



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ
ÚSTAV TELEKOMUNIKACÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION
DEPARTMENT TELECOMMUNICATIONS

OBJEKTIVNÍ MĚŘENÍ A POTLAČOVÁNÍ ŠUMU V HUDEBNÍM SIGNÁLU

OBJECTIVE ASSESSMENT AND REDUCTION OF NOISE IN MUSICAL SIGNAL

DISERTAČNÍ PRÁCE

AUTOR PRÁCE

Ing. ONDŘEJ RÁŠO

AUTHOR

VEDOUČÍ PRÁCE

Ing. MIROSLAV BALÍK, Ph.D.

SUPERVISOR

Vzor desek závěrečné práce (Adobe PDF)

Vzor titulního listu (Adobe PDF)

Abstrakt

Tato disertační práce pojednává o objektivním měření a potlačování rušivého šumu na pozadí hudebního signálu. V této práci je navrhnout nový algoritmus pro objektivní měření slyšitelnosti tohoto typu šumu. Provedenými poslechovými testy bylo prokázáno, že tento nový algoritmus lépe predikuje slyšitelnost šumu na pozadí než stávající algoritmy. Výhodou navrženého algoritmu je skutečnost, že lze tento algoritmus využít i na obecný zvukový signál, kdy se hodnotí slyšitelnost jednoho zvuku na pozadí jiného. U tohoto typu signálu stávající algoritmy často selhávají. Dále je v této práci navrhnout nový způsob adaptivní segmentace pro dělení dlouhotrvajícího zvukového signálu na krátké segmenty s proměnlivou délkou. Bylo ukázáno, že při použití tohoto nového způsobu segmentace v systémech pro potlačení šumu na pozadí, má výstupní zvukový signál vyšší subjektivně vnímanou kvalitu než ostatní testované způsoby segmentace.

Klíčová slova

Zvukové signály, Digitální filtry, Číslicové zpracování signálů, Potlačení šumu, Psychoakustický model, Restaurování zvukových signálů, Stanovení kvality zvukových signálů.

Abstract

The dissertation thesis focuses on objective assessment and reduction of disturbing background noise in a musical signal. In this work, a new algorithm for the assessment of background noise audibility is proposed. The listening tests performed show that this new algorithm better predicts the background noise audibility than the existing algorithms do. An advantage of this new algorithm is the fact that it can be used even in the case of a general audio signal and not only musical signal, i.e. in the case when the audibility of one sound on the background of another sound is assessed. The existing algorithms often fail in this case. The next part of the dissertation thesis deals with an adaptive segmentation scheme for the segmentation of long-term musical signals into short segments of different lengths. A new adaptive segmentation scheme is then introduced here. It has been shown that this new adaptive segmentation scheme significantly improves the subjectively perceived quality of the musical signal from the output of noise reduction systems which use this new adaptive segmentation scheme. The quality improvement is better than that achieved by other segmentation schemes tested.

Keywords

Audio restoration, Audio Signals, Digital Filters, Noise reduction, Digital Signal Processing, Audio Quality Assessment, Psychoacoustic model.

RÁŠO, O. *Objektivní měření a potlačování šumu v hudebním signálu*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2013. 148 s. Vedoucí disertační práce Ing. Miroslav Balík, Ph.D.

PROHLÁŠENÍ

Prohlašuji, že svou doktorskou práci na téma „Objektivní měření a potlačování šumu v hudebním signálu“ jsem vypracoval samostatně pod vedením vedoucího doktorské 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 doktorské práce dále prohlašuji, že v souvislosti s vytvořením této doktorské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následku porušení ustanovení §11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledku vyplývajících z ustanovení §152 trestního zákona č. 140/1961 Sb.

Brno

.....

(podpis autora)

PODĚKOVÁNÍ

Děkuji vedoucímu práce Ing. Miroslavu Balíkovi, Ph.D. za metodickou a odbornou pomoc při zpracování této disertační práce.

V neposlední řadě bych chtěl poděkovat rodičům a kolegům za podporu při tvorbě této práce.

Obsah disertační práce

1	ÚVOD	12
2	PŘEHLED SOUČASNÉHO STAVU PROBLEMATIKY	15
2.1	VYMEZENÍ POUŽITÝCH POJMŮ	15
2.2	POTLAČOVÁNÍ ADITIVNÍHO ŠUMU POZADÍ VE ZVUKOVÝCH SYSTÉMECH.	16
2.2.1	<i>Restaurování poškozených nebo nekvalitně pořízených zvukových záznamů.....</i>	<i>16</i>
2.2.2	<i>Systémy na potlačení aditivního šumu pozadí.</i>	<i>26</i>
2.3	MĚŘENÍ ŠUMU POZADÍ VE ZVUKOVÝCH SIGNÁLECH.....	30
2.3.1	<i>Úvod do subjektivního vnímání zvukových vjemů</i>	<i>30</i>
2.3.2	<i>Subjektivní metody a jejich škály vhodné pro měření šumu na pozadí ve zvukových signálech ...</i>	<i>31</i>
2.3.3	<i>Jednoduché objektivní metody</i>	<i>33</i>
2.3.4	<i>Moderní objektivní měřicí metody</i>	<i>34</i>
3	CÍLE DISERTAČNÍ PRÁCE	35
4	OBJEKTIVNÍ METODA PEAQ	37
5	OBJEKTIVNÍ METODA PRO MĚŘENÍ ÚROVNĚ SLYŠITELNÉHO ŠUMU POZADÍ V HUDEBNÍCH SIGNÁLECH – VLASTNÍ ŘEŠENÍ.....	42
5.1	VÝCHOZÍ ÚVAHY PRO NAVRHOVANOU METODU.....	42
5.2	PRINCIP NAVRHOVANÉ METODY	44
5.2.1	<i>Popis blokového schématu navrhované metody</i>	<i>44</i>
5.2.2	<i>Segmentace</i>	<i>45</i>
5.2.3	<i>Odhady výkonových spektrálních hustot.....</i>	<i>46</i>
5.2.4	<i>Výpočet maskujícího prahu</i>	<i>47</i>
5.2.5	<i>Transformace frekvenční osy v Hz na Barkova pásma</i>	<i>51</i>
5.2.6	<i>Stanovení odstupe šumu k masce</i>	<i>55</i>
5.2.7	<i>Stanovení relativní četnosti n emaskovaných komponent šumu na pozadí</i>	<i>57</i>
5.2.8	<i>Ukázky konkrétních příkladů šumem degradovaných zvukových záznamů</i>	<i>58</i>
5.3	POSLECHOVÝ EXPERIMENT.....	61
5.3.1	<i>Subjektivní měření prahu slyšitelnosti šumu pozadí pro testované zvukové signály</i>	<i>61</i>
5.3.2	<i>Charakteristika poslechového experimentu</i>	<i>63</i>
5.3.3	<i>Příprava a průběh poslechového experimentu</i>	<i>65</i>
5.3.4	<i>Shnutí výsledků realizovaného poslechového experimentu</i>	<i>67</i>
5.4	VÝSLEDNÉ ZHODNOCENÍ NAVRŽENÉ OBJEKTIVNÍ METODY.....	72
5.4.1	<i>Porovnání výsledků pro hudební a řečové signály.....</i>	<i>72</i>
5.4.2	<i>Porovnání výsledků pro umělé zvukové signály</i>	<i>75</i>
6	ZAVEDENÍ PARAMETRICKÉHO WIENEROVA FILTRU	78
6.1	WIENERŮV FILTR	78
6.2	KRÁTKODOBÝ WIENERŮV FILTR.....	80
6.3	TECHNIKY SPEKTRÁLNÍHO ODEČÍTÁNÍ	81
6.4	PARAMETRICKÝ WIENERŮV FILTR	82
7	ADAPTIVNÍ SEGMENTAČNÍ SCHÉMA	85
7.1	ADAPTIVNÍ SEGMENTACE	85
7.2	MOŽNOSTI VYUŽITÍ ADAPTIVNÍ SEGMENTACE	88
7.3	ALGORITMUS GREEDY SELECTION	92

8	SYSTÉM NA POTLAČENÍ ŠUMU V HUDEBNÍCH SIGNÁLECH – VLASTNÍ ŘEŠENÍ.....	95
8.1	MODIFIKOVANÁ VARIANTA ALGORITMU GREEDY SELECTION.....	95
8.1.1	<i>Snížení vlivu spektrálního prosakování.....</i>	<i>95</i>
8.1.2	<i>Algoritmus Maximum greedy selection.....</i>	<i>98</i>
8.2	PRAHOVACÍ PRAVIDLA.....	103
8.2.1	<i>Zavedení prahovacích pravidel</i>	<i>103</i>
8.2.2	<i>Souvislost prahovacích pravidel se ziskovými funkcemi</i>	<i>106</i>
8.3	POSLECHOVÝ EXPERIMENT.....	108
8.3.1	<i>Popis poslechového experimentu</i>	<i>108</i>
8.3.2	<i>Příprava a průběh poslechového experimentu</i>	<i>109</i>
8.4	VÝSLEDKY PROVEDENÉHO POSLECHOVÉHO EXPERIMENTU.....	110
9	ZÁVĚR.....	112
10	LITERATURA.....	115
11	PŘÍLOHY K DISERTAČNÍ PRÁCI	120

Seznam všech zkratek

ACR	Absolute Category Rating
BSD	Bark-Spectral Distortion
CD	Cepstral Distance
DAW	Digital Audio Workstation
DCR	Degradation Category Rating
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DI	Distortion Index
DS	Detectability Opinion Scale
DST	Discrete Sine Transform
EMBSD	Enhanced MBSD
FT	Fourier Transform
GSR	Global SNR
ITU	International Telecommunication Union
KEMAR	Knowles Electronics Manikin for Acoustic Research
LAR	Log-Area Distance
LP	Long Play
LPC	Linear Predictive Coding
LSNR	Local SNR
MBSD	Modified BSD
MOS	Mean Opinion Score
MOV	Model Output Values
MPEG	Moving Picture Experts Group
MS	Magnitude Subtraction
MSE	Mean Squared Error
NMR	Noise to Mask Ratio
ODG	Objective Different Grade
PCM	Pulse Code Modulation
PEAQ	Perceptual Evaluation of Audio Quality
PESQ	Perceptual Evaluation of Speech Quality
POLQA	Perceptual Objective Listening Quality Assessment
PS	Power Subtraction
PSD	Power Spectral Density
PSQM	Perceptual Speech Quality Measure
RASTA	RelAtive SpecTrAl processing
SDG	Subjective Difference Grade
SNR	Signal to Noise Ratio
SSNR	Segmental SNR
STFT	Short Time Fourier Transform
TDMA	Time Division Multiple Access
TOSQA	Telecommunication Objective Speech Quality Assessment
VAD	Voice Activity Detection
WT	Wavelet Transformation

Seznam obrázků

<i>Obr. 1.1. Frekvenční oblasti slyšení lidského sluchového ústrojí.....</i>	<i>13</i>
<i>Obr. 2.1. Spektrogram záznamu hry na kytaru s přítomností harmonického rušení typu Hum.</i>	<i>17</i>
<i>Obr. 2.2. Detail spektrogram harmonického rušení typu Hum].</i>	<i>17</i>
<i>Obr. 2.3. Spektrogram záznamu pěveckého sboru z LP vinyl desky z roku 1965.</i>	<i>18</i>
<i>Obr. 2.4. Spektrogram záznamu řečového signálu anglicky mluvícího mluvčího a zvuku cinkání.....</i>	<i>19</i>
<i>Obr. 2.5. Spektrogram řečového záznamu anglicky mluvícího mluvčího.....</i>	<i>20</i>
<i>Obr. 2.6. Clipping, časové průběhy.....</i>	<i>20</i>
<i>Obr. 2.7. Zarušení bílým šumem pozadí pro různé hodnoty SNR.....</i>	<i>22</i>
<i>Obr. 2.8. Příklad za rušení aditivním šumem pozadí.....</i>	<i>23</i>
<i>Obr. 2.9. Příklad nestacionárního šumu na pozadí.....</i>	<i>24</i>
<i>Obr. 2.10. Záznam hry na varhany, ve kterém je přítomen rušivý šum pozadí.....</i>	<i>25</i>
<i>Obr. 2.11. Detail úseku pro odhad PSD.....</i>	<i>25</i>
<i>Obr. 2.12. Spektrogram hudební ukázky s vybranými úseky s rozdílnou mírou stacionarity.....</i>	<i>28</i>
<i>Obr. 2.13. Graf typu Waterfall autokorelačních funkcí tří po sobě následujících segmentů z prvního úseku.</i>	<i>29</i>
<i>Obr. 2.14. Graf typu Waterfall autokorelačních funkcí tří po sobě následujících segmentů z druhého úseku.....</i>	<i>29</i>
<i>Obr. 2.15. Schématické znázornění hlavních oblastí zájmu z nové psychofyziky.....</i>	<i>30</i>
<i>Obr. 4.1. Základní koncept objektivního měření v metodě PEAQ.</i>	<i>37</i>
<i>Obr. 4.2. Pětistupňová hodnotící škála podle.....</i>	<i>38</i>
<i>Obr. 4.3. Pětistupňová hodnotící škály SDG a ODG.....</i>	<i>38</i>
<i>Obr. 4.4. Obecné blokové schéma metody PEAQ.....</i>	<i>39</i>
<i>Obr. 4.5. Detailní blokové schéma metody PEAQ.....</i>	<i>40</i>
<i>Obr. 5.1 Blokové schéma metody.....</i>	<i>44</i>
<i>Obr. 5.2 Lokální maskující práh, grafické naznačení lineární interpolace $LT_q(i_{sub})$ na $LT_q(k)$.....</i>	<i>47</i>
<i>Obr. 5.3 Odhad výkonové spektrální hustoty $P_{xx}(k)$ z jednoho náhodně vybraného segmentu $x(n)$ a maskující práh v tichu $LT_q(k)$.</i>	<i>48</i>
<i>Obr. 5.4 Vrcholy tónových a šumových komponent z $P_{xx}(k)$.....</i>	<i>48</i>
<i>Obr. 5.5 Zobrazení masek: a) tónových komponent, b) šumových komponent.....</i>	<i>49</i>
<i>Obr. 5.6 Zobrazení masek tónových i šumových komponent společně v jednom grafu.</i>	<i>49</i>
<i>Obr. 5.7 Lokální maskující práh $LT_i(k)$.....</i>	<i>50</i>
<i>Obr. 5.8 Závislost šířky Bärkového pásma na jeho poř. čísle.....</i>	<i>52</i>
<i>Obr. 5.9 Průběhy lokálních masek a výkonových spektrálních hustot vybraných průběhů před transformací frekvenční osy na Barkova pásma.</i>	<i>53</i>
<i>Obr. 5.10 Průběhy lokálních masek a výkonových spektrálních hustot vybraných průběhů po transformaci frekvenční osy na Barkova pásma.</i>	<i>54</i>
<i>Obr. 5.11 Modifikovaná varianta NMR.....</i>	<i>55</i>
<i>Obr. 5.12 Algoritmus výpočtu $NMR_i(b)$.....</i>	<i>56</i>
<i>Obr. 5.13 Záznam řeči.....</i>	<i>59</i>
<i>Obr. 5.14 Detail spektrogramu šumu pozadí, šum audio pásky.</i>	<i>59</i>
<i>Obr. 5.15 Výsledné charakteristiky záznamu řeči a šumu kazety při $SNR = 40$ dB.</i>	<i>60</i>
<i>Obr. 5.16 Příprava poslechového pracoviště.....</i>	<i>66</i>
<i>Obr. 5.17 Změřená přenosová charakteristika studiových sluchátek a zvukové karty.</i>	<i>66</i>
<i>Obr. 5.18 Intervalový odhad středních hodnot prahu slyšitelnosti šumu pozadí pro 1. část poslechového experimentu, tj. měření hudebních a řečových signálů.....</i>	<i>68</i>

Obr. 5.19 Výsledné subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro 1. část poslechového experimentu, tj. měření hudebních a řečových signálů.....	69
Obr. 5.20 Intervalový odhad středních hodnot prahu slyšitelnosti šumu pozadí pro 2. část poslechového experimentu, tj. měření umělých zvukových signálů	71
Obr. 5.21 Výsledné subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro 2. část poslechového experimentu, tj. měření umělých zvukových signálů	71
Obr. 5.22 Volba hraniční hodnoty $\min\{MSE(SNR_{OBJ})\}$ z 1. části poslechového experimentu, metoda PEAQ	74
Obr. 5.23 Volba hraniční hodnoty $\min\{MSE(SNR_{OBJ})\}$ z 1. části poslechového experimentu, metoda SpecNMR _{MAX}	74
Obr. 5.24 Volba hraniční hodnoty $\min\{MSE(SNR_{OBJ})\}$ z 2. části poslechového experimentu, metoda PEAQ	77
Obr. 5.25 Volba hraniční hodnoty $\min\{MSE(SNR_{OBJ})\}$ z 2. části poslechového experimentu, metoda SpecNMR _{MAX}	77
Obr. 6.1 Blokové schéma odhadu $\hat{x}(n)$ Wienerovým filtrem se zkreslujícím systémem $G(e^{j\Omega})$	80
Obr. 6.2 Blokové schéma odhadu $\hat{x}(n)$ Wienerovým filtrem bez zkreslujícího systému.	80
Obr. 6.3 Ziskové funkce pro Wienerův filtr H_W , amplitudové H_{MS} a výkonové H_{PS} spektrální odečítání	84
Obr. 7.1 Rozdíly v časo-frekvenčního rozlišení mezi pevnou a adaptivní segmentací.	86
Obr. 7.2 Adaptivní váhovací okno $w_A(n,q)$ pro $q = 1$ až 4.	87
Obr. 7.3 Hannova (nahore) a trojúhelníková (dole) adaptivní váhovací okna.	88
Obr. 7.4 Vstupní umělý nestacionární signál s výrazným šumem na pozadí.	88
Obr. 7.5 Pevná segmentace s dlouhým oknem, 4096 vzorků.	89
Obr. 7.6 Pevná segmentace s krátkým oknem, 256 vzorků.....	89
Obr. 7.7 Adaptivní segmentace s proměnou délkou oken.	89
Obr. 7.8 Detail adaptivní segmentace v okolí vzorku $n = 4.45 \times 10^4$	90
Obr. 7.9 Výstupní filtrované signály pro rozdílné typy segmentací a nastavení spektrogramů.....	91
Obr. 7.10 Algoritmus Greedy selection.....	93
Obr. 7.11 Segmentace dlouhotrvajícího umělého zvukového signálu $y(n)$ podle algoritmu Greedy selection.	94
Obr. 8.1 Vliv jevu spektrálního prosakování na hodnotu koeficientu lokální spektrální špičatosti C	96
Obr. 8.2 Snížení vlivu jevu spektrálního prosakování na koeficientu C pomocí techniky doplnění nulami.....	96
Obr. 8.3 Adaptivní segmentace podle původního algoritmu Greedy selection, tj. bez doplnění nulovými vzorky.....	97
Obr. 8.4 Adaptivní segmentace s doplněním nulovými vzorky.....	98
Obr. 8.5 Časové průběhy a amplitudové spektra náhodně vybraného segmentu o délce $L = 512$ vzorků (vlevo), a $L = 1024$ vzorků (vpravo) při $f_s = 8$ kHz.	99
Obr. 8.6 Algoritmus Maximum greedy selection.	100
Obr. 8.7 Greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, bez doplnění nulovými vzorky.	101
Obr. 8.8 Greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, s doplněním nulovými vzorky.	101
Obr. 8.9 Maximum greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, s doplněním nulovými vzorky.	101
Obr. 8.10 Ukázka segmentace signálu ze zvukového záznamu vážné hudby..	102
Obr. 8.11 Prahovací pravidla.....	105
Obr. 8.12 Prahovací pravidla základních ziskových funkcí.....	107
Obr. 8.13 Průběhy ziskových funkcí pro rozdílné násobky směrodatné odchylky σ_m	108
Obr. 8.14 Koláčkové grafy rozložení preferencí jednotlivých metod	111

Seznam tabulek

TAB. 2.1. SUBJEKTIVNÍ ŠKÁLA DS (DETECTABILITY OPINION SCALE)	31
TAB. 2.2. SUBJEKTIVNÍ ŠKÁLY MOS (MEAN OPINION SCORE) A SDG (SUBJECTIVE DIFFERENT GRADE)	32
TAB. 4.1. VELIČINY MOV ZE ZÁKLADNÍ VERZE METODY PEAQ	39
TAB. 4.2. VELIČINY MOV Z POKROČILÉ VERZE METODY PEAQ	40
TAB. 5.1. POUŽITÝ ZPŮSOB TRANSFORMACE NA BARKOVAPÁSMU	51
TAB. 5.2 SEZNAM HUDEBNÍCH A ŘEČOVÝCH ZVUKOVÝCH SIGNÁLŮ PRO 1. ČÁST EXPERIMENTU	62
TAB. 5.3 SEZNAM UMĚLÝCH ZVUKOVÝCH SIGNÁLŮ PRO 2. ČÁST EXPERIMENTU	63
TAB. 5.4 PARAMETRY POSLECHOVÉHO MATERIÁLU	63
TAB. 5.5 ZMĚŘENÉ SUBJEKTIVNÍ HODNOTY PRAHU SLYŠITELNOSTI ŠUMU POZADÍ PRO 1. ČÁST EXPERIMENTU	67
TAB. 5.6 ZMĚŘENÉ SUBJEKTIVNÍ HODNOTY PRAHU SLYŠITELNOSTI ŠUMU POZADÍ PRO 2. ČÁST EXPERIMENTU	70
TAB. 5.7 TABULKA S NAMĚŘENÝMI SUBJEKTIVNÍMI HODNOTAMI SNR_{OBJ} PRAHU SLYŠITELNOSTI ŠUMU POZADÍ PRO 1. ČÁST EXPERIMENTU ..	72
TAB. 5.8 TABULKA S NAMĚŘENÝMI OBJEKTIVNÍMI HODNOTAMI SNR_{OBJ} PRAHU SLYŠITELNOSTI ŠUMU POZADÍ PRO 1. ČÁST EXPERIMENTU PRO OBĚ POROVNÁVANÉ METODY	73
TAB. 5.9. OBJEKTIVNÍ HODNOTÍCÍ ŠKÁLA ODG	74
TAB. 5.10 TABULKA S NAMĚŘENÝMI SUBJEKTIVNÍMI HODNOTAMI SNR_{OBJ} PRAHU SLYŠITELNOSTI ŠUMU POZADÍ PRO 2. ČÁST EXPERIMENTU.	75
TAB. 5.11 TABULKA S NAMĚŘENÝMI OBJEKTIVNÍMI HODNOTAMI SNR_{OBJ} PRAHU SLYŠITELNOSTI ŠUMU POZADÍ PRO 1. ČÁST EXPERIMENTU PRO OBĚ POROVNÁVANÉ METODY	75
TAB. 8.1 TABULKA S VÝSLEDNÝM PREFERENČNÍM HODNOCENÍM PRO JEDNOTLIVÉ METODY	110
TAB. 8.2 DETAILNÍ TABULKA S PREFERENČNÍM HODNOCENÍM PRO JEDNOTLIVÉ METODY	111
TAB. C.1 HODNOTY PEARSONOVA KORELAČNÍHO KOEFICIENTU R_p A PŘÍSLUŠNÝCH T HODNOT	134
TAB. C.2 UPRAVENÉ HODNOTY PEARSONOVA KORELAČNÍHO KOEFICIENTU R_p A PŘÍSLUŠNÝCH T HODNOT	135
TAB. C.3 HODNOTY SPEARMANOVA KORELAČNÍHO KOEFICIENTU R_s A T HODNOTY PRO JEDNOTLIVÉ POSLUCHAČE	136
TAB. C.4 HODNOTY PEARSONOVA KORELAČNÍHO KOEFICIENTU R_p , T HODNOT A KRITICKÝCH HODNOT PRO HLADINU VÝZNAMNOSTI 1% A 5%	137
TAB. C.5. PRVNÍ ČÁST VÝSLEDNÉ TABULKY ZMĚŘENÝCH DAT Z 1. ČÁSTI POSLECHOVÉHO EXPERIMENTU	138
TAB. C.6. DRUHÁ ČÁST VÝSLEDNÉ TABULKY ZMĚŘENÝCH DAT Z 1. ČÁSTI POSLECHOVÉHO EXPERIMENTU	139

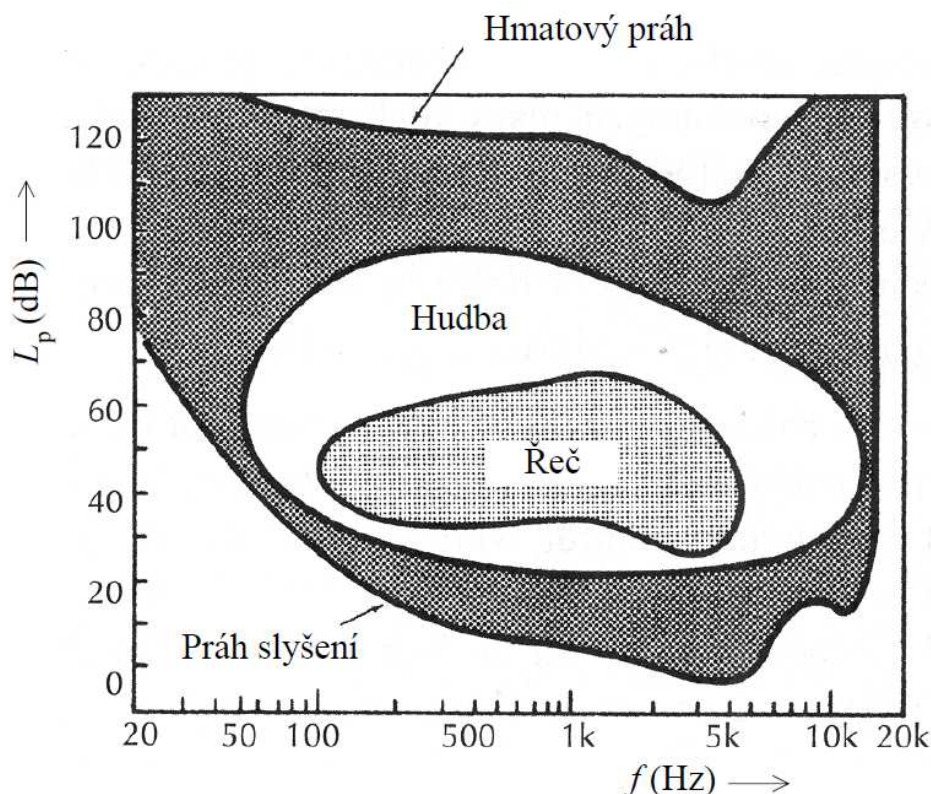
1 Úvod

V oblasti zpracování digitálního zvukového signálu nebyl v současné době stále uspokojivě vyřešen problém šumu na pozadí při přenosu nebo jeho záznamu. Potlačení šumu na pozadí je možno řešit mnoha způsoby, z více úhlů pohledu na řešenou problematiku a srozdílnou mírou úspěšnosti. Šum na pozadí v zvukovém signálu je z hlediska posluchače téměř vždy chápán jako prvek snižující výslednou poslechovou kvalitu. V případě řečového signálu je problém potlačení šumu spojován se systémy na zvýraznění tohoto řečového signálu (angl. *speech enhancement*) [25][26][34]. Rozvoj těchto systémů nastal především s rozvojem moderních telekomunikačních sítí v posledních letech, kde tyto systémy mají za úkol zvyšování kvality a srozumitelnosti přenášeného řečového signálu. Jejich přínos spočívá zejména ve zlepšení kvality hlasové komunikace v zarušených prostředích [25][26][77].

Další skupina zvukových signálů, kde je žádoucí potlačovat šum na pozadí, jsou hudební signály. Hudební signál je takový signál, který je vytvořen hudebními nástroji včetně lidského hlasového ústrojí a současně je tento signál nositelem hudební informace [74][75]. Systémy na potlačení šumu v hudebním signálu řeší, podobně jako systémy pro potlačení šumu v řečovém signálu, problém zvýšení kvality tohoto hudebního signálu a jeho záznamu. Příkladem činnosti, kde vzniká potřeba nežádoucí šum na pozadí, je digitalizace a následná restaurace starších analogových záznamů, viz např. v nedávně době provedené digitální restaurování záznamů z janáčkovských fonografických válců ze sbírek etnologického ústavu Akademie věd České republiky [40][41]. Podobně mohou být tyto systémy použity při restaurování zvukových stop filmových titulů různých filmových archivů. Další oblastí zájmu kde mohou být využity tyto systémy je zvyšování kvality různě nekvalitně pořízených amatérských zvukových záznamů z koncertů, hudebních představení, kina, rodinných oslav a podobně. Srozdvojem audio-vizuální záznamové techniky bývají tyto typy záznamů stále častější.

V současné době existuje poměrně velké množství různých systémů na potlačení šumu, které jsou zaměřeny na řečový signál. Množství systémů zaměřených na signál hudební je však mnohem menší, což je dáno zejména tím, že hudební signál je oproti řečovému mnohem komplexnějším signálem. Hudební signál totiž vykazuje větší variabilitu a subjektivní vnímání obou typů signálů je také odlišné [23]. Přizpůsobení systémů a algoritmů pro potlačení šumu v řečovém signálu na hudební signál je problematický a mnohdy zcela nemožný. Většinu systémů na potlačování šumu v řečovém signálu nelze jednoduše použít na hudební signál.

V části fyzikálního neboli podnětového kontinua [44] spočívají tyto rozdíly především v tom, že hudební signály mají širší frekvenční spektrum a dosahují větších úrovní hladin akustického tlaku SPL (Sound Pressure Level), viz Obr. 1.1, a dosahují vyšší dynamiky oproti řečovému signálu [74].



Obr. 1.1. Frekvenční oblasti slyšení lidského sluchového ústrojí.

Řečový i hudební signál lze považovat za nestacionární signály. Hudební signál ale bývá ve většině případů daleko více časově proměnlivý než řečový signál. To je dáno tím, že akustickým zdrojem je u řečového signálu hlasové ústrojí zpravidla pouze jednoho mluvčího a časová proměnlivost je zde omezena fyziologickými vlastnostmi hlasového ústrojí daného (konkrétního) mluvčího. Pro účely analýzy a syntézy řečového signálu se uvádí délka signálu 10 – 20 ms, pro kterou se řečový signál uvažuje jako stacionární signál [26]. Pro hudební signály toto ovšem neplatí, zvláště v případě, kdy obsahuje zvuky produkované více než jedním hudebním nástrojem či hlasy současně [1][2][12].

Dalším podstatným rozdílem v přístupu k popisu vlastností hudebního signálu je možnost použití specifických vlastností, jako jsou melodie, rytmus a harmonie. Melodie je organizovaná sekvence tónů s důrazem na vyjádření "hudební myšlenky", rytmus bývá charakterizován jako časová složka hudby a harmonie definuje stavbu, kladení a spojování akordů. Na zachování těchto vlastností je třeba brát velký ohled při potlačování šumu v hudebních signálech [44][74][75].

Důležitým faktem, který ztěžuje návrh systémů pro potlačení šumu je skutečnost, že v současné době neexistuje žádná objektivní metoda, která by dokázala měřit úroveň slyšitelného šumu na pozadí například v již pořízených záznamech, tedy veličinu, jež je velmi důležitá pro návrh těchto systémů v oblasti řečové i hudební. U takových záznamů je možné pouze realizovat subjektivní vyhodnocení v podobě poslechových testů, avšak pouze za cenu vyšší časové i finanční náročnosti vyplývající z nutnosti zajištění kvalitní a spolehlivé testovací skupiny posluchačů a provádění statistického vyhodnocení těchto poslechových testů. Na poli objektivních metod existují pouze metody pro hodnocení celkové kvality u řečových signálů, nebo pro hodnocení celkové míry degradace vůči referenčnímu signálu pro případ hudebních signálů [57][61][62]. Bohužel tyto

objektivní metody lze využít pro návrh a optimalizaci systémů pro potlačení šumu pouze v omezené míře. Lze například vysledovat závislost změny určitého parametru na výsledné kvalitě, ale již nelze například určit, zda je změna kvality zapříčiněna vyšším stupněm potlačení šumu na pozadí, nebo nižší mírou zkreslení užitečného signálu a podobně. Dalším problémem je nedostatečná přesnost těchto metod [57].

Stěžejní částí a tedy i hlavní motivací popisované disertační práce je nalezení objektivní metody pro měření velikosti slyšitelného šumu na pozadí v hudebním signálu, která posouvá současný stav poznání v oblasti návrhu a optimalizace systémů pro potlačení šumu v hudebním signálu.

2 Přehled současného stavu problematiky

2.1 Vymezení použitých pojmů

Vzhledem k nejednoznačnosti významů některých používaných odborných pojmů a názvosloví, budou tyto pojmy zde nyní vymezeny. Zmíněná nejednoznačnost vychází z rozdílných definic, které jsou používány v monografiích [44][74][75] a z české státní normy [17]. Proto pokládám za důležité zde všechny tyto pojmy explicitně definovat. Tato kapitola si klade za cíl předejít případným nedorozuměním při čtení dalších kapitol. Dále je potřeba poznamenat, že pro potřeby této disertační práce bude poněkud “netypicky” definovaný pojem *umělý zvuk*, viz seznam níže.

Definice používaných pojmů:

Zvuk – Mechanické vlnění v látkovém prostředí, které je schopno vyvolat sluchový vjem.

Hudba – Hudba je organizovaný systém zvuků, který je vytvořen hudebními nástroji a to včetně lidského hlasového ústrojí a současně je nositelem hudební informace.

Řeč – Řeč je artikulovaný a zvukový projev člověka, který má věcný obsah a lze jej vyjádřit písmem.

Umělý zvuk – Umělé vytvářené zvuky, které jsou vytvářené pro účely analýzy, syntézy a testování zvukového systému a nemají charakter hudby nebo řeči. Jako příklady umělých zvuků lze uvést například programově vygenerované směsi tónů, různé typy šumů (bílý šum, růžový šum), různé deterministické signály typu pila, obdélník a podobně.

Na základě těchto definic budou nyní definovány signály výše uvedených zvuků. Tedy například *hudební signál* bude definován jako reprezentace hudby v jiné než původní formě a to od jejího zaznamenání až po reprodukci. Podobně bude definován např. *zvukový signál* jako reprezentace zvuku (zvukového vlnění) v jiné než původní formě od jejího zaznamenání až po reprodukci. Podobným způsobem bude definován *řečový* a *umělý zvukový signál*.

Další definice používaných pojmů:

Šum – Šum je nežádoucí a rušivý zvuk.

Hluk – Hluk je jakýkoliv zvuk, který vyvolává nepříjemný nebo rušivý vjem a má škodlivý účinek. Hlukem jsou označovány ty šumy, které mají významný rušivý vjem, jako je například hluk automobilu.

2.2 Potlačování aditivního šumu pozadí ve zvukových systémech.

2.2.1 *Restaurování poškozených nebo nekvalitně pořízených zvukových záznamů*

Tato kapitola si klade za cíl seznámit čtenáře s nejčastější se vyskytujícími typy rušení nebo zkreslení, které se vyskytují při restaurování poškozených nebo nekvalitně pořízených řečových a hudebních záznamů. Stěžejní část kapitoly se přitom bude zaměřovat na zkreslení tzv. šumem na pozadí (angl. *background noise*). Právě o tomto typu zkreslení, resp. o metodách jeho objektivního měření a potlačování, pojednává tato disertační práce. Tematicky tato kapitola vychází z publikací [11][46][66].

Není zde snaha popsat kompletní seznam většiny v praxi se vyskytujících rušení nebo zkreslení, takovýto seznam by byla nad rámec této disertační práce, ale provést pouze základní dělení s cílem seznámit s touto problematikou. Ani v rámci jedné skupiny zkreslení či rušení není uveden kompletní výčet všech možných zkreslení či rušení, který spadá do této skupiny, nýbrž je vždy představen pouze jeden ilustrační příklad.

Dále v tomto výčtu zcela chybí tzv. "charakteristická zkreslení" způsobená konkrétním typem použitého zvukového systému, např. pro účely kvantování a kódování při digitalizaci, nebo pro kompresi zvukového signálu. Příkladem takového typu charakteristického zkreslení může být kvantizační zkreslení způsobené pulzně kódovou modulací (PCM), nebo zkreslení způsobené kompresním formátem MPEG [16][31] a podobně.

V neposlední řadě je potřeba zmínit skutečnost, že se v praxi spíše vyskytují kombinace více typů rušení a zkreslení současně. Z toho důvodů je také potřeba použít více typů technik k potlačení těchto rušení a zkreslení. Lze ale tvrdit, že šum pozadí je jeden z nejčastěji se vyskytujících typů rušení, který je často přítomný u většiny restaurovaných zvukových záznamů [46][66].

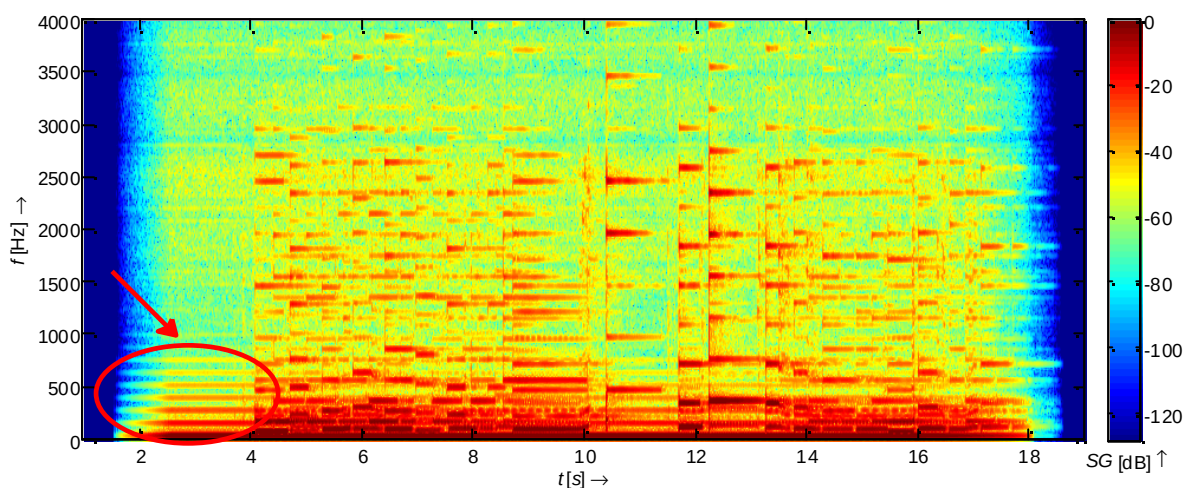
Názvy, které se používají pro většinu typů níže popisovaných zarušení či zkreslení, pocházejí z anglického jazyka. Následný překlad těchto názvů do češtiny bývá často problematický. Významově se totiž můžou lišit v závislosti na použitém jazyce. Z těchto důvodů budou dále v textu používány převážně jejich anglické názvy.

První část použitého zvukového materiálu z této kapitoly pochází z materiálů pro výuku softwarového nástroje iZotope RX™ [46][66], který se využívá pro restaurování zvukových záznamů. A druhá část použitého zvukového materiálu pochází z databáze SQAM [72], která je primárně určena jako materiál pro evaluaci zvukové kvality.

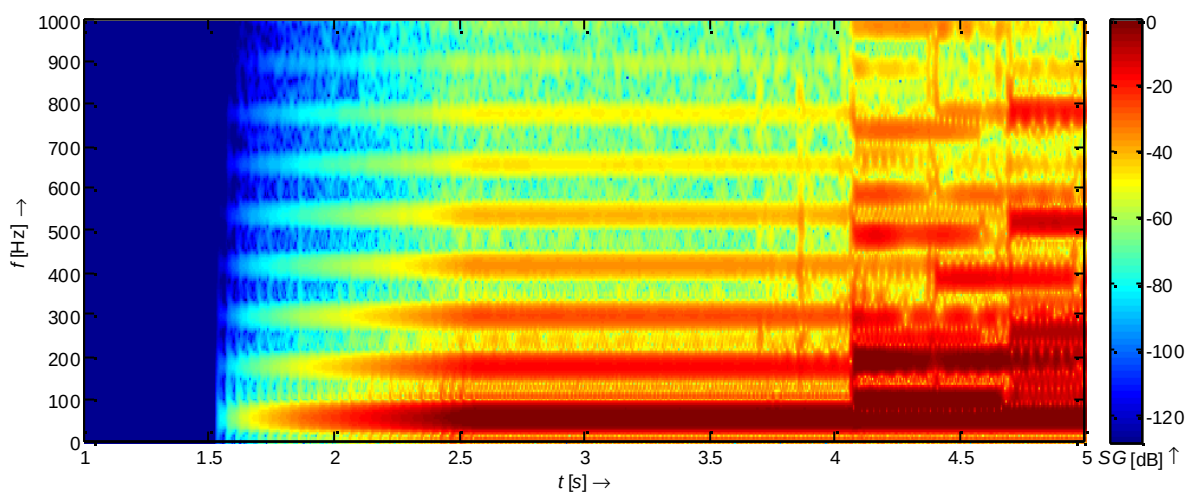
2.2.1.1 Harmonické rušení

Příkladem harmonického rušení může být *Hum*. Slovo *Hum* lze překládat jako hukot, nebo hučení. Je jím označováno rušení vzniklé elektromagnetickou indukcí harmonického signálu z rozvodné sítě, z AC transformátorů, popřípadě ze zemnicích smyček signálových spojení ve zvukových systémech. Charakteristickým rysem tohoto rušení je jeho harmonická struktura, kde se vyskytují násobky základního kmitočtu 50 nebo 60 Hz.

Na Obr. 2.1 je zobrazen spektrogram záznamu hry na kytaru. V tomto zvukovém záznamu je extrémně výrazné rušení typu *Hum* z použitého zesilovače při záznamu [66]. Dále je na tomto spektrogramu šipkou naznačena oblast s nejviditelnějším výskytem tohoto rušení. *Hum* je přítomný v celém signálu, v této úvodní části spektrogramu je pouze nejviditelnější, viz Obr. 2.2. Lze zde vidět zmiňovanou harmonickou strukturu tohoto rušení. Kmitočty základní harmonické je na 60 Hz, vyšší harmonické se vyskytují na celočíselných násobcích tohoto základního kmitočtu.



Obr. 2.1. Spektrogram záznamu hry na kytaru s přítomností harmonického rušení typu *Hum*. Šipkou je zde naznačena oblast s nejvýraznější oblastí výskytu tohoto rušení, *Hum* se zde vyskytuje po celou dobu trvání signálu [66].



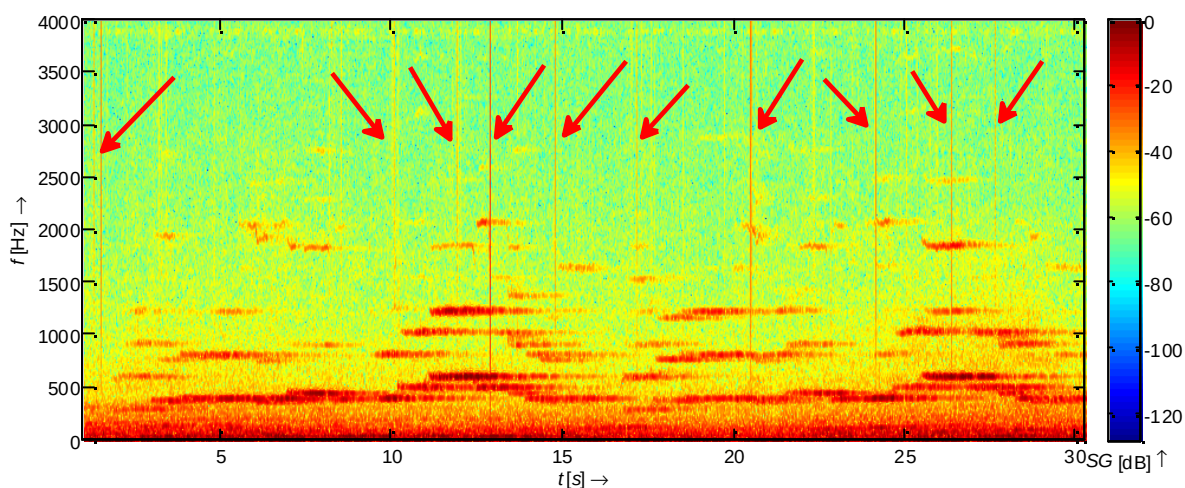
Obr. 2.2. Detail spektrogram harmonického rušení typu *Hum* [66].

Dalším příkladem harmonického rušení může být například *Buzz* (bzukot). Podobně jako Hum má signál tohoto rušení harmonickou strukturu. Rozdílem je, že základní frekvence tohoto rušivého signálu se vyskytuje na vyšších kmitočtech. Zdrojem tohoto typu rušení může být například zářivkové osvětlení.

2.2.1.2 Impulsní rušení

U impulsního rušení je užitečný signál zarušen signálem s velmi krátkou dobou trvání a velmi nepravidelným výskytem. Na spektrogramech lze tento rušivý signál vyzorovat jako úzké vertikální přímký nejčastěji přes celé uvažované frekvenční spektrum. Tento typ rušení je obvyklý na záznamech provedených na gramofonových deskách, nebo jako důsledek chyb při záznamu DAW systémy s nevhodně nastavenými velikostmi vyrovnávacích pamětí (bufferů) a podobně [46].

Na Obr. 2.3 je zobrazen spektrogram 30 sekund dlouhého záznamu pěveckého sboru z LP vinyl desky. Tento záznam byl provedený v roce 1965 na univerzitě v Cambridge [46]. Kromě výrazného šumu pozadí z nahrávací kaple z této univerzity jsou zde slyšitelné zejména zvukové artefakty v podobě lehkého klikání z LP vinyl desky. Tyto artefakty jsou na Obr. 2.3 vyznačeny šipkami.

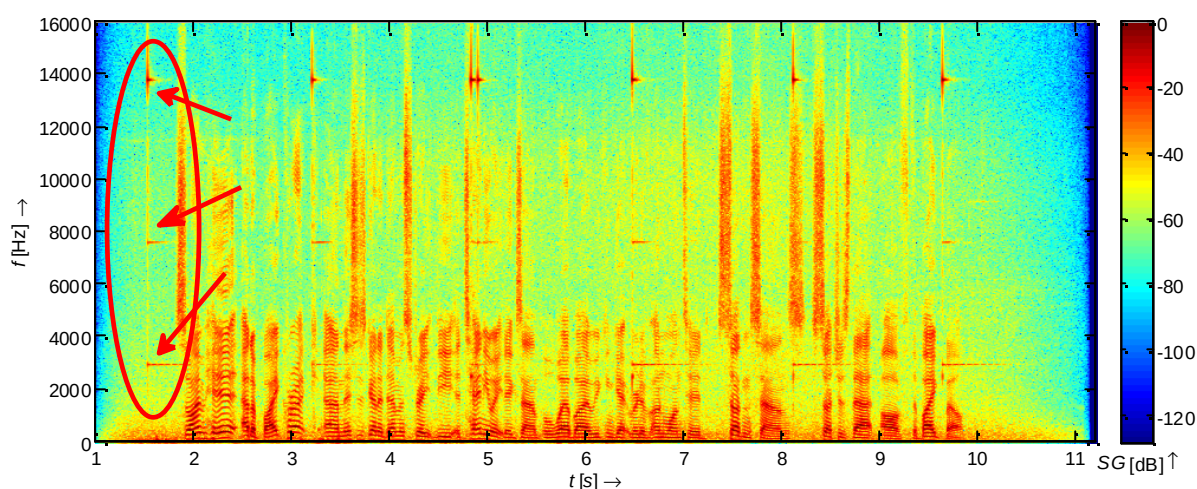


Obr. 2.3. Spektrogram záznamu pěveckého sboru z LP vinyl desky z roku 1965 [46]. Šipkami jsou vyznačeny komponenty impulsního rušení.

2.2.1.3 Intermitentní rušení

Intermitentní (ve významu občasné) rušení nastává, když se na pozadí užitečného signálu občasné vyskytuje jiný rušivý zvuk, který pochází nejčastěji z jiného akustického zdroje. Na následujícím spektrogramu, Obr. 2.4, je zobrazen záznam řečového signálu anglicky mluvícího mluvčího. Na pozadí tohoto záznamu je slyšet občasné cinkání skleniček. Spektrogram z Obr. 2.4 zobrazuje pouze první kanál.

Jednotlivé komponenty tohoto rušivého zvuku jsou na spektrogramu označeny šipkami, na tomto spektrogramu viditelné jako krátké horizontální přímkové. A tyto tři komponenty reprezentují tři harmonické signály na kmitočtech 3018 Hz, 7665 Hz a 13840 Hz [46]. Na spektrogramu z Obr. 2.4 je vyznačen pouze první výskyt tohoto rušivého zvuku, další 4 výskyty již nejsou nijak vyznačeny.

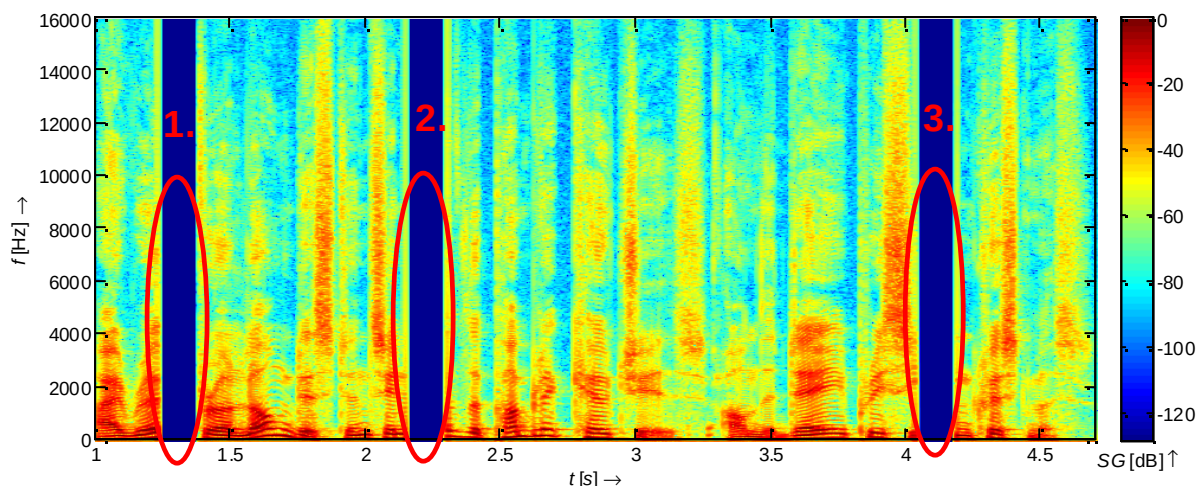


Obr. 2.4. Spektrogram záznamu řečového signálu anglicky mluvícího mluvčího a zvuku cinkání, tj. nežádoucího rušivého zvuku na pozadí. Šipkami jsou vyznačeny jednotlivé komponenty prvního výskytu tohoto rušivého zvuku [46].

2.2.1.4 Mezery a výpadky

Při záznamu nebo přenosu zvukového signálu bývá často tento signál degradován jeho výpadky nebo mezery. Tyto výpadky nebo mezery mohou být způsobeny například důsledkem ztrát datových paketů v multimediálních sítích, jako důsledek poškozených segmentů na digitálních zvukových nosičích, nebo jako důsledek poškozených oblastí na analogových záznamových médiích.

Příkladem takovéto degradace může být zvukový záznam ze spektrogramu z Obr. 2.5. Na tomto spektrogramu je zobrazen řečový signál anglicky mluvícího mluvčího z databáze PEAQ [72]. Tento signál obsahuje tři 150 ms dlouhé mezery. Tyto mezery byly vytvořeny uměle a mají právě za cíl simulovat tento typ degradace.



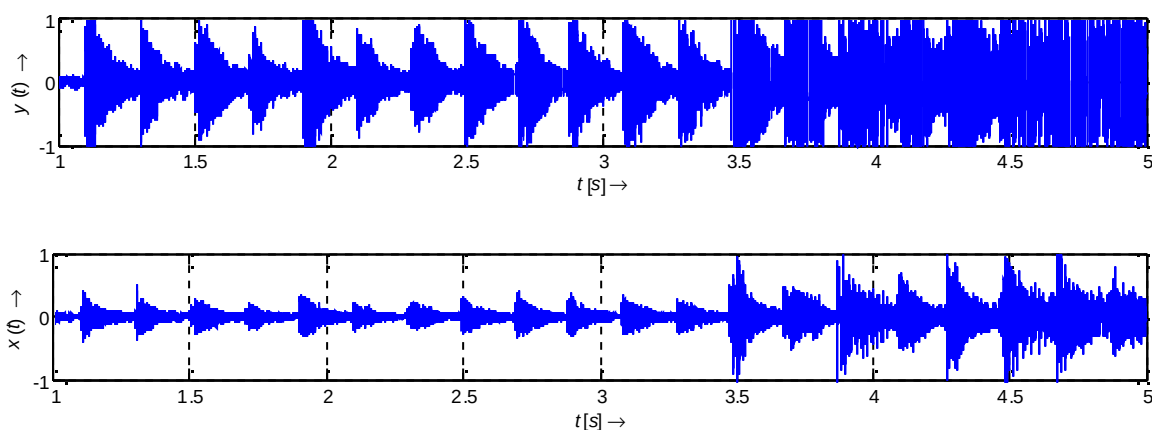
Obr. 2.5. Spektrogram řečového záznamu anglicky mluvícího mluvčího [72], tento záznam obsahuje tři mezery délky 150 ms.

Nelineární zkreslení

Mezi nelineární třídu zkreslení užitečného signálu může být například zařazen tzv. *clipping*, toto zkreslení bývá česky někdy nazýváno jako “zkreslení přebuzením”, nebo také “zkreslení oříznutím”.

Clipping například nastává, když je signál zesílen nad své možné maximum. To může například nastat, když je zesilovač při záznamu v saturaci a snaží se dodávat napětí, nebo proud za jeho výkonovými možnostmi. Příliš hlasitě zpívající zpěváci popř. hrající bubeníci bývají také často zdrojem tohoto zkreslení [46].

Na Obr. 2.6 (dole) je zobrazen časový průběh vstupního nezkresleného signálu $x(n)$ a na Obr. 2.6 (nahore) je zobrazen časový průběh výstupního zkresleného signálu $y(n)$.



Obr. 2.6. Clipping, časové průběhy vstupního nezkresleného signálu $x(t)$ (dole) a výstupního zkresleného signálu $y(t)$ (nahore) [72]

Šum na pozadí

Šumem na pozadí (angl. *background noise*), nebo také šum pozadí, je označované rušení, kdy se k užitečné složce zvukového signálu superponuje náhodný rušivý šum. To může nastat například při pořizování zvykových záznamů, kdy je kromě užitečného signálu snímán také okolní hluk nebo šum, například z ventilátorů a podobně. Dalším příkladem šumu na pozadí může být tzv. *Hiss*. Hiss (sykot) lze charakterizovat jako ostrý zvuk, podobný dlouho znějící hlásce s. Hiss je vysokofrekvenční šum pozadí, který bývá přítomný na analogových magnetických páskách a je způsoben použitými rozměry magnetických částic při jejich výrobě [46].

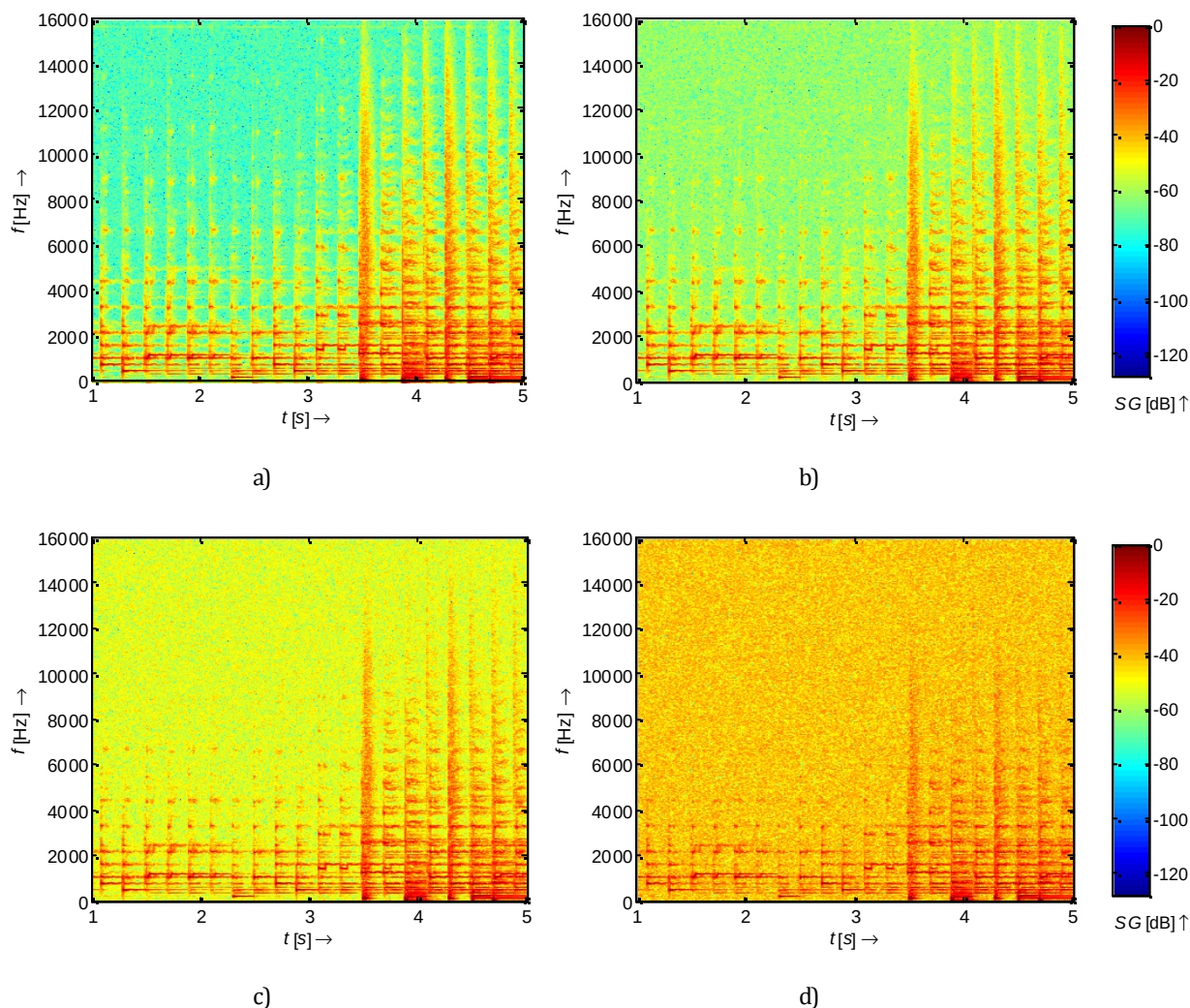
Z hlediska stacionarity lze tyto rušivé šumy signály rozdělit na stacionární a nestacionární a z hlediska rozložení výkonové spektrální hustoty je lze dále rozdělit na signály s rovnoměrně a nerovnoměrně rozloženou výkonovou spektrální hustotou. Zarušení šumem pozadí bývá charakteristické tím, že užitečný signál popředí má obvykle vyšší výkon než je výkon šumu pozadí, proto je také tento užitečný signál subjektivně vnímán jako hlasitější a rušivý šum je slyšitelný “na pozadí” tohoto užitečného signálu. Tento sluchový vjem reflektuje skutečnost, že ve vyšších frekvenčních pásmech, šum pozadí výkonově převyšuje užitečný signál popředí. Na nízkých kmitočtech se tento rušivý šum vyskytuje také, ale bývá maskován a proto se pro posluchače hůře postřehnutelným [23].

