus

V tejto kapitole prejdeme na samotnú realizáciu v Matlabe. Najprv popíšeme nepoškodené signály a spôsob, akým ich poškodíme. Následne ozrejmime nastavenie parametrov do algoritmu. V závere posúdime výsledky algoritmov objektívnymi metódami popísanými v kapitole 4.1 a subjektívne, vlastným sluchom.

### 4.2.1 Banka signálov a ich úprava

Do algoritmu vstúpi celkovo 15 signálov, 10 reálnych signálov s  $f_{vz} = 44\,100$  Hz z normy EBU-SQAM [25], z ktorých použijeme rovnaký užší výber ako v práci [24] a 5 syntetizovaných signálov.<sup>1</sup> Matematický predpis syntetizovaných signálov v čase  $t$  je:

$$y(t) = \sin(\omega t) + \frac{1}{2} \sin\left(\frac{\omega t}{2}\right) + \frac{1}{4} \sin\left(\frac{\omega t}{4}\right), \quad (4.3a)$$

$$y(t) = \sum_{n=0}^7 \left( \frac{1}{(2n+1)} \sin(\omega(2n+1)t) \right), \quad (4.3b)$$

$$y(t) = \sum_{n=0}^9 \left( \frac{1}{(2a_n+1)^2} \sin(\omega a_n t) \right), \text{ pre } a_n \text{ platí: } a_n \in \mathbb{N}; a_n = \mathcal{U}(1, 30), \quad (4.3c)$$

$$y(t) = \text{sgn}(\sin(\omega t)), \quad (4.3d)$$

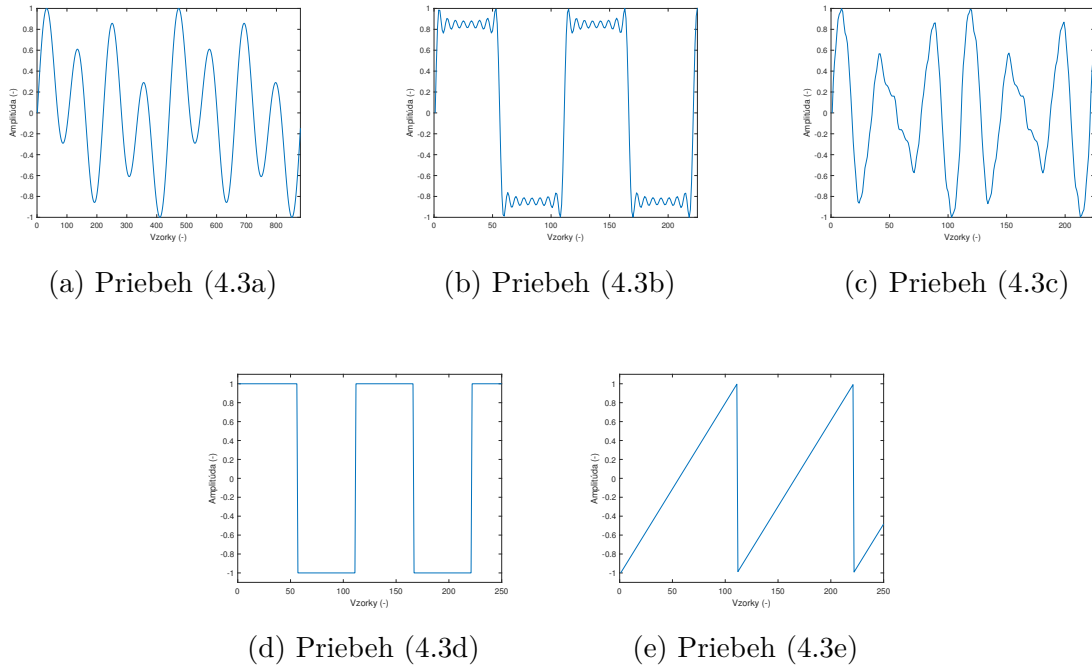
$$y(t) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(\omega n t). \quad (4.3e)$$

Kvôli prehľadnejšiemu zápisu využívame  $\omega = 2\pi f$ , kde  $f$  je frekvencia fundamentálnej frekvencie. V našom prípade  $f = 400$  Hz, teda  $\omega$  bude konštantou pre všetky umelé signály. Priebehy (4.3a) – (4.3c) sú postavené na súčte sínusových vln. Pre tretí

---

<sup>1</sup>Reálne signály dostupné z: [https://github.com/ondrejmokry/InpaintingRevisited/blob/master/matlab/EBU\\_SQAM.mat](https://github.com/ondrejmokry/InpaintingRevisited/blob/master/matlab/EBU_SQAM.mat)

signál (4.3c) sú koeficienty generované náhodne na základe rovnomerného rozdelenia (symbol  $\mathcal{U}$ ). Aby bola zaručená kontrolovaná náhodnosť (pri opakovanom spustení skriptu budú čísla náhodné, ale vždy rovnaké), je použitá funkcia `rng.m`. Za (4.3d) je skrytý obdĺžnikový signál so striedou 50 %, realizovaný funkciou `square.m` zo Signal Processing Toolbox (SPT). Posledný umelý signál (4.3e) je pílovitý, takisto realizovaný funkciou `sawtooth.m` z SPT.

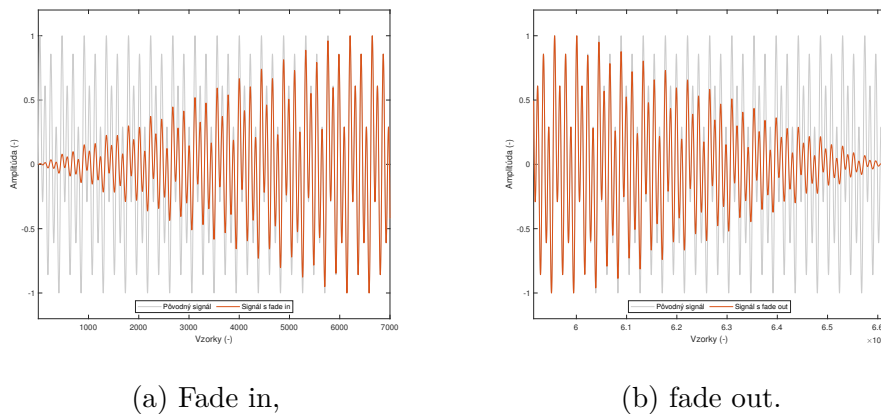


Obr. 4.2: Zobrazenie približne dvoch periód syntetizovaných signálov podľa (4.3).

