

# VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ  
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS

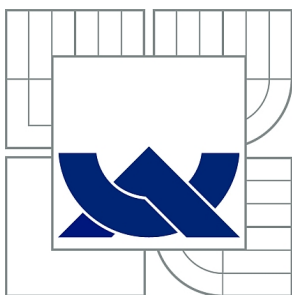
POKROČILÉ METODY INTERPOLACE ZVUKOVÝCH SIGNÁLŮ

DIPLOMOVÁ PRÁCE  
MASTER'S THESIS

AUTOR PRÁCE  
AUTHOR

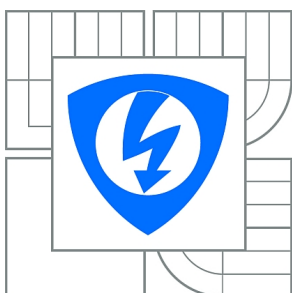
Bc. JIŘÍ POSPÍŠIL

BRNO 2014



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY



**FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH  
TECHNOLOGIÍ**

**ÚSTAV TELEKOMUNIKACÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION  
DEPARTMENT OF TELECOMMUNICATIONS

# **POKROČILÉ METODY INTERPOLACE ZVUKOVÝCH SIGNÁLŮ**

ADVANCED METHODS OF AUDIO SIGNALS INTERPOLATION

**DIPLOMOVÁ PRÁCE**

MASTER'S THESIS

**AUTOR PRÁCE**

AUTHOR

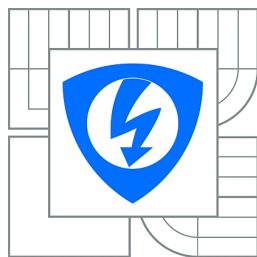
**Bc. JIŘÍ POSPÍŠIL**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**Ing. VÁCLAV MACH**

BRNO 2014



VYSOKÉ UČENÍ  
TECHNICKÉ V BRNĚ

Fakulta elektrotechniky  
a komunikačních technologií

Ústav telekomunikací

# Diplomová práce

magisterský navazující studijní obor  
Telekomunikační a informační technika

**Student:** Bc. Jiří Pospíšil

**ID:** 125605

**Ročník:** 2

**Akademický rok:** 2013/2014

## NÁZEV TÉMATU:

**Pokročilé metody interpolace zvukových signálů**

## POKYNY PRO VYPRACOVÁNÍ:

Implementujte metodu pro interpolaci zvukového signálu založenou na modelování parametrů signálu [1,2]. Pro hromadné testování metod pro doplňování chybějících vzorků v audio signálu sestavte testovací scénář a skript na výpočetním serveru Blade. Vyhodnoťte úspěšnost rekonstrukce různých rekonstrukčních technik objektivními metodami SNR a PEMO-Q.

## DOPORUČENÁ LITERATURA:

[1] M. Lagrange, S. Marchand, J.-b. Rault, Long interpolation of audio signals using linear prediction in sinusoidal modeling, J.Audio Eng. Soc 53 (10) (2005) 891–905. URL <http://www.aes.org/e-lib/browse.cfm?elib=13390>

[2] A. Lukin, J. Todd, Parametric interpolation of gaps in audio signals, in: Audio Engineering Society Convention 125, 2008. URL <http://www.aes.org/e-lib/browse.cfm?elib=14664>

[3] R.Huber, B.Kollmeier, PEMO-Q - A new method for objective audio quality assessment using a model of auditory perception, IEEE Trans. Audio Speech Language Proc. 14 (6) (2006) 1902–1911. doi:10.1109/TASL.2006.883259.

**Termín zadání:** 10.2.2014

**Termín odevzdání:** 30.5.2014

**Vedoucí práce:** Ing. Václav Mach

**Konzultanti diplomové práce:**

**doc. Ing. Jiří Mišurec, CSc.**

*Předseda oborové rady*

**UPOZORNĚNÍ:**

Autor diplomové práce nesmí při vytváření diplomové 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.

### **ANOTACE**

Tato diplomová práce se zabývá teoretickým rozbohem predikční metody interpolace signálu a modelováním signálu pomocí sinusoidálního modelu. Na základě této teorie je ve výpočetním prostředí MATLAB realizován algoritmus pro rekonstrukci chybějících úseků ve zvukovém signálu. Pomocí hromadného testování jsou zobrazeny výsledky rekonstrukce pomocí objektivních metod SNR a PEMO-Q. Dále jsou provedeny experimenty na konkrétních signálech a popsáno jejich vyhodnocení.

**Klíčová slova:** lineární predikce, Burgova metoda, sinusoidální modelování, interpolace, extrapolace, MATLAB

### **ABSTRACT**

This diploma thesis deals with the theoretical analysis of the predictive methods of signal interpolation and signal modeling using sinusoidal model. On the basis of this theory the algorithm for the reconstruction of the missing sections in the audio signal is implemented in computing environment MATLAB. Results of mass testing reconstructions are displayed using objective methods SNR and PEMO-Q. Further experiments are carried out on single signals and their evaluation is described.

**Keywords:** linear prediction, Burg method, sinusoidal modeling, interpolation, extrapolation, MATLAB

POSPÍŠIL, J. *Pokročilé metody interpolace zvukových signálů*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2014. 57 s. Vedoucí diplomové práce Ing. Václav Mach.

## **Prohlášení**

Prohlašuji, že svoji diplomovou práci na téma „Pokročilé metody interpolace zvukových signálů“ jsem vypracoval samostatně pod vedením vedoucího diplomové práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

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

V Brně dne .....

.....

podpis autora

## **Poděkování**

Děkuji vedoucímu práce Ing. Václavu Machovi za velmi užitečnou metodickou pomoc a cenné rady při zpracování diplomové práce.

V Brně dne .....

.....

podpis autora





Faculty of Electrical Engineering  
and Communication  
Brno University of Technology  
Technická 12, CZ-61600 Brno  
Czech Republic  
<http://www.six.feec.vutbr.cz>

## PODĚKOVÁNÍ

Výzkum popsany v této diplomové práci byl realizován v laboratořích podpořených z projektu SIX; registrační číslo CZ.1.05/2.1.00/03.0072, operační program Výzkum a vývoj pro inovace.

Brno .....

.....

(podpis autora)



EVROPSKÁ UNIE  
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ  
INVESTICE DO VAŠÍ BUDOUCNOSTI



# Obsah

SEZNAM OBRÁZKŮ.....	10
SEZNAM TABULEK.....	11
ÚVOD .....	12
1 INTERPOLACE AUDIO SIGNÁLU .....	13
2 PREDIKČNÍ METODY.....	14
2.1 Lineární predikce .....	15
2.2 Burgova metoda.....	16
2.2.1 Implementace Burgova algoritmu.....	17
3 SINUSOIDÁLNÍ MODELOVÁNÍ.....	18
3.1 Predikce kmitočtových složek.....	19
3.2 Spojování složek .....	20
3.3 Interpolace spojených kmitočtových složek.....	22
3.3.1 Interpolace frekvence.....	22
3.3.2 Interpolace amplitudy.....	22
3.3.3 Interpolace fáze .....	23
3.4 Extrapolace nespojených kmitočtových složek.....	25
3.5 Interpolace residua.....	26
4 Metody objektivního hodnocení kvality rekonstrukce .....	27
4.1 SNR.....	27
4.2 PEMO-Q .....	27
5 IMPLEMENTACE INTERPOLAČNÍ METODY.....	28
5.1 Popis vstupní struktury .....	28
5.2 Popis výstupní struktury .....	30
5.3 Popis funkce SinModInt .....	31
5.3.1 Realizace výpočtu fáze .....	35
.....	36

5.4 Spouštěcí skripty .....	36
6 VÝSLEDKY EXPERIMENTŮ .....	38
6.1 Výsledky SNR .....	38
6.1.1 Výsledky SNR řečových signálů .....	39
6.2.2 Výsledky SNR hudebních signálů .....	40
6.1.3 Shrnutí výsledků SNR .....	41
6.2 Výsledky PEMO-Q .....	41
6.2.1 Výsledky PEMO-Q řečových signálů .....	42
6.2.2 Výsledky PEMO-Q hudebních signálů .....	42
6.2.3 Shrnutí výsledků PEMO-Q .....	43
6.3 Praktické ukázky rekonstruovaných signálů .....	44
6.4 Shrnutí výsledků experimentů .....	50
7 ZÁVĚR .....	52
LITERATURA .....	53
SEZNAM ZKRATEK .....	56
SEZNAM PŘÍLOH .....	57
A OBSAH ELEKTRONICKÉ PŘÍLOHY .....	58

## SEZNAM OBRÁZKŮ

Obr. 1: Křížová struktura lineárního predikčního filtru 1. řádu .....	16
Obr. 2: Analýza složek signálu .....	18
Obr. 3: Algoritmus pro výpočet odhadovaných hodnot .....	19
Obr. 4: Časový průběh a spektrogram původního a rekonstruovaného signálu s posunutou fází .....	35
Obr. 5: Časový průběh a spektrogram původního a rekonstruovaného signálu bez posunu fáze .....	36
Obr. 6: Závislost SNR na délce mezery a parametru $\text{thr}_F$ u řečových signálů .....	39
Obr. 7: Závislost SNR na délce mezery a parametru $\text{thr}_F$ u hudebních signálů .....	40
Obr. 8: Závislost $\text{PSM}_t$ na prahu vzdálenosti vektorů frekvence protějšších složek u řečového signálu .....	42
Obr. 9: Závislost $\text{PSM}_t$ na prahu vzdálenosti vektorů frekvence protějšších složek u hudebního signálu .....	43
Obr. 10: Závislost aritmetických průměrů hodnoty $\text{PSM}_t$ na prahu vzdálenosti vektorů frekvence protějšších složek .....	44
Obr. 11: Rekonstrukce mezery s délkou $l = 160$ vzorků stacionárního úseku signálu .....	45
Obr. 12: Extrapolace signálu z levé strany do mezery délky $l = 1600$ vzorků .....	45
Obr. 13: Extrapolace signálu z obou stran do mezery délky $l = 1600$ vzorků .....	46
Obr. 14: Interpolace mezery s délkou $l = 1600$ vzorků .....	46
Obr. 15: Rekonstrukce úseku o délce $l = 1600$ vzorků u signálu s větší spektrální hustotou .....	47
Obr. 16: Rekonstrukce úseku o délce $l = 3200$ vzorků u signálu s větší spektrální hustotou .....	47
Obr. 17: Rekonstrukce mezery o délce $l = 1600$ vzorků u signálu s dlouhými kmitočtovými složkami .....	48
Obr. 18: Rekonstrukce mezery o délce $l = 6400$ vzorků u signálu s dlouhými kmitočtovými složkami .....	49
Obr. 19: Rekonstrukce mezery s délkou $l = 800$ vzorků v řečovém signálu .....	49
Obr. 20: Rekonstrukce mezery s délkou $l = 840$ vzorků v řečovém signálu .....	50

## SEZNAM TABULEK

Tab. 1: Struktura vstupního parametru.....	29
Tab. 2: Struktura výstupního parametru .....	30
Tab. 3: Názvy souborů pro hromadné testování .....	38
Tab. 4: Délky chybějících úseků .....	39
Tab. 5: Ukázka aritmetických průměrů výsledků hodnot SNR souboru music02_16kHz.wav.....	41

# ÚVOD

Jako transportní médium pro přenos zvuku se v dnešní době stále častěji využívá digitální počítačové sítě (internet, intranet). Analogovou telefonní sítí nahrazuje technika VoIP (Voice over IP). Při přenosu pomocí těchto technologií může nastat problém ve formě ztráty paketů v neoptimálních síťových podmínkách. Použitím protokolů pro přenos v reálném čase, kdy nedochází k opakovanému vysílání ztracených paketů, tedy přijímáme signál s chybějícími vzorky. Délka úseku chybějících vzorků se většinou pohybuje pouze v řádu do několika desítek milisekund, avšak i za takové situace pak při poslechu přeneseného signálu dojde k nepříjemnému zvukovému jevu. Pro odstranění těchto problémů proto existuje několik algoritmů, které jsou schopny odhadnout a dopočítat vzorky takového signálu a to s různými výsledky, výhodami a nevýhodami.

Tato práce se zaměřuje na odhad jednotlivých složek signálu (frekvence, amplituda, fáze) pomocí lineární predikce. Pro odhad predikčních koeficientů je použita Burgova metoda. Pomocí této metody získáme odhad jednotlivých složek signálu a residuálního šumu, z čehož získáme signál pro doplnění chybějícího úseku.

V kapitolách 1 až 3 je rozebrán teoretický základ rekonstrukční metody. V kapitole 4 jsou představeny metody, kterými jsou kvality rekonstrukce porovnávány. Následuje kapitola popisující realizaci algoritmu ve výpočetním prostředí MATLAB. V závěrečné části práce je ukázka rekonstrukce několika průběhů signálu pomocí implementované metody s různým nastavením parametrů a přehled výsledků rekonstrukce za použití těchto parametrů.

# 1 INTERPOLACE AUDIO SIGNÁLU

Jednou ze základních metod pro rekonstrukci chybějícího signálu či odstranění kliku je interpolace, která vymění poškozené vzorky za odhad jejich skutečné hodnoty. Úkolem je tedy vyplnit mezeru vhodným úsekem signálu, který má odpovídající délku jako poškozená část. Tyto úseky mohou obklopovat poškozený blok jak z levé tak i pravé strany. To znamená, že se interpolace dá provádět z předchozích i následujících vzorků, což lze popsat následujícím způsobem [2].

Vektor  $\mathbf{x}$  představuje signál o délce  $N$  vzorků. Tento vektor lze rozdělit na další dva vektory, a to vektor  $\mathbf{x}_-$ , jehož hodnoty známe, a vektor  $\mathbf{x}_m$ , který obsahuje neznámé prvky. Těmi může být úsek o délce  $l$  vzorků začínající na vzorku  $m$ , který obsahuje poškozené nebo chybějící vzorky. Tyto vektory jdou dále rozepsat na tři sekce. První je sekce s neznámými vzorky  $\mathbf{x}_m = (x[m], x[m+1], \dots, x[m+l-1])$ , dále  $m$  vzorků nalevo od mezery  $\mathbf{x}_{-l} = (x[1], x[2], \dots, x[m-1])$  a zbývající známé vzorky napravo od mezery  $\mathbf{x}_{-r} = (x[m+l], \dots, x[N])$ . Tyto sekce dohromady tvoří vektor  $\mathbf{x} = (\mathbf{x}_{-l} \mathbf{x}_m \mathbf{x}_{-r})$ . Interpolace tedy spočívá v odhadu správných hodnot vektoru  $\mathbf{x}_m$  z hodnot neporušených vzorků vektoru  $\mathbf{x}_- = (\mathbf{x}_{-l} \mathbf{x}_{-r})$ .

Existuje několik různých metod [1,10,12,13], kde nejjednodušší jsou schopny rekonstruovat pouze mezery o počtu několika málo vzorků. Jiné metody například mezeru od užitečného signálu odříznou a signál návazně spojí. U těchto metod je však kladen důraz na periodickou podstatu řečového a hudebního signálu a spolehlivý odhad délky opakujících se period. Zde se zaměříme na metody odhadu signálu pomocí lineární predikce a parametrického modelu signálu.

## 2 PREDIKČNÍ METODY

Predikční metody jsou rozděleny podle toho, jak přistupujeme ke zpracovávanému signálu. Jednou z nich je zpracovávání po jednotlivých vzorcích a to za použití přímé a křížové struktury. Druhou metodou je zpracování po blocích o délce  $N$ , kde k nalezení optimálních predikčních koeficientů je zapotřebí minimalizace predikční chyby za použití celého bloku, ne pouze předchozího vzorku.

