



TECHNICKÁ UNIVERZITA V LIBERCI  
Fakulta mechatroniky, informatiky  
a mezioborových studií ■

# Rozklad akustické impulzní odezvy na brzké odrazy a difúzní část

## Diplomová práce

*Studijní program:* N2612 – Elektrotechnika a informatika

*Studijní obor:* 1802T007 – Informační technologie

*Autor práce:* **Bc. Tomáš Franěk**

*Vedoucí práce:* doc. Ing. Zbyněk Koldovský, Ph.D.





TECHNICAL UNIVERSITY OF LIBEREC  
Faculty of Mechatronics, Informatics  
and Interdisciplinary Studies ■

# Decomposition of acoustical impulse responses into early reflection and diffused parts

**Master thesis**

*Study programme:* N2612 – Electrotechnology and informatics

*Study branch:* 1802T007 – Information technology

*Author:* **Bc. Tomáš Franěk**

*Supervisor:* doc. Ing. Zbyněk Koldovský, Ph.D.



## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Tomáš Franěk**  
Osobní číslo: **M15000163**  
Studijní program: **N2612 Elektrotechnika a informatika**  
Studijní obor: **Informační technologie**  
Název tématu: **Rozklad akustické impulzní odezvy na brzké odrazy a difúzní část**  
Zadávající katedra: **Ústav informačních technologií a elektroniky**

### Z á s a d y p r o v y p r a c o v á n í :

1. Proveďte rešerši tématu. Vyhledejte, jak je běžně definovaný rozklad akustických impulzních odezev a jak jsou tyto části měřeny nebo odhadovány.
2. Proveďte experimentální studii, tj. měření impulzních odezev v různých prostředích, jejich částí a ověření vlastností.
3. Navrhněte rozklad impulzní odezvy tak, aby konvoluce signálu s první částí byla nezávislá (nebo jenom nekorelovaná) na konvoluci stejného signálu s druhou částí.
4. Ověřte navržený rozklad v metodách beamformingu za účelem redukce šumu ve vícekanálovém signálu a dereverberaci.

Rozsah grafických prací: Dle potřeby dokumentace

Rozsah pracovní zprávy: cca 40-50 stran

Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

- [1] O. Schwartz, S. Gannot, and E. A. P. Habets, Multimicrophone speech dereverberation and noise reduction using relative early transfer functions, *IEEE/ACM Transactions on Audio, Speech, and Language*
- [2] J. Allen, D.A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, vol. 62, no. 4, pp. 912-915, 1977.
- [3] I. J. Tashev, *Sound Capture and Processing: Practical Approaches*, Wiley, 2009.

Vedoucí diplomové práce: doc. Ing. Zbyněk Koldovský, Ph.D.

Ústav informačních technologií a elektroniky

Konzultant diplomové práce: Ing. Jiří Málek, Ph.D.


Ústav informačních technologií a elektroniky

Datum zadání diplomové práce: 12. září 2016

Termín odevzdání diplomové práce: 15. května 2017

  
prof. Ing. Zdeněk Pluha, Ph.D.  
děkan



  
prof. Ing. Ondřej Novák, CSc.  
vedoucí ústavu

V Liberci dne 12. září 2016

## Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

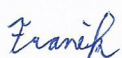
Beru na vědomí, že Technická univerzita v Liberci (TUL) nezasahuje do mých autorských práv užitím mé diplomové práce pro vnitřní potřebu TUL.

Užiji-li diplomovou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědom povinnosti informovat o této skutečnosti TUL; v tomto případě má TUL právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím mé diplomové práce a konzultantem.

Současně čestně prohlašuji, že tištěná verze práce se shoduje s elektronickou verzí, vloženou do IS STAG.

Datum: 15.5.2017

Podpis: 

## Abstrakt

Tato práce se zabývá rozdělením akustické impulsní odezvy místnosti na část s přímou vlnou a brzkými odrazy a na difuzní část. Cílem práce je navrhnout metodiku pro tento rozklad.

Pro vyřešení tohoto tématu jsem nejdříve v různých místnostech nahrál několik nahrávek. Z těchto nahrávek jsem odhadl akustické impulsní odezvy. Tyto akustické impulsní odezvy jsem postupně rozděloval na dvě části. Z každou částí jsem provedl konvoluci nahraného signálu. Mezi oběma vzniklými signály jsem testoval závislost pomocí korelačního koeficientu a vzájemné informace. Výslednou hranici mezi brzkými odrazy a difuzní částí jsem určil v místě, kde práh závislosti klesne pod určitou mez.

Výsledný rozklad jsem pak ověřil odstraněním difuzní části z reálné nahrávky pomocí Wienerova postfiltru.

## Klíčová slova:

akustická impulsní odezva, rozklad odezvy, dereverberace, zlepšení kvality řeči

## Abstract

Aim of this thesis is to decompose acoustic impulse response into two parts. First part belongs to direct wave and early reflections and second part belongs to diffusion. In this thesis there is found methodology of this decomposition.

In this thesis there are made several records in different rooms in the beginning. Acoustic impulse responses are estimated from these records. These acoustic impulse responses are decomposed into two parts sample by sample. Each part is convoluted with recorded signal. Dependence between new signals by correlation coefficient and mutual information is tested. As a result, we can find obtained information about border between early reflection and diffuse part when value of the correlation coefficient first gets below threshold.

Resulting decomposition is verified by dereverberation of real record by Wiener postfilter.

## Key words:

acoustic impulse response, response decomposition, dereverberation, speech enhancement

## Poděkování

Tímto bych chtěl poděkovat doc. Ing. Zbyňkovi Koldovskému, Ph.D. za cenné rady, vstřícnost při konzultacích a za zapůjčení profesionálních mikrofونů.



# Obsah

<b>1</b>	<b>Úvod</b>	<b>14</b>
<b>2</b>	<b>Měření impulsních odezev</b>	<b>18</b>
2.1	Tvorba nahrávek . . . . .	18
2.2	Odhad impulsních odezev . . . . .	21
2.3	Ověření vlastností . . . . .	22
<b>3</b>	<b>Použité algoritmy</b>	<b>27</b>
3.1	Overlap Add . . . . .	27
3.2	Vzájemná korelace . . . . .	29
3.3	Wienerův filtr . . . . .	30
3.4	Korelační koeficient . . . . .	30
3.5	Vzájemná informace . . . . .	31
3.6	PESQ . . . . .	32
<b>4</b>	<b>Rozdělení akustické impulsní odezvy</b>	<b>33</b>
4.1	Bez synchronizace mezi $\mathbf{s}_e$ a $\mathbf{s}_r$ . . . . .	33
4.2	Synchronizace $\mathbf{s}_e$ a $\mathbf{s}_r$ pomocí vzájemné korelace . . . . .	37
4.3	Synchronizace $\mathbf{s}_e$ a $\mathbf{s}_r$ pomocí Wienerova filtru . . . . .	39
<b>5</b>	<b>Dereverberace</b>	<b>42</b>
<b>6</b>	<b>Závěr</b>	<b>47</b>
	<b>Literatura</b>	<b>49</b>

<b>Přílohy</b>	<b>50</b>
<b>A Obsah přiloženého CD</b>	<b>51</b>
<b>B Spektrogramy nahraných signálů</b>	<b>52</b>
<b>C Průběhy</b>	<b>60</b>

## Seznam obrázků

1.1	Definice hlavních částí akustické impulsní odezvy s difuzní částí po 25 ms . . . . .	14
1.2	Spektrogram akustické impulsní odezvy . . . . .	15
1.3	Průběh hodnot korelací . . . . .	16
2.1	Rozmístění reproduktoru a mikrofonů v místnosti . . . . .	19
2.2	Příklad odhadnuté akustické impulsní odezvy . . . . .	22
2.3	Porovnání AIR s různými vzdálenostmi mikrofonů . . . . .	23
2.4	Porovnání AIR podle typu přehrávaného zvuku ve stejných podmínkách	24
2.5	Porovnání detailu AIR podle typu přehrávaného zvuku ve stejných podmínkách . . . . .	25
2.6	Průměrná kvadratická chyba mezi AIR s různými typy přehrávaných signálů . . . . .	25
2.7	Spektrogram signálu $r$ a $r - \hat{r}$ . . . . .	26
3.1	Princip Overlap add . . . . .	28
3.2	Porovnání rychlosti Overlap add a konvoluce . . . . .	28
3.3	Porovnání odchylky výsledků konvoluce a Overlap add . . . . .	29
4.1	Rozdělení $h$ na potenciální $h_e$ a $h_r$ . . . . .	34
4.2	Porovnání průběhů korelačního koeficientu . . . . .	35
4.3	Porovnání průběhů vzájemné informace . . . . .	36
4.4	Příklad detailu vzájemné korelace . . . . .	37
4.5	Porovnání průběhů korelačního koeficientu . . . . .	38

4.6	Porovnání průběhů vzájemné informace . . . . .	39
4.7	Porovnání průběhů korelačního koeficientu . . . . .	40
4.8	Porovnání průběhů vzájemné informace . . . . .	41
5.1	Porovnání reverberovaného signálu s ideálním $\mathbf{s}_e$ a získaným odfiltrovaným $\hat{\mathbf{s}}_e$ . . . . .	45
5.2	Porovnání ideální difuzní části a odfiltrované difuzní části . . . . .	46
B.1	Porovnání nahrávek šumu 1 m . . . . .	52
B.2	Porovnání nahrávek šumu 2 a 3 m . . . . .	53
B.3	Porovnání nahrávek sine sweepu 1 m . . . . .	54
B.4	Porovnání nahrávek sine sweepu 2 a 3 m . . . . .	55
B.5	Porovnání nahrávek ženského hlasu 1 m . . . . .	56
B.6	Porovnání nahrávek ženského hlasu 2 a 3 m . . . . .	57
B.7	Porovnání nahrávek mužského hlasu 1 m . . . . .	58
B.8	Porovnání nahrávek mužského hlasu 2 a 3 m . . . . .	59
C.1	Průběhy korelačního koeficientu a vzájemné informace pro místnost 1	60
C.2	Průběhy korelačního koeficientu a vzájemné informace pro místnost 2	61
C.3	Průběhy korelačního koeficientu a vzájemné informace pro místnost 3	62

## Seznam tabulek

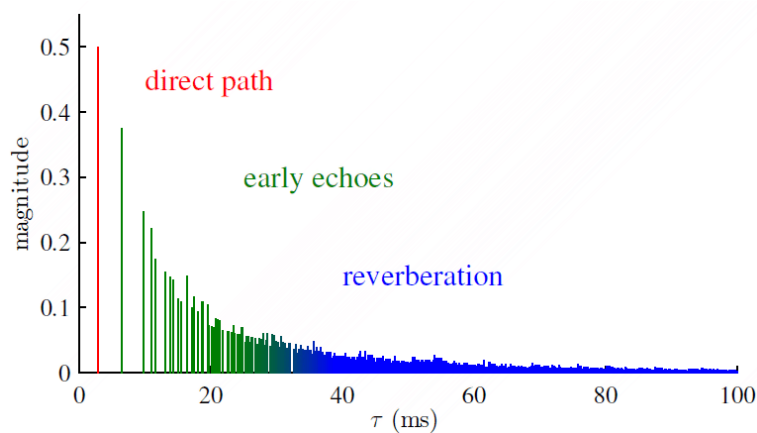
2.1	Rozměry místností . . . . .	18
2.2	Umístění reproduktoru . . . . .	19
2.3	Předpokládaná a naměřená doba šíření přímé vlny mezi mikrofony . .	23
3.1	Tabulka ohodnocení kvality řeči . . . . .	32
5.1	Hodnoty hranic $b$ pro jednotlivé místnosti a vzdálenosti mikrofonů . .	43
5.2	Hodnoty PESQ pro porovnání kvality řeči na originální nahrávce a na dereverbované nahrávce . . . . .	44

## Seznam zkratek

<b>AD</b>	Analogovo-digitální
<b>AIR</b>	Acoustic Impulse Response (akustická impulsní odezva)
<b>CD</b>	Compact Disc (kompaktní disk)
<b>FFT</b>	Fast Fourier Transformation (rychlá Fourierova transformace)
<b>LMS</b>	Least Mean Square (metoda nejmenších čtverců)
<b>MATLAB</b>	MATrix LABoratory (programové prostředí)
<b>PESQ</b>	Perceptual Evaluation of Speech Quality (standard pro porovnání kvality řeči)
<b>TDRTF</b>	Least-square time-domain estimator or relative transfer function (metoda pro odhad impulsní odezvy)
<b>TIMIT</b>	Databáze nahrávek s řečí a jejich přepisů

# 1 Úvod

Podle [1] se akustická impulsní odezva skládá ze tří hlavních částí. První část je přímá vlna ze zdroje k mikrofonu. Dále následují brzké odrazy korespondující s prvními odrazy ode stěn a nábytku. Následné několikanásobné odrazy jsou již tak husté a zdeformované, že je již nelze od sebe rozeznat a dále formují exponenciálně klesající chvost zvaný difuzní část. Tato část je tak chaotická, že odpovídá spíše šumu. Část s brzkými odrazy se pozvolna prolíná s difuzní částí, tudíž hranici mezi nimi nelze jednoznačně vypočítat tak jednoduše jako část akustické impulsní odezvy odpovídající hlavní vlně. Viz obrázek 1.1, který je též z [1].



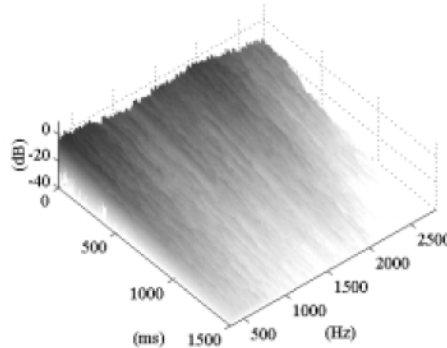
Obrázek 1.1: Definice hlavních částí akustické impulsní odezvy s difuzní částí po 25 ms

Hlavním cílem této práce je najít způsob jak rozdělit akustickou impulsní odezvu na dvě části. První část by měla obsahovat hlavní vlnu a brzké odrazy od překážek v místnosti. Druhá část by měla obsahovat vzdálené odrazy odpovídající difuzní části.

Informaci o hranici mezi brzkými odrazy a difuzní částí lze využít v aplikacích pro zlepšení řeči. Jelikož difuzní část odpovídá spíše šumu, tak nenese žádnou užitečnou informaci. Difuzní část v nahrávkách s řečí tedy spíše znesnadňuje aplikacím pro rozpoznání řeči jejich správnou funkci. S informací o hranici mezi brzkými odrazy a difuzní částí lze například vytvořit filtr, který difuzní část z nahrávky s řečí odstraní. Dále lze tuto informaci využít k redukci šumu ve vícekanálovém signálu a dereverberaci.

Podle [2] dříve existovaly názory, že hranice mezi brzkými odrazy a difuzní částí je vždy zhruba 80 ms. Někdy zase 100 ms a jindy v rozhraní mezi 50–200 ms. Poté v [2] byl vytvořen postup, jak dynamicky rozhodovat, kde se tato hranice nachází.

Z akustické impulsní odezvy byl vytvořen spektrogram (viz obrázek 1.2 [2]). Poté byla spočítána korelace mezi spektrem v počátečním čase ( $t = 0$  s) a spektrem v každém dalším čase. Z těchto hodnot poté byl vykreslen graf 1.3 [2]. Z grafu poté byl odečten čas  $t_L$ , kdy hodnota korelace poprvé poklesla pod hodnotu převráceného Eulerova čísla ( $\frac{1}{e} = 0,367$ ). Čas  $t_L$  byl poté považován za hranici mezi brzkými odrazy a difuzní částí.

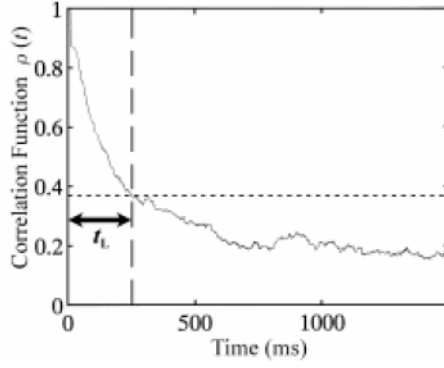


Obrázek 1.2: Spektrogram akustické impulsní odezvy

V [3] byl vytvořen následující postup pro hledání hranice mezi brzkými odrazy. Z nahrávky signálu z kulového pole mikrofónů bylo odhadnuto v každém čase množství difuzní složky  $\hat{\psi}$  podle

$$\hat{\psi}(k, t) = \sqrt{1 - \frac{\|E\{\mathbf{I}(k, t)\}\|}{E\{\|\mathbf{I}(k, t)\|\}}}. \quad (1.1)$$





Obrázek 1.3: Průběh hodnot korelací

Symbol  $\mathbf{I}(k, t)$  značí vektor intenzity signálu,  $E\{\cdot\}$  značí časové průměrování,  $\|\cdot\|$  značí  $l^2$  normu,  $t$  značí čas a  $k = 2\pi f/c$ . Symbol  $f$  je frekvence a  $c$  je rychlost zvuku. Symbol  $\mathbf{I} = \frac{1}{2}\Re\{p^* \cdot \mathbf{v}\}$ , kde  $p$  značí akustický tlak a  $\mathbf{v}$  je vektor rychlosti částic. Poté byl spočten poměr mezi energií přímého signálu a difuzní složky  $\hat{\Gamma}$  podle

$$\hat{\Gamma}(t) = 10 \log_{10} \left( \frac{1}{\hat{\psi}(t)} - 1 \right). \quad (1.2)$$

Nakonec hranice mezi brzkými odrazy a difuzní částí byla získána v místě, kdy  $\hat{\Gamma}(t)$  poklesl pod určitou hranici.