Reálne signály sú rôznej dĺžky, preto potrebujeme zaistiť, aby mali rovnakú dĺžku. Najprv ich skrátime na dĺžku 1,5 s, čo predstavuje 66 150 vzoriek. Realizujeme to výberom vzoriek s indexami  $\{f_{vz} + 1, \dots, 1,5 \cdot f_{vz}\}$ . Umelé signály sú generované funkciou `synthsignals.m`, ktorá má ako vstup dĺžku skráteného signálu. Preto budú mať signály požadovanú dĺžku 66 150 vzoriek. Táto dĺžka bude dostatočná na subjektívne zhodnotenie signálu. Následne všetky signály zoradíme do jednej matice, kde každý stĺpec predstavuje jeden signál a každý signál normalizujeme tak, aby bola najvyššia amplitúda zhodná.

Pri posluhu opravených nahrávok by mohlo vzniknúť na začiatku a na konci lupnutie, ktoré by ovplyvnilo poslucháča, pretože by bol upriamený viac na lupnutie ako na zrekonštruovaný úsek. Zabránilo tomu postupným zosilnením a zoslabením vzoriek (tzv. fade in a fade out). Realizáciu vykonávame pomocou vektora, ktorý obsahuje koeficienty lineárnej funkcie  $y = x$  od 0 do 1. Počet prvkov je rovný požadovanej dĺžke zoslabovania, v našom prípade 6 000 vzoriek. Súčinom koeficientov signálu s koeficientami vektora po zložkách upravíme amplitúdu na začiatku.

Na konci je operácia rovnaká s rozdielom, že otočíme hodnoty vektora od 1 po 0. Znázornenie postupného zosilnenia a zoslabenia je v nasledujúcom obrázku.



Obr. 4.3: Krátke úseky začiatku a konca, kde je aplikovaný fade, zobrazené na priebehu signálu (4.3a).

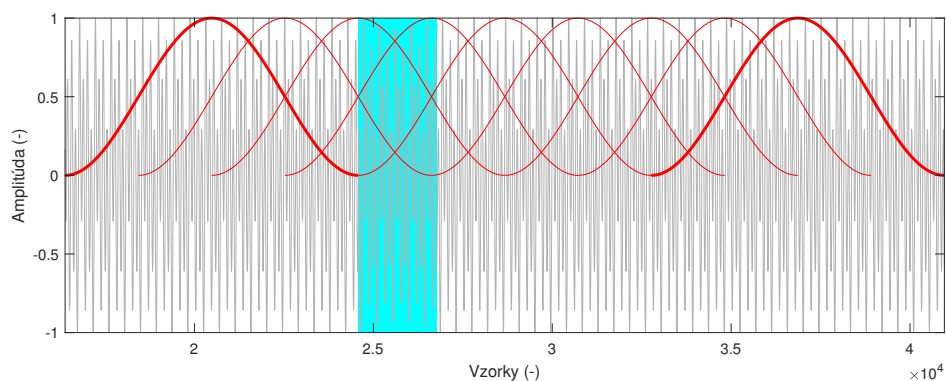
## 4.2.2 Vstupný signál do algoritmov

Signál o dĺžke 66 150 je ako vstup algoritmu zbytočne dlhý a zvyšoval by výpočetnú náročnosť algoritmu. Postačí, keď bude signál pozostávať zo vzoriek, ktoré sú v tesnej blízkosti. Naša úvaha je postavená na základoch, ktoré boli prezentované v práci [26]. Okno  $w$  v DGT volíme  $w = 8192$ , preto pred a za úsek diery vložíme 8192 vzoriek, čím zaručíme, že v prvom a poslednom okne sa budú vyskytovať výhradne nepoškodené vzorky. Následne fixne zvolíme začiatok diery na pozícii 8193 (prvá poškodená vzorka). Počet poškodených vzoriek  $n$  je dopočítaný na základe znalosti vzorkovacej frekvencie  $f_{vz}$  pomocou vzorca:

$$n = \frac{t}{1000} \cdot f_{vz}, \quad (4.4)$$

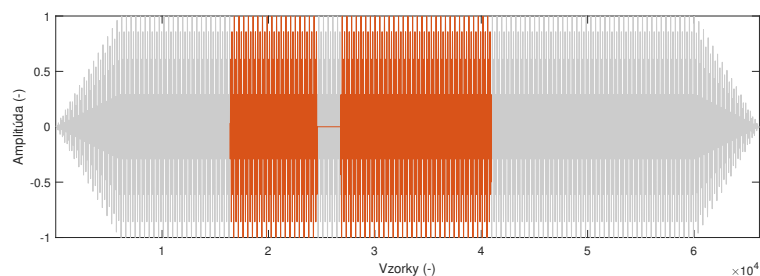
kde  $t$  je trvanie diery v ms. Dĺžku diery volíme v rozmedzí od 5 po 50 ms, v intervale po 5 ms. Celkovo budeme testovať 10 dier. Keďže vzorkovacia frekvencia našej banky signálov je 44,1 kHz, pri dĺžke diery napr. 5 ms by nebol výsledok (4.4) celé číslo. V takom prípade zaokrúhľime počet  $n$  na jednotky nahor. Finálna dĺžka  $d$  skráteného signálu by bola  $d = w \cdot 2 + n$ . Táto dĺžka ale nevyhovuje parametrom operátorov `frana` a `frsyn`, ktoré realizujú DGT a IDGT. Preto musíme použiť funkciu `framelength` z knižnice LTFAT<sup>2</sup>, ktorá vypočíta dĺžku signálu, aby mohli `frana` a `frsyn` prebehnúť. Znázornenie všetkých okien je v nasledujúcom grafe 4.4.

<sup>2</sup>Knižnica dostupná z: <https://ltfat.github.io/>



Obr. 4.4: Hannovo okno dĺžky  $w = 8192$  s posunom  $a = 2048$  na skrátenej signáli. Šedou farbou je zobrazený rekonštruovaný signál a tyrkysovou je zvýraznený doplnený úsek. Hrubou čiarou je zvýraznené prvé a posledné okno. Vidíme, že prvé okno nezasahuje do doplneného úseku. Posledné by sme dokázali posunúť smerom doľava (teda počítať s menším počtom vzoriek), ale výsledok syntézy po analýze by bol rôzne dlhý (doplnený nulami) a museli by sme to nejakým spôsobom riešiť.

Úsek, kde sa nachádza diera volíme v dlhšom úseku na pozíciu  $3 \cdot w + 1$ , teda skrátenej signál bude pozostávať zo vzoriek s indexami  $2 \cdot w + 1$  až  $5 \cdot w$ . Toto ilustruje obrázok 4.5



Obr. 4.5: Umiestenie skrátenej úseku vstupujúceho do algoritmu s dĺžkou  $d = 3 \cdot w = 24576$  vzoriek, dĺžka diery predstavuje 50 ms, teda 2205 vzoriek. Šedý signál reprezentuje pôvodný signál s použitím fade a oranžový signál predstavuje skrátenej signál vstupujúci do algoritmu.