Pro následující výklad bude zaveden aditivní model zkreslení, který je definovaný jako $y(n) = x(n) + m(n)$, kde $x(n)$ je užitečný signál popředí, $m(n)$ je (aditivní) šum pozadí a je $y(n)$ výstupní aditivním šumem pozadí zarušený signál. Symbol n je zde použit pro vyjádření diskrétního času. Pro stanovení rozdílů úrovní mezi výkony obou signálů bude použita veličina SNR definována jako

$$SNR = 10 \log \left(\frac{\sum_{n=0}^{N_c-1} x^2(n)}{\sum_{n=0}^{N_c-1} m^2(n)} \right), \quad (2.1)$$

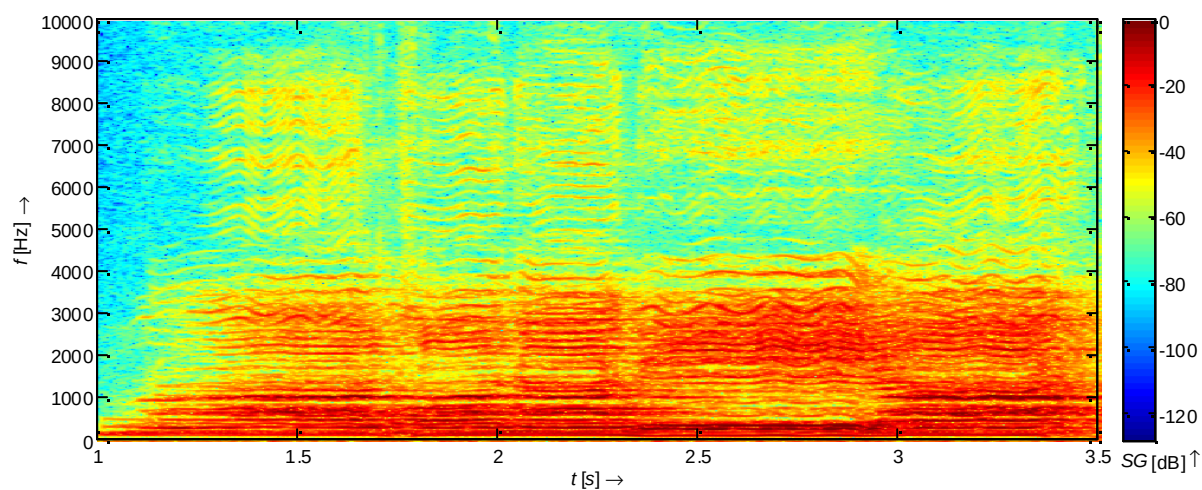
kde N_c je celková délka obou signálů, tj. platí, že oba signály mají stejnou délku.

Na sérii spektrogramů z Obr. 2.7a-d) je zobrazen příklad zarušení bílým šumem pozadí pro různé hodnoty SNR (2.1) v rozsahu 10 až 40 dB. Jako užitečný signál popředí byla na tomto příkladu použita 5 s dlouhá ukázka popové skladby od skupiny ABBA z databáze [72]. Zvyšující se úroveň šumu pozadí, tj. snižující se hodnota SNR , je na těchto spektrogramem patrná změnou barvy zejména v oblastech vyšších kmitočtů. Tento barevný přechod je od modro-zelené barvy z Obr. 2.7a) až po žluto-červenou barvu z Obr. 2.7d).

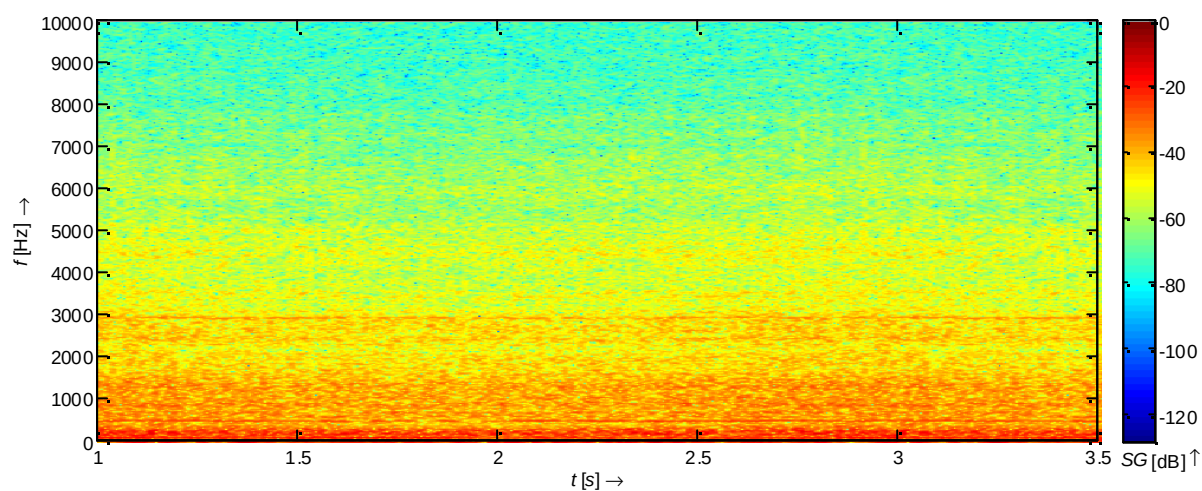


Obr. 2.7. Zarušení bílým šumem pozadí pro různé hodnoty SNR : a) $SNR = 40$ dB, b) $SNR = 30$ dB, c) $SNR = 20$ dB, d) $SNR = 10$ dB.

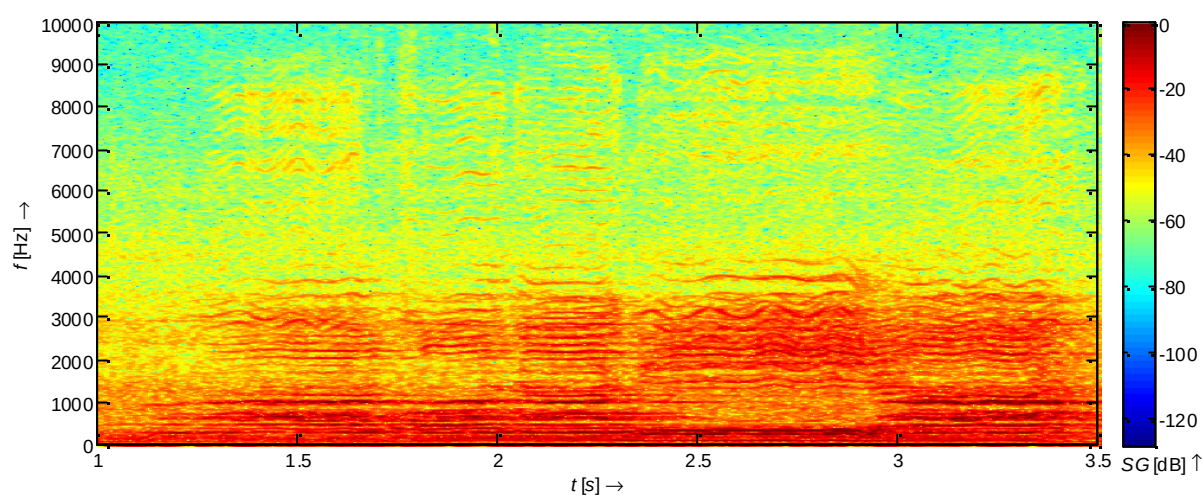
Zarušení bílým šumem ze spektrogramu z Obr. 2.7a-d) je příklad zarušení šumem srovnoměrně rozloženou výkonovou spektrální hustotou. Prakticky se ovšem vyskytují spíše případy šumu pozadí s nerovnoměrně rozloženou výkonovou spektrální hustotou, viz například šum ventilátoru z Obr. 2.8a-c). Na těchto obrázcích je na spektrogramech zobrazeno rušení zvukem z ventilátoru. Na Obr. 2.8a) je nejprve zobrazen spektrogram užitečného signálu kterým je 3,5 s dlouhý záznam sborového zpěvu (vokální kvartet) z databáze [72]. Tento signál je zarušen šumem z ventilátorů, Obr. 2.8b). Lze vidět, že výkonová spektrální hustota tohoto šumu není konstantní přes všechny kmitočty, ale je proměnná a největší část výkonu leží v pásmu do 3 kHz, výkonová špička leží v pásmu do cca 400 Hz. Na posledním spektrogramu z obrázky Obr. 2.8c) je zobrazen jejich součet pro $SNR = 15$ dB.



a)



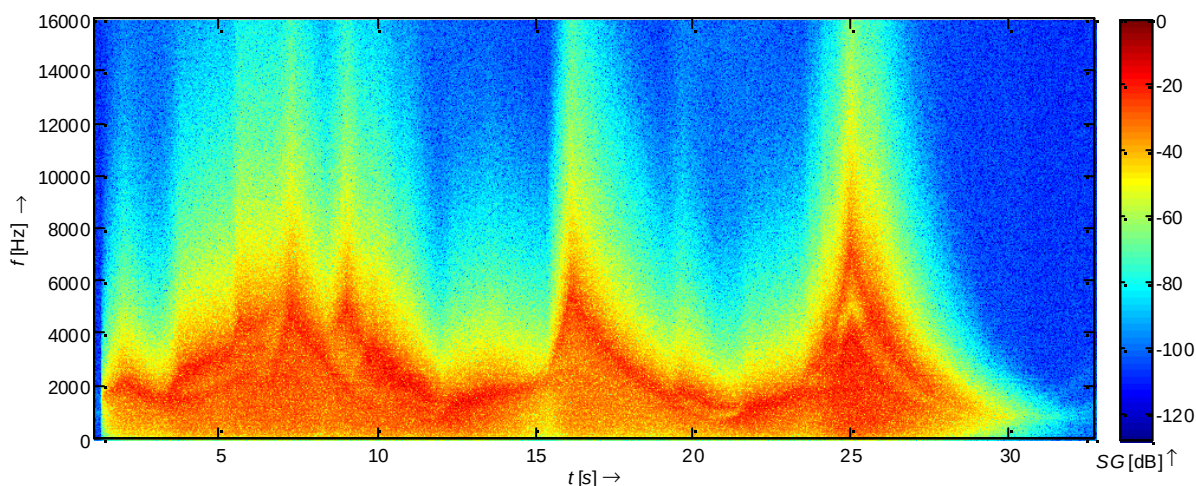
b)



c)

Obr. 2.8. Příklad zarušení aditivním šumem pozadí: a) užitečný signál -záznam sborového zpěvu [72], b) šum pozadí – šum z ventilátoru, c) zarušený signál při $SNR = 15$ dB.

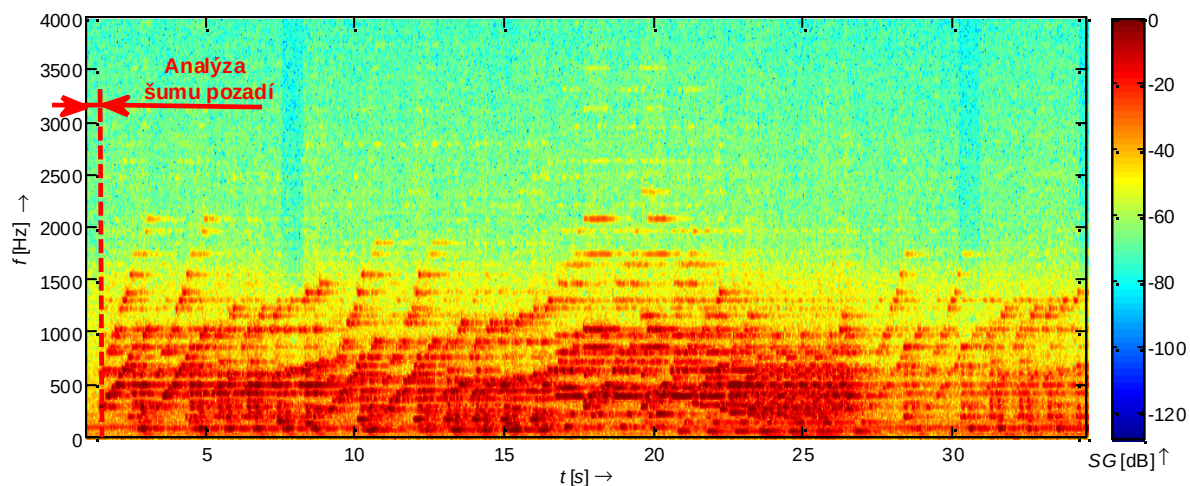
Na výše uvedených spektrogramech byly zobrazeny spektra zarušených signálů stacionárním šumem na pozadí. Velmi často se ale také vyskytuje rušení nestacionárním šumem na pozadí, tj. rušení náhodným signálem s normálním rozdělením, jehož statistické vlastnosti (střední hodnota a rozptyl) jsou časově proměnlivé. Zdrojem takového typu rušení může být například vítr. Příklad takového rušivého signálu je zobrazen na spektrogramu z Obr. 2.9.



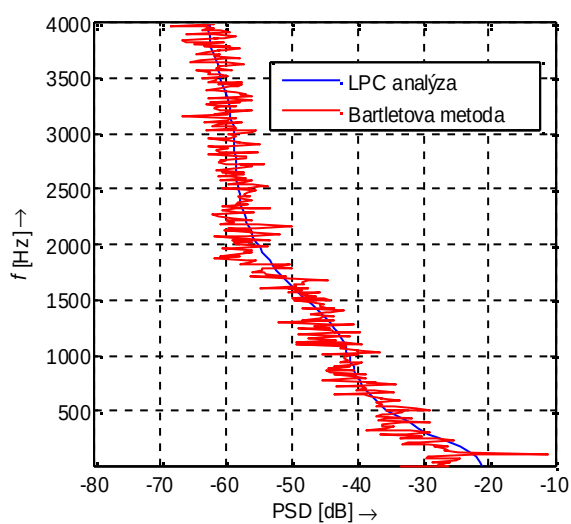
Obr. 2.9. Příklad nestacionárního šumu na pozadí, zvukový záznam proudění vzduchu s proměnlivou rychlostí.

Při restaurování zvukových záznamů, které jsou zarušeny šumem pozadí, je často potřeba provést odhad výkonové spektrální hustoty (PSD, angl. *power spectral density*) tohoto šumu pozadí. V případě stacionárního typu šumu pozadí, lze tento odhad provést v tzv. “tichých” úsecích, tj. v úsecích bez řečové nebo hudební aktivity. V případě řečových signálů lze k detekci takovýchto úseků s výhodou použít detektory řečové aktivity (VAD, angl. *voice activity detector*) a u hudebních signálů lze tento odhad provést například z krátkého úseku před začátkem skladby. Tento způsob odhadu bude demonstrován na příkladu z Obr. 2.10 a Obr. 2.11.

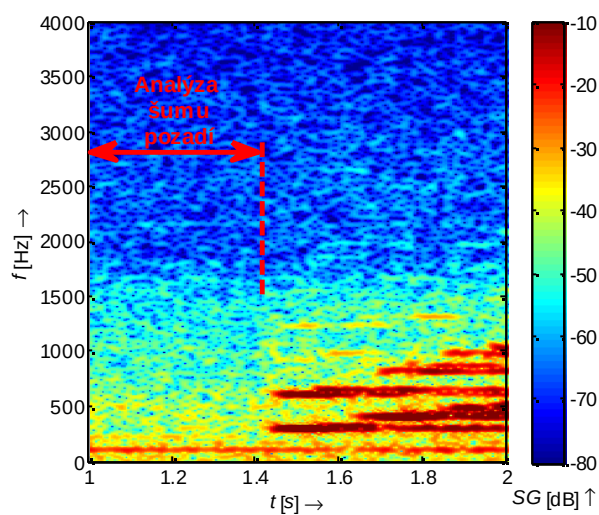
Na spektrogramu z Obr. 2.10 je zobrazen záznam hry na varhany. Tento záznam byl nahrán na analogovou $\frac{1}{4}$ magnetickou pásku. Při záznamu nebyl použit žádný systém pro redukci nebo potlačení šumu [46]. Hlavním problémem tohoto záznamu je přítomnost nízkourovňového šumu na pozadí. Odhad výkonové spektrální hustoty tohoto šumu pozadí může být proveden z krátkého úseku na začátku skladby, tento úsek trvá přibližně 1,4 s a je vyznačen šipkou na tomto spektrogramu, detail toho úseku je zobrazen na Obr. 2.11b). V dalším grafu z Obr. 2.11b) jsou zobrazeny průběhy odhadů PSD ze stejného úseku, ale vypočteny jinými metodami.



Obr. 2.10. Záznam hry na varhany, ve kterém je přítomen rušivý šum pozadí, šipkou je zde označen úsek “ticha” ze kterého bude proveden odhad PSD [46]



a)



b)

Obr. 2.11. Detail úseku pro odhad PSD: a) pomocí LPC analýzy a Bartletovy metody b) detail spektrogramu.

2.2.2 Systémy na potlačení aditivního šumu pozadí

Tato kapitola je zaměřena na shrnutí současného stavu problematiky systémů pro potlačování šumu pozadí ve zvukových systémech. V případě řečových signálů bývají tyto systémy spíše nazývány jako systémy pro zvyšování kvality řečových signálů. Výklad v této kapitole se bude zaměřovat na jednokanálové metody pro potlačení aditivního šumu.

Uvažovaný model zkreslení užitečného signálu je popsán rovnicí

$$y(n) = x(n) + m(n), \quad (2.2)$$

kde $x(n)$ je originální signál, $m(n)$ je aditivní šum a $y(n)$ je zkreslený signál. Symbol n je zde použit pro vyjádření diskrétního času, který je tvořen množinou $\{nT, n \in \mathbb{Z}\}$, kde T je perioda vzorkování.

Klasickým přístupem k filtraci šumem zarušeného signálu je tzv. Wienerova filtrace, zavedena N. Wienerem [79]. Jedná se o klasickou metodu odhadu originálního signálu na základě měřeného signálu za přítomnosti náhodného šumu [34]. Kritériem optimality tohoto odhadu je minimalizace střední kvadratické odchylky mezi odhadovaným signálem a jeho originálem a to pro všechny možné realizace obou signálů, oba signály jsou totiž chápány jako náhodné signály. Ve své klasické podobě lze Wienerovu filtraci použít pouze na stacionární signály. Hudební a řečové signály ovšem nelze považovat za stacionární signály. Měřený dlouhotrvající signál je v takovém případě rozdělen na krátké úseky a na tyto úseky je aplikována Wienerova filtrace s využitím krátkodobé Fourierovy analýzy. Takovýto postup je nazýván krátkodobá Wienerova filtrace [19].

Alternativou k Wienerové filtraci jsou techniky spektrálního odečítání. Jedná se o techniky amplitudového a výkonového spektrálního odečítání. Podobně jako Wienerova filtrace, jsou tyto techniky založené na odhadu originálního signálu na základě měřeného signálu za přítomnosti šumu. Jen typy odhadu jsou jiné. Výkonové spektrální odečítání se snaží odhadnout výkonové spektrum originálního signálu a amplitudové spektrální odečítání se snaží zase odhadnout amplitudové spektrum. Informace o fázi není odhadována, ale nejčastěji je použita fáze z měřeného signálu [34][19]. Wienerova filtrace, výkonové a spektrální odečítání jsou si tedy velmi blízké. Zobecněním těchto tří metod do jednoho společného schématu byl zavedený tzv. parametrický Wienerův filtr [70], někdy je toto schéma označováno jako parametrické odečítání [22]. Nevýhodou těchto metod je skutečnost, že ve výstupních signálech bývá velmi výrazný efekt hudebního šumu. Tento zvukový artefakt je pro posluchače velmi nepříjemným a rušivým vjemem. Hudební šum zde nastává především díky náhodnému charakteru zpracovávaných signálů a nepřesným odhadům výkonových spektrálních hustot těchto signálů [34].

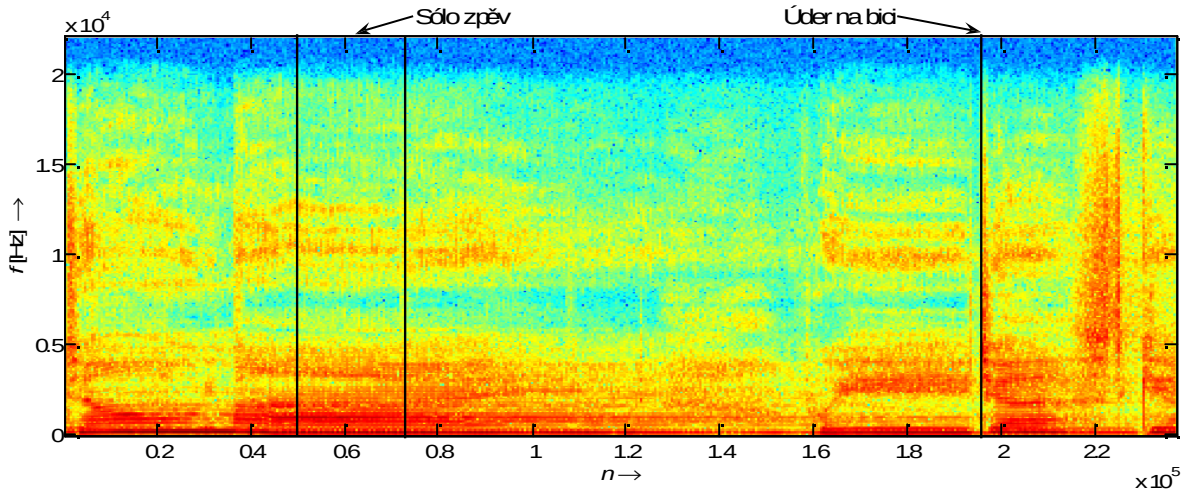
Příkladem systému, který je zaměřen výhradně na potlačení šumu pozadí v řečových signálech je metoda RASTA (RelAtive SpecTrAl processing) [27][51][38]. Tato metoda byla navržena pro jednotky handsfree mobilních telefonů firmu Motorola. Podstata metody spočívá ve filtraci rychlých a pomalých změn v časových trajektoriích logaritmických krátkodobých spekter řeči [14]. Spektrum časových trajektorií bývá také označováno jako modulační spektrum. V některých variantách jsou filtrovány časové trajektorie krátkodobých kepler namísto spekter [45]. Princip metody vychází z fyziologických schopností lidského řečového ústrojí, konkrétně z rychlosti změny polohy svalů mluvidel při pronášení promluvy. Typické mezní kmitočty pásmové propusti, která bývá použita pro filtraci zmíněných trajektorií, jsou 1 a 16 Hz [51][38]. Tato metoda dokáže potlačit trvalé nebo velmi rychlé parazitní signály. Hluky nebo šumy, jejichž časová změna

trajektorií spadá do pásma propustnosti použité pásmové propusti, nejsou potlačeny vůbec. Protože princip metody RASTA vychází z fyziologických schopností lidského řečového ústrojí, tak tato metoda není vhodná pro potlačení šumu v hudebních signálech. Časové trajektorie hudebních nástrojů spadají mimo oblast pásma propustnosti použité pásmové propusti.

Dalším nevhodným systémem pro potlačování šumu v hudebních signálech je metoda nazvaná *mapování spektrogramu* [51][38][50]. Tato metoda využívá obraz spektrogramu a obraz korelačního jádra. Následně je z těchto dvou obrazů vypočten dvojdimenzionální korelační obraz a tento obraz je poté prahován. Výsledkem prahování je binární šablona, kterou je násoben spektrogram původního zarušeného zvukového záznamu. Následnou zpětnou syntézou je získán výstupní signál. Při vzájemném srovnání, je metoda mapování spektrogramu účinnější než metoda RASTA [51][38]. Nevýhoda oproti metodě RASTA spočívá především v tom, že je potřeba přesně určit základní tón řeči. Tato metoda je efektivní pro potlačování šumu ve znělých úsecích s poměrem SNR větším než 3 dB. Metoda mapování spektrogramu je ve zmíněné formě nepoužitelná pro potlačování šumu v hudebních signálech, základní tón řeči zde totiž nedefinuje periodické úseky (neuvažujme případ zpěvu). Tyto periodické úseky jsou definovány základním tónem hudebního nástroje a jeho alikvotními tóny. Další problematickou částí této metody je nutnost dělit vstupní signál na znělé a neznělé úseky, pro případ hudebních signálů by se jednalo o na úseky tonální a netonální. Takovéto dělení je velmi obtížné, ne-li zcela nemožné.

Vlnková transformace je dalším typem transformace, kterou lze využít pro potlačení šumové složky ve zvukových signálech. Rozvoj vlnkové transformace začal v 80 letech a zasloužil se o něj především francouzský vědec Stéphane Mallat [42]. Stéphane Mallat poukázal především na vzájemnou provázanost mezi matematickou teorií vlnkové transformace a teorií Fourierových zobrazení. Výsledkem pak bylo vytvoření zcela nového typu reprezentace a analýzy signálu pomocí tzv. mateřských vlnek (angl. mother wavelets). V oblasti zpracování signálu našla vlnková transformace své uplatnění především pro komprese dat, viz formát JPEG2000 pro kompresi digitálních obrazů [76][32]. Dále byla vlnková transformace například úspěšně použita pro filtraci signálu EKG [39]. Oblasti zájmu této práce je ale zkoumání možností použití vlnkové transformace na potlačení šumu ve zvukových signálech a to se zaměřením na hudební signály. Základní technika, která se používá pro potlačení šumu, se nazývá *waveshrink* [28][35]. Jedná se o jednoduchou metodu, jejímž principem je vynulování koeficientů WT s nízkou hodnotou. Většina energie původního nezrušeného signálu je totiž koncentrována v malém množství koeficientů a odstraněním těchto koeficientů s nízkou hodnotou dojde k potlačení šumu. V literatuře [35] byla tato metoda nahrazena zobecněným postupem nazvaným *wavelet shrinkage*. Toto zobecnění spočívá v nahrazení prostého vynulování koeficientů WT, použitého v metodě *waveshrink*, tzv. prahováním koeficientů WT prahovacím pravidlem. Stručně lze popsat toto prahování jako jednoduchou matematickou operaci nad koeficienty WT aditivně zarušeného signálu s cílem odhadnout spektrum WT čistého signálu. Typ operace je pak definován již zmíněným prahovacím pravidlem [35]. Aby obě zmíněné metody pracovaly úspěšně, je potřeba zvolit vhodnou mateřskou vlnku. Tato mateřská vlnka by měla být volena s ohledem na třídu zpracovaných dat. Mezi třídou zpracovaných dat, konkrétním účelem a mateřskou vlnkou by měla vždy existovat nějaká spojitost. Pokud třída zpracovaných signálů budou hudební signály, tak by tyto báze funkce měli mít harmonický charakter a to zejména kvůli zvýšené citlivosti lidského sluchového ústrojí na tónovou strukturu hudebních signálů [23], podobně jako ji například mají báze funkce z transformací DFT, DCT nebo DST.

Většina zmíněných systémů pracuje v transformační oblasti a z důvodů zpracování dlouhotrvajících signálů jsou tyto systémy nuceny rozdělit vstupní dlouhotrvající signál na krátké segmenty. Typicky volené délky segmentů se pohybují přibližně od 10 ms až po 20 ms, čemuž odpovídají přibližné délky 512 vzorků až po 1024 vzorků při vzorkovacím kmitočtu 44.1 kHz [25][26]. Popřípadě jiné délky, které jsou ale vždy konstantní. Segmenty by při segmentaci měli mít vždy takovou délku, aby signály uvnitř jednotlivých segmentů bylo možno považovat za stacionární signály. Toho lze ale velmi těžko dosáhnout pro případ hudebních signálů. V těchto typech signálů se mohou ve velmi krátké době vystřídat dlouhé a pomalu se měnící úseky za krátké a rychle se měnící úseky. Lépe je tato skutečnost demonstrována na spektrogramu na Obr. 2.12.

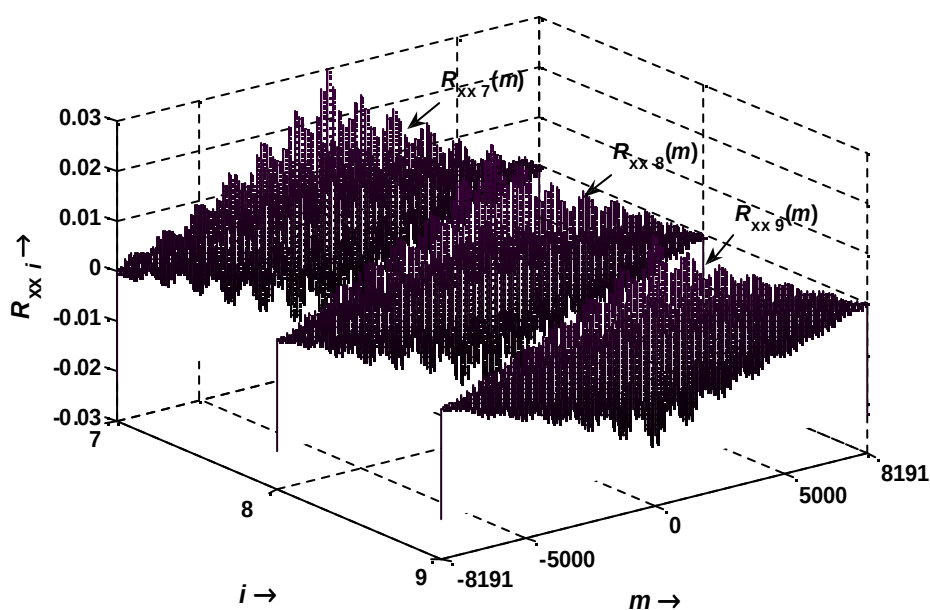


Obr. 2.12. Spektrogram hudební ukázky s vybranými úseky s rozdílnou mírou stacionarity [13].

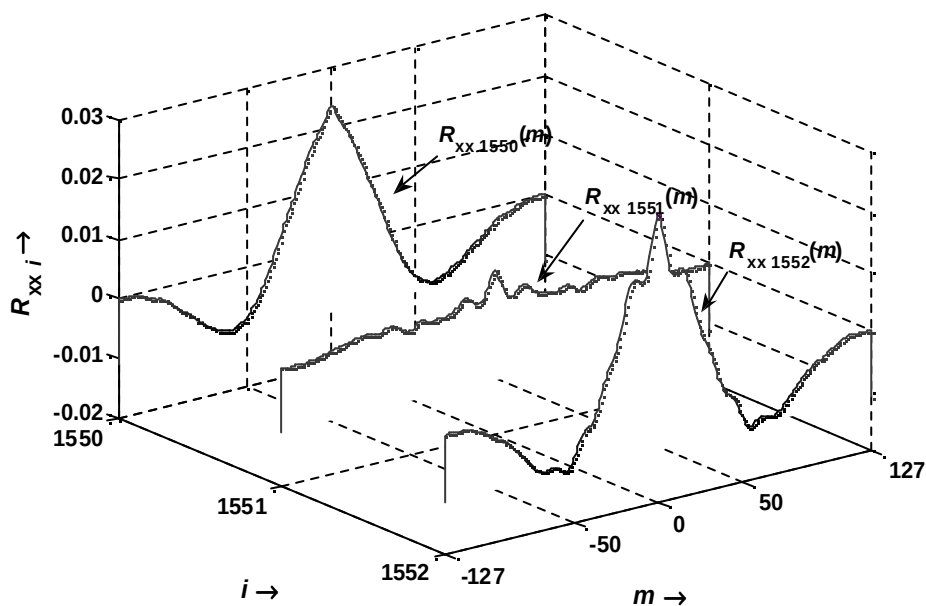
Tento spektrogram reprezentuje krátkou hudební ukázku délky přibližně 5 s s vzorkovacím kmitočtem $f_s = 44.1$ kHz. Na tomto spektrogramu jsou vyznačeny dva úseky, první delší úsek reprezentuje část hudebního ukázky se sólovým zpěvem. A druhý úsek reprezentuje část hudebního ukázky s úderem na bicí, tato část má velmi krátkou dobu trvání a svým charakterem je tento signál podobný impulznímu signálu. Tyto rozdílné úseky byly vybrány z důvodu jejich velmi rozdílné míry stacionarity. Na dalších obrázcích jsou zobrazeny autokorelační funkce segmentů z těchto vybraných úseků. Každý úsek je zde reprezentován třemi na sebe navazujícími segmenty, které se nepřekrývají. Autokorelační funkce tří segmentů z prvního úseku, tj. z úseku se sólovým zpěvem, jsou zobrazeny na Obr. 2.13. A autokorelační funkce z druhého úseku, tj. z úseku s úderem na bicí, jsou zobrazeny na Obr. 2.14. Autokorelační funkce $R_{xxi}(m)$ je definována

$$R_{xxi}(m) = \frac{1}{N} \sum_{k=0}^{N-|m|-1} x_i(k)x_i(k+m) \quad (2.3)$$

kde $x_i(n)$ značí i -tý segment vstupního signálu $x(n)$ o délce N vzorků.



Obr. 2.13. Graf typu Waterfall autokorelačních funkcí tří po sobě následujících segmentů z prvního úseku, tj. z úseku se sólovým zpěvem. Délka segmentu N je 8192 vzorků a $f_s = 44.1$ kHz [13].



Obr. 2.14. Graf typu Waterfall autokorelačních funkcí tří po sobě následujících segmentů z druhého úseku, tj. z úseku s úderem na bicí. Délka segmentu N je 128 vzorků a $f_s = 44.1$ kHz [13].

Na tomto demonstračním příkladu bylo poukázáno na fakt, že je velmi obtížné nalézt univerzální fixní délku segmentů pro segmentaci hudebních signálů. V některých úsecích lze použít délku segmentu $N = 8192$ vzorků, tj. jeden segment trvá 186 ms při $f_s = 44.1$ kHz. A autokorelační funkce sousedních segmentů se příliš neliší, viz Obr. 2.13. A někdy ani délka $N = 128$ vzorků, tj.

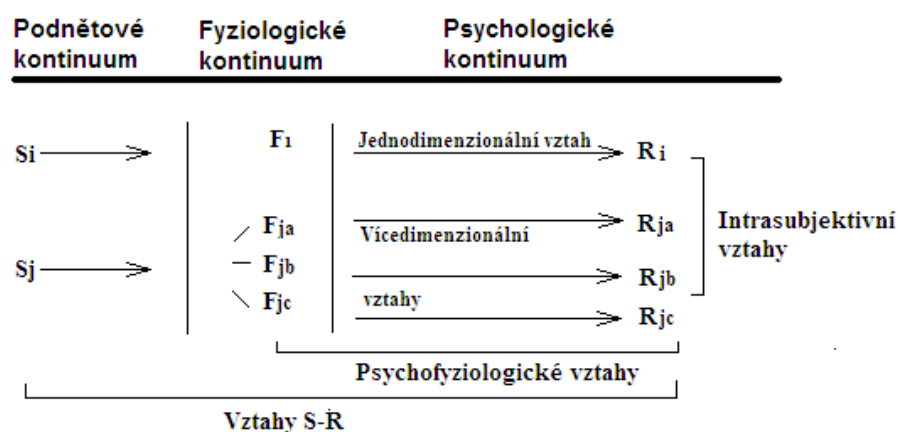
jeden segment trvá necelé 3 ms při $f_s = 44.1 \text{ kHz}$, není dostatečně nízká. Autokorelační funkce sousedních segmentů jsou velmi rozdílné, viz Obr. 2.14.

2.3 Měření šumu pozadí ve zvukových signálech

2.3.1 Úvod do subjektivního vnímání zvukových vjemů

Popsat subjektivní vnímání většiny zvukových vjemů objektivními metodami bývá velmi často obtížná úloha. Je potřeba zkoumat dílčí reakce jedince na podněty a současně tyto reakce správně interpretovat v širším měřítku. Vědní disciplína, která se tímto zabývá, se nazývá psychologická akustika, zkráceně pak psychoakustika. Jedna z definic psychoakustiky uvádí, že „v užším pojetí je cílem psychoakustiky zjišťování kvantitativních vztahů mezi akustickými podněty a sluchovými vjemy“ citováno z [23]. Touto definicí, má psychoakustika velmi blízko k psychofyzice, která je také mimo jiné jejím předchůdcem [44]. V současné době existují dva směry psychofyziky a tyto směry pak podstatně ovlivňují pohled na psychoakustiku jako takovou a na interpretaci jejich výsledků. Klasická psychofyzika zkoumá kvantitativní vztahy mezi fyzikálními podněty (stimuly) a jejími vyvolanými smyslovými počitky. Přičemž počitkem rozumí tato starší psychologie elementární a čistou (tj. zkušenostmi, postoji a myšlením neovlivněnou) reakci osoby. Naproti tomu moderní fyzika odmítá dělení na počitky a vjemy. Podle ní takovéto dělení neodpovídá skutečnosti, protože žádné „čisté počitky“ neexistují, moderní psychofyzika tvrdí, že jsou tyto počitky vždy ovlivněny prostředím, latentními účinky dřívějších podnětů a podobně [44]. Moderní psychofyzika se tedy přiklání k tzv. celostnímu pojetí vnímání, které tvrdí že „vjem není pouhou složeninou elementárních počitků, ale individuálním celostním prožitkem“ citováno z [44].

Pokud tedy bude vycházeno z moderní psychofyziky, tak lze zavést tři typy veličin a tři typy vztahů [44] mezi těmito veličinami pro popis zvukových vjemů. Tyto tři veličiny jsou rozděleny podle typu kontinua, kde je lze měřit. První typ veličin se nachází na podnětovém (fyzikálním) kontinuu. Druhy na fyziologickém kontinuu a třetí na psychologickém kontinuu. Vše je lépe zobrazeno na Obr. 2.15. Dále jsou na tomto obrázku naznačeny typy vztahů mezi těmito veličinami.



Obr. 2.15. Schématické znázornění hlavních oblastí zájmu z nové psychofyziky [43][44].

Příkladem veličiny z podnětového kontinua může být frekvence tónu zvukového signálu v Hz. Fyziologickou proměnou pak může být pozice maximálního podráždění na bazilární membráně a psychologickou proměnou bude subjektivní vjem výšky tónu, vyjádřeny v melech

[44][43]. Vztah S-R v takovém případě bude popisovat závislost vjemu výšky tónu v melech na frekvenci tónu v Hz. Na vjem výšky tónu má ovšem kromě frekvence také vliv například hladina akustického tlaku v dB, tedy další veličina z fyzikálního kontinua [43]. Cílem tohoto krátkého příkladu bylo demonstrovat složité a taky mnohdy nejednoznačné vztahy mezi těmito veličinami. Uvedený příklad se zabýval velmi jednoduchým typem zvukového signálu a to čistým tónem, který lze popsat dvěma veličinami, frekvencí a amplitudou. Situace se ovšem dále velmi komplikuje v případě tónu s časově proměnnými parametry, nebo složených tónů, popřípadě složených zvuků [23].

V dalších podkapitolách se bude rozbor současné problematiky zaměřovat na subjektivní metody, které lze použít pro měření šumu na pozadí ve zvukových signálech. Poté bude následovat přehled a stručný popis většiny objektivních metod. Bude zmíněn účel těchto metod a bude poukázáno na skutečnost, že v současné době neexistuje žádná objektivní metoda, která by dokázala měřit šum na pozadí způsobem podobným tomu, jakým jej subjektivně vyhodnocují posluchači. Většina těchto objektivních metod je totiž založena na klasických metodách zpracování signálu a ty ve většině případů postrádají jakýkoliv psychoakustický model, který by modeloval vztahy a vzájemné závislosti veličin z podnětového a psychotického kontinua podobné těm z Obr. 2.15. Z tohoto důvodu jsou tyto metody nepoužitelné pro predikci subjektivního vjemu slyšitelného šumu a hluku na pozadí ve zvukových signálech.

2.3.2 Subjektivní metody a jejich škály vhodné pro měření šumu na pozadí ve zvukových signálech

Smyslem subjektivních metod je ohodnocení nějakého zvukového vjemu nebo jiné poslechové vlastnosti statisticky významnou skupinou posluchačů. Tito posluchači mají možnost ohodnotit tento zvukový vjem na speciálně vytvořených subjektivních hodnoticích škálách. Subjektivní metody jsou nejčastěji realizovány formou psychologických měření, které jsou nazývány poslechové testy. Pro subjektivní měření šumu pozadí ve zvukových záznamech existuje několik subjektivních metod a škál. Příkladem první z nich může být subjektivní škála nazvaná *Detectability opinion Scale* (DS) definovaná standardem ITU (angl. International Telecommunication Union) [58]. Tato škála slouží pro detekci nějakého zvukového vjemu, např. ozvěny (echa) v telekomunikačním řetězci [58]. Posluchači na této škále při poslechových testech hodnotí, zda je tento zvukový vjem detekovatelný a mají k tomu k dispozici hodnoticí škálu o třech hodnotách, viz Tab. 2.1. Takovouto škálu lze použít i pro detekci šumu pozadí ve zvukových signálech. Výsledkem tohoto hodnocení by však byla pouze informace, zda šum pozadí je nebo není slyšitelný. A v případě že je slyšitelný, tak zda je nebo není rušivý.

Tab. 2.1. Subjektivní škála DS (Detectability opinion Scale) [58].

Škála DS, hodnotí se detekovatelnost a rušivost.	
A	Nedetkovatlené (angl. not detectable)
B	Detekovatlené, ale nerušící (angl. detectable)
C	Detekovatlené a rušící (angl. objectionable)

Další subjektivní škály, které mohou být využity v subjektivních metodách pro měření šumu pozadí v řečových a hudebních signálech, jsou škály *Mean Opinion Score* (MOS) [59] a *Subjective Different Grade* (SDG) [55]. Primární využití těchto škál ovšem leží v měření kvality, MOS škála, nebo velikosti degradace, případ škály SDG. Škála MOS hodnotí celkovou kvalitu a SDG hodnotí velikost poškození (degradace) měřeného zvukového signálu vůči jeho originální verzi. Škála SDG oproti škále MOS dosahuje vyšší citlivosti na míru zkreslení nebo zarušení [55]. Obě škály společně s popisem jsou zobrazeny v Tab. 2.2.

Tab. 2.2. Subjektivní škály MOS (Mean Opinion Score) a SDG (Subjective Different Grade) [59][55].

Škála SDG, hodnotí se velikost degradace.		Škála MOS, hodnotí se celková kvalita.	
0	Nevnímatelné (angl. imperceptible)	5	Výborná (angl. excellent)
-1	Vnímatelné, ale ne nepříjemné (angl. perceptible but not annoying)	4	Dobrá (angl. good)
-2	Lehce nepříjemné (angl. slightly annoying)	3	Slušná (angl. fair)
-3	Nepříjemné (angl. annoying)	2	Špatná (angl. poor)
-4	Velmi nepříjemné (angl. very annoying)	1	Nevhodná (angl. bad)

Tyto škály mohou měřit šum pozadí v zvukových signálech pouze nepřímou a tímto způsobem, že se měří výsledná kvalita, a z této výsledné kvality ohodnocené na MOS škále se predikuje úroveň slyšitelného šumu pozadí. Podobným způsobem lze použít získaná hodnocení na SDG škále pro evaluaci slyšitelného šumu pozadí. Opět lze predikovat úroveň slyšitelného šumu pozadí podle míry degradace vstupního signálu, vyjádřenou na SDG škále, vůči originálnímu signálu. Ani jedna z uvedených škál ovšem nedává informaci o úrovni slyšitelného šumu, ale pouze o tom jak moc šum pozadí degraduje originální signál, nebo jak moc šum pozadí snižuje kvalitu originálního signálu.

Další možností jak subjektivně vyhodnotit velikost šumu pozadí ve zvukových systémech je použít tzv. adjektivní posuzovací škály. Tento typ škál je často použit v technikách sémantického diferencíálu pro zjištění základních percepčních dimenzí vícerozměrných subjektivních veličin jako je například barva zvuku, celková kvalita reprodukce zvuku, rušivost hluku apod [44]. Příkladem adjektivní škály pro hodnocení drsnosti/jemnosti zvuku může být stupnice [44]:

mimořádně velmi poněkud ani jemný
ani drsný poněkud velmi mimořádně

jemný _____ _____ _____ _____ _____ _____ drsný

Podobná adjektivní škála, pro posuzování množství slyšitelného šumu, by byla jistě výhodnější, než škály DS z Tab. 2.1 nebo škály MOS a SDG, Tab. 2.2. Rozšířila by se totiž hodnotící stupnice na více, než 3 resp. 5 hodnot. A hlavně by takováto škála lépe vystihovala charakter měřené veličiny, kterou není kvalita, velikost degradace a ani detekovatelnost slyšitelného šumu, ale úroveň slyšitelného šumu na pozadí ve zvukových signálech. Bohužel v současné době

neexistuje žádná objektivní metoda, která by predikovala úroveň slyšitelného šumu pozadí, na podobném typu škály.

2.3.3 Jednoduché objektivní metody

Realizace poslechových testů je časově i finančně náročná činnost, proto se je snaha obcházet tyto testy pomocí tzv. objektivních metod. Jedná se o deterministické metody, jejímž cílem je predikce subjektivních soudů získaných poslechovými testy a v konečném výsledku nahrazením těchto poslechových testů objektivními metodami. Hodnotícím kritériem většiny objektivních metod bývá procentuální úspěšnost predikce subjektivních soudů, tj. míra shody mezi výsledky získaných objektivními metodami vůči výsledkům ze subjektivních metod. Objektivní metody tedy používají stejný typ hodnotících škál jako subjektivní metody. Jediným rozdílem je, že tyto objektivní škály bývají označeny adjektivem "objektivní", tedy například SDG škála z Tab. 2.2 je označována jako ODG (Objective Different Grade).

Nejstarší objektivní metody lze charakterizovat jako jednoduché míry, které kvantifikují rozdíly mezi čistým a degradovaným signálem, popř. jeho šumovou složkou. Tyto míry se vyznačují nejnižším stupněm predikce a absencí psychoakustického modelu [14]. Mezi tyto míry lze zařadit například různé varianty SNR (Signal-to-Noise Ratio), např.: GSNR (Global SNR), LSNR (Local SNR) [25], SSNR (Segmental SNR) [47]. Rozdíly mezi zmíněnými variantami spočívají v rozdílném způsobu výpočtu přírůstku SNR z jednotlivých segmentů do výsledné objektivní míry. Stímtypem měř úzce souvisí tzv. spektrální vzdálenostní míry. Konkrétně se jedná o míry *Linear Spectral Distance*, *Non-Linear Spectral Distance*, *Log-Spectral Distance* [25][26]. Tyto míry vždy počítají "vzdálenost" definovanou L_p normou [25] pro různé p mezi krátkodobými spektry čistého a zarušeného signálu. Tyto míry se liší tzv. vzdálenostní mírou, pokud se jedná o lineární vzdálenostní míru tak se mluví o *Linear Spectral Distance* a pokud o nelineární tak je to *Non-Linear Spectral Distance* atd. Kromě spektrálních vzdálenostních měř existují také kepsrální vzdálenostní míry (angl. cepstral distance [25]). Princip je stejný, jen je počítána vzdálenost mezi kepstrem čistého a zarušeného signálu. Další míry z této kategorie jsou: *log-likelihood ratio* [33], *log-area distance* [25]. Tyto míry jsou charakteristické tím, že jsou založené na koeficientech lineární predikce. Z výzkumu těchto objektivních měř [53] vyplývá, že tyto jednoduché míry jsou úspěšné ve své predikci jen částečně. Úspěšnost predikce vyjádřena hodnotou Pearsonova korelačního koeficientu [44] je rovna pro zmíněné objektivní míry hodnotě kolem 0.9 pro případy jednoduchých zkreslení a hodnotě kolem 0.7 pro komplexnější případy zkreslení [25][53]. Nejlepších výsledků dosahovaly míry: LAR, kepsrální vzdálenosti a log-area distance [25][53], tj. míry které kvantifikují rozdíly, popř. změny, mezi čistým a degradovaným signálem v prostorové (spektrální nebo kepsrální) oblasti.

Protože nejlepších výsledků dosahovaly míry, které operují v prostorové oblasti, tak se následující výzkum zaměřil právě na tuto oblast. Následující vylepšení spočívalo v nelineárním mapování frekvenční škály (angl. frequency axis warping) na bářková pásma a z těchto spektrálních komponent byly poté vypočteny tzv. dílčí hlasitosti v sonech [23][44]. Jedná se o první psychoakusticky motivované objektivní míry. V roce 1992 byla zavedena první takováto míra založená na popsáních principech nazvaná *Bark-Spectral Distortion* (BSD). Tato míra kvantifikuje rozdíl v hlasitostech mezi čistým a degradovaným signálem [25]. Změřená hodnota Pearsonova korelačního koeficientu mezi hodnocením BSD a MOS dosahovala hodnot 0.85 až 0.98 pro řečové kodéry s nízkým bitovým tokem, což byla řádově o 0.1 větší míra korelace než dosahovala CD. V

roce 1997 byla metoda BSD vylepšena o použití maskovacích prahů při výpočtu celkové hlasitosti, vznikla tak *modified BSD* (MBSD) [80][83][81]. Z dalších vylepšení MBSD lze uvést například [82], nebo *enhancement MBSD* (EMBSD) [81]. EMBSD byl optimalizován pro řečové signály v TDMA (angl. *Time Division Multiple Access*) sítích a zde také dosahuje lepších výsledků než MSDB [81]. Míra MBSD dosahuje lepších výsledků než BSD [80][83].

2.3.4 Moderní objektivní měřicí metody

Většina moderních objektivních metod je navržena pro objektivní varianty škály z Tab. 2.2. Škála MOS se používá pro hodnocení řečových signálů a SDG pro zvukové signály, nejčastěji se ale tato škála používá pro hudební signály. Toto dělení vychází z odlišné citlivosti obou škál na zkreslení měřeného signálu. Škála MOS hodnotí výslednou kvalitu a proto je méně citlivá na velikost zkreslení nebo zarušení měřeného signálu než například škála SDG, která měří velikost degradace měřeného zvukového signálu vůči jeho referenční verzi.

Objektivních metod navržených na evaluaci kvality řečových signálů na MOS škále [58] je relativně velké množství, např. PSQM [60], TOSQA, PESQ [61] nebo POLQA [62]. Tyto metody se liší především použitým psychoakustickým modelem a rozdílnými požadavky na měřený signál. Bohužel tyto metody nelze použít na hodnocení zvukových popř. hudebních signálů, protože jsou zaměřené pouze na signály řečové, a to například z důvodů omezené šířky pásma, viz např. metoda PSQM pracující v pásmu kmitočtu 300 – 3400 Hz [60]. Nebo kvůli použitým psychoakustickým modelům, které selhávají pro hudební signály. Primární využití těchto metod je hodnocení přenosu řečové kvality v telekomunikačních sítích a evaluace řečových kodeků.

Druhá skupina metod jsou objektivní metody zaměřené na evaluaci míry degradace měřeného zvukového signálu vůči jeho originální verzi podle ODG škály. Počet objektivních metod v této skupině je podstatně nižší než v předcházejícím případě. Hlavními zástupci této skupiny jsou metody PEAQ (angl. *Perceptual Evaluation of Audio Quality*) [57] a PEMO-Q [30]. Obě metody jsou shodně navrženy pro evaluaci zvukových signálů. Primární oblast použití těchto metod je evaluace zvukových kodeků. Tyto metody se navzájem odlišují způsobem jejich implementace. Metoda PEAQ je realizována tzv. toolbox přístupem, jádrem metody je neuronová síť, na jejíž vstup je přiváděno několik psychoakustických metrik z vybraných psychoakustických modelů [57]. Druhý algoritmus PEMO-Q je realizován tzv. psychoakustickým modelováním lidského sluchového ústrojí. Nevýhoda toolbox přístupu spočívá především v tom, že výsledný algoritmus neobsahuje validovaný model sluchového ústrojí, ale pouze adaptovanou neuronovou síť na předložený vzor [55].

Pro měření šumu pozadí v hudebních signálech lze tedy využít pouze objektivní metody PEAQ a PEMO-Q a to ale s několika omezeními. Tyto omezení vyplývají z použité hodnotící škály ODG. Tato škála je navržena pro hodnocení míry degradace nebo zkreslení měřeného signálu vůči jeho originální verzi. V případě měření slyšitelného šumu pozadí v hudebních signálech, lze těmito metodami pouze změřit, zda šum pozadí je nebo není slyšitelný. Nebo lze změřit velikost degradace způsobenou aditivním šumem. Nelze ale změřit úroveň tohoto slyšitelného šumu. Popřípadě na kterých frekvenčních pásmech a taky v jakých časových okamžicích se tento slyšitelný šum vyskytuje. Tedy informace, které jsou toliko potřebné pro návrh správně fungujících systému na potlačení šumu, jejímž konečným uživatelem je člověk.

3 Cíle disertační práce

Tato disertační práce se zabývá objektivním měřením a potlačování rušivého šumu na pozadí zvukových signálů se zaměřením na signály hudební. Tematicky ji lze rozdělit do dvou zdánlivě samostatných částí, které spolu však velmi úzce souvisí. Možnost objektivního měření tohoto rušivého šumu na pozadí, namísto provádění časově a také finančně náročných poslechových testů, dovoluje návrh mnohem lépe optimalizovaných systémů pro jeho potlačení. Cílovým uživatelem zde není stroj nebo další zařízení, ale člověk. Zdánlivě jednoznačná úloha potlačení šumu se totiž pro zvukové záznamy zacílené na lidské posluchače stává komplexnějším problémem potlačení šumu na pozadí způsobem, kdy je třeba uvažovat psychoakustické jevy jako je například maskování zvuku, nelineární charakteristiky různých psychoakustických veličin a další jevy.

Prvním cílem disertační práce je návrh metody pro objektivní stanovení odhadu subjektivní míry slyšitelného šumu pozadí v hudebních signálech. Konkrétně tato metoda má predikovat, zda šum pozadí je pro posluchače slyšitelný, či nikoliv. Dále by tato metoda měla poskytovat prostředky pro další hlubší analýzu tohoto šumu na pozadí, zejména lokalizaci a kvantifikaci slyšitelného šumu na pozadí v časo-frekvenční oblasti. Součástí návrhu této metody je i provedení poslechových testů. Cílem těchto testů je zjištění velikosti shody mezi predikovanými hodnotami slyšitelnosti šumu na pozadí a subjektivními hodnotami získanými z těchto poslechových testů.

Dalším cílem disertační práce je návrh samotného systému na potlačení aditivního šumu pozadí v hudebních signálech. Tento systém má být navržen s ohledem na psychoakustické aspekty vnímání zejména hudebních signálů posluchači. Kromě samotného potlačení nežádoucího slyšitelného šumu pozadí by tento systém měl zavádět minimální zkreslení užitečného signálu a také by měl minimalizovat množství vzniklého hudebního šumu.

Jak již bylo ukázáno v předchozí kapitole, nevýhodou většiny systémů na potlačení šumu je skutečnost, že zpracování vstupního dlouhotrvajícího signálu probíhá nad segmenty konstantní délky. Naráží se zde na problém omezeného časo-frekvenčního rozlišení použitých transformací a nestacionárního charakteru zvukových signálů. Proto je dalším cílem této disertační práce doplnit stávající systém na potlačení šumu novým algoritmem, který adaptivně rozdělí dlouhotrvající nestacionární signál na úseky (segmenty), které lze pro účely prahování považovat za úseky stacionární, tj. na ty úseky, ve kterých se statistické vlastnosti analyzovaného signálu budou měnit velmi pomalu. Dále by měl být tento algoritmus navržen takovým způsobem, aby byly zachovány podmínky perfektní rekonstrukce. Na začátku práce byl předpoklad, že doplnění stávajícího systému na potlačení šumu adaptivní segmentací povede k vyššímu potlačení nežádoucího šumu a současně ke snížení množství hudebního šumu a zvýšení výsledné subjektivně vnímané kvality výstupních signálů. Posledním cílem disertační práce je tuto domněnku potvrdit realizovaným poslechovým testem na statisticky významné skupině dobrovolníků.

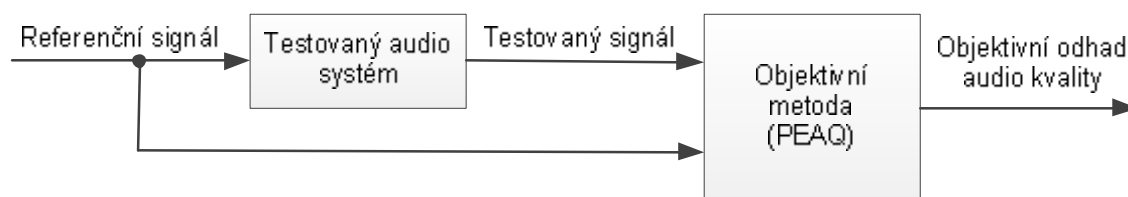
Konkrétní body zadání cílů disertační práce:

- 1) Proved'te **analýzu** současného stavu dané problematiky objektivního měření a potlačování aditivního šumu pozadí v hudebních signálech.
- 2) **Navrhňte** novou metodu pro objektivní stanovení množství slyšitelného šumu na pozadí v hudebních signálech.
- 3) **Proved'te** poslechové testy pro **zhodnocení** úspěšnosti metody z bodu č. 2.
- 4) **Navrhňte** systém na potlačení šumu na pozadí v hudebních signálech, který využívá adaptivní segmentační schéma.
- 5) **Proved'te** vhodný poslechový test pro **zhodnocení** navrženého systému z bodu č. 4

4 Objektivní metoda PEAQ

Metoda PEAQ (*Perceptual Evaluation of Audio Quality*) byla navržena pro objektivní hodnocení kvality zvukového signálu [57] [37]. Tato metoda nehodnotí tzv. absolutní kvalitu (ACR, *absolute category rating*) jako například v případě škály MOS, ale velikost degradace signálu (DCR, *Degradation Category Rating*) [58]. Jednoduchým příkladem hodnocení kvality zvukového signálu z kategorie DCR může být hodnocení, kdy posluchači jsou předloženy dvě zvukové ukázky. Jedna je původní "čistá" a druhá zvuková ukázka je degradována nějakým typem šumu nebo zkreslením. Tento posluchač ví, která zvuková ukázka je původní a hodnotí velikost degradace této druhé zvukové ukázky oproti té první. Obecně se dá říci, že hodnocení podle škál z kategorie DCR jsou citlivější na různá malá rušení, degradace nebo zkreslení užitečného signálu než stupnice z kategorie ACR. Norma [55] udává, že škály z kategorie ACR by měly být použity při středně velké degradaci čistého signálu a při nižší degradaci by se měly používat citlivější škály z kategorie DCR.

Základní koncept objektivního měření v metodě PEAQ je zobrazen na Obr. 4.1. Tento koncept tematicky vychází z norem [55] a [56] pro subjektivní hodnocení malého zkreslení ve zvukových signálech.

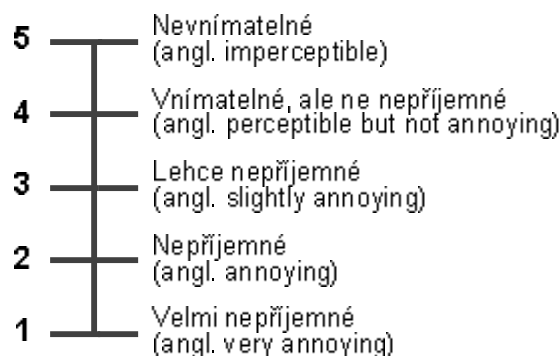


Obr. 4.1. Základní koncept objektivního měření v metodě PEAQ [57].

Referenční signál je přiváděn na vstup testovaného zvukového systému. V případě metody PEAQ se nejčastěji jedná o zvukové kodéry. Z výstupu testovaného systému je přiveden testovaný signál na vstup do objektivní metody. Souběžně s tímto signálem vstupuje také do tohoto bloku referenční signál. Z rozdílu mezi testovaným signálem a referenčním signálem, objektivní metoda určí výslednou odhadovanou kvalitu. Jedná se tedy o intrusivní objektivní metodu.

Objektivní měřicí metoda PEAQ, popsána normou [57], vychází při svém hodnocení v subjektivní doméně z normy [55]. Tato norma popisuje metody pro subjektivní hodnocení malých zkreslení (angl. *small impairments*) ve zvukových systémech včetně vícekanálových zvukových systémů.

Základní princip hodnocení kvality zvukových signálů podle [55], který se využívá při poslechových testech, je následující. Posluchačům jsou předloženy tři zvukové ukázky se třemi signály. Budou označeny jako "A", "B" a "C". Tito posluchači ví, že ukázka "A" vždy obsahuje referenční signál. Ukázky "B" a "C" jsou posluchačům předloženy náhodně, jedna z nich obsahuje testovaný signál a druhá obsahuje znova referenční signál. Před posluchači jsou ale tato označení skryta. Posluchači pouze ví, že jedna z předložených zvukových ukázek obsahuje znova referenční zvukovou ukázku a jejich cílem je subjektivně ohodnotit velikost degradace ukázek "B" a "C" vůči zvukové ukázce "A". Při svém hodnocení použijí škálu z normy [55], Obr. 4.2.