K dosažení odhadovaných hodnot  $\hat{x}[n]$  lze dojít pomocí určení vhodného modelu procesu, který vytvořil známou část signálu  $x[n]$ . Výsledek lze popsat jako superpozice předchozích vážených výstupních hodnot systému modelu a současné vstupní hodnoty  $x[n]$

$$\begin{aligned} y[n] &= \frac{1}{a_0} (a_1 y[n-1] + \dots + a_p y[n-p]) + x[n] \\ &= \frac{1}{a_0} \left( \sum_{i=1}^p a_i y[n-p] \right) + x[n], \end{aligned} \tag{1}$$

kde  $a_i$  jsou váhové koeficienty a  $p$  je řád modelu. Pomocí Z-transformace pak získáme přenosovou funkci

$$H(z) = \frac{1}{\sum_{i=1}^p a_i z^{-i}}, \tag{2}$$

což je zároveň přenosová funkce lineárního predikčního modelu [1]. Tento model se také používá například v LPC kóděrech pro syntézu řeči. Koeficienty  $a_i$  jsou v tomto případě lineární predikční koeficienty. Rovnice (2) však také představuje přenosovou funkci rekurzivního IIR filtru, jehož parametry jsou váhové koeficienty  $a_i$ . Odhadovaný signál  $\hat{x}[n]$  můžeme pomocí tohoto filtru získat jako extrapolaci známého úseku  $x[n]$ . Zásadní je tedy vybrat si správný parametrický model, jehož koeficienty aplikujeme jako parametry filtru. Poté na vstup filtru přivedeme nulový vektor o délce  $l$

$$[x(m), \dots, x(m+l-1)] = 0, \tag{3}$$



kdy proces začíná na  $m$ -tém vzorku.

## 2.1 Lineární predikce

Zde jako základní blok parametrického modelu používáme lineární predikční model [1]. V momentě, kdy chceme z předchozích získaných hodnot odhadnout následující hodnotu  $x[n]$ , použijeme lineární predikci řádu  $p$

$$\hat{x}[n] = - \sum_{i=1}^p a_i x[n-i]. \quad (4)$$

Vzorky  $x[n-1], x[n-2], \dots, x[n-p]$  vynásobíme váhovými koeficienty  $a_1, a_2, \dots, a_p$  a získáme tak odhad nových hodnot. Rozdíl mezi současnou hodnotou  $x[n]$  a odhadovanou hodnotou  $\hat{x}[n]$  se nazývá chyba predikce

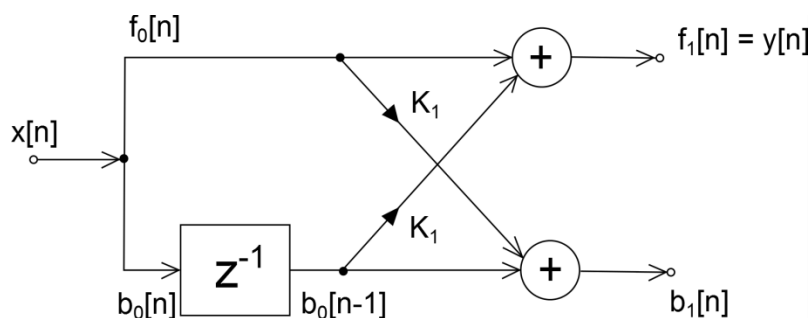
$$e[n] = x[n] - \hat{x}[n]. \quad (5)$$

Koeficienty  $a_i$  by měly být vybírány tak, aby energie predikční chyby  $e[n]^2$  byla co nejmenší. Podle [4] se chyba dopředné lineární predikce značí

$$f_p[n] = x[n] + \sum_{i=1}^p a_i x[n-i]. \quad (6)$$

Predikční koeficienty lze získat pomocí číslicového FIR filtru s přenosovou funkcí

$$H(z) = \sum_{i=1}^p a_i z^{-i}, \quad a_0 = 1. \quad (7)$$



Obr. 1: Křížová struktura lineárního predikčního filtru 1. řádu

Pro implementaci je vhodnější křížová struktura filtru, která je z přímé struktury odvozena [4]. Chceme-li vytvořit křížový filtr  $p$ -tého řádu, musíme do kaskády naskládat  $p - 1$  sekcí, což odpovídá FIR filtru řádu  $N - 1$ . Obecně tedy pro filtr  $p$ -tého řádu platí rovnice uvedena v [3]

$$\begin{aligned}
 f_0[n] &= b_0[n] = x[n], \\
 f_p[n] &= K_p b_{p-1}[n-1] + f_{p-1}[n], \\
 b_p[n] &= K_p^* f_{p-1}[n] + b_{p-1}[n-1], \quad \text{pro } p = 1, 2, 3, \dots, N-1.
 \end{aligned} \tag{8}$$

Koeficienty  $K_p$  jsou takzvané koeficienty odrazu. Mezi koeficienty  $a_i$  filtrů s přímou strukturou a koeficienty  $K_p$  existuje vztah převzatý z [4]

$$\begin{aligned}
 a_p[0] &= 1 \quad \text{pro } p = 1, 2, \dots, N-1 \\
 a_p[p] &= K_p \\
 a_p[i] &= a_{p-1}[i] + K_p a_{p-1}^*[p-i] \\
 \text{pro } 1 &\leq i \leq p-1, \quad p = 2, \dots, N-1.
 \end{aligned} \tag{9}$$

## 2.2 Burgova metoda

Tato metoda patří mezi ty, které signál zpracovávají po blocích a vychází z křížové formy lineární predikce.

K získání optimálních predikčních koeficientů v rámci bloku musíme dle [1] nalézt takové, aby suma energie chyb všech  $p$  křížových sekcí  $J_i = \sum_{n=i}^{N-1} (f_i^2[n] + b_i^2[n])$  byla minimální. Toho dosáhneme nastavením derivace podle požadovaného predikčního koeficientu na nulu

$$\frac{\partial J_i}{\partial k_i} \stackrel{\text{def}}{=} 0. \quad (10)$$

Vyřešením dostaneme tzv. Burgovu rovnici,

$$k_i = -\frac{2 \sum_{n=i}^{N-1} (f_{i-1}[n]b_{i-1}[n-1])}{\sum_{n=i}^{N-1} (f_{i-1}^2[n] + b_{i-1}^2[n-1])}, \quad (11)$$

kteřou můžeme rekurzivně použít k výpočtu koeficientů odrazu  $k_i$  pro všechna  $i = 1, \dots, p$ . Po každém výpočtu  $k_i$  se musí u všech křížových sekcí aktualizovat koeficient odrazu podle nově vypočítaného, dle (8).

Tato metoda je vhodná pro extrapolaci dat za pomoci pouze několika dostupných vzorků a to získáním predikčních koeficientů  $a_i$  následujícím algoritmem.

### 2.2.1 Implementace Burgova algoritmu

V zaváděcí fázi Burgova algoritmu načteme podle (8) počáteční hodnoty  $f_0$  a  $b_0$ . Také dle (9) nastavíme hodnotu  $a_0 = 1$ . Dále na začátku vytvoříme nulový vektor  $k$  o délce hodnoty řádu  $p$ , do kterého se budou ukládat koeficienty odrazu. Z (8) vyplývá, že maximální hodnota řádu  $p$  musí být menší než je počet vzorků signálu, a to  $N - 1$ . V každém kroku  $i$ , kdy  $i = 1, \dots, p$ , vytvoříme pomocný vektor pro hodnoty dopředné a zpětné chyby kroku předchozího, kde zpětná chyba podle (8) má hodnoty posunuté o jeden vzorek zpět. Dále provedeme výpočet koeficientu odrazu  $k_i$  pomocí Burgovy rovnice (11). Následuje výpočet nových vektorů hodnot predikčních chyb  $f_p$  a  $b_p$  s aktuálním koeficientem  $k_i$ . Na závěr musíme převést hodnoty vektoru koeficientů odrazu na hodnoty lineárních predikčních koeficientů, podle (9). Koeficient převádíme v každém kroku  $i$ , čímž je krok uzavřen.

### 3 SINUSOIDÁLNÍ MODELOVÁNÍ

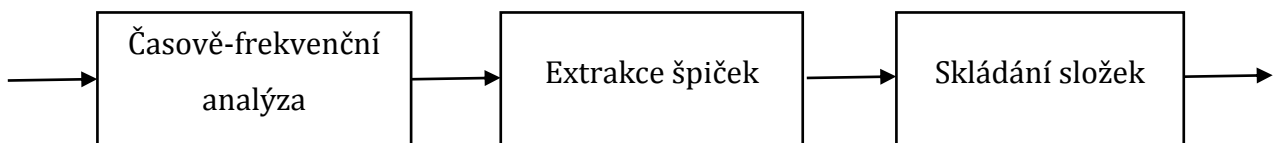
Sinusoidální model signálu interpretuje zvukový signál jako složení několika harmonických a neharmonických složek, které jsou popsány parametry frekvence, amplitudy a fáze v okamžitém čase. Výsledný signál  $s$  získáme podle [7] pomocí těchto parametrů

$$s(t) = \sum_{p=1}^P A_p(t) \cdot \cos[\Phi_p(t)], \quad (12)$$

$$\Phi_p(t) = \Phi_p(0) + 2\pi \int_0^t f_p(u) du, \quad (13)$$

kde  $P$  je počet složek a  $f_p$ ,  $A_p$ , a  $\Phi_p$  jsou parametry frekvence, amplitudy a fáze  $p$ -té složky v čase  $t$ . Jedná se tedy o aditivní syntézu.

Jednou z metod, jak rozložit zvukový signál na tyto časově proměnné parametry, je McAulay-Quatieriho (MQ) algoritmus [9] založený na Fourierově transformaci. Analýza signálu se skládá ze tří kroků, jak vidíme na obr. 2 [7]. Prvním krokem je časově-frekvenční analýza, která je provedena pomocí krátkodobé Fourierovy transformace (STFT), z níž získáme rámce parametrů popisující určitý počet vzorků signálu podle nastavené délky okna. Z této spektrální analýzy vybereme lokální maxima, tzv. špičky, k získání okamžitých hodnot parametrů kmitočtových složek. Jednotlivé celé složky poté získáme skládáním těchto hodnot v čase za sebou, rámeček po rámečku.



Obr. 2: Analýza složek signálu

Z různých důvodů mohou některé špičky chybět nebo obsahovat špatné hodnoty. Blok extrakce špiček může chybné špičky odstranit, ale z informace spektra je není schopen

obnovit. Interpolaci chybějících špiček by měl řešit blok skládání složek, abychom dostali složky rozumných délek.

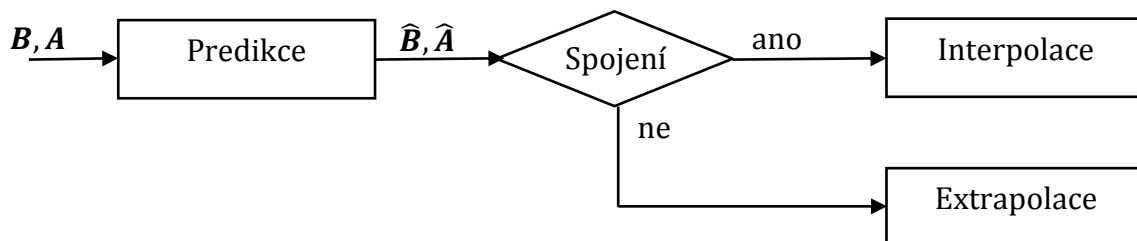
Na základě této analýzy můžeme signál zpětně získat pomocí jednoduchých sinusových oscilátorů [9], na jejichž vstupech jsou parametry jednotlivých kmitočtových složek.

Tohoto modelu využijeme v naší aplikaci, kdy k rekonstrukci chybějícího úseku v signálu použijeme odhad jednotlivých parametrů složek, místo výpočtu odhadu na základě jednotlivých vzorků signálu.

### 3.1 Predikce kmitočtových složek

Mějme tedy mezeru začínající na rámci  $n_1$  a končící na rámci  $n_2$ . Tato mezeru narušuje množinu kmitočtových složek  $S$ . Účelem našeho algoritmu je interpolovat hodnoty složek  $S$  v mezeře. K tomu nám slouží množina  $B$ , která představuje existující složky začínající před mezerou a končící na rámci  $n_1 - 1$ . Dále množina  $A$ , která leží napravo od mezery a začíná na rámci  $n_2 + 1$ .

Na blokovém diagramu vidíme postup algoritmu k interpolaci chybějícího úseku [5].



Obr. 3: Algoritmus pro výpočet odhadovaných hodnot

Odhadované parametry v chybějící oblasti jsou vypočítány z obou množin  $B$  a  $A$ , z nichž dostaneme odhadované množiny  $\hat{B}$  a  $\hat{A}$ . Některé kmitočtové složky množiny  $B$  mohou navazovat na složky v množině  $A$ . Odhadované hodnoty v mezeře tedy tvoří interpolace odhadovaných hodnot z těchto dvou navazujících množin. Nespojené kmitočtové složky, tzn. složky, které mají počátek či konec v mezeře, jsou do chybějícího úseku

extrapolovány z jejich odhadnutých hodnot v množinách  $\widehat{\mathbf{B}}$  nebo  $\widehat{\mathbf{A}}$ . Výsledkem je tedy množina odhadovaných kmitočtových složek  $\widehat{\mathbf{S}}$  v mezeře.

Jednotlivé složky množin  $\mathbf{B}$  a  $\mathbf{A}$  si označíme jako

$$\mathbf{P}_i = P_i(n), \text{ pro } n = n_1 - l_i + 1, \dots, n_1, \quad (14)$$

$$\mathbf{P}_j = P_j(n), \text{ pro } n = n_2, \dots, n_2 + l_j - 1, \quad (15)$$

$$P_k(n) = (f_k(n), A_k(n), \Phi_k(n)) \text{ pro všechna } k, \quad (16)$$

kde  $l_i$  a  $l_j$  jsou délky  $\mathbf{P}_i$  a  $\mathbf{P}_j$  a  $P_k(n)$  je trojice okamžitých parametrů složky  $\mathbf{P}_k$  rámce  $n$ .

Dále  $\widehat{\mathbf{P}}_i$  a  $\widehat{\mathbf{P}}_j$  představují odhadované frekvence a amplitudy složek ze soustav  $\mathbf{B}$  a  $\mathbf{A}$  v místě chybějícího úseku signálu:

$$\widehat{\mathbf{P}}_i = \widehat{P}_i(n_1 + k), \text{ pro } k = 1, \dots, n_2 - n_1 - 1, \quad (17)$$

$$\widehat{\mathbf{P}}_j = \widehat{P}_j(n_2 - k'), \text{ pro } k' = 1, \dots, n_2 - n_1 - 1, \quad (18)$$

$$\widehat{P}_k(n) = (\widehat{f}_k(n), \widehat{A}_k(n)) \text{ pro všechna } k, \quad (19)$$

kde  $\widehat{P}_k(n)$  je dvojice okamžitých odhadovaných parametrů. Fáze se zde neodhaduje z nepoškozených hodnot, ale odvozuje se z odhadnuté frekvence. K výpočtu odhadovaných hodnot amplitudy a frekvence se použije lineární predikce za pomoci predikčních koeficientů získaných Burgovou metodou.

### 3.2 Spojování složek

Prvním krokem k interpolaci poškozených složek v chybějícím úseku je rozhodnutí, které kmitočtové složky z množin  $\mathbf{B}$  a  $\mathbf{A}$  mají být spojeny, aby vytvořily jednotnou složku [5]. Hodnoty složek se můžou různě měnit a při delším časovém intervalu mezery můžou mít na obou stranách naprosto odlišné hodnoty, což bývá při správném spojení

problém. Toto rozhodnutí tedy provádíme pomocí odhadovaných hodnot  $\widehat{\mathbf{P}}_i$  a  $\widehat{\mathbf{P}}_j$ . Můžeme zde použít euklidovskou vzdálenost mezi vektory obsahující hodnoty odhadnutých frekvencí či amplitud. Ta však bohužel samotná nestačí, protože pokud se vektory odhadnutých hodnot od sebe příliš liší, musí být práh jejich vzdáleností více tolerantní, na rozdíl od téměř konstantních průběhů kmitočtových složek. A právě pokud by byl nastaven tolerantnější práh, mohlo by se stát, že se spojí dvě nesouvisející složky.