Jelikož difuzní část je tak chaotická, až téměř náhodná, tak by měla být teoreticky nezávislá na části s přímou vlnou a brzkými odrazy. V mém řešení se pokusím akustickou impulsní odezvu rozdělit na brzké odrazy a difuzní část tak, aby konvoluce signálu s první částí byla nekorelovaná nebo nezávislá na konvoluci stejného signálu s první částí. A to tak, že budu posouvat hranici mezi potencionálními brzkými odrazy a difuzní částí a poté budu testovat korelačním koeficientem nebo vzájemnou informací závislost mezi vzniklými signály. Ideálně nezávisle na použitých testovacích signálech. Případně se pokusím vzniklé signály optimalizovat v časové oblasti před porovnáním nezávislosti pomocí vzájemné korelace nebo Wienerova filtru. Pro výpočet konvoluce s menší náročností na výpočetní výkon se pokusím implementovat metodu Overlap add. Nakonec navrženou metodiku ověřím v metodě dereverberace. Pokusím se vytvořit Wienerův filtr s nalezenou informací o hranici mezi brzkými odrazy a difuzní částí, tak aby po zfiltrování s reverbovaným signá-

lem s řečí tuto část odstranil. Výsledný dereverberovaný signál porovnáám s ideálně dereverberovaným signálem pomocí metody pro porovnání kvality řeči PESQ.

Veškeré experimenty budu provádět v programovém prostředí MATLAB.

## 2 Měření impulsních odezev

### 2.1 Tvorba nahrávek

Pro měření akustických impulsních odezev jsem vybral celkem 3 reálné místnosti, jejichž rozměry jsou uvedené v tabulce 2.1. Do každé místnosti jsem umístil reproduktor, jehož souřadnice v místnosti jsou zobrazeny v tabulce 2.2, kde  $w$  značí vzdálenost od delší stěny,  $l$  značí vzdálenost od kratší stěny a  $h$  značí výšku reproduktoru a mikrofonů. Před reproduktor jsem umístil referenční mikrofon  $s$  do vzdálenosti cca 1 cm. Do vzdálenosti  $d = 1, 2$  a  $3$  m jsem před reproduktor postupně umístil druhý mikrofon  $r$  podle obrázku 2.1. V místnosti 3 je  $d$  pouze 1 a 2 m. Použité klopové mikrofony měly kulovou směrovou charakteristiku.

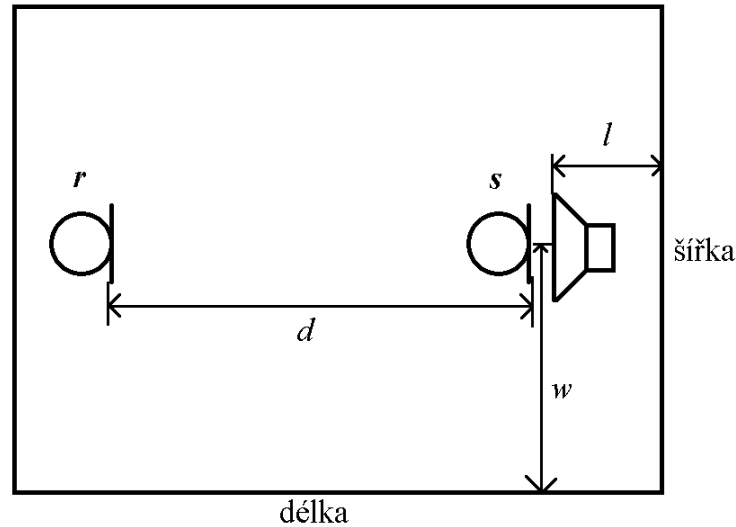
Tabulka 2.1: Rozměry místností

Místnost	Šířka [m]	Délka [m]	Výška [m]
Místnost 1	2,9	4,8	2,7
Místnost 2	4	4,4	2,7
Místnost 3	2,5	2,8	2,7

Použití referenčního mikrofonu a odhadnutí impulsní odezvy mezi ním a vzdáleným mikrofonem minimalizuje zanesení chyby do odhadu impulsní odezvy způsobenou zkreslením signálu použitým reproduktorem, než kdyby byla odezva odhadnuta mezi originálním pouštěným signálem a vzdáleným mikrofonem. Oba mikrofony mají stejné frekvenční charakteristiky, tudíž frekvenční zkreslení mikrofony se při výpočtu akustické impulsní odezvy vyruší. Stejně tak, pokud by reproduktor při některé

Tabulka 2.2: Umístění reproduktoru

Místnost	$w$ [m]	$l$ [m]	$h$ [m]
Místnost 1	1	0,8	0,9
Místnost 2	2,8	3,3	0,6
Místnost 3	1,4	0,2	0,6



Obrázek 2.1: Rozmístění reproduktoru a mikrofonů v místnosti

frekvenci rezonoval, projeví se to na obou mikrofonech a výsledek to také nijak neovlivní.

Poté jsem do reproduktoru pouštěl předem vygenerovaný bílý šum, sine sweep, mužský a ženský hlas. Bílý šum jsem použil, protože v každém čase rovnoměrně pokrývá celé frekvenční pásmo. Sine sweep pokrývá v čase vždy pouze jednu frekvenci a plynule stoupá od spodní hranice pásma až po horní hranici. Hlasy jsem použil pro porovnání, jelikož se používají v praktických aplikacích.

Šum jsem vygeneroval pomocí příkazu `wgn` se vzorkovací frekvencí 16 kHz a o délce 10 s.

Sine sweep jsem vygeneroval podle

$$x(t) = \sin \left[ T \cdot \omega_1 \cdot \left( e^{\frac{t}{T} \cdot \ln \left( \frac{\omega_2}{\omega_1} \right)} - 1 \right) / \ln \left( \frac{\omega_2}{\omega_1} \right) \right] \quad (2.1)$$

získaného z [4]. Symbol  $\omega_1$  značí počáteční úhlovou frekvenci podle vzorce  $\omega_1 = 2 \cdot \pi \cdot f_{poc}$ . Symbol  $\omega_2$  značí koncovou úhlovou frekvenci podle vzorce  $\omega_2 = 2 \cdot \pi \cdot f_{konc}$ . Symbol  $T$  značí celkovou délku signálu ve vteřinách. Výsledný signál je uložen do  $x$  v čase  $t$ . Jelikož vzorec 2.1 je definovaný pro spojitý signál, tak jsem ho upravil pro diskrétní signál na vzorec

$$x[t] = \sin \left[ T \cdot \omega_1 \cdot \left( e^{\frac{t \cdot (1/Fs) \cdot \ln(\frac{\omega_2}{\omega_1})}{T}} - 1 \right) / \ln \left( \frac{\omega_2}{\omega_1} \right) \right], \quad (2.2)$$

kde  $t$  značí aktuální vzorek a  $Fs$  vzorkovací frekvenci. Počáteční frekvenci  $f_{poc}$  jsem zvolil 10 Hz a koncovou frekvenci  $f_{konc}$  22 kHz. Vzorkovací frekvenci  $Fs$  jsem zvolil 44,1 kHz. Poté jsem vygenerovaný signál převzorkoval pomocí příkazu *resample* na 16 kHz. Délku signálu  $T$  jsem zvolil 10 s.

Mužský a ženský hlas jsem získal z databáze TIMIT. Nahrávky mužského a ženského hlasu jsou dlouhé 4 s a oba hlasy předčítají anglický text „she had your dark suit in greasy wash water all year“. Obě nahrávky mají vzorkovací frekvenci  $Fs$  16 kHz.

Pomocí rozmístěných mikrofónů jsem nahrál celkem 32 dvoukanálových nahrávek. Levý kanál obsahoval signál ze vzdáleného mikrofónu  $r$  a pravý kanál obsahoval signál z referenčního mikrofónu  $s$  u reproduktoru. Pro tvorbu záznamů jsem vytvořil skript, který přehrává postupně šum, sine sweep, mužský a ženský hlas. Zároveň při přehrávání skript nahrává signál z mikrofónů. Před jednotlivými signály je 5 s navíc nahráváno ticho pro stabilizaci AD (analogovo-digitálního) převodníku ve zvukové kartě. Před prvním přehrávaným signálem je 20 s prodleva pro opuštění místnosti, aby má přítomnost neovlivňovala tvar akustické impulsní odezvy. Z nahraného signálu toto ticho skript před uložením odstranil. Tvorba nahrávek najednou pomocí skriptu zajistila, že všechny druhy nahrávek pro jednu vzdálenost v konkrétní místnosti byly pořízeny za stejných podmínek s naprosto stejně rozmístěnými mikrofóny. Všechny záznamy byly pořízeny se vzorkovací frekvencí 16 kHz.

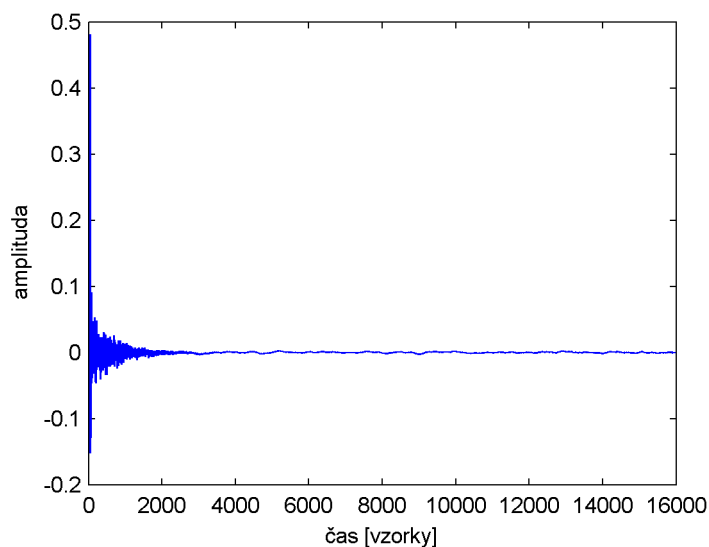
Příklad spektrogramu a signálu v časové oblasti pro přehrávaný šum a jeho nahrávek z obou mikrofónů pro  $d = 1$  m je v příloze B na obrázku B.1, pro  $d = 2$  a 3 m je na obrázku B.2, sine sweep pro  $d = 1$  m je na obrázku B.3, pro  $d = 2$  a 3 m je na obrázku B.4, ženský hlas pro  $d = 1$  m je na obrázku B.5, pro  $d = 2$  a 3 m je

na obrázku B.6, mužský hlas pro  $d = 1$  m je na obrázku B.7 a pro  $d = 2$  a  $3$  m je na obrázku B.8. Všechny příklady jsou pro místnost 1.

## 2.2 Odhad impulsních odezev

Akustické impulsní odezvy  $h$  mezi mikrofony  $\mathbf{s}$  a  $\mathbf{r}$  jsem odhadl pomocí skriptu, který jsem dostal od vedoucího práce a implementuje metodu TDRTF. Tento skript odhaduje relativní přenosovou funkci (impulsní odezvu) mezi dvěma signály pomocí metody nejmenších čtverců. Jako parametry přebírá délku odhadované impulsní odezvy  $N$ , dva signály  $xL$  a  $xR$ , mezi kterými odhaduje impulsní odezvu, globální zpoždění impulsní odezvy  $delay$  a regularizační parametr  $reg$ . Délku akustické impulsní odezvy  $N$  jsem zvolil 16 000 vzorků, tj. 1 s. Do parametru  $xL$  jsem vložil signál z mikrofonu  $\mathbf{s}$ . Do parametru  $xR$  jsem vložil signál z mikrofonu  $\mathbf{r}$ . Zpoždění  $delay$  jsem nastavil na 0 vzorků a regularizační parametr  $reg$  jsem nastavil na 0. Pro každou nahrávku jsem odhadl jednu akustickou impulsní odezvu. Vzniklo tedy celkem 32 akustických impulsních odezev ze všech místností a vzdáleností mezi mikrofony.

Na obrázku 2.2 je zobrazen příklad odhadnuté akustické impulsní odezvy z místnosti 1. Vzdálenost  $d$  mezi mikrofony  $\mathbf{s}$  a  $\mathbf{r}$  byl 1 m a pouštěný signál z reproduktoru byl bílý šum.



Obrázek 2.2: Příklad odhadnuté akustické impulsní odezvy

## 2.3 Ověření vlastností

Po odhadnutí všech akustických impulsních odezev jsem ověřil jejich předpokládané vlastnosti.

Nejdříve jsem ověřil vlastnost, že čím větší je vzdálenost mezi mikrofony, tím déle by mělo přímé zvukové vlně trvat tuto vzdálenost překonat. Na impulsní odezvě tuto vlastnost zobrazuje vzorek s nejvyšší hodnotou. V tabulce 2.3 je zobrazena předpokládaná doba, kdy zvuková vlna překoná vzdálenost 1, 2 a 3 m. Dále na kolikátém vzorku AIR (akustické impulsní odezvy) by se mělo nacházet maximum, reprezentující časový úsek, kdy by měla přímá vlna z reproduktoru dosáhnout vzdáleného mikrofonu  $r$ . V posledních dvou sloupcích jsou hodnoty odečtené z odhadnuté AIR. Teoretické hodnoty odpovídají rychlosti zvuku 343 m/s. Drobné odchylky uvedené v tabulce odpovídají nepřesnostem při rozmístění mikrofonů a reálně mírně odlišné rychlosti zvuku.

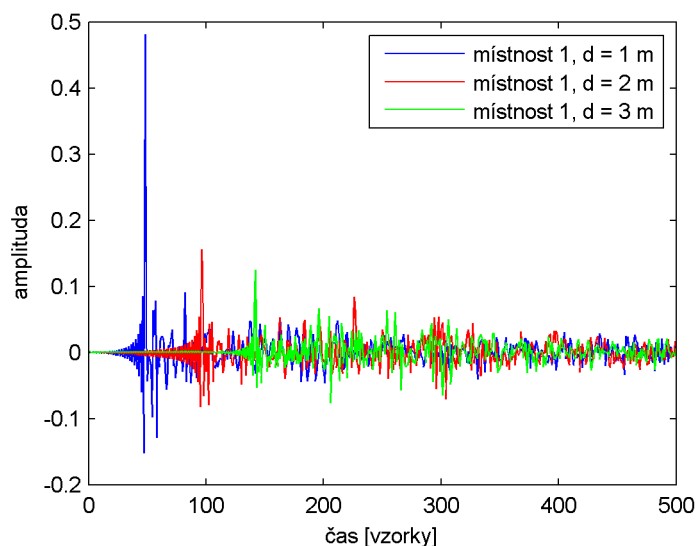
Na obrázku 2.3 jsou zobrazené odhadnuté akustické impulsní odezvy z místnosti 1 se vzdáleností mezi mikrofony 1, 2 a 3 m. Přehrávaný signál do reproduktoru byl bílý šum. Na obrázku je zobrazený úsek do 3 000. vzorku, aby byly patrné rozdíly mezi maximy. Maxima udávají časovou vzdálenost přímé vlny, která se šířila

Tabulka 2.3: Předpokládaná a naměřená doba šíření přímé vlny mezi mikrofony

vzdálenost mezi $s$ a $r$ [m]	teoretická doba [ms]	předpokládaný počet vzorků	naměřená doba z AIR [ms]	naměřený počet vzorků z AIR
1	2,92	47	3	48
2	5,83	93	6	96
3	8,75	140	8,88	142

od referenčního mikrofону ke vzdálenému. Ze zobrazených impulsních odezev jsou odečteny hodnoty uvedené v tabulce 2.3. Z obrázku je patrný pravidelný posun globálních maxim jednotlivých akustických impulsních odezev, odpovídající 1, 2 a 3 m, které musí přímá zvuková vlna překonat.

Na obrázku lze též pozorovat, že globální maxima jednotlivých akustických impulsních odezev s přibývajícím vzdáleností klesají. Odpovídá to postupnému snižování energie přímé zvukové vlny s přibývajícím vzdáleností.

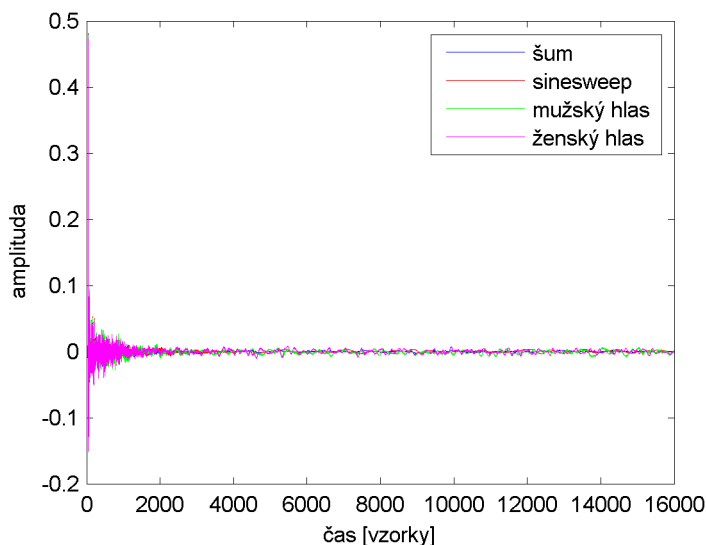


Obrázek 2.3: Porovnání AIR s různými vzdálenostmi mikrofónů

Poté jsem ověřil vlastnost, že akustické impulsní odezvy odhadnuté ve stejných podmínkách při různých reprodukováných signálech by měly být teoreticky podobné.



Na obrázku 2.4 jsou zobrazené akustické impulsní odezvy odhadnuté při vzdálenosti mikrofónů 1 m ze signálů s bílým šumem, sine sweepem, mužským a ženským hlasem. Na obrázku 2.5 je zobrazen detail akustických impulsních odezev. Na detailu mezi 170. a 270. vzorkem je patrné, že všechny čtyři akustické impulsní odezvy jsou téměř totožné. Na obrázku 2.6 je zobrazen graf průměrné kvadratické chyby mezi stejnými vzorky všech čtyř předešlých akustických impulsních odezev.