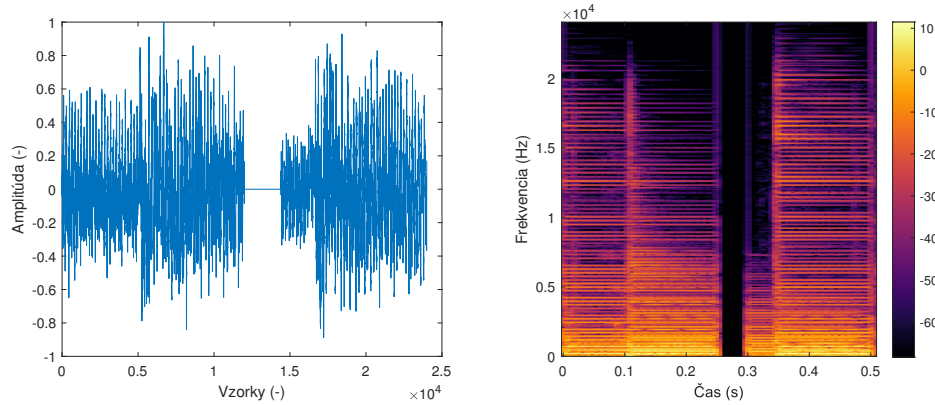
Po skončení algoritmu na rovnakú pozíciu ako signál skrátenej vložíme opravený signál.

### 4.2.3 Overenie funkčnosti algoritmov

Pred spustením testovania na celej banke signálov so zmenami vo váhových vektoroch a dĺžke vystrihnutého úseku otestujeme funkčnosť oboch algoritmov. Kon-



krátne budeme sledovať závislosť  $\ell_1$  normy frekvenčných koeficientov na počte iterácií u DRA a priebeh  $\ell_1$  normy analýzy signálu na počte iterácií u CPA. Očakávaným výsledkom je, že  $\ell_1$  norma bude klesať, čiže konvergovať k optimu úlohy. Tento test je prevedený na testovacej nahrávke akustickej gitary s  $f_{vz} = 48\,000$  Hz a dĺžkou diery 50 ms, ktorá nie je súčasťou banky z kapitoly 4.2.1 a slúži len na účely overenia funkčnosti algoritmov. Priebeh a spektrogram tohto signálu vidíme v obrázku 4.6.



Obr. 4.6: Časovo-amplitúdové zobrazenie a časovo-frekvenčná analýza umelo poškodeného signálu o dĺžke 500 ms s vloženými nulami v dĺžke 50 ms približne v strede úseku.

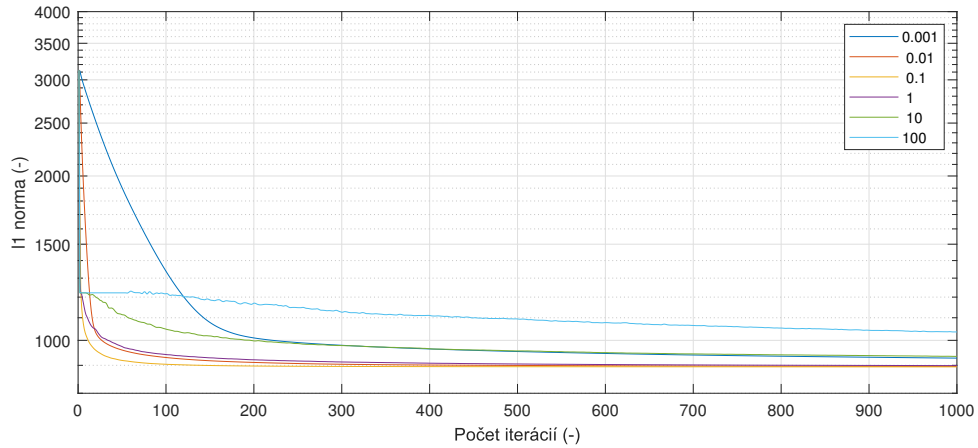
### Test funkčnosti DRA

Podľa článku [3] volíme  $\gamma > 0$ . Pre našu prácu sme si zvolili 6 parametrov  $\gamma$ , konkrétne:  $\{0,001, 0,01, 0,1, 1, 10, 100\}$ . Porovnanie bude spočívať v tom, že pri každom  $\gamma$  spočítame 1000 iterácií. Po každej iterácii spočítame  $\ell_1$  normu  $\tilde{c}$  v algoritme 1, ktorá má podľa [3] konvergovať k optimu úlohy. Zobrazíme, ako sa  $\ell_1$  norma vyvíja s počtom iterácií.

Výsledky porovnania sú v obrázku 4.7. Pre všetky  $\gamma$  riešenie spĺňa predpoklad minimalizačnej úlohy. Všetky  $\ell_1$  normy konvergujú (klesajú) k optimu úlohy. Vo vzťahu k počtu iterácií vidíme, že  $\ell_1$  norma pri nízkych  $\gamma$  nemá čisto klesajúci trend, ale v globále môžeme povedať, že klesá. Pri  $\gamma$  väčších ako 1 v rozsahu cca prvých 300 iterácií vidíme veľmi prudký pokles  $\ell_1$  normy, ale potom sa ustáluje a hodnota  $\ell_1$  normy klesá už pozvoľne.

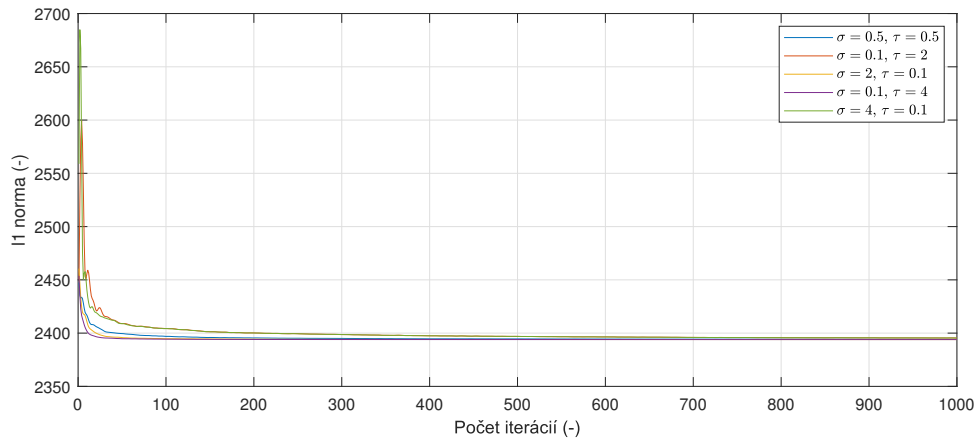
### Test funkčnosti CPA

Test priebeha rovnako pri DRA. Vyskúšali sme 5 dvojíc parametrov  $\sigma$  a  $\tau$ , aby spĺňali  $\sigma\tau < 1$ . Dvojice boli nasledovné:  $(\sigma, \tau) = [0,5; 0,5]; [0,1; 2]; [2; 0,1]; [0,1; 4];$



Obr. 4.7: Závislosť  $\ell_1$  normy na zvyšujúcom sa počte iterácií pre rôzne  $\gamma$ , počiatočný bod  $\mathbf{c}^{(0)}$  je matica Gaborovských koeficientov signálu s dierou.