Proto je euklidovská vzdálenost mezi dvěma odhadnutými vektory navíc normalizována součtem směrodatné odchylky obou odhadnutých vektorů.

Mějme tedy  $d_f(P_i, P_j)$ , což značí normalizovanou euklidovskou vzdálenost mezi odhadnutými frekvencemi  $\widehat{f}_i$  a  $\widehat{f}_j$ ,

$$d_f(P_i, P_j) = \sqrt{\frac{\sum_{n=n_1+1}^{n_2-1} (\widehat{f}_i(n) - \widehat{f}_j(n))^2}{n_2 - n_1 - 1}}. \quad (20)$$

Podobně vypočítáme normalizovanou euklidovskou vzdálenost  $d_A(P_i, P_j)$  mezi odhadnutými amplitudami. Složky můžeme poté úspěšně spojit, pokud splňují současně obě kritéria

$$\frac{d_f(P_i, P_j)}{1 + \sigma(\widehat{f}_i) + \sigma(\widehat{f}_j)} < T_f, \quad (21)$$

$$\frac{d_A(P_i, P_j)}{1 + \sigma(\widehat{A}_i) + \sigma(\widehat{A}_j)} < T_A, \quad (22)$$

kde  $\sigma(x)$  je směrodatná odchylka vektoru  $\mathbf{x}$  a  $T_f$  a  $T_A$  je práh vzdáleností frekvence a amplitudy.

Poté vznikne jednotná kmitočtová složka  $\mathbf{P}_m$ .  $\mathbf{P}_i$  a  $\mathbf{P}_j$  jsou pro další zpracování z výběru odstraněny. Tento krok se opakuje pro všechny kombinace protějšších složek, dokud nezůstane ani jedna vyhovující. Nespojené složky jsou poté do mezery extrapolovány, viz kapitola 3.4. Výsledné odhadované hodnoty v mezeře  $\widehat{\mathbf{P}}_m$  získáme interpolací popsanou v následující kapitole.

### 3.3 Interpolace spojených kmitočtových složek

Interpolované parametry frekvence a amplitudy složky  $\widehat{P}_m$ , začínající na rámci  $n_1 + 1$  a končící rámcem  $n_2 - 1$ , získáme spojením odhadovaných parametrů frekvence a amplitudy ze složek  $\widehat{P}_i$  a  $\widehat{P}_j$ . Jak již bylo zmíněno, fáze je zajištěna vypočítáním podle metody popsané v kapitole 3.3.3.

#### 3.3.1 Interpolace frekvence

K výpočtu výsledné odhadované hodnoty  $\widehat{f}_m(n)$  z hodnot odhadovaných frekvencí  $\widehat{f}_i(n)$  a  $\widehat{f}_j(n)$  se dostaneme sečtením těchto hodnot poté, co na jednotlivé vektory aplikujeme váhovací okno [5]. Vektor  $\widehat{f}_i$  je vynásoben funkcí  $w$ , zatímco vektor  $\widehat{f}_j$  je vynásoben funkcí  $1 - w$ ,

$$\widehat{f}_m(n) = w \left( \frac{n-n_1}{n_2-n_1} \right) \widehat{f}_i(n) + \left[ 1 - w \left( \frac{n-n_1}{n_2-n_1} \right) \right] \widehat{f}_j(n). \quad (23)$$

Vstupním parametrem funkce váhovacího okna je poměr indexu okamžitého rámce v mezeře a délky mezery. V našem případě je jako samotná váhovací funkce použito symetrické kosinové okno, které má v polovině úseku hodnotu 0.5,

$$c(t) = \frac{1 + \cos[\pi(1+t)]}{2}. \quad (24)$$

#### 3.3.2 Interpolace amplitudy

Výslednou amplitudu dostaneme podobným způsobem jako u frekvence. Na rozdíl od frekvence je však amplituda více modulována, odhadované hodnoty získané lineární predikcí tedy nejsou dostačující.

Předtím, než tedy odhadované amplitudy z protějších stran váhově sečteme, zajistíme, aby odhadovaná amplituda vypočítaná ze složky  $P_i$  končila na hodnotě, která je rovna střední hodnotě amplitudy protější složky  $P_j$  vypočítané z rozsahu  $n_2$  až  $\min(n_2 + M, n_2 + l_j - 1)$ . Parametr  $M$  by měl být zvolen tak, abychom získali správný



odhad energie začátku složky  $P_j$ . Omezuje tedy délku intervalu, z kterého určujeme střední hodnotu amplitudy tak, aby nezasahovala do úseku, kde může mít modulace jiný charakter, který by už neměl ovlivňovat výpočet odhadovaných hodnot v mezeře [5]. Tato podmínka je splněna tím, že k jednotlivým hodnotám vektoru odhadovaných amplitud  $\widehat{A}_i$  přičteme

$$\delta_i(n) = \frac{n-n_1}{n_2-n_1} \left[ \frac{\sum_{\tau=0}^{\min(M,l_j-1)} A_j(n_2+\tau)}{\min(M,l_j-1)+1} - \widehat{A}_i(n_2) \right]. \quad (25)$$

Podobný výpočet aplikujeme na vektor  $\widehat{A}_j$ , přičtením

$$\delta_j(n) = \frac{n_2-n}{n_2-n_1} \left[ \frac{\sum_{\tau=0}^{\min(M,l_i-1)} A_i(n_1-\tau)}{\min(M,l_i-1)+1} - \widehat{A}_j(n_1) \right]. \quad (26)$$

Výsledného odhadu amplitudy pak dosáhneme váhovým součtem odhadovaných amplitud z protějších složek aplikací stejného váhovacího okna jako u interpolace frekvence

$$\widehat{A}_m(n) = w \left( \frac{n-n_1}{n_2-n_1} \right) [\widehat{A}_i(n) + \delta_i(n)] + \left[ 1 - w \left( \frac{n-n_1}{n_2-n_1} \right) \right] [\widehat{A}_j(n) + \delta_j(n)]. \quad (27)$$

### 3.3.3 Interpolace fáze

Interpolaci fáze neprovádíme z odhadovaných hodnot získaných lineární predikcí jako u frekvence nebo amplitudy. Získáme ji ze znalosti již odhadnutých interpolovaných frekvencí a skutečných hodnot frekvence a fáze na hranicích chybějícího úseku  $f_m(n_1), \Phi_m(n_1)$  a  $f_m(n_2), \Phi_m(n_2)$ .

První hodnotu fáze [5] vypočítáme podle

$$\tilde{\varphi}(n_1 + 1) = \Phi(n_1) + \pi T [f(n_1) + \hat{f}(n_1 + 1)]. \quad (28)$$

Další hodnoty jsou následovně vypočítané jako

$$\tilde{\varphi}(n) = \tilde{\varphi}(n_1 + 1) + \pi T \sum_{\tau=n_1+1}^n [\hat{f}(\tau - 1) + \hat{f}(\tau)], \quad (29)$$

kde  $n$  je v intervalu  $n \in [n_1 + 2, n_2]$  a  $T$  je vzdálenost rámců v sekundách. Za  $\varphi(n)$  označujeme fázi, z které skutečnou dostaneme zbytkem po dělení  $\Phi(n) = \varphi(n) \bmod 2\pi$ . Index  $m$  je z rovnic vynechán pro zobecnění pro obě strany složek.

Výpočtem všech hodnot fáze podle (11) však může nastat nenávaznost fází na konci chybějícího úseku, kterou zjistíme nerovností odhadované hodnoty, která přesahuje chybějící úsek a první skutečné hodnoty fáze na hranici úseku  $\tilde{\Phi}(n_2) \neq \Phi(n_2)$ . Zavádíme proto chybu  $e_\phi$  extrapolované fáze rámce na indexu  $n_2$ ,

$$e_\phi = \tilde{\Phi}(n_2) - \Phi(n_2). \quad (30)$$

Ke každé hodnotě fáze v úseku mezery poté váhově přičteme konstantu  $\rho$ , která je vybrána tak, aby zajistila návaznost fáze na hranici konce chybějícího úseku, tzn. aby platila podmínka  $\tilde{\Phi}(n_2) - \Phi(n_2) = 0$ . Konstanta  $\rho$  však může mít nekonečně mnoho řešení v násobcích  $2\pi$ , nejnižší hodnotu tedy získáme z podmínky

$$\rho = \begin{cases} e_\phi + 2\pi & \text{pro } e_\phi < -\pi \\ e_\phi - 2\pi & \text{pro } e_\phi > \pi \\ e_\phi & \text{ostatní.} \end{cases} \quad (31)$$

Konstantu pak přičteme následovně:

$$\hat{\varphi}(n) = \tilde{\varphi}(n) + \frac{n - n_1}{n_2 - n_1} \rho, \quad (32)$$

kde  $n$  je v intervalu  $n \in [n_1 + 1, n_2]$ .

### 3.4 Extrapolace nespojených kmitočtových složek

Pokud předpokládáme, že všechny protější složky jsou správně interpolovány, zbývají nám nespojené složky z množiny  $\mathbf{B}$ , které tedy pravděpodobně končí v úseku mezery a nespojené složky z množiny  $\mathbf{A}$ , které naopak v úseku začínají. Určíme si parametry jejich maximální délky, a to  $l_B$ , což je maximální délka extrapolace nespojené složky z množiny  $\mathbf{B}$  a  $l_A$  je maximální délka extrapolace nespojené složky množiny  $\mathbf{A}$ . K extrapolaci nespojených složek  $\mathbf{P}_i$  nebo  $\mathbf{P}_j$  použijeme hodnoty odhadovaných parametrů v množině  $\widehat{\mathbf{P}}_i$  nebo  $\widehat{\mathbf{P}}_j$ . Odhadované hodnoty frekvence se nemění a k výpočtu fáze použijeme rovnice (28) a (29).

Je velice důležité zvolit vhodné parametry  $l_B$  a  $l_A$  podle typu rekonstruovaného signálu. Tyto parametry ovlivňují výpočet amplitudy, jejichž chování můžeme v určitých signálech přibližně předpovídat. Při extrapolaci kmitočtové složky z množiny  $\mathbf{B}$  můžeme z odhadovaných hodnot amplitudy určit, v kterém rámci by měla složka končit. Extrapolovaná amplituda  $\tilde{A}_i(n)$  [5] je tedy následně tlumená odhadovaná amplituda  $\hat{A}_i(n)$

$$\tilde{A}_i(n) = \hat{A}_i(n) - \gamma_i(n), \quad (33)$$

kde

$$\gamma_i(n) = \frac{n - n_1}{l_B} \max[\hat{A}_i(n_1 + l_B), 0]. \quad (34)$$

Pokud má extrapolovaná amplituda  $\tilde{A}_i(n)$  zápornou hodnotu ještě před koncem extrapolované složky nastavené parametrem  $l_B$ , tj. na rámci  $n < n_1 + l_B$ , extrapolovaná složka  $\tilde{\mathbf{P}}_i$  končí na rámci  $n - 1$ .

Při extrapolaci nespojené amplitudy napravo od mezery, tj. z množiny kmitočtových složek  $\mathbf{A}$ , už nelze jednoduše odhadnout nástup dané složky. Všechny nespojené složky množiny  $\mathbf{A}$  by tedy měly alespoň začínat na stejném rámci  $n_2 - l_A$ . Extrapolovaná amplituda  $\tilde{A}_j(n)$  je tedy následovně tlumená odhadovaná amplituda  $\hat{A}_j(n)$

$$\tilde{A}_j(n) = \hat{A}_j(n) - \gamma_j(n), \quad (35)$$

kde

$$\gamma_j(n) = \frac{n_2 - n}{l_A} \hat{A}_j(n_2 - l_B). \quad (36)$$

Extrapolovaná složka  $\tilde{\mathbf{P}}_j$  začíná, podobně jako v případě  $\tilde{\mathbf{P}}_i$ , na indexu rámce bližšího konci mezery  $n \geq n_2 - l_A$  tak, aby byla splněna podmínka  $\tilde{A}_j(n+k) > 0$  pro všechna  $k \geq 0$ .

### 3.5 Interpolace residua

Poté co je signál popsán sinusovým modelem a na jeho základě je provedena interpolace chybějícího úseku, zbývá ještě interpolovat residuum signálu [6], které sinusovým modelem reprezentováno není. Tím může být šum v pozadí a menší harmonické složky, které nejsou v sinusovém modelu signálu popsány. Pro některé, hlavně hudební signály nese tato složka významnou část energie signálu. Vynecháním interpolace residua by došlo k významnému poklesu hlasitosti a změně tónu v oblasti interpolovaného úseku.

Prvním krokem je získání residua z oblasti známého signálu. Provedeme tedy syntézu signálu podle parametrů modelu z množiny  $\mathbf{B}$  a syntetizovanou část signálu odečteme od skutečného signálu na levé straně od mezery. Stejně tak odečteme syntetizovanou část signálu podle parametrů množiny  $\mathbf{A}$  od skutečného signálu na pravé straně od mezery.

Druhým krokem je interpolace residua. To provedeme, stejně jako u interpolace jednotlivých složek sinusového modelu, pomocí lineární predikce s koeficienty získanými Burgovou metodou. Nevýhodou je však u interpolace delších úseků tendence poklesu energie směrem ke středu mezery a porušení fázové návaznosti skutečného signálu na interpolovaný.

## 4 Metody objektivního hodnocení kvality rekonstrukce

K zobrazení výsledků rekonstrukce nám kromě samotných grafů průběhu a spektra původního a rekonstruovaného signálu pro vzájemné porovnání poslouží také numerické hodnoty získané následujícími metodami. Tyto hodnoty nám poslouží k lepšímu přehledu výsledků rekonstrukce pro různé typy průběhů signálu, které budeme navzájem porovnávat.

### 4.1 SNR

Jednou z objektivních metod testování rekonstrukce signálu je metoda pro výpočet odstupů signálu od šumu, což je poměr výkonu signálu k výkonu šumu. V našem případě je za šum považován rozdíl amplitudy původního a rekonstruovaného signálu [10]

$$SNR(n) = 10 \log_{10} \frac{\sum_{j=1}^l x^2(n_1 + j)}{\sum_{j=1}^l (x(n_1 + j) - \hat{x}(n_1 + j))^2} \quad (1)$$

kde  $l$  je délka rekonstruovaného úseku, tj.  $l = n_2 - n_1 - 1$ .

### 4.2 PEMO-Q

Druhou z objektivních metod testování rekonstrukce signálu je metoda PEMO-Q [8]. Jedná se o novější metodu hodnocení zvukové kvality představenou v r. 2006. Tato metoda je rozšířením měření objektivní kvality řeči  $q_c$ . Slouží k předpovězení vnímání kvality rozdílu mezi referenčním a zkresleným signálem. K vyhodnocení kvality slouží hodnota  $PSM$ , která představuje koeficient křížové korelace vnitřních reprezentací signálů metodou PEMO. Je však ovlivněna typem vstupního signálu. Metoda tedy představuje také hodnotu  $PSM_t$ , která typem vstupního signálu ovlivněna není. Tato hodnota je v testování použita k hodnocení objektivní kvality rekonstrukce. Hodnota je omezena v intervalu  $PSM_t \in [-1,1]$ , kde 1 značí identické signály a menší hodnoty odpovídají větším odchylkám testovacího signálu od referenčního.

## 5 IMPLEMENTACE INTERPOLAČNÍ METODY

Použitá interpolační metoda je zrealizována ve výpočetním prostředí MATLAB na základě teorie v předchozích kapitolách.

Metoda je vytvořena jako univerzální funkce v souboru `SinModInt.m`. Vstupní a výstupní proměnná je shodná proměnná `inpaint` typu struktura. Tento krok byl zvolen kvůli univerzálnímu využití funkce, kdy není omezena počtem vstupních parametrů a zároveň spouštěcí funkce či skript může volat více podobných algoritmů, které sdílí např. podobné parametry a vstupní signál. Funkce je tedy použitelná pro Brno – Wien Inpainting Toolbox [16], v kterém jsou provedeny některé experimenty. K sinusoidálnímu modelování a zpětné syntéze signálu jsou použity funkce z toolboxu Sinewave and Sinusoid+Noise Analysis/Synthesis in Matlab [9]. K zobrazení spektra signálu je použitý The Large Time/Frequency Analysis Toolbox [14].