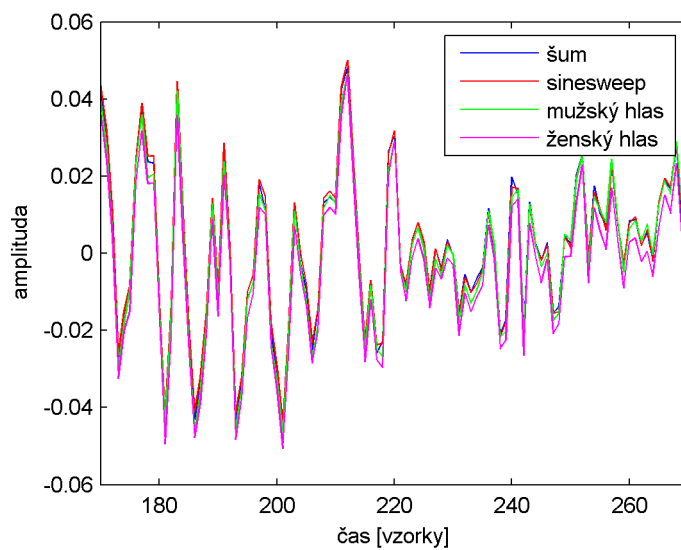


Obrázek 2.4: Porovnání AIR podle typu přehrávaného zvuku ve stejných podmínkách

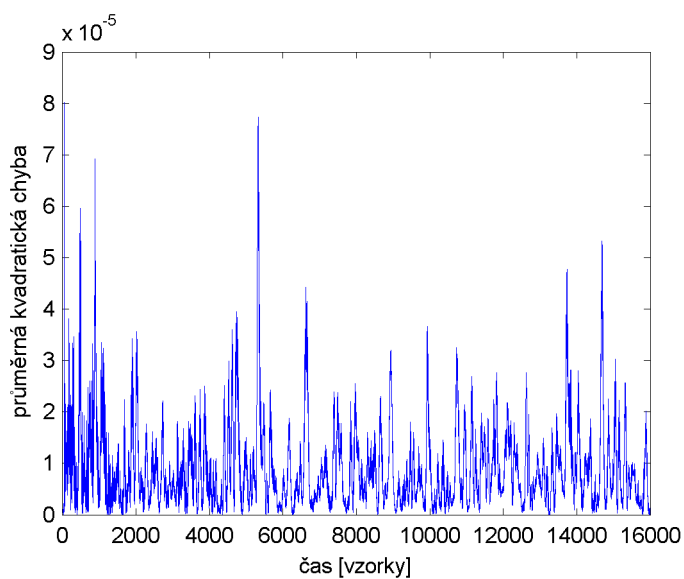
Nakonec jsem ověřil vlastnost, že pokud se provede konvoluce signálu z referenčního mikrofónu  $\mathbf{s}$  s příslušnou akustickou impulsní odezvou  $\mathbf{h}$ , vznikne signál  $\hat{\mathbf{r}}$  podle

$$\hat{\mathbf{r}} = \mathbf{s} * \mathbf{h}. \quad (2.3)$$

Ideálně rozdíl mezi  $\mathbf{r}$  a  $\hat{\mathbf{r}}$  je nulový. V reálném prostředí se však liší o šum vzniklý na mikrofonech, AD převodnících a vlivem rozdílu délek mezi odhadnutou a skutečnou akustickou impulsní odezvou, která je teoreticky nekonečná. Mírně se také liší o část zvukového záznamu v nahrávce vlivem nelinearit mikrofónů a o odrazy, které dopadly na mikrofón  $\mathbf{r}$  déle než za 1 s od reprodukce reproduktorem vlivem zvolené konečné délky akustické impulsní odezvy 1 s. Ve zvolených místnostech je dozvuk podle naměřených AIR mnohem kratší, tudíž by se tato odlišnost měla projevit

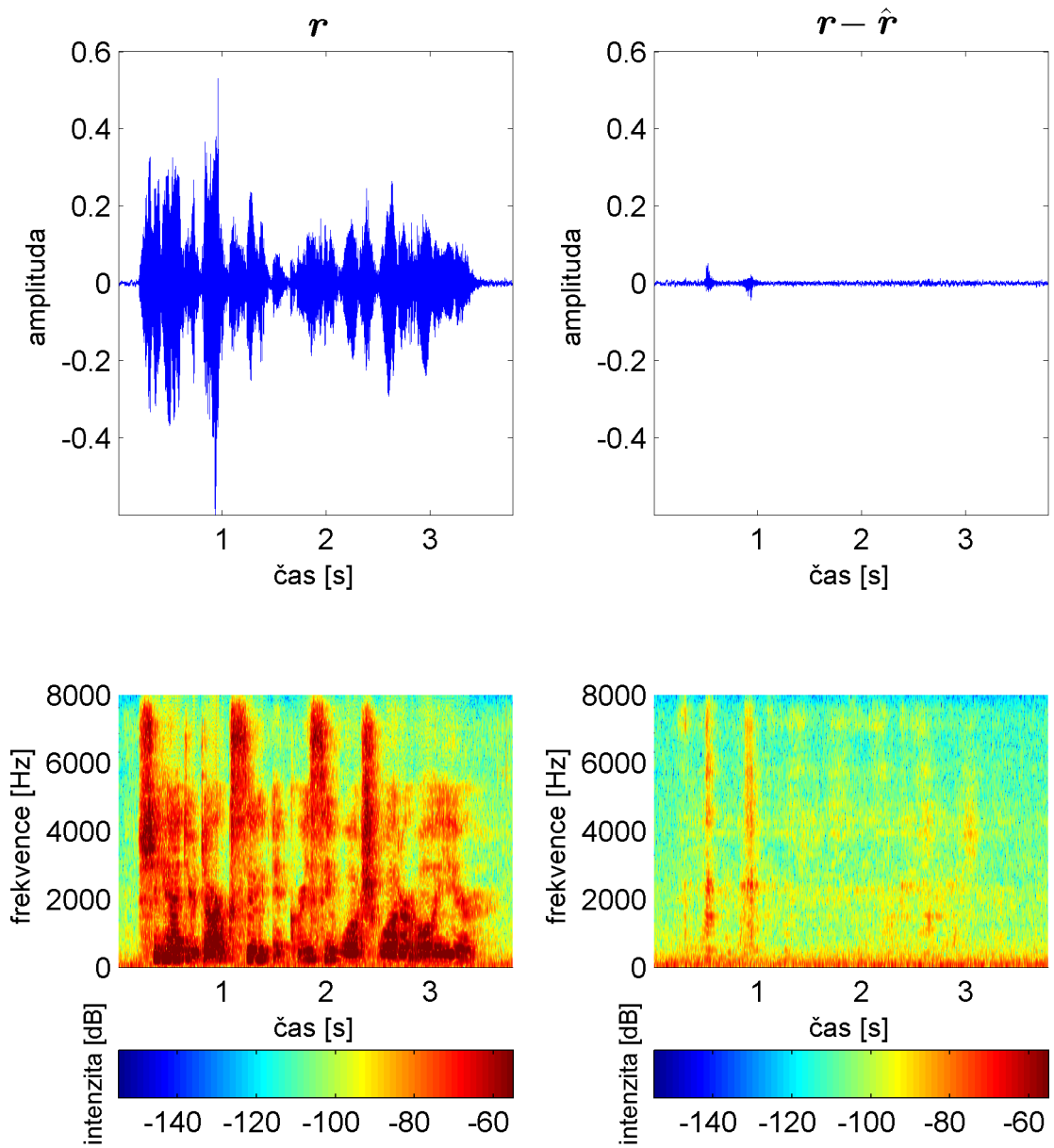


Obrázek 2.5: Porovnání detailu AIR podle typu přehrávaného zvuku ve stejných podmínkách



Obrázek 2.6: Průměrná kvadratická chyba mezi AIR s různými typy přehrávaných signálů

jen minimálně. Na obrázku 2.7 je zobrazen spektrogram a časový průběh signálu  $\mathbf{r}$  a  $\mathbf{r} - \hat{\mathbf{r}}$  pro nahrávku ženského hlasu se vzdáleností mikrofону  $\mathbf{r}$  1 m v místnosti 1. Na uvedeném obrázku jsou zřejmé předešlé předpoklady.



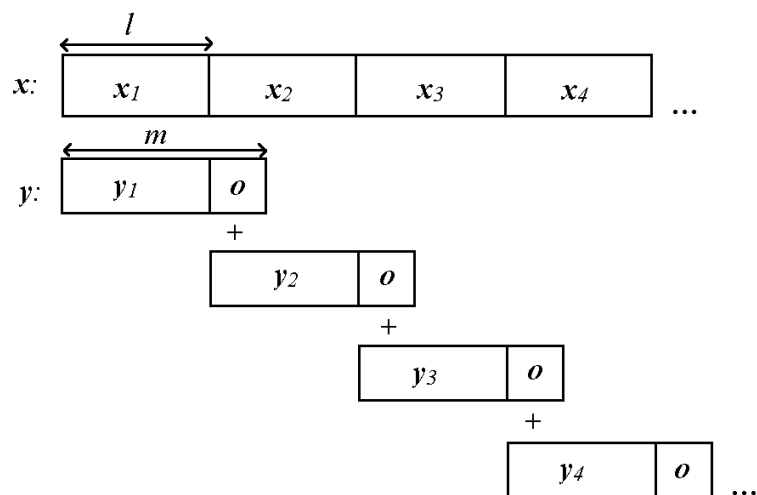
Obrázek 2.7: Spektrogram signálu  $r$  a  $r - \hat{r}$

## 3 Použité algoritmy

### 3.1 Overlap Add

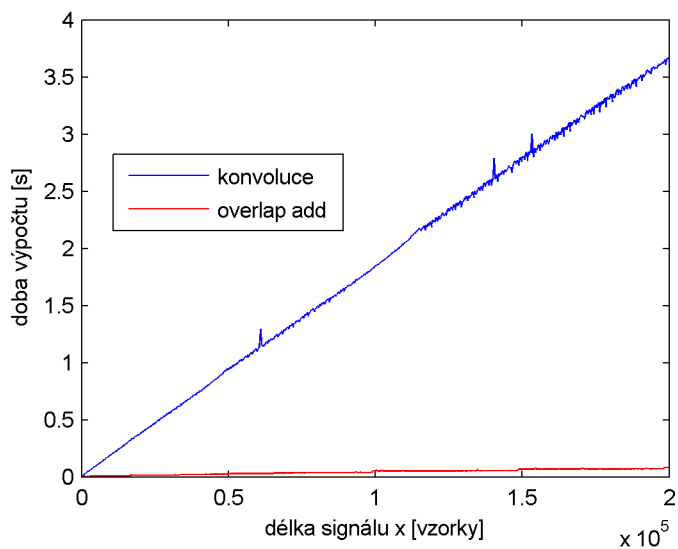
Pro výpočet konvoluce jsem vytvořil skript, který počítá konvoluci metodou Overlap add. Tuto metodu jsem implementoval podle [5]. Na obrázku 3.1 je zobrazen princip funkce metody Overlap add. Vstupem funkce je posloupnost  $\mathbf{x}$  a filtr  $\mathbf{h}$ . Funkce předpokládá, že posloupnost  $\mathbf{x}$  je delší než  $\mathbf{h}$ . V obráceném případě funkce oba parametry prohodí. Dále nastaví velikost bloku  $l$  takovou, aby součet  $l$  a velikosti přesahu  $\mathbf{o}$ , tedy  $m$ , byl druhou nejbližší vyšší mocninou čísla dvě než je délka filtru  $\mathbf{h}$ . Velikost bloku  $\mathbf{o}$  odpovídá délce filtru  $\mathbf{h} - 1$ . Funkce cyklicky vyjme blok  $x_n$  a doplní ho nulami na velikost  $m$ . Poté vezme filtr  $\mathbf{h}$  a také ho doplní nulami na velikost  $m$ . Dále obě posloupnosti skript převede do frekvenční oblasti pomocí FFT (Rychlé Fourierovy transformace) a po prvcích je vynásobí. Jejich součin je poté zpět převeden do časové oblasti pomocí zpětné FFT. K části  $\mathbf{o}$  v předchozím bloku je pak přičtena přední část nového bloku a zbytek bloku je pouze zkopírován za tuto část. Výsledná konvoluce má délku  $\mathbf{x} + \mathbf{h} - 1$ . Pokud je poslední blok  $\mathbf{x}_N$  kratší než  $l$ , tak je doplněn nulami na délku  $m$ . Po přičtení k předposlednímu bloku  $\mathbf{y}_{N-l}$  je pak výsledný signál oříznut na délku  $\mathbf{x} + \mathbf{h} - 1$ .

Metodu Overlap add jsem upřednostnil před standardní konvolucí pomocí příkazu `conv`, protože je efektivnější při použití dlouhých posloupností  $\mathbf{x}$  a  $\mathbf{h}$ . Na obrázku 3.2 je zobrazen graf doby výpočtu pomocí příkazu `conv` modře a červeně je doba výpočtu konvoluce pomocí metody Overlap add. Na ose  $x$  je délka vstupního signálu  $\mathbf{x}$ . Na ose  $y$  je doba výpočtu algoritmů v sekundách. Filtr  $\mathbf{h}$  jsem pro všechny výpočty použil akustickou impulsní odezvu odhadnutou z nahrávky bílého šumu v místnos-



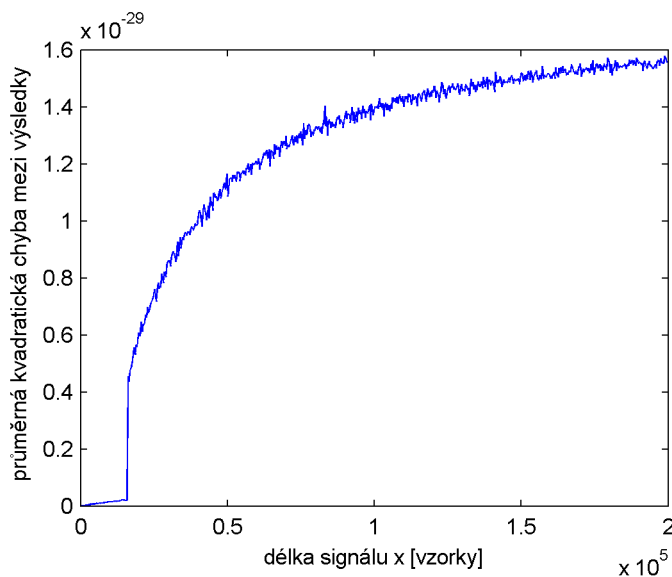
Obrázek 3.1: Princip Overlap add

ti 1 se vzdáleností mikrofonu  $r$  1 m. Porovnání jsem provedl pro délky signálu  $x$  od 300 do 199 800 vzorků. Signál  $x$  obsahoval vždy náhodně vygenerovaná čísla pomocí příkazu *randn*. Z uvedeného obrázku je patrné, že metoda Overlap add je pro výpočet konvoluce s dlouhými signály velmi efektivní. Například pro signál  $x$  o délce 160 000 je výpočet konvoluce pomocí metody Overlap add 40× rychlejší než pomocí příkazu *conv*.



Obrázek 3.2: Porovnání rychlosti Overlap add a konvoluce

Zároveň s porovnáním rychlosti obou algoritmů jsem ověřil i odchylku výsledků. Na obrázku 3.3 je zobrazen průběh průměrné kvadratické chyby výsledků po výpočtu konvoluce pomocí metody Overlap add a příkazu *conv*. Odchylka je spočtena jako druhá mocnina průměru rozdílů výstupních signálů. Délky a obsahy signálů  $\mathbf{x}$  a  $\mathbf{h}$  byl shodný s předešlým experimentem. Z uvedeného obrázku vyplývá, že průměr druhé mocniny rozdílů signálů roste logaritmicky s délkou signálu  $\mathbf{x}$ . Jeho velikost se nachází v řádu  $10^{-29}$ , což je zanedbatelný rozdíl.



Obrázek 3.3: Porovnání odchylky výsledků konvoluce a Overlap add

## 3.2 Vzájemná korelace

Pro zjištění vzájemného zpoždění dvou signálů jsem použil vzájemnou korelaci. Výpočet vzájemné korelace se provádí podle

$$\hat{R}_{x,y}[k] = \frac{1}{N} \sum_{n=0}^{N-k-1} x[n] y[n+k]. \quad (3.1)$$

Pro jeho výpočet jsem použil příkaz *xcorr*. Výstupem funkce je sekvence hodnot s délkou o jedna menší než součet délky obou vložených signálů. Symbol  $k$  označuje o kolik prvků je signál  $x$  posunutý proti  $y$  a může nabývat hodnot od  $-N$  do  $N$ .

Symbol  $N$  označuje délku delšího signálu. Výsledný průběh značí vzájemnou podobnost signálů pro různé vzájemné časové posunutí. Pokud je spočtena vzájemná korelace mezi identickými bílými šумы, které jsou od sebe posunuté v čase o  $l$  vzorků, pak je výsledná vzájemná korelace po celé délce přibližně nulová, kromě prvku, který je od středu vzdálen o  $l$ .

### 3.3 Wienerův filtr

Wienerův filtr funguje tak, že se snaží uchovat informaci o spektrálním rozložení nezávislé stacionární informace v signálu (například šumu) tak, že minimalizuje chybu mezi odfiltrovaným zarušeným signálem a originálním nezarušeným signálem pomocí metody nejmenších čtverců. Pro výpočet Wienerova filtru jsem použil skript, který je implementovaný podle [6].

### 3.4 Korelační koeficient

Pro porovnání závislosti dvou signálů jsem použil korelační koeficient, který programové prostředí MATLAB spočítá pomocí příkazu *corrcoef*. Korelační koeficient dvou libovolných stejně dlouhých signálů značí velikost jejich lineární závislosti. Korelační koeficient může nabývat hodnoty od  $-1$  do  $1$ . Pokud je hodnota  $\pm 1$ , pak jsou podle [7] lineárně závislé a nesou identickou informaci. Jestliže je hodnota nulová, pak jsou signály nekorelované (jsou si nejméně podobné). Neznamena to však, že signály nenesou žádnou společnou informaci. Pokud každá proměnná má  $N$  skalárních prvků, pak Pearsonův korelační koeficient je definován podle

$$\rho(x, y) = \frac{1}{N-1} \sum_{i=1}^N \left( \frac{\overline{x_i - \mu_x}}{\sigma_x} \right) \left( \frac{y_i - \mu_y}{\sigma_y} \right). \quad (3.2)$$

Proměnná  $\mu_x$  značí střední hodnotu signálu  $x$  (aritmetický průměr) a  $\sigma_x$  značí směrodatnou odchylku  $x$ . Ekvivalentně  $\mu_y$  a  $\sigma_y$  jsou střední hodnota a směrodatná odchylka  $y$ . Opruhování značí komplexní sdružení pokud jsou signály v  $x$  a  $y$  komplexní čísla.