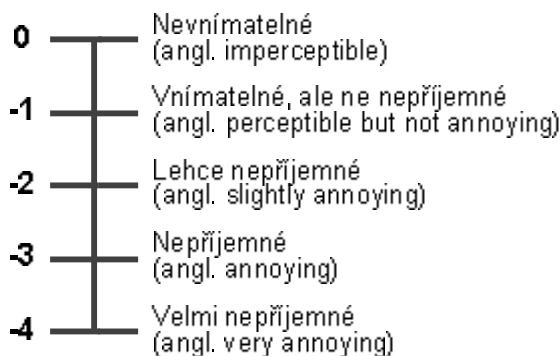


Obr. 4.2. Pětistupňová hodnotící škála podle [55].

Výsledné hodnocení SDG (*Subjective Different Grade*) je stanoveno z obou výsledných diferencí označených jako $GRADE_{\text{Testovaný signál}}$ a $GRADE_{\text{Referenční signál}}$

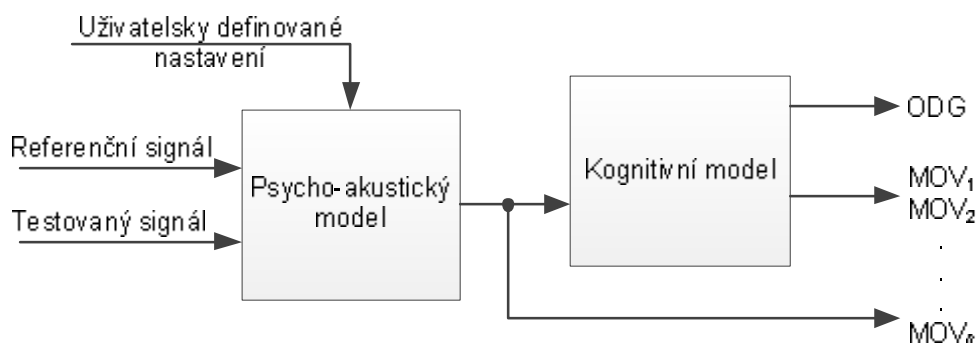
$$SDG = Grade_{\text{Testovaný signál}} - Grade_{\text{Referenční signál}} \quad (4.1)$$

Výsledné hodnocení SDG má stejný význam jako na stupnici z Obr. 4.2. Jen je použito jiné číselné označení jednotlivých škál, a to podle vztahu (4.1), viz Obr. 4.3. Tuto škálu (4.1) používá metoda PEAQ k objektivnímu hodnocení kvality zvukového signálu. Aby nedošlo k záměně za subjektivní hodnocení získané poslechovými testy, tak je škála SDG (4.1) označena jako ODG (*Objective Different Grade*) [57], namísto písmenka S (Subjektivní) je použito písmenko O (Objektivní).



Obr. 4.3. Pětistupňová hodnotící škály SDG a ODG [57].

Obecné blokové schéma metody PEAQ je zobrazeno na Obr. 4.4. Vstupními signály jsou referenční a testovaný signál. Výstupem je zejména výsledné hodnocení na ODG škále a pak také tzv. dílčí výstupní hodnoty modelu PEAQ označované jako MOV_1 až MOV_n (*Model Output Values*), kde index n značí pořadové číslo. Dále toto blokové schéma obsahuje dva bloky označené psychoakustický model a kognitivní model. Psychoakustický model se snaží modelovat psychoakustické vnímání průměrného lidského jedince. Výstupem tohoto modelu jsou první dílčí výstupní hodnoty MOV. Kognitivní model provádí extrakci a kombinaci jednotlivých MOV do výsledné hodnoty ODG. Je důležité zde říci, že tento kognitivní model je realizován umělou neuronovou sítí.



Obr. 4.4. Obecné blokové schéma metody PEAQ [57].

Objektivní metoda PEAQ existuje ve dvou verzích, v tzv. základní (angl. *basic*) a pokročilé (angl. *advanced*) verzi. Tyto verze se liší zejména ve způsobu realizace modelu periferního ucha (část psychoakustického modelu), ten je buď realizován DFT (základní model), nebo kombinovanou variantou DFT a bankou filtrů (pokročilý model). Z hlediska použití obou verzí je rozdíl zejména v tom, že pokročilá verze je navržena pro zpracování zvukových signálů v reálném čase. V následujících tabulkách Tab. 4.1 a Tab. 4.2 jsou uvedeny všechny dílčí výstupní hodnoty MOV a to pro obě verze včetně krátkého popisu jejich účelu. Tyto dílčí výstupní hodnoty MOV mohou být použity samostatně k měření nějaké konkrétní vlastnosti testovaného signálu, např. k měření velikosti lineárního/nelineárního zkreslení atd. Detailní popis, společně s matematickými definicemi, jednotlivých MOV lze nalézt např. v [57] [37].

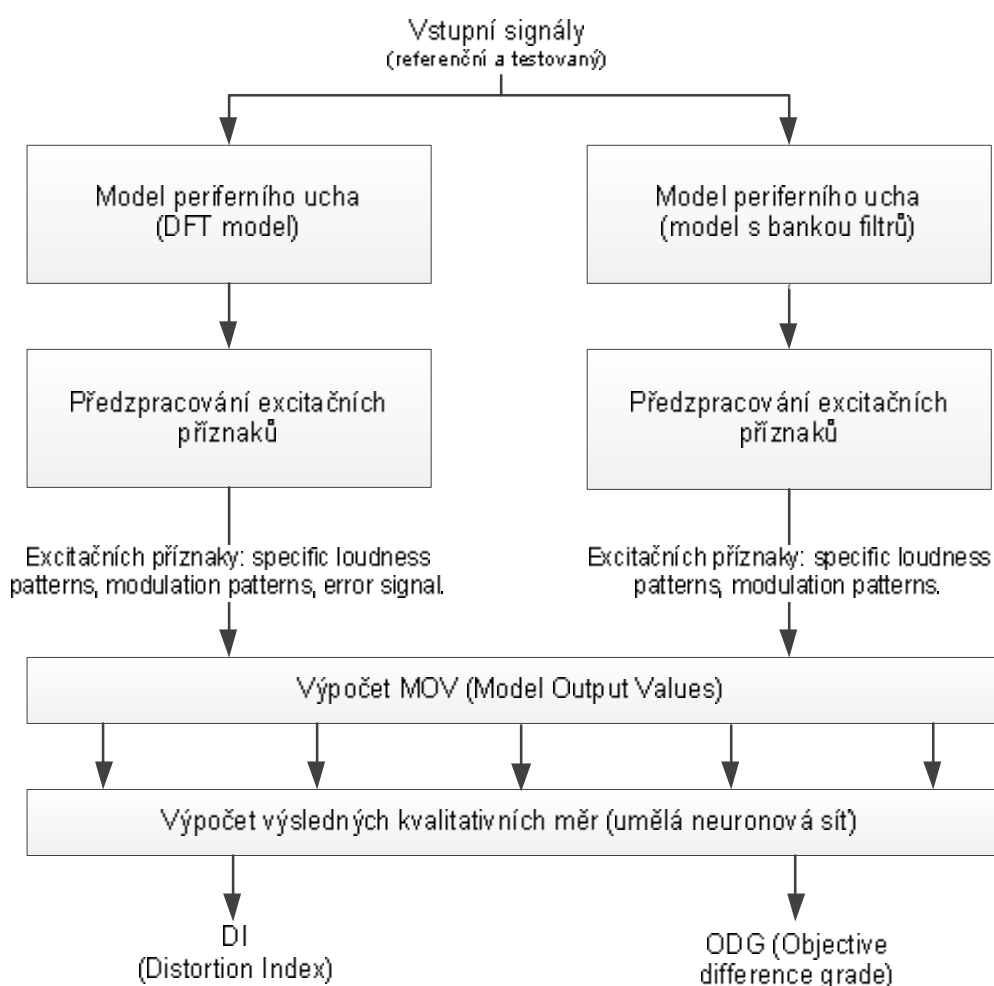
Tab. 4.1. Veličiny MOV ze základní verze metody PEAQ.

MOV (<i>Model Output Values</i>)	Účel těchto veličin
WinModDiff1 _B	Měření změn v modulaci (vzhledem k drsnosti zvuku)
AvgModDiff1 _B	
AvgModDiff2 _B	
RmsNoiseLoud _B	Měření hlasitosti zkreslení
BandwidthRef _B	Měření lineárního zkreslení
BandwidthTest _B	
RelDistFrames _B	Měření četnosti slyšitelných zkreslení
Total NMR _B	Měření odstup šumu k masce
MFPD _B	Měření pravděpodobnosti detekce
ADB _B	
EHS _B	Měření chyby harmonické struktury

Tab. 4.2. Veličiny MOV z pokročilé verze metody PEAQ.

MOV (Model Output Values)	Účel těchto veličin
$RmsNoiseLoudAsym_A$	Měření hlasitosti zkreslení
$RmsModDiff_A$	Měření změn v modulaci (vzhledem k drsnosti zvuku)
$AvgLinDist_A$	Měření lineárního zkreslení
$Segmental\ NMR_B$	Měření odstupe šumu k masce
EHS_B	Měření chyby harmonické struktury

Detailní blokové schéma metody PEAQ je zobrazeno na Obr. 4.5. Metoda PEAQ se skládá z modelu periferního ucha, několika následných kroků (zde naznačeno blokem “předzpracování excitačních příznaků”), blokem výpočtů psychoakusticky motivovaných dílčích výstupních hodnot MOV a mapování MOV do jediné hodnoty ODG reprezentující zvukovou kvalitu testovaného signálu.



Obr. 4.5. Detailní blokové schéma metody PEAQ [57].

Princip mapování do výsledné hodnoty ODG je naznačen vztahem

$$\text{ODG} = b_{\min} + (b_{\max} - b_{\min}) \text{sig}(DI), \quad (4.2)$$

kde $b_{\min}$ a $b_{\max}$ jsou experimentálně získané konstantní hodnoty měřítek pro každou variantu metody PEAQ a DI je hodnota zkreslujícího indexu (angl. *Distortion Index*). Hodnota DI je výstupní hodnotou umělé neuronové sítě [57]. Aktivační funkce této neuronové sítě je asymetrický sigmoid (4.3).

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \quad (4.3)$$

5 Objektivní metoda pro měření úrovně slyšitelného šumu pozadí v hudebních signálech – vlastní řešení

V této kapitole bude představeno vlastní řešení objektivní metody pro měření a stanovení odhadu slyšitelného šumu na pozadí v hudebních signálech, které jsem navrhnul. Toto vlastní řešení poskytuje prostředky a nástroje k odhadu subjektivního vjemu slyšitelnosti šumu pozadí. A dále poskytuje nástroje pro následnou hlubší analýzu tohoto slyšitelného šumu pozadí v časofrekvenční oblasti. Metoda je navržena i na další typy zvukových signálů, konkrétně se jedná o řečové a umělé zvukové signály, viz jejich definice v kapitole 2.1. Metoda se zaměřuje na nízký stupeň zarušení aditivním šumem. Tímto nízkým stupněm zarušení je myšleno takové zarušení, které je na hraně slyšitelnosti posluchači. A navíc bývá vnímatelné pouze v některých oblastech takto zarušeného signálu, například v tišších úsecích, na začátku skladby a podobně.

Tato kapitola je strukturována následovně. Nejprve jsou stanoveny a diskutovány výchozí úvahy, ze kterých bylo vycházeno při návrhu této objektivní metody. Následuje popis modelu zarušení originálního signálu aditivním šumem pozadí a blokové schéma navržené metody. Po té následuje detailní popis navržené metody. Dále je v této kapitole uveden příklad zvukové ukázky zarušené aditivním šumem pozadí. Na této ukázce je posléze demonstrována analýza slyšitelného šumu pozadí pomocí nově navržené objektivní metody. Další demonstrativní příklady jsou uvedeny v příloze A. Závěr kapitoly je věnován provedenímu poslechovému experimentu a celkovému zhodnocení navržené objektivní metody.

5.1 Výchozí úvahy pro navrhovanou metodu

Navrhovaná metoda vychází z několika následujících úvah (presumpcí):

- **Šum pozadí je slyšitelný pouze tehdy, když není maskován užitečným signálem.**

Maskování je proces v psychoakustice, který zvyšuje práh slyšení jednoho zvuku vlivem přítomnosti jiného (maskujícího) zvuku [17]. Maskování probíhá ve dvou hlavních oblastech a to v frekvenční oblasti, tj. simultánní maskování, a v časové oblasti, tj. nesimultánní maskování. Pokud je zvuk maskován jiným hlasitějším zvukem, tak se stává pro posluchače nepostřehnutelný, tj. nebude pro posluchače slyšitelný. Důsledek této presumpce pro navrhovanou objektivní metodu je takový, že je chybné do této metody zakomponovat vliv těch zvukových komponent, které jsou maskovány a tudíž irelevantní pro posluchače [23].

Příklady metod, kde je podle této presumpce chybně zakomponován vliv maskovaných zvukových komponent, mohou být metody $R_{\text{NMR}_{\text{tot}}}$ a $R_{\text{NMR}_{\text{seg}}}$ z metody PEAQ pro stanovení celkové a segmentální úrovně odstupů šumu k masce, angl. *Noise to Mask Ratio* (NMR) z celého analyzovaného signálu. Tyto varianty NMR jsou definovány [37]

$$R_{\text{NMR}_{\text{tot}}} = 10 \log_{10} \left(\frac{1}{I} \sum_{i=0}^{I-1} \frac{1}{N_k} \sum_{k=0}^{N_k-1} R_{\text{NMR}}(k, i) \right), \quad (5.1)$$

$$R_{\text{NMR}_{\text{seg}}} = \frac{1}{I} \sum_{i=0}^{I-1} 10 \log_{10} \left(\frac{1}{N_k} \sum_{k=0}^{N_k-1} R_{\text{NMR}}(k, i) \right), \quad (5.2)$$

kde N_k je počet filtračních pásem v dílčích kritických pásmech, I je celkový počet segmentů a $R_{\text{NMR}_{\text{seg}}}(k, i)$ je hodnota NMR pro každé k -té pásmo a i -tý segment.

Obě tyto varianty NMR ze vztahu (5.1) a (5.2) zahrnují vliv maskovaných zvukových komponent do svých výsledných hodnot. Záporné hodnoty $R_{\text{NMR}_{\text{seg}}}(k, i)$ zde značí ty komponenty, kde šum pozadí je maskován, tj. není slyšitelný. A přesto jsou tyto záporné hodnoty $R_{\text{NMR}_{\text{seg}}}(k, i)$ použity do výsledných průměrných hodnot celkového $R_{\text{NMR}_{\text{tot}}}$ (5.1) nebo segmentálního $R_{\text{NMR}_{\text{seg}}}$ (5.2) odstupu šumu k masce.

- **Šum pozadí je složený zvuk a pro jeho detekci stačí vyhodnotit pouze “nejhlasitější” komponentu.**

Pro výpočet hlasitosti složených zvuků se používají převážně dva postupy. První postup je nazýván výpočet hlasitosti podle Stevensa [73][18] a druhý postup je nazýván výpočet hlasitosti podle Zwicker [23]. Oba postupy výpočtů celkové hlasitosti, lze naznačit stejným vztahem

$$N_L = \int_{b=1}^{24} N_L'(b) db, \quad (5.3)$$

kde $N_L'(b)$ jsou tzv. dílčí hlasitosti v sonech v Barkových pásmech a N_L je výsledná hlasitost v sonech. Podstatnou roli při vzniku vjemu hlasitosti hrají také mechanismy maskování [44]. Jelikož je výsledná hlasitost N tvořena těmito dílčími hlasitostmi $N_L'(b)$, tak lze očekávat, že pokud pro nějaké $N_L'(b)$ platí, že $N_L'(b) > 0$, tak zvuk bude slyšitelný (detekovaný). K detekci slyšitelnosti šumu pozadí jako složeného zvuku stačí tedy vyhodnocovat dílčí hlasitosti s největší hodnotou.

- **Subjektivní veličina popisující úroveň slyšitelného šumu má nestacionární charakter.**

Oba signály, tj. šum pozadí a užitečný signál popředí jsou obecně nestacionárními signály a proto lze očekávat, že subjektivní veličina popisující úroveň slyšitelného šumu pozadí také bude nestacionární, tj. funkcí času. V řeči například je šum pozadí nejslyšitelnější v pauzách, v hudbě pak v tichých úsecích a podobně. Proto by objektivní metoda pro vyhodnocení slyšitelného šumu pozadí měla být funkcí času, nebo by měla nějakým jiným způsobem tuto skutečnost reflektovat.

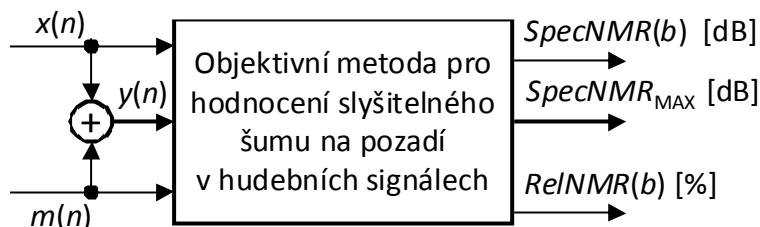
5.2 Princip navrhované metody

V této kapitole bude popsán princip metody pro objektivní měření slyšitelnosti šumu na pozadí v hudebních signálech, kterou jsem navrhl. Tato metoda hodnotí celý dlouhotrvající signál po částech, tj. po segmentech. Z každého segmentu jsou odhadnuty výkonové spektrální hustoty (PSD) vstupního užitečného signálu a šumu na pozadí. Z odhadu PSD užitečného signálu jsou vypočteny maskující prahy. Tyto prahy jsou vypočteny podle psychoakustického modelu z kompresního formátu MPEQ [31]. Byl zde použit psychoakustický model č. 1, jehož implementaci lze nalézt v [49]. Z odhadů PSD šumu na pozadí a vypočtených maskujících prahů jsou poté stanoveny hodnoty odstupu šumu k tomuto maskujícímu prahu (NMR). Tímto způsobem jsem se snažil zakomponovat do navrhované objektivní metody vliv maskování šumu pozadí užitečným signálem popředí. Oproti stávajícím variantám měření NMR je navržena metoda unikátní tím, že do výsledných hodnot NMR nejsou započítávány ty případy, kdy komponenty šumu na pozadí jsou maskovány komponentami užitečného signálu popředí, a tedy nejsou pro posluchače slyšitelné.

Zmíněný způsob výpočtu NMR se provádí v tzv. Barkových pásmech (kritické pásma slyšení). Barkové pásma jsou tvořeny 24 frekvenčními pásmy, která jsou přibližně logaritmičtě rozvržena a pokrývají oblast slyšitelných zvukových signálů do cca 15,5 kHz. Tyto Barkova pásma jsem zavedl z důvodu, aby navržená metoda lépe vystihovala charakter vnímání složených zvuků posluchači. Monografie [23] totiž uvádí, že výsledná hlasitost složeného zvuku je dána dílčími přírůstky hlasitostí z jednotlivých Barkových pásem. Přesnější popis lze nalézt v [23]. Aby bylo možno stanovit hodnoty NMR v Barkových pásmech, tak je nutné transformovat frekvenční osu na tyto Barkova pásma a současně upravit maskovací křivky a prahy slyšení. Při prvotním návrhu této metody jsem tuto transformaci neprovedl a dopustil jsem se chyby ve stanovení slyšitelnosti v případě šumů na pozadí, které měly PSD koncentrovanou na vyšších kmitočtech, popř. tyto šумы měly širokopásmový charakter. V takovém případě byly hodnoty NMR záporné, tj. PSD šumu na pozadí byla pod maskujícím prahem a tento šum na pozadí měl být pro posluchače neslyšitelný, poslechové testy ale prokázaly, že tomu tak není a tento šum pozadí slyšitelný je. Detailněji je tento použitý způsob transformace popsán v kapitole 5.2.5.

5.2.1 Popis blokového schématu navrhované metody

V této kapitole bude popsán princip navržené metody pro hodnocení slyšitelného šumu pozadí v hudebních signálech. Blokové schéma této metody je zobrazeno na Obr. 5.1.



Obr. 5.1 Blokové schéma metody.

Vstupními signály této metody jsou šum pozadí $m(n)$ a užitečný signál $x(n)$, jedná se tedy o intrusivní metodu. Lineární kombinací těchto dvou signálů vznikne vstupní šumem zarušený signál $y(n)$

$$y(n) = x(n) + m(n), \quad (5.4)$$

kde n je časový index, $n = 0, 1, \dots, N_c - 1$ a N_c je celková délka signálů $m(n)$, $x(n)$ a $y(n)$.

Výstupními signály (veličinami) budou $SpecNMR(b)$, $SpecNMR_{MAX}$ a $RelNMR(b)$. Tyto tři veličiny budou detailně popsány v následujících kapitolách. Nyní bude následovat pouze jejich stručný popis.

Veličina $SpecNMR(b)$ popisuje specifickou (ve smyslu výběrovou) průměrnou hodnotu odstupů šumu k masce (NMR, Noise to Mask Ratio). Tato veličina udává rozložení tohoto odstupů přes bářková pásma a to pro celý dlouhotrvající analyzovaný zvukový signál. Veličina $SpecNMR(b)$ je unikátní tím, že do výsledných hodnot NMR nejsou započítávány ty případy, kdy komponenty šumu na pozadí jsou maskovány komponentami užitečného signálu popředí. Tato vlastnost vychází ze zavedené úvahy z předchozí kapitoly, která zní: „Šum pozadí je slyšitelný pouze tehdy, když není maskován užitečným signálem.“

Veličina $SpecNMR_{MAX}$ je zavedena jako maximální prvek z $SpecNMR(b)$, tj. $SpecNMR_{MAX} = \max\{SpecNMR(b)\}$, a slouží pro detekci slyšitelnosti šumu na pozadí. Při návrhu bylo vycházeno z 2. výchozí úvahy z předchozí kapitoly, že: „Šum pozadí je složený zvuk a pro jeho detekci stačí vyhodnotit pouze „nejhlasitější“ komponentu.“

Poslední veličinou je $RelNMR(b)$, která popisuje tzv. relativní četnosti nemaskovaných komponent šumu pozadí. Tedy stavu, kdy komponenty šumu pozadí nejsou maskovány komponentami užitečného signálu popředí. Jedná se o doplňující údaj k veličině $SpecNMR(b)$. Tato veličina vychází z poslední zavedené úvahy, že subjektivní veličina popisující úroveň slyšitelného šumu má nestacionární charakter.

5.2.2 Segmentace

Z důvodu lepší rozlišitelnosti v časové oblasti při časově-frekvenční analýze, budou dlouhotrvající signály $m(n)$, $x(n)$ a $y(n)$ rozděleny na segmenty podle vztahu

$$y_i(n) = w(n)y(n + i384 - 64), \quad (5.5)$$

kde indexem i bude značen časový index segmentu, tj. $i = 0, \dots, I - 1$, kde I je celkový počet segmentů

$$I = \left\lfloor \frac{N_c}{384} \right\rfloor, \quad (5.6)$$

kde symbol $\lfloor \cdot \rfloor$ značí zaokrouhlení směrem k zápornému nekonečnu a $w(n)$ je Hanningovo okno

$$w(n) = \begin{cases} \sqrt{\frac{8}{3}} 0.5 \left(1 - \cos\left(2\pi \frac{n}{N}\right) \right), & n = 0, \dots, N - 1 \\ 0, & n \neq 0, \dots, N - 1 \end{cases} \quad (5.7)$$

kde N je délka jednoho segmentu a platí $N = 512$ vzorků. Výkon tohoto okna je roven

$$\frac{1}{N} \sum_{n=0}^{N-1} w^2(n) = 1. \quad (5.8)$$

Způsob této segmentace (5.5) je totožný se segmentací použitou ve standardu MPEG-1 [31]. Vstupní signály jsou segmentovány na úseky o délce 512 vzorků s překryvem 128 vzorků, vždy se překrývá prvních a posledních 64 vzorků.

Pro segmentované $y_i(n)$ pak platí

$$y_i(n) = x_i(n) + m_i(n). \quad (5.9)$$

5.2.3 Odhady výkonových spektrálních hustot

Modifikovaný periodogram [34] [52] je jedna z metod, kterou lze použít pro odhad krátkodobé výkonové spektrální hustoty z jednoho segmentu analyzovaného signálu.

$$PSD_{mmi}(k) = \frac{1}{N} \left| \sum_{n=0}^{N-1} m_i(n) e^{-j\frac{2\pi}{N}kn} \right|^2 \quad (5.10)$$

kde k je číslo spektrálního koeficientu a platí $k = 0, \dots, N-1$.

Protože bude dále v navrhované metodě použit upravený psychoakustický model ze standardu MPEG-1 vrstva 1 (model č. 1) [31], tak pro odhad výkonové spektrální hustoty z i -tého segmentu bude použit vztah z této normy, jedná se o mírnou modifikaci (5.10).

$$P'_{mmi}(k) = \left| \frac{1}{N} \sum_{n=0}^{N-1} m_i(n) e^{-j\frac{2\pi}{N}kn} \right|^2 \quad (5.11)$$

Pro vzájemný přepočít lze použít

$$\frac{P'_{mmi}(k)}{PSD_{mmi}(k)} = N. \quad (5.12)$$

Při délce segmentu $N = 512$ vzorků je tedy odhad $P'_{mmi}(k)$ přibližně o $10 \log(512) \approx 27.1$ dB větší, než "správná" hodnota $PSD_{mmi}(k)$.

Poslední krokem je normalizace hlasitosti $P'_{mmi}(k)$ na hladinu hlasitosti 92 dB SPL (Sound Pressure Level)

$$P_{mmi}(k) = P'_{mmi}(k) 10^{L_{\text{norm}}/10}, \quad (5.13)$$

kde L_{norm} je normalizační faktor a platí $L_{\text{norm}} = 92$ dB.

5.2.4 Výpočet maskujícího prahu

Pro výpočet maskujícího prahu ve spektrální oblasti je použit psychoakustický model z MPEG-1 vrstvy 1 (model č. 1) [31][49]. Maskující práh je vždy vypočten pro každý segment (tj. pro každé $x_i(n)$), a proto takovýto práh bude označován jako lokální maskující práh $LT_i(k)$ (angl. *local masking threshold*)

$$LT_i(k) = LT_q(k) + LT_{tm,i}(k) + LT_{nm,i}(k) \quad [\text{dB}], \quad (5.14)$$

kde $LT_q(k)$ je maskující práh v tichu, $LT_{tm,i}(k)$ je maskující práh tónových komponent a $LT_{nm,i}(k)$ je maskující práh komponent šumu. Všechny tyto prahy jsou v dB.

V normě [31] se namísto spektrálních koeficientů k používají koeficienty i_{sub} . Proto jsou všechny tři maskující prahy z rovnice (5.14) funkcí i_{sub} , např. $LT_q(i_{\text{sub}})$ [31].

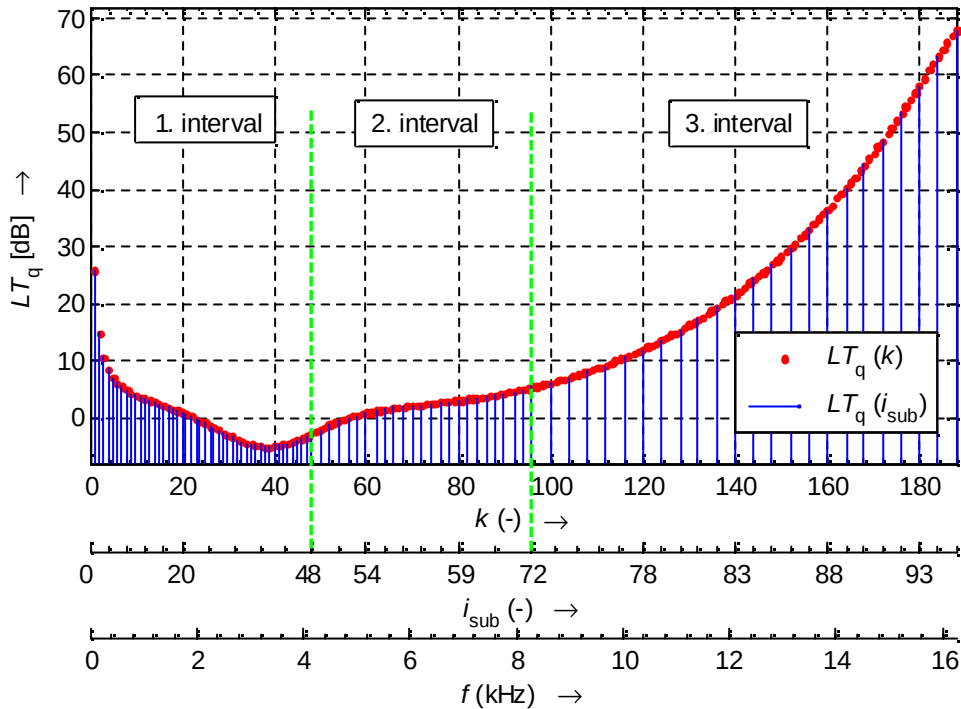
Lze říci, že koeficienty i_{sub} jsou podmnožinou koeficientů k . Při výpočtu lokálního maskujícího prahu se operuje v pásmu $<86.13, 16192>$ Hz, čemuž odpovídá lineárně rozložených 188 koeficientů k a 93 nelineárně rozložených koeficientů i_{sub} . Z důvodu zachování lineárního rozložení přes již zmíněné pásmo, a také kvůli zjednodušení následujících výpočtů, je potřeba transformovat koeficienty i_{sub} na koeficienty k . Způsob této transformace, pro případ $LT_q(i_{\text{sub}})$, je popsán vztahem (5.15) a graficky naznačen na Obr. 5.2.

$$LT_q(i_{\text{sub}}) \xrightarrow{\text{LI}} LT_q(k) \quad [\text{dB}]$$

pro $i_{\text{sub}} = 1, 2, \dots, 95$; $k = 1, 2, \dots, 188$

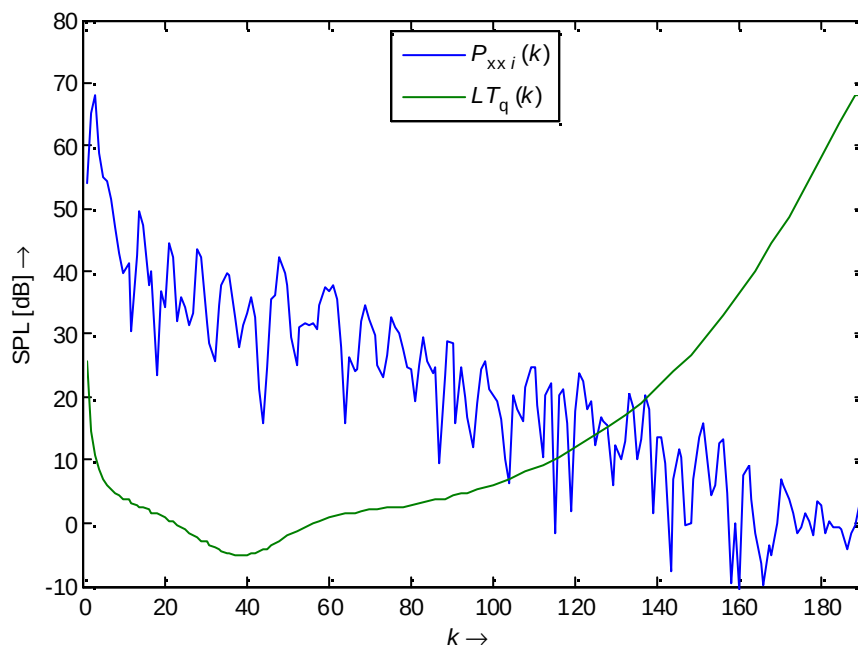
(5.15)

kde symbol LI značí lineární interpolaci.



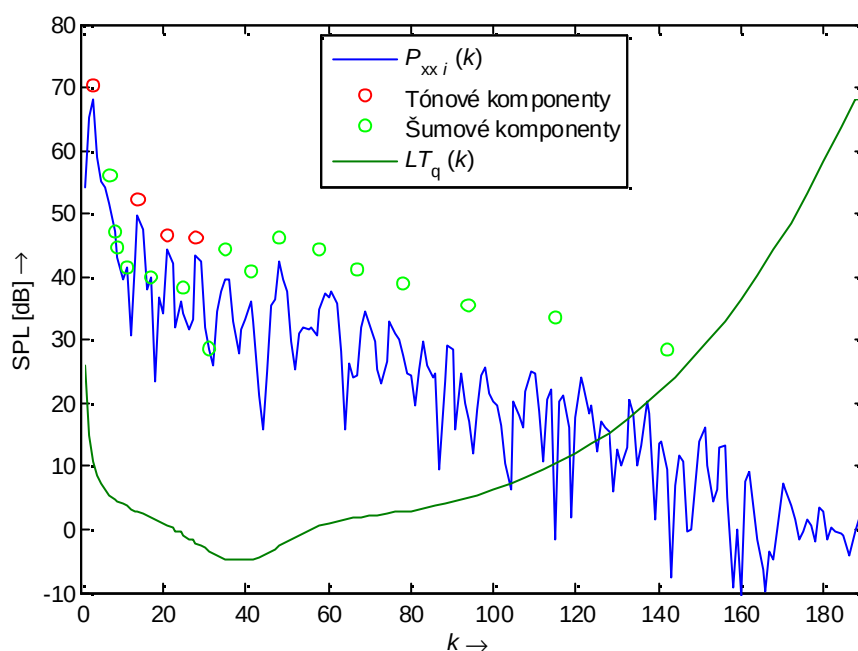
Obr. 5.2 Lokální maskující práh, grafické naznačení lineární interpolace $LT_q(i_{\text{sub}})$ na $LT_q(k)$.

Matematické definice $LT_{tm\ i}(i_{sub})$ a $LT_{nm\ i}(i_{sub})$ lze nalézt v normě [31]. Transformace na spektrální koeficienty k je opět provedena lineární interpolací. Na následujících obrázcích je graficky naznačen postup při výpočtu lokálního maskujícího práhu $LT_i(k)$. Na Obr. 5.3 je zobrazen odhad výkonové spektrální hustoty $P_{xx\ i}(k)$ (5.11) z jednoho náhodně vybraného segmentu signálu $x(n)$. Současně je v tomto grafu vyneseno $LT_q(k)$, tj. maskující práh v tichu. Lze vidět, že frekvenční pásmo užitečného signálu leží přibližně do koeficientu $k = 130$ (cca 11 kHz) nad maskou $LT_q(k)$.

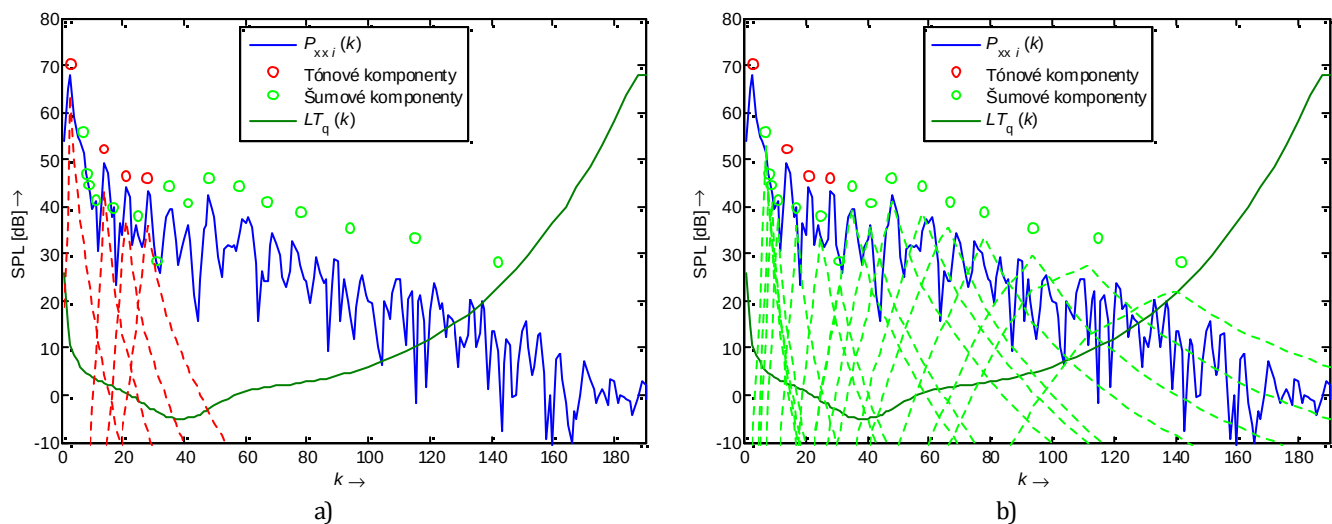


Obr. 5.3 Odhad výkonové spektrální hustoty $P_{xx\ i}(k)$ z jednoho náhodně vybraného segmentu $x(n)$ a maskující práh v tichu $LT_q(k)$ [49].

Na dalším grafu na Obr. 5.4 jsou navíc vyznačeny vrcholy tzv. tónových a šumových komponent podle [31] a na Obr. 5.5 jsou zobrazeny jednotlivé dílčí masky pro tyto komponenty.

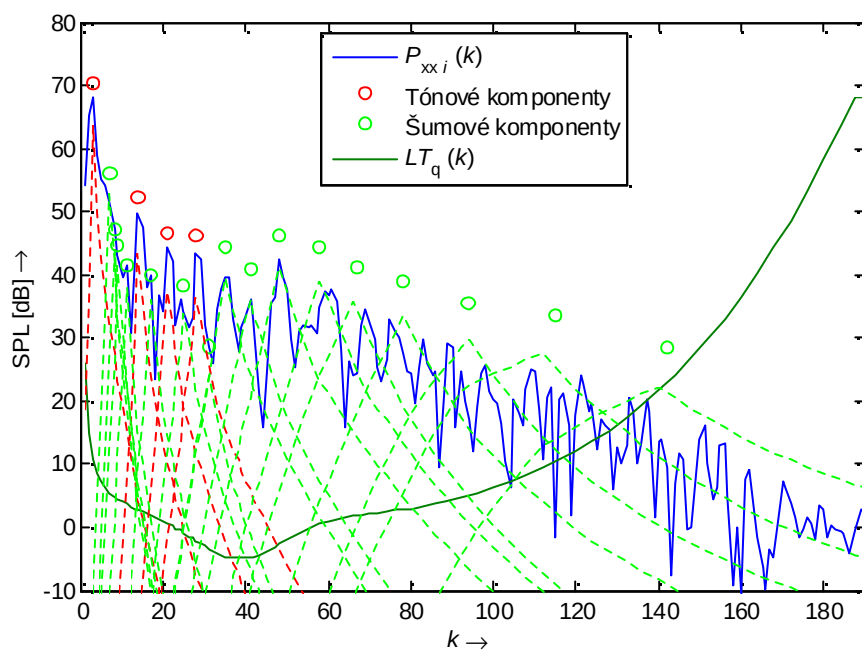


Obr. 5.4 Vrcholy tónových a šumových komponent z $P_{xx\ i}(k)$ [49].



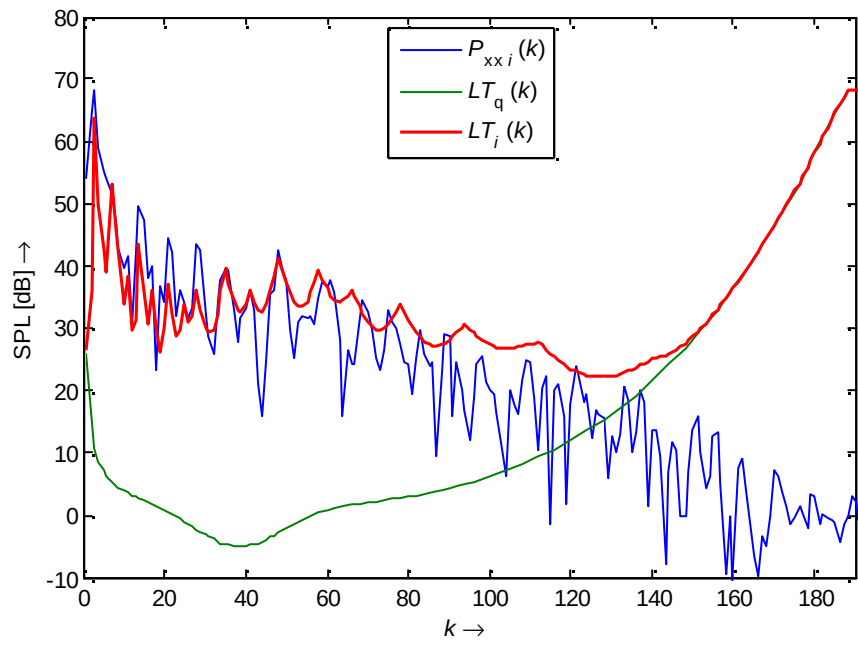
Obr. 5.5 Zobrazení masek: a) tonových komponent, b) šumových komponent [49].

Zobrazení masek tónových i šumových komponent společně v jednom grafu je zobrazeno na následujícím Obr. 5.6.



Obr. 5.6 Zobrazení masek tónových i šumových komponent společně v jednom grafu [49].

Lokální maskující práh $LT_i(k)$ vypočtený ze všech tří masek, $LT_q(k)$, $LT_{tm,i}(k)$, $LT_{nm,i}(k)$ podle vztahu (5.16), je vyobrazen červeným průběhem na Obr. 5.7.



Obr. 5.7 Lokální maskující práh $LT_i(k)$ [49].

5.2.5

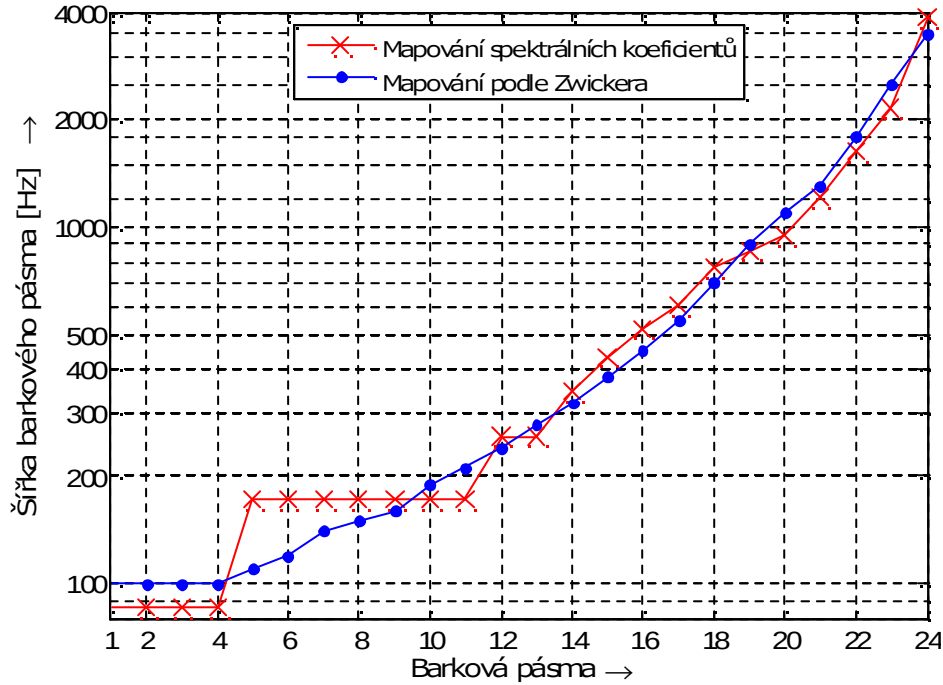
Transformace frekvenční osy v Hz na Barkova pásma

Dalším krokem při návrhu této objektivní metody je transformace frekvenční osy na Bárkova pásma (kritické pásma slyšení). Jelikož byla použita segmentace podle (5.5) tak je kmitočtové rozlišení rovno 86.13 Hz, 512 vzorků při $f_s = 44100$ Hz tj. $44100 / 512 = 86.13$ Hz. Díky tomuto omezení, bude transformace na Bárkova pásma pouze přibližná. Tab. 5.1 zobrazuje způsob transformace (mapování) spektrálních koeficientů k na pořadová čísla b příslušných Barkových pásem.

Tab. 5.1. Použitý způsob transformace na Barkova pásma.

Barkové pásma	Frekvenční pásma		Spektrální koeficienty		Barkové pásma	Frekvenční pásma		Spektrální koeficienty	
b	f		k		b	f		k	
[-]	[Hz]		[-]		[-]	[Hz]		[-]	
poř. číslo	od do	Šířka pásma	od do	Šířka pásma	poř. číslo	od do	Šířka pásma	od do	Šířka pásma
1	43,07	86,13	1	1	13	1680	258,4	20	3
	129,2		1			1938		22	
2	129,2	86,13	2	1	14	1938	344,5	23	4
	215,3		2			2283		26	
3	215,3	86,13	3	1	15	2283	430,7	27	5
	301,5		3			2713		31	
4	301,5	86,13	4	1	16	2713	516,8	32	6
	387,6		4			3230		37	
5	387,6	172,3	5	2	17	3230	602,9	38	7
	559,9		6			3833		44	
6	473,7	172,3	6	2	18	3833	775,2	45	9
	646		7			4608		53	
7	646	172,3	8	2	19	4608	861,3	54	10
	818,3		9			5469		63	
8	818,3	172,3	10	2	20	5469	947,5	64	11
	990,5		11			6417		74	
9	904,4	172,3	11	2	21	6417	1206	75	14
	1077		12			7623		88	
10	1077	172,3	13	2	22	7623	1637	89	19
	1249		14			9259		107	
11	1249	172,3	15	2	23	9259	2153	108	25
	1421		16			11413		132	
12	1421	258,4	17	3	24	11413	3876	133	45
	1680		19			15289		177	

Na dalším grafu na Obr. 5.8, je graficky zobrazený použitý způsob transformace na Barkova pásma podle Tab. 5.1. Společně s tímto průběhem je na tomto grafu také zobrazen mapování na Bárkova pásma podle Zwicker [23].



Obr. 5.8 Závislost šířky Bárkového pásma na jeho poř. číslu.

Z grafu lze vyčíst, že mapování spektrálních koeficientů k podle tabulky Obr. 5.8 dosahuje největších relativních odchylek vůči naměřeným hodnotám podle Zwicker [23] na nízkých kmitočtech. A to právě díky již zmíněnému omezenému kmitočtovému rozlišení při krátkodobé analýze vstupních segmentů.

Transformace frekvenční osy výkonové spektrální hustoty $P_{mm\ i}(k)$ na $P_{mm\ i}(b)$ je definována

$$P_{mm\ i}(b) = \sum_{k \in b} P_{mm\ i}(k) \quad [-], \quad (5.16)$$

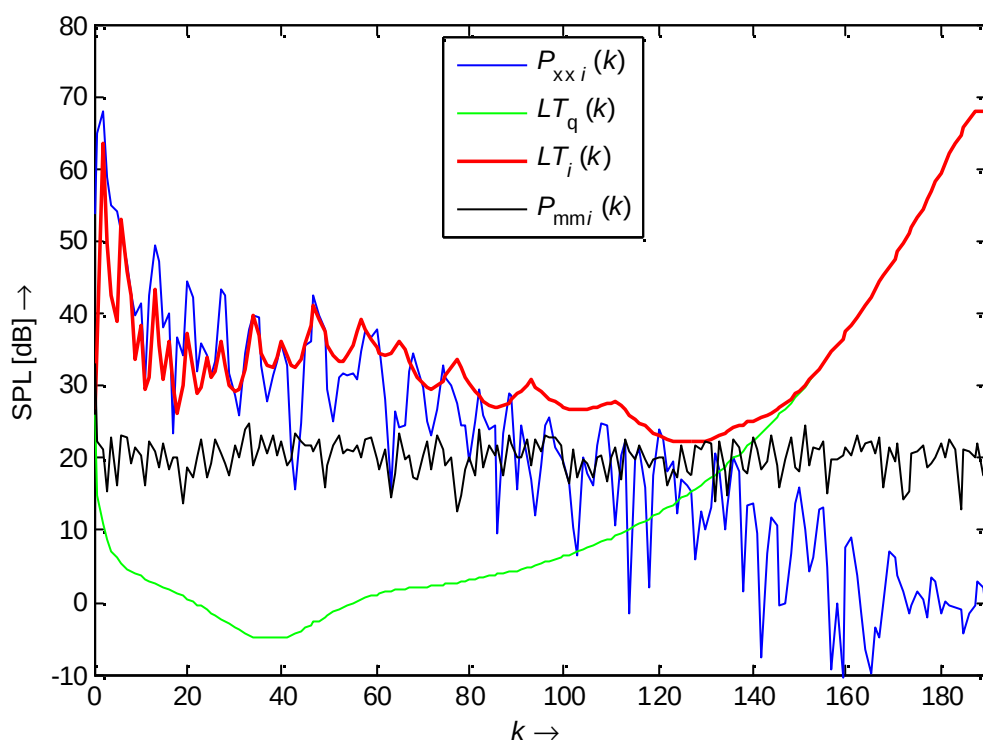
kde b jsou pořadové čísla Bárkových pásem a platí $b = 1, 2, \dots, 24$. Symbol $\sum_{k \in b}$ je použit pro součet všech hodnot spektrálních koeficientů, které přísluší danému Barkovému pásmu podle Tab. 5.1. Tento způsob sčítání výkonu přes Barkova pásma byl použit v souladu se způsobem výpočtu celkové hlasitosti podle Zwicker, nebo podle Stevense [73] [18] [23]. V obou případech se totiž sčítají výkony uvnitř jednotlivých Barkových pásem.

Pro výpočet maskujícího prahu v tichu $LT_q(b)$ z $LT_q(k)$ byl použit odlišný postup. Nesčítají se zde výkony této masky přes jednotlivé Barkova pásma (5.16), ale je vybrán medián z hodnot $LT_q(k)$ pro k ležící uvnitř daného pásma b

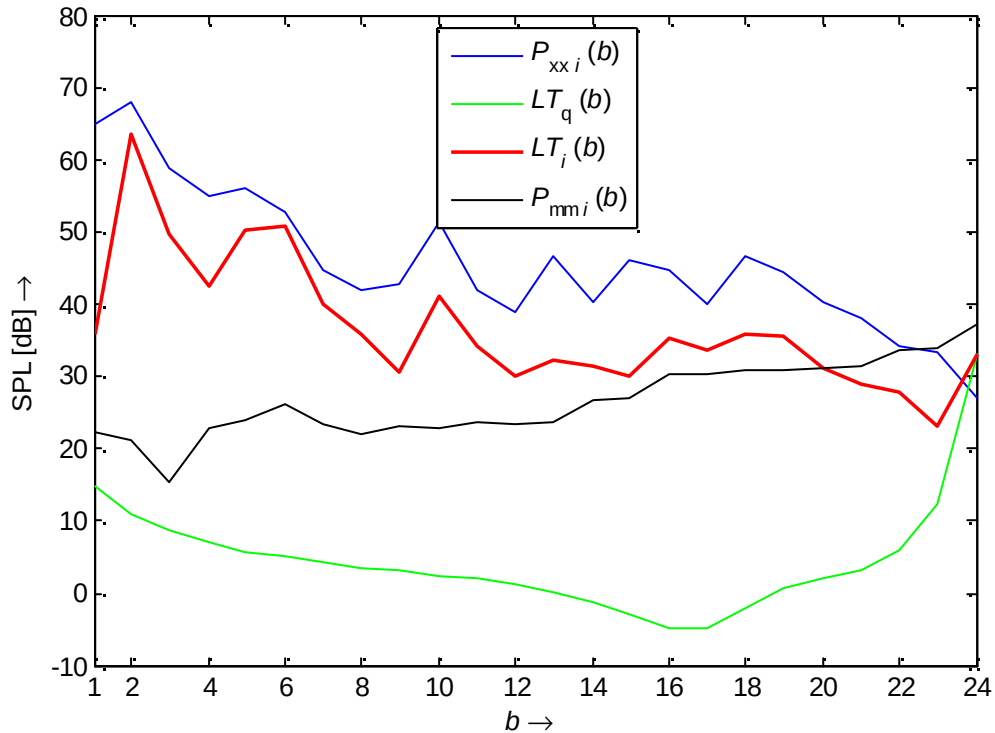
$$LT_q(b) = \text{median}\{LT_q(k)\}_{k \in b}. \quad (5.17)$$

Tento postup byl zvolen, protože křivka popisující maskující práh v tichu byla změřena pro maskování jednoho tónu a ne nějakého obecného složeného zvuku [44] [23]. Tedy maska v tichu $LT_q(k)$ popisuje maskující práh jednoho jediného tónu. Pro skupinu tónů, nebo nějaký úzkopásmový šum, tuto masku již nelze použít [23]. Pokud má tón například svoji úroveň menší než je maskující práh v tichu, tak to lze interpretovat tak, že tento tón není slyšitelný. Pokud ovšem jsou tyto tóny dva, tak může nastat situace, že budou slyšitelné. Budou například velmi blízko sebe, řekněme uvnitř jednoho kritického pásma, tak takovéto tóny budou splývat v jeden tón [23]. Křivka popisující maskující práh v tichu byla získána metodou nastavování podnětu na skupině lidí, kterou lze označovat za náhodný výběr. Tedy typ výběru, kdy testovacími subjekty jsou náhodně vybraní posluchači z celkové populace, tj. mladí – staří, ženy – muži, dospělí – děti atd. Křivka popisující maskující práh v tichu, například ta z Obr. 5.2, byla získána jako mediánová hodnota z hodnot ze subjektivních soudů jednotlivých posluchačů. Individuální hodnoty této masky se ale velmi liší, příkladem může být horní hranice slyšitelnosti tónů o vysokých kmitočtech, která klesá s přibývajícím rokem z původní hranice cca 20 kHz na cca 16 kHz v dospělosti. Z těchto důvodů byl zvolen vztah pro výpočet $LT_q(b)$ z $LT_q(k)$ podle (5.17), kde se vybírá median ze všech $LT_q(k)$ pro k ležící uvnitř daného pásma b . Stejným způsobem je vypočten $LT_i(b)$ z $LT_i(k)$.

Na následujících obrázcích, Obr. 5.9 a Obr. 5.10, jsou vykresleny příklady typických průběhů masek a výkonových spektrálních hustot před a po transformaci frekvenční osy na Barkova pásma.



Obr. 5.9 Průběhy lokálních masek a výkonových spektrálních hustot vybraných průběhů před transformací frekvenční osy na Barkova pásma.



Obr. 5.10 Průběhy lokálních masek a výkonových spektrálních hustot vybraných průběhů po transformaci frekvenční osy na Barkova pásma.

Na těchto grafech si lze všimnout několika zajímavých skutečností. První z nich je zobrazení maskujícího prahu v tichu před a po transformaci, zelený průběh, kde se “posunulo” lokální minimum (oblast kolem 3 kHz) této funkce z levé půlky grafy, Obr. 5.9, na pravou, Obr. 5.10. To přibližně odpovídá zlogaritmováním x-ové osy spektrálních koeficientů k .

Dále zobrazení rozložení výkonové spektrální hustoty maskujícího signálu $P_{xx_i}(k)$ a $P_{xx_i}(b)$, modrý průběh, a jeho lokálního maskujícího prahu $LT_i(k)$ a $LT_i(b)$, červený průběh. V prvním grafu, Obr. 5.9, je přibližně od $k > 50$ lokální maskující práh nad maskujícím signálem. Z toho typu zobrazení to vypadá, že tyto maskující komponenty maskují samy sebe, což samozřejmě není možné. Lépe je vše zobrazeno na následujícím grafu, Obr. 5.10, kde červený průběh, reprezentující lokální maskující práh $LT_i(b)$, leží správně pod výkonovou spektrální hustotou $P_{xx_i}(b)$ maskujícího signálu $x_i(n)$, modrý průběh.

Posledním průběhem z těchto obrázků je odhad výkonové spektrální hustoty $P_{mm_i}(k)$ šumu pozadí. Šum pozadí je v tomto případě bílý šum svýkonem 20 dB SPL. $P_{mm_i}(k)$ leží pod lokálním maskujícím prahem $LT_i(k)$, Obr. 5.9. Po transformaci frekvenční osy na Barkova pásma se $P_{mm_i}(b)$ dostává pro $b > 20$ nad lokální maskující práh. Zobrazení z grafu Obr. 5.10 odpovídá více skutečnosti, že tento šum pozadí je slyšitelný na vyšších kmitočtech. Spektrum bílého šumu po této transformaci již není rovnoměrně rozloženo přes všechny Barkova pásma, uvažujme $b = 1, 2, \dots, 24$, ale od cca 4-6 pásma vzrůstá výkonová spektrální hustota přibližně 10 dB na oktávu. V psychoakustice se používá pojem *rovnoměrně rozložený maskující šum* (angl. *uniform masking noise*) pro označení takového typu šumu, který je rovnoměrně rozložený přes uvažovaná Barkova pásma [23].

5.2.6 Stanovení odstupu šumu k masce

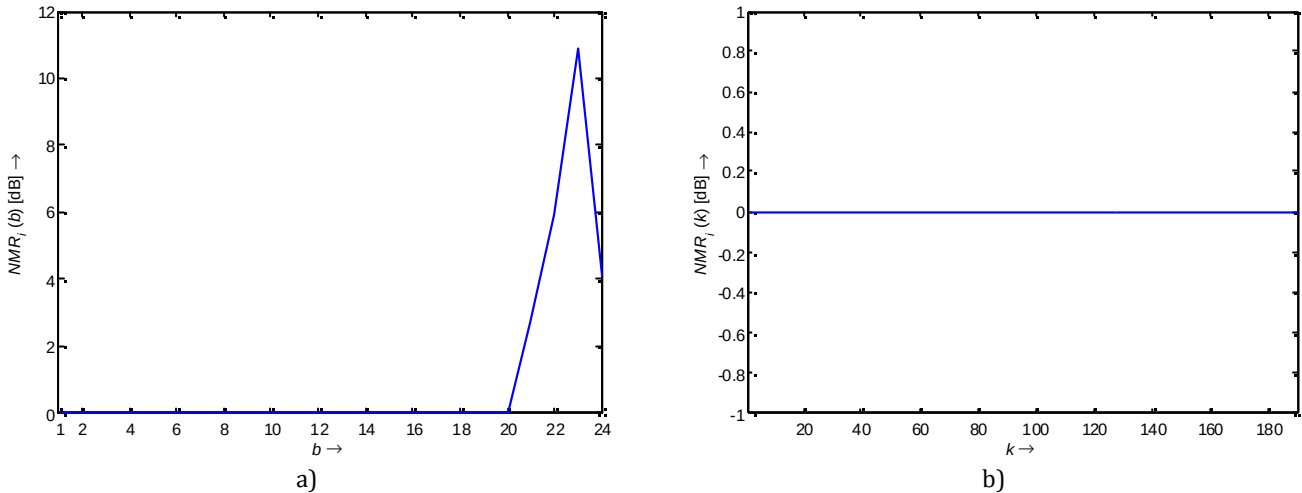
Prvním krokem ke stanovení slyšitelnosti šumu pozadí je zavedení modifikované varianty odstupu úrovně šumu k úrovni masky (angl. *noise to mask ratio*, NMR). Tato objektivní míra byla zavedená v roce 1987 [16][57]. Modifikace této metody spočívá v nelineární charakteristice NMR mezi úrovní šumu a masky, a to v závislosti na výsledku podmínky, zda úroveň šumu je nebo není větší než úroveň masky, tedy

$$NMR_i(b) = \begin{cases} \frac{P_{mmi}(b)}{LT_i(b)}, & P_{mmi}(b) \geq LT_i(b) \\ 1, & P_{mmi}(b) < LT_i(b). \end{cases} \quad (5.18)$$

Tento vztah se snaží reflektovat skutečnost, že u objektivních metod nemá význam pro stanovení slyšitelného šumu pozadí v zvukových signálech, vyhodnocovat úroveň neslyšitelného šumu pozadí. Pokud pro nějaké i , nebo b platí, že $P_{mmi}(b) < LT_i(b)$ tak $NMR_i(b) = 1$ a tento výsledek lze interpretovat tak, že komponenta šumu pozadí z $P_{mmi}(b)$ pro dané i a b není slyšitelná.