### 5.1 Popis vstupní struktury

První z povinných vstupních parametrů je `inpaint.miss`, který obsahuje vektor dvou hodnot definující vlastnosti chybějícího úseku. Na prvním indexu vektoru je hodnota vzorku, na kterém začíná mezera, na druhém indexu je počet vzorků určující délku mezery. Dalším povinným vstupem je struktura `inpaint.signal`, která obsahuje informace o vstupním signálu. Obsahuje proměnné `inpaint.signal.samples` obsahující vektor hodnot vzorků vstupního nepoškozeného zvukového signálu a `inpaint.signal.fs` obsahující hodnotu vzorkovacího kmitočtu signálu.

Další vstupní struktura `inpaint.solver.param` je nepovinná a obsahuje hodnoty parametrů ovlivňující výpočet rekonstruovaného signálu. Pokud nejsou na vstupu zadány žádné hodnoty parametrů, funkce jim přiřadí vlastní standardní hodnoty. Přehled jednotlivých parametrů je uveden v tab. 1.

Proměnná `KNB` ve struktuře `inpaint.solver.param` obsahuje hodnotu řádu AR modelu residuálního signálu nalevo od mezery, tzn. počet predikčních koeficientů získaných Burgovou metodou. Stejně tak obsahuje proměnná `KNA` hodnotu řádu AR modelu residuálního signálu napravo od mezery. Standardní hodnoty jsou u obou proměnných nastaveny funkcí na polovinu počtu vzorků odpovídajících stran signálu.

Při volbě jiných hodnot musíme brát na vědomí délku odpovídajících modelovaných úseků a zvolit vždy řád s maximální hodnotou o jednu menší než je počet vzorků úseku. Původním záměrem byla také možnost volby řádu AR modelu kmitočtových složek určených k interpolaci, avšak z praktického hlediska byla tato volba vypuštěna, viz kapitola 5.3.

**Tab. 1: Struktura vstupního parametru**

název proměnné	standardní hodnota	popis proměnné
<code>inpaint.signal.samples</code>	-	vektor vzorků vstupního signálu
<code>inpaint.signal.fs</code>	-	hodnota vzorkovací frekvence
<code>inpaint.miss</code>	-	vektor definující chybějící úsek
<code>inpaint.solver.param.KNB</code>	délka residua/2	řád AR modelu residua signálu nalevo od mezery
<code>inpaint.solver.param.KNA</code>	délka residua/2	řád AR modelu residua signálu napravo od mezery
<code>inpaint.solver.param.thrF</code>	2	práh rozdílu vektorů predikčních hodnot frekvence
<code>inpaint.solver.param.thrA</code>	0.5	práh rozdílu vektorů predikčních hodnot amplitudy
<code>inpaint.solver.param.parM</code>	30	délka vektoru pro výpočet střední hodnoty amplitudy
<code>inpaint.solver.param.lB</code>	délka mezery - 1	délka extrapolované složky z levé strany
<code>inpaint.solver.param.lA</code>	$(\text{délka mezery} - 1)/2$	délka extrapolované složky z pravé strany

Dalším nastavitelným parametrem je `thrF`, který představuje práh vzdálenosti vektorů s hodnotami frekvencí protějšších kmitočtových složek, jak je uvedeno v rovnici (21). Čím větší práh nastavíme, tím větší je pravděpodobnost že dojde k interpolaci složek, které na sebe původně nenavazují. Pro spojení složek s větší modulací je naopak vhodné zvolit větší práh. Volba tohoto parametru je tedy velmi citlivá a může značnou mírou ovlivnit výsledky rekonstrukce. Bude tedy jedním z hlavních předmětů testování.

Podobným parametrem je `thrA`, viz rovnice (22). Rozdíl amplitud protějšších složek však nebývá tak různorodý jako u frekvence, volí se proto menší hodnota. Dalším

důvodem této volby je také splnění obou podmínek, jak vzdálenost amplitudy, tak frekvence, pro spojení protějšších složek. S větším prahem vzdálenosti amplitud by se totiž šance na interpolaci původně nenavazujících složek ještě zvýšila.

Dalším parametrem je `parM`, který při interpolaci amplitudy ovlivňuje výpočet střední hodnoty amplitudy (25) a (26) určením maximálního počtu rámců s hodnotami amplitud nepoškozeného úseku u delších kmitočtových složek, z kterých jsou střední hodnoty vypočítány.

Proměnné `lB` a `lA` ve struktuře `inpaint.solver.param` obsahují maximální délky extrapolovaných nespojených složek z levé a pravé strany od mezery. Standardní nastavení těchto parametrů vidíme v tab. 1. Délka mezery představuje počet rámců s hodnotami parametrů odhadovaných složek.

## 5.2 Popis výstupní struktury

Výstupem funkce `SinModInt` jsou tři vektory ve výstupní struktuře `inpaint.sol`. První vektor obsahuje hodnoty vzorků poškozeného signálu s mezerou. Tento vektor se nachází v proměnné `err`. Další vektor obsahuje hodnoty vzorků signálu s rekonstruovaným úsekem, ve kterém je však pouze syntetizovaná část signálu podle interpolovaných parametrů sinusoidálního modelu. Vektor se nachází v proměnné `dr1`. Posledním výstupem je vektor hodnot vzorků rekonstruovaného signálu včetně interpolovaného residua. Tento vektor se nachází v proměnné `samples`.

Tab. 2: Struktura výstupního parametru

název proměnné	popis proměnné
<code>inpaint.sol.err</code>	Vektor signálu s mezerou
<code>inpaint.sol.dr1</code>	Vektor signálu se syntetizovaným rekonstruovaným úsekem
<code>inpaint.sol.samples</code>	Vektor signálu s rekonstruovaným úsekem včetně residua



### 5.3 Popis funkce SinModInt

Účelem funkce je tedy vytvořit na vstupním signálu mezeru podle parametrů zadaných na vstupu a poté se pokusit mezeru rekonstruovat za použití zadaných parametrů ovlivňující rekonstrukci.

Prvním krokem je tedy vytvoření mezery vložení nul do úseku vektoru určeným vstupními parametry.

```
d(1:l+Me-1)=0; %vytvoření mezery,
```

kde index  $l$  značí počátek mezery a index  $Me$  značí délku mezery. Dále získáme funkci `sinmodel` popis signálu pomocí parametrů sinusoidálního modelu za účelem zjištění počtu rámců celého signálu pro výpočet počtu rámců v mezeře. Funkce se volá pomocí

```
[F, M, P]=sinmodel(d,sr); %sinusoidální model,
```

kde vstupními parametry jsou vektor vzorků signálu  $d$  a hodnota vzorkovacího kmitočtu  $s_r$ . Na výstupu dostaneme matice jednotlivých parametrů, a to frekvence  $F$ , amplitudy  $M$  a fáze  $P$ . Každý řádek představuje jednu plynulou kmitočtovou složku a každý rámeček sloupce popisuje časový úsek v signálu podle nastavené velikosti okna u STFT (128 vzorků). Rámce, které představují časový úsek, v němž kmitočtová složka nemá hodnoty, tzn. složka zde již nepokračuje nebo ještě nezačala, obsahuje hodnotu NaN (Not-a-Number). Tato funkce obsahuje další funkce MQ algoritmu [9].

Dalším krokem je popis obou stran signálu od mezery pomocí sinusoidálního modelu. Tímto získáme parametry, které jsou později určeny k výpočtu odhadovaných hodnot parametrů z obou stran od mezery. Rozdíly fáze v matici  $P_B$ ,  $P_A$ , které obsahují hodnoty fáze složek jednotlivých stran, a v matici nově odvozených hodnot nesmí být v absolutní hodnotě větší než  $\pi$ , hodnoty v maticích proto přepočítáme algoritmem rozbalení fáze (angl. unwrapping) [5].

Syntetizovaný signál pomocí získaných parametrů je následovně od skutečného signálu dané strany odečten, pro získání residuální části signálu. Do proměnné  $F_{Plen}$  je poté uložen počet rámců v mezeře.

Následuje lineární predikce parametrů frekvence a amplitudy všech kmitočtových složek, jejichž hodnoty jsou na hranici s mezerou, což značí pravděpodobné pokračování složek v mezeře. To platí pro obě strany signálu. Řádky matic, jejichž hodnoty na hranici mezery jsou NaN, zpracované nejsou. Predikční koeficienty jsou získány funkcí `BurgMeth` která ze signálu získá koeficienty Burgovou metodou. Implementace je popsána v kapitole 2.2.1. Volá se pomocí

```
a = BurgMeth(X, K);,
```

kde  $X$  je vektor hodnot signálu a  $K$  je řád AR modelu signálu. Na výstupu  $a$  dostaneme vektor hodnot predikčních koeficientů. Jak již bylo zmíněno v kapitole 5.1, řád AR modelu  $K$  je standardně nastavován uvnitř funkce `SinModInt` a to na hodnotu poloviny počtu rámců kmitočtové složky, maximální hodnota je však  $K = 40$ . Je to z toho důvodu, že nemůžeme odhadnout nejmenší počet rámců kmitočtové složky v nepoškozeném úseku signálu, nastavený parametr může být v některých případech maximálně  $K = 1$ . Ruční nastavení tohoto parametru tedy nemá smysl. Nové odhadované hodnoty jsou získány funkcí `interleft`, která představuje výpočet podle (4). Funkci voláme jako

```
I = interleft(X, a, K, FPlen);,
```

kde na výstupu dostaneme vektor odhadovaných hodnot  $I$ . Stejnou predikcí také získáme odhadované vzorky residua signálu z obou stran od mezery.

Zvláštním případem je situace, kdy je na hranici skutečného signálu s mezerou pouze jeden rámeček kmitočtové složky. V tomto případě nelze predikční metodu použít, hodnoty proto pouze do rámců v mezeře nakopírujeme. V případě, že na protější straně se najde navazující složka, je případná modulace zajištěna pomocí predikčních hodnot zjištěných z této protější složky. V případě extrapolace, tzn. pokud se navazující protější složka nenajde, bude tato kmitočtová složka v mezeře zanikat. Tomu také bude odpovídat hodnota amplitudy a složka tedy pravděpodobně zanikne ještě před nastavenou délkou extrapolovaných složek parametrem  $l_B$  nebo  $l_A$ . U delších mezer, v případě kdy složka bude zanikat těsně před koncem mezery a původní složka v tomto úseku je modulována, nelze bohužel původní modulaci odhadnout. V tomto případě se tedy počítá s velkou nepřesností rekonstrukce.

Další část zdrojového kódu funkce se zabývá interpolací spojených protějšších složek. Funkce *okno* provede výpočet vektoru s hodnotami kosinového váhovacího okna podle (24) závislého na délce mezery, tedy vstupní hodnotě proměnné  $FPlen$ . Nyní dochází ke kroku spojování odhadovaných vektorů protějšších kmitočtových složek kvůli rozhodnutí pro jejich interpolaci, jak je zobrazeno v diagramu na obr. 3. Řádky matic odhadovaných hodnot kmitočtových složek jsou zpracovány pouze v případě, že řádek není nulový nebo nebyl již dříve zpracován. Index již zpracovaných řádků se zapisuje do vektorů  $IndexI$  pro řádky matic levé strany a  $IndexJ$  pro řádky matic pravé strany.

Vzdálenost vektorů hodnot frekvence a amplitudy protějšších složek vypočítá funkce

```
matchvect (BI, AI) ; ,
```

podle (20), (21), (22). Vstupní proměnné  $BI$  a  $AI$  jsou vektory odhadovaných parametrů jednotlivých kmitočtových složek z levé a pravé strany od mezery. Po splnění podmínky, kdy obě vzdálenosti nepřesahují hodnotu parametrů  $thrF$  a  $thrA$  dojde k interpolaci složek. Interpolace frekvence se provádí podle (23). Výpočet (25) a (26) provedou funkce

```
vypDelI (MBI, MA, parM, FPlen) ;
vypDelJ (MAI, MB, parM, FPlen) ;.
```

$MBI$  a  $MAI$  jsou vektory odhadovaných amplitud z levé a pravé strany od mezery.  $MB$  a  $MA$  jsou vektory skutečných amplitud levé a pravé strany od mezery.

Poté podle (27) proběhne interpolace amplitudy kmitočtové složky. Výpočet vektoru hodnot fáze pro interpolovanou složku je podle kapitoly 3.3.3 realizován ve funkci

```
interphase (PB, PA, FB, FBI, FA(1), FPlen, sr) ; ,
```

kde  $PB$  a  $PA$  jsou vektory skutečných hodnot fáze složky na levé a pravé straně od mezery,  $FB$  je vektor skutečných hodnot frekvence složky na levé straně a  $FA(1)$  je

skutečná hodnota frekvence prvního rámece na pravé straně od mezery. Realizací výpočtu odvození fáze se zabývá kapitola 5.3.1.

Interpolované hodnoty frekvence a amplitudy včetně vypočítaných hodnot fáze se zapíše do nových odpovídajících matic `FI`, `MI` a `PI` na nový řádek. Zápisem zpracovaných řádků matic odhadovaných hodnot je ukončen krok interpolace.

Následuje extrapolace nezpracovaných řádků matic představujících odhadované hodnoty parametrů kmitočtových složek obou stran od mezery. Pro extrapolaci složek levé strany upravíme odhadované hodnoty amplitudy funkcí `extraampB`. Funkce se volá jako

```
[M,i] = extraampB(MBI,lB);,
```

kde vstupní hodnoty jsou vektor odhadovaných hodnot amplitudy `MBI` a délka vektoru extrapolované složky `lB`. Na výstupu dostaneme upravený vektor amplitudy `M` a index `i`, který představuje poslední rámeček extrapolované složky. Jak je popsáno v kapitole 3.4, kmitočtová složka může končit před nastavenou délkou parametrem `lB`. Vektory amplitudy a fáze extrapolované složky poté za indexem `i` doplníme hodnotami `NaN`. To provedeme funkcí `vecnanB`. Funkce `extraphase` vypočítá hodnoty fáze extrapolované složky a vektor je opět za indexem `i` doplněn hodnotami `NaN`. Vektory parametrů extrapolované složky se opět zapíše na nový řádek do odpovídajících matic. Krok je opět ukončen zápisem zpracovaných řádků. Pro extrapolaci složek z pravé strany je analogicky stejný postup za použití funkcí `extraampA` a `vecnanA`.

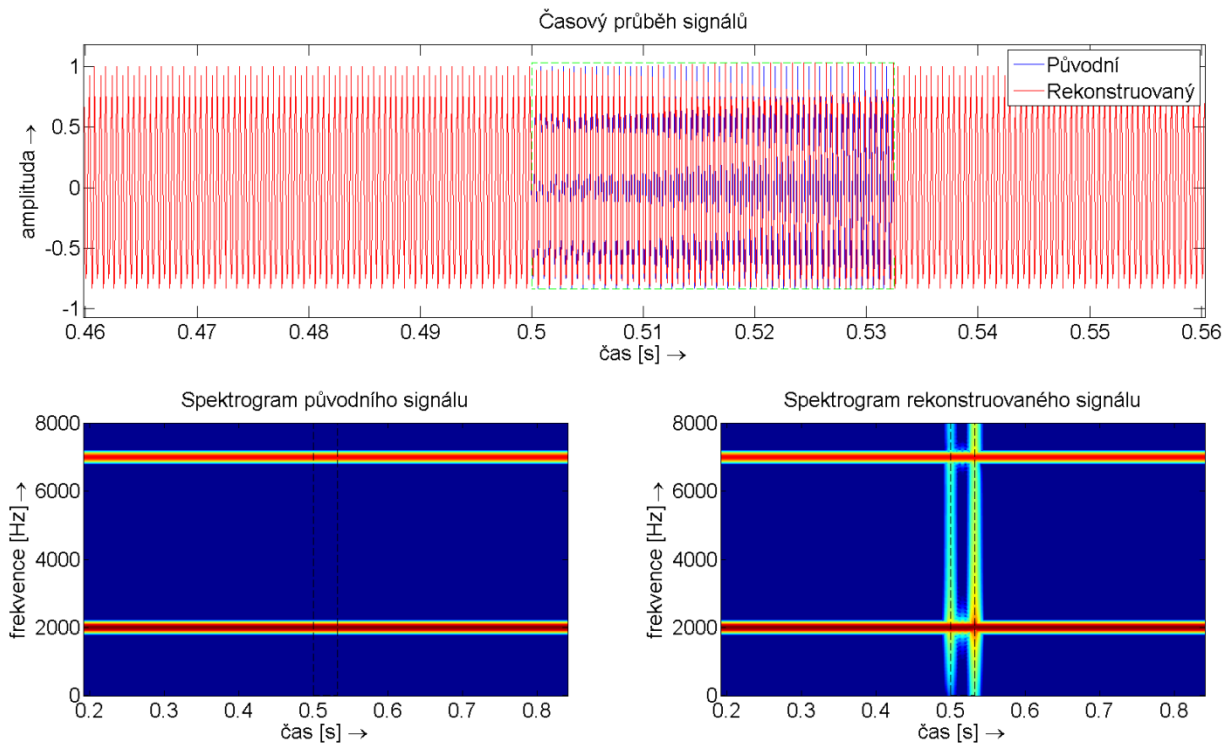
Pro interpolaci residua je pomocí funkce `okno` vypočítán nový vektor hodnot váhovacího okna a to podle délky mezery nastavené vstupním parametrem `inpaint.miss`.