Směrodatná odchylka určuje jak moc jsou hodnoty rozptýleny, či odchýleny od průměru hodnot. Směrodatná odchylka se spočítá podle

$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}, \quad (3.3)$$

kde hodnota  $\bar{x}$  značí průměrnou hodnotu signálu  $x$ . Symbol  $N$  pak značí počet prvků v signálu  $x$ .

Dále podle [8] lze korelační koeficient alternativně definovat na základě kovariance mezi signály  $x$  a  $y$  podle

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}, \quad (3.4)$$

jelikož kovariance se podle [7] spočte podle

$$\text{cov}(x, y) = \hat{C}_{xy} = \frac{1}{N} \sum_{n=1}^N (x[n] - \hat{\mu}_x) \overline{(y[n] - \hat{\mu}_y)}. \quad (3.5)$$

Výstup příkazu *corrcoef* je matice korelačních koeficientů pro každou kombinaci signálu  $x$  a  $y$ , viz vzorec 3.6.

$$R = \begin{pmatrix} \rho(x, x) & \rho(x, y) \\ \rho(y, x) & \rho(y, y) \end{pmatrix} \quad (3.6)$$

Jelikož  $x$  a  $y$  jsou vždy sami se sebou korelované, tak na diagonále je vždy hodnota 1, viz vzorec 3.7. Pro získání korelačního koeficientu jsem tedy vždy vybral hodnotu z prvního řádku a druhého sloupce.

$$R = \begin{pmatrix} 1 & \rho(x, y) \\ \rho(y, x) & 1 \end{pmatrix} \quad (3.7)$$

### 3.5 Vzájemná informace

Jak jsem uvedl v předchozí podkapitole, korelační koeficient má tu vlastnost, že pokud vyjde nulový, jsou signály nekorelované, ale nemusí být nezávislé. Použil jsem tedy ještě jednu metodu pro porovnání podobnosti dvou signálů a to vzájemnou



informaci. Vzájemná informace vyjde nulová pouze pokud jsou signály navzájem nezávislé, jinak je hodnota kladná. Množství informace je udáváno v bitech. Výpočet vzájemné informace se provede podle vzorce 3.8, který je z [9].

$$I(X : Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (3.8)$$

Funkce  $p(x, y)$  značí sdruženou distribuční funkci a  $p(x)p(y)$  jsou marginální distribuční funkce. Pro výpočet vzájemné informace jsem použil skript, který je implementovaný podle [10].

## 3.6 PESQ

Ohodnocení vnímané kvality řeči PESQ (Perceptual Evaluation of Speech Quality) je standard pro objektivní porovnávání kvality hlasu. Algoritmus porovnává referenční hlasový signál se zarušeným signálem a zarušený signál poté ohodnocuje čísly 1–4 podle tabulky 3.1. Pro výpočet hodnoty PESQ jsem použil skript, který je implementovaný podle [11].

Tabulka 3.1: Tabulka ohodnocení kvality řeči

Kvalita řeči	Ohodnocení PESQ
Špatná	1
Podprůměrná	2
Průměrná	3
Nadprůměrná	4
Perfektní	5

## 4 Rozdělení akustické impulsní odezvy

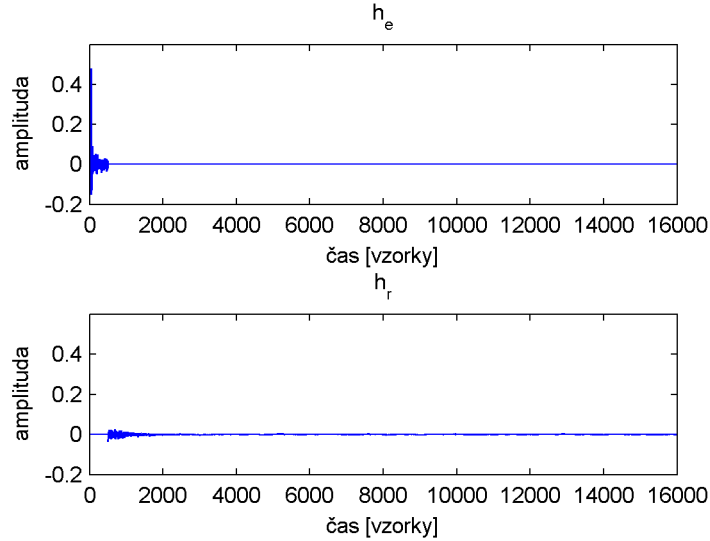
### 4.1 Bez synchronizace mezi $s_e$ a $s_r$

Pro rozdělení akustické impulsní odezvy na brzké odrazy a difuzní část tak, aby konvoluce signálu s první částí byla nezávislá nebo jen nekorelovaná na konvoluci stejného signálu s druhou částí jsem vytvořil následující skript. Rozdělil jsem akustickou impulsní odezvu  $\mathbf{h}$  na část s přímou vlnou a s potenciálními brzkými odrazy  $\mathbf{h}_e$  a potenciálně difuzní část  $\mathbf{h}_r$  s hranicí  $b$ . Akustická impulsní odezva s potenciálními brzkými odrazy a přímou vlnou  $\mathbf{h}_e$  obsahuje na prvcích od počátku až do hranice  $b$  (včetně) stejné hodnoty jako originální akustická impulsní odezva  $\mathbf{h}$ . Od prvku na indexu  $b + 1$  obsahuje  $\mathbf{h}_e$  nulové hodnoty. Akustická impulsní odezva s potenciálně difuzní částí  $\mathbf{h}_r$  obsahuje prvky od počátku do hranice  $b$  (včetně) nulové hodnoty a od prvku na indexu  $b + 1$  obsahuje stejné prvky jako originální akustická impulsní odezva  $\mathbf{h}$ . Názorný příklad rozdělené akustické impulsní odezvy je na obrázku 4.1. Uvedená akustická impulsní odezva je odhadnutá z místnosti 1 se vzdáleností mezi mikrofony 1 m a s reprodukováným bílým šumem. Hranice  $b$  jsem pro ukázkou zvolil 500.

Dále jsem vytvořil signál s potenciálními brzkými odrazy  $\mathbf{s}_e$ , který vznikl konvolucí signálu z referenčního mikrofону  $\mathbf{s}$  s akustickou impulsní odezvou  $\mathbf{h}_e$  popisující potenciálně brzké odrazy a přímou vlnu podle

$$\mathbf{s}_e = \mathbf{h}_e * \mathbf{s}. \quad (4.1)$$

Konvoluci jsem spočítal pomocí metody Overlap add. Poté jsem vytvořil signál  $\mathbf{s}_r$ , který obsahuje potenciální mnohonásobné odrazy. Tento signál lze analogicky



Obrázek 4.1: Rozdělení  $\mathbf{h}$  na potenciální  $\mathbf{h}_e$  a  $\mathbf{h}_r$

získat konvolucí  $\mathbf{h}_r$  s  $\mathbf{s}$  podle

$$\hat{\mathbf{s}}_r = \mathbf{h}_r * \mathbf{s}. \quad (4.2)$$

Signál  $\mathbf{s}_r$  lze získat také odečtením signálu  $\mathbf{s}_e$  od vzdáleného mikrofonu  $\mathbf{r}$  podle

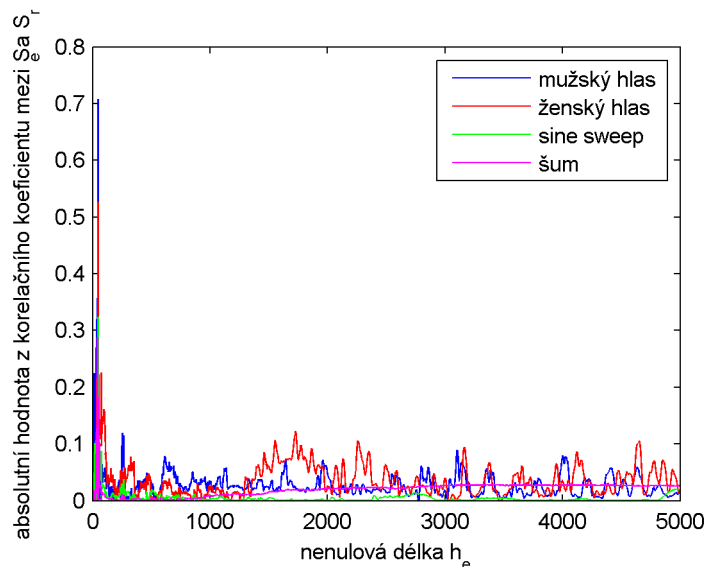
$$\mathbf{s}_r = \mathbf{r} - \mathbf{s}_e. \quad (4.3)$$

Pro výpočet  $\mathbf{s}_r$  jsem raději zvolil druhou možnost, jelikož není třeba vytvářet  $\mathbf{h}_r$  a výpočetně méně náročnější je odečítání, než počítání konvoluce. Signál  $\mathbf{s}_r$  se také od  $\hat{\mathbf{s}}_r$  liší o šum, odrazy, které jsou vzdálenější než je délka celé odhadnuté akustické impulsní odezvy a o nelinearity mezi mikrofony.

Následně jsem vytvořil skript, který postupně posouvá hranici  $b$  mezi  $\mathbf{h}_e$  a  $\mathbf{h}_r$  od počátku do konce akustické impulsní odezvy  $\mathbf{h}$  po jednom prvku. Pro každou hranici  $b$  jsem změřil korelační koeficient a vzájemnou informaci mezi vzniklým  $\mathbf{s}_e$  a  $\mathbf{s}_r$ . Výsledkem bylo 16 000 hodnot pro každou impulsní odezvu v každé místnosti pro každý druh nahraného signálu, v každé místnosti se všemi vzdálenostmi mikrofonu. Pro každou akustickou impulsní odezvu jsem vytvářel signály  $\mathbf{s}_e$  a  $\mathbf{s}_r$ , vždy z originálních záznamů, ze kterých byla odhadnuta konkrétní akustická impulsní odezva.

Na obrázku 4.2 je zobrazen detail čtveřice průběhů korelačního koeficientu mezi  $s_e$  a  $s_r$  pro rozsah hranice  $b$  od 1 do 5 000. Růžově je zobrazen průběh pro šum, zeleně pro sine sweep, modře pro mužský hlas a červeně pro ženský hlas. Průběhy jsou zobrazeny pro nahrávky z místnosti 2 se vzdáleností vzdáleného mikrofonu  $r$  od referenčního mikrofonu  $s$  1 m. Podle uvedeného obrázku jsou průběhy téměř nekorelované prakticky okamžitě pro hodnotu hranice  $b$  větší než je pozice v akustické impulsní odezvě s maximem značící přímou vlnu. Po této hranici průběhy pro mužský a ženský hlas chaoticky kolísají mezi hodnotami 0 a 0,1. Průběh pro šum po této hranici postupně stoupá až k hranici  $b = 3 000$ . Poté je hodnota korelačního koeficientu přibližně konstantní na hodnotě 0,02 až do konce. Hodnoty průběhu reprezentující sine sweep poté nabývají hodnot blízkých k nule.

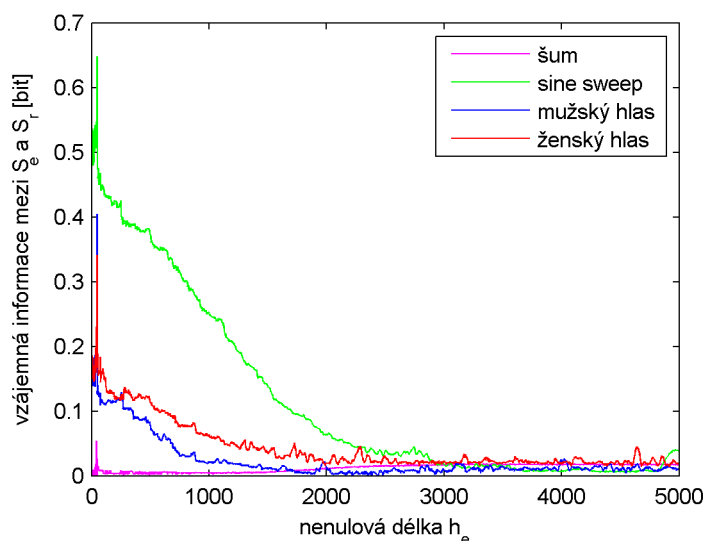
Závěr tohoto experimentu je, že z těchto průběhů není možné vyčíst žádnou informaci o hranici  $b$ , která by určovala, že po této hodnotě by byly signály  $s_e$  a  $s_r$  navzájem nekorelované.



Obrázek 4.2: Porovnání průběhů korelačního koeficientu

Na obrázku 4.3 je zobrazen detail čtveřice průběhů vzájemné informace mezi signály  $s_e$  a  $s_r$  stejných jako v předchozím případě pro hodnotu  $b$  od 1 do 5 000. Oproti průběhům korelačního koeficientu je z obrázku patrný pozvolný pokles vzájemné informace se zvyšující se hodnotou hranice  $b$  odpovídající počtu nenulových

prvků v  $\mathbf{h}_e$  následně po hodnotě  $h$  odpovídající pozici s maximem v akustické impulsní odezvě. Průběh se sine sweepem je nejstrmější ze všech průběhů. Jeho pokles končí zhruba od hodnoty  $b = 3\,000$  na hodnotě vzájemné informace zhruba  $0,3$  b. Průběhy pro mužský a ženský hlas mají podobný tvar. Jejich postupný pokles končí zhruba na hodnotě  $b = 2\,000$ . Poté kolísají podobně jako průběh pro sine sweep kolem hodnoty vzájemné informace zhruba  $0,3$  b. Průběh pro bílý šum je prakticky po celé délce téměř nulový.



Obrázek 4.3: Porovnání průběhů vzájemné informace

Z výsledku uvedeného experimentu se vzájemnou informací je patrné, že průběh vzájemné informace je silně závislý na druhu nahraného signálu. Jak strmější, tak míšnější, kde pokles končí. Tudíž z grafu nelze spolehlivě vyčíst hranici mezi brzkými odrazy a difuzní částí. Z obou výsledků tedy vyplývá, že takto nelze získat spolehlivou informaci o této hranici, jelikož pro každý měřený signál by vyšla jiná hodnota.

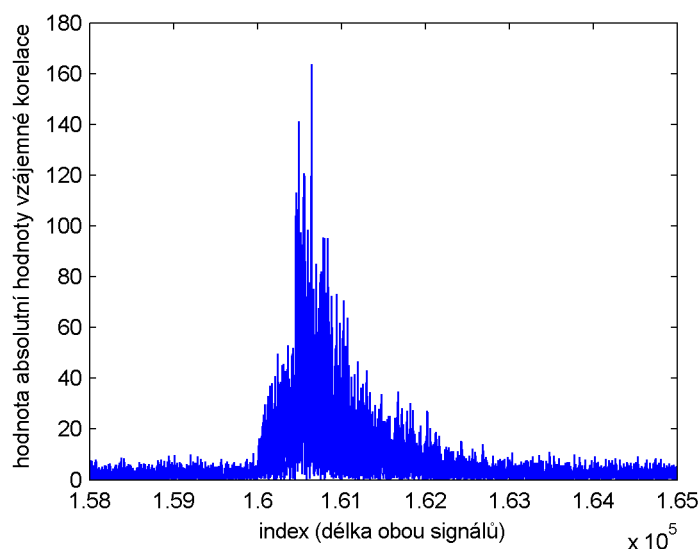
Korelační koeficient, ani vzájemná informace neberou v potaz, pokud jsou porovnávané signály posunuty v čase. Například pokud spočteme korelační koeficient nebo vzájemnou informaci mezi identickými bílými šumy, posunuté mezi sebou byť o jediný vzorek, pak hodnoty vyjdou nulové. Signál  $\mathbf{s}_r$  je od signálu  $\mathbf{s}_e$  časově posu-

nutý minimálně o délku nenulové části  $\mathbf{h}_e$ , tedy o velikost  $b$ . Z tohoto důvodu jsem se v dalších experimentech pokusil signály  $\mathbf{s}_e$  a  $\mathbf{s}_r$  mezi sebou časově synchronizovat.

## 4.2 Synchronizace $\mathbf{s}_e$ a $\mathbf{s}_r$ pomocí vzájemné korelace

V tomto experimentu jsem zopakoval stejný postup jako v předchozím experimentu až na časovou synchronizaci mezi signály  $\mathbf{s}_e$  a  $\mathbf{s}_r$ . Mezi oběma signály jsem spočetl vzájemnou korelaci. Poté jsem našel její maximum a zjistil o kolik prvků vpravo se toto maximum nalézá od středu průběhu vzájemné korelace. O tento počet prvků jsem pak posunul signál  $\mathbf{s}_r$  zpět.

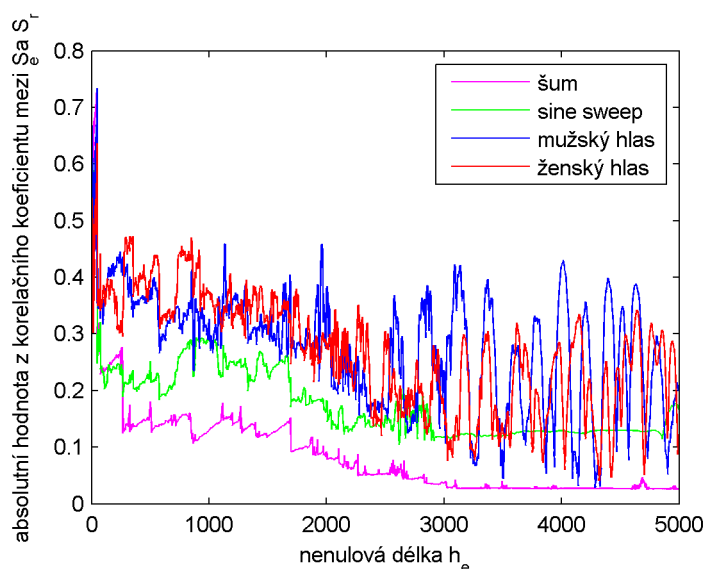
Na obrázku 4.4 je zobrazen detail vzájemné korelace mezi prvky 15 800 a 16 500. Celková délka vzájemné korelace je rovna součtu délek signálů  $\mathbf{s}_e$  a  $\mathbf{s}_r - 1$  tedy 31 999 pro délku  $\mathbf{s}_e$  a  $\mathbf{s}_r = 10$  s. Obrázek byl vygenerován ze signálů  $\mathbf{s}_e$  a  $\mathbf{s}_r$ , vzniklých podle vzorce 4.1 a 4.3, s impulsní odezvou  $\mathbf{h}$  odhadnutou z nahrávky šumu v místnosti 1 se vzdáleností mezi mikrofony 1 m. Hranici  $b$  jsem pro ukázkou zvolil 500. Z obrázku je patrné, že se maximum nalézá na 16 600. vzorku. Po odečtení délky jednoho signálu (16 000) zůstane číslo 600, které odpovídá časovému posunutí signálu  $\mathbf{s}_r$  dále v čase oproti signálu  $\mathbf{s}_e$ .