[4; 0,1].

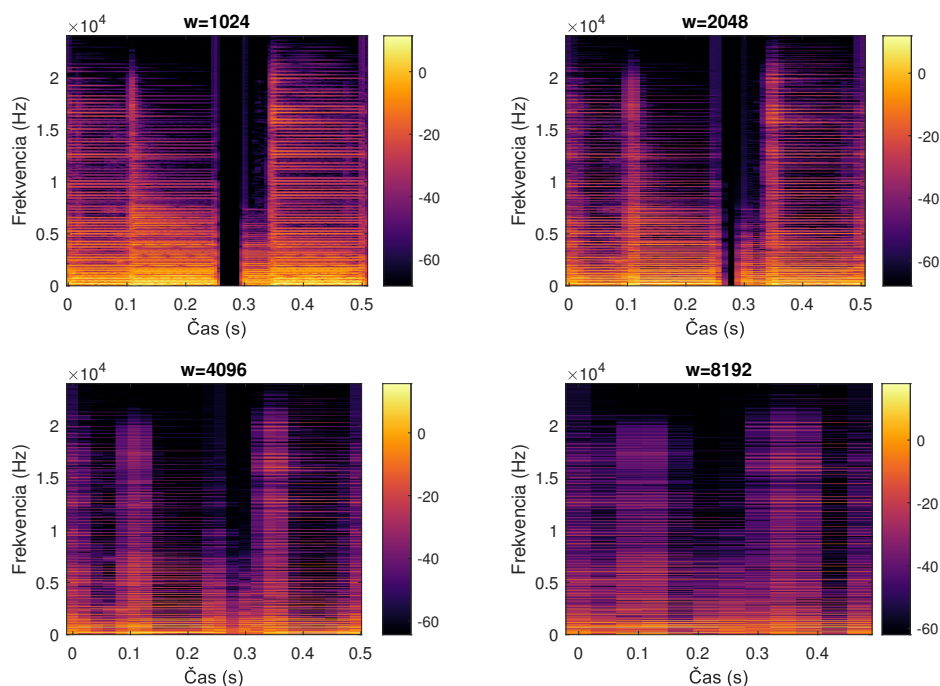


Obr. 4.8: Závislosť  $\ell_1$  normy na počte iterácií pre rôzne  $\sigma$  a  $\tau$ .

Pre všetky varianty premenných môžeme na základe grafu 4.8 potvrdiť, že algoritmus konverguje a považujeme ho za funkčný pre ďalšie výpočty.

#### 4.2.4 Nastavenie algoritmov

Na rekonštrukciu využijeme oba algoritmy popísané v kapitole 2, DRA a CPA. Nastavenie transformácií (DGT a IDGT) je v oboch prípadoch rovnaké, a to:  $w = 8192$ ,  $a = 2048$ ,  $M = 8192$ . Na vysvetlenie použitia  $w = 8192$  posluží obrázok 4.9, kde zobrazujeme časovo frekvenčnú analýzu testovacieho signálu akustickej gitary s nulami, pričom meníme dĺžku okna  $w$ , pričom  $w = M$ .



Obr. 4.9: Časovo-frekvenčná analýza umelo poškodeného signálu s meniacou sa šírkou okna  $w$ ,  $a = w/4$ .

Keďže parameter  $w$  udáva dĺžku v počte vzoriek, potrebujeme na zdôvodnenie skrytej diery v spektrograme poznať počet vzoriek diery. Počet nulových vzoriek za 50 ms sa podľa (4.4) rovná  $\frac{50}{1000} \cdot 48\,000 = 2\,400$  vzoriek. Pri  $w = 1024$  určite nastane situácia, že v oblasti okna budú všetky koeficienty nulové a aj FFT vypočíta nulové koeficienty, pretože dĺžka diery je 2400. Čím je  $w$  vyššie, tým väčšie množstvo vzoriek zaberie. Pri analýze diery to znamená, že do výpočtu zahrnie aj nepoškodené vzorky. Preto v spektrograme zaniká čierny pás. Prekryjú ho vypočítané koeficienty pred a za dierou. Pri rekonštrukčnom algoritme túto vlastnosť využijeme. Vzorky, ktoré sú pred a po diere by sa mali podobieť vzorkám, ktoré chceme dopočítať.

Pre DRA nastavíme hodnotu  $\gamma$  na 0,1 a u CPA  $\sigma = 0,5$ ,  $\tau = 0,5$ . Pre ukončenie sme zvolili metódu fixného počtu iterácií. Na základe 4.7 a 4.8 volíme počet iterácií 700. V oboch prípadoch sa priebeh na pohľad mení minimálne od cca tritej iterácie, no počítame aj s nejakou rezervou.

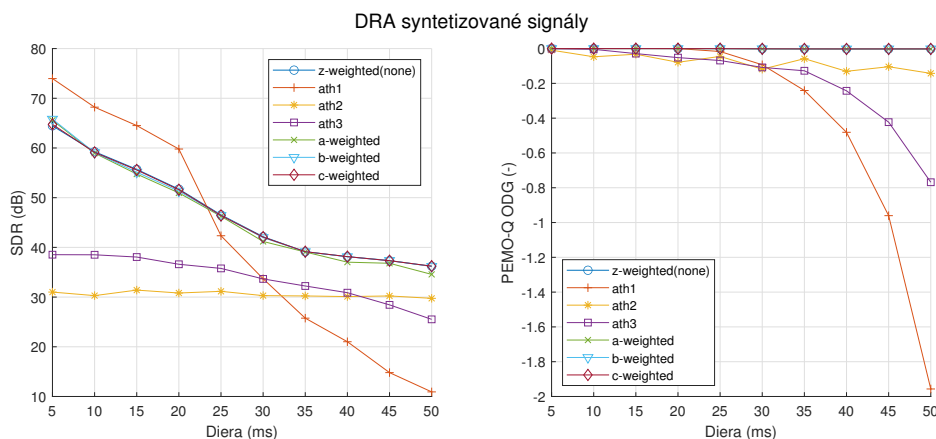
#### 4.2.5 Výsledky algoritmov

Výsledky objektívnych parametrov (SDR a ODG) pre obe metódy sú uložené v súbore `results.mat`. Pri pohľade na výsledky sú medzi reálnymi a umelými značné

rozdiely. Rekonštrukcia dopadla podľa predpokladu lepšie pre umelé signály, pretože majú väčšinou obmedzený počet harmonických zložiek, a aj ak nemajú (ako napríklad obdĺžnikový a pilovitý signál), tak sú s určitostou cez celý úsek nemenné. Fourierova analýza vráti rovnaké frekvenčné koeficienty len s rozdielnou amplitúdou v každom okne podľa toho, koľko z diery sa bude nachádzať v okne. Toto jednoznačne zvyšuje šance na úspešnosť algoritmu. Druhým predpokladom, ktorý hovorí v prospech simulovaných signálov je úplna alebo čiastočná absencia šumu. Reálne signály sú pri vzniku spracované zariadeniami, ktoré produkujú šum (i keď minimálny). Aj nahrávací priestor má svoj šum, ktorý sa do nahrávok dostane. V umelej nahrávke tieto elementy nemáme a prejaví sa to najmä na parametri SDR, ktorý je o cca 50 dB vyšší pri umelých signáloch pri najkratšej diere. Z týchto dôvodov súbor umelých a reálnych priebehov oddelíme a budeme hodnotiť separátne.