Na závěr je provedena syntéza signálu podle interpolovaných a extrapolovaných parametrů v mezeře a spolu s interpolovaným residuem přičtena k poškozenému signálu v místě mezery. Tento rekonstruovaný signál získáme na výstupu funkce.

### 5.3.1 Realizace výpočtu fáze

Při realizaci algoritmu bylo zjištěno, že problémem u rekonstruovaného signálu byl nežádoucí posun fáze. Funkce `interphase` prováděla výpočty podle teoretického základu v kapitole 3.3.3 včetně zpětného přičítání chyby extrapolované fáze  $e_\phi$ .

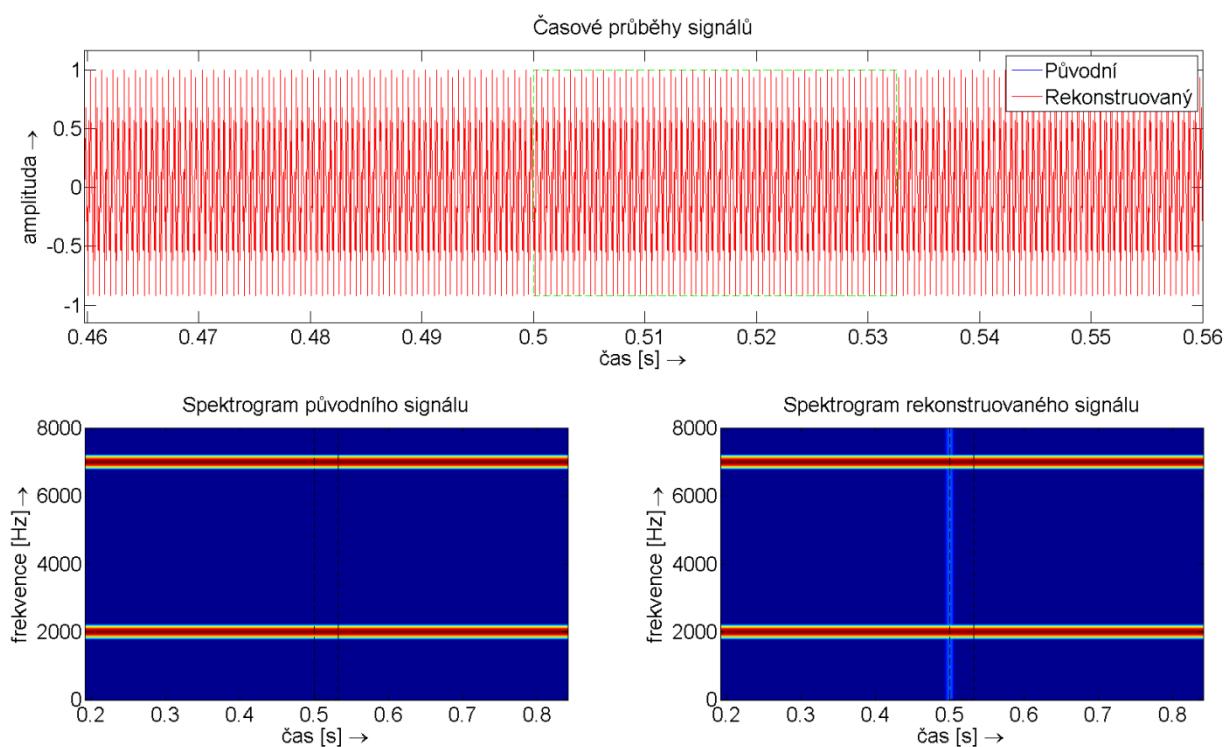
Byla provedena rekonstrukce jednoduchého uměle vygenerovaného signálu s dvěma harmonickými složkami o kmitočtu  $f_1 = 2$  kHz a  $f_2 = 7$  kHz a náhodně generovanou amplitudou i fází. Na časovém průběhu signálů na obr. 4 vidíme nežádoucí posun fáze. Ze spektrogramu lze na okrajích rekonstruované mezery zpozorovat, že na sebe původní a rekonstruované složky díky fázovému posuvu nenasazují.



Obr. 4: Časový průběh a spektrogram původního a rekonstruovaného signálu s posunutou fází

U druhé rekonstrukce bylo z funkce `interphase` zpětné přičítání chyby odstraněno, hodnoty fáze rekonstruovaných kmitočtových složek se tedy získají pouze výpočtem (28) a (29). Z časového průběhu signálů na obr. 5 vidíme, že se původní i rekonstruovaný signál v místě mezery překrývají. Stejně tak vidíme ve spektrogramu rekonstruovaného signálu navazující průběh původních a rekonstruovaných složek.

V realizaci funkce pro výpočet hodnot fáze rekonstruovaných složek je tedy zpětné přičítání chyby nežádoucí a pro další testování bylo z funkce odstraněno.



Obr. 5: Časový průběh a spektrogram původního a rekonstruovaného signálu bez posunu fáze

## 5.4 Spouštěcí skripty

Pro účely testování byly vytvořeny dva různé testovací skripty.

První je `start.m`, který slouží pro jednorázové testování rekonstrukce jednoho zvukového souboru a zobrazení průběhu původního a rekonstruovaného signálu. Úpravou skriptu lze pomocí funkce `wavread` načíst jako vstupní signál soubor typu `.wav` s libovolným názvem, umístěným v podsložkách `./signals/music_wav/` nebo `./signals/speech_wav/`. Do struktury proměnné `param` se vkládají hodnoty parametrů, viz tab. 1. Ve skriptu se také voláním funkce

`SNRCmpt(xRef, xEst);`

provádí výpočet hodnoty SNR. Vstupními parametry funkce jsou vektory vzorků referenčního signálu `xRef` a rekonstruovaného signálu `xEst`. Skript také uloží za účelem poslechu poškozený a rekonstruovaný signál ve formátu `.wav` funkcí `wavwrite`. Hlavním účelem skriptu pro jednorázové testování je zobrazení spektra původního, poškozeného a rekonstruovaného signálu a časový průběh původního a poškozeného signálu.

Druhý skript `testskript.m` slouží k testování rekonstrukce několika souborů s různým nastavením mezer a parametrů rekonstrukce. Do pole buněk `filelist` se zapisují názvy zvukových souborů, do matice `gaps` se zapisují parametry mezer a do dalších vektorů s názvy parametrů podle potřeby testování se zapisují hodnoty odpovídajících parametrů. Účel první části skriptu je výpočet hodnot SNR s nastavením různých parametrů rekonstrukce. Poškozené a rekonstruované signály jsou uloženy s popisem podle nastavených parametrů. Ve skriptu se volá další skript `starttest.m`, což je upravená verze skriptu `start.m`. V tomto skriptu se signál kvůli urychlení výpočtu zkracuje signál na délku  $t = 2$  s. Druhá část skriptu slouží pro rekonstrukci několika úseků za sebou v jednom zvukovém signálu a uložení tohoto rekonstruovaného signálu za účelem testování metodou PEMO-Q.

## 6 VÝSLEDKY EXPERIMENTŮ

Hromadné testování metody neproběhlo podle zadání na výpočetním serveru Blade, ale na osobním počítači. Testovala se rekonstrukce deseti souborů s hudebním signálem a deseti souborů s řečovým signálem viz tab. 3. Vzorkovací kmitočet všech signálů je  $f_{vz} = 16$  kHz a bitová hloubka je  $n = 16$  b. Testování probíhalo pro různé délky mezer a pro několik hodnot parametru  $thr_F$ , jelikož se předpokládá, že tento parametr nejvíce ovlivňuje výsledek rekonstrukce. Hromadné testování tedy sloužilo i k nalezení ideální hodnoty tohoto parametru. Změna parametru  $thr_A$  neměla na výsledky metod objektivního hodnocení vliv. Ostatní parametry mají standardní hodnotu podle tab. 1.

Některé jednorázové experimenty jsou provedeny na souborech z databáze GTZAN music/speech collection [15].

Tab. 3: Názvy souborů pro hromadné testování

Názvy souborů s hudebním signálem		Názvy souborů s řečovým signálem	
music02_16kHz.wav	music08_16kHz.wav	female01_16kHz.wav	male01_16kHz.wav
music03_16kHz.wav	music09_16kHz.wav	female02_16kHz.wav	male02_16kHz.wav
music04_16kHz.wav	music10_16kHz.wav	female03_16kHz.wav	male03_16kHz.wav
music06_16kHz.wav	music11_16kHz.wav	female04_16kHz.wav	male04_16kHz.wav
music07_16kHz.wav	music12_16kHz.wav	female05_16kHz.wav	male05_16kHz.wav

### 6.1 Výsledky SNR

V části testovacího skriptu pro získání výsledků kvality rekonstrukce metodou SNR bylo pro každý zvukový soubor nastaveno parametrem `gaps` osm různých délek mezer ve třech úsecích s počátkem na indexech vzorků  $Me = [8000 \ 10000 \ 15000]$ .



Tab. 4: Délky chybějících úseků

Počet chybějících vzorků [-]	160	320	640	1280	1600	2500	3200	6400
Délka chybějícího úseku [ms]	10,00	20,00	40,00	80,00	100,00	156,25	200,00	400,00

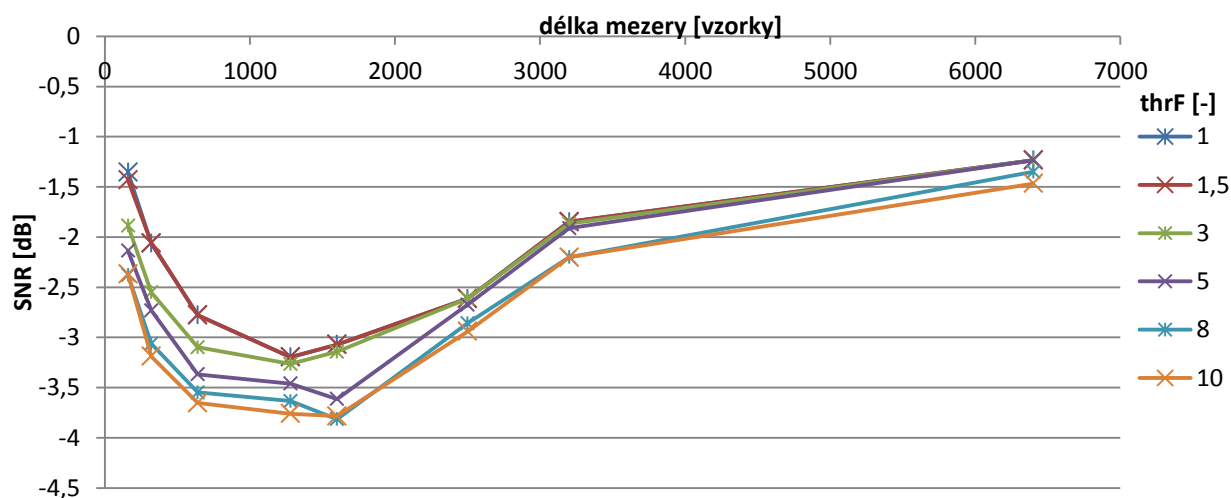
Každý úsek byl rekonstruován s nastavením parametru

$\text{thrF} = [1 \ 1.5 \ 3 \ 5 \ 8 \ 10];$ .

U obou typů signálu bylo tedy pro každý parametr  $\text{thrF}$  získáno třicet hodnot SNR. Výpočet hodnot probíhal pouze v úseku mezery, ne v rámci celého signálu. Nakonec byl vypočítán aritmetický průměr pro každou délku mezery a každý parametr  $\text{thrF}$ . Výsledky vidíme v grafech v následujících kapitolách 6.1.1 a 6.1.2.

### 6.1.1 Výsledky SNR řečových signálů

Z průběhu grafu na obr. 6 vidíme z počátku klesající tendenci hodnoty SNR s rostoucí mezerou, což bývá u jiných metod rekonstrukce běžným jevem. U délek mezer nad 100 ms však vidíme, že podle hodnocení SNR se kvalita rekonstrukce opět zlepšuje.



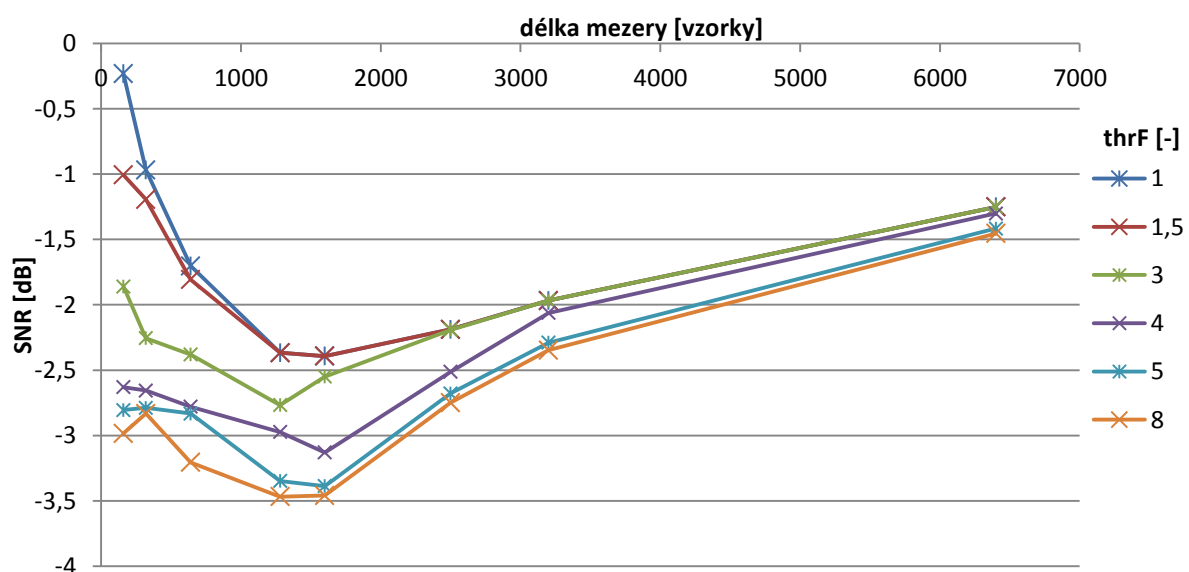
Obr. 6: Závislost SNR na délce mezery a parametru  $\text{thrF}$  u řečových signálů

Vidíme také, že s rostoucím prahem vzdáleností vektorů hodnot frekvence protějších kmitočtových složek kvalita rekonstrukce podle hodnocení SNR klesá. Rozdíl kvalit se však se zvětšující se mezerou nad 150 ms zmenšuje. Doporučená volba parametru  $\text{thr}_F$  je tedy pro rekonstrukci řečového signálu  $\text{thr}_F = 1$ .