Na Obr. 5.11 je zobrazen průběh $NMR_i(b)$ pro průběhy $LT_i(b)$ a $P_{mmi}(b)$ z Obr. 5.9 a Obr. 5.10. Navíc je zde zobrazen průběh $NMR_i(k)$ vypočtený podle vztahu (5.19). Tedy jednou je zobrazen průběh pro případ transformované frekvenční oblasti na Barkova pásma a jednou netransformované. Cílem těchto zobrazení je porovnat obě varianty NMR a poukázat na skutečnost, že pokud není provedena transformace na Barkova pásma tak hodnota NMR může být nulová v celém průběhu, viz Obr. 5.11 b), a to i v případě, že šum pozadí je slyšitelný na vyšších kmitočtech, viz Obr. 5.11 a).

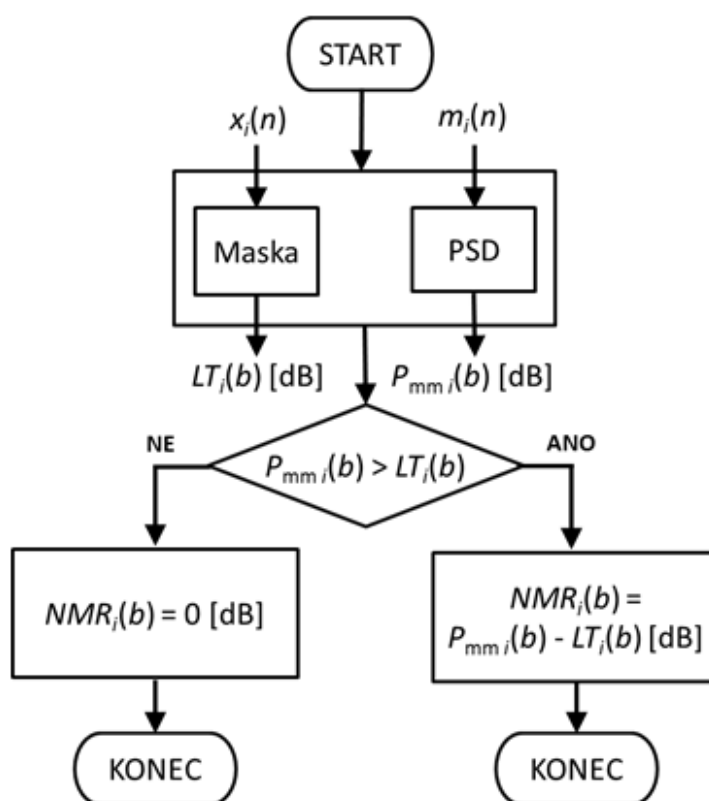
$$NMR_i(k) = \begin{cases} \frac{P_{mmi}(k)}{LT_i(k)}, & P_{mmi}(k) \geq LT_i(k) \\ 1, & P_{mmi}(k) < LT_i(k). \end{cases} \quad (5.19)$$



Obr. 5.11 Modifikovaná varianta NMR: a) charakteristika $NMR_i(b)$, b) charakteristika $NMR_i(k)$.

Na Obr. 5.11 a) lze vidět, že v pásnu $b = 1$ až 20 je hodnota $NMR_i(b) = 0$ dB, protože $LT_i(b)$ je větší než $P_{mmi}(b)$. Šum pozadí lze zde správně interpretovat jako neslyšitelný. V pásmech $b = 21$ až 24 je hodnota $NMR_i(b)$ větší než 0 dB, $P_{mmi}(b)$ je zde nad lokálním maskujícím prahem $LT_i(b)$. Proto lze zde očekávat, že v tom pásnu bude šum pozadí slyšitelný. Podobné úvahy o slyšitelnosti některých komponent šumu pozadí nelze získat, pokud frekvenční osa není transformovaná na Barkova pásma.

Na Obr. 5.12 je zobrazen navržený algoritmus pro výpočet $NMR_i(b)$. Vstupními signály do tohoto algoritmu jsou $x_i(n)$ a $m_i(n)$, dále je zde graficky naznačen postup výpočtu $P_{mmi}(b)$ a $LT_i(b)$ a nakonec je zobrazen rozhodovací blok pro porovnání velikosti $P_{mmi}(b)$ a $LT_i(b)$ a stanovení výsledné hodnoty $NMR_i(b)$.



Obr. 5.12 Algoritmus výpočtu $NMR_i(b)$.

5.2.7 Stanovení relativní četnosti nemaskovaných komponent šumu na pozadí

Velkým problémem při stanovení slyšitelnosti šumu pozadí v zvukových signálech je obecně jejich nestacionární charakter. Je tím myšlena skutečnost, že šum pozadí bývá slyšitelný v rozdílných úsecích zvukových signálu rozdílně. V řečových signálech je nejhlasitější v úsecích bez řečové aktivity a v hudebních signálech je šum pozadí nejhlasitěji vnímán nejčastěji na začátku, na konci, popřípadě také v tichých úsecích. Důsledkem této skutečnosti pak je, že subjektivní vnímání slyšitelnosti šumu pozadí je taktéž časově proměnné. V této kapitole je snaha zakomponovat vliv časově proměnného subjektivního vnímání slyšitelnosti šumu pozadí do navrhované metody. Prvním krokem je zavedení veličiny popisující absenci masky, tedy stav, kdy je $P_{mmi}(b) \geq LT_i(b)$. Absence masky $MA_i(b)$ je definována jako

$$MA_i(b) = \begin{cases} 1, & P_{mmi}(b) \geq LT_i(b) \\ 0, & P_{mmi}(b) < LT_i(b). \end{cases} \quad (5.20)$$

A celkový počet segmentů kdy tento stav nastává je

$$I_{MA}(b) = \sum_{i=0}^{I-1} MA_i(b). \quad (5.21)$$

Dále bude zavedena veličina $SpecNMR(b)$, tj. *specifické* (ve smyslu výběrové) *NMR*, pro popis průměrné hodnoty $NMR_i(b)$ přes ty segmenty, kdy je výkonová spektrální hustota větší než lokální maskující práh

$$SpecNMR(b) = \frac{1}{I_{MA}(b)} \sum_{i=0}^{I-1} NMR_i(b) MA_i(b). \quad (5.22)$$

Pro samotnou detekci slyšitelnosti šumu pozadí bude zavedena veličina

$$SpecNMR_{MAX} = \max\{ SpecNMR(b) \}, \quad (5.23)$$

kde symbol $\max\{ \}$ značí výběr největšího prvku z posloupnosti $SpecNMR(b)$.

Protože $SpecNMR(b)$ vyjadřuje velikost průměrné hodnoty $NMR_i(b)$ přes ty segmenty, kdy je výkonová spektrální hustota šumu pozadí větší než lokální maskující práh, tak je potřeba doplnit tento údaj o informaci, z kolika segmentů je $SpecNMR(b)$ vypočteno. Může se jednat například jen o několik chybně určených segmentů díky špatnému odhadu výkonové spektrální hustoty $P_{mmi}(b)$ a tedy desinformací, nebo o korektní údaj způsobený malým množstvím "slyšitelných" segmentů. Z těchto důvodů bude zavedena veličina $RelNMR(b)$, *Relative NMR*, popisující relativní četnosti nemaskovaných komponent šumu na pozadí

$$RelNMR(b) = \frac{I_{MA}(b)}{I} 100 \quad [\%]. \quad (5.24)$$

Veličiny $SpecNMR(b)$, $SpecNMR_{MAX}$ a $ReINMR(b)$ jsou výstupními charakteristikami navržené objektivní metody pro hodnocení slyšitelného šumu pozadí v hudebních signálech, viz Obr. 5.1.

5.2.8 Ukázky konkrétních příkladů šumem degradovaných zvukových záznamů

V této podkapitole budou popsány konkrétní příklady degradovaných záznamů zvukových signálů. Tyto signály budou degradovány aditivním šumem pro různě zvolené hodnoty SNR . SNR bude vždy experimentálně volena tak, aby se šum pozadí vyskytoval na mezi slyšitelnosti, popř. v jeho blízkém okolí. To znamená, že pokud například bude většina výkonu šumu pozadí ležet mimo frekvenční pásma užitečného signálu, tak budou hodnoty SNR nabývat vysokých hodnot, klidně i přes 50 dB a šum pozadí může být pořád slyšitelný. V opačném případě, tedy pokud je šum pozadí výrazně maskován, budou hodnoty SNR podstatně menší.

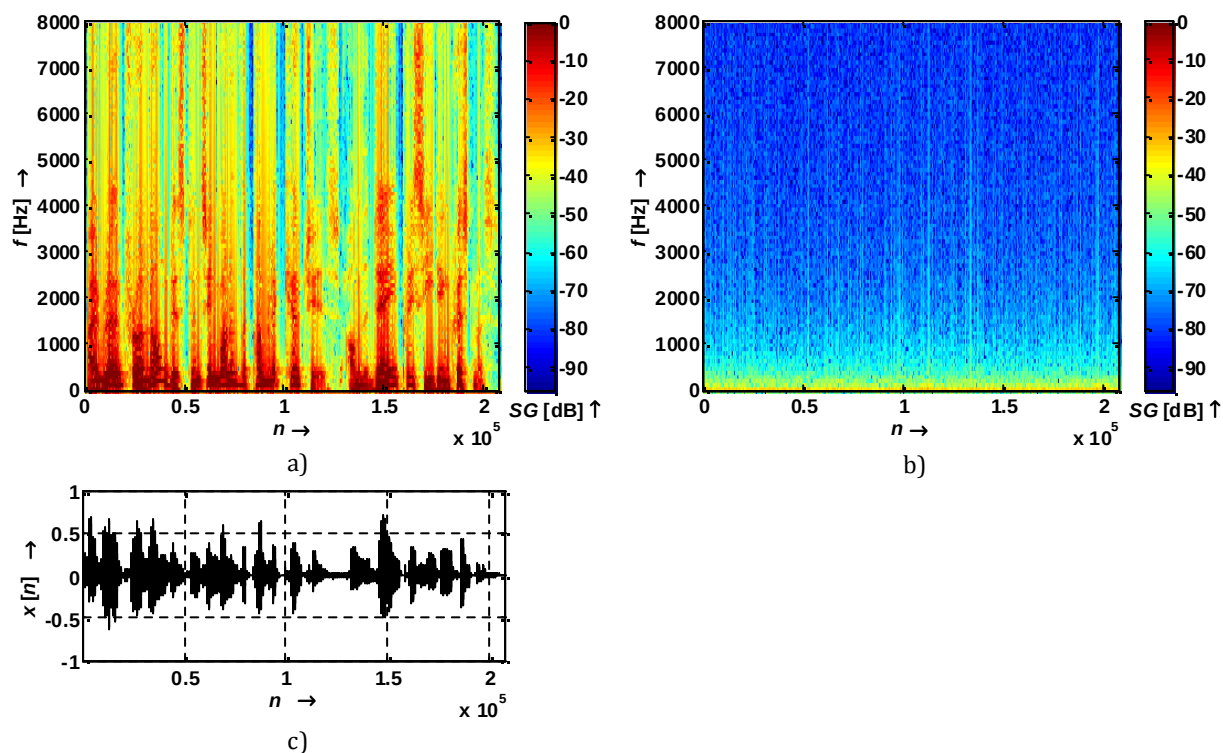
Cílem této podkapitoly není stanovit, zda šum pozadí je pro konkrétní příklad slyšitelný či nikoliv, ale spíše demonstrovat navrženou objektivní metodu. Každý takovýto demonstrační příklad bude obsahovat kombinaci užitečného hudebního nebo řečového signálu z databáze [72] a šumu pozadí. Tyto šumy pozadí budou reálně se vyskytující typy šumů, konkrétně: šum větráku, šum poškrábané vinylové desky a šum audio pásky.

Celková délka všech signálů bude vždy 5 sekund, $f_s = 44\,100$ Hz čemuž odpovídá celková délka signálu $N_c = 220\,500$ vzorků. Pro každou kombinaci zvukový signál + šum pozadí bude volena jiná hodnota SNR definovaná jako

$$SNR = 10 \log \left(\frac{\sum_{n=0}^{N_c-1} x^2(n)}{\sum_{n=0}^{N_c-1} m^2(n)} \right). \quad (5.25)$$

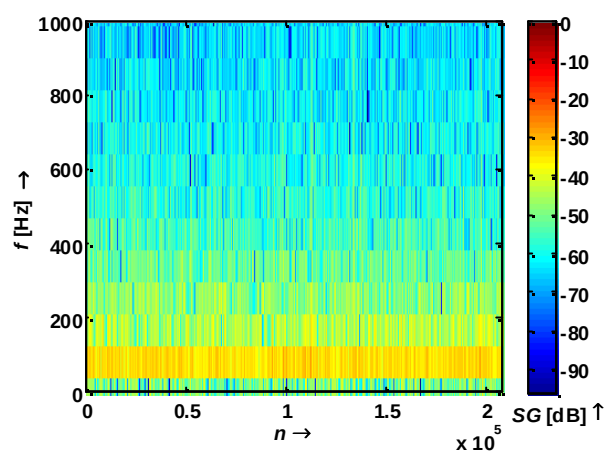
Příklad č. 1, záznam řeči a šumu kazety při $SNR = 40$ dB.

Prvním demonstrativním příkladem bude záznam promluvy anglicky hovořícího mužského mluvčího [72]. Spektrogram tohoto zvukového signálu je zobrazen na Obr. 5.13 a) a časový průběh na Obr. 5.13 c). Z obou obrázků jsou viditelné krátké úseky ticha, např. v okolí $n = 1.2 \times 10^5$, v této promluvě. Lze očekávat, že v těchto úsecích bude šum pozadí nejvýraznější.



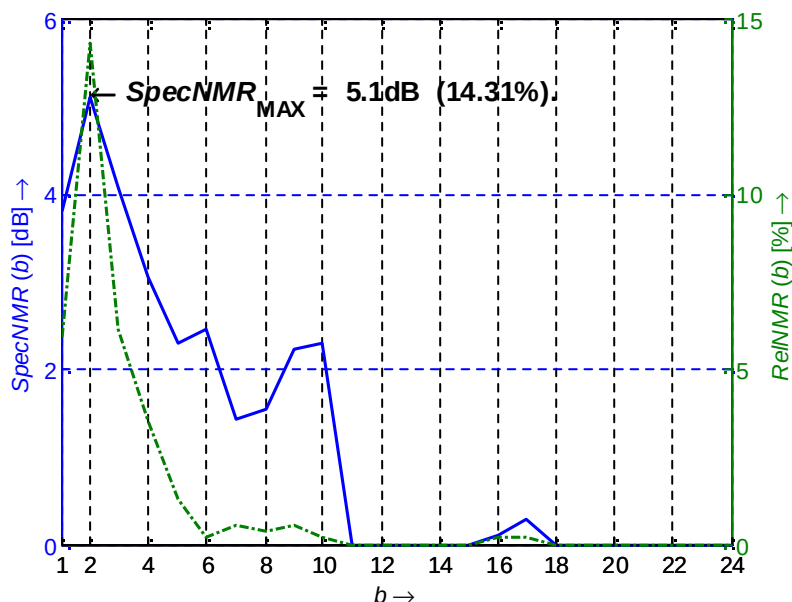
Obr. 5.13 Záznam řeči [72]: a) spektrogram, b) spektrogram šumu pozadí, šum audio pásky, c) časový průběh.

Šumem pozadí je šum audio pásky, viz spektrogram na Obr. 5.13 c). Většina výkonu tohoto šumu je koncentrována v pásmu v okolí 100 Hz, viz detailní spektrogram na Obr. 5.14.



Obr. 5.14 Detail spektrogramu šumu pozadí, šum audio pásky.

objektivní vyhodnocení velikosti slyšitelného šumu pozadí pomocí navržené objektivní metody pomocí veličin $SpecNMR(b)$, $SpecNMR_{MAX}$ a $RelNMR(b)$ je zobrazeno v grafu na Obr. 5.15.



Obr. 5.15 Výsledné charakteristiky záznamu řeči a šumu kazety při $SNR = 40$ dB.

Osa x je vynesena v Barkových pásmech v rozmezí $b = 1$ až 24, čemuž odpovídá přibližně logaritmické rozdělení kmitočtů od cca 43 Hz až po cca 15 288 Hz, viz Tab. 5.1. Největších hodnot $SpecNMR(b)$ je dosaženo pro pásma $b = 1$ až 3. A relativní četnost nemaskovaných komponent šumu na pozadí je přibližně 10 % v tomto pásmu. To znamená, že počet segmentů, kde platí $P_{mmi}(b) \geq LT_i(b)$, tj. segmentů ze kterých byly vypočteny hodnoty $SpecNMR(b)$, je přibližně 10 % ze všech segmentů. Maximální hodnoty $SpecNMR(b)$ je dosaženo v 2. Barkovém pásmu, kde $SpecNMR(2) = 5.1$ dB a $RelNMR(2) = 14.31$ %. Tato maximální hodnota je označena jako $SpecNMR_{MAX}$ (5.23).

Graf z Obr. 1.1 lze interpretovat tak, že slyšitelná část šumu pozadí je koncentrována na nízkých kmitočtech, jeho velikost je přibližně o 5 dB větší, než je maskující práh a vyskytuje se přibližně ve 14 % procentech signálu. Tento závěr není až tak překvapivý, když se porovná rozložení výkonové spektrální hustoty šumu, Obr. 5.13 c), a celková délka “tichých” úseků v promluvě, Obr. 5.13 a-b).

V příloze A jsou uvedeny další příklady aditivním šumem degradovaných hudebních a řečových signálů, na kterých je demonstrována navržená objektivní metoda pro objektivní vyhodnocení velikosti slyšitelného šumu pozadí.

5.3 Poslechový experiment

5.3.1 Subjektivní měření prahu slyšitelnosti šumu pozadí pro testované zvukové signály

Cílem tohoto experimentu bylo stanovit subjektivní práh slyšitelnosti šumu pozadí na testovaných zvukových ukázkách. Výsledky tohoto testu byly následně použity při hodnocení navržené objektivní metody. Srovnávali se výsledky získané z poslechových testů a z navržené objektivní metody a poté se měřila míra shody těchto výsledků. Dále byly tyto výsledky porovnávány s výsledky získanými z objektivní metody PEAQ z implementace [37].

Posluchačům bylo v tomto poslechovém experimentu předloženo několik sad zvukových ukázek s různými hodnotami SNR (5.25) mezi užitečným signálem a šumem pozadí, a tito posluchači subjektivně určili tu ukázkou, ve které již šum pozadí neslyší. Tyto hodnoty SNR budou dále v práci nazývány jako *prahy slyšitelnosti šumu pozadí*. Rozsah hodnot SNR pro každou sadu zvukových ukázek byl v rozsahu 10 dB až 70 dB s krokem 5 dB, tj. 13 ukázek. Takovýto rozsah hodnot SNR byl zvolen, protože vždy platilo, že šum pozadí byl pro $SNR = 10$ dB velmi výrazný a pro hodnotu 70 dB nebyl slyšitelný. Posluchačům byla dále předložena originální zvuková ukázka a ukázka, která obsahuje pouze šum pozadí. Posluchačům bylo tedy předloženo celkem 15 zvukových ukázek v každé sadě.

Takto velký rozsah hodnot SNR byl zvolen také proto, aby pokryl rozdíly v hodnocení prahu slyšitelného šumu pozadí v dB pro různé kombinace užitečného signálu popředí a šumu pozadí. Hodnota prahu slyšitelnosti šumu pozadí je totiž výrazně závislá na tom, jak moc je šum pozadí maskován užitečným signálem popředí. Pokud je například šum pozadí velmi málo maskován, tak práh slyšitelnosti tohoto šumu pozadí může dosahovat hodnoty i přes 50 dB, tj. i při takto vysokých hodnotách bude šum pozadí pro posluchače slyšitelný. V opačném případě, tj. kdy šum pozadí je výrazně maskován, tak se tyto prahy slyšitelnosti pohybují v okolí hodnot 20 dB.

Realizovaný poslechový experiment byl rozdělen na dvě části. První část se zaměřuje na stanovení prahu slyšitelnosti šumu pozadí u hudebních a řečových signálů a v druhé části u umělých zvukové signálů. Seznamy použitých signálů pro obě části experimentu lze nalézt v Tab. 5.2 a Tab. 5.3. V prvním sloupci obou tabulek jsou vypsány použité užitečné signály popředí a v druhém sloupci vždy tři šumy pozadí pro konkrétní užitečný signál popředí. Tyto tři šumy pozadí degradují užitečný signál pro výše zmíněné hodnoty SNR v rozsahu 10 až 70 dB. V Tab. 5.2 jsou tyto tři šumy pozadí vždy stejné, u Tab. 5.3 to ale již neplatí.

Tab. 5.2 Seznam hudebních a řečových zvukových signálů pro 1. část experimentu.

Užitečné signály popředí	Šum pozadí
Vážná hudba (Mozart)	Šum audio kazety
	Šum větráku
	Šum vinylové desky
Pop music (Abba)	Šum audio kazety
	Šum větráku
	Šum vinylové desky
Country (Eddie Rabbit)	Šum audio kazety
	Šum větráku
	Šum vinylové desky
Hra na kytaru	Šum audio kazety
	Šum větráku
	Šum vinylové desky
Sborový zpěv (vokál)	Šum audio kazety
	Šum větráku
	Šum vinylové desky
Mužská řeč	Šum audio kazety
	Šum větráku
	Šum vinylové desky

Užitečné typy signály z Tab. 5.2 byly vybrány z databáze SQAM, *Sound Quality Assessment Material* [72]. Spektrogramy všech signálů z této tabulky jsou zobrazeny v příloze B. Při výběru vhodných signálů pro tuto část experimentu byla snaha zvolit takové signály, aby pokryly co největší oblast hudebních a řečových zvukových signálů. V užitečných signálech jsou hudební ukázky rozdílných hudebních stylů, hry na samostatný hudební nástroj, sborový zpěv a řeč. Stejným způsobem byly vybrány šumy pozadí, které reprezentují často se vyskytující typy šumů ve zvukových systémech.

Tab. 5.3 Seznam umělých zvukových signálů pro 2. část experimentu.

Užitečný signál popředí	Šum pozadí
Růžový šum	Bílý šum
	Tón 1 kHz
	Deset kmitočtů v pásmu 10 kHz až 14 kHz
Bílý šum	Růžový šum
	Tón 1 kHz
	Deset kmitočtů v pásmu 10 kHz až 14 kHz
Tón 1 kHz	Růžový šum
	Bílý šum
	Deset kmitočtů v pásmu 10 kHz až 14 kHz
Deset kmitočtů v pásmu 10 kHz až 14 kHz	Růžový šum
	Bílý šum
	Tón 1 kHz

Pro druhou část experimentu byly použity celkem 4 signály, viz Tab. 5.3. Vždy je jeden z těchto signálů zvolen jako užitečný signál popředí a zbytek jako šum pozadí a poté se signály prohodí. Spektrogramy, odhady PSD a časové průběhy těchto umělých zvukových signálů jsou zobrazeny v příloze B.

Parametry poslechového materiálu jsou přehledně vyneseny v Tab. 5.4.

Tab. 5.4 Parametry poslechového materiálu.

Parametry poslechového materiálu	
Vzorkovací kmitočet	44 100 Hz
Délka signálu v sekundách	5 s
Délka signálu ve vzorcích	220 500

5.3.2 Charakteristika poslechového experimentu

Při charakteristice a klasifikaci tohoto poslechového experimentu bude použita odborná terminologie z monografií a učebnic [44] [43] [21].

Pro měření prahu slyšitelnosti šumu pozadí v obou částech experimentu byla použita metoda *nastavování zvukových podmětů*. Standardně jsou v této metodě posluchačům předkládány dva podněty: *standardní podnět* a *variabilní podnět*. Velikosti obou podnětů se na začátku experimentu většinou velmi výrazně liší a úkolem posluchače je nastavit variabilní podnět tak, aby byl “shodný” se standardním podnětem. V provedeném experimentu nebylo použito standardního podnětu ale pouze variabilního podnětu. Variabilním podnětem byly písmeny označené zvukové ukázky, písmena A až M. Posluchače si postupně přehrávali všechny ukázky a označili tu ukázku, kde šum pozadí pro ně nebyl slyšitelný. Podobnou metodou, tj. nastavování zvukových podnětů bez

použití standardního podnětu, byly například zjištěny hodnoty hladiny akustického tlaku absolutního prahu slyšení čistých tónů, viz Békésyho audiometrie [44] [48].

Z pohledu toho, jakým způsobem byli posluchači řízeni, lze experiment klasifikovat jako *částečně řízený experiment*. K řízení posluchačů byl totiž použit software, ve kterém si posluchači sami aktivně spouštěli testované zvukové ukázky.

Podle počtu posluchačů přítomných při jednom experimentálním sezení, lze experiment klasifikovat jako *individuální experiment*. Experiment byl totiž vždy prováděn s každým posluchačem jednotlivě.

Z hlediska podmínek, v nichž se experiment uskutečňuje, lze experiment klasifikovat jako *laboratorní experiment*, protože se experiment prováděl na měřícím pracovišti pro provádění poslechových testů, viz Obr. 5.16 b).

Z hlediska časové posloupnosti, lze experiment klasifikovat, jako *projektový experiment*, protože tento experiment byl předem záměrně připravený pokusem.

A podle cílů experimentu, lze experiment klasifikovat, jako *kritický experiment*. V tomto experimentu byly totiž ověřovány hypotézy založené na teoretických úvahách z kapitoly 5.1.

Podle výběru posluchačů, lze experiment klasifikovat jako *experiment s libovolným (příležitostným) výběrem*, neboli také *výběr na základě dostupnosti*. Většina posluchačů jsou totiž studenti doktorského nebo magisterského studia na ústavu telekomunikací na VUT v Brně, kteří byli ochotní se tohoto experimentu zúčastnit.

5.3.3 Příprava a průběh poslechového experimentu

Poslechový experiment byl prováděn pomocí kvalitních audio soustavy tvořené zejména kvalitními studiovými sluchátky *Beyerdynamic DT 770 pro* a externí zvukovou kartou *M-AUDIO proFire 610* na pracovišti z Obr. 5.16 b). Změřená přenosová charakteristika této audio soustavy je zobrazena na Obr. 5.17. Toto měření bylo prováděno na figuríně KEMAR (*Knowles Electronics Manikin for Acoustic Research*), Obr. 5.16 a).

Použité programové vybavení bylo velmi jednoduché. Obsahovalo sérii tlačítek, které sloužili k přehrání zvukových ukázek. Posluchači si stiskem těchto tlačítek přehrají testované zvukové ukázky označené písmeny A až M, kde písmeno A značí nejhlasitější šum pozadí, tj. $SNR = 10$ dB. Další písmeno B značí $SNR = 15$ dB, C značí $SNR = 20$ dB až M značí $SNR = 70$ dB, tj. zvukovou ukázku s nejnižším šumem pozadí. Další dvě tlačítka přehrají pouze šum pozadí, nebo užitečný signál popředí. Tyto dvě tlačítka slouží hlavně k lepší orientaci v přehrávání testovaných zvukových ukázek.

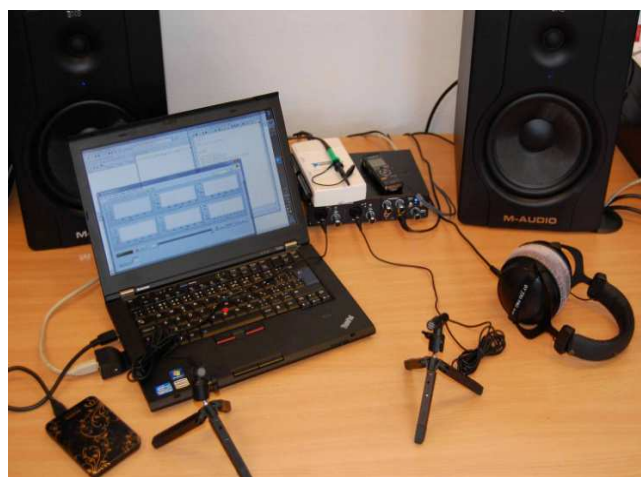
Po výběru zvukových ukázky s šumem pozadí, který není pro posluchače slyšitelný, je označeno vybrané hodnocení a tlačítkem "Potvrdit" je potvrzen tento výběr. Program automaticky průběžně ukládá vyhodnocené výsledky a po ohodnocení všech zvukových ukázek je program ukončen.

Hodnocení celé série zvukových ukázek probíhá vždy dvakrát. Nejprve v tzv. *testu* a poté tzv. *retestu*. Rozdíl mezi hodnocením v testu a retestu slouží hlavně k vyhodnocení stability celého měření. Například je zde testováno, zda posluchači nehodnotí příliš rozdílně při prvním a druhém poslechu stejné zvukové ukázky, tj. v testu a retestu.

Při spuštění tohoto softwaru, se vždy automaticky generuje náhodné pořadí přehrávaných zvukových ukázek. Tímto přístupem je snaha omezit vliv paměti na subjektivní soudy posluchačů. Přesněji řečeno, tento přístup omezuje *dlouhodobé časové vlivy* ovlivňující kvalitu hodnocení posluchačů a současně také vyvažuje i dopad nejsilnějšího z *krátkodobých vlivů*, tzv. *residuální účinek 1. řádu* [44].

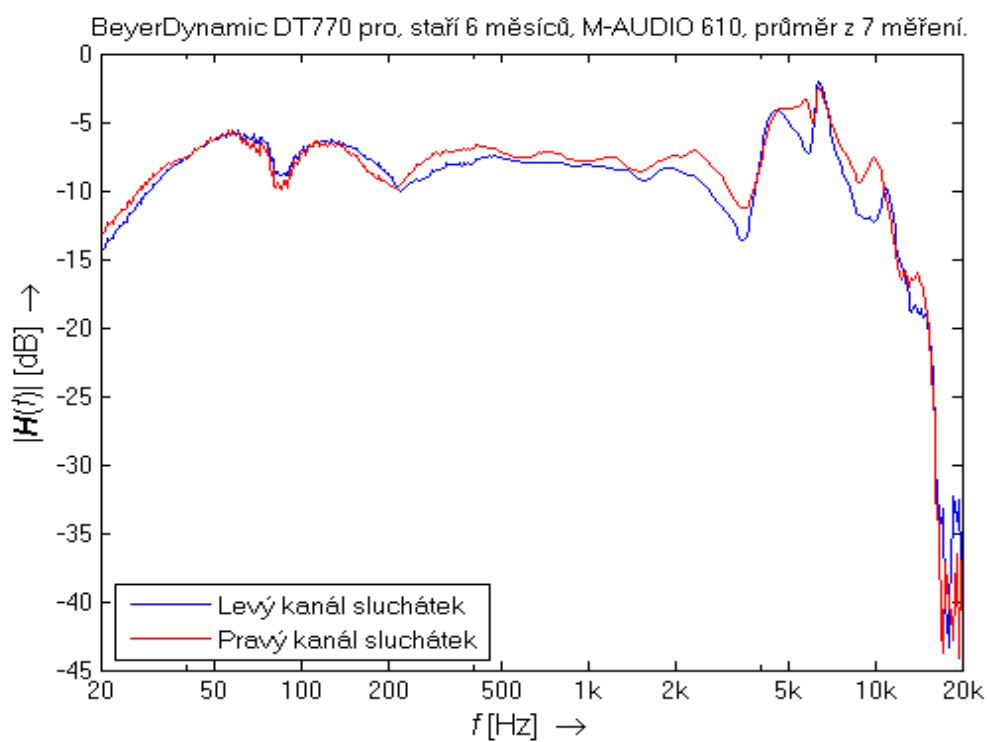


a)



b)

Obr. 5.16 Příprava poslechového pracoviště: a) měření přenosové charakteristiky studiových sluchátek, b) měřicí pracoviště pro provádění poslechových testů.



Obr. 5.17 Změřená přenosová charakteristika studiových sluchátek a zvukové karty.

5.3.4 Shrnutí výsledků realizovaného poslechového experimentu

5.3.4.1 Shrnutí výsledku z 1. části, měření prahu slyšitelnosti šumu pozadí u hudebních a řečových zvukových signálů

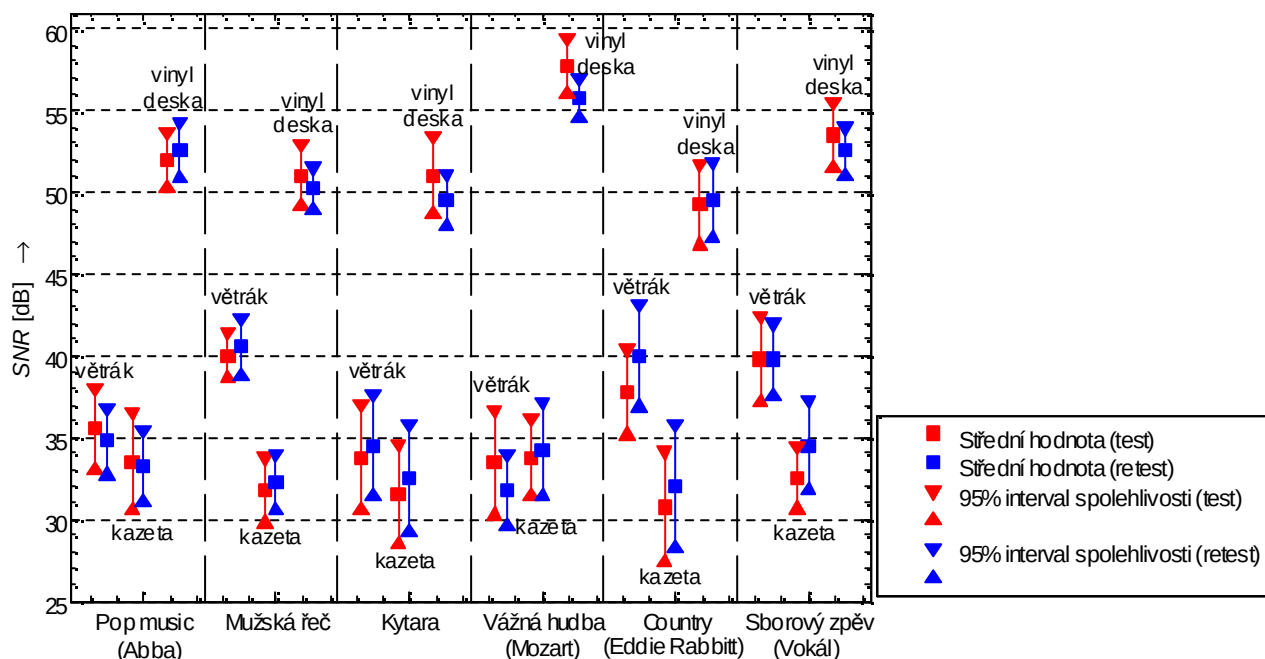
V této podkapitole budou shrnuty výsledky 1. části poslechového experimentu. Detailnější statistické vyhodnocení lze nalézt v příloze C. V této příloze je zjišťována spolehlivost měření, nejprve spolehlivost měření celé skupiny posluchačů jako celku a pak spolehlivost měření jednotlivých posluchačů. A také je zjišťována krátkodobá stabilita měření pro jednotlivé zvukové ukázky. Tato spolehlivost měření je stanovena ve srovnávání rozdílných odpovědí jednotlivců, celé skupiny posluchačů a podobně, z testu a retestu.

Této části experimentu se zúčastnilo celkem 20 posluchačů ve věku 20 až 35 let, z toho byly 4 ženy a 16 mužů. Naměřené subjektivní hodnoty prahu slyšitelnosti šumu pozadí jsou vyneseny v Tab. 5.5. Hodnoty SNR v testu a retestu byly vypočteny jako průměrná hodnota z odpovědí všech posluchačů. Hodnota celkem, je průměrem z hodnot SNR z testu a retestu. Kompletní data všech změřených subjektivních hodnot prahu slyšitelnosti šumu pozadí lze nalézt v příloze C v Tab. C.5 a Tab. C.6.

Tab. 5.5 Změřené subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro 1. část experimentu.

Užitečný signál popředí	Šum pozadí	Test	Retest	Celkem
		SNR_{TEST} [dB]	SNR_{RETEST} [dB]	SNR_{CELKEM} [dB]
Vážná hudba (Mozart)	Šum audio kazety	33,75	34,25	34,00
	Šum větráku	33,50	31,75	32,63
	Šum vinylové desky	57,75	55,75	56,75
Pop music (Abba)	Šum audio kazety	33,50	33,25	33,38
	Šum větráku	35,50	34,75	35,13
	Šum vinylové desky	52,00	52,50	52,25
Country (Eddie Rabbit)	Šum audio kazety	30,75	32,00	31,38
	Šum větráku	37,75	40,00	38,88
	Šum vinylové desky	49,25	49,50	49,38
Hra na kytaru	Šum audio kazety	31,50	32,50	32,00
	Šum větráku	33,75	34,50	34,13
	Šum vinylové desky	51,00	49,50	50,25
Sborový zpěv (vokál)	Šum audio kazety	32,50	34,50	33,50
	Šum větráku	39,75	39,75	39,75
	Šum vinylové desky	53,50	52,50	53,00
Mužská řeč	Šum audio kazety	31,75	32,25	32,00
	Šum větráku	40,00	40,50	40,25
	Šum vinylové desky	51,00	50,25	50,63

Interval spolehlivosti těchto středních hodnot z Tab. 5.5 je přehledně vyneseno v grafu na Obr. 5.18.



Obr. 5.18 Intervalový odhad středních hodnot prahu slyšitelnosti šumu pozadí pro 1. část poslechového experimentu, tj. měření hudebních a řečových signálů

Intervalový odhad střední hodnoty μ při neznámém rozptylu σ^2 je definován

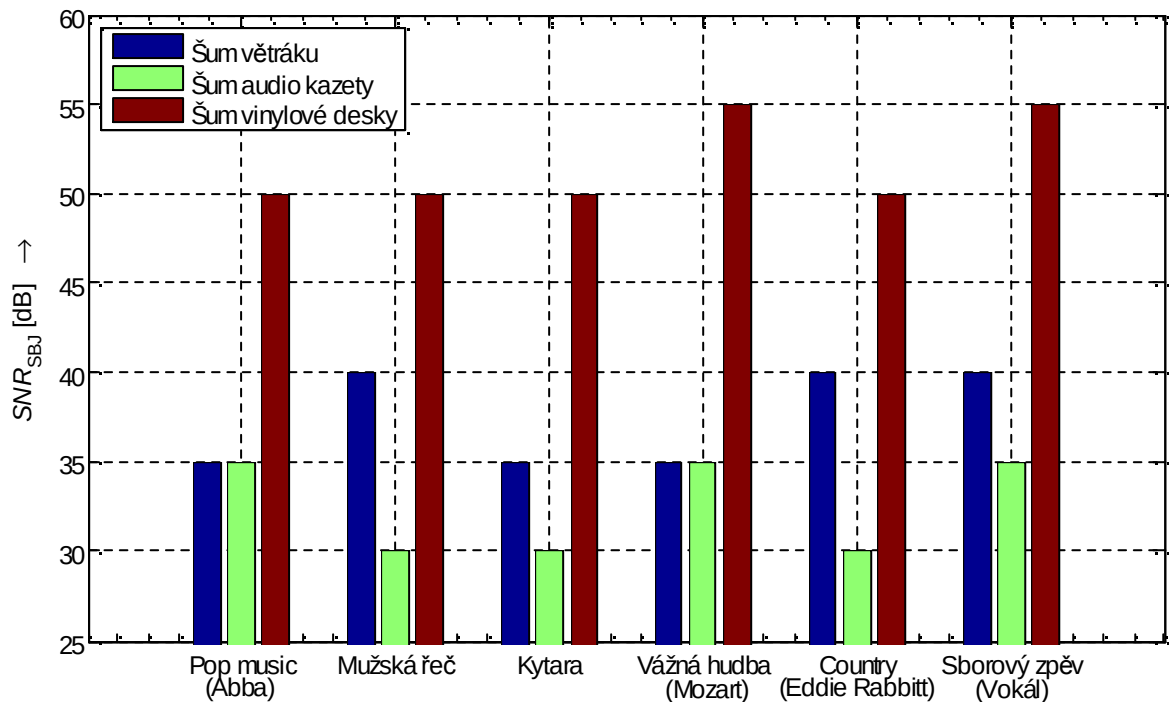
$$\bar{x} - \frac{s}{\sqrt{n-1}} t \left(1 - \frac{\alpha}{2}\right) \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n-1}} t \left(1 - \frac{\alpha}{2}\right) \quad (5.26)$$

kde $t(1 - \alpha/2)$ je kritická hodnota studentova rozdělení, α je hladina spolehlivosti, n je rozsah výběru, $\bar{x}$ je výběrový průměr a s je výběrová směrodatná odchylka.

Výsledné hodnoty získané v 1. části poslechového experimentu jsou zobrazeny v grafu na Obr. 5.19. Tyto výsledné subjektivní hodnoty prahu slyšitelnosti šumu pozadí byly získány jako průměrné hodnoty hodnot SNR v testu a retestu, Tab. 5.5 sloupec celkem, ale navíc byly tyto hodnoty zaokrouhleny na nejbližší hodnotu v bezzbytku dělitelnou 5, tj.

$$SNR_{\text{SBJ}} = 5 \left\lfloor \frac{SNR_{\text{CELKEM}} + 2.5}{5} \right\rfloor, \quad (5.27)$$

kde symbol $\lfloor \cdot \rfloor$ značí zaokrouhlení směrem dolů k zápornému nekonečnu. Toto zaokrouhlení bylo použito kvůli lepší fyzikální interpretaci změřených výsledků. Při poslechovém experimentu mohli posluchači totiž stanovit práh slyšitelnosti šumu pozadí pouze v krocích 5 dB, a proto bude výsledná subjektivní hodnota SNR_{SBJ} prahu slyšitelnosti šumu pozadí také zaokrouhlená na nejbližší hodnotu dělitelnou 5.



Obr. 5.19 Výsledné subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro 1. část poslechového experimentu, tj. měření hudebních a řečových signálů

Z grafu z obrázku Obr. 5.19 vyplývá, že subjektivní veličina SNR_{SBJ} , popisující subjektivní práh slyšitelnosti šumu pozadí, je velmi variabilní veličinou. Pro testované signály se pohybuje v rozmezí hodnot 30 dB až 55 dB. Z tohoto grafu dále vyplývá, že SNR_{SBJ} je závislá na užitečném signálu popředí a šumu pozadí. Nejvyšších hodnot SNR_{SBJ} je dosaženo u šumu vinylové desky, hodnoty až 55 dB. Tedy až při takto vysoké hodnotě SNR se šum pozadí stává neslyšitelným. Pokud je šum pozadí výrazně maskován stává se neslyšitelný pro posluchače v okolí hodnot 30 dB, např. šum audio kazety. Dále je potřeba poznamenat, že tento práh slyšitelnosti je zcela jistě individuální záležitostí. Práh slyšitelnosti je např. závislý na poslechových schopnostech, ale také i na věku, posluchačů. Posluchači snadprůměrným sluchem budou zcela jistě dosahovat vyšších hodnot SNR_{SBJ} . Stejně tak mladší osoby budou dosahovat vyšších hodnot SNR_{SBJ} než osoby starší a podobně [23].

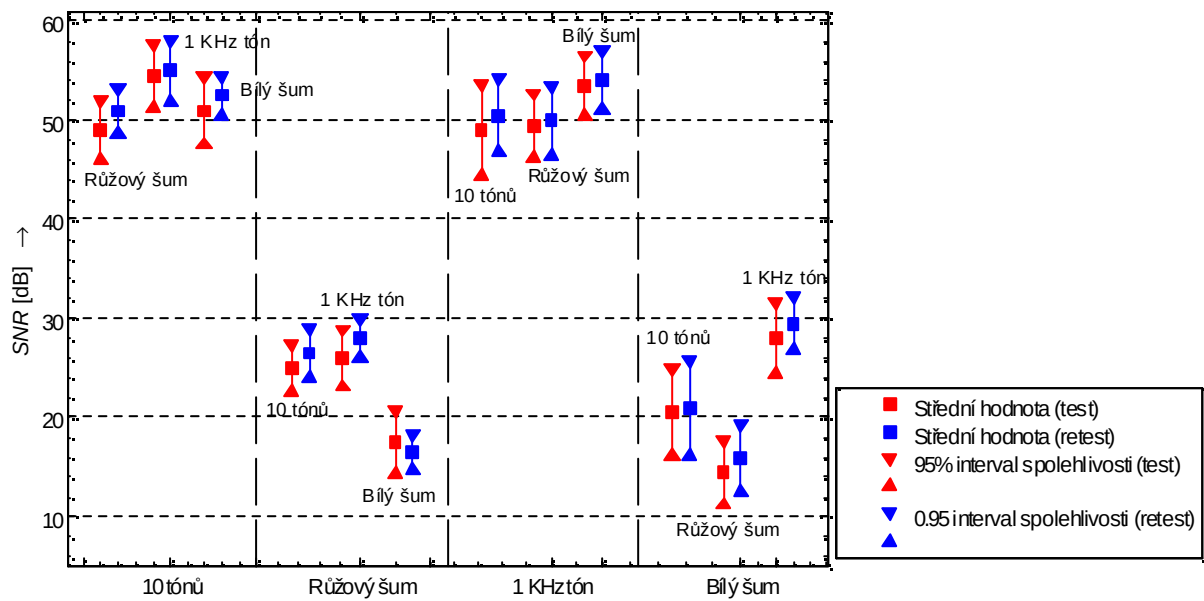
5.3.4.2 Shrnutí výsledku z 2. části, měření prahu slyšitelnosti šumu pozadí u umělých zvukových signálů

V této podkapitole budou shrnuty výsledky 2. části poslechového experimentu. Detailnější statistické vyhodnocení lze opět nalézt v příloze D. Této části experimentu se zúčastnilo celkem 10 posluchačů ve věku 20 až 40 let, z toho byly 2 ženy a 8 mužů. Naměřené subjektivní hodnoty prahu slyšitelnosti šumu pozadí jsou vyneseny v Tab. 5.6. Hodnoty SNR v testu a retestu byly vypočteny stejným způsobem jako v přecházející kapitole. Kompletní data všech změřených subjektivních hodnot prahu slyšitelnosti šumu pozadí lze nalézt v příloze D v Tab. D.4 a Tab. D.5.

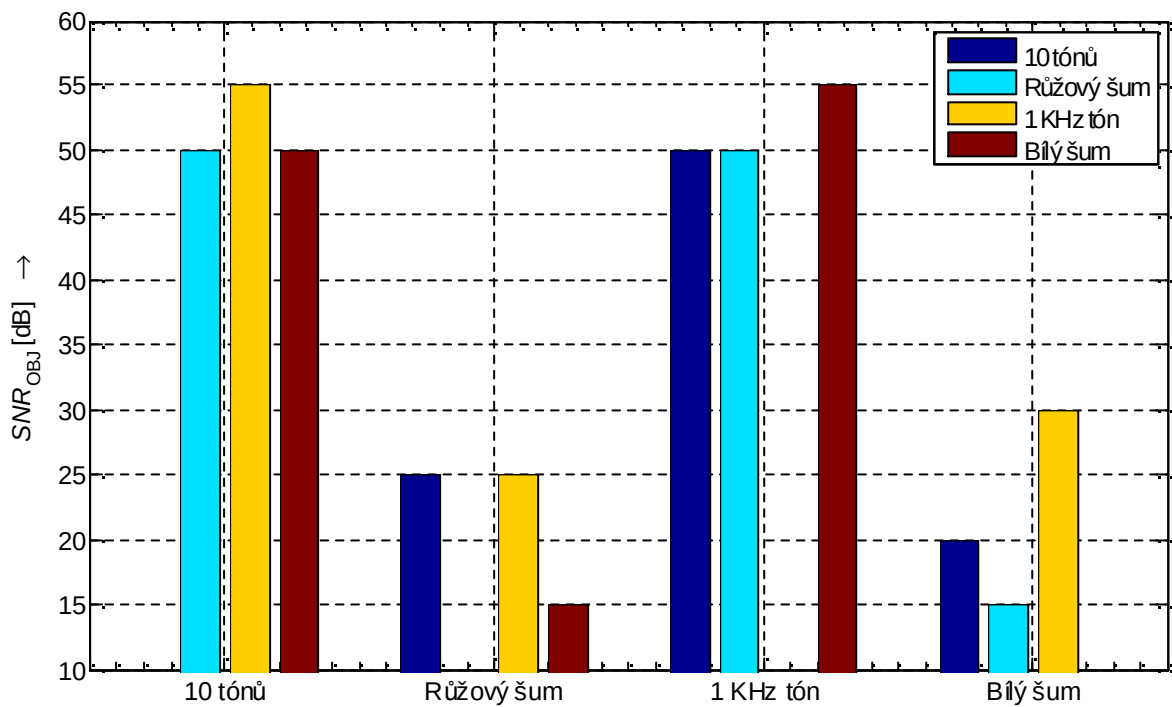
Tab. 5.6 Změřené subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro 2. část experimentu.

Užitečný signál popředí	Šum pozadí	Test	Retest	Celkem
		SNR_{TEST} [dB]	SNR_{RETEST} [dB]	SNR_{CELKEM} [dB]
Tóny v pásmu 10 až 14 kHz (v textu zkráceně označováno jako 10 tónů)	Růžový šum	49,00	51,00	50,00
	1 kHz tón	54,50	55,00	54,75
	Bílý šum	51,75	55,75	51,75
Růžový šum	10 tónů	25,00	26,50	25,75
	1 kHz tón	26,00	28,00	27,00
	Bílý šum	17,50	16,50	17,00
1 kHz tón	10 tónů	49,00	50,50	49,75
	Růžový šum	49,50	50,00	49,75
	Bílý šum	53,50	54,00	53,75
Bílý šum	10 tónů	20,50	21,00	20,75
	Růžový šum	14,50	16,00	15,25
	1 kHz tón	28,00	29,50	28,75

Interval spolehlivosti těchto středních hodnot z Tab. 5.6 je přehledně vynesena v grafu na Obr. 5.20 a výsledné subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro tuto 2. část poslechového experimentu je zobrazeno v grafu na Obr. 5.21.



Obr. 5.20 Intervalový odhad středních hodnot prahu slyšitelnosti šumu pozadí pro 2. část poslechového experimentu, tj. měření umělých zvukových signálů



Obr. 5.21 Výsledné subjektivní hodnoty prahu slyšitelnosti šumu pozadí pro 2. část poslechového experimentu, tj. měření umělých zvukových signálů

5.4 Výsledné zhodnocení navržené objektivní metody

Testování navržené objektivní metody probíhalo ve třech fázích. V první fázi byly získány subjektivní výsledky slyšitelnosti šumu pozadí z poslechových testů, viz předcházející kapitola 5.3. V druhé fázi byly získány objektivní hodnoty prahů slyšitelnosti z testovaných metod, testované metody byly metoda PEAQ [57] a mnou navržená metoda z kapitoly 5.2. V poslední fázi byly tyto tři sady výsledků vzájemně porovnány. Toto porovnání je právě náplní této kapitoly a probíhalo na dvou typech zvukových signálů. Prvním typem byly hudební a řečové signály a druhým typem byly umělé zvukové signály.

5.4.1 Porovnání výsledků pro hudební a řečové signály

V této kapitole budou porovnány dvě objektivní metody pro predikci prahu slyšitelnosti šumu pozadí. Využity přitom budou výsledky z poslechového experimentu z kapitoly 5.3. Porovnávána bude metoda PEAQ popsána v kapitole 4 a navržená metoda z kapitoly 5. Testovacím kritériem pro toto porovnání bude střední kvadratická chyba MSE (mean squared error) definována

$$MSE(SNR_{OBJ}) = E[(SNR_{SBJ} - SNR_{OBJ})^2], \quad (5.28)$$

kde SNR_{SBJ} jsou subjektivní prahy slyšitelnosti šumu pozadí získané z provedeného experimentu a SNR_{OBJ} jsou objektivní prahy slyšitelnosti získané metodami PEAQ a $SpecNMR_{MAX}$. V Tab. 5.7 jsou zobrazeny hodnoty SNR_{SBJ} a v Tab. 5.8 jsou zobrazeny hodnoty SNR_{OBJ} pro obě objektivní metody. Zeleným a červeným písmem jsou v těchto tabulkách zvýrazněny rozdíly od změřených subjektivních hodnot. V dolní části těchto tabulek jsou vypsány naměřené hodnoty MSE .

Tab. 5.7 Tabulka s naměřenými subjektivní hodnotami SNR_{OBJ} prahu slyšitelnosti šumu pozadí pro 1. část experimentu.

	Poslech		
	Vinyl	Kazeta	Větrák
	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]
Kytara	50	30	35
Sborový zpěv (Vokál)	55	35	40
Mužská řeč	50	30	40
Vážná hudba (Mozart)	55	35	35
Pop music (Abba)	50	35	35
Country (Mozart)	50	30	40

Tabulka s naměřenými subjektivní hodnotami prahu slyšitelnosti šumu pozadí pro 1. část experimentu. V levém sloupci jsou názvy užitečných signálů popředí a na druhém řádku jsou názvy šumu pozadí.

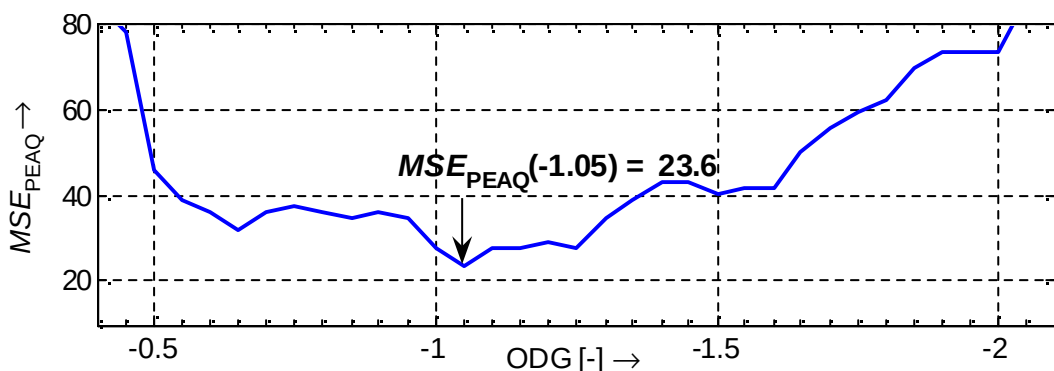
Tab. 5.8 Tabulka s naměřenými objektivními hodnotami SNR_{OBJ} prahu slyšitelnosti šumu pozadí pro 1. část experimentu pro obě porovnávané metody.

PEAQ = -1.05 ODG				<i>SpecNMR</i> _{MAX} = 8 dB (24 %)			
	Vinyl	Kazeta	Větrák		Vinyl	Kazeta	Větrák
	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]		SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]
Kytara	50	25 (-5dB)	25 (-10dB)	Kytara	50	35 (+5dB)	35
Sborový zpěv (Vokál)	50 (-5dB)	30 (-5dB)	35 (-5dB)	Sborový zpěv (Vokál)	50 (-5dB)	35	35 (-5dB)
Mužská řeč	55 (+5dB)	35 (+5dB)	45 (+5dB)	Mužská řeč	50	35 (+5dB)	40
Vážná hudba (Mozart)	55	35	30 (-5dB)	Vážná hudba (Mozart)	50 (-5dB)	35	35
Pop music (Abba)	45 (-5dB)	40 (+5dB)	40 (+5dB)	Pop music (Abba)	45 (-5dB)	40 (+5dB)	40 (+5dB)
Country (Mozart)	50	25 (-5dB)	35 (-5dB)	Country (Mozart)	45 (-5dB)	30	35 (-5dB)
$MSE_{PEAQ} = 23.6 [-]$				$MSE_{NMR} = 13.9 [-]$			

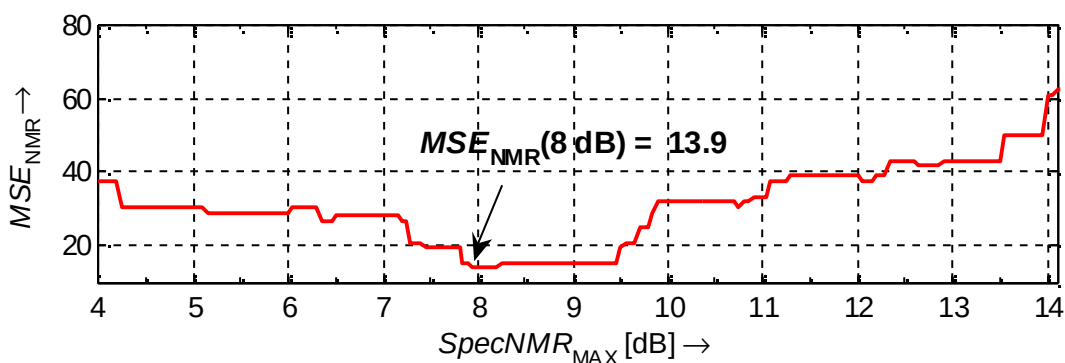
V těchto tabulkách jsou objektivní hodnoty prahu slyšitelnosti šumu pozadí SNR_{OBJ} pro metodu PEAQ (levá tabulka) a metodu *SpecNMR*_{MAX} (tabulka vpravo). V levém sloupci obou tabulek jsou názvy užitečných signálů popředí a na druhém řádku jsou názvy šumu pozadí. Zeleným a červeným písmem jsou v těchto tabulkách zvýrazněny rozdíly od změřených subjektivních hodnot. V dolní části těchto tabulek jsou vypsány naměřené hodnoty MSE .

Při vzájemném srovnání obou objektivních metod lze říci, že navržená objektivní metoda *SpecNMR*_{MAX} lépe predikuje subjektivní hodnoty SNR_{SBJ} , protože dosahuje totiž nižších hodnot MSE než metoda PEAQ. Už při vizuální kontrole velikosti shody mezi SNR_{SBJ} a SNR_{OBJ} obou objektivních metod je vidět, že *SpecNMR*_{MAX} správně predikuje správnou SNR_{SBJ} v 8 případech z 18 případů a ve zbytku se liší o 5 dB. Kdežto PEAQ predikuje správnou hodnotu SNR_{SBJ} pouze ve 4 případech, jednou se liší o 10 dB a ve zbytku případů o 5 dB.

Jako hraniční hodnoty pro určení objektivních hodnot prahu slyšitelnosti byly zvoleny hodnoty $PEAQ = -1.05$ ODG a $SpecNMR_{MAX} = 8$ dB. Tedy například pro metodu PEAQ byly všechny zvukové ukázky, které měli hodnocení v rozsahu 0 až -1.05 ODG, označeny jako zvukové ukázky s neslyšitelným šumem pozadí. Pro *SpecNMR*_{MAX} byla tato hraniční hodnota nastavena na 8 dB a $ReINMR(b)$ pro tuto komponentu musela být více než 24 %. Tyto hraniční hodnoty byly stanoveny jako $\min\{MSE(SNR_{OBJ})\}$, $\forall SNR_{OBJ}$. Tedy ze všech hodnot MSE bylo nalezeno to s nejmenší hodnotou. A právě tato hodnota byla stanovena jako hraniční hodnota pro určení objektivních hodnot prahu slyšitelnosti. Tento experimentální postup je lépe zobrazen na grafech na Obr. 5.22 a Obr. 5.23. Na těchto grafech jsou vykresleny průběhy MSE v závislosti na měnicích se velikostech těchto prahů. Šípkami jsou označeny zvolené hraniční hodnoty. Z obrázku lze vidět, že se jedná o globální minima. Hodnoty MSE mimo zobrazenou oblast z těchto grafů nabývají velmi vysokých hodnot.



Obr. 5.22 Volba hraniční hodnoty $\min\{MSE(SNR_{Obj})\}$ z 1. části poslechového experimentu, metoda PEAQ



Obr. 5.23 Volba hraniční hodnoty $\min\{MSE(SNR_{Obj})\}$ z 1. části poslechového experimentu, metoda $SpecNMR_{MAX}$

V případě metody PEAQ lze získanou hraniční hodnotu -1.05 ODG interpretovat pomocí Tab. 5.9. V této tabulce je vynesena pětistupňová hodnotící škála ODG a společně s touto škálou, je zde uveden krátký slovní popis pro každý hodnotící stupeň této škály. Z této tabulky pak vyplývá, že hodnocení velikosti degradace vůči originálnímu signálu v rozsahu od 0 ODG do -1 ODG je pro posluchače nevnímatelné, tedy neslyšitelné. Stejného závěru bylo dosaženo experimentálním postupem z grafu z Obr. 5.22 když jako hraniční hodnota byla stanovena hodnota -1.05 ODG vztahem (5.28).