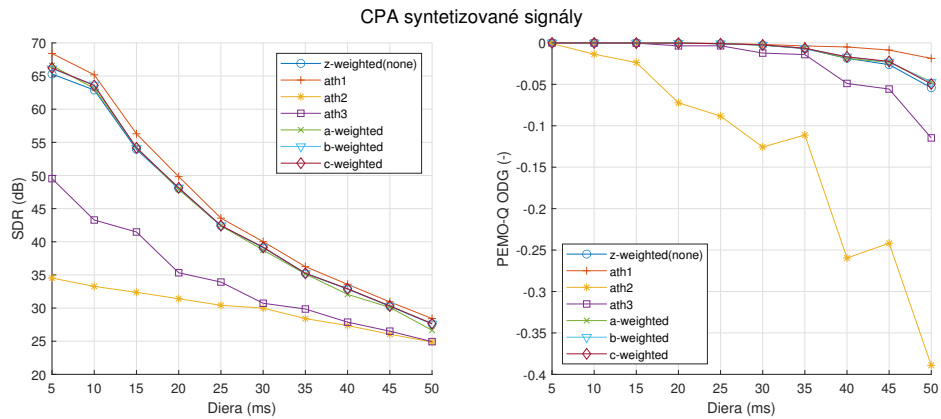
#### 4.2.6 Porovnanie syntetizovaných signálov

V obrázkoch 4.10 a 4.11 vidíme výsledky po spriemerovaní objektívnych hodnotení pre všetky syntetické signály (prvý až piaty z banky), aby sme pre každý váhový vektor zachytili približný trend v závislosti na dĺžke výrezu.



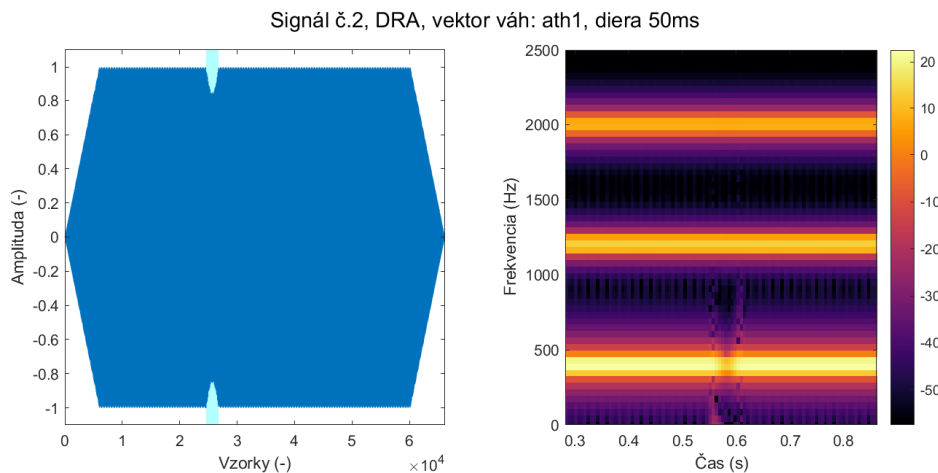
Obr. 4.10: Priebeh SDR a ODG pre umelé signály, Douglasov–Rachfordov algoritmus.

Pre váhové vektory A, B, C, Z-weighted dopadla rekonštrukcia podobne v oboch algoritmoch, a to veľmi dobre. SDR aj pri najdlhšej diere dosahuje vyššie hodnoty ako reálne signály pri najkratšom výreze. Parameter ODG je takmer stále rovný nule, teda stavu bez poškodenia. Správanie  $w_{ATH1}$  je neobvyklé. V prípade DRA začne ODG prudko klesať od 30 ms, naopak v prípade CPA dosahuje najlepšie výsledky zo všetkých vektorov. Pri detailnejšom pohľade na 3.1 môžeme vidieť, že nízkofrekvenčné koeficienty majú malé váhy (menšie ako 0,2), teda su málo kontrolované.



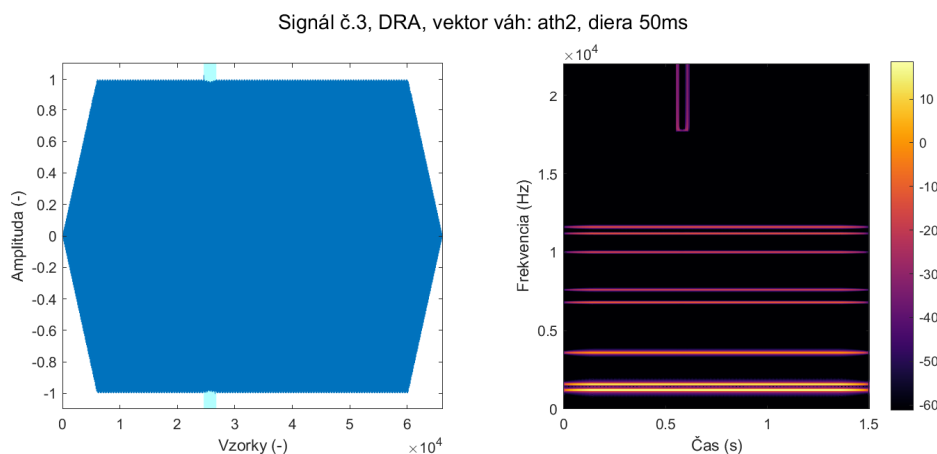
Obr. 4.11: Priebek SDR a ODG pre umelé signály, Chambollov–Pockov algoritmus.

Najvýraznejšie je to pozorovateľné na príklade druhého signálu v obrázku 4.12. Zaujímavé je, že v prípade CPA sa táto vlastnosť tak veľmi neprejaví. Pri DRA je lupnutie počutelné, no v CPA nie. Väčším problémom je pokles energie, ktorý sa prejaví najmä v strede doplneného úseku.



Obr. 4.12: Neobvyklé správanie rekonštrukcie v oblasti nízkych frekvencií pri DRA. S kratšou dierou je tento problém rovnaký, ale nie tak viditeľný. Spektrogram je priblížený na oblasť diery.