Obrázek 4.4: Příklad detailu vzájemné korelace

Následně jsem opět spočetl průběh korelačního koeficientu a vzájemné informace mezi  $s_e$  a  $s_r$  pro velikost hranice  $b$  od 1 do 16 000. Pro každou velikost  $b$  jsem  $s_e$  a  $s_r$  časově synchronizoval pomocí vzájemné korelace.

Na obrázku 4.5 je zobrazen výsledný průběh korelačního koeficientu pro stejné signály jako v předchozích experimentech. Na všech průbězích je patrný průběžný pokles, avšak průběhy pro mužský a ženský hlas jsou stále chaoticky probíhající. Průběhy pro bílý šum a sine sweep končí svůj pokles okolo velikosti  $d = 3\ 000$ . Sine sweep dále zůstává téměř konstantní okolo hodnoty velikosti korelačního koeficientu 1,5 a bílý šum okolo hodnoty 0,3. Z uvedeného obrázku je patrné mírné zlepšení oproti průběhům bez synchronizace, avšak jsou stále závislé na druhu nahraného signálu.

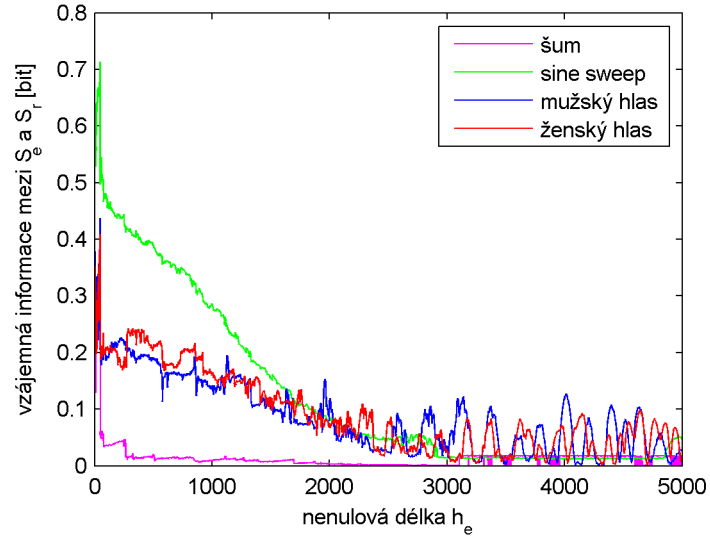


Obrázek 4.5: Porovnání průběhů korelačního koeficientu

Na obrázku 4.6 je zobrazen průběh vzájemných informací. Na rozdíl od průběhů korelačních koeficientů jsou průběhy pro mužský a ženský hlas méně chaotické. Výsledné strmosti průběhů jsou však závislejší na druhu signálu než v případě korelačních koeficientů. Průběh pro bílý šum je prakticky od začátku na téměř nulových hodnotách.

Podle výsledků tohoto experimentu je metoda synchronizace v čase mezi signály  $s_e$  a  $s_r$  málo dostatečná, protože každý odraz v místnosti od každé překážky přichází

na vzdálený mikrofon  $\mathbf{r}$  s různým časovým zpožděním. Proto jsem v následujícím experimentu zvolil metodu synchronizace pomocí Wienerova filtru.



Obrázek 4.6: Porovnání průběhů vzájemné informace

### 4.3 Synchronizace $\mathbf{s}_e$ a $\mathbf{s}_r$ pomocí Wienerova filtru

V tomto experimentu jsem nahradil časovou synchronizaci Wienerovým filtrem místo pouhého posunutí mezi signály  $\mathbf{s}_e$  a  $\mathbf{s}_r$  o hodnotu danou vzájemnou korelací. Mezi signály  $\mathbf{s}_e$  a  $\mathbf{s}_r$  jsem odhadl Wienerův filtr  $\mathbf{w}$  délky 5 000. Konvoluce  $\mathbf{s}_e$  s filtrem  $\mathbf{w}$  vytvoří signál  $\hat{\mathbf{s}}_r$ , který se podobá signálu  $\mathbf{s}_r$ , viz vzorec 4.4.

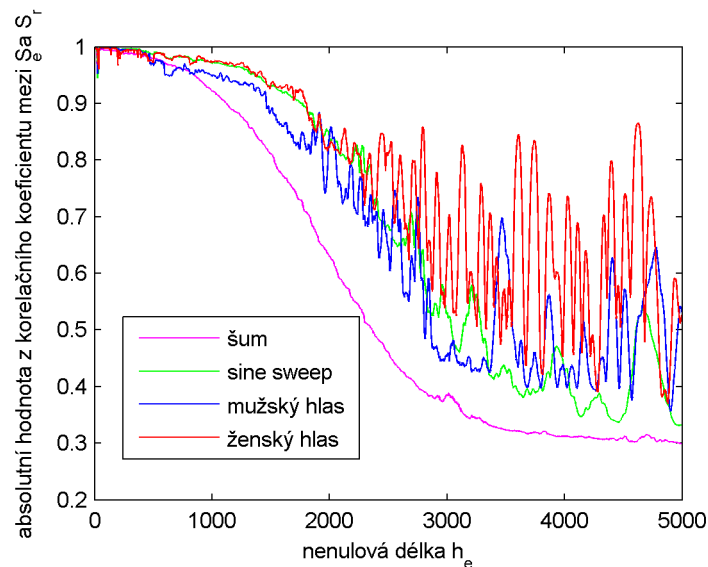
$$\hat{\mathbf{s}}_r = \mathbf{w} * \mathbf{s}_e \quad (4.4)$$

Filtr  $\mathbf{w}$  o délce 5 000 časově přizpůsobí odrazy vzdálené od reprodukce maximálně o 5 000 vzorků. Průběh korelačního koeficientu a vzájemné informace mezi  $\mathbf{s}_e$  a  $\mathbf{s}_r$  jsem tedy spočítal pro rozsah  $b$  mezi 1 a 5 000. Pro větší  $b$  by byly výsledky nerelevantní a delší filtr  $\mathbf{w}$  by neúměrně zvýšil výpočetní náročnost pro výpočet průběhů až na několik dní. Pro další úsporu výpočetního výkonu jsem 10s záznam bílého šumu zkrátil na 2 s. Ze záznamu sine sweepu jsem ponechal část, která je



vybuzená reprodukováným signálem, tj. 200–7 500 Hz s délkou 4,87 s. Z mužského a ženského hlasu jsem vybral počáteční 2 s záznamu.

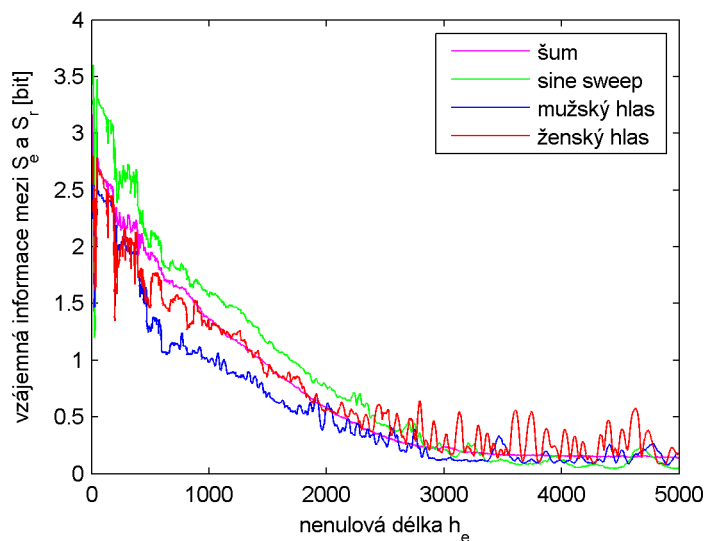
Na obrázku 4.7 je zobrazena čtveřice průběhů korelačního koeficientu mezi stejnými signály  $s_e$  a  $s_r$  jako v předchozích pokusech. Zde už je patrný postupný pokles korelačního koeficientu od 1, tj. úplně korelovaných signálů, až po zhruba 0,3. Průběh pro mužský a ženský hlas mají v místě průběhu značný rozkmit. Průběh sine sweepu je s mešním rozkmitem. Průběh pro bílý šum je nejhladší ze všech průběhů a lze z něho vyčíst postupný pokles, který končí okolo 3 000. prvku. Dále pak klesá jen minimálně. Možný důvod plynulého poklesu je, že bílý šum rovnoměrně vybuzuje celé spektrum v nahrávkách na rozdíl od sine sweepu a řeči.



Obrázek 4.7: Porovnání průběhů korelačního koeficientu

Na obrázku 4.8 je zobrazena čtveřice průběhů vzájemné informace mezi  $s_e$  a  $s_r$ . Z uvedeného grafu vyplývá, že tento pokus má výsledné průběhy nejméně závislé na druhu nahraného signálu. Všechny průběhy začínají v okolí hodnoty vzájemné informace 3 b. V ostatních pokusech začínaly na hodnotách mezi 0–0,7 b, což dělalo signály  $s_e$  a  $s_r$  téměř nezávislé už od malých hodnot  $b$ . Na počátku se chovají trochu chaoticky, ale poté klesají relativně stejně až k hodnotě vzájemné informace 0,2 b u hodnoty  $b = 3 000$ . Mužský a ženský hlas mají mnohem menší rozkmit než v případě průběhu korelačního koeficientu. Stejně jako v předchozím experimentu

průběh pro bílý šum vyšel nejhladší. Z tohoto důvodu jsem vybral průběhy pro bílý šum pro porovnání s ostatními naměřenými případy v následující části.



Obrázek 4.8: Porovnání průběhů vzájemné informace

V příloze C na obrázcích C.1, C.2 a C.3 jsou zobrazeny průběhy korelačního koeficientu a vzájemné informace pro bílý šum v místnostech 1, 2 a 3 pro porovnání vlastností v různých místnostech. Na každém obrázku jsou porovnány průběhy se vzdálenostmi mezi mikrofony 1, 2 a 3 m. Pro místnost 3 jen 1 a 2 m. Z uvedených obrázků vyplývá, že průběhy korelačního koeficientu pro vzdálenost mezi mikrofony 1 m končí pokles průběhu dříve než pro ostatní vzdálenosti. Průběhy pro vzdálenosti 1 a 2 m jsou téměř totožné a není z nich patrná žádná závislost na vzdálenosti mikrofonů. Na průbězích vzájemné informace kromě počátečních mírně chaotických průběhů jsou průběhy prakticky nezávislé na vzdálenosti mikrofonů v libovolné místnosti. Pro každou místnost však pokles končí při různých hodnotách nenulové délky  $h_e$ . Kompletní průběhy pro všechny místnosti, vzdálenosti a druhy signálů jsou uloženy na příloženém CD.

Výsledkem tohoto experimentu jsou již celkem dostačující průběhy pro získání informace o hranici mezi brzkými odrazy a difuzní částí. Výsledné průběhy potvrzují, že hranice mezi těmito částmi je plynulá. V následujícím experimentu se pokusím s pomocí získaných informací dereverberovat signál z nahrávky.

## 5 Dereverberace

Pro ověření správného rozdělení akustické impulsní odezvy na brzké odrazy a difuzní část jsem vytvořil následující experiment. Vybral jsem všechny průběhy vzájemné informace získané z nahrávek šumu z posledního experimentu, protože jejich průběh je nejplynulejší. Pro každou místnost a vzdálenost mezi mikrofony jsem z grafu vzájemné informace odečetl hodnotu  $b$ , kdy hodnota vzájemné informace klesne pod polovinu maximální hodnoty průběhu. Jelikož brzké odrazy a difuzní část se podle grafu plynule prolínají během celé délky vybuzené akustické impulsní odezvy, vybral jsem hodnotu, kdy je jejich poměr v signálu stejný. Poté jsem rozdělil nahrávky s ženským a mužským hlasem na signály  $\mathbf{s}_e$  a  $\mathbf{s}_r$  podle příslušné získané hranice  $b$ . K rozdělení jsem použil akustické impulsní odezvy získané z konkrétních signálů s řečí. Následně jsem pro každý signál vytvořil Wienerův postfilter ve frekvenční oblasti podle

$$W(k, l) = \frac{|S_e(k, l)|^2}{|S_e(k, l)|^2 + |S_r(k, l)|^2}. \quad (5.1)$$

Symbol  $W(k, l)$  značí Wienerův postfilter ve frekvenční oblasti,  $S_e(k, l)$  značí signál  $\mathbf{s}_e$  ve frekvenční oblasti a  $S_r(k, l)$  značí signál  $\mathbf{s}_r$  ve frekvenční oblasti.

V tabulce 5.1 jsou zobrazeny hranice  $b$  v akustické impulsní odezvě mezi brzkými odezvami a difuzní částí, kdy hodnota vzájemné informace poklesla pod polovinu maximální hodnoty. Hodnoty jsou uvedeny pro všechny místnosti a vzdálenosti mezi mikrofony.

Poté jsem se pomocí získaného Wienerova postfiltru pokusil odstranit difuzní část ze vzdáleného mikrofonu  $\mathbf{r}$  pomocí

$$\hat{S}_e(k, l) = R(k, l) \cdot W(k, l), \quad (5.2)$$

Tabulka 5.1: Hodnoty hranic  $b$  pro jednotlivé místnosti a vzdálenosti mikrofonů

Místnost	$d$ [m]	$b$ [vzorky]	$b$ [ms]
Místnost 1	1	826	52
Místnost 1	2	881	55
Místnost 1	3	923	58
Místnost 2	1	739	46
Místnost 2	2	677	42
Místnost 2	3	844	53
Místnost 3	1	779	49
Místnost 3	2	1017	64

kde  $R$  je signál ze vzdáleného mikrofonu ve frekvenční oblasti. Výsledný signál  $\hat{S}_e$  je získaný odfiltrovaný signál s brzkými odrazy ve frekvenční oblasti.

Následně jsem signál  $\hat{S}_e$  převedl do časové oblasti na signál  $\hat{s}_e$ . Poté jsem kvalitu tohoto signálu porovnal s ideálním signálem s brzkými odrazy  $s_e$  pomocí metody PESQ. Zkusil jsem též porovnat kvalitu řeči pomocí metody PESQ mezi signály  $s_e$  a  $r$ , zda se po odstranění difuzní části z  $r$  kvalita řeči v nahrávce zlepšila.

V tabulce 5.2 jsou uvedeny hodnoty PESQ pro originální záznam řeči ze vzdáleného mikrofonu  $r$  a pro dereverberovaný signál s odstraněnou difuzní složkou pomocí Wienerova postfiltru. Z tabulky je patrné zlepšení hodnoty PESQ po odstranění difuzní složky. Na příloženém CD jsou umístěné originální a dereverberované signály, z kterých je dobře slyšet odstranění vzdálených odrazů a difuzní složky.

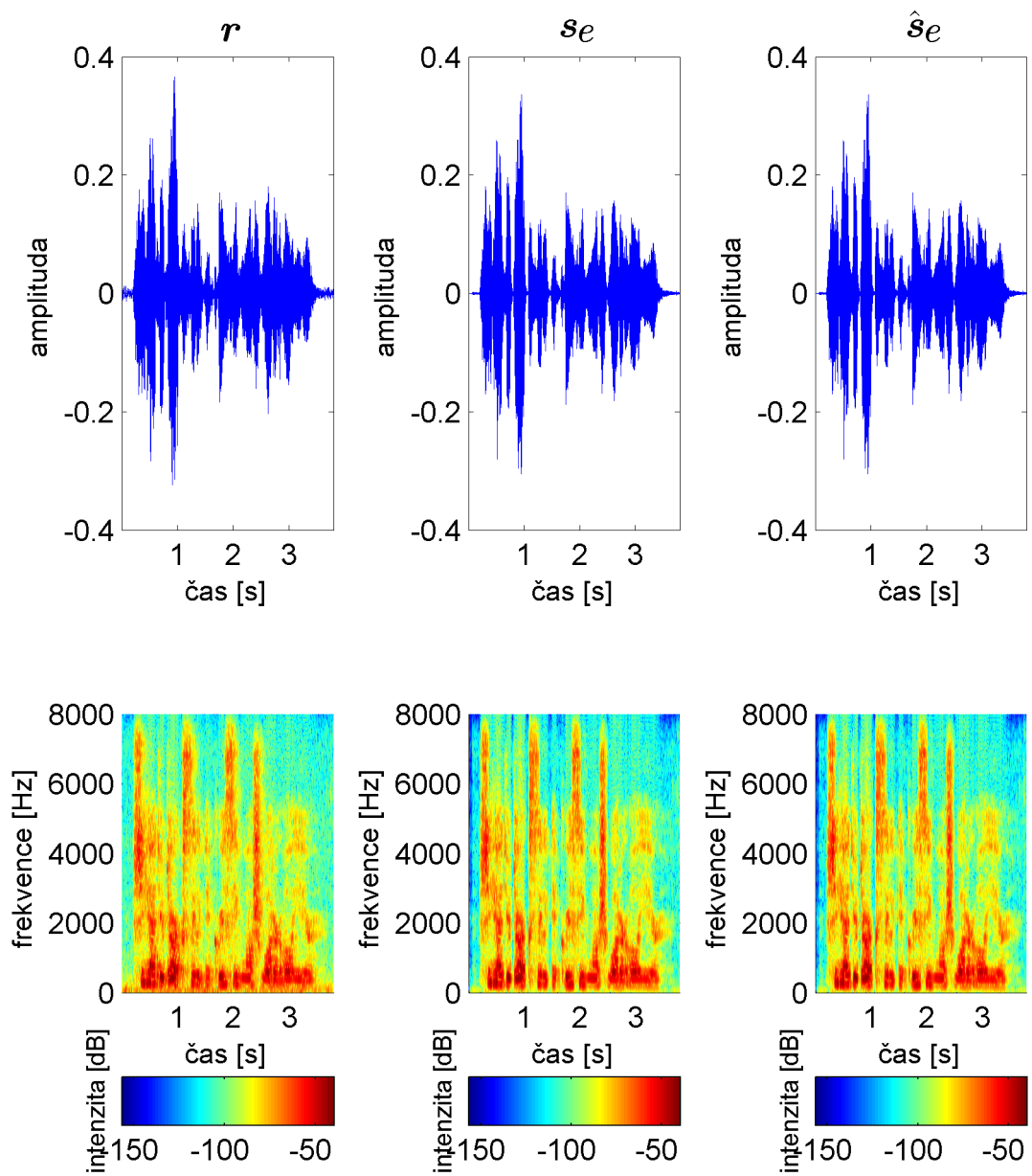
Na obrázku 5.1 je zobrazen příklad signálu ze vzdáleného mikrofonu  $r$  s nahrávkou ženského hlasu v místnosti 1 se vzdáleností mezi mikrofony 3 m. Spolu s ním je vykreslen i signál ideálního signálu s brzkými odrazy  $s_e$  a získaný signál  $\hat{s}_e$  vzniklý odfiltrováním difuzní části ze vzdáleného mikrofonu  $r$  pomocí vzorce 5.2 a převedením do časové oblasti. Hodnota hranice  $b$  pro získání Wienerova postfiltru a signál  $s_e$  je 923 vzorků.