Tab. 5.9. Objektivní hodnotící škála ODG [57].

ODG - hodnotící škála (PEAQ)	Slovní popis Hodnotí se velikost degradace nebo zhoršení (angl. impairment) signálu vůči originálu.
0	Nevnímatelné (angl. imperceptible)
-1	Vnímatelné, ale ne nepříjemné (angl. perceptible but not annoying)
-2	Lehce nepříjemné (angl. slightly annoying)
-3	Nepříjemné (angl. annoying)
-4	Velmi nepříjemné (angl. very annoying)

5.4.2 Porovnání výsledků pro umělé zvukové signály

Stejným způsobem jako v minulé kapitole, budou v této kapitole porovnávány objektivní metody pro predikci prahu slyšitelnosti šumu pozadí. Opět budou využity výsledky z poslechového experimentu z kapitoly 5.3 a to z 2. části tohoto experimentu, který byl zaměřen na umělé zvukové signály. Porovnávány budou metody PEAQ (kapitola 4) a nově navržená objektivní metoda z kapitoly 5. Testovacím kritériem pro toto porovnání bude střední kvadratická chyba MSE (5.28). Subjektivní prahy slyšitelnosti šumu pozadí SNR_{OBJ} získané z této 2. části provedeného experimentu jsou zobrazeny v Tab. 5.10.

Tab. 5.10 Tabulka s naměřenými subjektivní hodnotami SNR_{OBJ} prahu slyšitelnosti šumu pozadí pro 2. část experimentu.

Poslech				
	10 tónů	Růžový šum	1 kHz tón	Bílý šum
	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	
10 tónů	x	50	55	50
Růžový šum	25	x	25	15
1 kHz tón	50	50	x	55
Bílý šum	20	15	30	x

V Tab. 5.11 jsou zobrazeny hodnoty SNR_{OBJ} pro obě objektivní metody. Zeleným a červeným písmem jsou v těchto tabulkách zvýrazněny rozdíly od změřených subjektivních hodnot. V dolní části těchto tabulek jsou vypsány naměřené hodnoty MSE .

Tab. 5.11 Tabulka s naměřenými objektivními hodnotami SNR_{OBJ} prahu slyšitelnosti šumu pozadí pro 1. část experimentu pro obě porovnávané metody.

PEAQ = -2.9 ODG				
	10 tónů	Růžový šum	1 kHz tón	Bílý šum
	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]
10 tónů	x	70 (+20dB)	40 (-15dB)	65 (+15dB)
Růžový šum	10 (+15dB)	x	10 (+15dB)	15
1 kHz tón	60 (+10dB)	60 (+10dB)	x	60 (+5dB)
Bílý šum	10 (-10dB)	15	10 (-10dB)	x
$MSE_{PEAQ} = 126.6 [-]$				

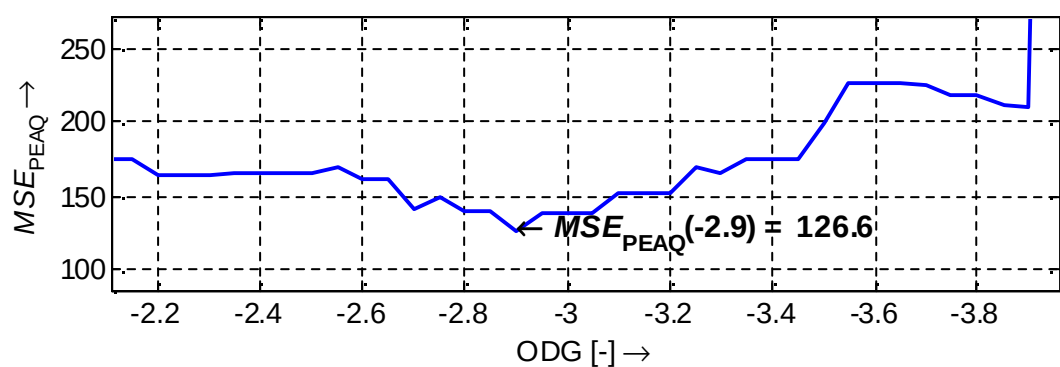
SpecNMR _{MAX} = 7.9 dB (100 %)				
	10 tónů	Růžový šum	1 kHz tón	Bílý šum
	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]	SNR_{OBJ} [dB]
10 tónů	x	55 (+5dB)	60 (+5dB)	55 (+5dB)
Růžový šum	20 (-5dB)	x	20 (-5dB)	10 (-5dB)
1 kHz tón	55 (-5dB)	50	x	55
Bílý šum	15 (-5dB)	30 (+15dB)	25 (-5dB)	x
$MSE_{NMR} = 37.5 [-]$				

Lze vidět, že metoda PEAQ zcela selhává pro odhady subjektivního vjemu slyšitelnosti šumu pozadí pro případ umělých zvukových signálů. Je to pravděpodobně dáno způsobem realizace této metody, kde hlavním výpočetním algoritmem této metody je neuronová síť, která je právě adaptována na hudební a řečové typy signálů a ne na umělé zvukové signály [57]. Metoda PEAQ pouze ve dvou případech správně odhadla subjektivní práh slyšitelného šumu pozadí SNR_{OBJ} . A v 9 případech ze 12 se lišila minimálně o 10 dB, někdy ale dokonce i o 20 dB. Největším problémem je ovšem nesprávné určení hraniční hodnoty pro určení objektivních hodnot prahu slyšitelnosti. Tato hraniční hodnota je -2.9 SDG. Takováto hodnota zkreslení nebo degradace podle ODG škály odpovídá vnímatelnému zkreslení, které je pro posluchače nepříjemné, viz Tab. 5.9. Realizovaný poslechový experiment z kapitoly 5.3 ale dokázal, že toto zkreslení, způsobené aditivním šumem je nevnímavé, tj. neslyšitelné. Z tohoto důvodu je tato hraniční hodnota nesprávná. Správně by tedy měla být v rozsahu 0 ODG až -1 ODG, podobně jako tomu je u hudebních a řečových signálů.

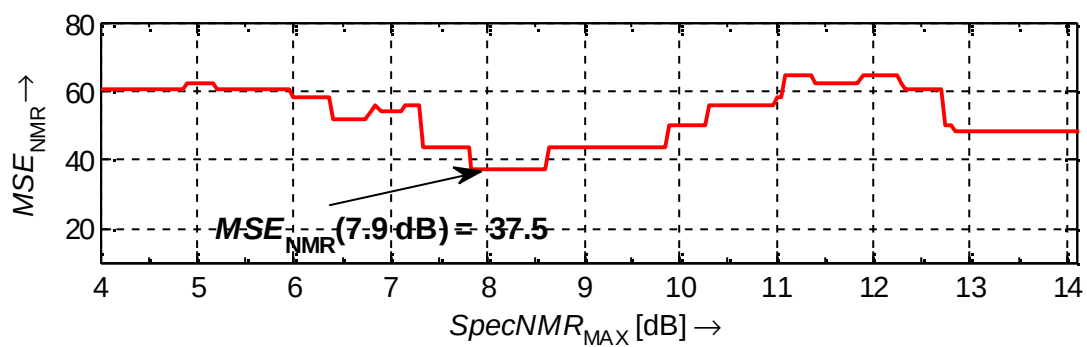
Naproti tomu odhad subjektivního vjemu slyšitelnosti šumu pozadí podle nově navržené metody z kapitoly 5 dosahuje relativně přesných výsledků. V 11 z celkových 12 případů, byl proveden odhad s chybou maximálně 5 dB. V jednom případě se navržená metoda dopustila velmi špatného odhadu a to s chybou 15 dB. To bylo velmi pravděpodobně způsobeno podobnými typy zvuků bílého a růžového šumu. Posluchači je pak nebyli schopni vzájemně od sebe rozlišit. Posluchači nejspíše nesprávně považovali šum pozadí za užitečný signál popředí.

Hodnota střední kvadratické odchylky MSE je ale výrazně lepší pro navrženou metodu z kapitoly 5 než pro případ metody PEAQ, $MSE_{NMR} = 37.5$ a $MSE_{PEAQ} = 126.6$.

Hraniční hodnota pro určení objektivních hodnot prahu slyšitelnosti je zde téměř stejná jako u reálných typů signálů. Nyní má $SpecNMR_{MAX}$ hodnotu 7.9 dB a u reálných typů signálů byla $SpecNMR_{MAX}$ hodnota rovna 8 dB. Experimentální postup stanovení obou těchto hraničních hodnot je zobrazen na Obr. 5.24 a Obr. 5.25.



Obr. 5.24 Volba hraniční hodnoty $\min\{MSE(SNR_{OBJ})\}$ z 2. části poslechového experimentu, metoda PEAQ



Obr. 5.25 Volba hraniční hodnoty $\min\{MSE(SNR_{OBJ})\}$ z 2. části poslechového experimentu, metoda SpecNMR_{MAX}

6 Zavedení parametrického Wienerova filtru

Tato kapitola pojedná o standardních metodách na potlačení aditivního šumu. Konkrétně se jedná o techniky spektrálního odečítání a Wienerův filtr. Dále je v této kapitole zaveden tzv. parametrický Wienerův filtr. Tento filtr byl zaveden z důvodu tohoto, že integruje zmíněné standardní metody do jediného (parametrického) restauračního filtru a lze ho použít i na nestacionární třídu signálů, jakou jsou např. hudební a řečové signály. Hlavním cílem této kapitoly je vytvoření teoretického pozadí pro následné zavedení prahovacích pravidel, které budou použita ve vlastním řešení systému na potlačení šumu na pozadí.

6.1 Wienerův filtr

Uvažujme stochastické diskrétní signály $x(n)$ a $m(n)$, kde $x(n)$ reprezentuje užitečný signál a $m(n)$ je rušivý šum. Vstupní, tj. šumem zarušený, signál $y(n)$ je definován

$$y(n) = x(n) + m(n). \quad (6.1)$$

Ve spektrální oblasti pak lze psát [34][77]

$$Y(e^{j\Omega}) = X(e^{j\Omega}) + M(e^{j\Omega}), \quad (6.2)$$

kde $Y(e^{j\Omega})$, $X(e^{j\Omega})$, $M(e^{j\Omega})$ jsou spektra signálů $y(n)$, $x(n)$, $m(n)$ definované Fourierovou transformací $\mathcal{F}\{ \}$

$$Y(e^{j\Omega}) = \mathcal{F}\{y(n)\} = \sum_{n=-\infty}^{+\infty} y(n)e^{-j\Omega n}, \quad (6.3)$$

kde $\Omega = 2\pi f/f_s$ je normovaný úhlový kmitočet, f_s je vzorkovací kmitočet. Funkce $Y(e^{j\Omega})$ je periodická s periodou 2π . Výkonové spektrum signálu $y(n)$ bude zavedeno jako [19][34]

$$S_{yy}(\Omega) = E \left\{ |Y(e^{j\Omega})|^2 \right\}. \quad (6.4)$$

Pokud $x(n)$ a $m(n)$ jsou lineárně nezávislé procesy, tak lze psát

$$S_{yy}(\Omega) = S_{xx}(\Omega) + S_{mm}(\Omega). \quad (6.5)$$

Výkonové spektrum $S_{xx}(\Omega)$ lze obnovit z výkonového spektra $S_{yy}(\Omega)$ modifikací předchozího vztahu jako [15][19] [20][25]

$$S_{xx}(\Omega) = S_{yy}(\Omega) - S_{mm}(\Omega). \quad (6.6)$$

Tento postup ovšem obnoví pouze výkonové spektrum $S_{xx}(\Omega)$. Jedním z přístupů, jak odhadnou $\hat{x}(n)$ z $y(n)$ je Wienerova filtrace [34][79]. Kritérium optimality tohoto odhadu je minimalizace střední kvadratické odchylky mezi odhadem $\hat{x}(n)$ a jeho originálem $x(n)$, tj. minimalizace výrazu $E\{[\hat{x}(n) - x(n)]^2\}$ pro všechny realizace náhodných procesů $x(n)$ a $m(n)$. Odhad $\hat{X}(e^{j\Omega})$ z $Y(e^{j\Omega})$ je proveden [79]

$$\hat{X}(e^{j\Omega}) = H_W(\Omega)Y(e^{j\Omega}), \quad (6.7)$$

kde $H_W(\Omega)$ bude dále v textu nazývána jakou zisková funkce (Gain Function) [19][77] Wienerova filtru a tato zisková funkce je definována jako

$$H_W(\Omega) = \frac{S_{xx}(\Omega)}{S_{xx}(\Omega) + S_{mm}(\Omega)}, \quad (6.8)$$

nebo častěji používané vyjádření

$$H_W(\Omega) = \frac{S_{yy}(\Omega) - S_{mm}(\Omega)}{S_{yy}(\Omega)}. \quad (6.9)$$

Nyní lze s využitím vztahů (6.7) a (6.9) psát

$$\hat{X}(e^{j\Omega}) = \left[\frac{S_{yy}(\Omega) - S_{mm}(\Omega)}{S_{yy}(\Omega)} \right] Y(e^{j\Omega}), \quad (6.10)$$

Tato forma rovnice znázorňuje způsob obnovy zarušeného signálu formou spektrálního odečítání, viz $S_{yy}(\Omega) - S_{mm}(\Omega)$ v této rovnici. Výkonové spektrální hustoty $S_{yy}(\Omega)$ a $S_{mm}(\Omega)$ jsou podle (6.4) rovny $E\{|Y(e^{j\Omega})|^2\}$ a $E\{|M(e^{j\Omega})|^2\}$ a proto lze pro $H_W(\Omega)$ psát

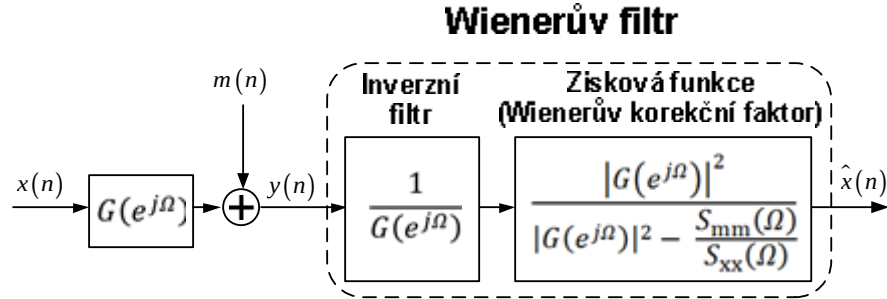
$$H_W(\Omega) = \frac{E\{|Y(e^{j\Omega})|^2\} - E\{|M(e^{j\Omega})|^2\}}{E\{|Y(e^{j\Omega})|^2\}}. \quad (6.11)$$

Po nahrazení odhadů $|Y(e^{j\Omega})|^2$ a $|M(e^{j\Omega})|^2$ za $E\{|Y(e^{j\Omega})|^2\}$ a $E\{|M(e^{j\Omega})|^2\}$ lze rovnici (6.11) přepsat jako

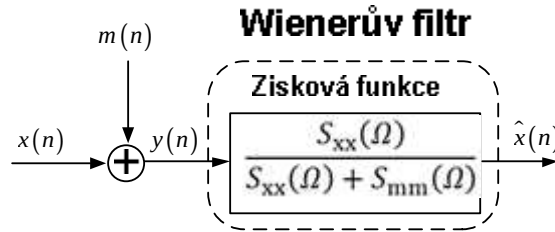
$$H_W(\Omega) \cong \frac{|Y(e^{j\Omega})|^2 - |M(e^{j\Omega})|^2}{|Y(e^{j\Omega})|^2}. \quad (6.12)$$

Tedy namísto souborových středních hodnot byly do předchozího vztahu dosazeny kvadráty amplitudových spekter, které byly získané pouze z jediné (dostupné) realizace $y(n)$ a

$m(n)$. Na Obr. 6.1 je zobrazeno blokové schéma odhadu $\hat{x}(n)$ Wienerovým filtrem, navíc je zde zobrazen zkreslující systémem $G(e^{j\Omega})$. Po dosažení $G(e^{j\Omega}) = 1$ je získán odhad podle (6.10), viz Obr. 6.2.



Obr. 6.1 Blokové schéma odhadu $\hat{x}(n)$ Wienerovým filtrem se zkreslujícím systémem $G(e^{j\Omega})$.



Obr. 6.2 Blokové schéma odhadu $\hat{x}(n)$ Wienerovým filtrem bez zkreslujícího systému.

6.2 Krátkodobý Wienerův filtr

Krátkodobý Wienerův filtr (Short-Time Wiener Filter [19]) je rozšíření Wienerova filtru i na nestacionární třídu signálů. Wienerův filtr definovaný v minulé kapitole lze totiž aplikovat pouze na stacionární třídu signálů. Hudební signály ovšem nelze považovat za stacionární [13][77]. V takovém případě je potřeba rozdělit dlouhotrvající signál na kratší úseky. A na těchto úsecích následně provádět odhady výkonových spektrálních hustot. K tomu účelu jsou využity nástroje analýzy a syntézy krátkodobé Fourierovy transformace STFT (Short-Time Fourier Transform).

Dlouhotrvající signál $y(n)$ bude rozdělen na kratší úseky délky N vzorků

$$y_i(n) = \begin{cases} w(n)y(n + N(i-1) - N(v-1)) & \text{pro } n = 0, 1, \dots, N-1 \\ 0 & \text{pro } n \neq 0, 1, \dots, N-1 \end{cases} \quad (6.13)$$

kde symbol i značí pořadové číslo segmentu a platí $i = 1, \dots, I$ a I je celkový počet segmentů. Symbol v je zde použit pro vyjádření velikosti překryvu sousedních segmentů, $v = 0,5$ značí 50% překryv a $v = 0,75$ značí 75% překryv atd., $w(n)$ je analyzující okno.

Takto segmentovaný signál $y_i(n)$ bude reprezentován STFT [77]

$$Y_i(k) = \sum_{n=0}^{N-1} y_i(n) e^{-j2\pi kn/N} \quad (6.14)$$

kde k je pořadové číslo k -tého spektrálního koeficientu STFT, $k = 0, \dots, K - 1$ a K je celkový počet těchto koeficientů. Protože je STFT typ zobrazení $\mathbb{C}^N \rightarrow \mathbb{C}^N$, kde $\mathbb{C}$ značí množinu komplexních čísel, tak platí, že $K = N$.

Vztah (6.10) pro stanovení odhadu $\hat{X}_{W_i}(k)$ lze nyní s využitím (6.13) a (6.14) přepsat jako

$$\hat{X}_{W_i}(k) = \left[\frac{|Y_i(k)|^2 - |M_i(k)|^2}{|Y_i(k)|^2} \right] Y_i(k) \quad (6.15)$$

Část v hranaté závorce výrazu z předchozího vztahu je právě nazývána krátkodobý Wienerův filtr. Zisková tohoto filtru funkce je definována následovně [19]

$$H_{W_i}(k) = \frac{|Y_i(k)|^2 - |M_i(k)|^2}{|Y_i(k)|^2} \quad (6.16)$$

6.3 Techniky spektrálního odečítání

Tato kapitola bude pojednávat o technikách amplitudového a spektrálního odečítání [15][19][34]. Tyto techniky odhadují, podobně Wienerův filtr, originálního signálu ze vstupního signálu zarušeného aditivním šumem, je uvažován model zkreslení (6.1). Typy odhadu jsou zde ale jiné. Výkonové spektrální odečítání se snaží odhadnout výkonové spektrum originálního signálu a amplitudové spektrální odečítání se snaží odhadnout amplitudové spektrum [19][34].

Spektrum vstupního signálu $X(e^{j\Omega})$ lze rozdělit na jeho modulové spektrum $|X(e^{j\Omega})|$ a fázové spektrum $\phi_x(\Omega)$ podle vztahu

$$X(e^{j\Omega}) = |X(e^{j\Omega})| e^{j\phi_x(\Omega)}. \quad (6.17)$$

Za předpokladu lineární nezávislosti signálů $x(n)$ a $m(n)$ lze psát

$$|\hat{X}_{PS}(e^{j\Omega})|^2 = |Y(e^{j\Omega})|^2 - |M(e^{j\Omega})|^2. \quad (6.18)$$

A po dosazení (6.18) do (6.17) dostáváme

$$\hat{X}_{PS}(e^{j\Omega}) = \sqrt{|Y(e^{j\Omega})|^2 - |M(e^{j\Omega})|^2} e^{j\phi_y(\Omega)}, \quad (6.19)$$

kde symbol $\hat{X}_{PS}(e^{j\Omega})$ je použit pro odhad $X(e^{j\Omega})$ technikou výkonového spektrálního odečítání, angl. *Power Spectral Subtraction*, proto je v symbolu $\hat{X}_{PS}(e^{j\Omega})$ použit index PS. Pro případ

krátkodobého výkonového spektrálního odečítání lze vztah (6.19) vyjádřit pomocí (6.13) a (6.14) jako

$$\hat{X}_{PS_i}(k) = \sqrt{|Y_i(k)|^2 - |M_i(k)|^2} e^{j\phi_{y_i}(k)} \quad (6.20)$$

Amplitudové odečítání je velmi podobné výkonovému spektrálnímu odečítání. Namísto vztahu (6.18) pro odečítání výkonových spekter je použit vztah pro odečítání modulových (amplitudových) spekter

$$|\hat{X}_{MS_i}(k)| = |Y_i(k)| - |M_i(k)|. \quad (6.21)$$

Po dosazení do (6.17) je získán výsledný vztah pro odhad $X_i(k)$ technikou amplitudového spektrálního odečítání, angl. *Magnitude Spectral Subtraction*, proto je také použit index MS v $\hat{X}_{MS_i}(k)$

$$\hat{X}_{MS_i}(k) = [|Y_i(k)| - |M_i(k)|] e^{j\phi_{y_i}(k)}. \quad (6.22)$$

6.4 Parametrický Wienerův filtr

Techniky spektrálního odečítání a Wienerova filtru jsou si velmi blízké metody. Několik autorů [19][36][71] využilo tento fakt a integrovalo tyto techniky pro potlačení šumu (resp. odhad originální složky) do jediného obecného filtru, nazvaného parametrický Wienerův filtr (Parametric Wiener Filter [19]).

Jak již bylo ukázáno v (6.15), odhad $\hat{X}_{W_i}(k)$ z $Y_i(k)$ lze provést podle vztahu

$$\hat{X}_{W_i}(k) = H_{W_i}(k) Y_i(k), \quad (6.23)$$

kde $H_{W_i}(k)$ je zisková funkce Wienerova krátkodobého filtru

$$H_{W_i}(k) = \frac{|Y_i(k)|^2 - |M_i(k)|^2}{|Y_i(k)|^2}. \quad (6.24)$$

Podobným způsobem lze zavést i ziskovou funkci pro amplitudové spektrální odečítání $H_{MS_i}(k)$ (6.26) a podle ní provést odhad $\hat{X}_{MS_i}(k)$ (6.25).

$$\hat{X}_{MS_i}(k) = H_{MS_i}(k) Y_i(k) \quad (6.25)$$

$$H_{MS_i}(k) = \frac{|Y_i(k)| - |M_i(k)|}{|Y_i(k)|} \quad (6.26)$$

Zisková funkce pro výkonové spektrální odečítání $H_{PS_i}(k)$ je uvedena v (6.27) a odhad $\hat{X}_{MS_i}(k)$ právě s využitím je uveden ve vztahu (6.28).

$$\hat{X}_{PS_i}(k) = H_{PS_i}(k) Y_i(k) \quad (6.27)$$

$$H_{PS_i}(k) = \left[\frac{|Y_i(k)|^2 - |M_i(k)|^2}{|Y_i(k)|^2} \right]^{1/2} \quad (6.28)$$

Záměrně byly voleny podobné tvary těchto rovnic, aby byla patrná velká podobnost mezi technikami spektrálního odečítání a Wienerovým filtrem. Rovnice pro stanovení odhadu (6.23)(6.25)(6.27) jsou téměř shodné, mění se zde pouze typ ziskové funkce. Nahrazením této ziskové funkce lze tedy měnit typ odhadu. Dalším zkoumáním vztahů pro jednotlivé ziskové funkce $H_{W_i}(k)$, $H_{MS_i}(k)$ a $H_{PS_i}(k)$ byla zavedena obecná zisková funkce $H_{G_i}(k)$ (6.30), kde index G zde značí obecný angl. *general*. Odhad pomocí této obecné ziskové funkce je proveden obdobně jako v minulých příkladech

$$\hat{X}_{G_i}(k) = H_{G_i}(k) Y_i(k), \quad (6.29)$$

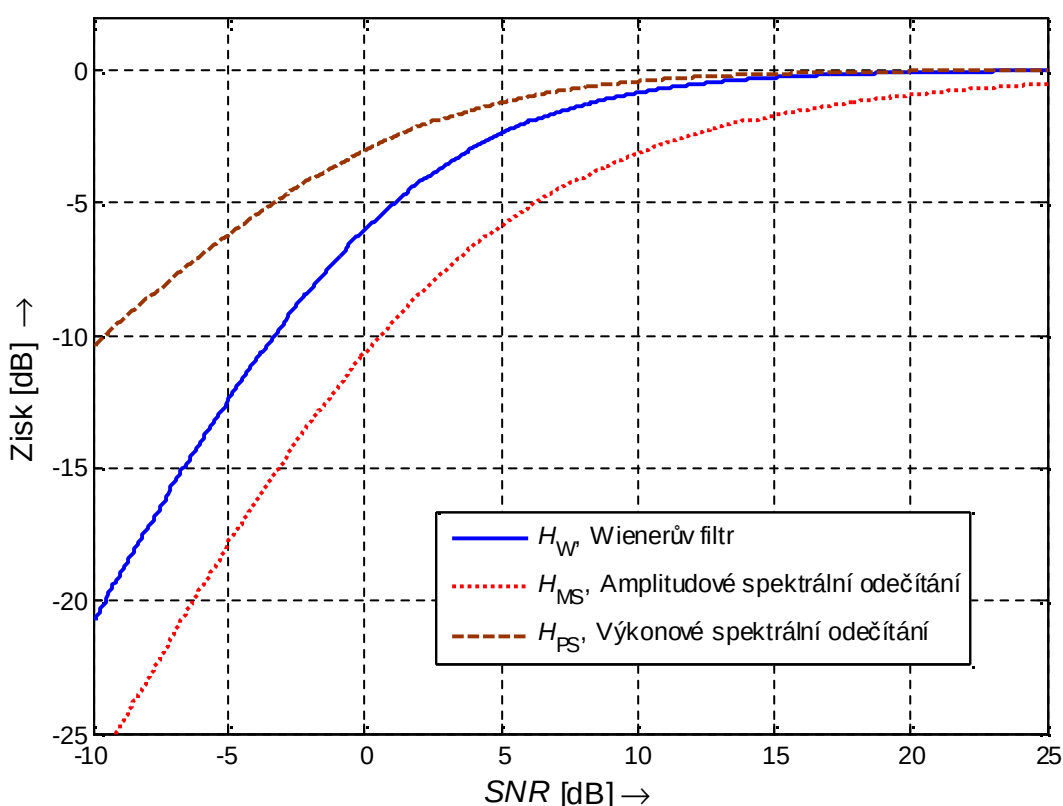
$$H_{G_i}(k) = \left[1 - \left(\frac{|M_i(k)|}{|Y_i(k)|} \right)^\gamma \right]^\beta. \quad (6.30)$$

Parametry (γ, β) jsou zde využity k vyjádření typu odhadu. Při nastavení $(\gamma, \beta) = (2, 1)$ vztah (6.30) přechází v ziskovou funkci $H_{W_i}(k)$ (6.24) Wienerova krátkodobého filtru, $(\gamma, \beta) = (1, 1)$ odpovídá amplitudovému spektrálnímu odečítání a $(\gamma, \beta) = (2, 1/2)$ výkonovému spektrálnímu odečítání [19][20][26].

Bohužel obecné parametry (γ, β) nesledují žádné kritérium optimality jako například minimalizace MSE u Wienerova filtru a podobně. Tato skutečnost ovšem nesnižuje užitečnost obecné formulace parametrického Wienerova filtru podle (6.30) pro výběr metody pro potlačení šumu. V některých případech kritérium optimality nelze jednoznačně určit, v takových případech bývá vyšší stupeň volnosti při návrhu metod a systému pro potlačení šumu výhodou. Takovým případem může být hodnocení kvality hudebních a řečových signálů, kde subjektivní vjem kvality je tvořen několika subjektivními psychoakustickými veličinami [44][23]. Definice těchto veličin a definice jejich vzájemných vztahů je velmi problematická, proto je problematické i stanovení kritéria této optimality. Z toho důvodů se například v moderních systémech pro potlačení šumu ve zvukových signálech od tohoto přístupu upouští. Potlačení šumu se pro jednotlivé záznamy a pro konkrétní případy zkreslení a zarušení řeší individuálně. Operátor takového systému mění vhodně vybrané parametry s cílem dosáhnout maximální subjektivní kvality výsledného zvukového záznamu. Takováto práce je podle [75] dokonce označována za uměleckou tvorbu.

V kapitole č. 8, ve které je představeno vlastní řešení systému na potlačení aditivního šumu pozadí, je zisková funkce nahrazena obecnějším přístupem nazvaným *prahovací pravidlo*. Toto prahovací pravidlo pak dovoluje ještě větší stupeň volnosti při návrhu systému a metod pro potlačení šumu, resp. stanovení odhadu originálního signálu, než například parametrický Wienerův filtr z definice (6.30).

Na následujícím Obr. 6.3 jsou zobrazeny rozdílné průběhy ziskových funkcí: Wienerova krátkodobého filtru $H_{W_i}(k)$ (6.24), amplitudového $H_{MS_i}(k)$ (6.26) a výkonového $H_{PS_i}(k)$ (6.28) spektrálního odečítání. Tyto ziskové funkce byly získány pro různé vstupní hodnoty SNR , kde $SNR = \sigma_x^2 / \sigma_m^2$ a σ_x^2 , σ_m^2 jsou rozptyly signálů $x(n)$ a $m(n)$. A kde $x(n)$ a $m(n)$ jsou bílé šумы, tj. platí $x(n) \sim \text{No}(0, \sigma_x^2)$ a $m(n) \sim \text{No}(0, \sigma_m^2)$. Pro účely zobrazení ziskových funkcí z tohoto obrázku a z charakteru signálů $x(n)$ a $m(n)$, byly druhé mocniny modulových spekter z (6.24), (6.26) a (6.28) nahrazeny rozptyly σ_x^2 a σ_m^2 .



Obr. 6.3 Ziskové funkce pro Wienerův filtr H_W , amplitudové H_{MS} a výkonové H_{PS} spektrální odečítání [19].

Průběhy z tohoto grafu zobrazují zvyšující se útlum ziskových funkcí se snižující se hodnotou vstupního SNR . Například útlum ziskové funkce pro Wienerův filtr H_W je roven hodnotě 6 dB při $SNR = 0$ dB. Při této hodnotě SNR jsou si výkony signálů $x(n)$ a $m(n)$ rovny a to ve všech frekvenčních pásmech. Dále jde z toho grafy vypožorovat, že nejvyššího útlumu dosahuje amplitudové spektrální odečítání a nejnižší výkonové spektrální odečítání.

7 Adaptivní segmentační schéma

Jedním se standardní přístupů jak zpracovávat dlouhotrvající signál v transformační oblasti je jeho dělení, neboli segmentace na kratší úseky nazývané segmenty. U řečových signálů je standardní délka jednoho segmentu cca 10 až 20 ms, tato doba odpovídá fyziologickým možnostem lidského hlasového ústrojí změnit postavení mluvidel při promluvě [23]. Takto dlouhé segmenty řečových signálů jsou považovány za stacionární. U hudebních signálů ovšem nelze nalézt nějakou optimální délku segmentů, protože tyto signály jsou mimo zpěv tvořeny převážně hudebními nástroji a pro ně, žádné takové omezení neplatí.

Při návrhu systému na potlačení šumu v hudebních signálech, který pracuje v transformační oblasti, se často řeší problém vhodného výběru oblastí nebo segmentů, nad kterými bude prováděna tato transformace. Větší délka segmentu znamená lepší rozlišení v frekvenční oblasti a kratší délka segmentu naopak lepší rozlišení v časové oblasti, kratší segmenty dokáží lépe sledovat časové změny v analyzovaném signálu. U systému na potlačení aditivního šumu, které byly popsány v minulé kapitole, je zpravidla žádoucí, aby toto frekvenční rozlišení bylo co možná největší. A to hlavně z důvodu, aby potlačení šumu prováděné ziskovými funkcemi (6.30) nad koeficienty příslušné transformace, bylo co možná nejdetailejší. A taky aby se do výstupního signálu zavádělo co nejmenší množství různých zkreslení a zvukových artefaktů, které jsou často důsledkem právě nízkého frekvenčního rozlišení [4][7][13].

Ve většině systému na potlačení šumu v hudebních signálech je segmentace fixní (pevná) a to i naproti skutečnosti, že nelze najít pro tuto třídu signálu optimální délku segmentu jako v případě řečových signálů. V hudebních signálech se často vyskytují úseky, které mají relativně velmi dlouhou dobu trvání, např. v záznamech vážné hudby, operního zpěvu a podobně. Druhým extrémem jsou úseky, které ve velmi krátké době, změní svůj tvar a velikost. V řeči se může jednat o plovivní souhlásky a v hudebních signálech například o údery na bicí a podobně. V případě pevné segmentace tyto krátké úseky aditivně zaruší svým širokopásmovým spektrem okolní kmitočty, podobně jako například bílý šum. A ve výsledku tím znemožní jakékoliv další operace nad koeficienty příslušné transformační oblasti v daném segmentu.

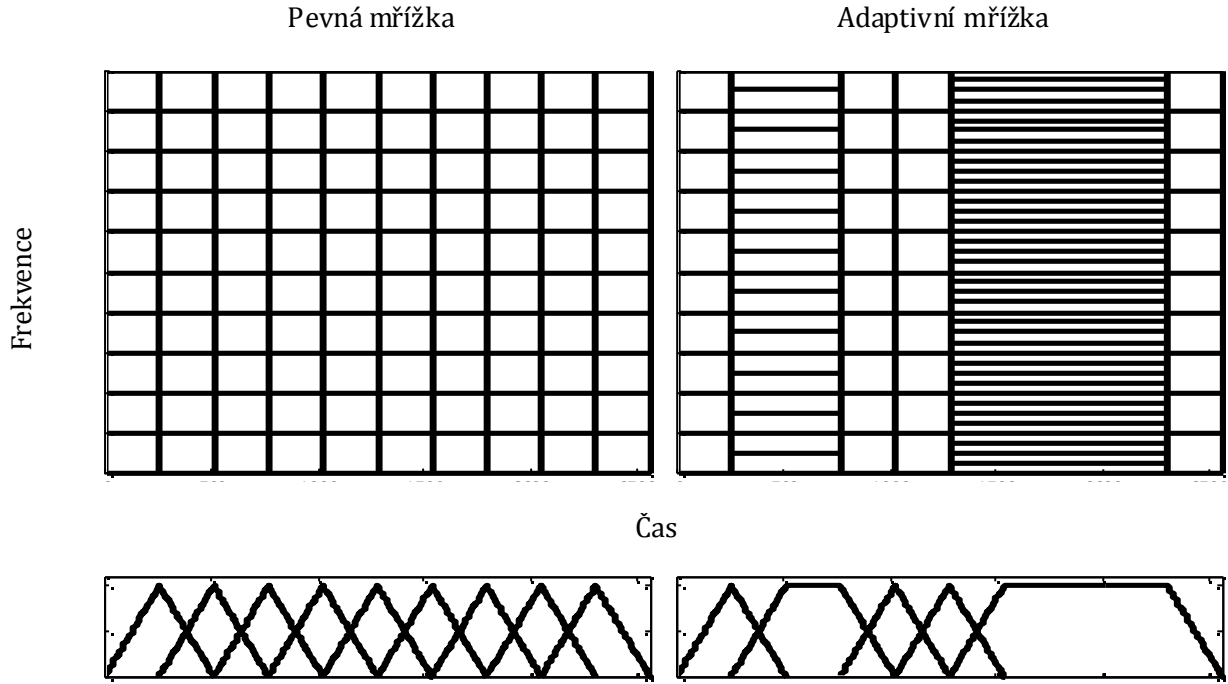
Myšlenkou adaptivní segmentace je nalézt vhodnou posloupnost po sobě jdoucích segmentů proměnné délky a to v celém analyzovaném signálu. A to tak, aby co nejlépe respektovali míru časové proměnlivosti analyzovaného signálu. Nad těmito segmenty pak bude v transformační oblasti prováděna filtrace pomocí ziskových funkcí (6.30) a tzv. *prahovacích pravidel*. Ty budou zavedeny v kapitole č. 8

7.1 Adaptivní segmentace

Na Obr. 7.1 jsou schématicky znázorněny rozdíly mezi pevnou a adaptivní segmentací z pohledu rozdílného časo-frekvenčního rozlišení [52]. V levém sloupci dole je nejprve zobrazena posloupnost váhovacích oken délky N vzorků. Na tomto obrázku jsou zobrazena trojúhelníková váhovací okna. Jelikož mají všechna okna konstantní délku, tak se jedná o pevnou segmentaci. Rozlišení v časo-frekvenční oblasti je také konstantní, to symbolizují obdélníky konstantní velikosti, viz pevná mřížka. Vpravo dole je na tomto obrázku zobrazena adaptivní segmentace. Adaptivní váhovací okna jsou vytvořena sloučením sousedních oken. Díky správnému tvaru těchto oken a

velikosti překryvu lze zachovat podmínky perfektní rekonstrukce. Adaptivní mřížka zobrazuje variabilní rozlišení v časo-frekvenční oblasti. Pokud se délka okna zdvojnásobí, tak se také zdvojnásobí rozlišení ve frekvenční oblasti ovšem za cenu polovičního rozlišení v časové oblasti. Plocha v jednotlivých obdélnících na této mřížce musí být vždy stejná, viz Heisenberg - Gaborův princip neurčitosti [29].

V následujícím textu bude použit symbol N pro délku nejkratšího adaptivního segmentu. Symbol L pak bude použit pro aktuální délku adaptivního segmentu. Je zřejmé, že tato délka bude většinou pro každý segment jiná. Překryv mezi sousedními segmenty bude mít vždy velikost $N/2$.



Obr. 7.1 Rozdíly v časo-frekvenčním rozlišení mezi pevnou a adaptivní segmentací [65].

Uvažujme takovou segmentaci, kdy dlouhotrvající signál $y(n)$ bude rozdělen na segmenty $y_{\text{SEG}}(n, p, q)$

$$y_{\text{SEG}}(n, p, q) = \begin{cases} w_A(n, q) y\left(n + (p-1)\frac{N}{2}\right) & \text{pro } n = 0, 1, \dots, (q-1)\frac{N}{2} + N - 1 \\ 0 & \text{pro } n \neq 0, 1, \dots, (q-1)\frac{N}{2} + N - 1 \end{cases}, \quad (7.1)$$

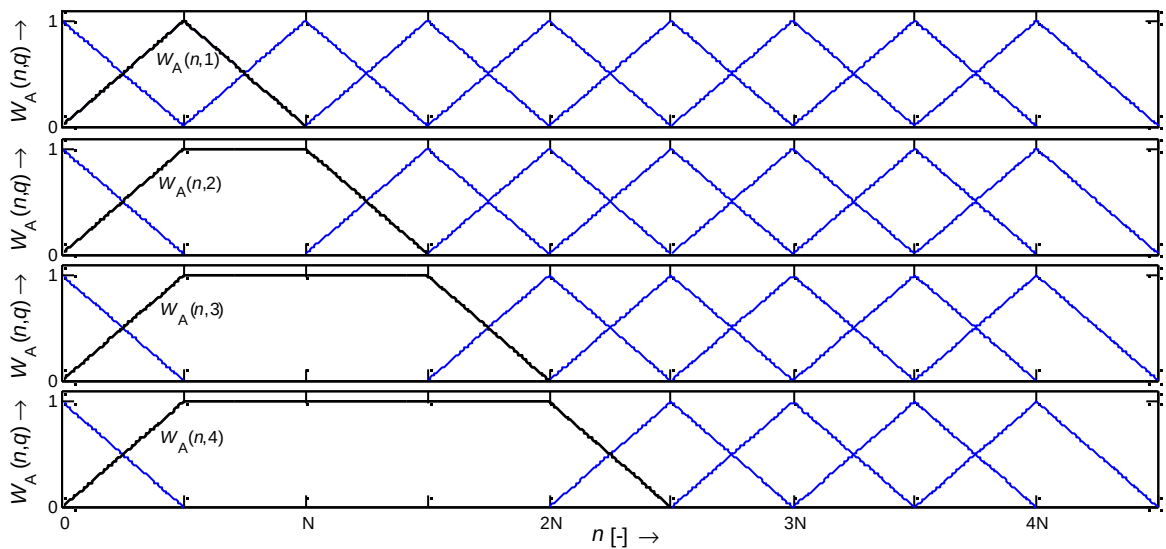
kde p je časový index toho segmentu, resp. posunutí toho segmentu vůči začátku $y(n)$ a symbol q vyjadřuje výslednou délku toho segmentu. Mezi sousedními segmenty $y_{\text{SEG}}(n, p, q)$ dochází k překryvu o délce $N/2$. Symbol $w_A(n, q)$ je použit pro adaptivní váhovací okno o délce q s počátkem $n = 0$

$$w_A(n, q) = \sum_{l=0}^{q-1} w\left(n - l\frac{N}{2}\right) \quad (7.2)$$

kde $w(n)$ je základní váhovací okno délky N , např. Hannovo okno

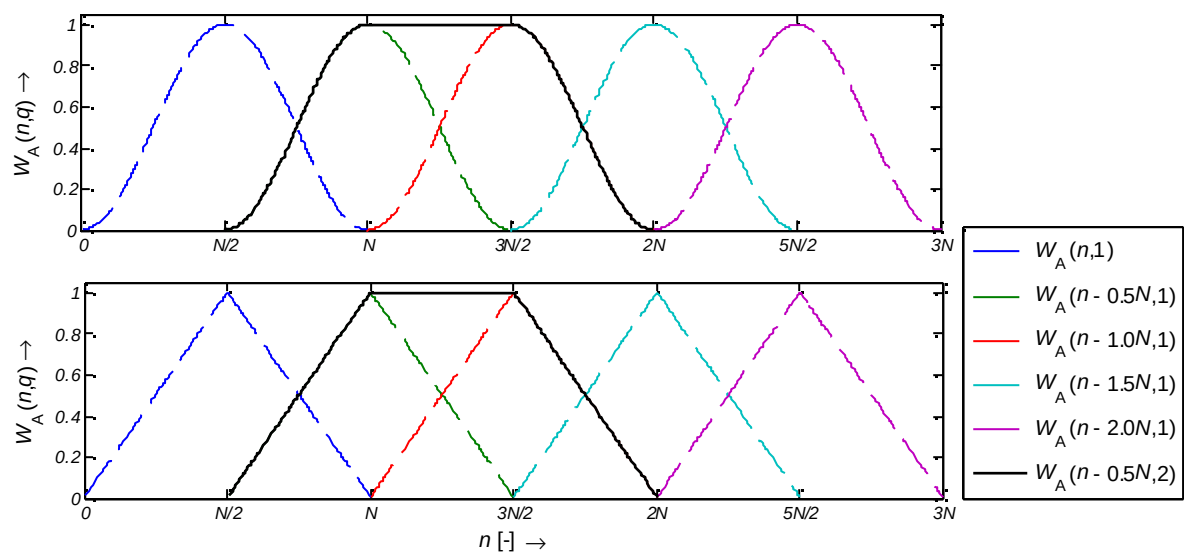
$$w(n) = \begin{cases} 0.5 \left(1 - \cos\left(2\pi\frac{n}{N}\right) \right), & \text{pro } n = 0, \dots, N-1 \\ 0, & \text{pro } n \neq 0, \dots, N-1 \end{cases} \quad (7.3)$$

Na Obr. 7.2 je graficky znázorněn význam parametru q . Lze tvrdit, že $q = 1$ odpovídá oknu o délce N vzorků, $q = 2$ odpovídá oknu o délce $3N/2$ vzorků, $q = 3$ délce $4N/2$ vzorků atd.



Obr. 7.2 Adaptivní váhovací okno $w_A(n, q)$ pro $q = 1$ až 4.

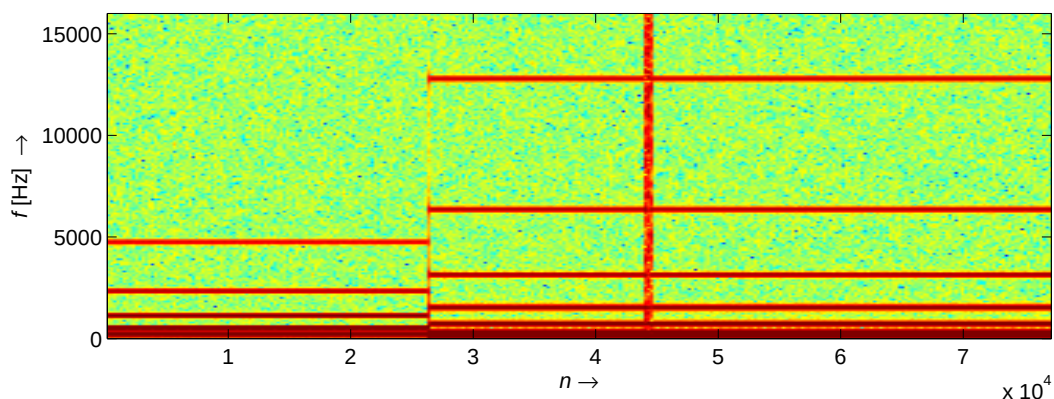
V případě jiného typu okna, než je trojúhelníkové okno, je význam parametru q stejný, viz Obr. 7.3 pro případ Hannova okna. Výsledné okno je zde značeno černou barvou a vždy se jedná o součet q oken $w(n)$ o délce N vzorků.



Obr. 7.3 Hannova (nahore) a trojúhelníková (dole) adaptivní váhovací okna.

7.2 Možnosti využití adaptivní segmentace

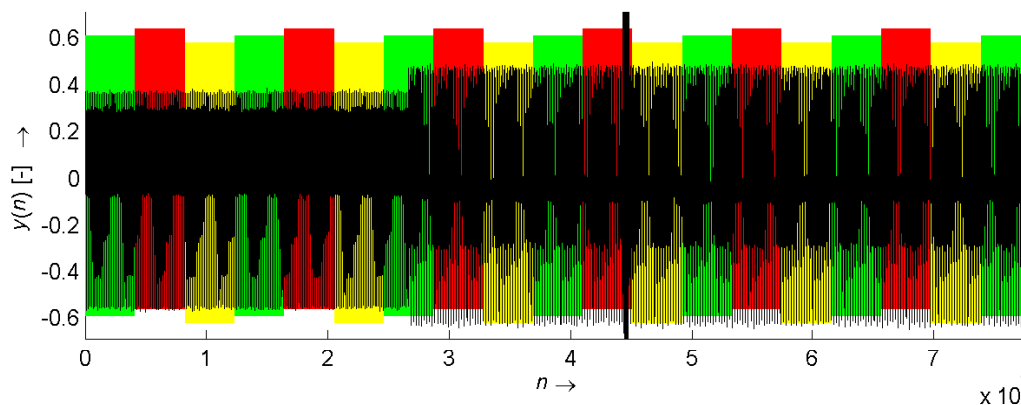
V této kapitole budou diskutovány výhody využití adaptivního segmentačního schématu, tj. adaptivní segmentace, při odhadech originálního signálu ze vstupního signálu zarušeného aditivním šumem. Tyto odhady budou prováděny pomocí krátkodobého Wienerova filtru (6.16). Budou zde srovnávány tři způsoby segmentace, dvě pevné a jedna adaptivní. Konkrétně se jedná o pevnou segmentaci skrátkým oknem (256 vzorků), dlouhým oknem (4096 vzorků) a adaptivní segmentaci.



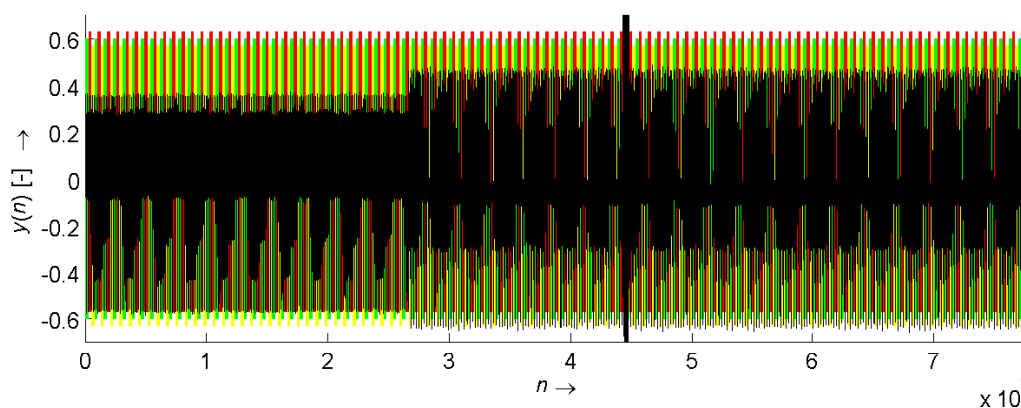
Obr. 7.4 Vstupní umělý nestacionární signál s výrazným šumem na pozadí.

Vstupním signál je zobrazen na spektrogramu z Obr. 7.4. Jedná se o umělý nestacionární signál, který byl zarušen aditivním bílým šumem. Lze vidět, že tento signál má výraznou harmonickou strukturu. Tato harmonická struktura je konstantní pro $n \in \langle 0, 2.6 \times 10^4 \rangle$ a pro $n \in \langle 2.6 \times 10^4, 7.7 \times 10^4 \rangle$, tedy v čase $n = 2.6 \times 10^4$ dochází ke změně v této struktuře. Další změna se nachází přibližně v čase $n = 4.45 \times 10^4$. V tomto okamžiku se vyskytuje širokopásmový šum s velmi krátkou dobou trvání (576 vzorků). Další parametry tohoto signálu jsou: celková délka signálu je $N_C = 7.7 \times 10^4$ vzorků a vzorkovací kmitočet je $f_s = 44\,100$ Hz.

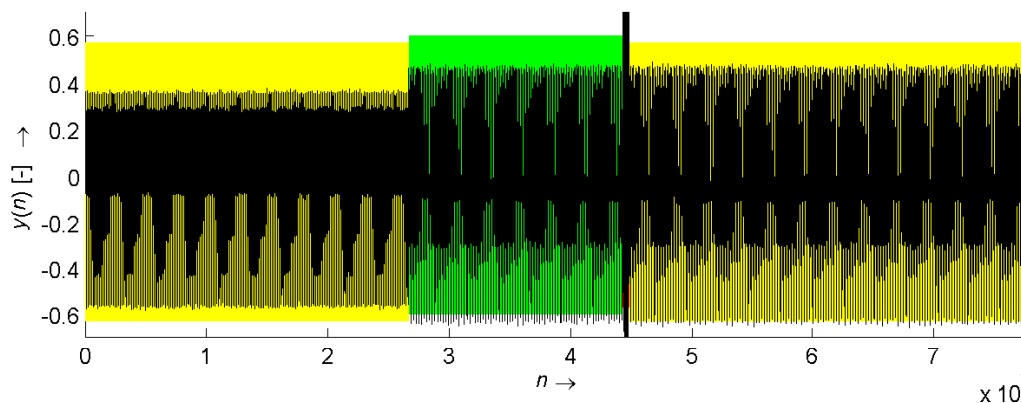
Před samotnou filtrací krátkodobým Wienerovým filtrem (6.16) s cílem potlačit šum na pozadí, bude tento dlouhotrvající signál segmentován na kratší úseky. Budou přitom využity tři rozdílné způsoby segmentace. Tyto tři rozdílné způsoby jsou graficky naznačeny na Obr. 7.5 až Obr. 7.8. Barevné obdélníky zde symbolizují jednotlivé po sobě jdoucí segmenty. V žádné z těchto segmentací nebyl použit překryv sousedních segmentů a jako váhovací okno bylo použito obdélníkové okno.



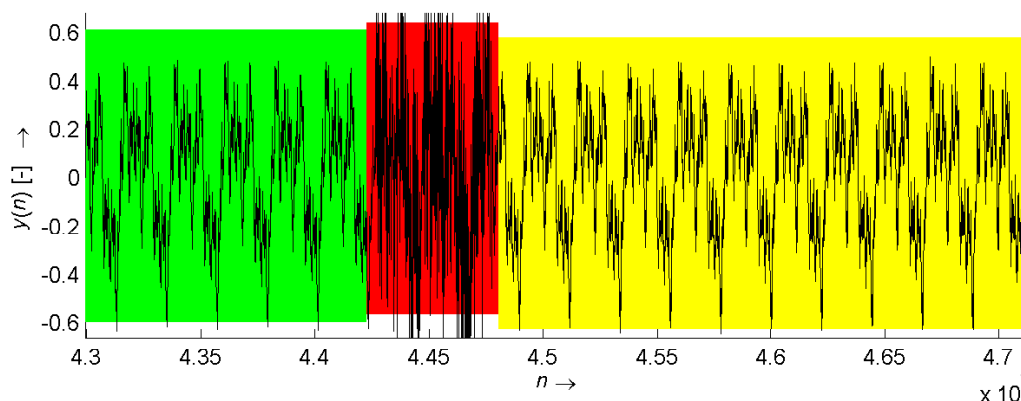
Obr. 7.5 Pevná segmentace s dlouhým oknem, 4096 vzorků.



Obr. 7.6 Pevná segmentace s krátkým oknem, 256 vzorků.



Obr. 7.7 Adaptivní segmentace s proměnou délkou oken.



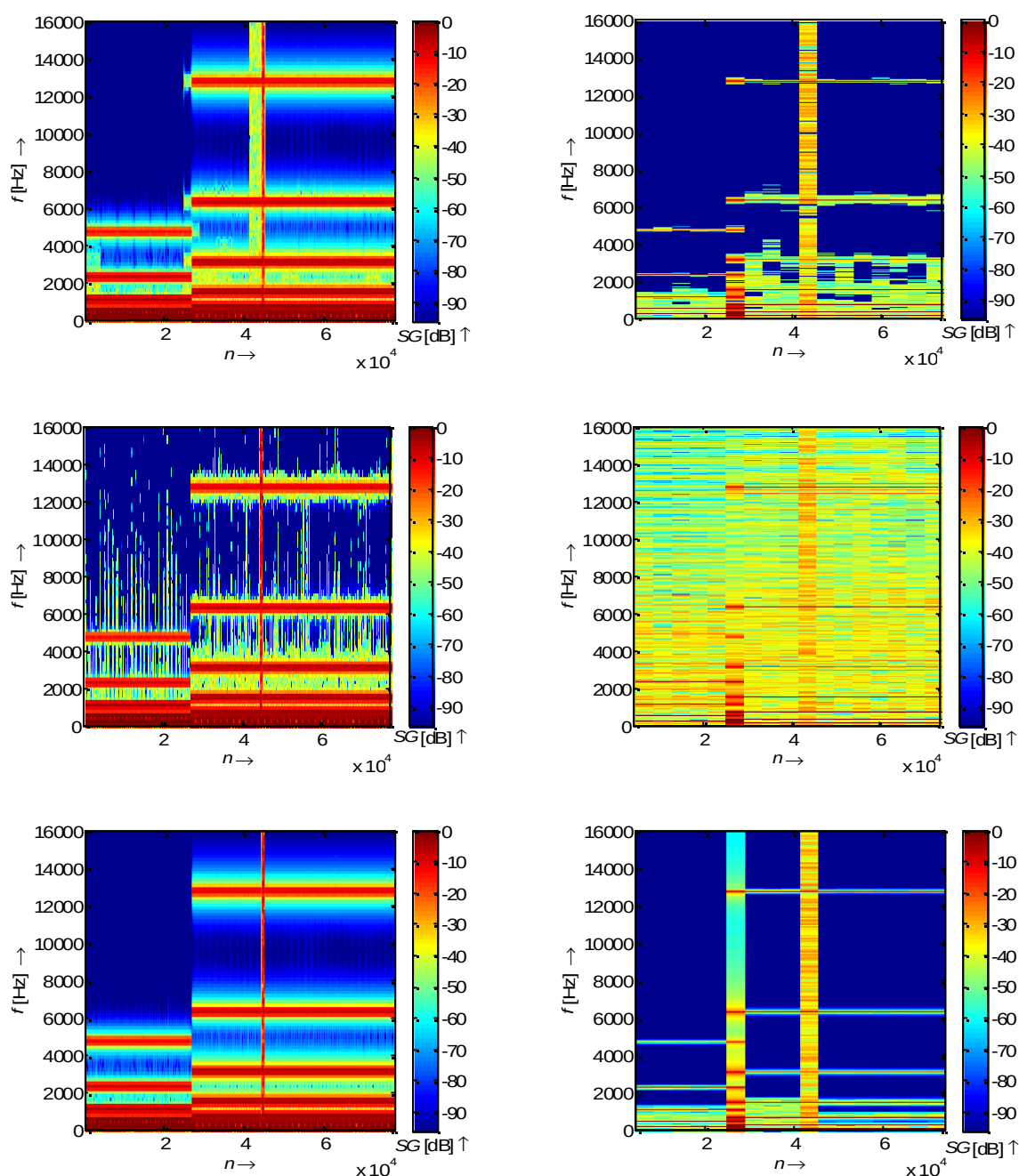
Obr. 7.8 Detail adaptivní segmentace v okolí vzorku $n = 4.45 \times 10^4$.

U adaptivní segmentace byly použity takové délky segmentů, aby pokryly stacionární úseky vstupního nestacionárního signálu.

Po aplikaci krátkodobého Wienerova filtru byly získány tři výstupní signály s potlačeným šumem pozadí. Spektrogramy těchto výstupních signálů jsou zobrazeny na Obr. 7.9. Jak lze vypozorovat z těchto spektrogramů, tak míra potlačení šumu pozadí je závislá na typu použité segmentace. Delší okna dovolují lépe potlačit šum ve stacionárních částech, ale hůře v oblastech kde dochází k rychlé změně, např. v časech $n = 2.6 \times 10^4$ a $n = 4.45 \times 10^4$. V těchto okamžicích zde dochází ke spektrálnímu prosakování v časové oblasti. Tento typ prosakování je velmi zřetelný na spektrogramu z prvního řádku a levého sloupce. Naproti tomu krátké délky oken dovolují lépe sledovat časové změny, ovšem za cenu nižšího potlačení ve stacionárních částech. Uplatňuje se zde jev spektrálního prosakování ve frekvenční oblasti, viz spektrogram z druhého řádku a pravého sloupce.

Adaptivní segmentační schéma spojuje výhody obou délek oken. V stacionárních částech dokáže potlačit šum pozadí podobně jako dlouhé typy oken u pevné segmentace a v úsecích s rychlou změnou signálu zase dokáže sledovat tyto změny podobně jako pevná segmentace, která využívá krátké délky oken. Spektrogramy výstupního signálu pro případ adaptivní segmentace jsou zobrazeny na posledním řádku.

Nevýhodou adaptivní segmentace je především skutečnost, že je potřeba nalézt vhodný algoritmus, který by dokázal tuto segmentaci provádět automatizovaně. Jako příklad takového algoritmu může být zmíněn algoritmus *Greedy selection*. Tento nový a perspektivní algoritmus bude představen v následující kapitole. Dále pak v kapitole č. 8 bude představeno vlastní řešení algoritmu pro adaptivní segmentaci, které je právě založeno na modifikaci algoritmu *Greedy selection*.



Obr. 7.9 Výstupní filtrované signály pro rozdílné typy segmentací a pro rozdílné nastavení spektrogramů. Na prvním řádku je pevná segmentace s dlouhým oknem (4096 vzorků), na druhém řádku je pevná segmentace s krátkým oknem (256 vzorků) a na posledním řádku je adaptivní segmentace. V levém sloupci bylo použito nastavení spektrogramu s vysokým stupněm rozlišení v časové oblasti. A v pravém sloupci pak nastavení pro vysoké rozlišení ve frekvenční oblasti.