## 6.2.2 Výsledky SNR hudebních signálů

V grafu na obr. 7 vidíme podobné průběhy výsledků hodnot SNR jako u průběhů výsledků u řečových signálů. Z počátku u kratších mezer kvalita rekonstrukce podle hodnocení SNR opět klesá. Stejně jako u výsledků v předchozí kapitole s délkou mezer nad 100 ms začne kvalita rekonstrukce stoupat. U mezer do 20 ms je však u hudebních signálů, na rozdíl od řečových, dosaženo lepších výsledků než u nejdlejší testované mezery o délce 400 ms. I celkově dosahuje kvalita rekonstrukce podle hodnocení SNR lepších výsledků než u řečového signálu.

Rekonstrukci hudebního signálu ovlivňuje parametr  $\text{thr}_F$  stejně jako rekonstrukci řečového signálu. S rostoucím prahem vzdálenosti vektorů hodnot frekvence protějších kmitočtových složek kvalita rekonstrukce podle hodnocení SNR opět klesá. S rostoucí mezerou nad 200 ms však výsledky rekonstrukce neovlivňuje takovým rozdílem jako u kratších mezer. Nejvhodnější volbou parametru podle hodnocení SNR je opět  $\text{thr}_F = 1$ .



Obr. 7: Závislost SNR na délce mezery a parametru  $\text{thr}_F$  u hudebních signálů

### 6.1.3 Shrnutí výsledků SNR

Z výsledků objektivního testování metodou SNR u obou typů signálu vyplynula ideální hodnota nastavení parametru  $thr_F$ . Z obou grafů také vidíme, že lepších výsledků je dosaženo u hudebního signálu. Z průběhů závislostí vyplývá, že interpolace je také vhodnější pro rekonstrukci delších úseků nad 150 ms, kde ostatní metody mají klesající tendenci kvality rekonstrukce. Při testování však nebylo dokázáno, při jaké délce mezery začne kvalita opět klesat.

Aritmetické průměry hodnot výsledků SNR však mají zápornou hodnotu, což znamená výrazné rozdíly v časovém průběhu původního a rekonstruovaného signálu. Výsledky SNR u některých jednotlivých souborů však dosáhly kladných hodnot, viz tab. 5.

Tab. 5: Ukázka aritmetických průměrů výsledků hodnot SNR souboru music02\_16kHz.wav

$thr_F$ [-]	Délka chybějícího úseku [ms]							
	10,00	20,00	40,00	80,00	100,00	156,25	200,00	400,00
1,0	2,36629	1,79673	0,38127	-0,04819	-0,52318	-1,41301	-0,86430	-0,64246
1,5	1,79962	1,79806	-0,13486	-0,04819	-0,52318	-1,41301	-0,86430	-0,64246
3,0	1,53345	-1,47179	-0,30738	-0,34801	-0,52021	-1,41301	-0,86430	-0,64228
5,0	-1,54757	-1,53884	-1,67510	-0,34895	-1,51502	-2,51687	-0,87991	-0,64228
8,0	-2,01420	-1,53884	-1,68104	-1,37923	-1,87993	-2,48787	-0,98673	-0,95874
10,0	-2,38987	-0,93119	-2,65174	-1,59689	-1,87993	-2,49012	-1,01229	-1,03302

K hodnocení vnímání rozdílu mezi původním a rekonstruovaným signálem slouží následující metoda.

### 6.2 Výsledky PEMO-Q

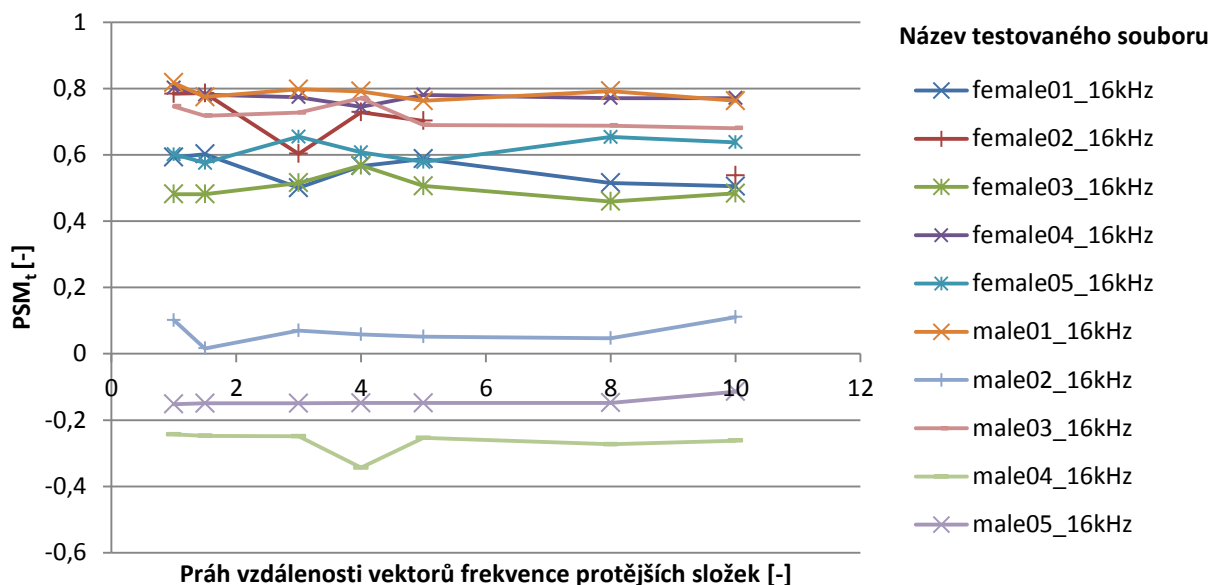
V části testovacího skriptu určené pro získání výsledků rekonstrukce objektivní metodou PEMO-Q bylo parametrem `gaps` v každém zvukovém souboru vytvořeno dvanáct mezer s různými délkami od 10 ms do 100 ms. Z výsledků testování kvality rekonstrukce objektivní metodou SNR vyplývá, že v tomto rozmezí má nastavení parametru  $thr_F$  největší vliv na kvalitu rekonstrukce. Parametr byl opět nastaven na hodnoty

$\text{thr}_F = [1 \ 1.5 \ 3 \ 5 \ 8 \ 10];$

K výpočtu hodnot  $\text{PSM}_t$  byl použit program PEMO-Q v1.3 demo [11] s defaultním nastavením. Referenční soubory byly zkráceny na stejnou délku jako rekonstruované soubory na výstupu testovacího skriptu, tj. 2 s.

### 6.2.1 Výsledky PEMO-Q řečových signálů

Na grafu v obr. 8 vidíme, že podle objektivního hodnocení PEMO-Q nelze z výsledků u jednotlivých souborů jednoznačně určit ideální hodnotu parametru  $\text{thr}_F$ . Závislost se zvětšující se vzdáleností vektorů frekvence protějších složek má kolísavý charakter i přesto, že podle výsledků objektivního hodnocení SNR by měla kvalita rekonstruovaného signálu se zvětšujícím se parametrem klesat.

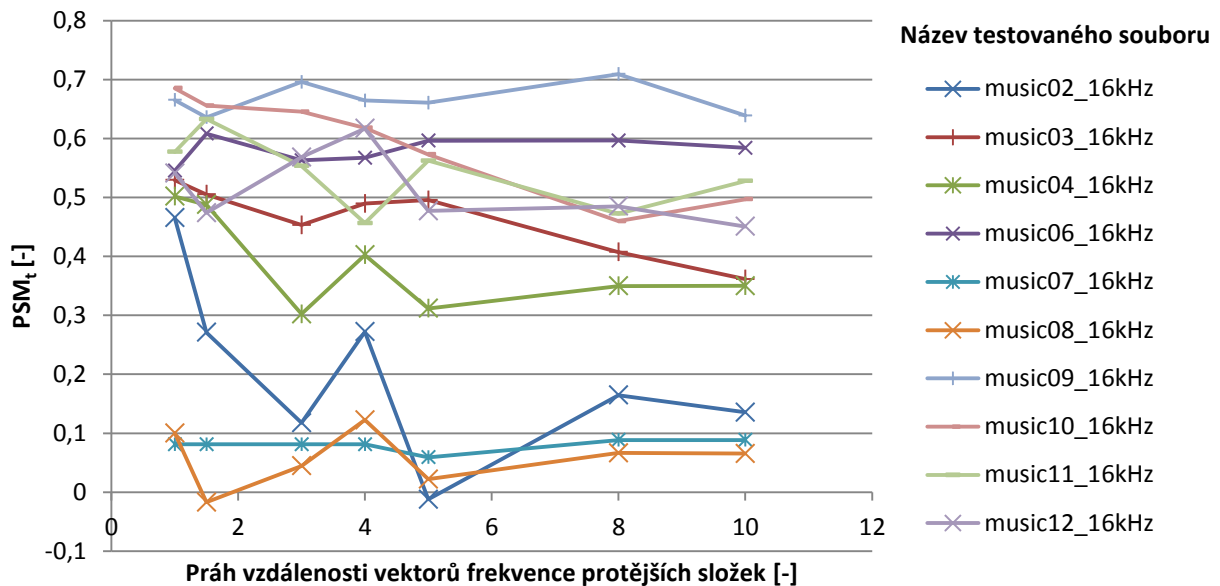


Obr. 8: Závislost  $\text{PSM}_t$  na prahu vzdálenosti vektorů frekvence protějších složek u řečového signálu

### 6.2.2 Výsledky PEMO-Q hudebních signálů

U některých rekonstruovaných hudebních souborů již vidíme na obr. 9 očekávaný výsledek, kdy hodnota  $\text{PSM}_t$  je u menší vzdálenosti vektorů frekvence protějších složek

vyšší, než u větší vzdálenosti. Za ideální hodnotu vzdálenosti vektorů frekvence protějších složek tedy můžeme určit opět  $\text{thr}_F = 1$ .



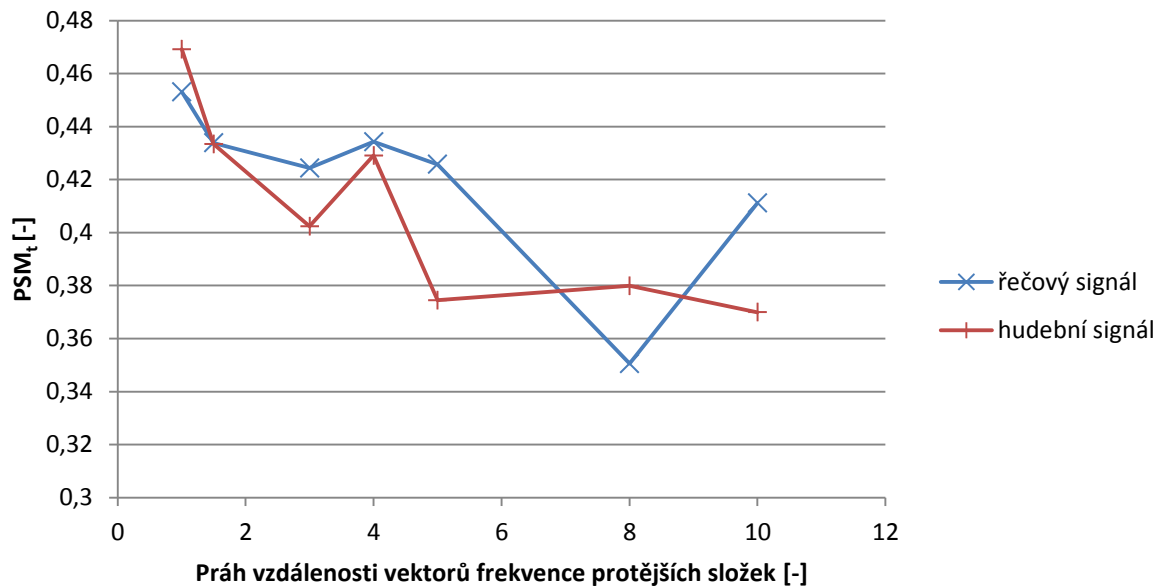
Obr. 9: Závislost  $\text{PSM}_t$  na prahu vzdálenosti vektorů frekvence protějších složek u hudebního signálu

### 6.2.3 Shrnutí výsledků PEMO-Q

Pro jednoznačnější určení kvality rekonstrukce podle objektivní metody PEMO-Q je opět vypočítán aritmetický průběh hodnot  $\text{PSM}_t$  zvláště pro řečové a hudební signály. Z grafu na obr. 10 lze vyčíst že nejlepší kvalita rekonstrukce je s nastavením parametru  $\text{thr}_F=1$  u obou typů signálu.

U některých souborů je však kvalita rekonstrukce podle hodnocení PEMO-Q hodně nízká, protože získané hodnoty  $\text{PSM}_t$  jsou záporné, což by se podle [8] v praxi nemělo stát.

Ze závislosti výsledných hodnot metody PEMO-Q vyplývá, že tato metoda je na hodnocení rozdílů kvality rekonstrukce v tak krátkých úsecích méně vhodná, protože u většiny testovaných souborů je nepatrný rozdíl výsledných hodnot  $\text{PSM}_t$  v závislosti na parametru  $\text{thr}_F$ , na rozdíl od závislosti hodnot SNR.



Obr. 10: Závislost aritmetických průměrů hodnoty  $PSM_t$  na prahu vzdálenosti vektorů frekvence protějších složek

### 6.3 Praktické ukázky rekonstruovaných signálů

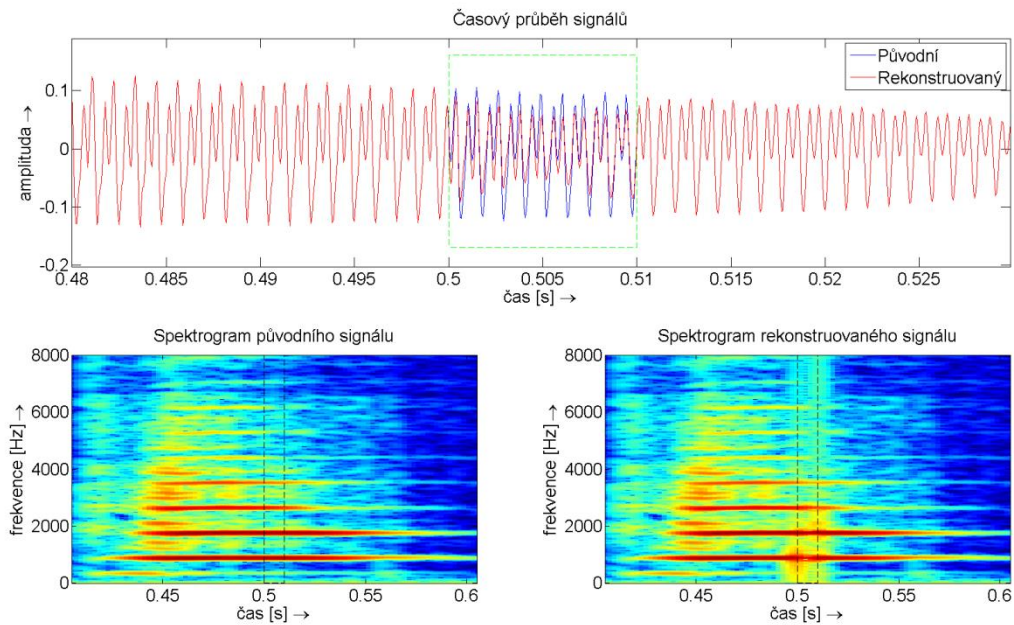
Zde jsou pomocí časových průběhů a spektrogramů původních a rekonstruovaných signálů zobrazeny konkrétní případy rekonstrukce chybějících úseků.

Prvním testovaným signálem je nahrávka hry na klarinet v souboru clar.wav s vzorkovací frekvencí  $f_{vz} = 16$  kHz. Ze spektrogramů si lze všimnout, že tento signál má malou hustotu kmitočtových složek. Na obr. 11 je ukázka rekonstrukce krátkého stacionárního úseku o délce  $l = 160$  vzorků. Vidíme zde pokles energie rekonstruovaného úseku.