Tabulka 5.2: Hodnoty PESQ pro porovnání kvality řeči na originální nahrávce a na dereverbované nahrávce

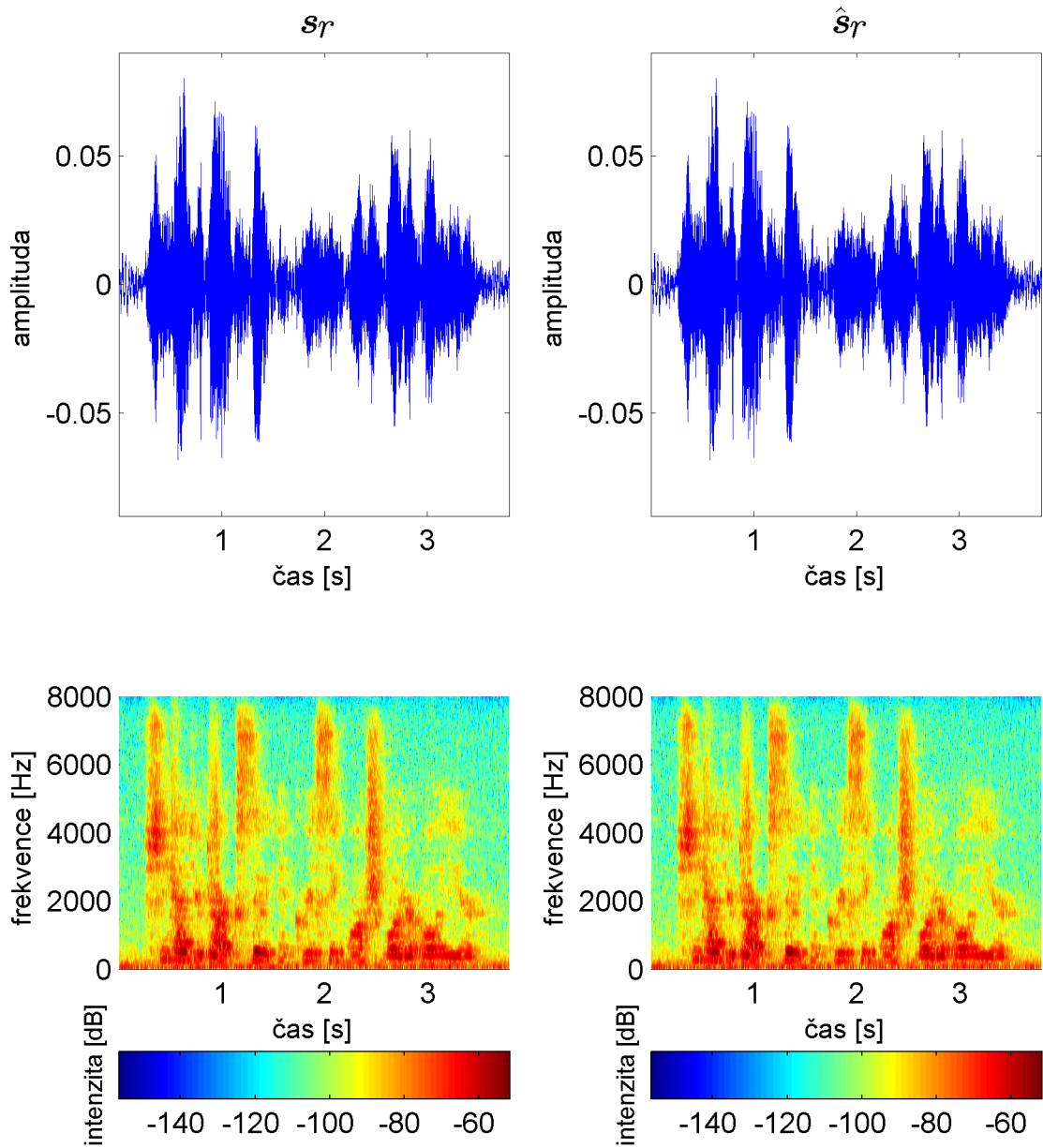
Místnost	hlas	$d$ [m]	PESQ mezi $\mathbf{s}_e$ a $\hat{\mathbf{s}}_e$	PESQ mezi $\mathbf{s}_e$ a $\mathbf{r}$
Místnost 1	mužský	1	3,900297	3,082772
Místnost 1	ženský	1	3,758413	3,105735
Místnost 1	mužský	2	3,810592	3,067256
Místnost 1	ženský	2	3,835885	3,048227
Místnost 1	mužský	3	3,921298	2,964174
Místnost 1	ženský	3	3,956037	3,017171
Místnost 2	mužský	1	3,742683	2,949935
Místnost 2	ženský	1	3,904854	3,140365
Místnost 2	mužský	2	3,686581	2,716610
Místnost 2	ženský	2	3,682700	2,662985
Místnost 2	mužský	3	3,746492	3,061412
Místnost 2	ženský	3	3,917903	3,019751
Místnost 3	mužský	1	3,772499	2,896500
Místnost 3	ženský	1	3,969102	2,957266
Místnost 3	mužský	2	3,778734	2,891793
Místnost 3	ženský	2	3,857240	2,909738

Z uvedeného obrázku je patrné, že filtr  $\mathbf{w}$  dobře odstranil difuzní složku a vzdálené odrazy ze signálu  $\mathbf{r}$ .

Na obrázku 5.2 je zobrazen signál ideálního  $\mathbf{s}_r$  získaného ze stejného signálu jako v předchozím obrázku a difuzní složky  $\hat{\mathbf{s}}_r$ , která byla odstraněna ze signálu  $\mathbf{r}$  pomocí Wienerova postfiltru podle vzorce  $\hat{\mathbf{s}}_r = \mathbf{r} - \hat{\mathbf{s}}_e$ .



Obrázek 5.1: Porovnání reverberovaného signálu s ideálním  $s_e$  a získaným odfiltrovaným  $\hat{s}_e$



Obrázek 5.2: Porovnání ideální difuzní části a odfiltrované difuzní části

## 6 Závěr

V první části této práce jsem se zaměřil na řešení zadaného tématu, kde jsem se seznámil s metodikami hledání hranice v akustické impulsní odezvě mezi brzkými odrazy a difuzní částí. Poté jsem navrhl vlastní metodiku rozdělení akustické impulsní odezvy tak, aby konvoluce signálu s první částí byla nezávislá nebo nekorelovaná na konvoluci stejného signálu s druhou částí.

Ve druhé části jsem vytvořil nahrávky v různých místnostech s různými druhy signálů a odhadl z nich akustické impulsní odezvy. Dále jsem ověřil jejich vlastnosti.

Ve třetí části jsem se seznámil s algoritmy pro porovnávání dvou signálů, jejich konvoluci a filtrování pro jejich aplikaci v následující části.

V další nejdůležitější části jsem aplikoval vlastní návrh rozdělení akustické impulsní odezvy na brzké odrazy a difuzní část. Nejdříve jsem postupně rozděloval po vzorcích AIR a porovnával vzniklé signály pomocí korelačního koeficientu a vzájemné informace. Výsledek tohoto experimentu byl nedostačující, protože výsledné průběhy závislosti signálu vyšly chaotické a závislé na druhu signálu. Poté jsem před porovnáváním vzniklých signálů tyto signály časově přizpůsobil vzájemným posunutím signálů o hodnotu, která vyšla vzájemnou korelací. Výsledek tohoto experimentu byl méně chaotický než v předešlém pokusu, nicméně průběhy závislosti byly stále silně závislé na druhu signálu. Nakonec jsem před porovnáváním závislosti tyto signály přizpůsobil pomocí Wienerova filtru, kde po zfiltrování prvního signálu vznikl tvarově podobný signál druhý. Výsledkem byly už dostačující průběhy závislosti, které již nebyly tolik závislé na druhu signálu. Tyto průběhy měly postupný pokles hodnot vzájemné informace i korelačního koeficientu. Z toho jsem usoudil, že poměr brzkých odrazů a difuzní části se během celé délky vybuze akustické impulsní



odezvy mění. S rostoucím časem klesá podíl brzkých odezev a naopak stoupá podíl difuzní části.

V posledním experimentu jsem ověřil správnost výsledků pomocí dereverberace. Z výsledných průběhů závislosti jsem zvolil ty, které odpovídají šumu, protože měly nejhladší průběh. Poté jsem zvolil práh pro průběh vzájemné informace podle toho, kde průběh nabývá poloviny maximální hodnoty. V tomto místě je podíl brzkých odezev a difuzní části stejný. Výsledné časy, kdy je podíl brzkých odezev vyšší než difuzní složka, byly mezi 42–64 ms. Tyto hodnoty se pohybují v rozmezí daném v člancích, které řešily stejné téma jinými metodami. Následně jsem pomocí vzniklých signálů vytvořil Wienerův postfiltr, který z reverberované nahrávky řeči odstranil difuzní část. Výsledný dereverberovaný signál jsem porovnal pomocí metody PESQ s ideálně dereverberovaným signálem. Výsledkem bylo zlepšení srozumitelnosti řeči v nahrávce.

## Literatura

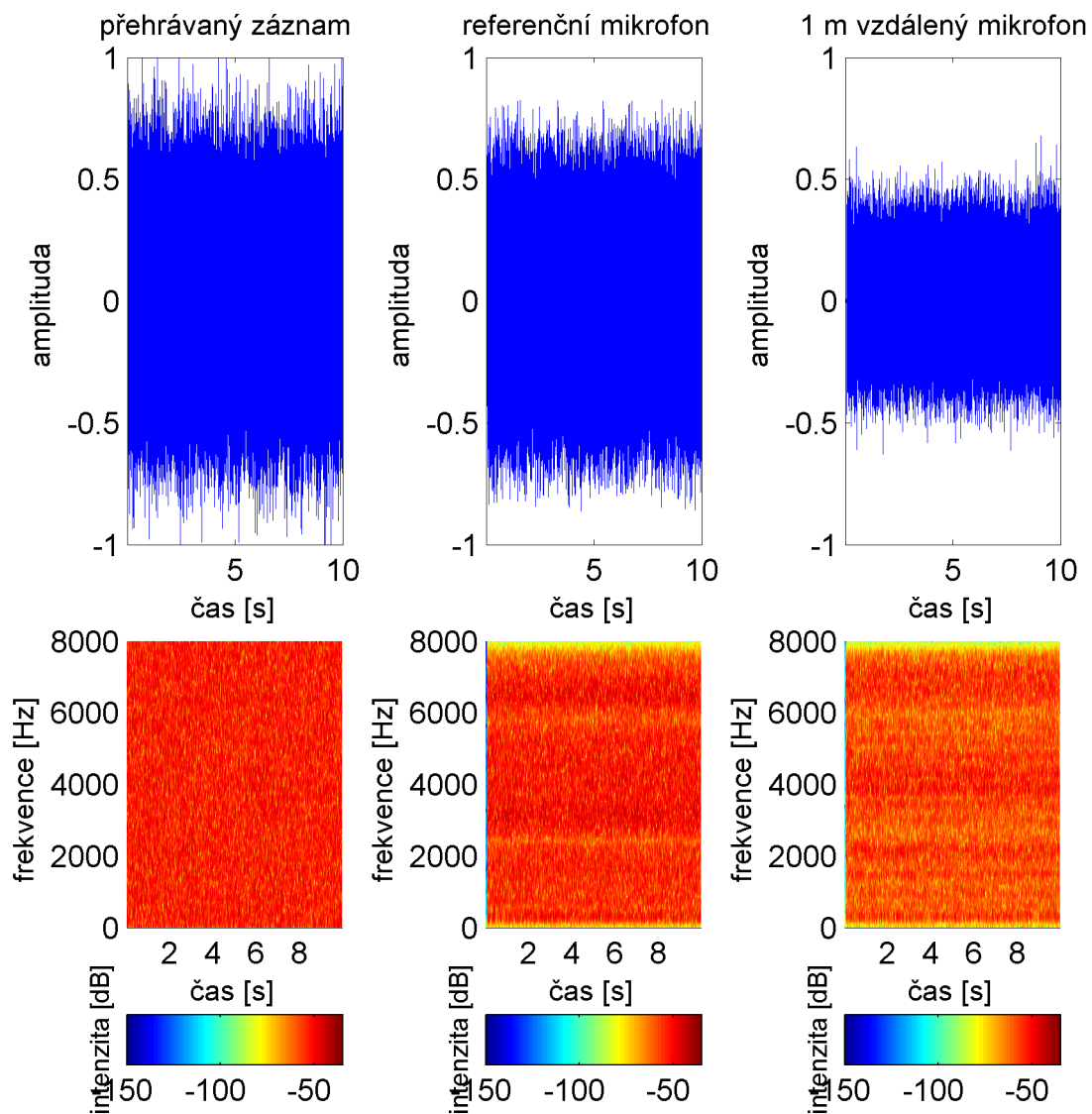
- [1] GANNOT, Sharon, Emmanuel VINCENT, Shmulik MARKOVICH-GOLAN a Alexey OZEROV. A Consolidated Perspective on Multimicrophone Speech Enhancement and Source Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* [online]. 2017, **25**(4), 692-730 [cit. 2017-05-02]. DOI: 10.1109/TASLP.2016.2647702. ISSN 2329-9290. Dostupné z: <http://ieeexplore.ieee.org/document/7805139/>
- [2] HIDAKA, Takayuki, Yoshinari YAMADA a Takehiko NAKAGAWA. A new definition of boundary point between early reflections and late reverberation in room impulse responses. *The Journal of the Acoustical Society of America* [online]. 2007, **122**(1), 326-332 [cit. 2017-05-02]. DOI: 10.1121/1.2743161. ISSN 0001-4966. Dostupné z: <http://asa.scitation.org/doi/10.1121/1.2743161>
- [3] GÖTZ, Philipp, Konrad KOWALCZYK, Andreas SILZLE a Emanuël A. P. HABETS. Mixing time prediction using spherical microphone arrays. *The Journal of the Acoustical Society of America* [online]. 2015, **137**(2), EL206-EL212 [cit. 2017-05-10]. DOI: 10.1121/1.4907547. ISSN 0001-4966. Dostupné z: <http://asa.scitation.org/doi/10.1121/1.4907547>
- [4] CULDA, Tudor M., Victor POPA, Dumitru STANOMIR a Cristian NEGRESCU. Reducing time in acoustic impulse response measurements using exponential sine sweeps. In: *International Symposium on Signals, Circuits and Systems ISSCS2013* [online]. IEEE, 2013, s. 1-4 [cit. 2017-04-20]. DOI: 10.1109/ISSCS.2013.6651220. ISBN 978-1-4673-6143-9. Dostupné z: <http://ieeexplore.ieee.org/document/6651220/>

- [5] MÁLEK, Jiří. *Počítačové zpracování signálů - Přednáška VI*. Liberec, 2015.
- [6] HUANG, Yiteng, Jacob. BENESTY a J. CHEN. *Acoustic MIMO signal processing*. New York: Springer, 2006. ISBN 978-3-540-37630-9.
- [7] KOLDOVSKÝ, Zbyněk. *Biologické a akustické signály: Kvadratická kritéria pro porovnávání signálů*. Liberec, 2017.
- [8] Correlation coefficients - MATLAB corrcoef. *MathWorks - Makers of MATLAB and Simulink* [online]. [cit. 2017-04-29]. Dostupné z: <https://www.mathworks.com/help/matlab/ref/corrcoef.html?requestedDomain=www.mathworks.com&nocookie=true#bunkanr>
- [9] COVER, T. M. a Joy A. THOMAS. *Elements of information theory*. 2nd ed. Hoboken, N.J.: Wiley-Interscience, c2006. ISBN 978-0-471-24195-9.
- [10] DARBELLAY, G.A. a I. VAJDA. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory* [online]. **45**(4), 1315-1321 [cit. 2017-05-09]. DOI: 10.1109/18.761290. ISSN 00189448. Dostupné z: <http://ieeexplore.ieee.org/document/761290/>
- [11] ITU-T RECOMMENDATION P.862. *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Switzerland, 2001.