7.3 Algoritmus Greedy selection

Jedním z algoritmů pro adaptivní segmentaci je algoritmus nazvaný *Greedy selection*. Tento algoritmus byl poprvé publikován v roce 2008 v [64] a v roce 2011 byl úspěšně použit pro schéma adaptivní časově-frekvenční analýzy a rychlou rekonstrukci tzv. *framů* [65]. Jedná se tedy o relativně nový algoritmus.

Tento algoritmus zavádí koeficient C pro měření tzv. *lokální spektrální špičatosti* (angl. local spectral kurtosis)

$$C(y_{\text{SEG}}) = \frac{\sum_{k=0}^{L-1} |\sum_{n=0}^{L-1} y_{\text{SEG}}(n) e^{-j2\pi kn/L}|^4}{(\sum_{k=0}^{L-1} |\sum_{n=0}^{L-1} y_{\text{SEG}}(n) e^{-j2\pi kn/L}|^2)^2}, \quad (7.4)$$

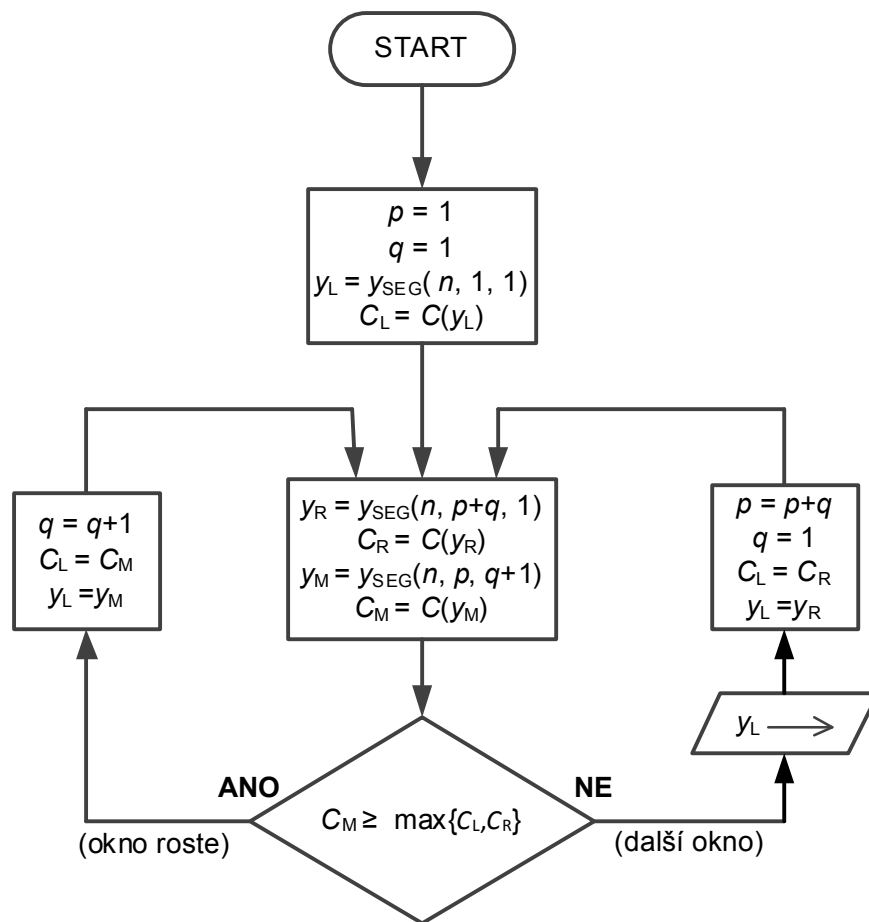
kde y_{SEG} je jeden segment dlouhotrvajícího signálu $y(n)$ podle vztahu (7.1) o délce $L = (q-1)\frac{N}{2} + N$ vzorků. Rovnici (7.4) lze přepsat jako

$$C(y_{\text{SEG}}) = \frac{\sum_{k=0}^{L-1} |\text{DFT}\{y_{\text{SEG}}(n)\}|^4}{(\sum_{k=0}^{L-1} |\text{DFT}\{y_{\text{SEG}}(n)\}|^2)^2}, \quad (7.5)$$

kde $\text{DFT}\{\}$ značí diskretní Fourierovu transformaci nad segment y_{SEG} délky L

$$\text{DFT}\{y_{\text{SEG}}\} = \sum_{n=0}^{L-1} y_{\text{SEG}}(n) e^{-j2\pi kn/L}. \quad (7.6)$$

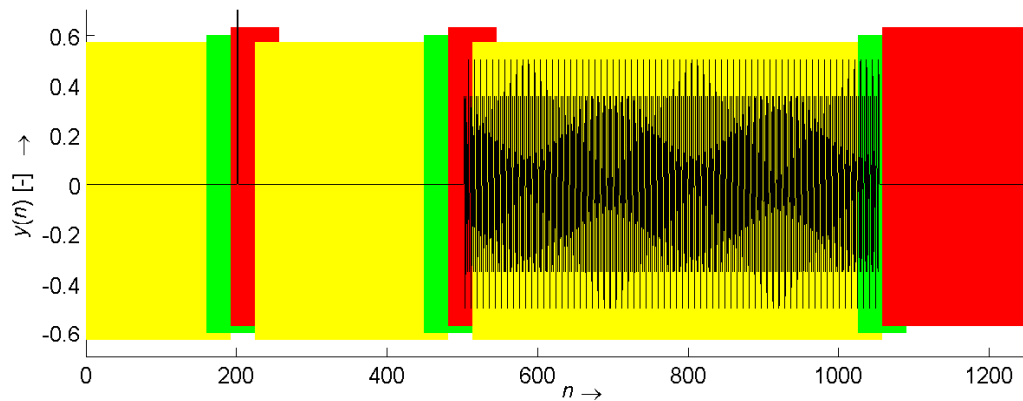
Principem tohoto algoritmu je měření koeficientů lokální spektrální špičatosti C (7.5) a to vždy z dvou po sobě jdoucích segmentů. Pro větší názornost budou tyto dva segmenty označovány jako levý y_L a pravý y_R segment. Oba tyto segmenty mají na začátku shodnou délku N vzorků, délka segmentu y_L se bude zvětšovat skrokem $N/2$ vzorků a délka segmentu y_R bude mít vždy konstantní délku N vzorků. Sloučením těchto dvou segmentů y_L a y_R do jednoho, vznikne segment y_M (z angl. *merge* - splynout.). Z každého z těchto tří segmentů jsou změřeny koeficienty lokální spektrální špičatosti C , tyto koeficienty budou značeny C_L , C_R a C_M . Následně jsou tyto koeficienty vzájemně porovnávány. Pokud je koeficient C_M větší než C_L a C_R pak se délka segmentu y_L zvětší o $N/2$ vzorků, tedy o velikost překryvu. Segment y_L “roste”. Poté se tento proces opakuje, vybere se nový pravý segment y_R a nový sloučený segment y_M . Pokud opět platí, že $C_M \geq \max\{C_L, C_R\}$, tak segment y_L opět poroste. V opačném případě se růst segmentu y_L zastaví a jako nový levý segment y_L se označí následující segment. Detailněji je tento algoritmus popsán v blokovém schématu na Obr. 7.10.



Obr. 7.10 Algoritmus Greedy selection. Symbol $y_L \rightarrow$ značí uložení y_L do výstupní paměti, pro další zpracování.

Symbol $y_L \rightarrow$ v tomto obrázku značí uložení y_L do výstupní paměti, pro další zpracování. Tento algoritmus skončí v momentě, kdy skončí signál $y(n)$.

Na Obr. 7.11 je zobrazena segmentace dlouhotrvajícího umělého zvukového signálu $y(n)$ podle algoritmu Greedy selection. Tento signál je tvořen jednotkovým impulsem v čase $n = 200$ a harmonickým signálem. Černou barvou je zde zobrazen časový průběh tohoto signálu a barevnými obdélníky jsou vyznačeny jednotlivé segmenty. Na rozhraní těchto segmentů si lze všimnout vzájemného překryvu. Ten je vždy konstantní a má v tomto příkladu délku 32 vzorků. Ostatní parametry jsou: základní délka okna (7.3) je $N = 64$, bylo použito Hannovo okno, vzorkovací kmitočet $f_s = 8$ kHz a frekvence harmonické části $f_1 = 3$ kHz.



Obr. 7.11 Segmentace dlouhotrvajícího umělého zvukového signálu $y(n)$ podle algoritmu Greedy selection.

Algoritmus Greedy selection v uvedeném příkladu pracuje správně, protože zde správně označí stacionární úseky. Místa, kde dochází ke změně signálu, se vyznačují malou hodnotou koeficientu C_M , a proto zde začne nový segment, viz Obr. 7.10. Tento algoritmus samozřejmě nedokáže sledovat rychlejší změny, než je délka použitého základního okna N . Proto je jednotkový impulz v čase $n = 200$ s délkou 1 vzorek uvnitř dvou segmentů o mnohem větších délkách $N = 64$ vzorků. Tyto segmenty jsou dva, protože zde dochází vzájemnému překryvu.

8 Systém na potlačení šumu v hudebních signálech – vlastní řešení

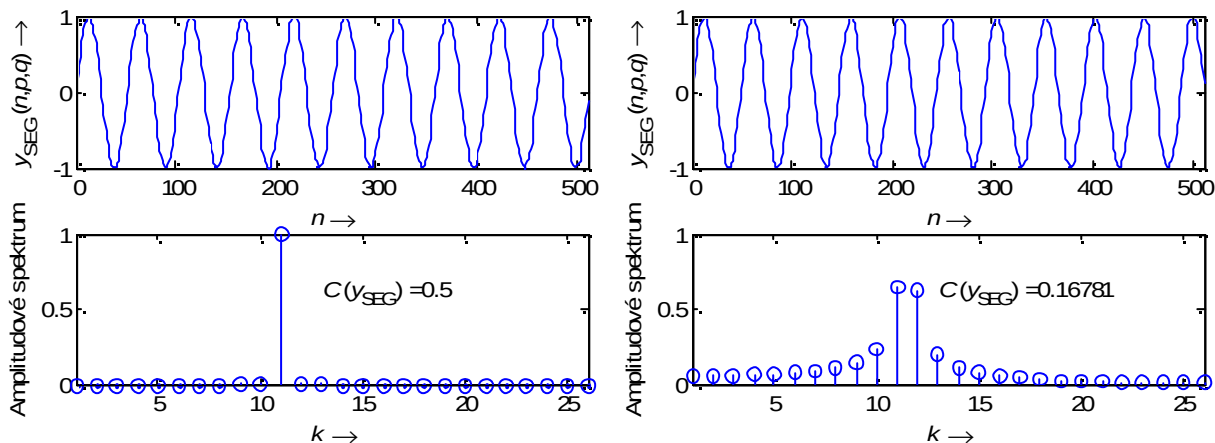
V této kapitole bude představeno vlastní řešení systému na potlačení aditivního šumu pozadí v hudebních signálech technikami prahování. V tomto řešení bude použit modifikovaný způsob adaptivní segmentace z kapitoly 7, který jsem nazval Maximum greedy selection. Dále v této kapitole bude poukázáno na vzájemnou podobnost ziskových funkcí zavedených v kapitole 6 a nově zavedených prahovacích pravidel. Závěr kapitoly je věnován provedenému poslechovému experimentu, jeho popisu a diskuzi nad získanými výsledky. Tento experiment si klade za cíl porovnat výslednou subjektivně vnímanou kvalitu posluchači pro rozdílné způsoby použité segmentace a při rozdílných velikostech zarušení užitečného signálu šumem pozadí.

8.1 Modifikovaná varianta algoritmu Greedy selection

8.1.1 Snížení vlivu spektrálního prosakování

Jedna z nevýhod algoritmu Greedy Selection pro adaptivní segmentaci je fakt, že dělení do segmentu je náchylné na jev spektrálního prosakování (angl., spectral leakage). Je to důsledek toho, že koeficienty lokální spektrální špičatosti jsou vypočteny z odhadů výkonových spektrálních hustot. Ke snížení vlivu tohoto spektrálního prosakování budou proto analyzované segmenty doplněny nulovými vzorky, s cílem vyhladit obálku, tvořící odhad výkonové spektrální hustoty. Na takovémto odhadu jsou poté viditelné jednotlivé laloky (hlavní a postranní), tvořené použitým váhovacím oknem. Rozdíly v hodnotách koeficientů lokálních spektrálních špičatostí nejsou tak velké jako v případě spekter, kde nedošlo k tomuto prodloužení nulami.

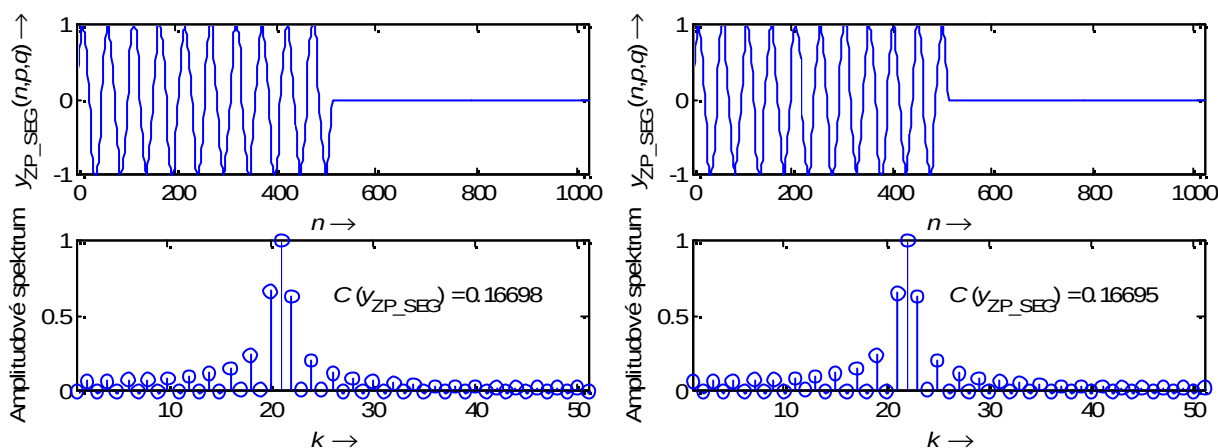
Na Obr. 8.1 jsou zobrazeny časové průběhy dvou segmentů $y_{\text{SEG}}(n, p, q)$ o délce 500 vzorků. U prvního segmentu (sloupec vlevo) nedochází k spektrálnímu prosakování. Lze vidět, že jeho amplitudové spektrum je vyjádřeno jediným spektrálním koeficientem ($k=11$). Kdyby se zperiodizoval časový průběh tohoto signálu, tak nenastanou ostré přechody mezi segmenty a tudíž taky nenastane jev spektrálního prosakování. Hodnota koeficientu lokální spektrální špičatosti $C(y_{\text{SEG}})$ je 0.5, to je jeho maximální hodnota. Na dalších grafech (sloupec vpravo) je zobrazen časový průběh dalšího segmentu harmonického signálu, kdy toto spektrální prosakování nastává a navíc je velmi výrazné. Amplitudové spektrum tohoto signálu je vyjádřeno více než jedním koeficientem. Lze vidět, že toto spektrum má dva velmi výrazné koeficienty ($k=11$ a 12) a u ostatních koeficientů je vidět jev prosakování energie z těchto dvou výrazných segmentů. Koeficient $C(y_{\text{SEG}})$ je zde roven hodnotě 0.16781. Vliv spektrálního prosakování tedy snižuje hodnotu koeficientů lokálních spektrálních špičatostí. Takto velký rozdíl v hodnotách těchto koeficientů zapříčiní to, že následující segment bude uvažován jako nový segment a nedojde ke sloučení. Zde je potřeba říci, že i pro segmenty ze stejného signálu, není hodnota koeficientu C v jednotlivých segmentech stejná. Je spektrálního prosakování je u každého ze segmentů jinak veliký.



Obr. 8.1 Vliv jevu spektrálního prosakování na hodnotu koeficientu lokální spektrální špičatosti C . Vlevo je zobrazen případ, kdy nenastává spektrální prosakování a vpravo případ s velmi výrazným jevem spektrálního prosakování.

Vliv spektrálního prosakování by se samozřejmě snížil, pokud by bylo použito jiné než obdélníkové váhovací okno. Bohužel při použití adaptivní segmentace podle (7.1) s tímto snížením u větších délek oken nelze příliš počítat. Pokud totiž poroste délka okna z délky $q = 1$ např. na hodnotu $q = 4$, nebo ještě vyšší, tak tyto adaptivní váhovací okna se budou svým tvarem blížit obdélníkovému váhovacímu oknu, viz srovnání tvaru okna z Obr. 7.2 pro různé hodnoty q .

Ke snížení vlivu spektrálního prosakování na hodnoty koeficientů C bude použita technika doplnění nulami za každý uvažovaný segment. Délka tohoto segmentu se tedy zdvojnásobí a přesně polovina tohoto segmentu bude tvořena nulovými vzorky. Tyto nulami doplněné segmenty budou označovány symbolem $y_{ZP_SEG}(n,p,q)$, kde index ZP zde značí doplnění nulami z angl. *zero padding*. Stejná situace jako v případě Obr. 8.1, jen po doplnění nulovými vzorky, je zobrazena na Obr. 8.2. Lze vidět, že rozdíly v hodnotách koeficientů C pro tyto nulou doplněné segmenty $y_{ZP_SEG}(n,p,q)$ jsou velmi malé.



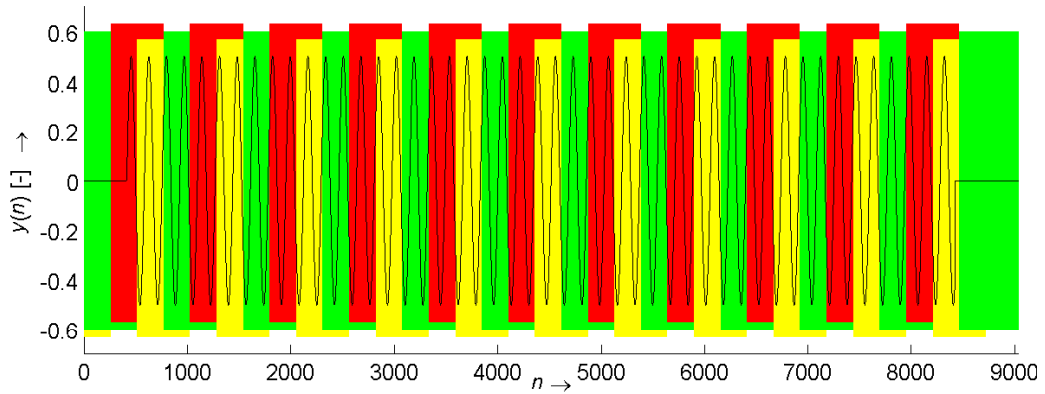
Obr. 8.2 Snížení vlivu jevu spektrálního prosakování na koeficientu C pomocí techniky doplnění nulami.

Segment $y_{\text{SEG}}(n, p, q)$ o délce L vzorků bude doplněn L nulovými vzorky tak, aby jeho výsledná délka byla $2L$ vzorků

$$y_{\text{ZP_SEG}}(n, p, q) = \sum_{n=0}^{2L} (y_{\text{SEG}}(n, p, q) + o(n, 2L)). \quad (8.1)$$

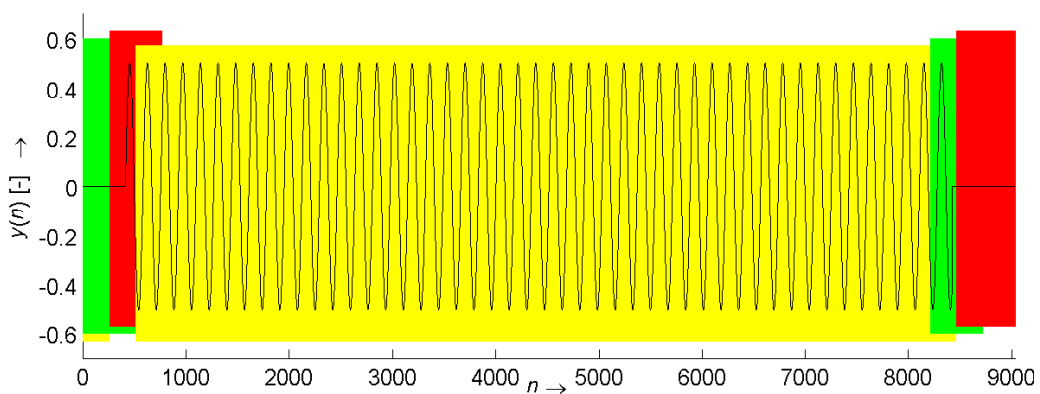
kde $o(n, 2L)$ je nulový vektor délky $2L$ a $y_{\text{ZP_SEG}}(n, p, q)$ má proto také délku $2L$. Koeficient lokální spektrální špičatosti $C(y_{\text{ZP_SEG}})$ bude vypočten stejně jako v (7.5) jen je potřeba uvažovat větší délku segmentu, tedy operace suma a DFT jsou nyní provedeny nad segmenty délky $2L$ vzorků.

Na Obr. 8.3 je zobrazen příklad adaptivní segmentace podle původního algoritmu Greedy selection, tj. bez doplnění nulovými vzorky. Segmentován je zde stacionární (harmonický) signál skmitočtem 46,875 Hz při vzorkovacím kmitočtu $f_s = 8 \text{ kHz}$ a délce základního váhovacího okna $N = 512$ vzorků a bylo použito Hannovo okno. Kmitočet byl záměrně zvolen takový, aby nastal velmi výrazný jev spektrálního prosakování. Lze vidět, že algoritmus Greedy selection pro tento typ signálu selhává. Stacionární část, tvořená tímto harmonickým signálem, zde totiž není vybrána jako jeden dlouhý segment.



Obr. 8.3 Adaptivní segmentace podle původního algoritmu Greedy selection, tj. bez doplnění nulovými vzorky.

Pokud dojde k doplnění nulovými vzorky podle (8.1), tak se sníží vliv spektrálního prosakování na hodnoty koeficientu C a adaptivní segmentace bude pro tento příklad fungovat správně, viz Obr. 8.4.



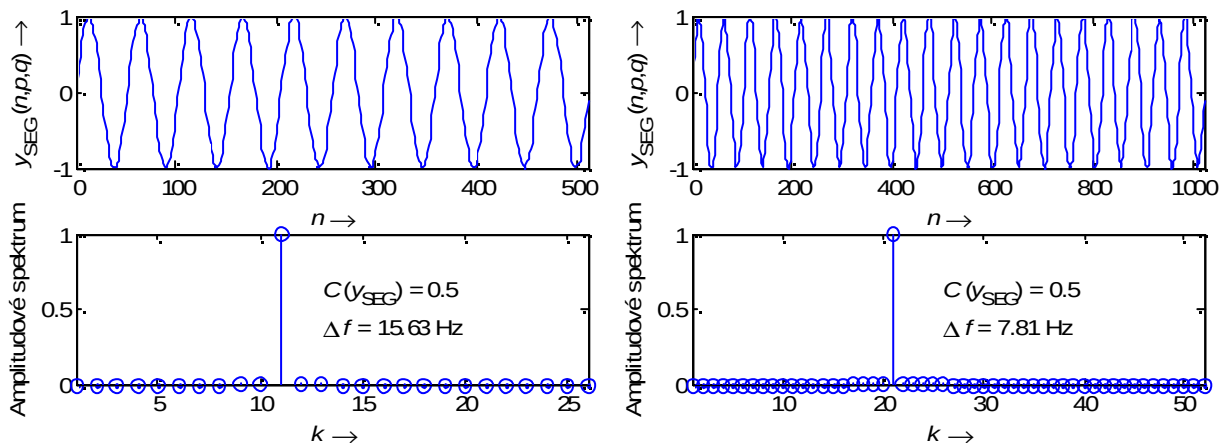
Obr. 8.4 Adaptivní segmentace s doplněním nulovými vzorky.

8.1.2 Algoritmus Maximum greedy selection

Dalším problémem algoritmu Greedy selection je skutečnost, že tento algoritmus nerespektuje zvyšující se kmitočtové rozlišení s rostoucí délkou segmentu. Pokud budou mít například dva sousední segmenty podobnou velikost koeficientů lokální spektrální špičatosti C , tak bude záležet na vnějších, zpravidla náhodných vlivech, zda se tyto segmenty sloučí v jeden. Je potřeba si uvědomit, že hudební a řečové signály jsou náhodné signály, tvořené několika dominantními harmonickými komponentami, např. základní a alikvotní tóny u hudebních nástrojů, nebo základní tón řeči a jeho vyšší harmonické u řečových signálů a dále pak komponent náhodného charakteru tvořící “zvukové pozadí” těchto harmonických komponent. A právě tyto komponenty náhodného charakteru mohou (při malých rozdílech mezi koeficienty lokálních spektrálních špičatosti C sousedních segmentů) ovlivnit rozhodovací proces o sloučení v jeden segment.

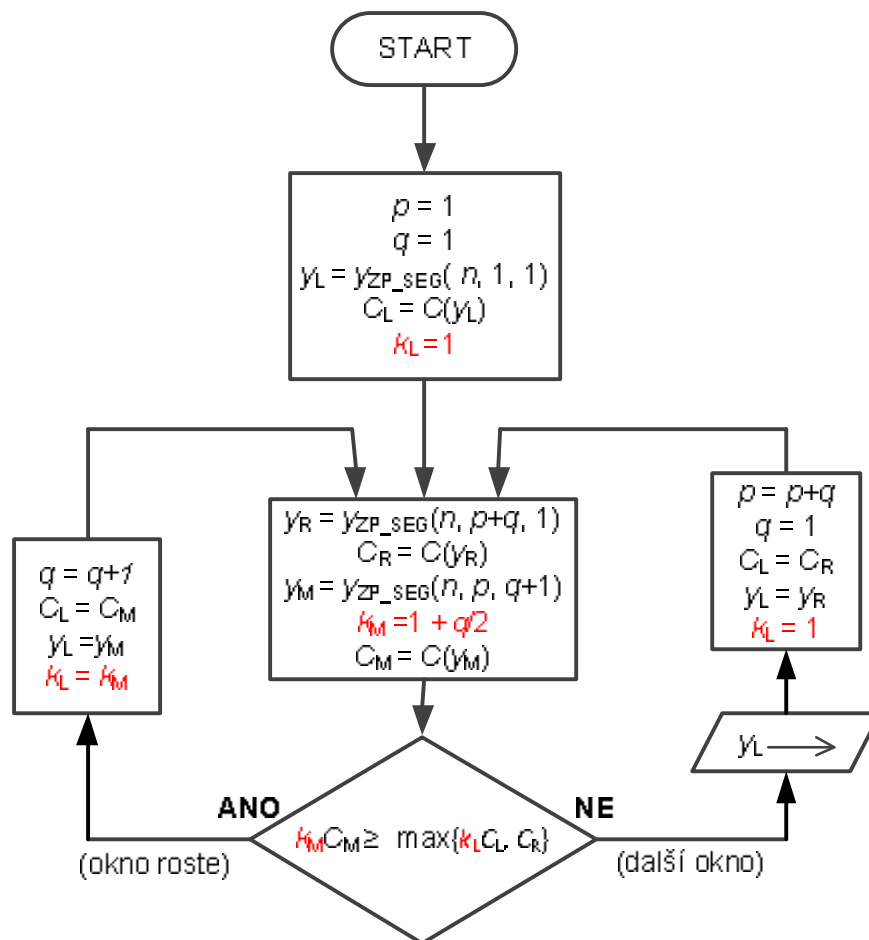
Aby se snížil vliv těchto komponent náhodného charakteru na proces slučování sousedních segmentů, tak byl algoritmus Greedy selection doplněn o započítání vlivu kmitočtového rozlišení DFT do svých rozhodovacích procesů. Tuto modifikaci jsem pojmenoval jako *Maximum greedy selection* (maximální nenasytný výběr).

Kmitočtové rozlišení $\Delta f = f_s/L$ jednoho segmentu je při konstantní hodnotě vzorkovacího kmitočtu f_s funkcí délky L tohoto segmentu. S rostoucí délkou segmentu L klesá hodnota Δf a tedy se zvyšuje kmitočtové rozlišení. Na Obr. 8.5 jsou zobrazeny časové průběhy náhodně vybraného segmentu $y_{\text{SEG}}(n, p, q)$, vč. jeho amplitudového spektra, pro rozdílné hodnoty délek L tohoto segmentu. V levém sloupci je $L = 512$ vzorků v pravém $L = 1024$ při $f_s = 8 \text{ kHz}$. Dále je v těchto grafech vynesena příslušná hodnota koeficientů lokální spektrální špičatosti C a kmitočtového rozlišení Δf a to vždy pro konkrétní délku segmentu L . Lze vidět, že pro rostoucí L roste kmitočtové rozlišení a hodnota koeficientu C je neměnná.



Obr. 8.5 Časové průběhy a amplitudové spektra náhodně vybraného segmentu o délce $L = 512$ vzorků (vlevo), a $L = 1024$ vzorků (vpravo) při $f_s = 8$ kHz.

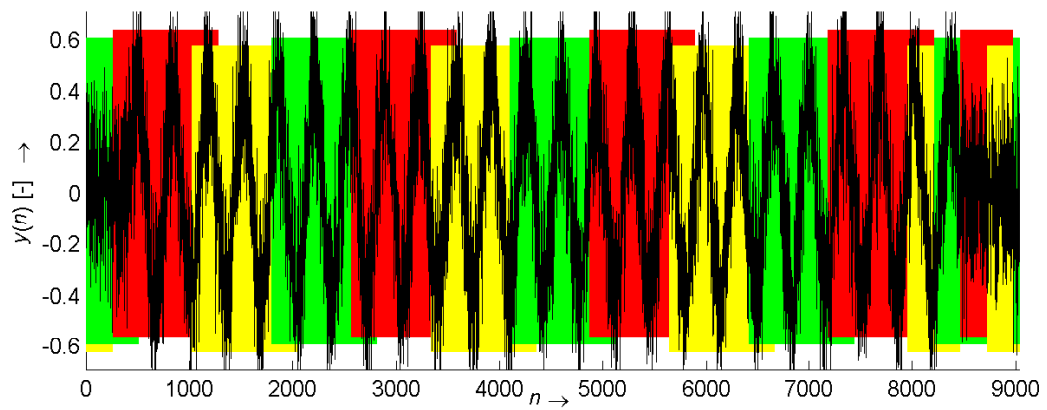
Výchozí úvaha při návrhu algoritmu Maximum greedy selection byla taková, že vliv rozdílných délek L segmentů y_L , y_L a y_M by se měl promítnout do rozhodovacích procesů o sloučení do výsledného segmentu y_M . A to z důvodu, že i když neroste hodnota koeficientu lokální spektrální špičatosti C , tak lze lépe lokalizovat jednotlivé harmonické komponenty zvukových signálů a současně se sníží vliv komponent náhodného charakteru na tento rozhodovací proces. Způsob, jakým je započítán vliv rozdílných délek L segmentů y_L , y_L a y_M do rozhodovacích procesů o sloučení, je zobrazen v algoritmu na Obr. 8.6. Modifikace, oproti původní verzi z Obr. 7.10, jsou zobrazeny červeným písmem. Dalším rozdílem oproti původnímu algoritmu Greedy selection z Obr. 7.10 je ten, že hodnoty koeficientů C_L , C_L a C_M jsou vypočteny ze segmentů $y_{ZP_SEG}(n, p, q)$, tj. ze segmentů doplněných nulovými vzorky.



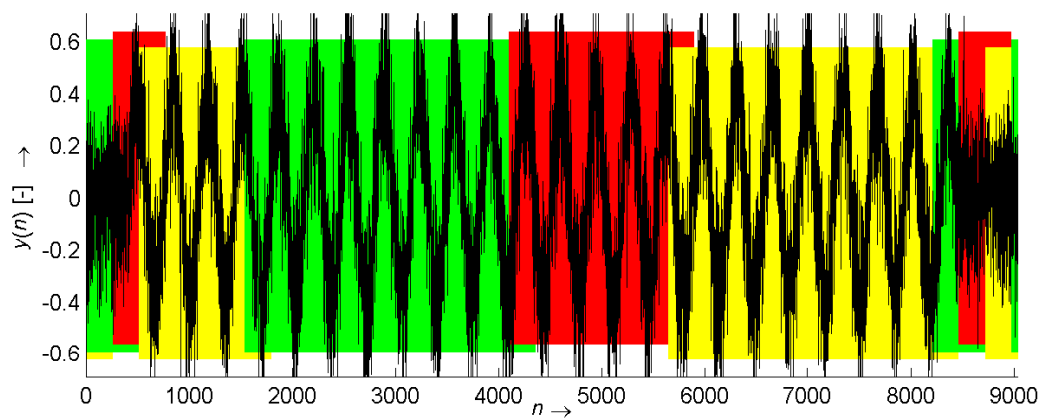
Obr. 8.6 Algoritmus Maximum greedy selection. Symbol $y_L \rightarrow$ značí uložení y_L do výstupní paměti pro další zpracování. Červeným písmem jsou vyznačeny změny oproti původní variantě algoritmu Greedy selection.

Na následujících obrázcích Obr. 8.7 až Obr. 8.9 jsou zobrazeny příklady adaptivní segmentace harmonického signálu zarušeného aditivním šumem. Barevně jsou zde naznačeny jednotlivé po sobě jdoucí segmenty. Jelikož se statické parametry tohoto signálu nemění, tak lze tento signál považovat za stacionární. Správně fungující algoritmus by měl tedy tento signál rozdělit do jediného segmentu a to i za přítomnosti aditivního šumu. Na Obr. 8.7 je zobrazena adaptivní segmentace podle algoritmu Greedy selection bez doplnění nulovými vzorky. Tedy ve své původní variantě podle algoritmu z Obr. 7.10. Na následujícím obrázku, Obr. 8.8, je zobrazen opět příklad segmentace podle algoritmu Greedy selection, ale tentokrát s doplněním nulovými vzorky. Segmenty mají v tomto příkladě větší délky, ale i tady algoritmus Greedy selection selhává. V posledním příkladě, Obr. 8.9, je zobrazena adaptivní segmentace modifikované varianty Greedy selection, nazvaná Maximum greedy selection. Díky provedeným modifikacím, tj. doplnění nulovými vzorky a započítání vlivu rozdílných délek sousedních segmentů, zde tento algoritmus pro adaptivní segmentaci funguje správně. Signál je správně rozdělen do jediného segmentu, bez uvažování přechodných jevů na začátku a konci tohoto signálu.

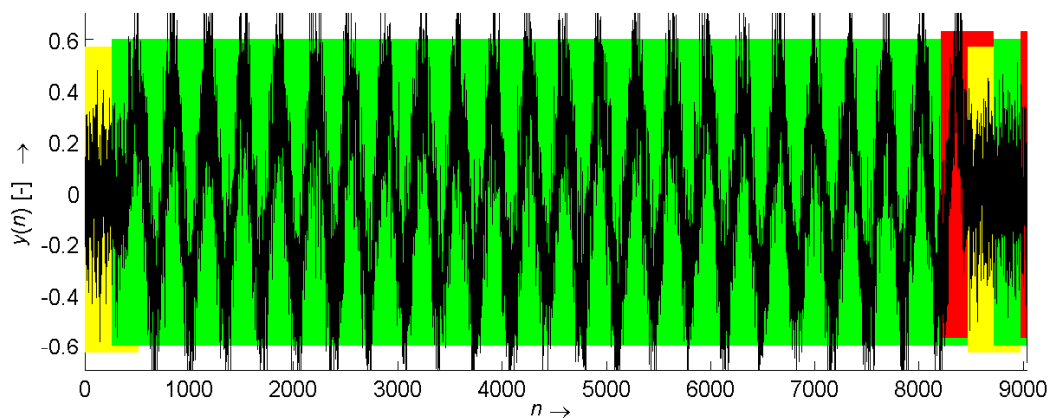
Společnými parametry pro všechny tři příklady z Obr. 8.7 až Obr. 8.9 jsou: kmitočet harmonického signálu $f = 46.875$ Hz, vzorkovací kmitočet $f_s = 8$ kHz, $SNR = 3$ dB (5.25), $N = 512$. Dále bylo použito Hannovo okno (7.3) s překryvem $N/2 = 216$ vzorků.



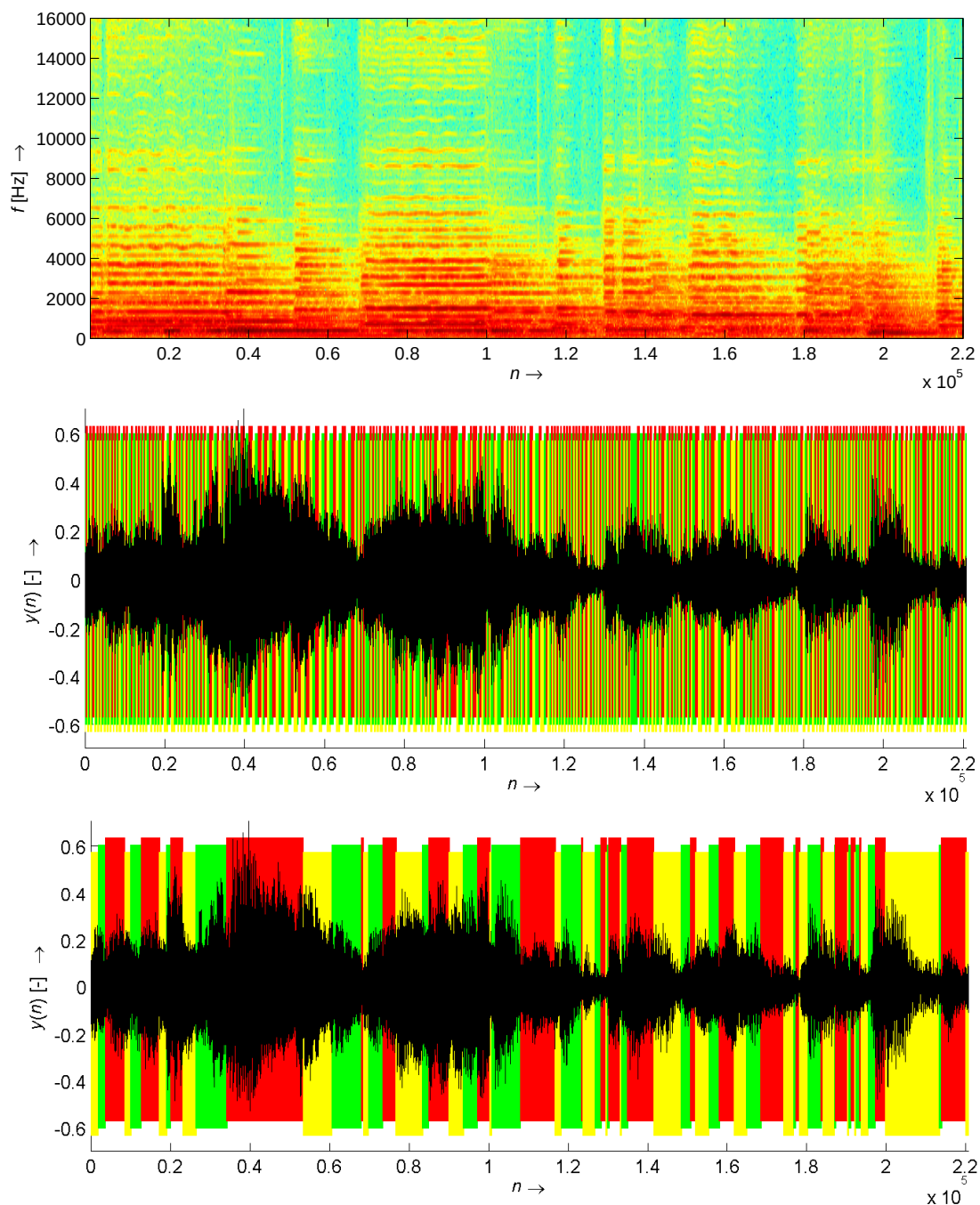
Obr. 8.7 Greedy selection , adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, bez doplnění nulovými vzorky.



Obr. 8.8 Greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, s doplněním nulovými vzorky.



Obr. 8.9 Maximum greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, s doplněním nulovými vzorky.



Obr. 8.10 Ukázka segmentace signálu ze zvukového záznamu vážné hudby. V prvním grafu je spektrogram tohoto zvukového signálu, v prostředním grafu je segmentace algoritmem Greedy selection a ve spodním grafu je segmentace algoritmem Maximum greedy selection. Délka záznamu je 5 s při $f_s = 44100$ Hz a délka základního okna $N = 512$ vzorků.

Na Obr. 8.10 je zobrazena ukázka segmentace signálu ze zvukového záznamu vážné hudby. Ze vzájemného srovnání předložených grafů lze vidět, že algoritmus Greedy selection ve své původní variantě (graf uprostřed) je zcela nepoužitelný pro adaptivní segmentaci tohoto signálu. Naproti tomu modifikovaný algoritmus provedl segmentaci mnohem úspěšněji (graf dole).

8.2 Prahovací pravidla

8.2.1 Zavedení prahovacích pravidel

Dalším krokem při návrhu vlastního řešení při návrhu systému pro potlačení nežádoucího aditivního šumu pozadí je zavedení *prahovacích pravidel*.

Model zkreslení užitečného signálu je stejně jako v kapitole 6 popsán rovnicí

$$y(n) = x(n) + m(n), \quad (8.2)$$

kde $x(n)$ reprezentuje užitečný signál a $m(n)$ je rušivý šum. Jejich součtem pak vznikne vstupní, tj. šumem zarušený, signál $y(n)$

Pokud je provedena segmentace podle (7.1) tak pro jednotlivé segmenty platí

$$y_{\text{SEG}}(n, p, q) = x_{\text{SEG}}(n, p, q) + m_{\text{SEG}}(n, p, q) \quad (8.3)$$

kde n reprezentuje diskrétní čas, p je parametr posunutí segmentu vůči počátku časové osy a parametr q vyjadřuje délku adaptivního délku okna. Význam těchto parametrů je shodný jako v kapitole 7, která právě zavádí tento způsob segmentace. Další důležitými parametry jsou délka základního okna N a délka adaptivního okna L . Vztah (8.3) vyjadřuje schéma pro adaptivní segmentaci. Způsob volby parametrů p a q v něm není nijak uveden a závisí pouze na použitém algoritmu pro adaptivní segmentaci. Pokud je například použita pevná segmentace tak vždy platí $p = 1$ a $q = 1$ a $N = L$.

Spektrum diskrétní Fourierovy transformace segmentu $x_{\text{SEG}}(n, p, q)$ o délce L je získáno jako

$$X_{\text{SEG}}(k, p, q) = \text{DFT}\{x_{\text{SEG}}(n, p, q)\} = \sum_{n=0}^{L-1} x_{\text{SEG}}(n, p, q) e^{-j2\pi kn/L} \quad (8.4)$$

Podobně je získáno DFT spektrum $M_{\text{SEG}}(k, p, q)$ segmentu $m_{\text{SEG}}(n, p, q)$. Za podmínky lineární nezávislosti signálů $x(n)$, $m(n)$ lze psát

$$Y_{\text{SEG}}(k, p, q) = X_{\text{SEG}}(k, p, q) + M_{\text{SEG}}(k, p, q) \quad (8.5)$$

Nyní bude zavedena operace prahování koeficientů DFT jako transformace nad koeficienty $Y_{\text{SEG}}(k, p, q)$ s cílem odhadnout DFT spektrum původního nezarušeného signálu $X_{\text{SEG}}(k, p, q)$. Tento odhad bude značen $\hat{X}_{\text{SEG}}(k, p, q)$

$$\hat{X}_{\text{SEG}}(k, p, q) = \delta\{Y_{\text{SEG}}(k, p, q)\}. \quad (8.6)$$

kde $\delta\{\}$ značí použité prahovací pravidlo. Kvůli zjednodušení následujících vztahů budou všechny segmenty značeny jako $Y_{\text{SEG}}(k)$ a prahování podle (8.6) nahrazeno zkrácenou formou jako $\hat{X}_{\text{SEG}}(k) = \delta\{Y_{\text{SEG}}(k)\}$.

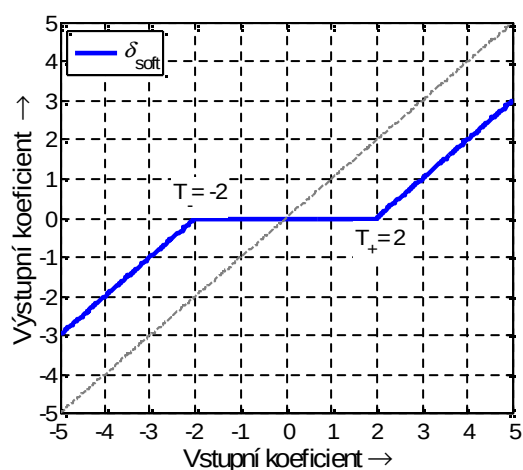
Následující vztahy (8.7) až (8.9) popisují nejzákladnější prahovací pravidla. Těmito základními prahovacími pravidly jsou: měkké prahovací pravidlo $\delta_{\text{soft}}\{\}$, tvrdé prahovací pravidlo $\delta_{\text{hard}}\{\}$ a poloměkké prahovací pravidlo $\delta_{\text{halfsoft}}\{\}$, kterého jsem autorem [35].

$$\delta_{\text{soft}}\{Y_{\text{SEG}}(k)\} = \begin{cases} 0 & \text{pro } |Y_{\text{SEG}}(k)| \leq T \\ Y_{\text{SEG}}(k) & \text{pro } |Y_{\text{SEG}}(k)| > T \end{cases} \quad (8.7)$$

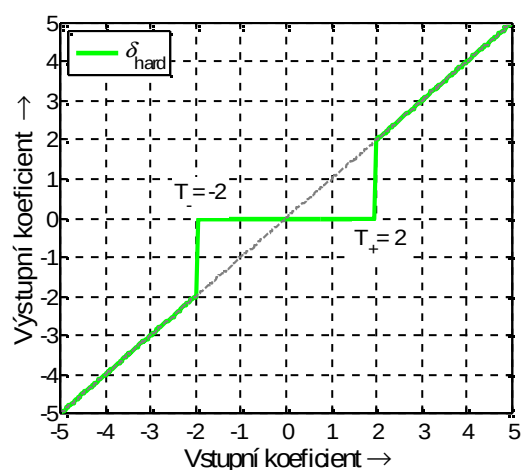
$$\delta_{\text{hard}}\{Y_{\text{SEG}}(k)\} = \begin{cases} 0 & \text{pro } |Y_{\text{SEG}}(k)| \leq T \\ [|Y_{\text{SEG}}(k)| - T]e^{j\varphi\{Y_{\text{SEG}}(k)\}} & \text{pro } |Y_{\text{SEG}}(k)| > T \end{cases} \quad (8.8)$$

$$\delta_{\text{halfsoft}}\{Y_{\text{SEG}}(k)\} = \begin{cases} 0 & \text{pro } |Y_{\text{SEG}}(k)| \leq T_1 \\ \left[\frac{T_2(|Y_{\text{SEG}}(k)| - T_1)}{T_2 - T_1} \right] e^{j\varphi\{Y_{\text{SEG}}(k)\}} & \text{pro } T_1 < |Y_{\text{SEG}}(k)| \leq T_2 \\ Y_{\text{SEG}}(k) & \text{pro } |Y_{\text{SEG}}(k)| > T_2 \end{cases} \quad (8.9)$$

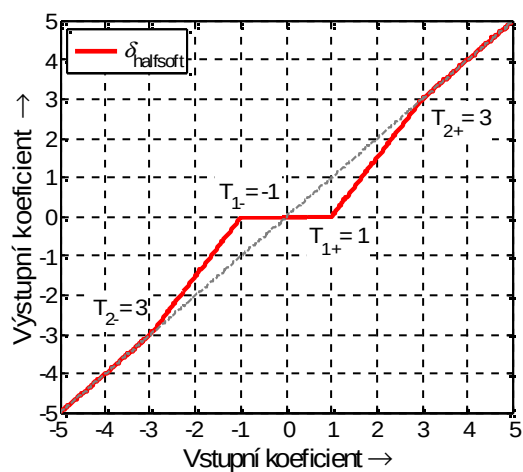
Na následujících obrázcích Obr. 8.11a-d) jsou tato pravidla zobrazena graficky ve formě přenosových funkcí mezi velikostí vstupního a výstupního koeficientu. Symbol T je zde použit pro vyjádření prahu v těchto prahovacích pravidlech. Pokud má některé prahovací pravidlo více prahů tak jsou značeny číselnými indexy, např. T_1 a T_2 u pravidla $\delta_{\text{halfsoft}}\{\}$. Prahy prahovacích pravidel jsou po celý průběh prahování konstantní a proto nejsou značeny kurzívou. Statistické analýze prahovacích pravidel, konkrétně jejich rizikových funkcí, se věnuje například disertační práce [54]. Dále se prahovacími pravidly pro potřeby prahování koeficientů vlnkové transformace zabývá monografie [35]. Vzájemné souvislosti mezi technikami prahování DFT koeficientů, technikami spektrálního odečítání a krátkodobým Wienerovým filtrem jsou detailněji popsány v [5][6][8][10].



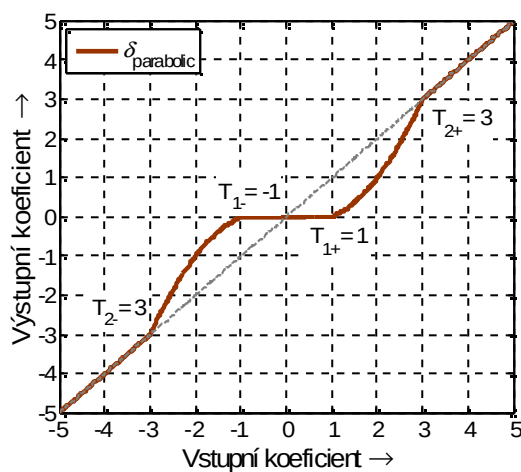
a)



b)



c)



d)

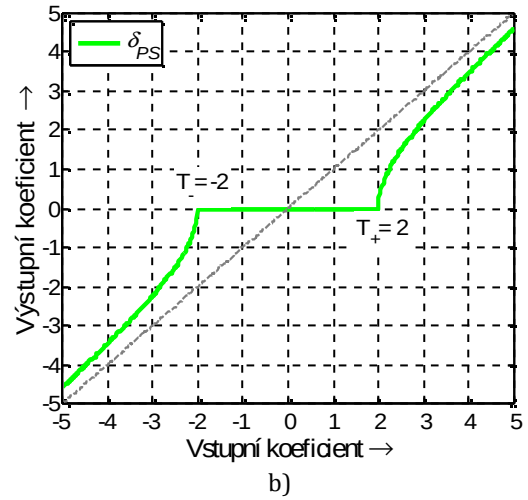
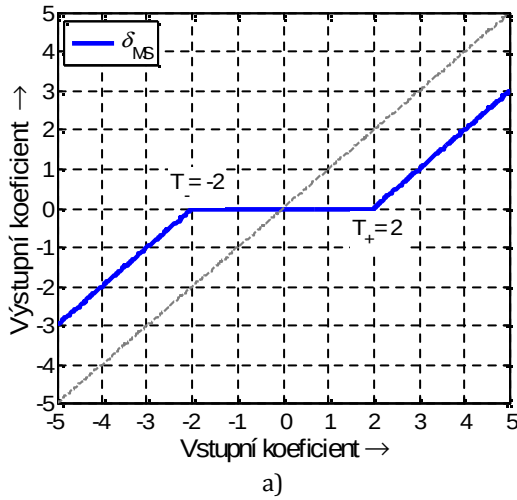
Obr. 8.11 Prahovací pravidla: a) měkké prahovací pravidlo, b) tvrdé prahovací pravidlo, c) poloměkké prahovací pravidlo, d) parabolické prahovací pravidlo [6][9].

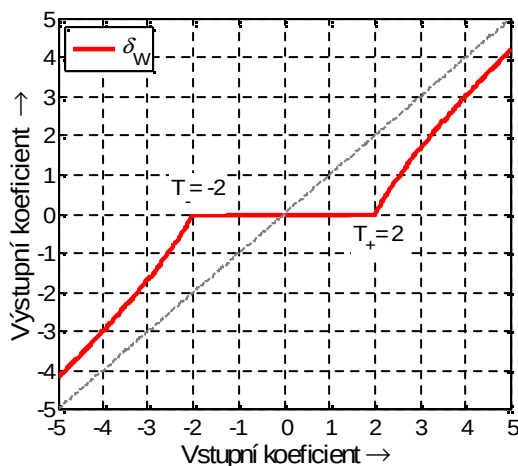
8.2.2 Souvislost prahovacích pravidel se ziskovými funkcemi

V této kapitole bude poukázáno na vzájemnou podobnost ziskových funkcí zavedených v kapitole 6 a prahovacích pravidel zavedených v předchozí kapitole. Pokud například bude velikost prahu $T = |M_{SEG}(k)|$ a pokud platí podmínka $|Y_{SEG}(k)| > |M_{SEG}(k)|$, tak lze psát

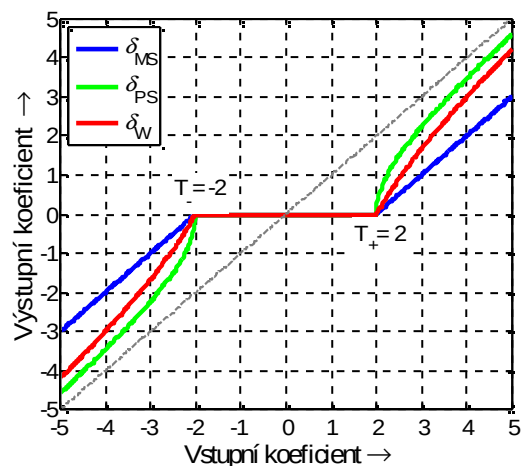
$$\begin{aligned} \frac{\delta_{\text{hard}}\{Y_{SEG}(k)\}}{Y_{SEG}(k)} &= \frac{[|Y_{SEG}(k)| - |M_{SEG}(k)|]e^{j\varphi\{Y_{SEG}(k)\}}}{Y_{SEG}(k)} = \\ &= \frac{|Y_{SEG}(k)|e^{j\varphi\{Y_{SEG}(k)\}}}{|Y_{SEG}(k)|e^{j\varphi\{Y_{SEG}(k)\}}} - \frac{|M_{SEG}(k)|e^{j\varphi\{Y_{SEG}(k)\}}}{|Y_{SEG}(k)|e^{j\varphi\{Y_{SEG}(k)\}}} = \\ &= 1 - \frac{|M_{SEG}(k)|}{|Y_{SEG}(k)|} = H_{MS}(k), \end{aligned} \quad (8.10)$$

Tedy tvrdé prahování při správně volbě prahu přechází až na člen $\frac{1}{Y_{SEG}(k)}$ v amplitudové spektrální odečítání. Způsobem podobný (8.6) jdou odvodit prahovací pravidla ze všech ziskových funkcí definované vztahy (6.24)(6.26)(6.28). Na následujících obrázcích Obr. 8.12a-b) jsou tyto prahovací funkce zobrazeny graficky.





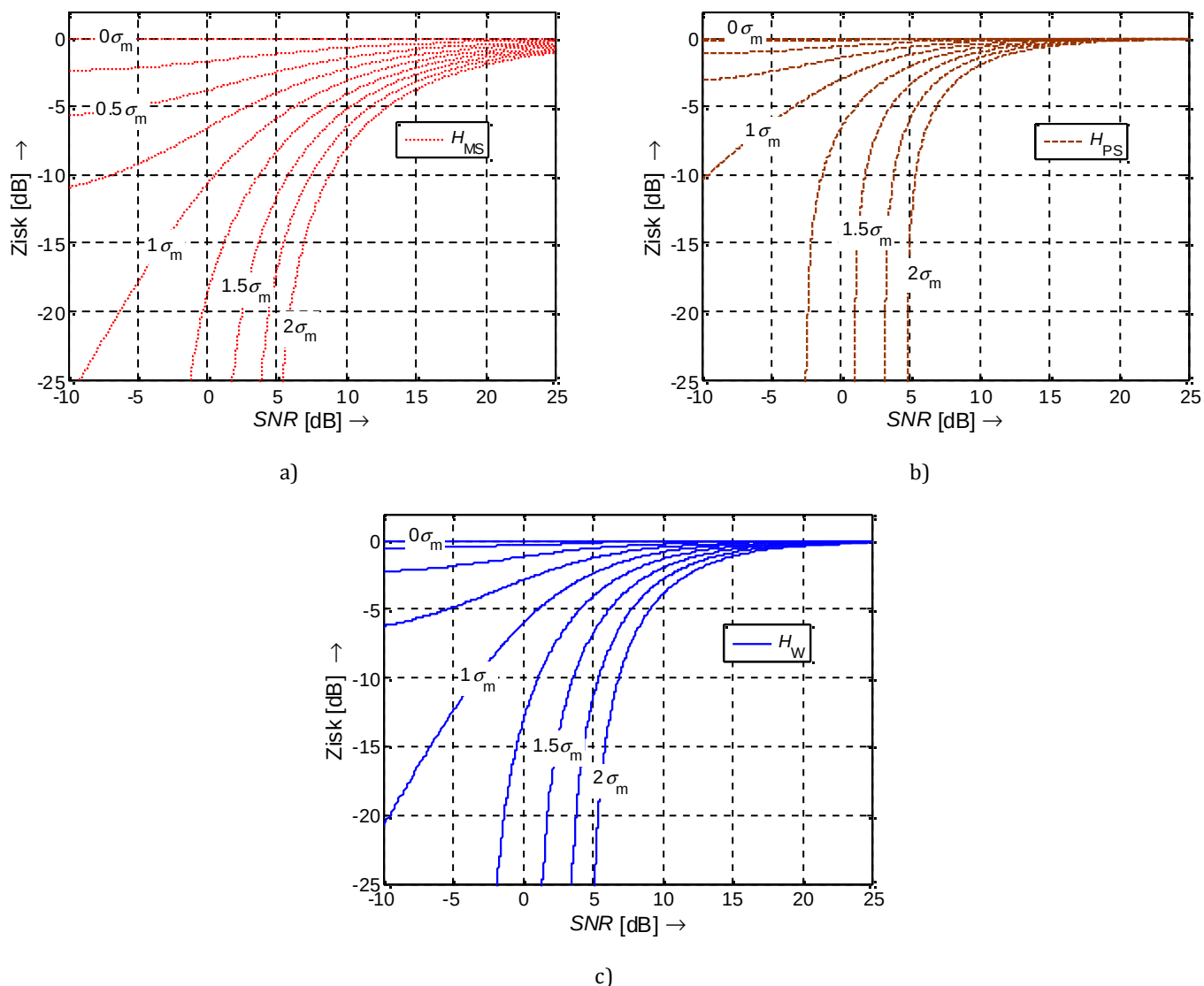
c)



d)

Obr. 8.12 Práhovací pravidla základních ziskových funkcí: a) amplitudové spektrální odečítání, b) výkonové spektrální odečítání, c) krátkodobý Wienerův filtr, d) všechny ziskové funkce.

Vliv změny velikosti prahu T v ziskových funkcích je graficky zobrazen na obrázcích Obr. 8.13a-c). Práh T je zde reprezentován násobkem směrodatné odchylky σ_m , kde σ_m je směrodatná odchylka signálu $m(n)$. Signály $x(n)$ a $m(n)$ jsou v tomto případě bílé šумы s rozptyly σ_x^2 a σ_m^2 a SNR je definováno jako $SNR = \sigma_x^2 / \sigma_m^2$, viz stejně jako v grafu na Obr. 6.3. Pokud $\sigma_m = 1$ pak jsou tyto ziskové funkce stejné jako v grafu na Obr. 6.3. Lze říci, že s rostoucí hodnotou násobku σ_m roste útlum všech ziskových funkcí a se snižující se hodnotou tohoto násobku naopak útlum klesá, až pro $0 \cdot \sigma_m$ má hodnotu 0 dB, tedy jednotkový přenos.



Obr. 8.13 Průběhy ziskových funkcí pro rozdílné násobky směrodatné odchylky σ_m : a) amplitudové spektrální odečítání, b) výkonové spektrální odečítání, c) krátkodobý Wienerův filtr.