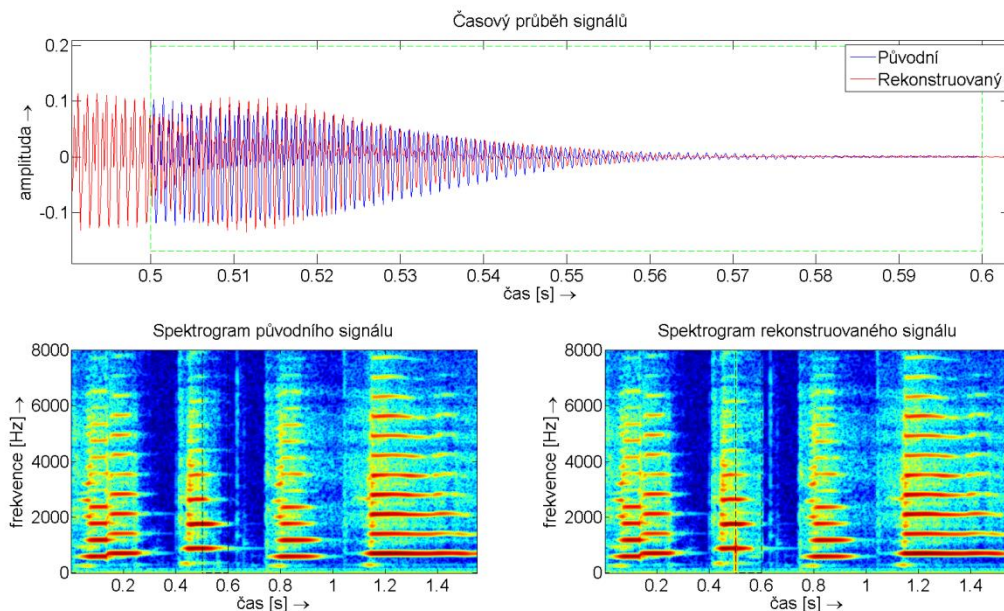
Na obr. 12 je ukázka rekonstrukce v místě, kde kmitočtové složky zasahují do mezery o délce  $l = 1600$  vzorků pouze z jedné strany. V algoritmu je provedena extrapolace těchto složek.

Na obr. 13 je ukázka rekonstrukce mezery o délce  $l = 1600$  vzorků v místě, kde na sebe kmitočtové složky původního signálu navenazují, je tedy provedena extrapolace složek z obou stran. V porovnání s extrapolací z jedné strany zde došlo k výraznému poklesu energie signálu směrem ke středu mezery.

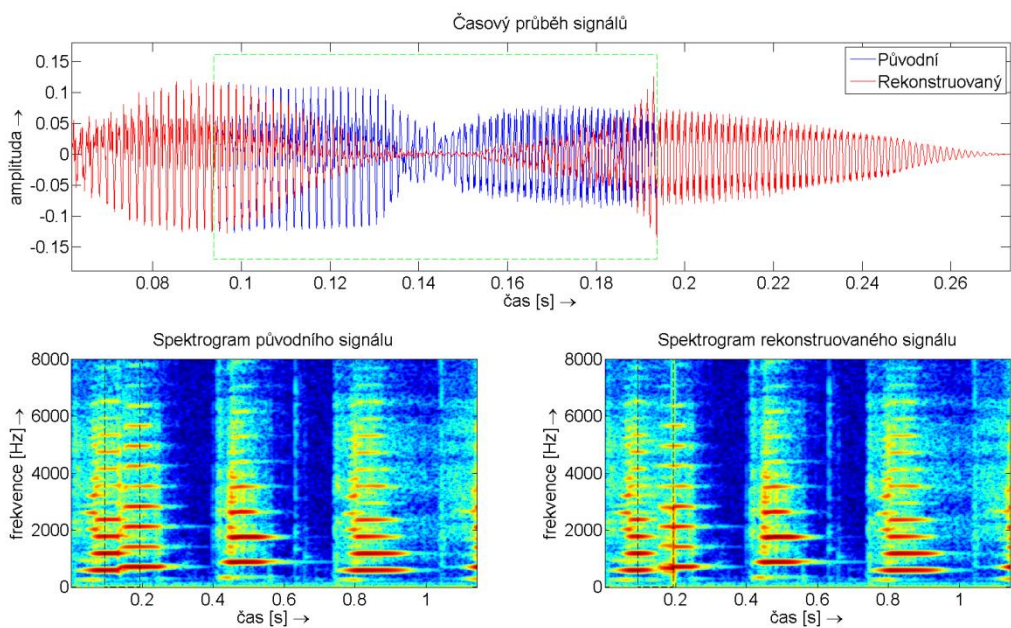
Na obr. 14 je ukázka rekonstrukce mezery o délce  $l = 1600$  vzorků v místě, kde na sebe původní složky navazují. Algoritmus však tyto složky nespojil a dochází tím opět k výraznému poklesu energie směrem ke středu mezery.



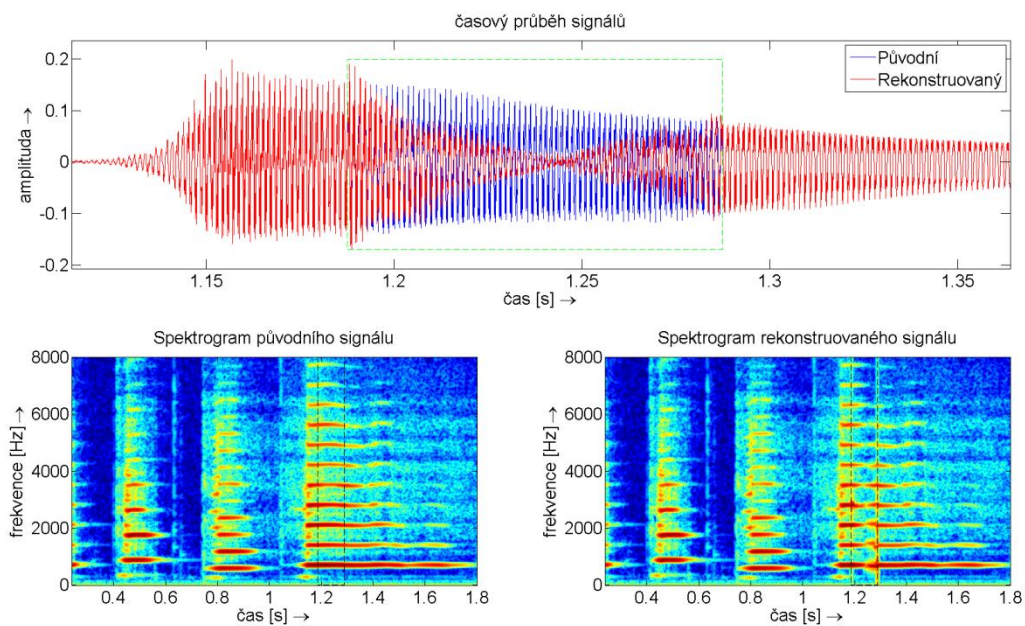
Obr. 11: Rekonstrukce mezery s délkou  $l = 160$  vzorků stacionárního úseku signálu



Obr. 12: Extrapolace signálu z levé strany do mezery délky  $l = 1600$  vzorků



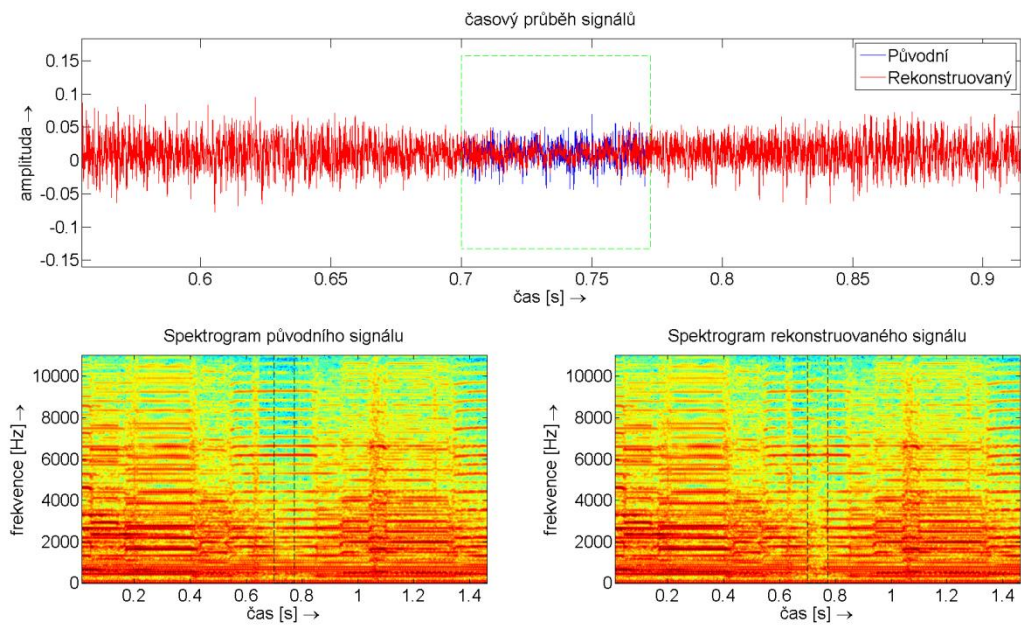
**Obr. 13: Extrapolace signálu z obou stran do mezery délky  $l = 1600$  vzorků**



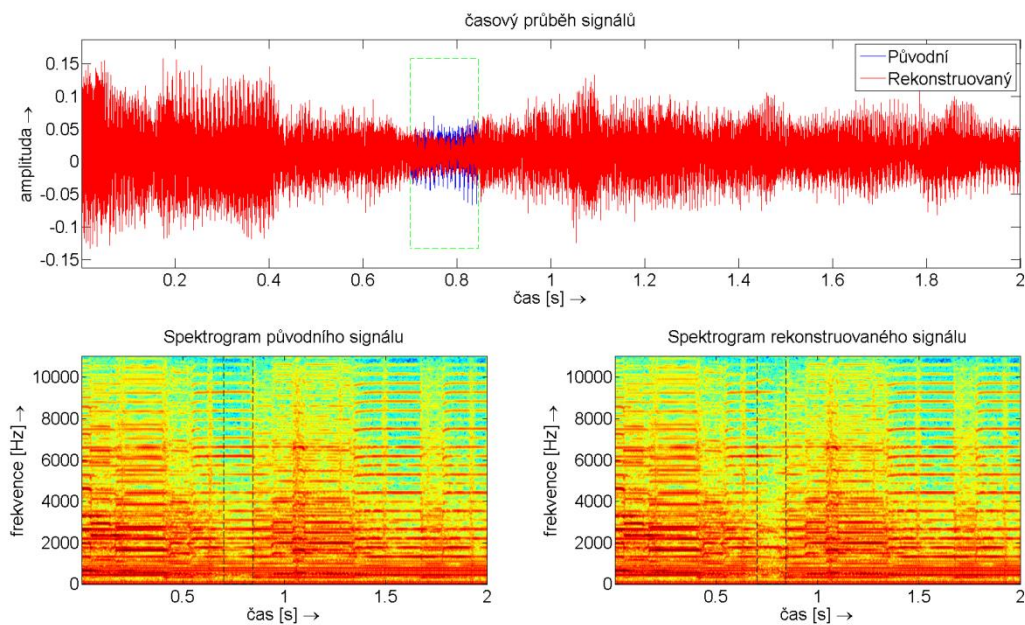
**Obr. 14: Interpolace mezery s délkou  $l = 1600$  vzorků**



Dalším testovaným signálem byl zvukový záznam hry na dudy v souboru bagpipe.wav s vzorkovací frekvencí  $f_{vz} = 22,05$  kHz. Na spektrogramu signálu vidíme výraznější hustotu kmitočtových složek.



Obr. 15: Rekonstrukce úseku o délce  $l = 1600$  vzorků u signálu s větší spektrální hustotou



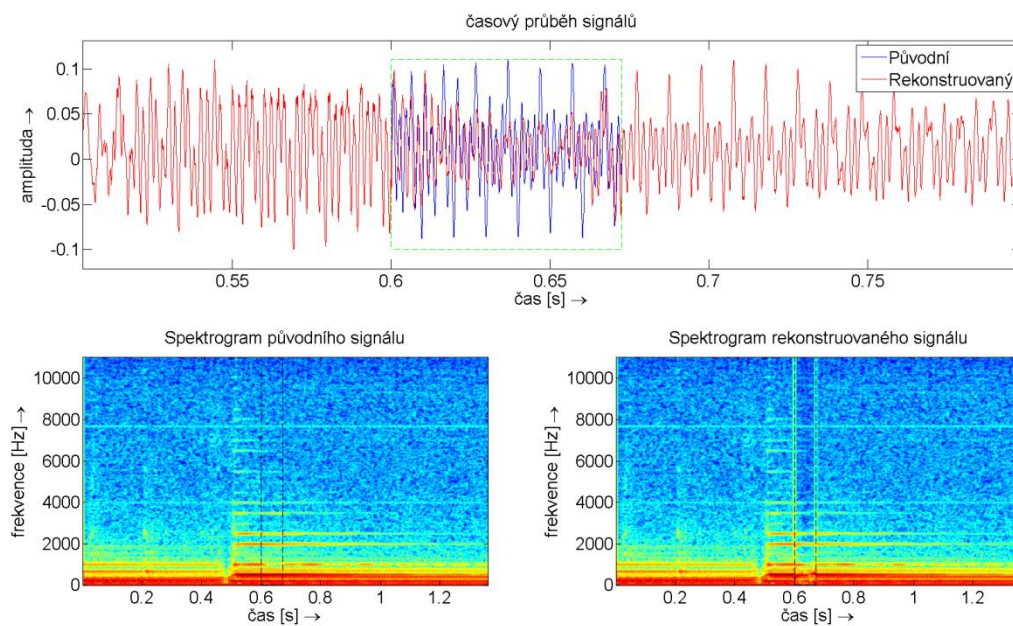
Obr. 16: Rekonstrukce úseku o délce  $l = 3200$  vzorků u signálu s větší spektrální hustotou

U rekonstruovaných úseků obou délek na obr. 15 a 16 vidíme, že pokles energie rekonstruovaného úseku je menší než v případě signálu s menší spektrální hustotou a také že nedošlo k výraznému poklesu při dvojnásobné délce mezery. Na spektrogramu rekonstruovaného signálu však vidíme, že při dané spektrální hustotě nižších kmitočtů má algoritmus opět problém se spojováním protějšších kmitočtových složek.

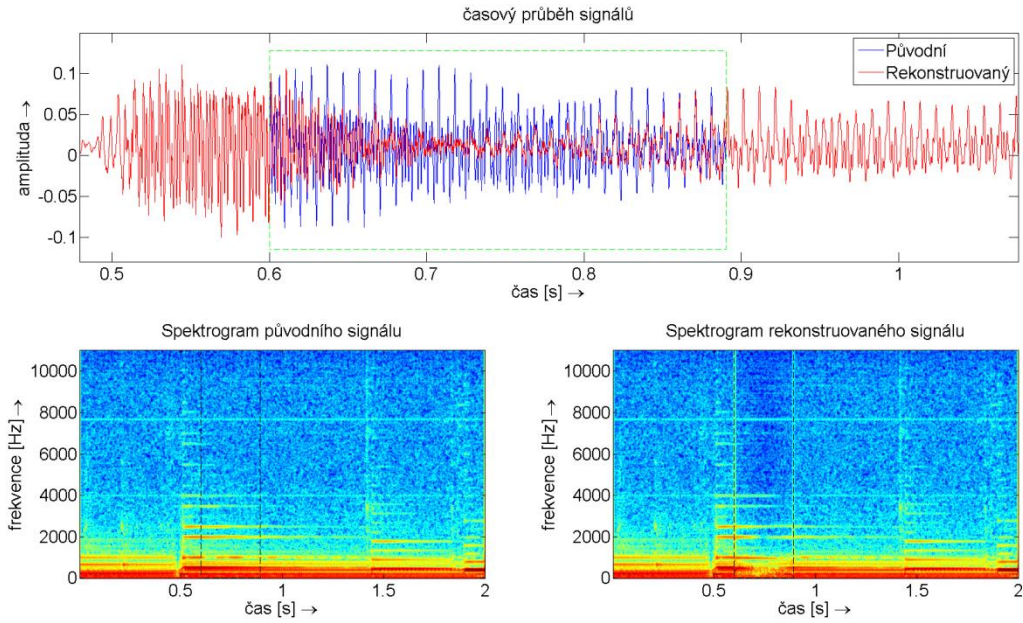
Následujícím testovaným signálem je nahrávka hry na kytaru v souboru guitar.wav s vzorkovací frekvencí  $f_{vz} = 22,05$  kHz. Vyskytují se zde dlouhé kmitočtové složky díky doznívajícím tónům kytary, proto byla testována rekonstrukce delší mezery.