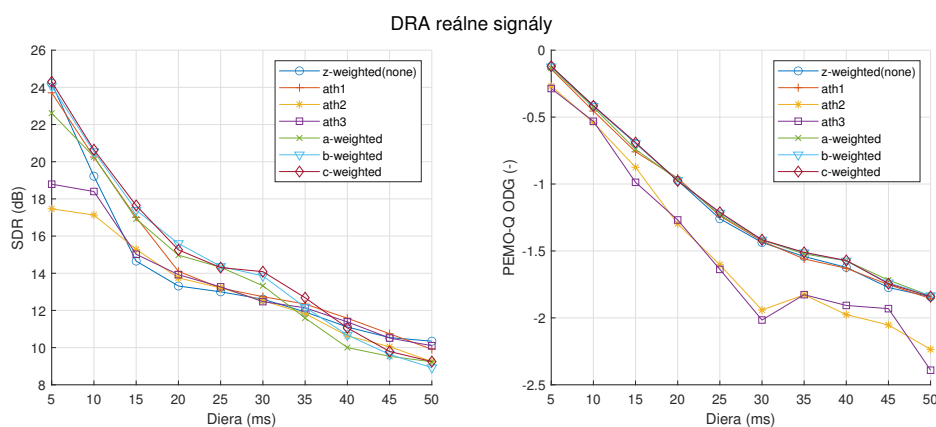
Správanie rekonštrukcie s použitím  $\mathbf{w}_{ATH2}$  a rovnako  $\mathbf{w}_{ATH3}$  má podobný priebeh, a to postupne zhoršujúci, čo je očakávané, avšak nie tak výrazne ako s použitím  $\mathbf{w}_{ATH1}$ . Problémom sú tentokrát vysokofrekvenčné elementy takisto plynúce z váhových vektorov 3.1. Približne od 15 kHz sú váhy rovné 0, čo znamená, že tieto koeficienty sú nekontrolované. V spektrograme 4.13 zreteľne vidno tieto vysokofrekvenčné zložky, ktoré vznikli práve nulovými koeficientami vo váhovom vektore.



Obr. 4.13: Zreteľné poškodenie v oblasti vysokých frekvencií.

### 4.2.7 Porovnanie reálnych signálov

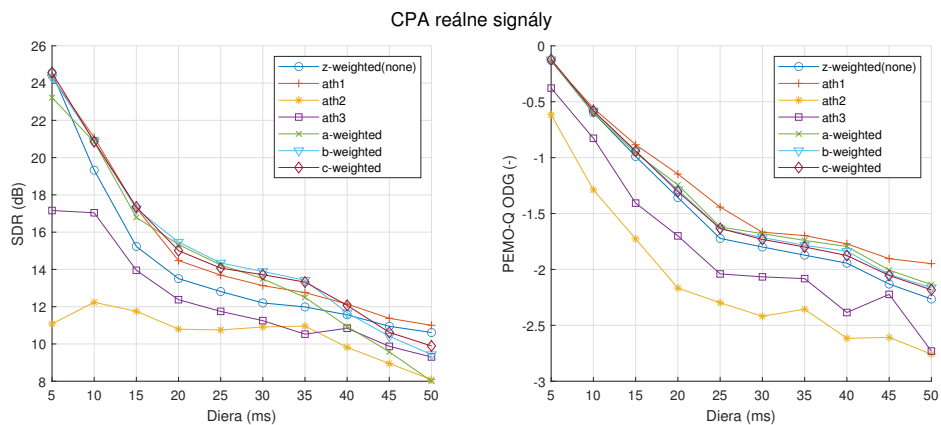
Rovnako priemerujeme výsledky všetkých reálnych signálov pre získanie trendu. Výsledky môžeme vidieť v obrázkoch 4.14 a 4.15. V oblasti reálnych signálov je správanie rekonštrukcie mierne odlišné. Medzi algoritmiami nie sú veľké rozdiely a správajú sa viacmenej obdobne. Hodnoty parametrov pri  $w_{ATH2}$  a  $w_{ATH3}$  vykazujú na prvý pohľad horšie výsledky oproti ostatným váhovým vektorom.



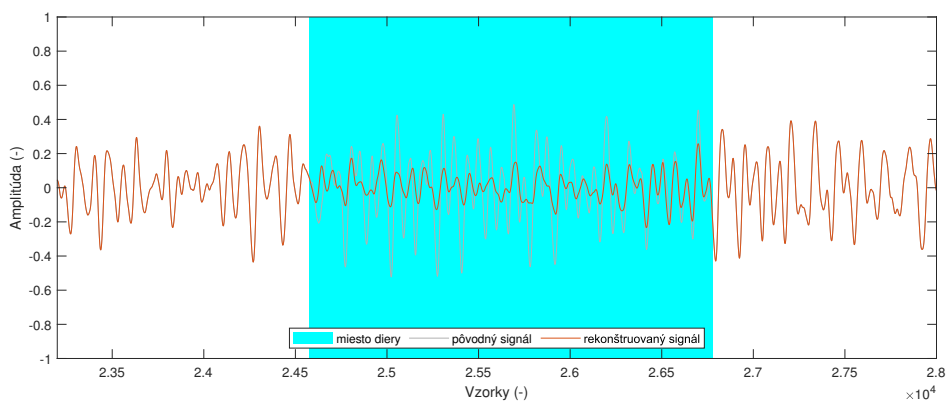
Obr. 4.14: Pribeh SDR a ODG pre umelé signály, Douglasov - Rachfordov algoritmus.

Problém s nulovými váhami v prípade  $w_{ATH2}$  a  $w_{ATH3}$  je ešte komplexnejší ako pri umelých signáloch. Pre zlepšenie by bola potrebná úprava týchto vektorov minimálne z dôvodu zmeny nulových koeficientov na nejaké malé hodnoty.

Problémom je pokles energie, ktorý ilustrujeme napríklad na signáli 8 v obrázku 4.16. Jeho riešenie je s veľkou pravdepodobnosťou s využitím psychoakustiky nedosiahnuteľné.



Obr. 4.15: Priebeh SDR a ODG pre reálne signály, Chambollov–Pockov algoritmus.



Obr. 4.16: Pokles energie pri váhovom vektore ATH1, dĺžka diery 50 ms.

## 5 Záver

Táto práca sa zaoberá rekonštrukciou krátkeho úseku zvukového signálu, ktorý je poškodený alebo chýba. Miesto chýbajúcich vzoriek je identifikované už pred rekonštrukciou a nehľadáme ho.