8.3 Poslechový experiment

8.3.1 Popis poslechového experimentu

Cílem tohoto experimentu je zjistit, zda navržený systém pro potlačení šumu na pozadí adaptivní segmentací dosahuje lepších výsledků, než stávající systémy. Všechny testované systémy využívají techniky prahování koeficientů DFT (8.6). Konkrétně bude v tomto experimentu zkoumáno, zda se typ použité segmentace nějak projeví na výsledné subjektivně vnímané kvalitě výstupních signálů. Výstupními signály jsou myšleny signály s potlačeným aditivním šumem, tj. z výstupu systému na potlačení šumu. Zkoumanými typy segmentací jsou:

- **Metoda 1, pevná segmentace.**

- Definována vztahem (7.1) s hodnotami $p = 1$ a $q = 1$. Délka okna je vždy konstantní s délkou $N = 1024$ vzorků s překryvem 50 %.

- **Metoda 2, adaptivní segmentace podle algoritmu Greedy selection.**
 - Definována vztahem (7.1) s hodnotami p a q nastavenými podle algoritmu Greedy selection, viz Obr. 7.10. Délka okna je proměnná a platí $L \geq N$, kde $N = 512$ vzorků. Velikost překryvu je vždy 256 vzorků.
- **Metoda 3, adaptivní segmentace podle algoritmu Maximum greedy selection.**
 - Definována vztahem (7.1) s hodnotami p a q nastavenými podle algoritmu Maximum greedy selection, viz Obr. 8.6. Délka okna je proměnná a platí $L \geq N$, kde $N = 512$ vzorků. Velikost překryvu je vždy 256 vzorků.

Všechny tři metody shodně používají měkké prahování, velikost prahu je rovna odhadnuté směrodatné odchylce $\sigma_m(k)$ pro jednotlivé k přes segmenty se stejnou délkou L . Tento odhad je proveden ve fázi analýzy šum pozadí. Délka maximální délka adaptivního okna L není nijak omezena. Ve vztahu (7.2) bylo použito Hannovo okno (7.3).

Tyto tři metody budou testovány na připravené sadě zarušených zvukových ukázek. Tato sada ukázek se skládá z kombinací užitečných signálů popředí $x(n)$ a signálů šumu pozadí $m(n)$. Vstupní signál $y(n)$ je dán součtem těchto signálů pro různé hodnoty SNR (5.25). Rozsah hodnot SNR je 10, 20 a 30 dB. Z výstupu jednotlivých metod jsou získány výstupní signály. Tyto signály mají potlačenou šumovou složku. Vlivem jednotlivých metod je do užitečných signálů velmi často zaveden nějaký typ zkreslení. Odhad užitečného signálu $x(n)$ ze zarušeného signálu $y(n)$ je značen jako $\hat{x}(n)$, symbol stříšky je zde použit pro vyjádření odhadu. V takovém případě lze pro lineární třídu systémů psát:

$$\hat{x}(n) = m_{\text{red}}(n) + x_{\text{red}}(n) \quad (8.11)$$

kde $m_{\text{red}}(n)$ je redukovaný (tj. potlačený) šum a $x_{\text{red}}(n)$ je zkreslený užitečný signál $x(n)$. Uvažujeme zde, že systémy na potlačení šumu ovlivní obě složky tj. $x(n)$ a $m(n)$, filtrací vstupního, tj. šumem zarušeného, signálu $y(n)$. V ideálním případě je $m_{\text{red}}(n) \rightarrow 0$, $x_{\text{red}}(n) \rightarrow x(n)$ a proto $\hat{x}(n) \rightarrow x(n)$, tedy systémy na potlačení šumu zcela odstraní šum a nezavádí žádné lineární zkreslení do užitečné složky signálů. Takováto situace nastává ale zcela výjimečně, např. v úsecích bez řečové aktivity u řečových signálů a podobně.

Vliv zkreslení užitečného signálů na výslednou subjektivní kvalitu je velmi významný. Často bývá významnější než vliv přítomnosti šumu pozadí na tuto kvalitu [26]. Z těchto důvodů bude výsledná kvalita výstupních signálů $\hat{x}(n)$ hodnocena jako celek a ne pouze vzhledem k celkovému potlačení šumové složky $m_{\text{red}}(n)$. Typickým příkladem takového chybného přístupu je měření zlepšené hodnoty SNR , tj. měření rozdílu hodnot SNR před a po zpracování systémem na potlačení šumu [3][78].

8.3.2 Příprava a průběh poslechového experimentu

V tomto poslechovém experimentu posluchači hodnotili výslednou kvalitu předložených zvukových ukázek. Tyto ukázky vznikly z výstupních signálů z metod 1 až 3. Posluchačům byly tedy předloženy 3 ukázky a každá zvuková ukázka reprezentovala jednu metodu. Tyto zvukové ukázky byly předkládány v náhodném pořadí. Dále posluchači neměli informaci, která ukázka přísluší jaké metodě. Cílem posluchačů bylo preferovat jednu, popřípadě více ukázek, které se jim jevily jako subjektivně nejkvalitnější. Posluchač mohl preferovat více ukázek v případě, kdy nedokázal rozlišit

mezi předloženými zvukovými ukázkami. Například v případě, kdy se mu všechny tři ukázky zdály stejné. Podobný typ poslechového experimentu byl například realizován v [64].

V experimentu byly použity zvukové ukázky označené jako “Abba”, “Mozart” a “Hra na kytaru” z databáze SQAM [72]. Každá z těchto ukázek byla aditivně zarušena (8.2) šumy “šum audio kazety” a “šum větráků”, při hodnotách $SNR = 10, 20$ a 30 dB (5.25). Spektrogramy, časové průběhy a odhady výkonových spektrálních hustot signálů z těchto zvukových ukázek jsou uvedeny v příloze B. Celkem tedy bylo posluchačům předloženo 18 sad po třech zvukových ukázkách, reprezentující každou z testovaných metod. Těchto 18 sad bylo předloženo posluchačům vždy dvakrát. Poprvé v tzv. *testu* a podruhé v tzv. *retestu*. Mezi testem a retestem byla krátká pauza.

Měření probíhalo na pracovišti vybavené kvalitními studiovými sluchátky a kvalitní externí zvukovou kartou, viz Obr. 5.16. Změřená přenosová charakteristika celé této audio soustavy je zobrazena na Obr. 5.17. Poslechového experimentu se zúčastnilo celkem 10 posluchačů, ve věku 20 až 30 let. Popis pracoviště lze nalézt v kapitole 5.3.3.

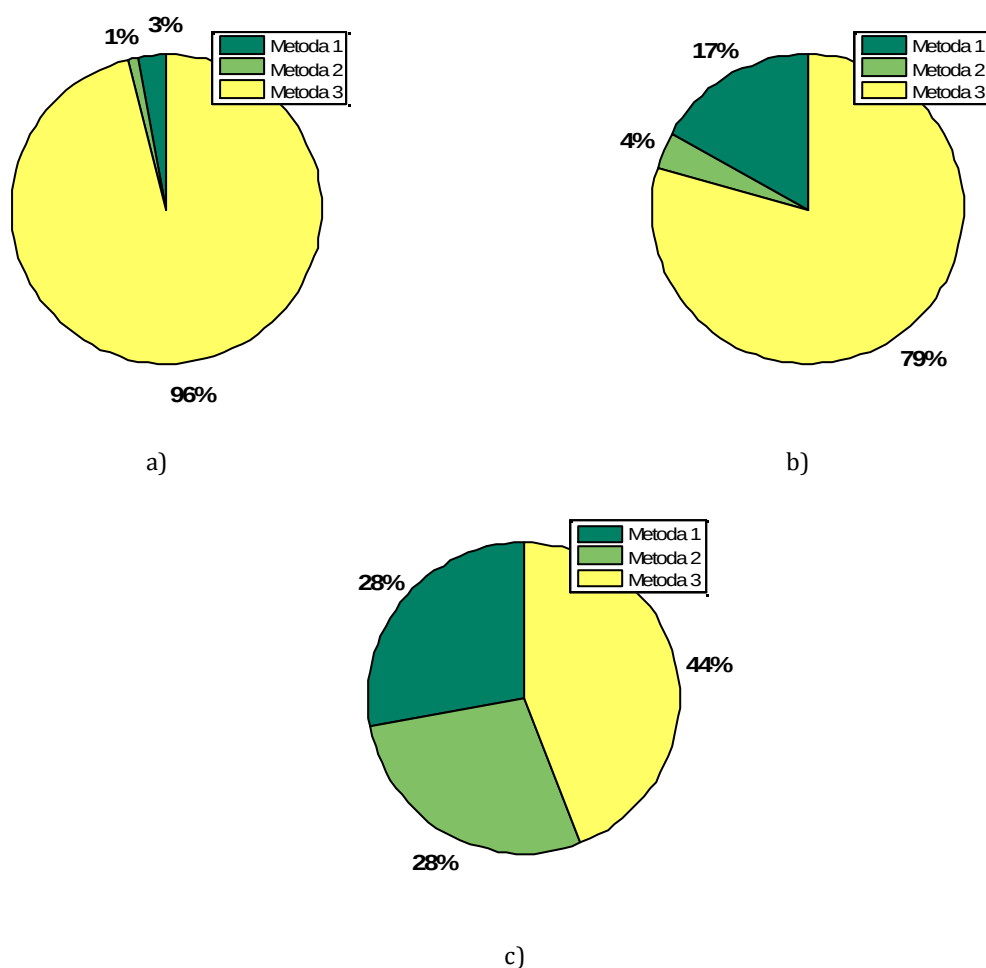
8.4 Výsledky provedeného poslechového experimentu

Tabulku s výsledným preferenčním hodnocením lze nalézt v Tab. 8.1. V jednotlivých sloupcích jsou vypsané procentuální preference, které odpovídají jednotlivým metodám a na řádcích jsou vyneseny tyto preference v závislosti na velikosti SNR . Lze vidět, že pro hodnotu $SNR = 10$ dB bylo preferenční hodnocení nejednoznačnější. Posluchači zde v 96% ohodnotili jako subjektivně nej kvalitnější zvukové ukázky ty ukázky, kde byl šum pozadí potlačen metodou 3. Tedy metodou kde byl použit nový algoritmus Maximum greedy selection. Pro hodnotu $SNR = 20$ dB už toto preferenční hodnocení kleslo z 96% na 79 %. Po poslední hodnotu $SNR = 30$ dB rozdíly mezi jednotlivými metodami nejsou příliš velké. To vychází ze skutečnosti, že posluchači již nedokázali při takto nízkém stupni zarušení rozlišit mezi testovanými zvukovými ukázkami v experimentu.

Tab. 8.1 Tabulka s výsledným preferenčním hodnocením pro jednotlivé metody.

	Metoda 1 (pevná segmentace)	Metoda 2 (Greedy selection)	Metoda 3 (Maximum greedy selection)
$SNR = 10$ dB	3%	1%	96%
$SNR = 20$ dB	17%	4%	79%
$SNR = 30$ dB	28%	28%	44%

Na následujících obrázcích Obr. 8.14a-c) jsou přehledně v koláčových grafech graficky vyneseny procentuální preferenční hodnocení pro rozdílné hodnoty SNR z Tab. 8.1.



Obr. 8.14 Koláčové grafy rozložení preferencí jednotlivých metod: a) $SNR = 10$ dB, b) $SNR = 20$ dB, $SNR = 30$ dB.

Detailnější tabulku s preferenčním hodnocením pro jednotlivé metody lze nalézt v Tab. 8.2. V této tabulce jsou navíc oproti Tab. 8.1 uvedeny preferenční hodnocení pro jednotlivé kombinace užitečného signálu a šumu pozadí včetně dělení na měření v testu a retestu. Kompletní výsledky z tohoto poslechového experimentu pak lze nalézt v příloze E.

Tab. 8.2 Detailní tabulka s preferenčním hodnocením pro jednotlivé metody.

		Kytara						Mozart						Abba					
		Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
		M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
		Preference [%]			Preference [%]			Preference [%]			Preference [%]			Preference [%]			Preference [%]		
SNR 10 dB	Test	0	5	95	0	0	100	0	0	100	0	0	100	0	10	90	0	0	100
	Retest	25	0	75	15	0	85	0	0	100	0	0	100	0	0	100	0	0	100
	Celkem	13	3	85	8	0	93	0	0	100	0	0	100	0	5	95	0	0	100
SNR 20 dB	Test	3	8	88	33	13	53	30	0	70	15	0	85	3	13	83	30	0	70
	Retest	20	10	70	20	0	80	18	3	88	15	0	85	0	0	100	10	0	90
	Celkem	12	9	79	27	7	67	24	2	79	15	0	85	2	7	92	20	0	80
SNR 30 dB	Test	23	33	43	22	27	52	27	17	57	35	30	35	30	25	45	23	18	58
	Retest	20	35	45	37	37	27	37	27	37	27	32	42	25	35	40	32	22	47
	Celkem	22	34	44	29	32	39	32	22	47	31	31	38	28	30	43	28	20	53

9 Závěr

Tuto disertační práci lze tematicky rozdělit do dvou vzájemně zdánlivě samostatných částí. První část práce se zabývá objektivním měřením šumu na pozadí v hudebních signálech a druhá část této práce se zabývá jeho potlačováním. Po úvodních teoretické kapitole č. 2, která si kladla za cíl analyzovat současný stav této problematiky, a tedy splnit 1. cíl disertační práce, následuje popis vlastního řešení objektivní metody pro měření rušivého šumu na pozadí, viz kapitola č. 5. Metoda navržená v této kapitole se zaměřuje na velmi malé zkreslení šumem na pozadí, tedy na ty případy, kdy tento šum je na hranici vnímatelnosti posluchači. Šum pozadí je zde slyšitelný jen v některých částech hudebního signálu, např. v tichých pasážích, na vyšších frekvencích a podobně.

Typickou oblastí použití této metody je analýza šumů a hluků na pozadí při restaurování starších hudebních záznamů [40]. Tato metoda poskytuje prostředky k lokalizaci slyšitelného šumu na pozadí v časo-frekvenční oblasti a k objektivnímu stanovení jeho velikosti.

Protože se však tato metoda nezaměřuje pouze na řečové a hudební signály, ale také na různé umělé typy zvukových signálů (např. bílý šum, harmonický signál, směs harmonických signálů atd.), tak lze tuto metodou využít i k odhadu slyšitelnosti jednoho umělého zvukového signálu na pozadí jiného, opět umělého zvukového signálu. Tato potřeba nastává například při volbě "optimálních" parametrů pro syntézu pásmově neomezených signálů typu píla nebo obdélník v digitálních syntezátorech u digitálních hudebních nástrojů [67][68][69]. Těmito optimálními parametry je myšleno nalezení kompromisu mezi výpočetní náročností algoritmů pro syntézu těchto typů signálů a jejich výslednou subjektivně vnímanou kvalitou. Kapitola č. 5, která popisuje realizaci této objektivní metody, splňuje zadání stanovené cílem č. 2 v cílech disertační práce.

V další části této kapitoly se výklad zaměřuje na ověření funkčnosti a správnosti navržené objektivní metody. Tato kapitola sleduje zadání č. 3 v cílech této disertační práce. V této části kapitoly byla testována míra predikce slyšitelného šumu na pozadí navrženou metodou. Nejprve byl proveden poslechový experiment a získaná data z tohoto experimentu byla využita ke stanovení míry úspěšnosti této predikce. Míra úspěšnosti této predikce byla následně porovnána s mírou úspěšnosti predikce slyšitelného šumu pozadí metodou PEAQ, viz kapitola č. 4. Metoda PEAQ je v dnešní době standardní objektivní metodou pro evaluaci kvality zvukových signálů [57]. Toto testování proběhlo nejprve na hudebních a řečových signálech a po té i na umělých zvukových signálech. Hudebními a řečovými signály byly vhodně vybrané záznamy hudebních nástrojů, hudby, zpěvu, řeči a také reálně se vyskytující typy šumů, jako je například šum z magnetické pásky, praskání vinylové desky a podobně. Naproti tomu umělé typy zvukových signálů obsahovaly signály typu bílý šum, růžový šum, směs harmonických tónů atd. Ze získaných dat z provedeného experimentu bylo zjištěno, že obě metody lze použít pro predikci slyšitelnosti šumu pozadí pro případ řečových nebo hudebních signálů. Při vzájemném porovnání ale navržena metoda z kapitoly č. 5 dosahuje lepších výsledků. U hudebních a řečových signálů šlo použít obě metody, ale u umělých zvukových signálů toto již neplatí. Metoda PEAQ zde selhává. Na rozdíl od metody PEAQ, lze navrženou metodu z kapitoly 5 lze využít i pro tuto třídu signálů.

Přínos této navržené metody spočívá především v možnosti objektivně měřit slyšitelný šum pozadí přesnějším způsobem, než který dosahují v současnosti používané metody. Dále je možno

provádět měření slyšitelnosti jednoho zvukového signálu na pozadí druhého a to i pro případ různých umělých zvukových signálů. Toho nebylo možno dosáhnout metodou PEAQ. Nyní je možno pro tyto případy využít navrženou metodu namísto realizace časově i finančně náročných poslechových experimentů. Další práce na této metodě bude spočívat v zavedení nové hodnotící škály. Tato škála bude použita k evaluaci míry (velikosti) slyšitelného šumu na pozadí. Navržená metoda totiž v současné době používá velmi jednoduchou hodnotící škálu. Tato škála se omezuje pouze na dvě úrovně hodnocení, které by se daly slovně charakterizovat jako “šum pozadí je slyšitelný” a “šum pozadí není slyšitelný”. Pokud je šum pozadí slyšitelný, tak nelze stanovit jeho subjektivní velikost, tedy určit jak moc se posluchačům jeví hlasitě. Zavedením nové hodnotící škály s více než dvěma stupni hodnocení by toto ale bylo možné.

Druhá část disertační práce se zabývala návrhem systému na potlačení šumu na pozadí v hudebních signálech. Tento systém měl být navržen s ohledem na psychoakustické aspekty vnímání tohoto typu signálů posluchači. Navržený systém je charakteristický zejména tím, že k samotnému potlačení šumu využívá techniku prahování koeficientů DFT (Diskrétní Fourierova transformace) a adaptivní segmentační schéma.

Techniky prahování totiž dovolují vyšší stupeň volnosti, při návrhu a obsluze systému pro potlačení šumu pozadí. Oproti parametrickému Wienerovu filtru, který byl zaveden v kapitole č. 6, a který integruje krátkodobý Wienerův filtr a techniky spektrálního odečítání, lze technikami prahování navíc také regulovat velikost útlumu a tvar přenosové charakteristiky restauračního filtru. Tento vyšší stupeň volnosti na jednu stranu vyžaduje nutnost nastavení většího množství parametrů, ale na stranu druhou umožňuje dosáhnout lépe optimalizovaný restaurační filtr pro konkrétní sadu restaurovaných hudebních záznamů. Kritérium optimality zde není striktně definováno, jako je například u Wienerova filtru, ale “volně” a je pouze na operátorovi takového systému, aby vhodně měnil tyto parametry s cílem dosáhnout maximální subjektivní kvality výsledných hudebních záznamů. Diskrétní Fourierova transformace (DFT) byla zvolena z toho důvodů, že básové funkce DFT jsou komplexní harmonické funkce a proto jsou vhodné pro analýzu hudebních signálů. Tyto hudební signály se právě vyznačují výraznou harmonickou strukturou. Další výhodou je možnost rychlé implementace DFT pomocí algoritmu FFT. Mezi další důvody patří skutečnost, že DFT je základním typem transformace, na kterém staví další typy transformací, které se často používají při zpracování číslicových zvukových signálů. Jako příklad takovýchto transformací lze uvést DCT nebo MDCT. Přechod z DFT na tyto transformace je v navrženém systému pro potlačení šumu na pozadí bez komplikací.

Adaptivní segmentační schéma řeší problém segmentace dlouhotrvajících a nestacionárních hudebních signálů. V kapitole č. 7 bylo ukázáno, že pro tuto skupinu signálů je velmi obtížné najít jedinou konstantní délku segmentu a to pro celý analyzovaný dlouhotrvající signál. U hudebních signálů lze relativně často nalézt pasáže, ve kterých se střídají velmi dlouhé úseky, které mohou být považovány za stacionární a úseky s krátkou dobou trvání, které jsou nestacionární. Dále byly v této kapitole ukázány výhody využití adaptivního segmentačního schématu pro adaptivní segmentaci. Zejména se jedná o dynamické změny časo-frekvenčního rozlišení a to v závislosti na míře stacionarity analyzovaného signálu.

Vlastní přínos při návrhu tohoto systému na potlačení šumu na pozadí v hudebních signálech spočíval především v modifikaci algoritmu pro adaptivní segmentaci *Greedy selection*, viz kapitola č. 7. Byly zavedeny dvě úpravy tohoto algoritmu s cílem zlepšit jeho vlastnosti u

segmentace hudebních signálů. Konkrétně se jedná o započítání vlivu rozdílného kmitočtového rozlišení každého segmentu do rozhodovacích struktur tohoto algoritmu a také snížení vlivu komponent náhodného charakteru na tyto rozhodovací procesy. Tyto modifikace byly detailně popsány v kapitole č. 8. Náplň této kapitoly pojednávající o vlastním řešení systému na potlačení aditivního šumu pozadí v hudebních signálech s využitím adaptivního segmentačního schématu splňuje zadání č. 4, které bylo vytyčeno v zadání cílů disertační práce.

Poslední část kapitoly č. 8 se zabývá výsledným zhodnocením navržené metody a porovnáním provedených modifikací s původní variantou algoritmu Greedy selection a klasickou (fixní) segmentací. Porovnáván je zde vliv rozdílných způsobů segmentací pro potřeby prahování koeficientů DFT na subjektivně vnímanou kvalitu výstupních hudebních signálů. Výstupními signály jsou myšleny ty signály, které vystupují ze systému na potlačení šumu pozadí. Toto porovnávání probíhalo formou poslechových testů, kde posluchači preferenčním hodnocením hodnotili jednotlivé výstupní signály. Z výsledků tohoto experimentu vyplývá, že uvedené modifikace algoritmu Greedy selection zvyšují pro vyšší stupeň zarušení poslechovou kvalitu těchto výstupních signálů. Posluchači tyto signály v průměru více preferovali než ostatní testované signály. Pro nižší stupeň zarušení, tedy pro hodnoty $SNR = 30$ dB a více, nelze výsledky provedeného poslechového testu takto jednoznačně interpretovat, protože vlivem nízkého stupně zarušení posluchači nebyli schopni vzájemně rozlišit testované výstupní signály. Tato poslední část kapitoly č. 8, která se zabývá zhodnocením navržené metody, splňuje poslední z vytyčených cílů zadání této disertační práce.

10 Literatura

Vybrané publikace autora

- [1] RÁŠO, O.; BALÍK, M. Comparison of Optimum Filter Length in Linear Prediction between Speech and Musical Signals. *In 34th International Conference on Telecommunications and Signal Processing*. 2011. s. 1-6. ISBN: 978-1-4577-1409- 2.
- [2] RÁŠO, O.; BALÍK, M. Optimum Filter Length in Linear Prediction of Musical Signals with Pop Music Content. *In The 13th International Conference on Research in Telecommunication Technologies RTT*. 2011. s. 1-4. ISBN: 978-80-214-4283- 2.
- [3] RÁŠO, O.; BALÍK, M.; HALAS, M. New Approach to Quality Measuring of Noise Reduction System. *In 6th International Conference on Teleinformatics*. 2011. s. 203-208. ISBN: 978-80-214-4231- 3.
- [4] RÁŠO, O.; BALÍK, M.; KOUŘIL, J. Origin and Suppression of Musical Tonal Noise in Thresholding Methods. *In 33rd International Conference on Telecommunications and Signal Processing*. 2010. s. 84-89. ISBN: 978-963-88981-0- 4.
- [5] RÁŠO, O.; BALÍK, M. Relation of Wiener Filter and Spectral Subtraction with Thresholding Methods. *In 33rd International Conference on Telecommunications and Signal Processing*. 2010. s. 78-83. ISBN: 978-963-88981-0- 4.
- [6] RÁŠO, O.; BALÍK, M. Adaptive Thresholding of DFT Coefficients based on Probability Distribution of Additive Noise. *WSEAS Transactions on Signal Processing*, 2010, roč. 2009, č. 12, s. 390-399. ISSN: 1790- 5052.
- [7] RÁŠO, O.; BALÍK, M. Method for Suppression of Musical Tonal Noise in Audio Signals. *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), 2010, roč. 2010, č. 11, s. 1-6. ISSN: 1213- 1539.
- [8] RÁŠO, O.; BALÍK, M. Special Types of Thresholding Rules Based on Wiener Filer and Spectral Subtraction. *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), 2010, roč. 2010, č. 11, s. 1-6. ISSN: 1213- 1539.
- [9] RÁŠO, O.; BALÍK, M. Half- soft parabolic thresholding of DFT coefficients. *In 32nd International Conference on Telecommunications and Signal Processing - TSP'2009*. Dunakiliti, Hungary: Asszisztencia Szervezo Kft., 2009. s. 1-6. ISBN: 978-963-06-7716- 5.
- [10] RÁŠO, O.; BALÍK, M. Polynomial thresholding based on probability distribution of noise signal. *In RECENT ADVANCES in CIRCUITS, SYSTEMS, ELECTRONICS, CONTROL and SIGNAL PROCESSING. WSEAS Applied Informatics & Communications*. Puerto de la Cruz, Spain: WSEAS Press, 2009. s. 314-319. ISBN: 978-960-474-139- 7. ISSN: 1790- 5117.
- [11] RÁŠO, O.; BALÍK, M.; KOVAŘÍK, M. Subjective Evaluation of Noise Pleasantness in Audio Systems. *In Research in Telecommunication Technologies 2012*. Žilina, Slovenska Republika: 2012. s. 1-5. ISBN: 978-80-554-0570- 4.
- [12] RÁŠO, O.; BALÍK, M. Educational Software for Optimal Order Determination of LP Speech Modeling in Telecommunication. *Proceedings of the 16th WSEAS International Conference on Computers* (part of CSCC 12). 2012. s. 106-109. ISBN: 978-1-61804-109- 8.
- [13] BALÍK, M.; RÁŠO, O. Sound artifacts caused by the thresholding method. *In Proceedings of the IASTED International Conference pn Modelling, Simulation, and Identification*. ACTA Press, B6, Suite #101, 2509 Dieppe Ave SW, Calgary, Alberta T3E 7J9, Canada: ACTA Press, 2009. s. 1-7. ISBN: 978-0-88986-810- 6.

Použité zdroje cizích autorů

- [14] AVENDANO, C.: *Temporal Processing of Speech in a Time-Feature Space*. Disertační práce, Oregon Graduate Institute, 1997. 118 s.
- [15] BOLL, S.F.: Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Tran. on Acoustic, Speech and Signal Processing ASSP-27*, s.. 113-120 1979
- [16] BRANDENBURG, K., Evaluation of quality for audio encoding at low bit rates. *Contribution to the 82nd AES Convention*, London, 1987. London.
- [17] ČSN 01 1600 *Akustika – Terminologie*. Český normalizační institut, Praha 2003. 80 s.
- [18] ČSN ISO 532 (011602), *Metoda pro výpočet hlasitosti*, Český normalizační institut, Praha, brazen 1994. 24 s.
- [19] DIETHORN, E. J.; Subband noise reduction methods for speech enhancement, *Acoustic signal processing for telecommunication*, s. 159-179, Kluwer Academic Publishers Hingham, MA, USA, 2000, ISBN:0-7923-7814-8
- [20] DINIZ, P.S.R.: *Adaptive filtering: Algorithms and practical implementation*. 3th edition, Springer, New York 2008, 627 s. ISBN:978-0-387-31274-3
- [21] EDWARDS, A. L.; Experiments , their planning and execution. *Handbook of social psychology*. Addison-Wesley, Reading, Mass 1954.
- [22] ETTER, W.; MOSCHYTZ, G. S.; Noise Reduction by Noise-Adaptive Spectral Magnitude Expansion, *J. Audio Eng. Soc*, sv. 42, č. 5, s. 341 – 349, 1994,
- [23] FASTL, H., ZWICKER E., *Psychoacoustic: Facts and Models*. 3rd ed. Berlin-Heidelberg-New York: Springer, 2006. 462 s. ISBN 978-3-540-23159-2.
- [24] FERGUSON, G. A.: *Statistical analysis in psychology & education*. McGraw-Hill, New York 1989. 6 vyd. ISBN: 0070204853.
- [25] HÄNSLER, E. S.; SCHMIDT, G.: *Speech and Audio Processing in Adverse Environments*. Berlin-Heidelberg-New York: Springer, 2008. 735 s. ISBN 978-3-540-70601-4.
- [26] HÄNSLER, E., SCHMIDT G., *Topics in Acoustic Echo and Noise Control: Selected Methods for the Cancellation of Acoustic Echoes, the Reduction of Background Noise, and Speech Processing*. Berlin-Heidelberg-New York: Springer, 2006. 642 s. ISBN 10 3-540-33212-x.
- [27] HERMANSKY, H., MORGAN, N.: RASTA Processing of Speech. *IEEE Trans. Speech and Audio Processing*, č. 4, s. 578-589, 1994.
- [28] HU, Y.; LOIZOU, C.; Speech Enhancement Based on Wavelet Thresholding the Multitaper Spectrum, *IEEE Transactions on Speech and Audio Processing*, sv. 12, č. 1, leden 2004. s. 59-67, ISSN: 1063-6676.
- [29] HUANG, N. E.; The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. Londýn*, 454 s., s. 903-993, 1998
- [30] HUBER, R.; Kollmeier, B.; PEMO-Q, *A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception*, Audio, Speech, and Language Processing, IEEE Transactions on, vol.14, no.6, pp.1902-1911, Nov. 2006, ISSN: 1558-7916
- [31] Implementation of ISO/IEC 11172-3, *Information technology: Coding of moving pictures and associated audio for digital storage media at up to 1,5 Mbits/s, Part3: audio*. British standard. BSI, London. říjen 1993.
- [32] Implementation of ISO/IEC 15444-1:2004, *Information technology: JPEG 2000 image coding system: Core coding system*. British standard. BSI, London. 2004.

- [33] ITAKURA, F.; Minimum prediction residual principle applied to speech recognition, *IEEE Transactions on Speech and Audio Processing*, sv. 23, č. 1, květen 1975. s. 67-72, ISSN: 0096-3518.
- [34] JAN, J.; *Číslíková filtrace, analýza a restaurace signálů*. 2nd edition. Brno, Vuitium, 2002. 427 s. ISBN: 80-214-1558-4
- [35] JANSEN, M.: *Noise reduction by wavelet thresholding*. New York : Springer, 2001. 191 s. ISBN 0-387-95244-6.
- [36] JINGDONG, C.; BENESTY, J.; YITENG, H.; DOCLO, S.; New insights into the noise reduction Wiener filter, *IEEE Transactions on Speech and Audio Processing*, sv. 14, č. 4, červenec 2006. s. 1218-1234, ISSN: 1063-6676.
- [37] KABAL, P.: *An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality*, Department of Electrical & Computer Engineering, McGill University, 87 s., 2002 www: <http://WWW.TSP.ECE.McGill.CA>
- [38] KÉPESI M.; NAGY Z.; Potlačení šumu pozadí řečového signálu mapováním spektrogramu, *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), roč. 2000, č. 37, ISSN: 1213-1539.
- [39] KOZUMPLÍK, J., *Vlnkové transformace a jejich využití pro filtraci signálů EKG*. Habilitační práce, Brno University of technology, 2004. 81 s.
- [40] MACH, V. Digitální restaurace záznamů z janáčkovských fonografických válců a jejich kopií. In *Vzaty do fonografu : slovenské a moravské písně v nahrávkách Hynka Bíma, Leoše Janáčka a Františky Kyselkové z let 1909- 1912*. 1. Brno: Etnologický ústav AV ČR, v.v.i., 2012. s. 153-163. ISBN: 978-80-87112-62- 5.
- [41] MACH, V.; Digital Restoration of Recordings from the Phonograph Cylinders and Their Copies. In *As recorded by the phonograph: Slovak and Moravian songs recorded by Hynek Bíma, Leoš Janáček and Františka Kyselková in 1909- 1912*. 1. Brno: Etnologický ústav AV ČR, v.v.i., 2013. s. 165-176. ISBN: 978-80-87112-63- 2.
- [42] MALLAT, S., A., *Wavelet Tour of Signal Processing*. 2nd ed. New York: Academic Press, 1998. 637 s. ISBN 0-12-466606-x.
- [43] MEILI, R. ROHRACHER, H.; *Učebnice experimentální psychologie*. Státní pedagogické nakladatelství, Praha 1969.
- [44] MELKA A., *Základy experimentální psychoakustiky*. 1 vydání. Akademie múzických umění v Praze, 2005. 327 s. ISBN 80-7331-043-0
- [45] NADEU, C.; JUANG, B.H.; Filtering of spectral parameters for speech recognition. *Proceedings of the International Conference on Speech and Language Processing*, 1994.
- [46] NAT, J.; *Restoration Audio with iZotope RX™*, Restoration Guide, 1. verze, 74 s., www: <http://izotope.com/products/audio/rx/support.asp>
- [47] NOLL, P.; Adaptive quantization in speech coding systems, *Proc. Int. Zürich Sem. '74*, 1974.
- [48] NOVÁK, A.; *Příručka základů klinické audiologie*. 170 s. Praha 1998. ISBN 80-238-2090-7
- [49] PETITCOLAS, F.; *MPEQ for Matlab*, software, poslední aktualizace prosinec 2012, www: <http://www.petitcolas.net/fabien/software/mpeg/>
- [50] PLŠEK, M.: *Extrakce řečového signálu z hluku pozadí ve spektrální oblasti*. Disertační práce, Vysoké učení technické v Brně, 2004.
- [51] PLŠEK, M.; KÉPESI M.; Jednokanálová separační technika mapování spektrogramu, *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), roč. 2001, č. 41, ISSN: 1213-1539.

- [52] PROAKIS, J. G., MANOLAKIS D. G., *Digital Signal Processing*. 4th ed. New Jersey: Parson Education, 2007. 1084 s. ISBN 0-13-187374-1.
- [53] QUACKENBUSH, S. R.; BARNWELL, T. P.; CLEMENTS, M. A.; *Objective Measures Of Speech Quality*, Prentice Hall, s. 377, 1988, ISBN-10: 0136290566.
- [54] RAJMIC, P.: *Využití waveletové transformace a matematické statistiky pro separaci signálu a šumu*. Disertační práce, Školitel: Prof. Ing. Zdeněk Smékal, CSc., Vysoké učení technické v Brně, 2004. 106 s.
- [55] Recommendation ITU-R BS.1116, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, 1997.
- [56] Recommendation ITU-R BS.1284, *General methods for the subjective assessment of sound quality*, 2003.
- [57] Recommendation ITU-R BS.1387-1 (PEAQ). *Method for objective measurements of perceived audio quality*, 2001.
- [58] Recommendation ITU-T P.800. *Methods for subjective determination of transmission quality*, 1996.
- [59] Recommendation ITU-T P.800.1. *Mean Opinion Score (MOS) terminology*, 2006.
- [60] Recommendation ITU-T P.861. *Objective quality measurement of telephone-band (300-3400 Hz) speech codecs*, 1996.
- [61] Recommendation ITU-T P.862. *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*, 2001.
- [62] Recommendation ITU-T P.863. *Perceptual objective listening quality assessment*, 2011.
- [63] REISENAUER, R.; *Metody matematické statistiky a jejich aplikace*. 2 vyd. SNTL – Práce, 239 s. Praha 1970.
- [64] RUDOY, D.; BASU, P.; QUATIERI, T. F.; DUNN, B.; WOLFE, P. J.; Adaptive Short-Time Analysis-Synthesis For Speech Enhancement, *International Conference on Acoustics, Speech, and Signal Processing - ICASSP*, s. 4905-4908, 2008
- [65] RUDOY, D.; BASU, P.; WOLFE, P. J.; Superposition Frames for Adaptive Time-Frequency Analysis and Fast Reconstruction, *IEEE Transactions On Signal Processing*, sv. 58, č. 5, Květen 2010, ISSN: 1053-587X
- [66] *RX2™ Tips & Tricks*, Restoration Guide, 20 s., www: <http://www.izotope.com/rx/tips>
- [67] SCHIMMEL, J. Audible Aliasing Distortion in Digital Audio Synthesis. *Radioengineering*, 2012, roč. 21, č. 1, s. 56-62. ISSN: 1210- 2512.
- [68] SCHIMMEL, J. Objective Evaluation of Audible Aliasing Distortion in Digital Audio Synthesis. *In Proceedings of 34th International Conference on Telecommunications and Signal Processing TSP 2011*. 2011. s. 343-347. ISBN: 978-1-4577-1409- 2.
- [69] SCHIMMEL, J., SERRA, J., P. Comparison of Real-Time Syntheses of Band- Limited Periodic Digital Audio Signals. *In Proceedings of 6th International Conference on Teleinformatics ICT 2011*. Brno: VUT v Brně, 2011. s. 197-202. ISBN: 978-80-214-4231- 3.
- [70] SLIM, B., L.; OPPENHEIM, A. V.; Enhancement and bandwidth compression of noisy speech, *Proc. of the IEEE*, sv. 67, prosinec 1979.
- [71] SLIM, B., L.; TONG, Y. C.; CHANG, J. S.; A parametric formulation of the generalized spectral subtraction method, *IEEE Transactions on Speech and Audio Processing*, sv. 6, č. 4, červenec 1998. s. 328-337, ISSN: 1063-6676.
- [72] *Sound Quality Assessment Material recordings for subjective tests*, Users' handbook for the EBU SQAM CD, Geneva, 2008, www: <http://tech.ebu.ch/publications/sqamcd>

- [73] STEVENS, S., S.; Procedure for calculating loudness: Mark VI. *Journal of the Acoustical Society of America*, sv. 33, č. 11, srpen 1961, s. 1577-1585, ISSN: 0001-4966.
- [74] SYROVÝ, V. *Hudební akustika*. Praha : Akademie múzických umění v Praze, 2003. 427 s. ISBN 80-7331-901-2.
- [75] SYROVÝ, V. *Hudební zvuk*. Praha : Akademie múzických umění v Praze, 2009. 303 s. ISBN 80-7331-161-2.
- [76] UNSER, M.; BLU, T.; Mathematical Properties of the JPEG2000 Wavelets of the Summer School ROBUST'96 (J. Antoch and G. Dohnal), s. 1080 – 1090, 2003.
- [77] VASEGHI, S. V.: *Advanced Digital Signal Processing and Noise Reduction*. 2nd edition. New York : Wiley, 2000. 473 s. ISBN 0-471-62692-9.
- [78] VONDRASEK M., POLLAK P., Methods for Speech SNR estimation: Evaluation Tool and Analysis of VAD Dependency, *Radioengineering*, sv.. 14, č. 1, s. 6-11, 2005. ISSN 1210-2512
- [79] WIENER, N.; *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*, Cambridge, Technology Press of Massachusetts Institute of Technology, and New York, Wiley, 1949.
- [80] WONHO Y.; DIXON, M.; YANTORNO, R.; A modified bark spectral distortion measure which uses noise masking threshold, *IEEE Workshop on Speech Coding For Telecommunications Proceeding*, 1997, s.55-56, ISBN: 0-7803-4073-6.
- [81] WONHO, Y.; *Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model*. Disertační práce, Temple University in Philadelphia, květen 1999. 163 s.
- [82] WONHO, Y.; YANTORNO, R.; Improvement of MBSD by scaling noise masking threshold and correlation analysis with MOS difference instead of MOS, *International Conference on Acoustics, Speech, and Signal Processing, ICASSP* 1999. sv. 2, březen 1999. s. 673-676, ISBN: 0-7803-5041-3.
- [83] YANG, W.; BENBOUCHTA, M.; YANTORNO, R.; Performance of a modified bark spectral distortion measure as an objective speech quality measure, *IEEE ICASSP*, s.541-544, Seattle, 1998.
- [84] ZOLZER, U.; DAFX – *Digital Audio Effects*. John Willey & Sons Ltd., 2005, ISBN 0792381300.

Významné v práci necitované publikace autora

- [85] MARTINÁSEK, Z.; MÁCHA, T.; RÁŠO, O.; MARTINÁSEK, J.; ŠILHAVÝ, P. Optimization of differential power analysis. *Przegląd Elektrotechniczny*, 2011, roč. 2011, č. 12, s. 140-144. ISSN: 0033- 2097. IF:0,242.
- [86] RÁŠO, O. Programování funkcí v jazyce symbolických adres podle STDCALL konvence. *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), 2009, roč. 2009, č. 8, s. 1-4. ISSN: 1213- 1539.
- [87] RÁŠO, O.; SOBOTKA, J. Základní techniky skenování TCP portů. *Elektrorevue - Internetový časopis* (<http://www.elektrorevue.cz>), 2010, roč. 2010, č. 11, s. 1-6. ISSN: 1213- 1539.
- [88] SOBOTKA, J.; ZEMAN, V.; RÁŠO, O. Implementation of Universal Authentication Framework in the Electronic Archive Management. In *RECENT RESEARCHES in COMMUNICATIONS and COMPUTERS (Proceedings of the 16th WSEAS International Conference on Communications (part of CCCC 12)*. 2012. s. 110-115. ISBN: 978-1-61804-109- 8.

11 Přílohy k disertační práci

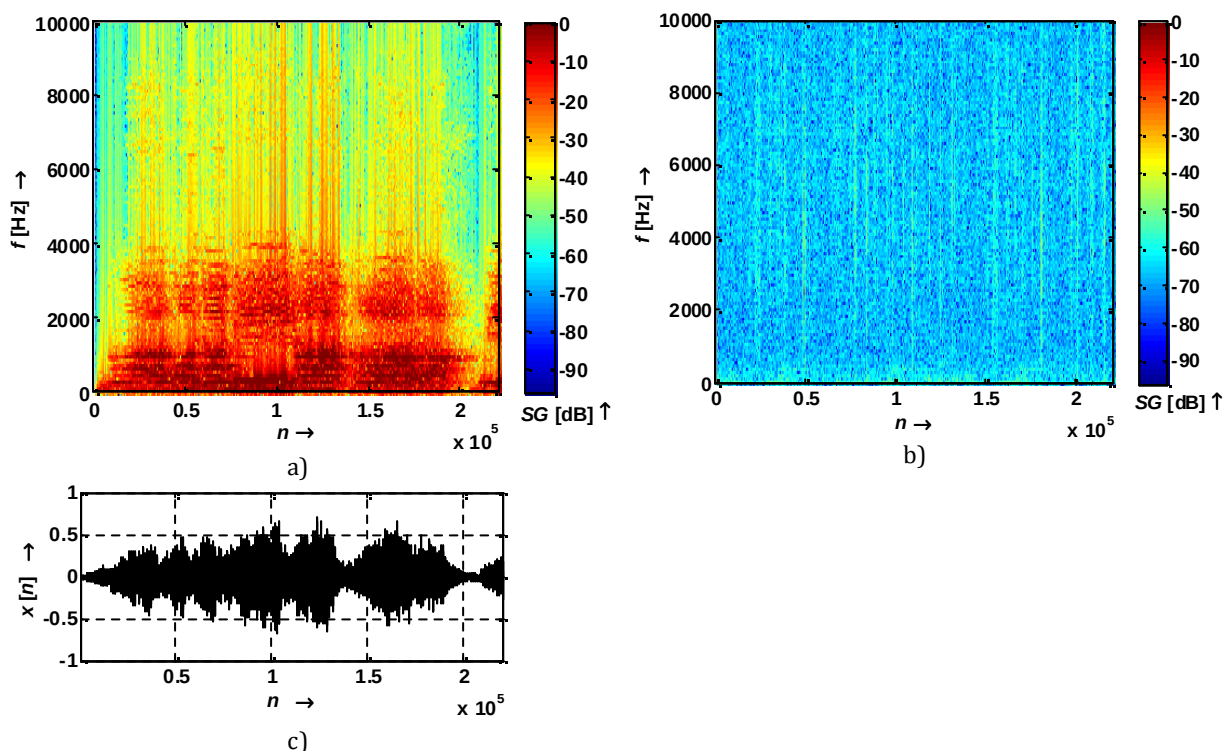
Seznam příloh:

PŘÍLOHA A: DALŠÍ DEMONSTRAČNÍ PŘÍKLADY ADITIVNÍM ŠUMEM DEGRADOVANÝCH ZVUKOVÝCH ZÁZNAMŮ	121
A.1 SBOROVÝ ZPĚVA ŠUM VINYL DESKY PŘI $SNR = 50$ DB.....	121
A.2 ZÁZNAM HRY NA KYTARU A ŠUMU VĚTRÁKU PŘI $SNR = 20$ DB	123
A.3 SKLADBA SKUPINY ABBA A ŠUM VINYL DESKY PŘI $SNR = 50$ DB	125
A.4 COUNTRY UKÁZKA A ŠUM AUDIO PÁSKY PŘI $SNR = 15$ DB.	127
PŘÍLOHA B: POUŽITÝ POSLECHOVÝ MATERIÁL	129
B.1 POSLECHOVÝ MATERIÁL Z 1. ČÁSTI POSLECHOVÉHO EXPERIMENTU, SUBJEKTIVNÍ MĚŘENÍ PRAHU SLYŠITELNOSTI HUDEBNÍCH A ŘEČOVÝCH SIGNÁLŮ.....	129
B.2 POSLECHOVÝ MATERIÁL Z 2. ČÁSTI POSLECHOVÉHO EXPERIMENTU, SUBJEKTIVNÍ MĚŘENÍ PRAHU SLYŠITELNOSTI UMĚLÝCH ZVUKOVÝCH SIGNÁLŮ.....	131
PŘÍLOHA C: STATISTICKÉ VYHODNOCENÍ 1. ČÁSTI POSLECHOVÉHO EXPERIMENTU, PRAHY SLYŠITELNOSTI HUDEBNÍCH A ŘEČOVÝCH SIGNÁLŮ	133
C.1 KRÁTKODOBÁ STABILITA MĚŘENÍ PRO JEDNOTLIVÉ ZVUKOVÉ UKÁZKY.....	134
C.2 ZJIŠTĚNÍ SPOLEHLIVOSTI JEDNOTLIVÝCH POSLUCHAČŮ.....	136
C.3 ZJIŠTĚNÍ SPOLEHLIVOSTI MĚŘENÍ CELÉ SKUPINY.....	136
PŘÍLOHA D: STATISTICKÉ VYHODNOCENÍ 2. ČÁSTI POSLECHOVÉHO EXPERIMENTU, PRAHY SLYŠITELNOSTI UMĚLÝCH ZVUKOVÝCH SIGNÁLŮ.....	140
D.1 KRÁTKODOBÁ STABILITA MĚŘENÍ PRO JEDNOTLIVÉ ZVUKOVÉ UKÁZKY.....	140
D.2 ZJIŠTĚNÍ SPOLEHLIVOSTI JEDNOTLIVÝCH POSLUCHAČŮ.....	140
D.3 ZJIŠTĚNÍ SPOLEHLIVOSTI MĚŘENÍ CELÉ SKUPINY.....	141
PŘÍLOHA E: TABULKY S NAMĚŘENÝMI DATY, PREFERENČNÍ HODNOCENÍ KVALITY VÝSTUPNÍCH HUDEBNÍCH ZÁZNAMŮ TESTOVANÝCH METOD.....	143
E.1 INDIVIDUÁLNÍ HODNOCENÍ POSLUCHAČŮ PRO $SNR = 10$ DB.....	144
E.2 INDIVIDUÁLNÍ HODNOCENÍ POSLUCHAČŮ PRO $SNR = 20$ DB.....	145
E.3 INDIVIDUÁLNÍ HODNOCENÍ POSLUCHAČŮ PRO $SNR = 30$ DB.....	146

PŘÍLOHA A: Další demonstrační příklady aditivním šumem degradovaných zvukových záznamů

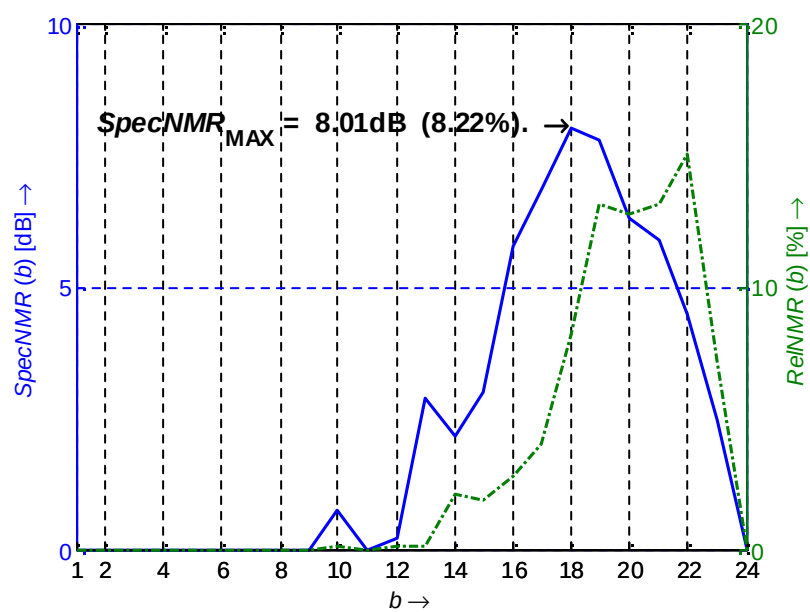
A.1 Sborový zpěv a šum vinyl desky při $SNR = 50$ dB

V tomto příkladu je degradován záznam sborového zpěvu aditivním šumem praskáním vinylové desky při relativně vysoké hodnotě $SNR = 50$ dB. Bude ukázáno, že i při takto vysoké hodnotě SNR je v některých oblastech výkonová spektrální hustota šumu nad maskujícím prahem. Záznam tohoto sborového zpěvu je tvořen čtyřmi vokalisty (sopránem, altem, tenorem a basem), jeho spektrogram je zobrazen na Obr. A.1 a) a časový průběh na Obr. A.1 c). Spektrogram šumu vinylové desky je zobrazen na Obr. A.1 b). Díky vysoké hodnotě SNR jsou jednotlivé rušivé komponenty ("škrábance") šumu vinylové desky málo výrazné na tomto spektrogramu.



Obr. A.1 Časový průběh záznamu sborového zpěvu: a) spektrogram, b) Spektrogram šumu pozadí, šum vinylové desky, c) časový průběh.

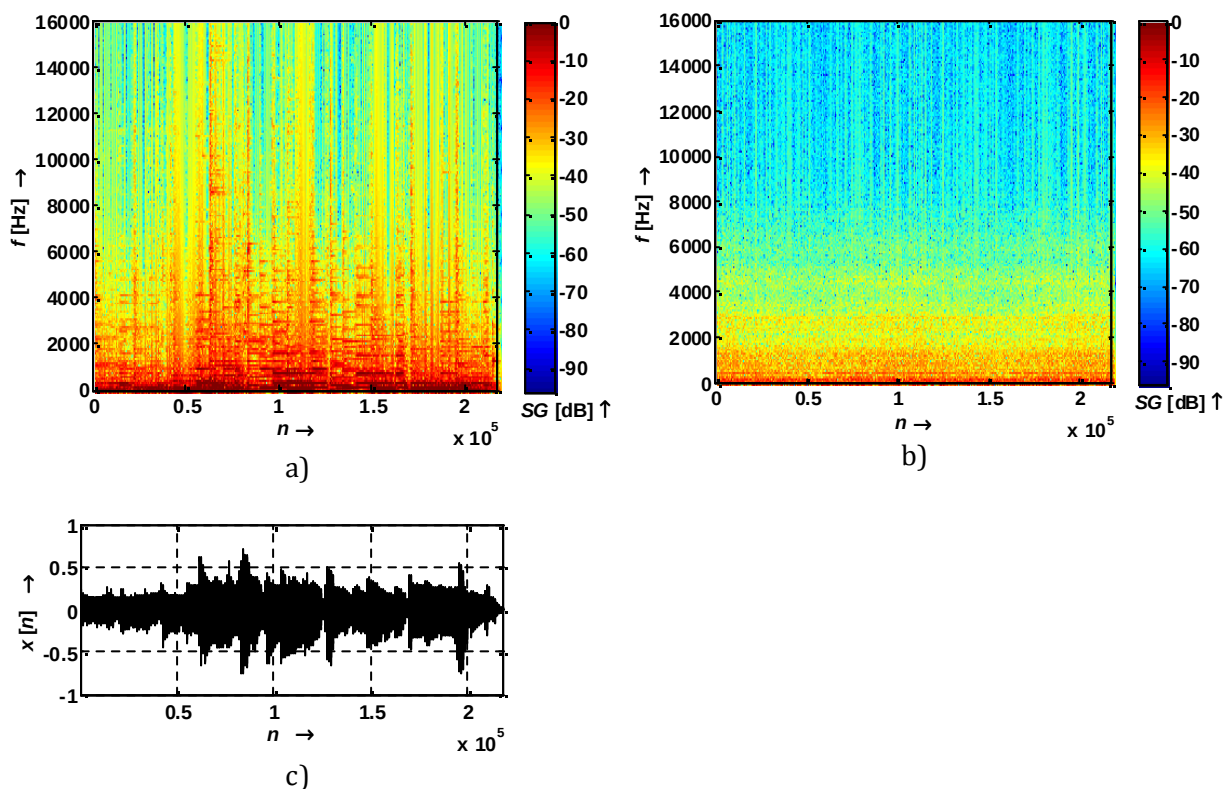
Na grafu z Obr. A.2 jsou zobrazeny průběhy $SpecNMR(b)$, $SpecNMR_{MAX}$ a $RelNMR(b)$. Nejvyšších hodnot $SpecNMR(b)$ je dosaženo v okolí 18. Barkového pásma (tj. kolem kmitočtu 4.2 kHz). $SpecNMR_{MAX}$ je rovno 8.01 dB a četnost výskytu nemaskovaných komponent šumu na pozadí je přibližně 8.22 %. Tato relativní četnost přibližně odpovídá relativnímu počtu výskytu jednotlivých "škrábanců" v šumu pozadí.



Obr. A.2 Výsledné charakteristiky záznamu sborového zpěvu a šumu vinylové desky při $SNR = 50$ dB.

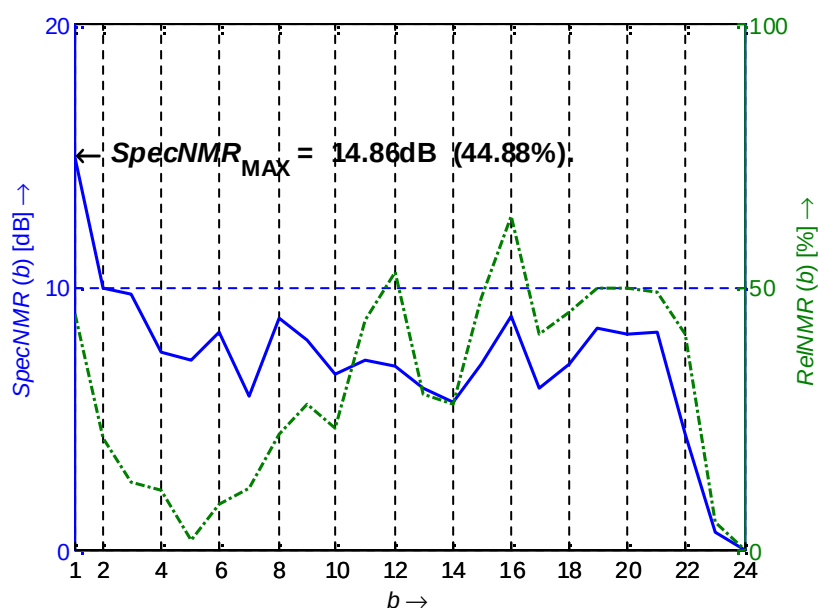
A.2 Záznam hry na kytaru a šumu větráku při $SNR = 20$ dB

Dalším příkladem je záznam hry na kytaru a šumu větráku při $SNR = 20$ dB. Hodnota SNR oproti minulým příkladům nabývá nejnižších hodnot, a proto lze očekávat, že degradace šumem této hudební ukázky bude největší. Příslušné spektrogramy a časový průběh originálního signálu jsou zobrazen na Obr. A.3a-c).



Obr. A.3 Časový průběh záznamu hry na kytaru [23]: a) spektrogram, b) spektrogram šumu pozadí, šum větráku, c) časový průběh.

Výsledné charakteristiky $SpecNMR(b)$, $SpecNMR_{MAX}$ a $RelNMR(b)$ jsou dále zobrazeny na Obr. A.3.



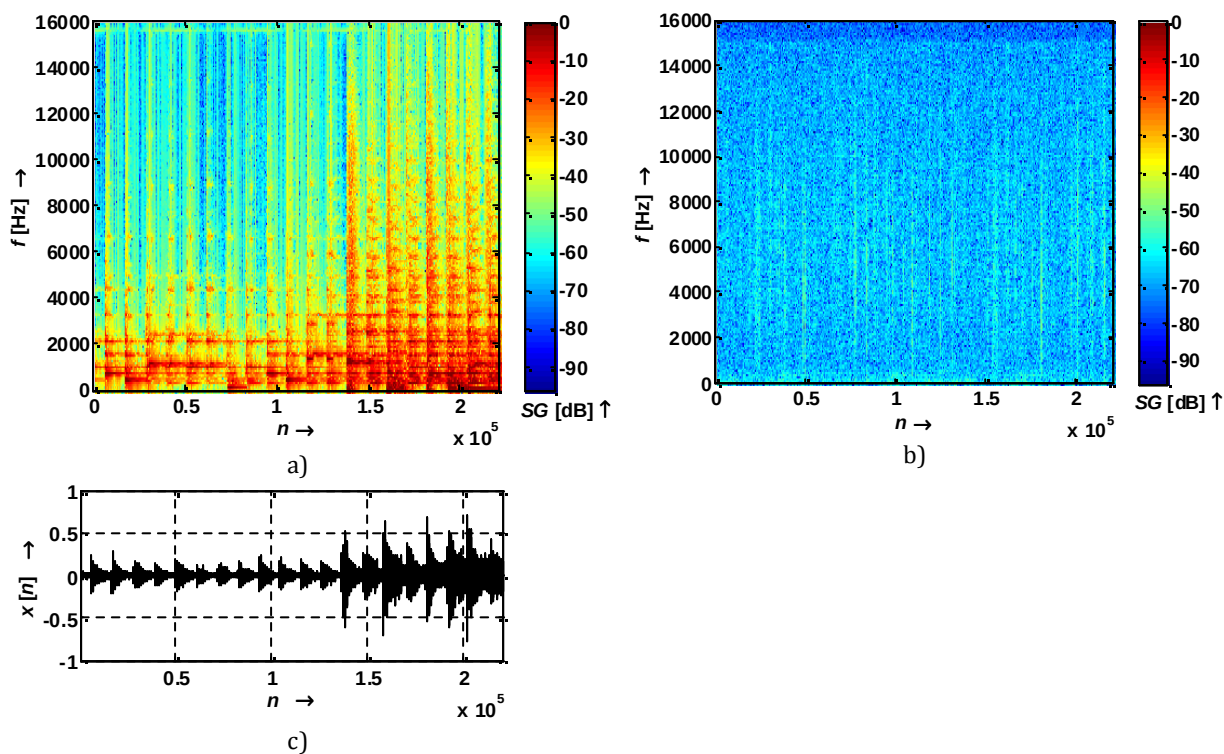
Obr. A.4 Výsledné charakteristiky záznamu hry na kytaru a šumu větráku při $SNR = 20$ dB

Z těchto průběhů lze vidět, že i když je hodnota SNR o 30 dB menší než v minulem příkladu tak $SpecNMR_{MAX}$ je větší pouze o cca 7 dB, tedy je $SpecNMR_{MAX} = 14.86$ dB a v min. příkladu při $SNR = 50$ dB bylo 8.01 dB. To je dáno zejména tím, že spektra obou signálů se překrývají a dochází zde k výraznějšímu efektu maskování komponent šumu na pozadí než v minulém příkladu.

Relativní četnosti nemaskovaných komponent šumu na pozadí $RelNMR(b)$ se pohybují kolem hodnoty 40 % v pásmech $b = 1$ a 11 až 22. Téměř polovina komponent šumu na pozadí pro tyto pásma leží nad maskou, a proto lze tvrdit, že tento šum není tvořen osamocenými komponentami jako v případě šumu vinylové desky.