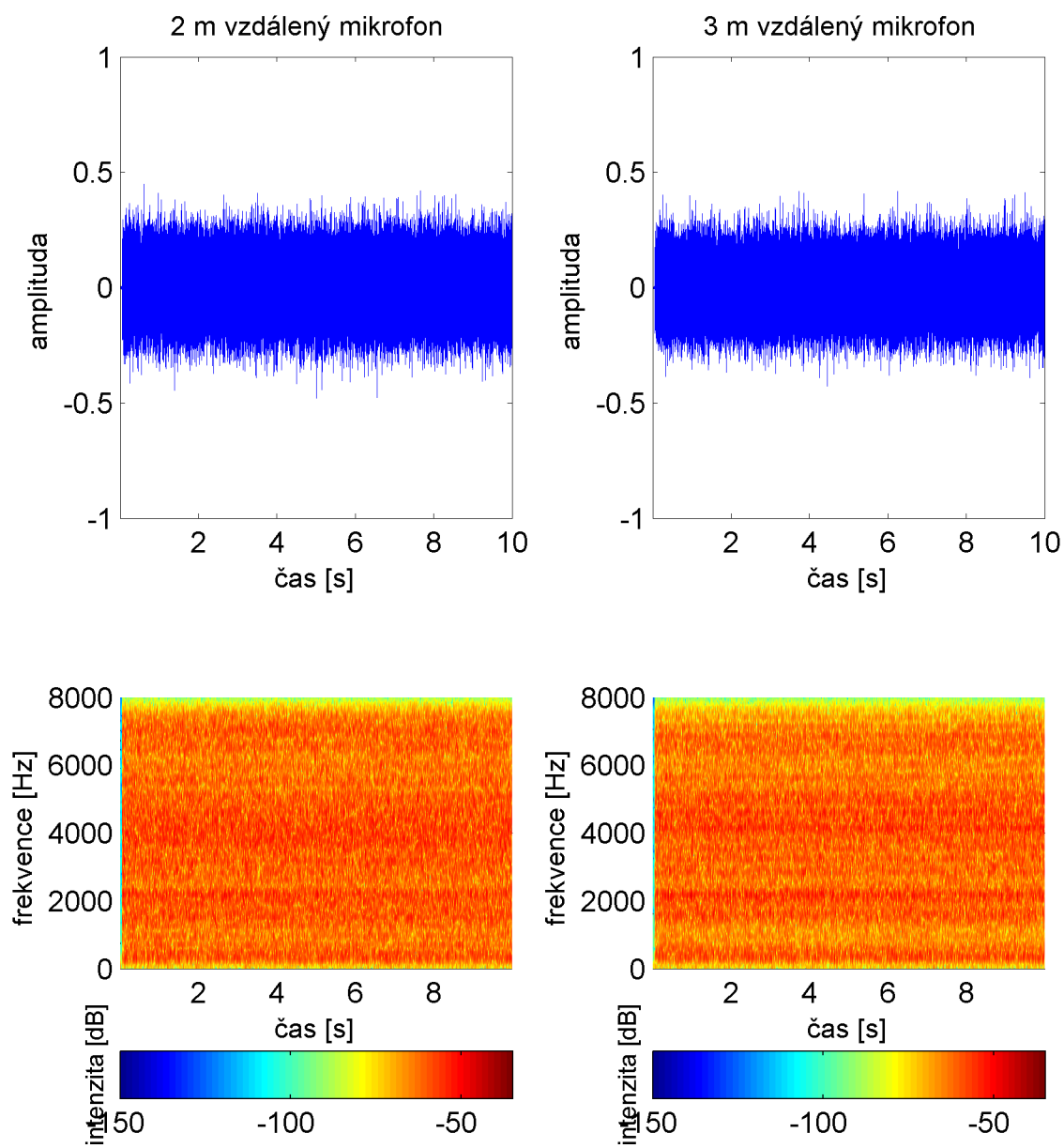
## A Obsah přiloženého CD

- text diplomové práce
  - diplomova\_prace\_2017\_Tomas\_Franek.pdf
  - diplomova\_prace\_2017\_Tomas\_Franek.tex
  - kopie\_zadani\_diplomova\_prace\_2017\_Tomas\_Franek.pdf
- dereverberace – skripty a dereverberované signály
- impulsni\_odezvy – skripty a spočtené impulsní odezvy
- nahravky – skripty, zdrojové signály, nahrané signály
- overlap add – skripty implementace a test metody
- spoctene\_koeficienty – skripty pro výpočet a spočtené průběhy korelačního koeficientu a vzájemné informace bez synchronizace
- spoctene\_koeficienty\_firwiener\_5000 – skripty pro výpočet a spočtené průběhy korelačního koeficientu a vzájemné informace se synchronizací pomocí Wienerova filtru
- spoctene\_koeficienty\_xcorr – skripty pro výpočet a spočtené průběhy korelačního koeficientu a vzájemné informace se synchronizací pomocí vzájemné korelace

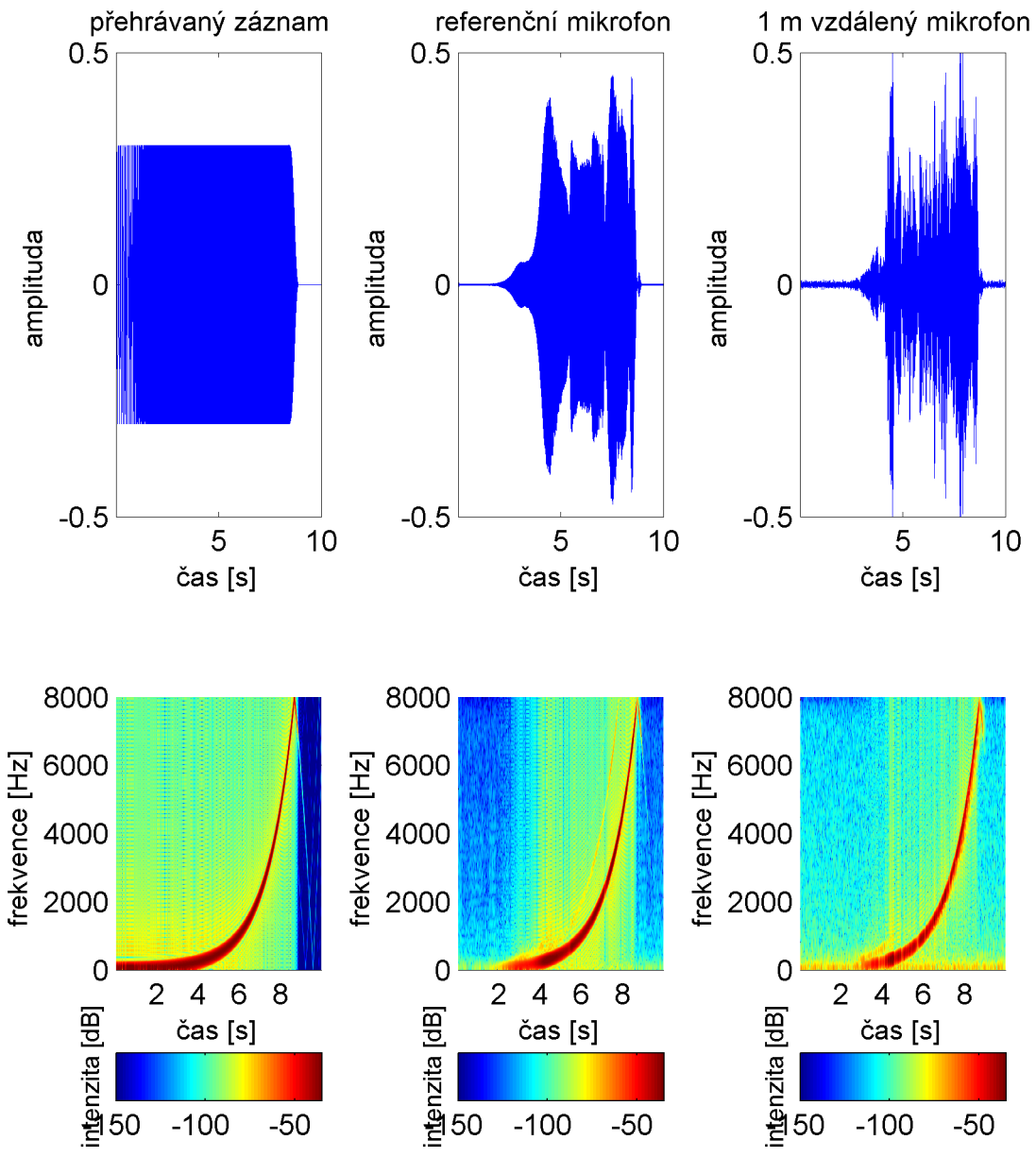
## B Spektrogramy nahraných signálů



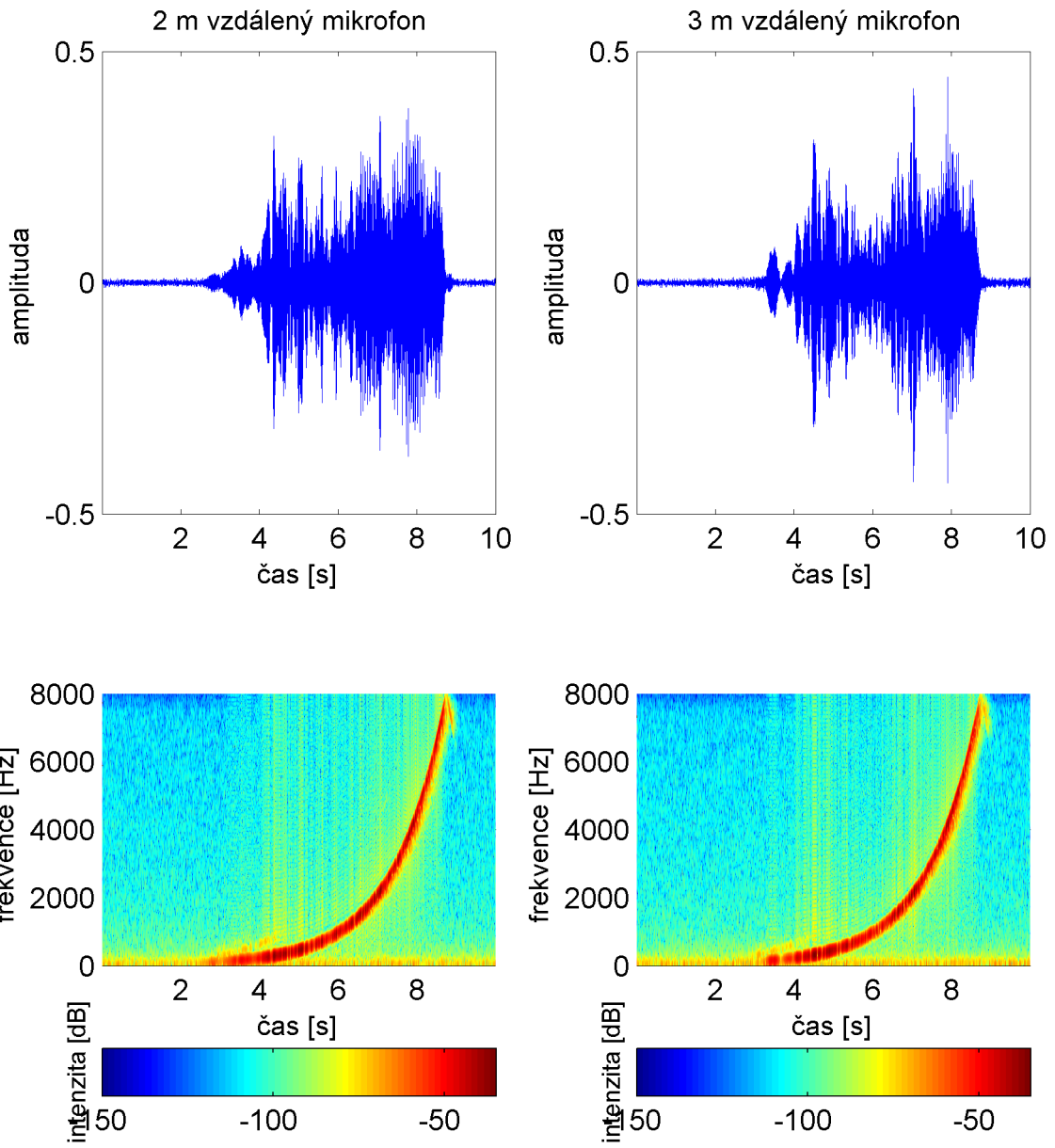
Obrázek B.1: Porovnání nahrávek šumu 1 m