Na obr. 17 je ukázka rekonstrukce mezery o délce  $l = 1600$  vzorků. Na spektrogramu vidíme, že došlo opět k problému s interpolací i u delších složek a rekonstruovaný úsek se od původního signálu liší.

U delší mezery na obr. 18 došlo kvůli stejnému jevu ještě k výraznějšímu poklesu energie rekonstruovaného signálu.

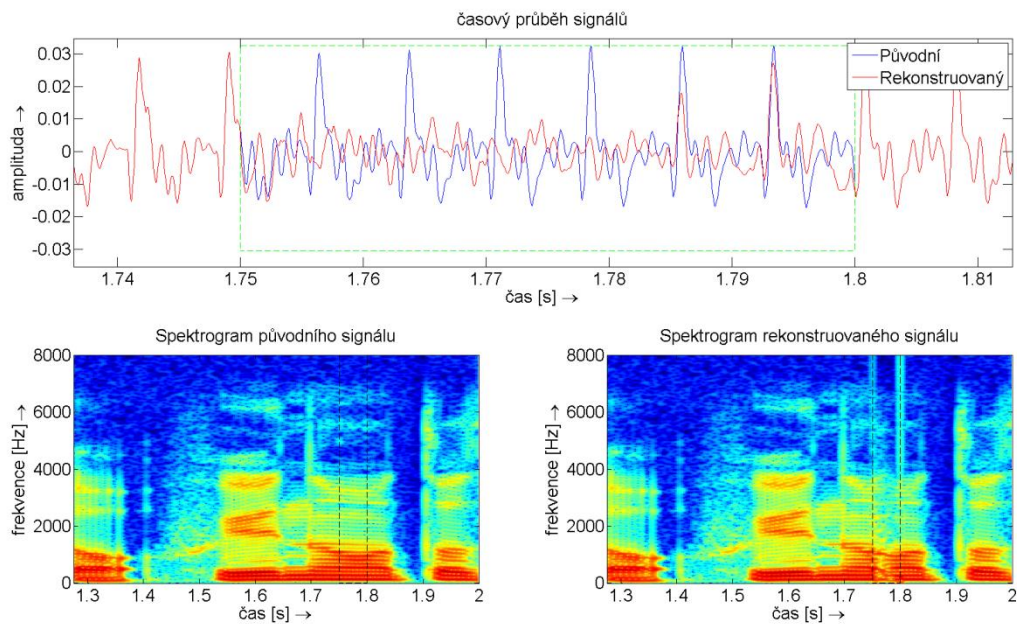


**Obr. 17: Rekonstrukce mezery o délce  $l = 1600$  vzorků u signálu s dlouhými kmitočtovými složkami**

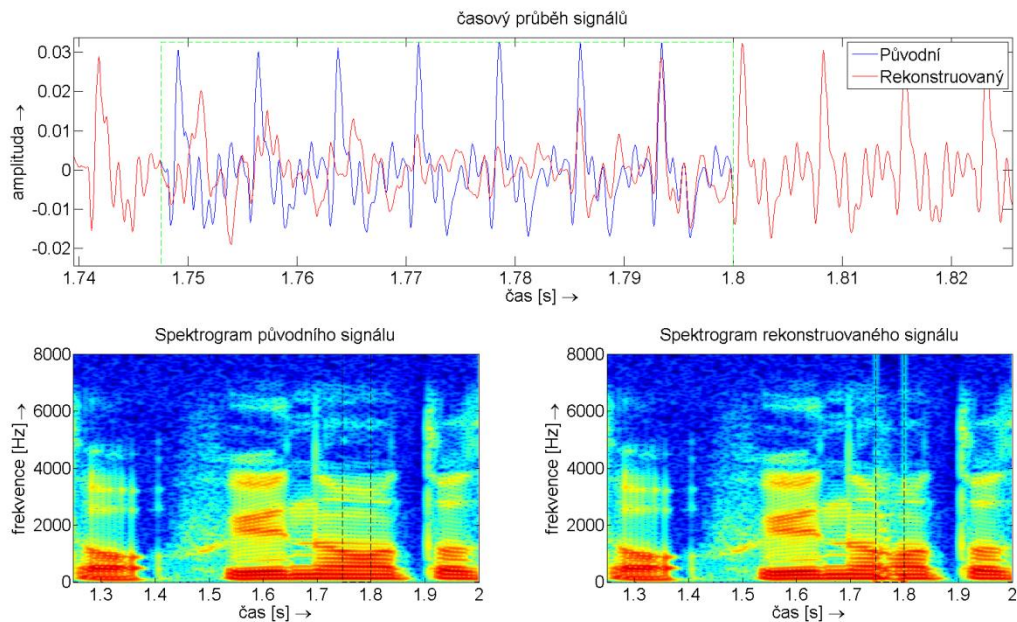


**Obr. 18: Rekonstrukce mezery o délce  $l = 6400$  vzorků u signálu s dlouhými kmitočtovými složkami**

Na testu rekonstrukce řečového signálu je ukázka vlivu polohy mezery a průběhu okolního signálu na výsledný časový průběh rekonstruovaného signálu. Rekonstrukce byla testována na souboru `male02_16kHz.wav` s vzorkovacím kmitočtem  $f_{vz} = 16$  kHz.



**Obr. 19: Rekonstrukce mezery s délkou  $l = 800$  vzorků v řečovém signálu**



**Obr. 20: Rekonstrukce mezery s délkou  $l = 840$  vzorků v řečovém signálu**

Na obr. 19 vidíme, že díky poloze mezery je podoba rekonstruovaného signálu s původním přesnější na pravé straně mezery, zatímco na levé straně je velká odchylka. Na obr. 20 je mezera na levé straně posunuta o 40 vzorků, tedy těsně za špičku signálu v daném úseku. Z průběhu rekonstruovaného signálu vidíme díky změně polohy zlepšení kvality rekonstrukce.

## 6.4 Shrnutí výsledků experimentů

Z výsledků hodnocení podle objektivní metody SNR a PEMO-Q jsme zjistili doporučené nastavení parametrů rekonstrukčního algoritmu. Podle získaných hodnot bylo zjištěno, že kvalita rekonstrukce s rostoucí délkou mezery z počátku klesá, u mezer s délkou nad 150 ms začne kvalita podle hodnocení SNR opět stoupat. Z toho tedy vyplývá, že algoritmus je vhodnější pro rekonstrukci delších úseků. Podle hodnot SNR bylo také zjištěno, že lepší kvality rekonstrukce je dosaženo u hudebních signálů. S doporučeným nastavením prahu vzdálenosti vektoru frekvence protějšších složek toto tvrzení potvrzují i výsledky podle hodnocení objektivní metodou PEMO-Q.

Na konkrétních příkladech experimentu rekonstruování signálu však bylo dokázáno, že algoritmus zvládá rekonstrukci mezer stacionárního signálu s krátkými úseky v řádech desítek milisekund. S rostoucí délkou mezery nad 100 ms dochází k výraznému poklesu energie rekonstruovaného signálu. Vhodné použití algoritmu se ukázalo v případě, kdy kmitočtové složky v chybějícím úseku zanikají, tzn. odhadované hodnoty kmitočtových složek jsou extrapolovány z jedné strany. Největším problémem je spojování protějších kmitočtových složek u mezer delších než 100 ms. U signálů s větší spektrální hustotou však docházelo k menšímu poklesu energie rekonstruovaného signálu než u signálů s menší spektrální hustotou a to i u mezer delších než 100 ms.

Zásadním faktorem ovlivňující výsledek rekonstrukce jsou však parametry a délka kmitočtových složek na hranicích s mezerou, z kterých se provádí výpočet odhadovaných hodnot. Kvalita rekonstrukce chybějícího úseku tedy závisí na její poloze v signálu.

## 7 ZÁVĚR

Podle teoretického základu v kapitolách 1, 2 a 3 byla zrealizována metoda pro rekonstrukci chybějících úseků v signálu. Výsledky rekonstrukce byly v hromadném testu hodnoceny objektivními metodami uvedenými v kapitole 4 pro různé délky úseků a s různým nastavením parametrů ovlivňujících výsledky interpolace.

Dále byly provedeny testy rekonstrukce na jednotlivých zvukových souborech v úsecích s odlišnými průběhy signálů a byly zobrazeny výsledky rekonstruovaných úseků v časové a časově-frekvenční rovině. Bylo zjištěno, že se výsledky v mnoha případech výrazně liší v závislosti na okolním signálu mezery. Kvality rekonstrukcí pomocí realizovaného algoritmu však nedosahují takových hodnot jako jiné, v praxi již nasazené metody.

Ze shrnutí výsledků experimentů vyplývá, že je realizace algoritmu vhodnější pro rekonstrukci mezer s délkou v řádu stovek vzorků u signálů s nízkou hustotou kmitočtových složek. Dá se tedy použít k rekonstrukci krátkých úseků signálu řeči či nahrávek jednotlivých nástrojů. V případě hudebních nahrávek, kde se předpokládá větší frekvenční rozsah a větší spektrální hustota, je algoritmus vhodný i pro rekonstrukci delších úseků. Z praktického hlediska je rekonstrukce delších úseků zvukového signálu vhodná většinou právě pro hudební signály v případě, že nedochází k častým změnám signálu v čase, např. u klasické hudby v pomalém tempu či u nahrávek jednotlivých nástrojů hrajících déle trvající tóny. Pro obnovu ztrát v přenosu řečového signálu nemá rekonstrukce delších úseků velký význam, protože dochází k rychlým změnám hlásek a prodloužením mezeře předcházející či za mezerou následující hlásky ztratíme původní informaci v řeči. Můžeme se však vyhnout alespoň na poslech nepříjemnému lupnutí.

## LITERATURA

- [1] FINK, Marco, Martin HOLTERS a Udo ZÖLZER. Comparison of Various Predictors for Audio Extrapolation. *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13)* [online]. 2013, 1 - 8 [cit. 2014-05-23]. Dostupné z: [http://dafx13.nuim.ie/papers/42.dafx2013\\_submission\\_27.pdf](http://dafx13.nuim.ie/papers/42.dafx2013_submission_27.pdf)
- [2] GODSILL, Simon J. *Digital audio restoration: a statistical model based approach*. New York: Springer, c1998, 328 s. ISBN 35-407-6222-1.
- [3] HAYKIN, Simon S. *Adaptive filter theory: a statistical model based approach*. 2nd ed. New Jersey,: Prentice-Hall, 1991, xx, 854 s. ISBN 01-300-5513-1.
- [4] SMĚKAL, Zdeněk. *Číslicové zpracování signálů*. Brno: Vysoké učení technické v Brně, 2012. ISBN 978-80-214-4639-7.
- [5] LAGRANGE, Mathieu, Sylvain MARCHAND a Jean-Bernard RAULT. Long Interpolation of Audio Signals Using Linear Prediction in Sinusoidal Modeling. *J. Audio Eng. Soc.* [online]. 2005, roč. 53, č. 10 [cit. 2014-05-23]. Dostupné z: <http://www.aes.org/e-lib/browse.cfm?elib=13390>
- [6] LUKIN, Alexey a Jeremy TODD. *Parametric Interpolation of Gaps in Audio Signals* [online]. Audio Engineering Society Convention 125, 2008 [cit. 2014-05-23]. Dostupné z: <http://www.aes.org/e-lib/browse.cfm?elib=14664>
- [7] LAGRANGE, Mathieu, Sylvain MARCHAND a Jean-Bernard RAULT. Using linear prediction to enhance the tracking of partials [musical audio processing]. *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing* [online]. IEEE, 2004, iv-241-iv-244 [cit. 2014-05-23]. DOI: 10.1109/ICASSP.2004.1326808. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1326808>

- [8] HUBER, Rainer a Birger KOLLMEIER. 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]. 2006, vol. 14, issue 6, s. 1902-1911 [cit. 2014-05-23]. DOI: 10.1109/TASL.2006.883259. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1709880>
- [9] ELLIS, Dan. Sinewave and Sinusoid+Noise Analysis/Synthesis in Matlab. *The Fu Foundation School of Engineering & Applied Science - Columbia University* [online]. 2011 [cit. 2014-05-22]. Dostupné z: <http://www.ee.columbia.edu/ln/rosa/matlab/sinemodel/>
- [10] ETTER, Walter. Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters. *IEEE Transactions on Signal Processing* [online]. vol. 44, issue 5, s. 1124-1135 [cit. 2014-05-23]. DOI: 10.1109/78.502326. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=502326>
- [11] PEMO-Q. *HörTech gmbH* [online]. 2004 [cit. 2014-05-28]. Dostupné z: [http://www.hoertech.de/web\\_en/produkte/pemo-q.shtml](http://www.hoertech.de/web_en/produkte/pemo-q.shtml)
- [12] JANSSEN, A., R. VELDHUIS a L. VRIES. Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing* [online]. 1986, vol. 34, issue 2, s. 317-330 [cit. 2014-05-28]. DOI: 10.1109/TASSP.1986.1164824. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1164824>
- [13] 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, vol. 20, issue 3, s. 922-932 [cit. 2014-05-28]. DOI: 10.1109/TASL.2011.2168211. Dostupné z: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6020748>



[14] The Large Time-Frequency Analysis Toolbox. *SourceForge: Download, Develop and Publish Free Open Source Software* [online]. 2014 [cit. 2014-05-28]. Dostupné z: <http://lftfat.sourceforge.net/>

[15] Data Sets. GEORGE TZANETAKIS. *Marsyas: Music Analysis, Retrieval and Synthesis for Audio Signals* [online]. 2009 [cit. 2014-05-28]. Dostupné z: [http://marsyas.info/download/data\\_sets/](http://marsyas.info/download/data_sets/)

[16] MACH, V., C. WIESMEYR, Methods of Completing Missing Samples in Audio Signal. *Conference Proceedings: International Masaryk Conference for Ph.D. Students and Young Researchers* 2013. Hradec Králové: MAGNANIMITAS, 2013. ISBN: 978-80-87952-00-9.

## **SEZNAM ZKRATEK**

AR Autoregresní

FIR Finite Impuls Response – Konečná impulsní odezva

SNR Signal-to-Noise Ratio – Poměr odstupů signálu od šumu

STFT Short-Time Fourier Transform – krátkodobá Fourierova transformace

## **SEZNAM PŘÍLOH**

A Obsah elektronické přílohy

57

## A OBSAH ELEKTRONICKÉ PŘÍLOHY

Přiložené CD obsahuje elektronickou verzi diplomové práce ve formátu PDF a zdrojové kódy algoritmu.

Důležité soubory a složky:

- ./SinModInt.m - hlavní funkce rekonstrukčního algoritmu
- ./start.m - skript pro testování jednotlivých souborů
- ./testskript.m - skript pro spuštění hromadného testování
- ./funkce/ - obsahuje pomocné funkce
- ./scripts/ - obsahuje skript potřebný ke spuštění v Brno -Wien  
Inpainting Toolbox
- ./signals/ - obsahuje testovací zvukové soubory
- ./sinemodel/ - obsahuje funkce toolboxu Sinewave and Sinusoid+Noise  
Analysis/Synthesis in Matlab