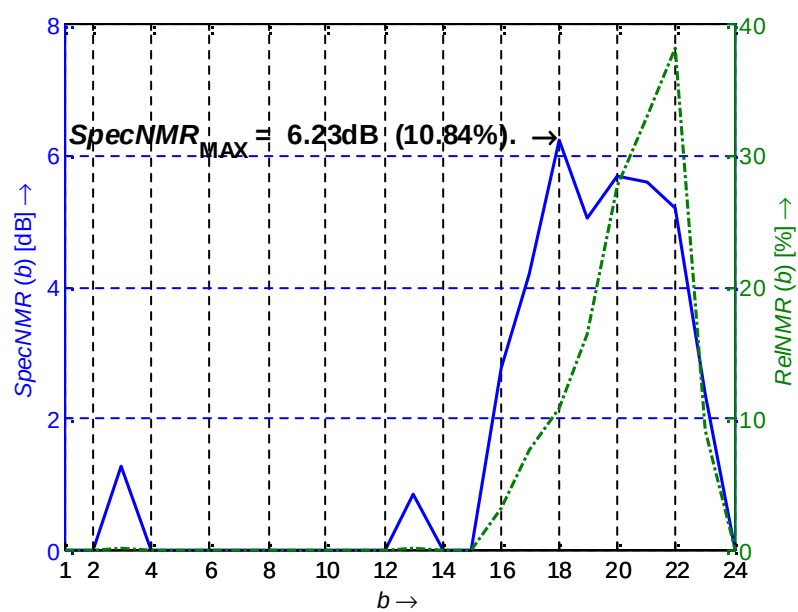
A.3 Skladba skupiny ABBA a šum vinyl desky při $SNR = 50$ dB

V tomto příkladu bude degradována krátká ukázka písně “*Head Over Heels*” od skupiny ABBA [23]. Tato hudební ukázka má výraznou rytmickou strukturu vytvořenou hrou na piano, viz spektrogram a časový průběh z Obr. A.5 a-c).



Obr. A.5 Krátká hudební ukázka skladby “*Head Over Heels*” od skupiny ABBA [23]: a) spektrogram, b) spektrogram šumu pozadí, šum vinyl desky c) časový průběh.

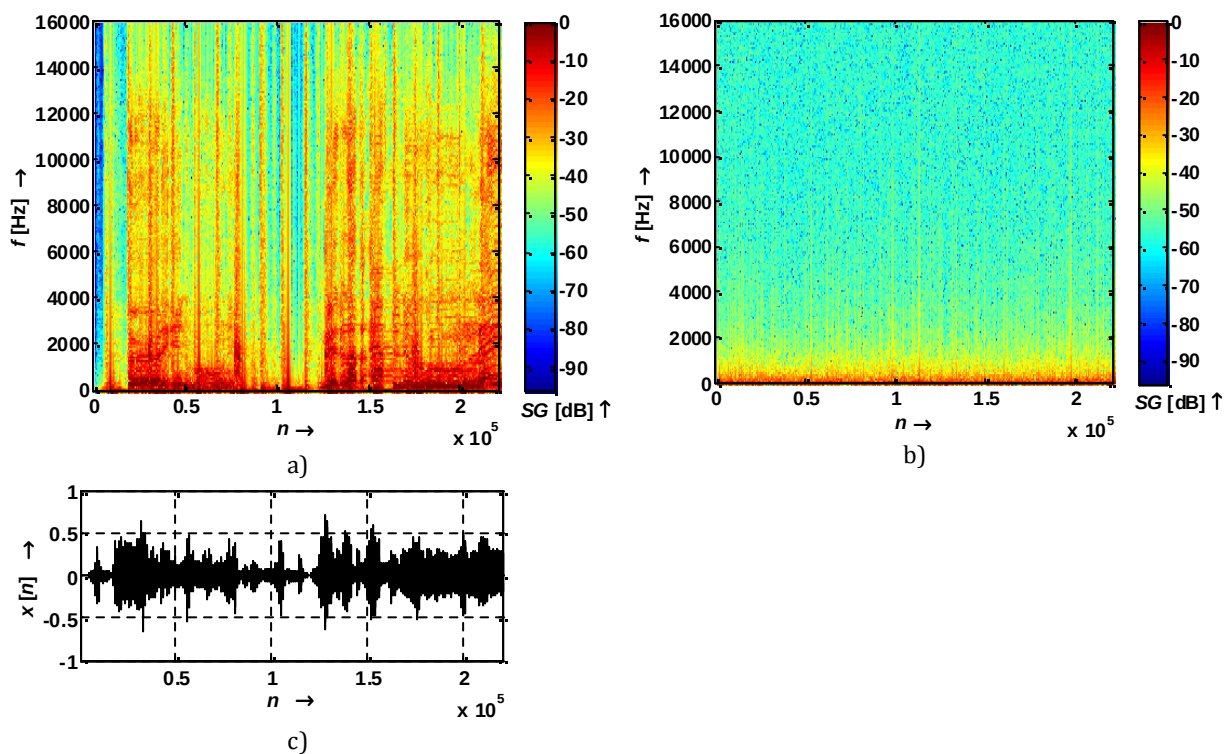
Spektrogram šumu pozadí, který je tvořen šumem vinylové desky, je zobrazen na Obr. A.5 c). Výsledné charakteristiky $SpecNMR(b)$, $SpecNMR_{MAX}$ a $RelNMR(b)$ jsou zobrazeny na Obr. A.5.



Obr. A.6 Výsledné charakteristiky záznamu skladby od skupiny ABBA a šumu vinylové desky při $\text{SNR} = 50$ dB

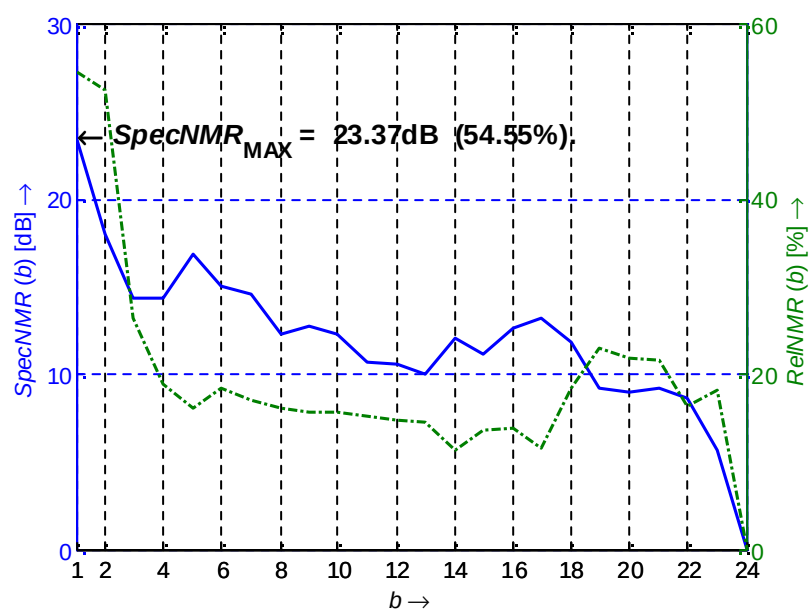
A.4 Country ukázka a šum audio pásky při SNR = 15 dB.

V tomto příkladu bude degradována hudební ukázka krátké country hudební ukázky “*Early in the morning*” od zpěváka Eddie Rabbitt [23]. Tato hudební ukázka je tvořena zpěvem a hrou na kytaru. Spektrogram je zobrazen na Obr. A.7 a) a časový průběh na Obr. A.7 b).



Obr. A.7 Hudební ukázka skladby “*Early in the morning*” od zpěváka Eddie Rabbitt [23]: a) spektrogram, b) spektrogram šumu pozadí, šum audio pásky. b) časový průběh.

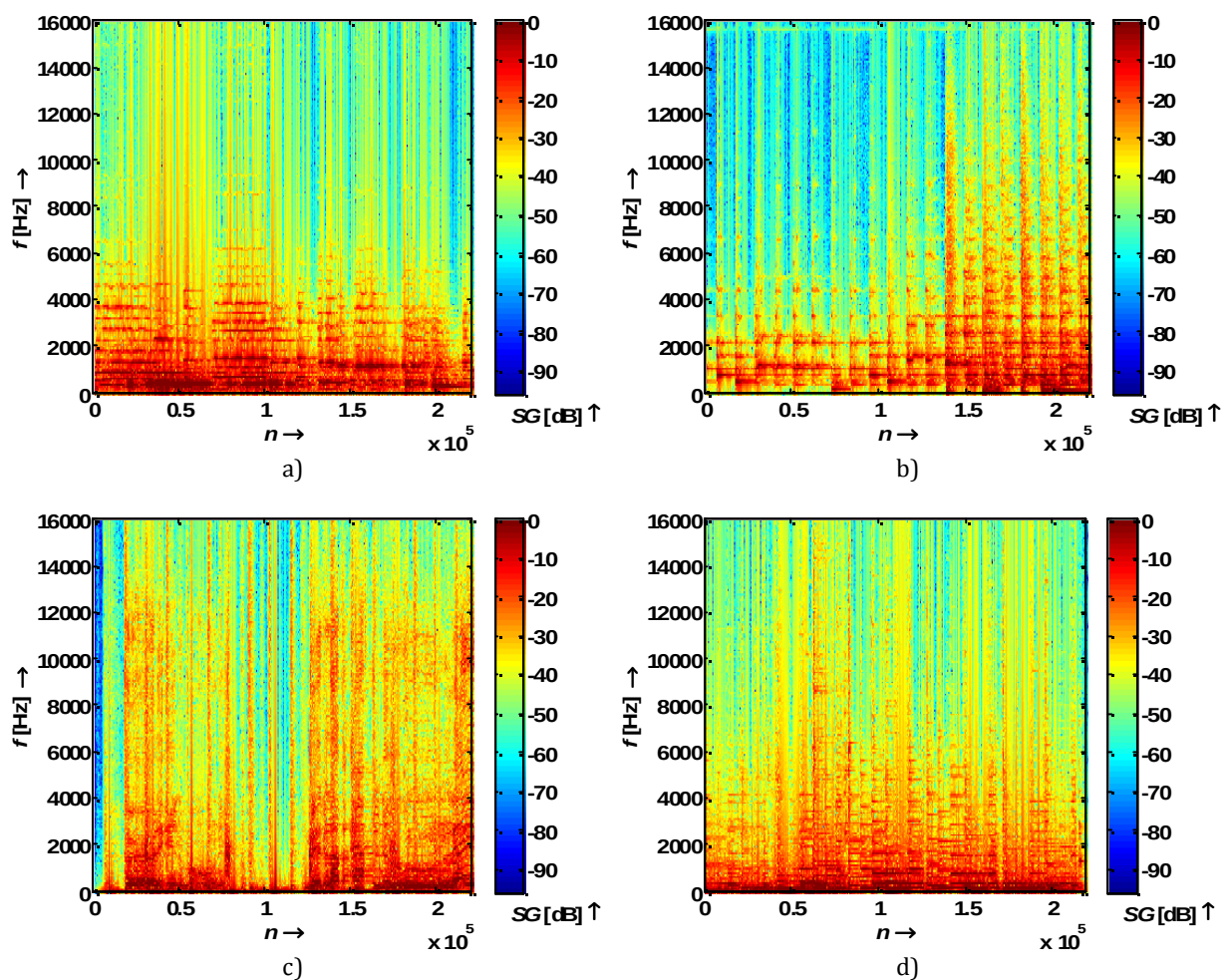
Spektrogram šumu pozadí, který je tvořen šumem audio pásky, je zobrazen na Obr. A.7 b). Výsledné charakteristiky $SpecNMR(b)$, $SpecNMR_{MAX}$ a $RelNMR(b)$ jsou zobrazeny na Obr. A.8.



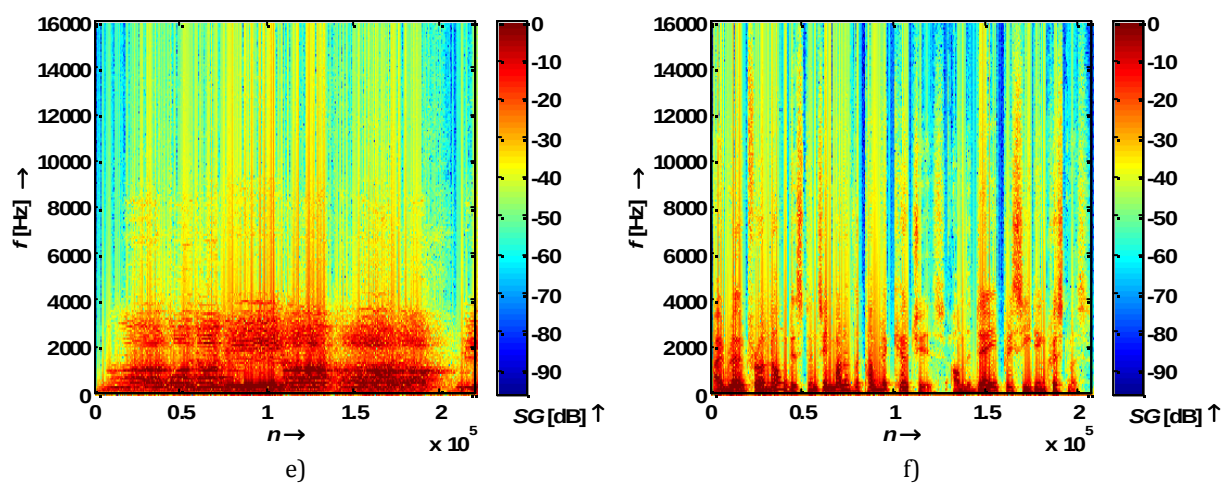
Obr. A.8 Výsledné charakteristiky záznamu country ukázky a šum audio kazety při $SNR = 15$ dB

PŘÍLOHA B: Použitý poslechový materiál

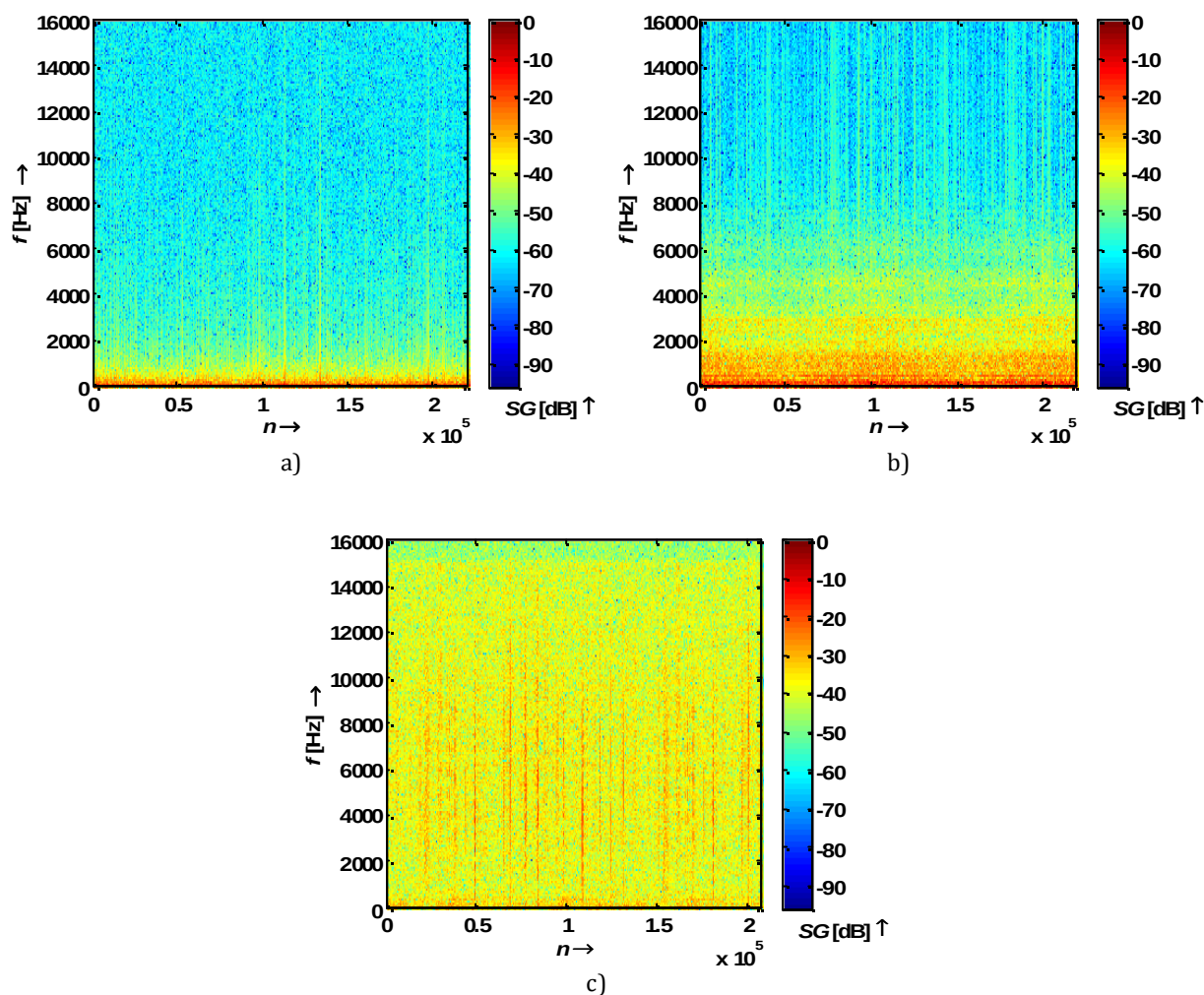
B.1 Poslechový materiál z 1. části poslechového experimentu, subjektivní měření prahu slyšitelnosti hudebních a řečových signálů



Obr. B.1 Spektrogramy hudebních a řečových signálů: a) vážná hudba (Mozart), b) populární hudba (Abba), c) country (Eddie Rabbit), d) hra na kytaru.

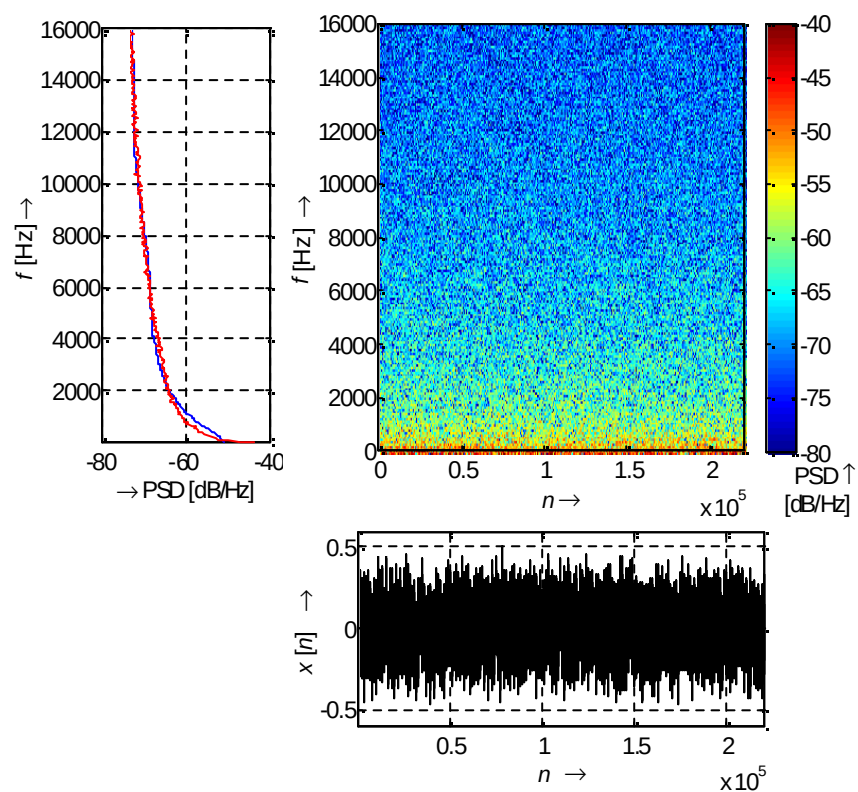


Obr. B.2 Spektrogramy hudebních a řečových (pokračování): e) sborový zpěv (vokál), f) mužská řeč.

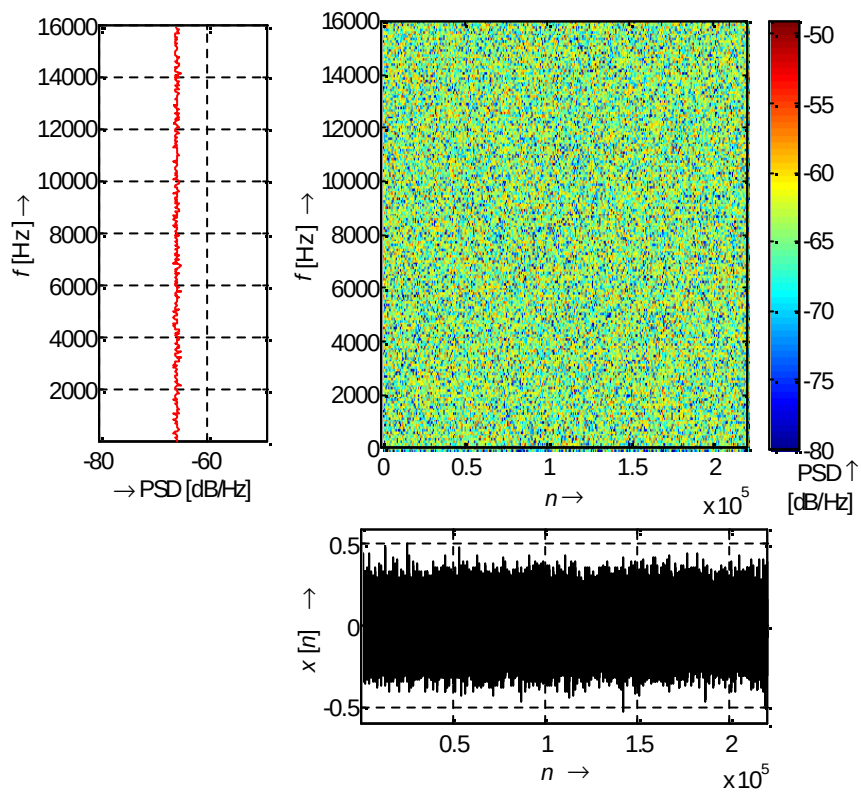


Obr. B.3 Spektrogramy šumů pozadí: a) šum audio kazety, b) šum větráku, c) šum vinylové desky.

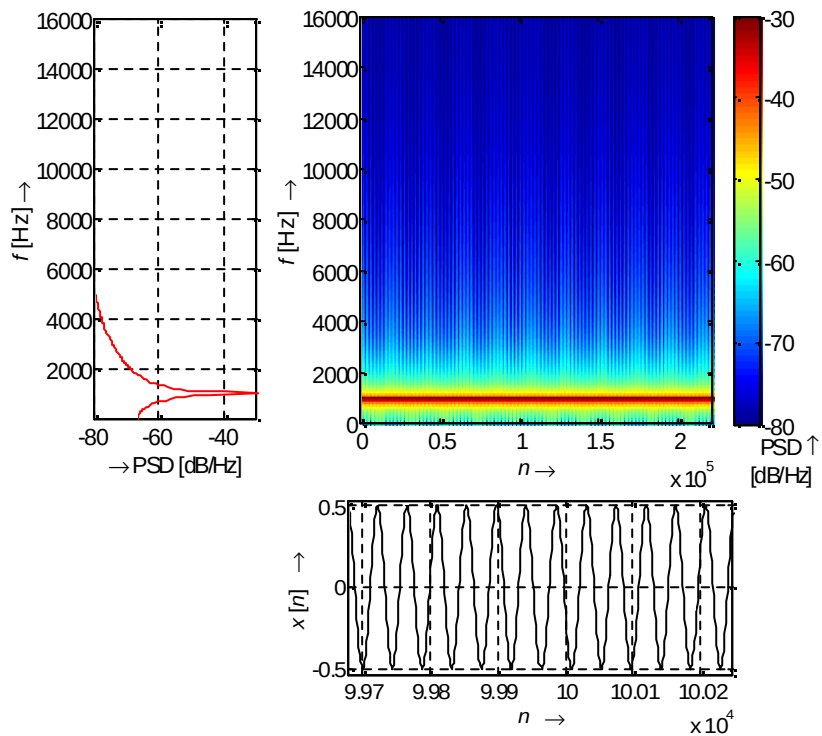
B.2 Poslechový materiál z 2. části poslechového experimentu, subjektivní měření prahu slyšitelnosti umělých zvukových signálů.



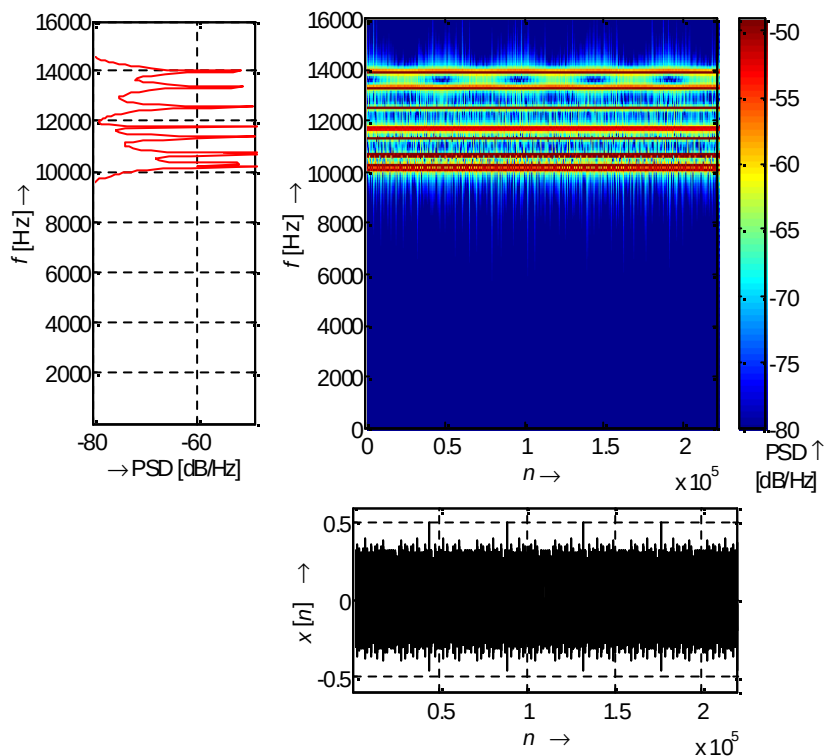
Obr.B.4 Růžový šum.



Obr. B.5 Bílý šum.



Obr. B.6 Tón 1 kHz.



Obr. B.7 Deset kmitočtů v pásmu 10 kHz až 14 kHz.

POZN.: Vlevo nahoře je vždy odhad PSD (červený průběh je odhad podle Barlettovy metody a modrý průběh je LPC modelu 6.řádu), v pravo nahoře spektrogram a dole jsou časové průběhy.

PŘÍLOHA C: Statistické vyhodnocení 1. části poslechového experimentu, prahy slyšitelnosti hudebních a řečových signálů

Tato příloha se zabývá statistickým vyhodnocením změřených dat z 1. části poslechového experimentu z kapitoly 5.3. Kompletní tabulku se všemi naměřenými daty z této části experimentu lze nalézt v Tab. C.5 a Tab. C.6. Cílem tohoto statistického vyhodnocení je testování hypotéz o lineární závislosti mezi změřenými daty z testu a retestu. Obě skupiny změřených dat z testu i retestu lze považovat za náhodná data a proto je potřeba pomocí testování zmíněných hypotéz ověřit stabilitu měření. Jelikož čas mezi testem a retestem byl relativně krátký, posluchači měli jen pár minut, než si celé změření zopakovali, tak se spíše jedná o ověřování krátkodobé stability celého měření [44].

Nulová hypotéza H_0 tvrdí, že $r = 0$, tj. mezi srovnávanými daty neexistuje žádná lineární závislost. Alternativní hypotéza H_1 zní, že $r \neq 0$ (pro oboustranný test) [44]. Symbol r zde reprezentuje korelační koeficient. Při výpočtech budou použity dva typy korelačních koeficientů. Prvním typem je Pearsonův součinný koeficient korelace r_p

$$r_p = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}, \quad (\text{A.1})$$

kde $\bar{x}$ a $\bar{y}$ jsou výběrové průměry, s_x a s_y jsou výběrové směrodatné odchylky a N je počet členů v každé z korelovaných řad x_i a y_i .

Druhým použitým typem korelačního koeficientu je Spearmanův pořadový koeficient korelace r_s

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1) - (T_x + T_y)}, \quad (\text{A.2})$$

kde d_i je rozdíl mezi hodnotami pořadí ležícími v obou řadách na i -tém místě, T_x a T_y jsou definovány

$$T_x = 0.5 \sum (t_x^3 - t_x), \quad (\text{A.3})$$

$$T_y = 0.5 \sum (t_y^3 - t_y), \quad (\text{A.4})$$

kde t_x a t_y jsou počty sloučených pořadí v jednotlivých skupinách (tj. počty dvojic, trojic) u řad x_i a y_i .

Testovaná nulová hypotéza H_0 bude zamítnuta ve prospěch hypotézy H_1 tehdy, když vypočtena hodnota t bude větší než kritická hodnota kvantilu studentova rozdělení t_{krit} pro zvolenou hladinu významnosti α . Hodnota t_{krit} byla nalezena v tabulkách kritických hodnot pro $sv = N - 2$ stupňů volnosti. Hodnota t je definována

$$t = r \sqrt{\frac{N-2}{1-r^2}}, \quad (\text{A.5})$$

kde r je hodnota korelačního koeficientu, tj. Pearsonův součinný korelační koeficient r_p (A.1), nebo Spearmanův koeficient pořadové korelace r_s (A.2). Nahrazením r_s za r do vztahu (A.3) je možno jen pouze tehdy, když $N > 8$ [63], podle [24] pak $N \geq 10$.

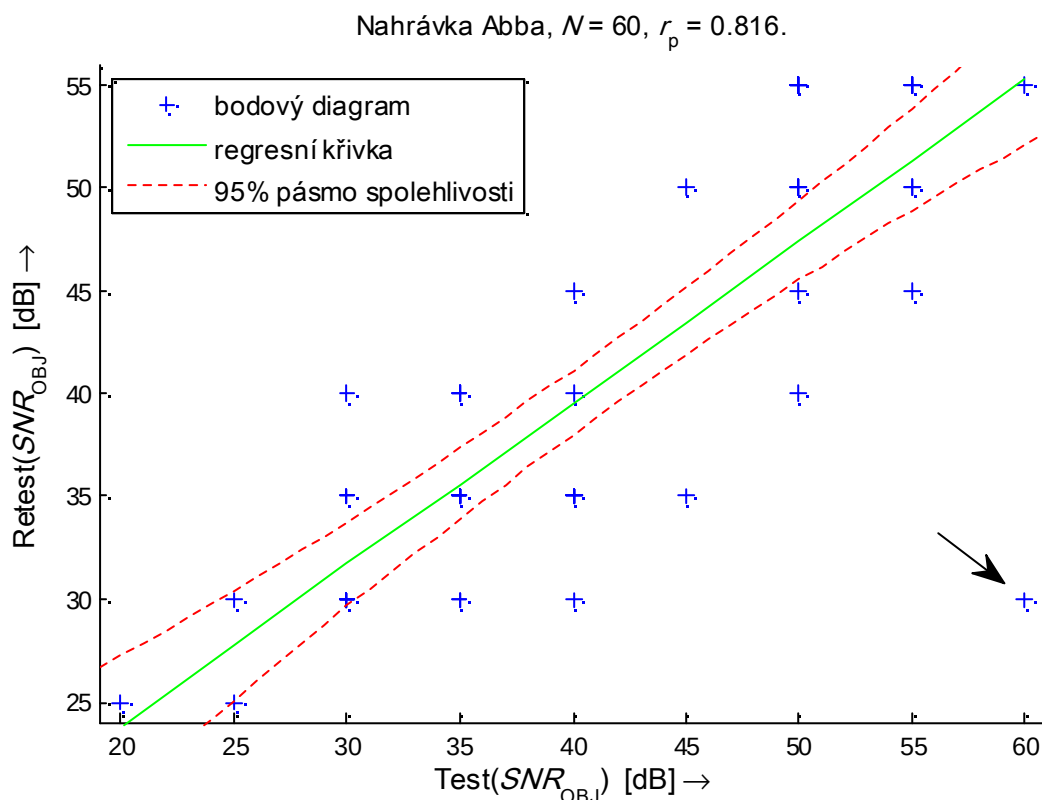
C.1 Krátkodobá stabilita měření pro jednotlivé zvukové ukázky

Cílem této statistiky je zjistit, zda při hodnocení nějakého posluchače nedošlo k extrémní odchylce pro některou z testovaných zvukových ukázek. Nejčastěji takováto odchylka bývá způsobena chybným nastavením výsledku v důsledku nedbalosti, nepozornosti popř. únavy hodnotícího posluchače. Bude zde testována lineární závislost mezi daty pro jednotlivé zvukové ukázky z testu a retestu. Korelovány budou dvě řady o délce $N = 60$ členů, tj. 20 posluchačů a 3 typy šumů pozadí. Vypočtené hodnoty Pearsonova korelačního koeficientu r_p (A.1) jsou zobrazeny v Tab. C.1. Na dalším řádku této tabulky jsou zobrazeny hodnoty t podle vztahu (A.5).

Tab. C.1 Hodnoty Pearsonova korelačního koeficientu r_p a příslušných t hodnot

	Abba	Řeč	Kytara	Mozart	Country	Vokál
r_p	0,816	0,916	0,849	0,901	0,894	0,829
t	10,739	17,333	12,237	15,790	15,158	11,302

Nejnižší hodnotu Pearsonova korelačního koeficientu z této tabulky mají data pro zvukovou ukázku Abba. Následnou detailní analýzou bylo zjištěno, že tato nejmenší hodnota byla zapříčiněna jedním extrémním nastavením jednoho z posluchačů, toto extrémní nastavení bylo nejspíš způsobeno chybným označením odpovědi toho posluchače. Na Obr. C.1 je zobrazen bodový diagram závislosti SNR_{OBJ} z retestu na SNR_{OBJ} z testu. Dále je zde zobrazena regresní křivka $y = 0.788x + 8.054$ a 95% pásmo spolehlivosti pro korelační vztah mezi hodnotami SNR_{OBJ} z testu a retestu. Šipkou je dále naznačen tento extrémně nastavený bod.



Obr. C.1 Bodový diagram, regresní přímka a 95% pásmo spolehlivosti pro korelační vztah mezi hodnotami SNR_{OBJ} z testu a retestu u hudební ukázky Abba se všemi typy šumu pozadí.

Po nahrazení tohoto extrémního bodu průměrnou hodnotou 35 dB se zvýší hodnota Pearsonova korelačního koeficientu r_p z 0.816 na hodnotu 0.887. Pro ostatní zvukové ukázky nebyl nalezen, žádný takto podobně vychýlený bod. Tabulka hodnot korelačních koeficientu r_p a t hodnot po této úpravě je zobrazena v Tab. C.2.

Tab. C.2 Upravené hodnoty Pearsonova korelačního koeficientu r_p a příslušných t hodnot.

	Abba	Řeč	Kytara	Mozart	Country	Vokál
r_p	0,887	0,916	0,849	0,901	0,894	0,829
t	14,625	17,333	12,237	15,790	15,158	11,302

Kritické hodnoty t_{krit} pro $sv = 58$ jsou $t_{krit}(0,05) = 2.01$ a $t_{krit}(0,01) = 2,68$. V obou případech jsou tyto hodnoty podstatně menší než t hodnoty z Tab. C.2 a proto lze zamítnout hypotézu H_0 o neexistenci lineární závislosti na hladinách významnosti 5% a 1%.

C.2 Zjištění spolehlivosti jednotlivých posluchačů

Krátkodobá stabilita měření bude také posuzována na úrovni posluchačů. Budou zde korelovány individuální nastavení posluchačů při testu a retestu. Cílem je posoudit, zda hodnocení některých posluchačů nebylo příliš rozdílné a pokud ano, tak pak tyto posluchače vyřadit z poslechového experimentu.

Korelované řady v tomto případě mají relativně nízký počet členů $N = 18$ a proto bude namísto Pearsonova korelačního koeficientu použit Spearmanův pořadový korelační koeficient. Literatura [44] [63] totiž udává, že pro posouzení korelačního vztahu s tak malým počtem členů je vhodné používat Spearmanův korelační koeficient, namísto Pearsonova korelačního koeficientu.

V Tab. C.3 jsou uvedeny hodnoty Spearmanova korelačního koeficientu r_s a t hodnoty pro jednotlivé posluchače. Kritické hodnoty t_{krit} pro $sv = 16$ jsou $t_{\text{krit}}(0,05) = 2,120$ a $t_{\text{krit}}(0,01) = 2,921$. Většina t hodnot má mnohem větší hodnoty než tyto kritické hodnoty, takže lze zamítnout nulovou hypotézu H_0 o lineární nezávislosti korelovaných dat. V jednom případě je ale t hodnota posluchače č. 17 velmi blízká $t_{\text{krit}}(0,01)$. Žádný z posluchačů ale nebyl vyřazený z poslechového experimentu.

Tab. C.3 Hodnoty Spearmanova korelačního koeficientu r_s a t hodnoty pro jednotlivé posluchače.

Posluchač číslo	r_s	t	Posluchač číslo	r_s	t
1	0,922	9,542	11	0,918	9,265
2	0,864	6,862	12	0,883	7,509
3	0,883	7,537	13	0,956	13,103
4	0,832	5,997	14	0,691	3,819
5	0,905	8,509	15	0,962	14,087
6	0,924	9,657	16	0,695	3,871
7	0,947	11,794	17	0,597	2,975
8	0,918	9,252	18	0,700	3,922
9	0,767	4,781	19	0,917	9,209
10	0,954	12,709	20	0,910	8,756

C.3 Zjištění spolehlivosti měření celé skupiny

Spolehlivost měření celé skupiny bude ověřována korelováním všech dat z testu a retestu. Každá z řad bude mít tedy délku $N = 360$ členů. Vypočtené hodnoty Pearsonova korelačního koeficientu r_p , t hodnot a kritických hodnot pro hladiny významnosti 1% a 5% jsou vyneseny v Tab. C.4. Z této tabulky vyplývá, že lze zamítnout nulovou hypotézu H_0 o lineární nezávislosti korelovaných dat, protože vypočtená t hodnota je mnohem větší než obě kritické hodnoty.

Tab. C.4 Hodnoty Pearsonova korelačního koeficientu r_p , t hodnot a kritických hodnot pro hladiny významnosti 1% a 5%.

r_p	t	$t_{\text{krit}}(0,05)$	$t_{\text{krit}}(0,01)$
0,878	34,764	1,960	2,576

Tab. C.5.První část výsledné tabulky změřených dat z 1. části poslechového experimentu.

Název / Hodnocení posluchačů	Abba						Řeč						Kytara					
	Větrák		Kazeta		Vinyl deska		Větrák		Kazeta		Vinyl deska		Větrák		Kazeta		Vinyl deska	
	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest
	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]
1	40	35	35	35	50	55	40	40	35	30	50	50	40	40	35	35	50	50
2	35	35	35	35	55	55	40	40	30	35	55	55	35	40	30	30	50	50
3	30	30	20	25	45	50	40	35	25	30	50	45	25	30	25	25	45	45
4	35	30	35	35	50	55	35	45	30	35	50	50	30	35	30	30	50	50
5	40	35	35	40	55	55	40	45	35	30	55	50	35	30	35	30	55	50
6	35	35	40	35	50	50	40	45	35	35	50	55	30	35	30	30	50	50
7	30	30	25	30	50	55	35	35	25	25	45	50	25	25	30	30	50	50
8	35	35	30	30	50	55	40	40	35	30	50	50	35	35	30	30	50	50
9	40	45	45	35	50	55	45	40	40	35	50	50	35	45	35	45	50	50
10	30	30	25	25	50	50	35	35	25	30	55	50	30	30	20	25	45	45
11	30	35	30	30	55	50	40	40	30	30	55	50	30	30	30	30	50	45
12	40	35	40	35	50	50	40	40	35	35	50	50	35	35	30	30	45	50
13	35	35	35	30	50	45	40	40	30	30	45	45	30	30	25	25	50	45
14	35	40	35	40	50	55	40	40	35	35	50	50	35	40	35	40	60	45
15	35	35	40	35	55	55	40	40	30	30	50	50	30	30	30	30	50	55
16	35 (60)	30	40	30	55	50	40	40	30	30	50	50	35	25	40	30	45	50
17	50	40	35	35	60	55	45	40	35	35	55	50	55	45	50	40	60	50
18	30	35	30	40	50	55	40	40	30	35	50	50	40	45	30	50	60	55
19	40	40	30	35	55	55	45	50	35	40	60	55	30	35	30	30	55	55
20	30	30	30	30	55	45	40	40	30	30	45	50	35	30	30	35	50	50
Průměr	35,50	34,75	33,50	33,25	52,00	52,50	40,00	40,50	31,75	32,25	51,00	50,25	33,75	34,50	31,50	32,50	51,00	49,50
Celkový průměr	35,13		33,38		52,25		40,25		32,00		50,63		34,13		32,00		50,25	

Tab. C.6.Druhá část výsledné tabulky změřených dat z 1. části poslechového experimentu.

Název / Hodnocení posluchačů	Mozart						Country						Vokál					
	Větrák		Kazeta		Vinyl deska		Větrák		Kazeta		Vinyl deska		Větrák		Kazeta		Vinyl deska	
	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest
	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]
1	30	30	30	35	55	55	35	40	25	25	45	45	35	35	30	30	50	55
2	35	35	35	35	60	60	40	45	30	40	50	50	50	40	35	30	55	55
3	25	25	30	30	60	55	35	40	20	25	45	55	35	50	30	45	50	50
4	30	30	30	30	55	55	35	50	30	30	45	50	35	35	30	35	55	50
5	30	30	35	35	60	55	40	40	30	30	50	50	45	40	35	40	60	55
6	30	30	40	35	55	55	40	40	25	30	50	55	40	35	35	35	50	50
7	25	25	30	30	55	55	25	25	35	35	55	55	35	35	30	25	50	50
8	40	35	35	40	60	60	40	45	35	35	50	55	45	40	35	35	55	50
9	40	40	35	30	60	55	45	50	50	50	50	50	45	45	40	35	55	50
10	30	25	30	30	55	55	35	35	25	25	45	45	35	35	30	30	50	50
11	30	30	30	30	60	55	45	45	35	30	55	55	40	40	30	35	60	50
12	30	35	35	35	50	55	40	40	35	30	50	50	40	40	40	35	55	50
13	30	30	30	30	55	50	35	35	25	25	45	40	30	35	25	25	50	50
14	35	35	30	40	60	55	45	40	35	40	50	50	45	40	30	40	45	55
15	30	30	35	35	55	55	35	35	25	25	45	45	40	35	35	30	55	55
16	45	30	45	30	60	55	35	35	30	25	45	45	35	40	30	35	55	55
17	50	40	30	55	65	60	40	35	40	45	60	55	45	45	35	45	60	55
18	35	35	45	35	60	55	45	50	30	40	55	45	40	45	30	40	55	55
19	40	35	35	35	60	60	30	35	25	25	55	50	45	45	35	35	50	60
20	30	30	30	30	55	55	35	40	30	30	40	45	35	40	30	30	55	50
Průměr	33,50	31,75	33,75	34,25	57,75	55,75	37,75	40,00	30,75	32,00	49,25	49,50	39,75	39,75	32,50	34,50	53,50	52,50
Celkový průměr	32,63		34,00		56,75		38,88		31,38		49,38		39,75		33,50		53,00	

PŘÍLOHA D: Statistické vyhodnocení 2. části poslechového experimentu, prahy slyšitelnosti umělých zvukových signálů.

D.1 Krátkodobá stabilita měření pro jednotlivé zvukové ukázky

Při testování krátkodobé stability měření bylo opět objeveno jedno extrémní odchýlení u jedné z testovaných zvukových ukázek. Jednalo se o ukázkou s kombinací bílý šum a růžový šum jako šum pozadí. Toto extrémní odchýlení mělo hodnotu 45 dB a bylo nahrazeno průměrnou hodnotou 15 dB. V Tab. D.5 je toto extrémní odchýlení označeno červenou hodnotou na příslušném řádku a sloupci. Zřejmě byla tato chyba způsobena velmi podobným zvukem obou typů šumů. Při vyšších hodnotách SNR bylo pro posluchače obtížnější zvuky těchto dvou šumů od sebe odlišit.

V tomto testu byly korelovány dvě řady o délce $N = 30$ členů, tj. 10 posluchačů a 3 typy šumů pozadí. Této délce odpovídá 28 stupňů volnosti, tj. $sv = 28$. Kritické hodnoty t_{krit} pro $sv = 28$ jsou $t_{krit}(0,05) = 2,048$ a $t_{krit}(0,01) = 3,047$. Hodnoty Pearsonova korelačního koeficientu r_p a příslušných t hodnot jsou zobrazeny v Tab. D.1. Protože jsou t -hodnoty z této tabulky větší než kritické hodnoty $t_{krit}(0,05)$ a $t_{krit}(0,01)$ pro obě hladiny významnosti, tak byla zamítnuta nulová hypotéza o lineární nezávislosti obou korelovaných řad.

Tab. D.1 Hodnoty Pearsonova korelačního koeficientu r_p a příslušných t hodnot.

	10 tónů	Růžový šum	1 kHz tón	Bílý šum
r_p	0,542	0,700	0,619	0,667
t	3,415	5,188	4,168	4,737

D.2 Zjištění spolehlivosti jednotlivých posluchačů

Konzistentnost hodnocení posluchačů v tomto experimentu byla testována korelováním jejich individuálních hodnocení v testu a retestu. Sledovalo se, zda nejsou tyto hodnocení příliš rozdílné. Příliš nízké hodnoty korelačního koeficientu r_s pro některého z posluchačů znamenaly, že tento posluchač hodnotil příliš rozdílně v testu a retestu.

Korelované byly dvě řady o velikosti $N = 12$ členů, této velikosti řad odpovídá $sv = 10$ stupňů volnosti. Z důvodu takto nízkého počtu členů jednotlivých řad byl použit Spearmanův korelační koeficient r_s (A.2) namísto Pearsonova korelačního koeficientu r_p (A.1) [44] [63]. V Tab. D.2 jsou uvedeny hodnoty Spearmanova korelačního koeficientu r_s a t hodnoty pro jednotlivé posluchače. Kritické hodnoty t_{krit} pro $sv = 10$ jsou $t_{krit}(0,05) = 2,228$ a $t_{krit}(0,01) = 3,581$. Všechny t -hodnoty mají mnohem větší hodnoty než obě kritické hodnoty t_{krit} , takže lze zamítnout nulovou hypotézu H_0 o lineární nezávislosti korelovaných dat.

Tab. D.2 Hodnoty Spearmanova korelačního koeficientu r_s a t hodnoty pro jednotlivé posluchače.

Posluchač číslo	r_s	t
1	0,885	6,012
2	0,894	6,300
3	0,948	9,385
4	0,966	11,787
5	0,951	9,691
6	0,942	8,855
7	0,804	4,282
8	0,938	8,564
9	0,975	13,972
10	0,951	9,755

D.3 Zjištění spolehlivosti měření celé skupiny

Spolehlivost měření celé skupiny byla ověřována korelováním všech dat z testu a retestu. Každá z řad měla délku $N = 120$ členů. Vypočtené hodnoty Pearsonova korelačního koeficientu r_p , t hodnot a kritických hodnot pro hladiny významnosti 1% a 5% jsou vyneseny v Tab. D.3. Z této tabulky vyplývá, že lze opět zamítnout nulovou hypotézu H_0 o lineární nezávislosti korelovaných dat, protože vypočtena t hodnota je mnohem větší než obě kritické hodnoty.

Tab. D.3. Hodnoty Pearsonova korelačního koeficientu r_p , t hodnot a kritických hodnot pro hladiny významnosti 1% a 5%.

r_p	t	$t_{\text{krit}}(0,05)$	$t_{\text{krit}}(0,01)$
0,954	34,566	1,980	2,617

Tab. D.4. .První část výsledné tabulky změřených dat z 2. části poslechového experimentu

Název / Hodnocení posluchačů	Tóny v pásmu 10 až 14kHz (v textu zkráceně označováno jako 10 tónů)						Růžový šum					
	Růžový šum		1 kHz tón		Bílý šum		10 tónů		1 kHz tón		Bílý šum	
	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest
	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]
1	45	45	55	60	50	50	25	20	25	30	20	15
2	50	50	55	55	50	50	20	25	20	25	20	15
3	50	50	50	55	45	55	25	25	20	25	20	20
4	45	50	55	55	50	50	25	30	25	30	10	15
5	50	55	60	60	60	55	20	30	30	30	10	15
6	45	55	55	60	55	55	25	25	30	25	20	20
7	50	50	55	50	50	55	25	30	25	25	20	15
8	45	50	60	55	50	50	25	25	30	30	15	20
9	55	50	55	50	45	50	30	30	25	30	20	15
10	55	55	45	50	55	55	30	25	30	30	20	15
Průměr	49	51	54,5	55	51	52,5	25	26,5	26	28	17,5	16,5
Celkový průměr	50		54,75		51,75		25,75		27		17	

Tab. D.5. Druhá část výsledné tabulky změřených dat z 2. části poslechového experimentu

Název / Hodnocení posluchačů	1 kHz tón						Bílý šum					
	10 tónů		Růžový šum		Bílý šum		10 tónů		Růžový šum		1 kHz tón	
	test	retest	test	retest	test	retest	test	retest	test	retest	test	retest
	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]
1	50	50	45	45	50	50	30	15	10	20	25	30
2	45	45	45	50	50	55	15	25	15 (45)	20	25	25
3	45	45	50	55	55	60	25	35	10	10	25	35
4	45	50	45	45	50	50	20	20	20	15	25	30
5	50	50	50	55	50	55	10	15	10	10	40	30
6	60	60	55	55	55	60	25	25	20	20	30	25
7	45	55	55	55	60	50	25	20	10	15	30	25
8	45	50	55	50	50	55	15	15	20	20	25	30
9	60	55	45	45	55	50	20	15	15	10	25	30
10	45	45	50	45	60	55	20	25	15	20	30	35
Průměr	49	50,5	49,5	50	53,5	54	20,5	21	14,5	16	28	29,5
Celkový průměr	49,75		49,75		53,75		20,75		15,25		28,75	

PŘÍLOHA E: Tabulky s naměřenými daty, preferenční hodnocení kvality výstupních hudebních záznamů testovaných metod.

V této příloze jsou uvedeny podrobné tabulky s výsledky poslechového experimentu pro stanovení subjektivní kvality testovaných hudebních ukázek. Cílem tohoto testu bylo porovnat tři rozdílné způsoby segmentace v technikách prahování a stanovit jejich vliv na výslednou subjektivní kvalitu výstupních hudebních signálů. Prvním způsobem segmentace byla pevná segmentace, druhý způsob byla adaptivní segmentace podle algoritmu Greedy selection a třetí způsob segmentace byla adaptivní segmentace podle vlastního řešení nazvaná Maximum greedy selection, detailněji jsou tyto metody popsány v kapitole 8.3. Metoda pro potlačení šumu s pevnou segmentací je označována jako M1, metoda s adaptivní segmentací podle algoritmu Greedy selection je označována jako M2 a třetí metoda s vlastním řešením s adaptivní segmentací podle algoritmu Maximum greedy selection je značena jako M3.

Posluchači v tomto testu volbou preferenčního hodnocení volili ty hudební ukázky, které se jim zdály jako "subjektivně nejkvalitnější". Posluchači byli nuceni vždy označit alespoň jednu ukázku. Přičemž nemuseli hodnotit pouze jednu hudební ukázku, ale klidně více ukázek. To nastávalo zejména v případech, kdy nedokázali rozlišit mezi testovanými hudebními ukázkami. Například z důvodu nízkého stupně zarušení ve vstupním hudebním signálu. Zejména tento jev nastával, při hodnotách $SNR = 30$ dB (5.25) mezi užitečným signálem popředí a šumem pozadí. U takto vysokých hodnot SNR bylo preferenční hodnocení pro posluchače velmi obtížné, hlavně kvůli velké podobnosti testovaných hudebních ukázek.

Způsob provedení experimentu, jeho charakteristika, použitý poslechový materiál a diskuze nad výsledky tohoto experimentu, jsou uvedeny v kapitole 8.3, v části s poslechovým experimentem. V této příloze jsou uvedeny pouze kompletní tabulky s individuálními hodnoceními jednotlivých posluchačů.

Tabulky Tab. E.1 až Tab. E.6 jsou strukturovány následovně. Na prvním řádku je název užitečného signálu popředí a na druhém řádku jsou názvy šumu pozadí. Názvy těchto signálů odpovídají zavedenému značení z přílohy B. Na dalších řádcích následují individuální preferenční hodnocení každého z 10 posluchačů. Preferenční hodnocení je zde značeno symbolem "X". Pokud posluchač preferoval více metod, pak je použit tento symbol vícekrát. Na posledním řádku je nakonec vypočtena procentuální hodnota preferencí pro každou z testovaných metod a pro každou možnou kombinaci užitečného signálu popředí a šumu pozadí.

E.1 Individuální hodnocení posluchačů pro $SNR = 10$ dB.

Tab. E.1. Individuální hodnocení v testu, pro $SNR = 10$ dB.

Posluchač č.	Kytara						Mozart						Abba					
	Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1			X			X			X			X			X			X
2			X			X			X			X			X			X
3			X			X			X			X			X			X
4			X			X			X			X			X			X
5			X			X			X			X			X			X
6			X			X			X			X		X				X
7			X			X			X			X			X			X
8		X	X			X			X			X			X			X
9			X			X			X			X			X			X
10			X			X			X			X			X			X
Preference	0%	5%	95%	0%	0%	100%	0%	0%	100%	0%	0%	100%	0%	10%	90%	0%	0%	100%

Tab. E.2. Individuální hodnocení v **retestu**, pro $SNR = 10$ dB.

Posluchač č.	Kytara						Mozart						Abba					
	Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1			X			X			X			X			X			X
2			X			X			X			X			X			X
3			X	X		X			X			X			X			X
4			X	X					X			X			X			X
5	X					X			X			X			X			X
6	X					X			X			X			X			X
7			X			X			X			X			X			X
8			X			X			X			X			X			X
9			X			X			X			X			X			X
10	X		X			X			X			X			X			X
Preference	25%	0%	75%	15%	0%	100%	0%	0%	100%	0%	0%	100%	0%	0%	100%	0%	0%	100%

E.2 Individuální hodnocení posluchačů pro $SNR = 20$ dB.

Tab. E.3. Individuální hodnocení v **testu**, pro $SNR = 20$ dB.

Posluchač č.	Kytara						Mozart						Abba					
	Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1			X	X					X			X			X	X		
2			X	X					X			X			X			X
3			X	X					X			X	X	X	X			X
4		X	X			X			X			X			X			X
5			X		X				X			X		X				X
6			X			X	X			X					X	X		
7			X			X			X			X			X			X
8			X	X	X	X						X			X			X
9	X	X	X			X	X			X		X			X			X
10			X			X	X		X			X			X	X		
Preference	3%	8%	88%	33%	13%	53%	30%	0%	70%	15%	0%	85%	3%	13%	83%	30%	%	70%

Tab. E.4. Individuální hodnocení v **retestu**, pro $SNR = 20$ dB.

Posluchač č.	Kytara						Mozart						Abba					
	Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1			X	X			X	X	X			X			X			X
2			X	X			X		X			X			X			X
3			X			X			X			X			X			X
4	X		X			X			X			X			X			X
5			X			X			X			X			X			X
6			X			X			X	X					X			X
7			X			X			X			X			X			X
8		X				X			X			X			X			X
9	X					X	X			X		X			X			X
10	X		X			X			X			X			X	X		
Preference	20%	10%	70%	20%	0%	80%	18%	3%	88%	15%	0%	85%	0%	0%	100%	10%	0%	90%

E.3 Individuální hodnocení posluchačů pro $SNR = 30$ dB.

Tab. E.5. Individuální hodnocení v **testu**, pro $SNR = 30$ dB.

Posluchač č.	Kytara						Mozart						Abba					
	Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2			X			X			X	X	X	X	X	X	X			X
3	X	X	X		X	X	X	X	X	X	X	X			X			X
4	X	X	X	X		X			X	X	X	X			X	X		
5		X	X			X			X	X		X		X	X			X
6	X	X	X		X	X	X	X	X	X	X	X	X				X	X
7	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X
10	X	X	X	X	X	X	X		X	X	X	X	X	X	X			X
Preference	23%	33%	43%	22%	27%	52%	27%	17%	57%	35%	30%	35%	30%	25%	45%	23%	18%	58%

Tab. E.6. Individuální hodnocení v **retestu**, pro $SNR = 30$ dB.

Posluchač č.	Kytara						Mozart						Abba					
	Šum kazety			Šum větráku			Šum kazety			Šum větráku			Šum kazety			Šum větráku		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X			X
2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3			X	X	X	X	X	X	X	X	X	X	X	X	X			X
4	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
5		X	X	X	X		X		X		X	X			X	X		
6		X	X	X	X		X	X	X	X	X	X		X		X	X	
7		X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
8	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X
10	X	X	X	X	X	X	X		X	X	X	X	X		X			X
Preference	20%	35%	45%	37%	37%	27%	37%	27%	37%	27%	32%	42%	25%	35%	40%	32%	22%	47%

Curriculum Vitae

Ondřej Rášo

Osobní údaje:

Jméno: Ondřej Rášo
Bydliště: Pod Hůrkou 227E, Hranice na Moravě, 753 01
Datum narození: 13.12.1983
E-mail: raso@feec.vutbr.cz

Vzdělání:

- 2003 - 2006 **VUT FEKT v Brně**, bakalářské studium, Fakulta Elektrotechniky a komunikačních technologií, Obor: Teleinformatika
Bakalářská práce: Bezpečný přístup do webového rozhraní
Ukončeno státní závěrečnou zkouškou (průměr 2,12)
- 2006 -2008 **VUT FEKT v Brně**, magisterské studium, Fakulta Elektrotechniky a komunikačních technologií, Obor: Telekomunikační a informační technika
Diplomová práce: Programování procesorů v 32- nebo 64- bitových operačních systémech
Ukončeno státní závěrečnou zkouškou, (průměr 2,19, cena děkana za diplomovou práci)
- od 2008 **VUT FEKT v Brně**, doktorské studium
Fakulta Elektrotechniky a komunikačních technologií, Obor: Teleinformatika
Dizertační práce: Potlačování šumu v akustických signálech s minimálním zkreslením

Současná pozice:

- od 2012 asistent, Ústav telekomunikací, VUT v Brně
- od 2012 technický pracovník, Ústav telekomunikací, VUT v Brně
- od 2012 administrátor projektu CZ 1.07/2.2.00/28.0098

Specializace:

Číslicové zpracování zvukových signálů, potlačení šumu pozadí v hudebních signálech, měření kvality hudebních signálů.

Účast na vědeckých projektech:

Spoluřešitel:

- **TACR 02020856** (2012-2015), *Aplikovaný výzkum inteligentních systémů pro sledování energetických sítí*. Příjemce: MEGA - Měřicí Energetické Aparáty, s.r.o..
- **FEKT-S-11-17** (2011-2014), *Výzkum sofistikovaných metod číslicového zpracování obrazu a zvuku*. Navrhovatel: prof. Z. Smékal
- **MPO: FR-TI1/075** (2009-2012)- Aplikovaný výzkum a vývoj systémů dálkového měření kvality dodávky elektrické energie.
- **MSM21630513** (2005-2011), *Elektronické komunikační systémy a technologie nových generací* (ELKOM). Řešitele: prof. Z. Raida, prof. K. Vrba, prof. J. Jan

Řešitel pedagogických projektů:

- 2563/2011 Zavedení výsledků výzkumu zpracování akustických signálů v reálném čase do výuky, zahájení.

Vyžádané recenze pro vědecké časopisy a konference:

- Conference on Telecommunications and Signal Processing (TSP)
- Elektrorevue
- International Conference on Teleinformatics (ICT)

Výsledky vědecké činnosti:

- Počet článků v impaktovaných časopisech: 1
- Počet článků v neimpaktovaných časopisech indexovaných v databázích SCOPUS: 1
- Počet příspěvků na konferencích indexovaných ve Web of Science: 8
- Ostatní odborné časopisy a konference: 7
- Celkový počet citací včetně autocitací podle Web of Science: 0
- H-index podle Web of Science: 0

Poslední aktualizace: 10. března 2013