Obrázek B.2: Porovnání nahrávek šumu 2 a 3 m

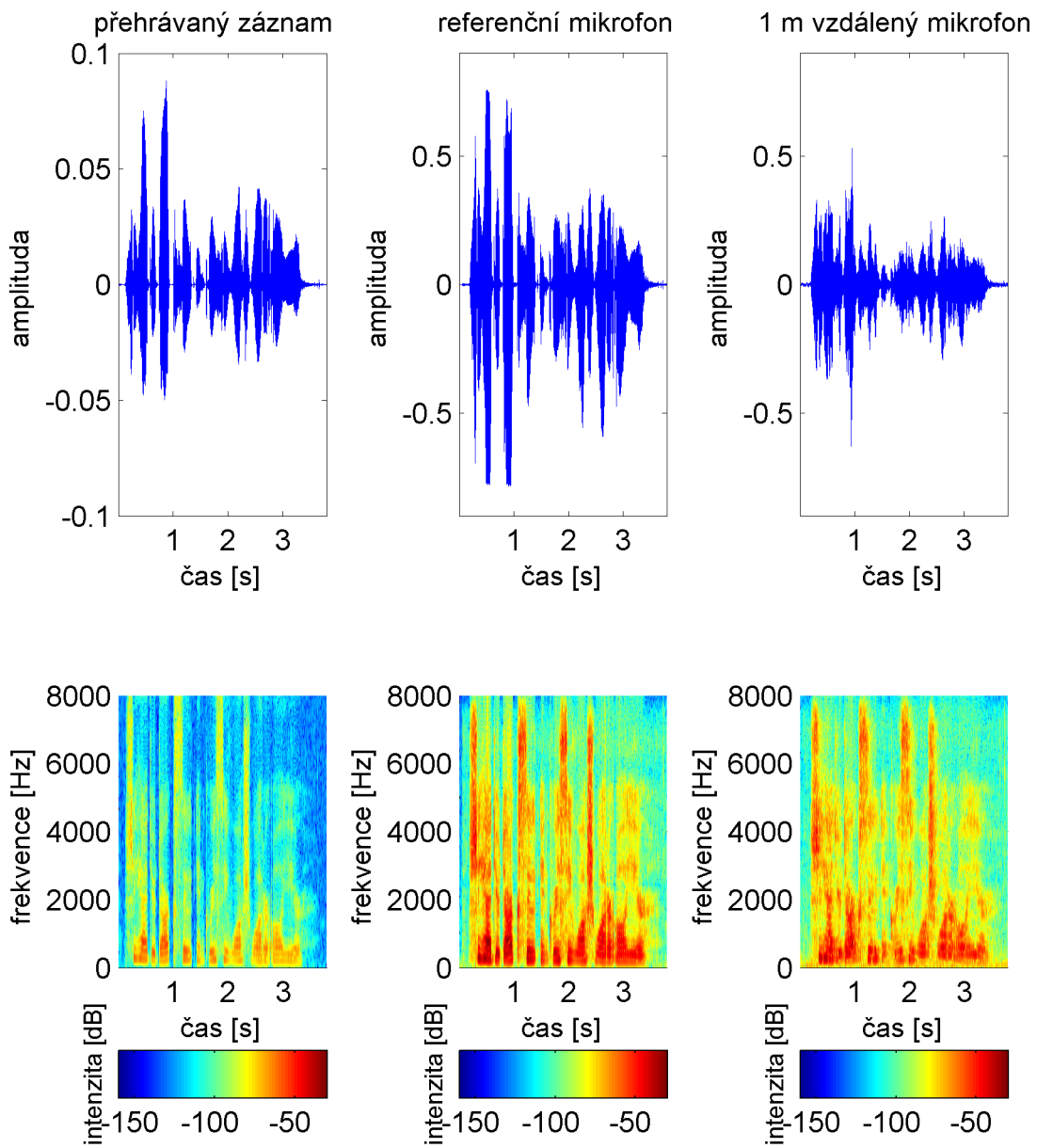


Obrázek B.3: Porovnání nahrávek sine sweepu 1 m

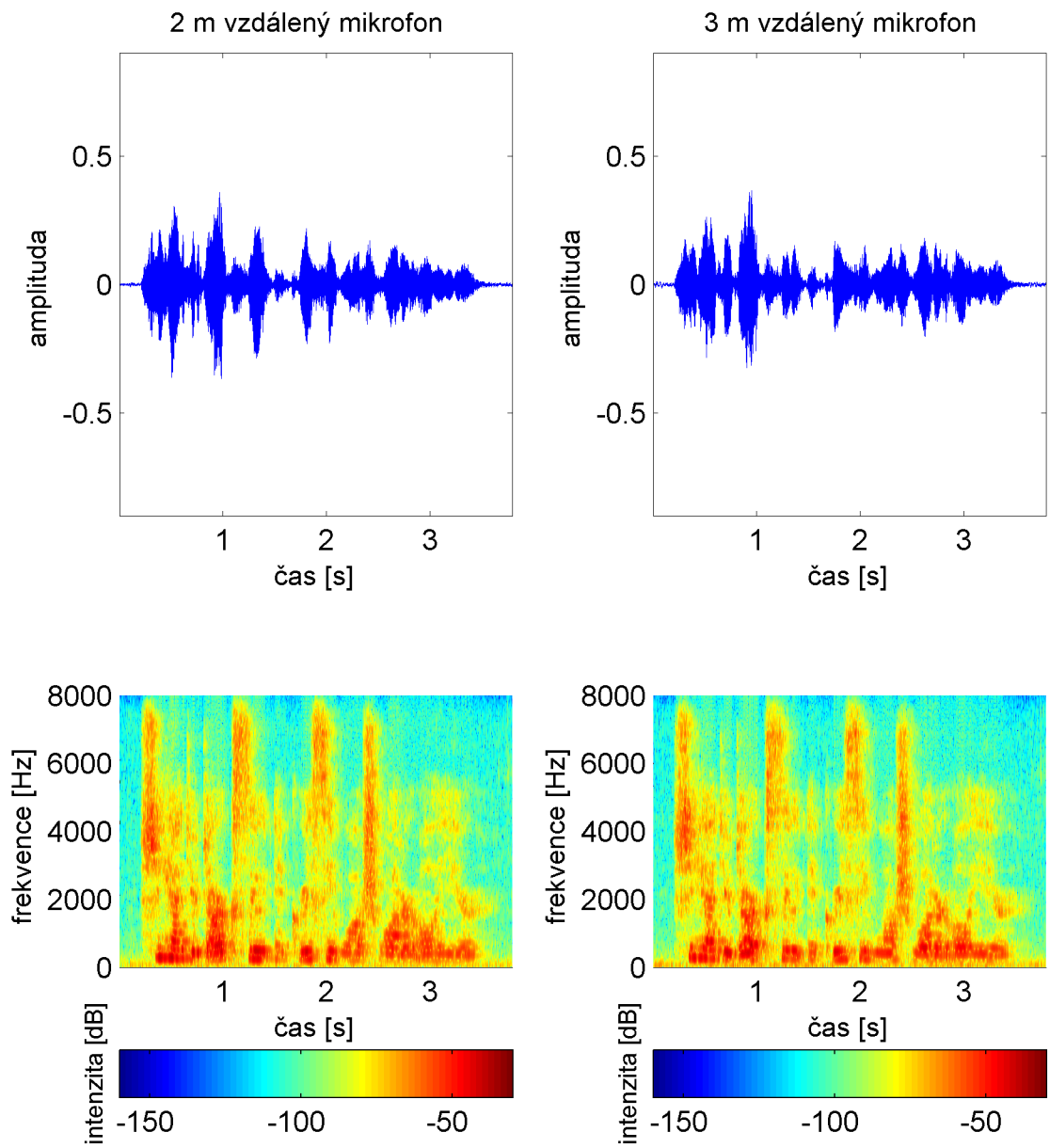


Obrázek B.4: Porovnání nahrávek sine sweepu 2 a 3 m

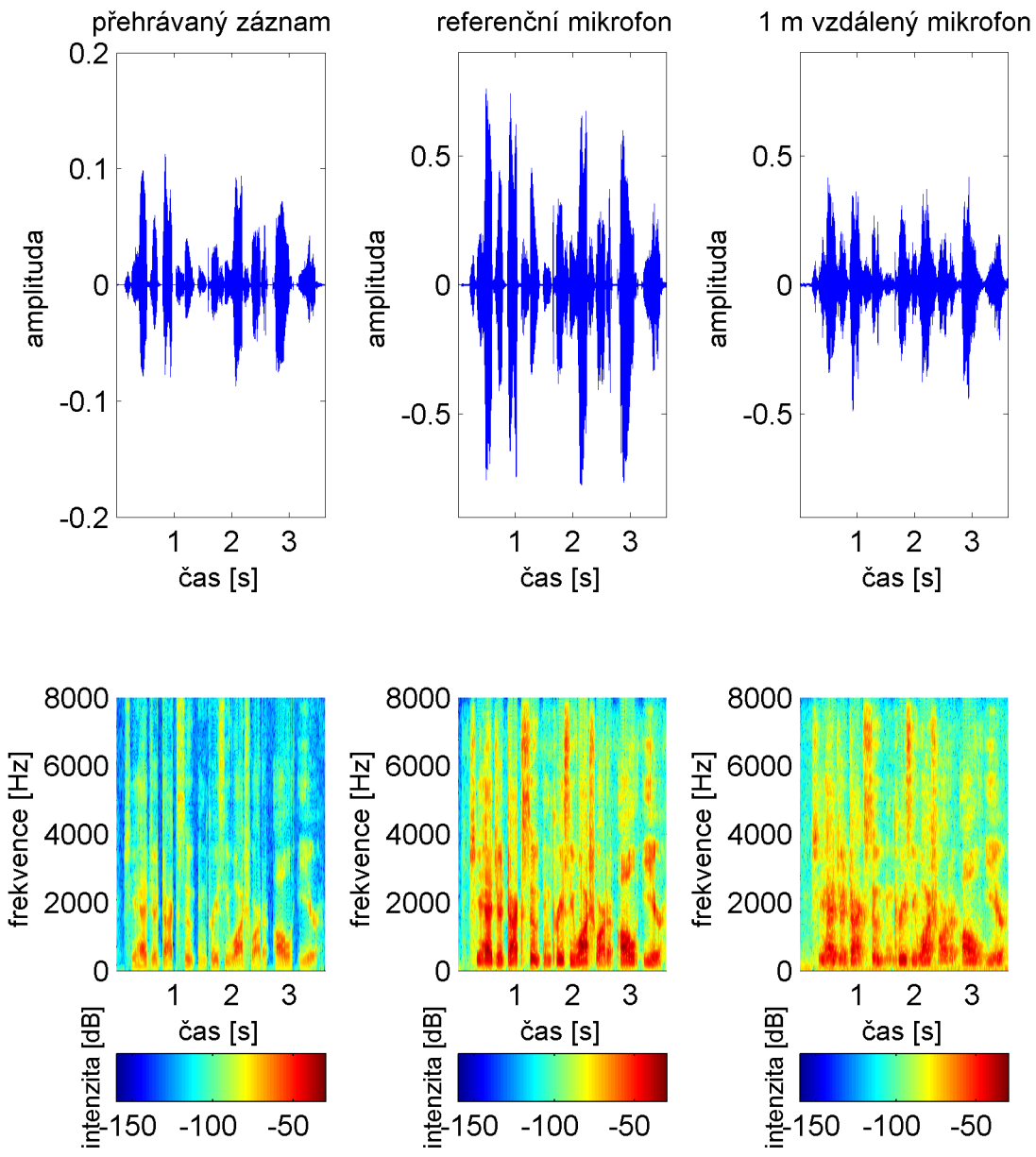




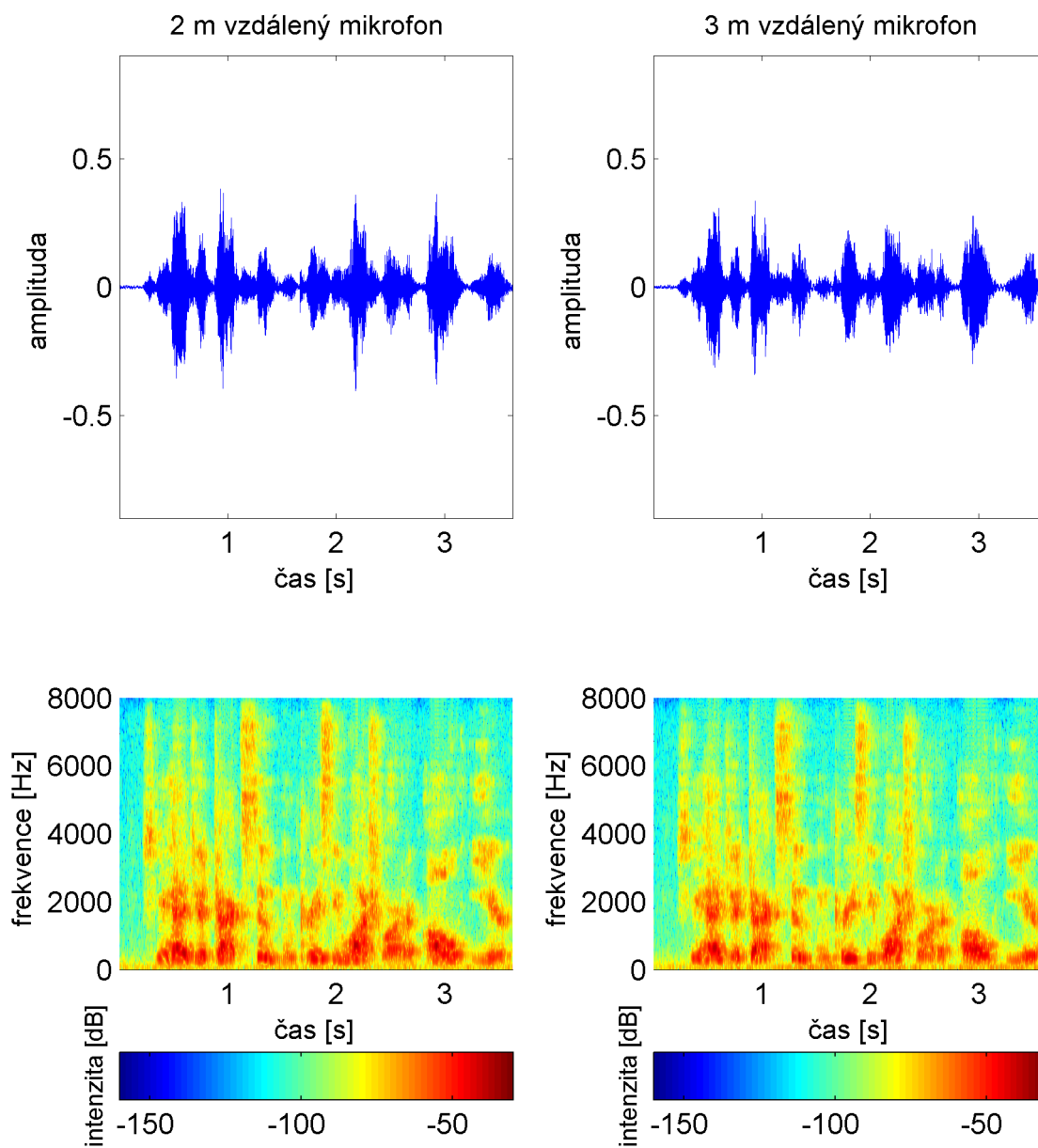
Obrázek B.5: Porovnání nahrávek ženského hlasu 1 m



Obrázek B.6: Porovnání nahrávek ženského hlasu 2 a 3 m

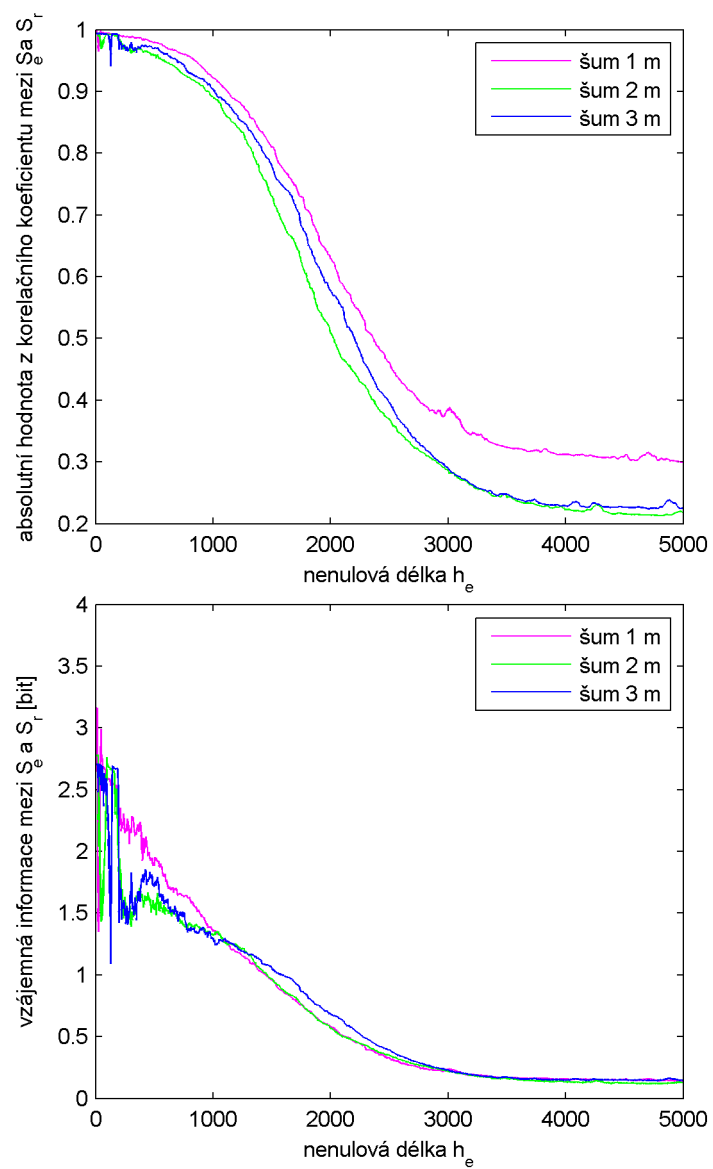


Obrázek B.7: Porovnání nahrávek mužského hlasu 1 m

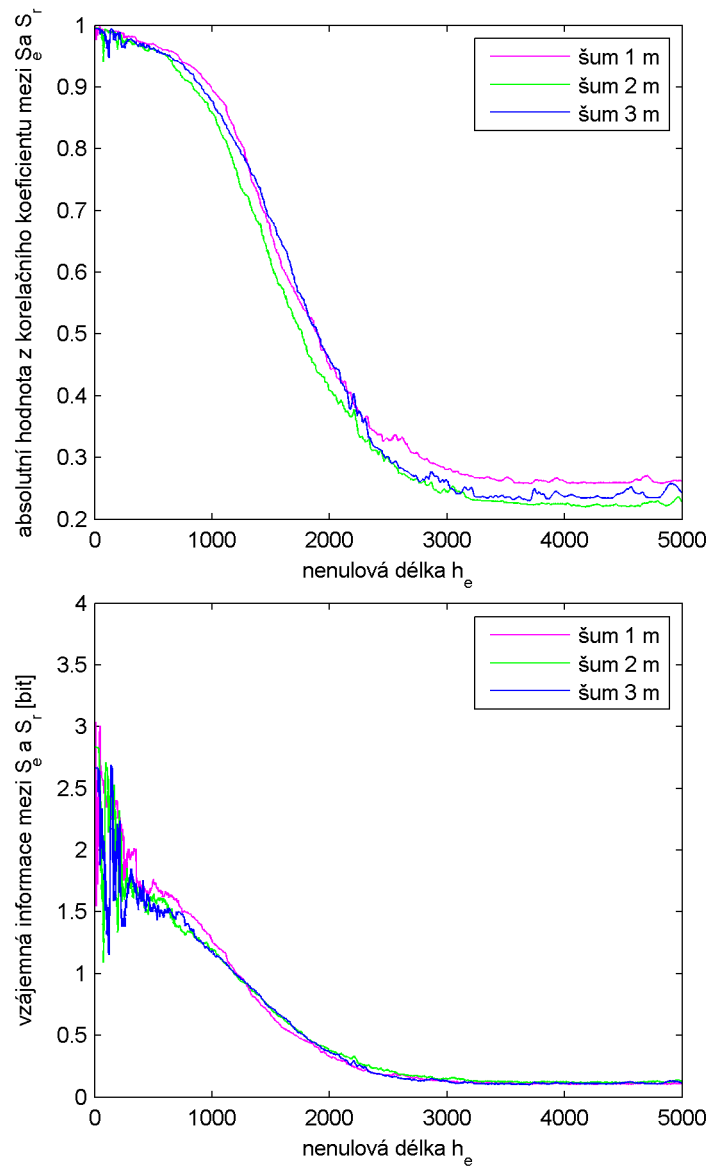


Obrázek B.8: Porovnání nahrávek mužského hlasu 2 a 3 m

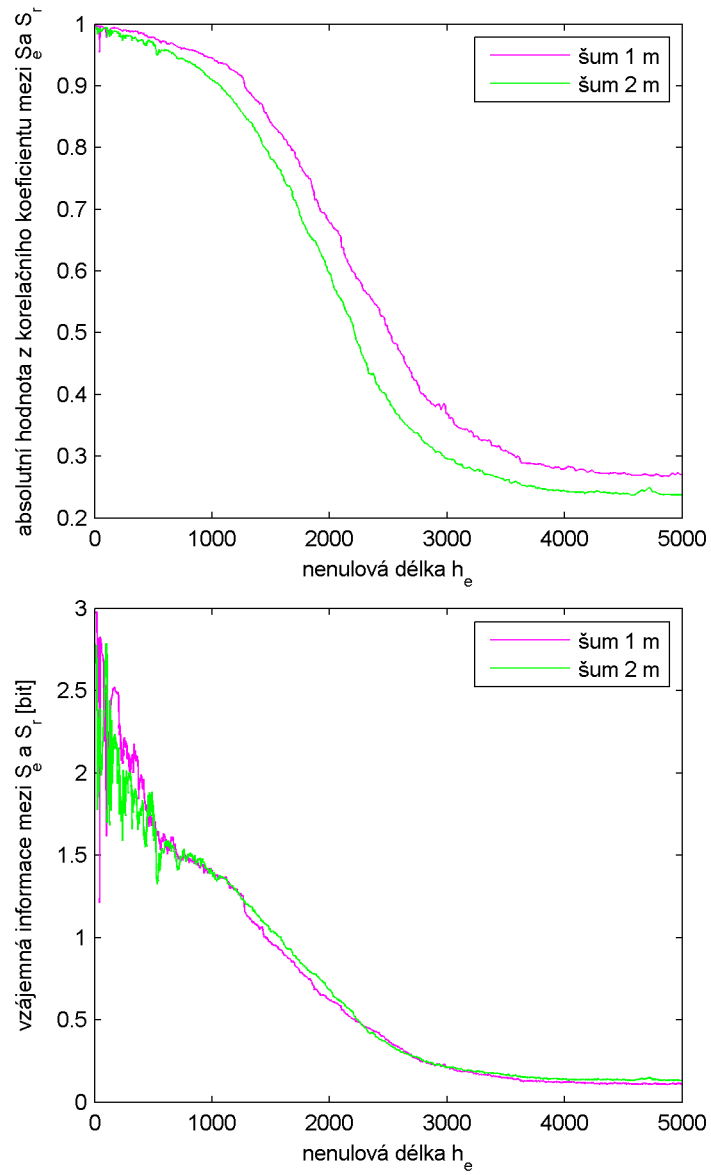
## C Průběhy



Obrázek C.1: Průběhy korelačního koeficientu a vzájemné informace pro místnost 1



Obrázek C.2: Průběhy korelačního koeficientu a vzájemné informace pro místnost 2



Obrázek C.3: Průběhy korelačního koeficientu a vzájemné informace pro místnost 3