Vytvorili sme funkčný rekonštrukčný algoritmus v Matlabe, ktorý využíva ako jadro Douglasov–Rachfordov algoritmus a Chambollov–Pockov algoritmus. Rekonštruovali sme súbor pätnástich signálov, z ktorých sme vyberali krátke úseky s dĺžkou 5 až 50 ms a umelo simulovali výpadok signálu vložением vzoriek s nulovou amplitúdou. Ako objektívne hodnotiteľa boli použité: jednoduchšia metóda SDR a sofistikovanejšia PEMO-Q, z ktorej sme používali parameter ODG. Subjektívne hodnotenie poslucháča slúžilo len na potvrdenie hodnotenia ODG.

Ak by sme mali zhodnotiť výsledky oboch algoritmov, tak sú si podobné a každý vynikol v inej oblasti. V syntetických signáloch mal lepšie výsledky CPA, naopak pri reálnych DRA. Zapojenie psychoakustického modelu prebiehalo pomocou váhovania frekvenčných koeficientov. Cieľom bolo kontrolovať, čo je počuteľné. Celkovo bolo predstavených 6 váhových vektorov. V globále tieto vektory nepriniesli zásadné zlepšenie výsledkov v porovnaní s variantom bez váhovania koeficientov. Váhové vektory  $\mathbf{w}_{\text{ATH}}$  priniesli vo výsledkoch skôr zhoršenie. Na zlepšenie by bolo potrebné krivky pozmeniť a zbaviť sa nulových koeficientov na začiatku a konci frekvenčného spektra. Váhové vektory  $\mathbf{w}_{\text{A,B,C-weighted}}$  mierne zlepšenie výsledkov priniesli, ale nejedná sa o tak markantné zlepšenie, aby sa to prejavilo subjektívne lepšie.

Očakávaním bolo, že váhovanie prinesie zlepšenie, no v prípade doplnenia chýbajúceho úseku sa to neprejavilo. Pri úlohe declippingu [7] bol problém s vysoko-frekvenčným šumom, ktoré váhy dokázali eliminovať. V našom prípade je najväčším problémom pokles energie, čo ilustruje obrázok 4.12 a 4.16 a nie je úplne zreteľné, prečo by malo práve psychoakustické váhovanie koeficientov pomôcť.



# Literatúra

- [1] ADLER, Amir, Valentin EMIYA, Maria G. JAFARI, Michael ELAD, Rémi GRIBONVAL a Mark D. PLUMBLEY. Audio Inpainting. *IEEE Transactions on Audio, Speech, and Language Processing* [online]. 2012, **20**(3), 922-932 [cit. 2020-11-19]. ISSN 1558-7916. Dostupné z: doi:10.1109/TASL.2011.2168211
- [2] RAJMIC, Pavel a Marie DAŇKOVÁ. *Úvod do řídkých reprezentací signálů a komprimovaného snímání* [online]. 1. Brno: Vysoké učení technické v Brně Fakulta elektrotechniky a komunikačních technologií Ústav telekomunikací, 2014 [cit. 2020-11-18]. ISBN 978-80-214-5169-8. Dostupné z: [https://www.utko.fekt.vut.cz/~rajmic/skripta/Skripta-Ridke\\_reprezentace-Rajmic\\_Dankova.pdf](https://www.utko.fekt.vut.cz/~rajmic/skripta/Skripta-Ridke_reprezentace-Rajmic_Dankova.pdf)
- [3] COMBETTES, Patrick L. a Jean-Christophe PESQUET. Proximal Splitting Methods in Signal Processing. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering* [online]. New York, NY: Springer New York, 2011, 2011-5-9, , 185-212 [cit. 2020-11-30]. Springer Optimization and Its Applications. ISBN 978-1-4419-9568-1. Dostupné z: doi:10.1007/978-1-4419-9569-8\_10
- [4] KITIĆ, Srđan, Nancy BERTIN a Rémi GRIBONVAL. Sparsity and Cosparsity for Audio Declipping: A Flexible Non-convex Approach. *Latent Variable Analysis and Signal Separation* [online]. Cham: Springer International Publishing, 2015, 2015-8-15, , 243-250 [cit. 2020-12-10]. Lecture Notes in Computer Science. ISBN 978-3-319-22481-7. Dostupné z: doi:10.1007/978-3-319-22482-4\_28
- [5] MOKRY, Ondrej, Pavel ZAVISKA, Pavel RAJMIC a Vitezslav VESELY. Introducing SPAIN (SParse Audio INpainter). *2019 27th European Signal Processing Conference (EUSIPCO)* [online]. IEEE, 2019, 2019, , 1-5 [cit. 2020-12-10]. ISBN 978-9-0827-9703-9. Dostupné z: doi:10.23919/EUSIPCO.2019.8902560
- [6] MOKRÝ, Ondřej. *Moderní metody restaurace poškozených audiosignálů* [online]. Brno, 2019 [cit. 2021-5-18]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/113107>. Diplomová práce. Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav matematiky. Vedúci práce Doc. Mgr. Pavel Rajmic, Ph.D.
- [7] ZAVISKA, Pavel, Pavel RAJMIC a Jiri SCHIMMEL. Psychoacoustically Motivated Audio Declipping Based on Weighted  $l_1$  Minimization. *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)* [online]. IEEE, 2019, 2019, , 338-342 [cit. 2020-11-18]. ISBN 978-1-7281-1864-2. Dostupné z: <https://ieeexplore.ieee.org/document/8769109/>

- [8] KOVAČEVIĆ, Jelena a Amina CHEBIRA. An Introduction to Frames. *Foundations and Trends® in Signal Processing* [online]. 2007, **2**(1), 1-94 [cit. 2020-12-04]. ISSN 1932-8346. Dostupné z: doi:10.1561/20000000006
- [9] ZÖLZER, Udo, 2008. *Digital audio signals processing*. 2nd ed. Chichester: John Wiley & Sons. ISBN 978-0-470-99785-7.
- [10] SPANIAS, Andreas, Ted PAINTER a Venkatraman ATTI, 2007. *Audio signal processing and coding*. 1. Hoboken: John Wiley. ISBN 978-0-471-79147-8.
- [11] ZWICKER, E. Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). *The Journal of the Acoustical Society of America* [online]. 1961, **33**(2), 248-248 [cit. 2021-5-18]. ISSN 0001-4966. Dostupné z: doi:10.1121/1.1908630
- [12] FLETCHER, Harvey a W. A. MUNSON. Loudness, Its Definition, Measurement and Calculation. *Bell System Technical Journal* [online]. 1933, **12**(4), 377-430 [cit. 2021-5-17]. ISSN 00058580. Dostupné z: doi:10.1002/j.1538-7305.1933.tb00403.x
- [13] *Acoustics - Normal equal-loudness-level contours*. ISO 226. Geneva, Switzerland: International Standardization Organization, 2003 [cit. 2021-16-5]
- [14] BUREAU OF INDIAN STANDARDS. *Electroacoustics - sound level meters: Part 1 : Specifications*. 1. Delhi: Simco Printing Press, 2005. 15575. Dostupné z: <http://www.proviento.com.co/IEC61672.pdf>
- [15] HUSNÍK, Libor a Marcelo HERRERA MARTÍNEZ. *Comparison of Two Methods Used for Subjective Evaluation of Compressed Sound Signals* [online]. 2005 [cit. 2021-5-8]. Dostupné z: [https://www.researchgate.net/publication/238111375\\_Comparison\\_of\\_Two\\_Methods\\_Used\\_for\\_Subjective\\_Evaluation\\_of\\_Compressed\\_Sound\\_Signals](https://www.researchgate.net/publication/238111375_Comparison_of_Two_Methods_Used_for_Subjective_Evaluation_of_Compressed_Sound_Signals)
- [16] ITU-T. *Methods for subjective determination of transmission quality*. P.800. Udelené 1996 [cit. 2021-17-5]. Dostupné z: <http://handle.itu.int/11.1002/1000/3638>
- [17] THOMA, Herbert. A system for subjective evaluation of audio, video and audiovisual quality using MUSHRA and SAMVIQ methods. In: *2012 Fourth International Workshop on Quality of Multimedia Experience* [online]. IEEE, 2012, 2012, s. 31-32 [cit. 2021-5-17]. ISBN 978-1-4673-0726-0. Dostupné z: doi:10.1109/QoMEX.2012.6263877

- [18] XIAOPING, Hu, He GUIMING a Zhu XIAOPING. PEAQ-based psychoacoustic model for perceptual audio coder. In: *2006 8th International Conference Advanced Communication Technology* [online]. IEEE, 2006, 2006, 5 pp.-1823 [cit. 2021-5-17]. ISBN 89-5519-129-4. Dostupné z: doi:10.1109/ICACT.2006.2063444
- [19] HUBER, Rainer a Birger KOLLMEIER, 2006. PEMO-Q-A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Transactions on Audio, Speech and Language Processing* [online]. **14**(6), 1902-1911 [cit. 2021-03-08]. ISSN 1558-7916. Dostupné z: doi:10.1109/TASL.2006.883259
- [20] TAN, Chin-Tuan a Brian C.J. MOORE. Perception of nonlinear distortion by hearing-impaired people. *International Journal of Audiology* [online]. 2009, **47**(5), 246-256 [cit. 2021-5-17]. ISSN 1499-2027. Dostupné z: doi:10.1080/14992020801945493
- [21] GAOXIONG, Yi a Zhang WEI. The Perceptual Objective Listening Quality Assessment algorithm in telecommunication: Introduction of ITU-T new metrics POLQA. In: *2012 1st IEEE International Conference on Communications in China (ICCC)* [online]. IEEE, 2012, 2012, s. 351-355 [cit. 2021-5-17]. ISBN 978-1-4673-2815-9. Dostupné z: doi:10.1109/ICCChina.2012.6356906
- [22] CHINEN, Michael, Felicia S. C. LIM, Jan SKOGLUND, Nikita GUREEV, Feargus O'GORMAN a Andrew HINES, 2020. ViSQOL v3: An Open Source Production Ready Objective Speech and Audio Metric. In: *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)* [online]. IEEE, 2020, s. 1-6 [cit. 2021-5-3]. ISBN 978-1-7281-5965-2. Dostupné z: doi:10.1109/QoMEX48832.2020.9123150
- [23] ITU-R. *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*. BS.1116-1. Udelené 1997. Dostupné z: [https://www.itu.int/dms\\_pubrec/itu-r/rec/bs/R-REC-BS.1116-1-199710-S!!PDF-E.pdf](https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-1-199710-S!!PDF-E.pdf)
- [24] MOKRY, Ondrej a Pavel RAJMIC. Audio Inpainting: Revisited and Reweighted. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* [online]. 2020, **28**, 2906-2918 [cit. 2020-12-10]. ISSN 2329-9290. Dostupné z: doi:10.1109/TASLP.2020.3030486
- [25] *Sound Quality Assessment Material recordings for subjective tests*. 3. doplnené vydanie. Geneve, Switzerland: EBU – TECH, 2008. Dostupné z: <https://tech.ebu.ch/docs/tech/tech3253.pdf>

- [26] RAJMIC, Pavel, Hana BARTLOVA, Zdenek PRUSA a Nicki HOLIGHAUS. Acceleration of audio inpainting by support restriction. In: *2015 7th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* [online]. Brno: IEEE, 2015, 2015, s. 325-329 [cit. 2021-5-24]. ISBN 978-1-4673-9283-9. ISSN 2157-0221. Dostupné z: doi:10.1109/ICUMT.2015.7382451

## Zoznam symbolov, veličín a skratiek

<b>CPA</b>	Chambollov–Pockov algoritmus
<b>DGT</b>	diskrétna Gaborova transformácia
<b>DRA</b>	Douglasov–Rachfordov algoritmus
<b>FFT</b>	fast Fourier transform - rýchla Fourierova transformácia
<b>IDGT</b>	inverzná diskretná Gaborova transformácia
<b>ODG</b>	objective difference grade
<b>PEMO-Q</b>	Perception model – Quality assesment
<b>SDR</b>	signal-to-distortion ratio – odstup signálu od poškodenia
<b>SPL</b>	sound pressure level
<b>SDG</b>	subjective difference grade

# A Zoznam príloh

V adresári `priloha.zip` nájdeme:

```
/.....kořenový adresár priloženého archívu
├── data.....výsledky algoritmov
│   ├── results.mat
│   └── resultsmean.mat
├── PemoQ.....Balíček pre perceptívne motivované hodnotenie
├── ltfat.....toolbox LTFAT
├── comparealgo.m
├── DouglasRachford.m
├── gapmaker.m
├── ChambollePock.m
├── Main.m..... Hlavný skript na spustenie algoritmov
├── Maintest.m..... Hlavný skript na prezeranie výsledkov
├── SDR.m
├── soundspectro.m
├── synthsignals.m
└── weights.m
```

V priečinku nie sú dostupné tri súbory, kvôli veľkosti, preto sú dostupné spoločne s celým vlastným kódom v repozitári:

<https://github.com/michalsvento/RestorationPsychoacousticmodel>.

Súbor `EBU_SQAM.mat` je potrebný na spustenie hlavného skriptu, je potrebné nakopírovať ho do koreňového adresára. Súbory `CPsignals.mat` a `DRsignals.mat` by sa mali nachádzať v priečinku `data` a tam je potrebné ich skopírovať. Bez týchto súborov nie je možné spustenie funkcie `soundspectro.m`. Je potrebné tieto dva súbory nakopírovať do priečinku `data`. Detailnejší popis všetkých funkcií je v README na GitHubu a takisto vždy na začiatku každého kódu.

Na tvorbu kódu sme využili nasledujúci software:

- MATLAB R2019b
- toolbox LTFAT vo verzii 2.4.0, dostupné z: <https://ltfat.github.io/>
- Pemo-Q vo verzii 1.4.0