

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

ÚSTAV TELEKOMUNIKACÍ

Ing. Ondřej Rášo

OBJEKTIVNÍ MĚŘENÍ A POTLAČOVÁNÍ ŠUMU V HUDEBNÍM SIGNÁLU

OBJECTIVE ASSESSMENT AND REDUCTION OF NOISE
IN MUSICAL SIGNAL

ZKRÁCENÁ VERZE PH.D. THESIS

Obor: Teleinformatika

Školitel: Ing. MIROSLAV BALÍK, Ph.D.

KLÍČOVÁ SLOVA

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

KEYWORDS

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

OBSAH

ÚVOD.....	4
1 CÍLE DIZERTACE.....	5
2 OBJEKTIVNÍ METODA PRO MĚŘENÍ ÚROVNĚ SLYŠITELNÉHO ŠUMU POZADÍ V HUDEBNÍCH SIGNÁLECH – VLASTNÍ ŘEŠENÍ.....	6
2.1 VÝCHOZÍ ÚVAHY PRO NAVRHOVANOU METODU	6
2.2 PRINCIP NAVRHOVANÉ METODY.....	7
2.2.1 <i>Popis blokového schématu navrhované metody</i>	7
2.2.2 <i>Demonstrační příklad</i>	8
2.3 VÝSLEDNÉ ZHODNOCENÍ NAVRŽENÉ OBJEKTIVNÍ METODY	10
2.3.1 <i>Porovnání výsledků pro hudební a řečové signály</i>	11
2.3.2 <i>Porovnání výsledků pro umělé zvukové signály</i>	12
3 SYSTÉM NA POTLAČENÍ ŠUMU V HUDEBNÍCH SIGNÁLECH – VLASTNÍ ŘEŠENÍ	15
3.1 ADAPTIVNÍ SEGMENTACE.....	15
3.2 ALGORITMUS MAXIMUM GREEDY SELECTION	17
3.2.1 <i>Snížení vlivu spektrálního prosakování</i>	17
3.2.2 <i>Započítání vlivu kmitočtového rozlišení</i>	19
3.3 POSLECHOVÝ EXPERIMENT	24
3.3.1 <i>Popis poslechového experimentu</i>	24
3.3.2 <i>Příprava a průběh poslechového experimentu</i>	25
3.3.3 <i>Výsledky provedeného poslechového experimentu</i>	25
4 ZÁVĚR	27
5 LITERATURA.....	28

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 s rozdílnou 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) [17][18][21]. 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 [17][18][31].

Další skupina zvukových signálů, kde je žádoucí potlačovat šum na pozadí, jsou hudební signály. Hudební signál je takový signál, který je vytvořený hudebními nástroji včetně lidského hlasového ústrojí a současně je tento signál nositelem hudební informace [29][30]. 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.

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 [16]. 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.

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.

1 CÍLE DIZERTACE

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 na 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.

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

2 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. Metoda se zaměřuje na nízký stupeň zarušení šumem na pozadí. 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.

2.1 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 [14]. 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 [16].

- **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 [15] a druhý postup je nazýván výpočet hlasitosti podle Zwickera [16]. Oba postupy výpočtů celkové hlasitosti, lze naznačit stejným vztahem

$$N_L = \int_{b=1}^{24} N_L'(b)db, \quad (2.1)$$

kde $N_L'(b)$ jsou tzv. dílčí hlasitosti v sonech v Barkových pásmech a N_L je výsledná hlasitost v sonech. 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.

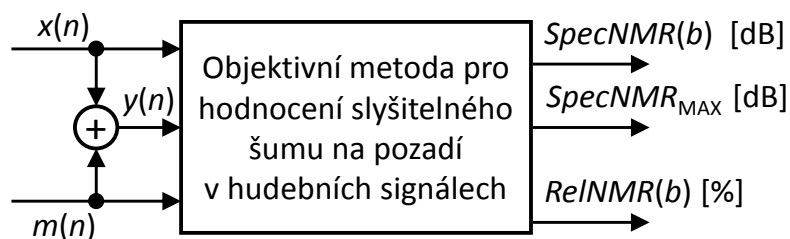
2.2 Princip navrhované metody

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 [20]. Byl zde použit psychoakustický model č. 1, jehož implementaci lze nalézt v [23]. 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). Bárkové pásma jsou tvořeny 24 frekvenčními pásmy, která jsou přibližně logaritmicky 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 [16] totiž uvádí, že výsledná hlasitost složeného zvuku je dána dílčími přírůstky hlasitostí z jednotlivých Barkových pásem. Přesnější popis lze nalézt v [16]. 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 disertační práci.

2.2.1 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. 2.1.



Obr. 2.1 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 $y(n)$

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

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 jsou detailně popsány v disertační práci. Nyní bude následovat pouze jejich stručný popis.

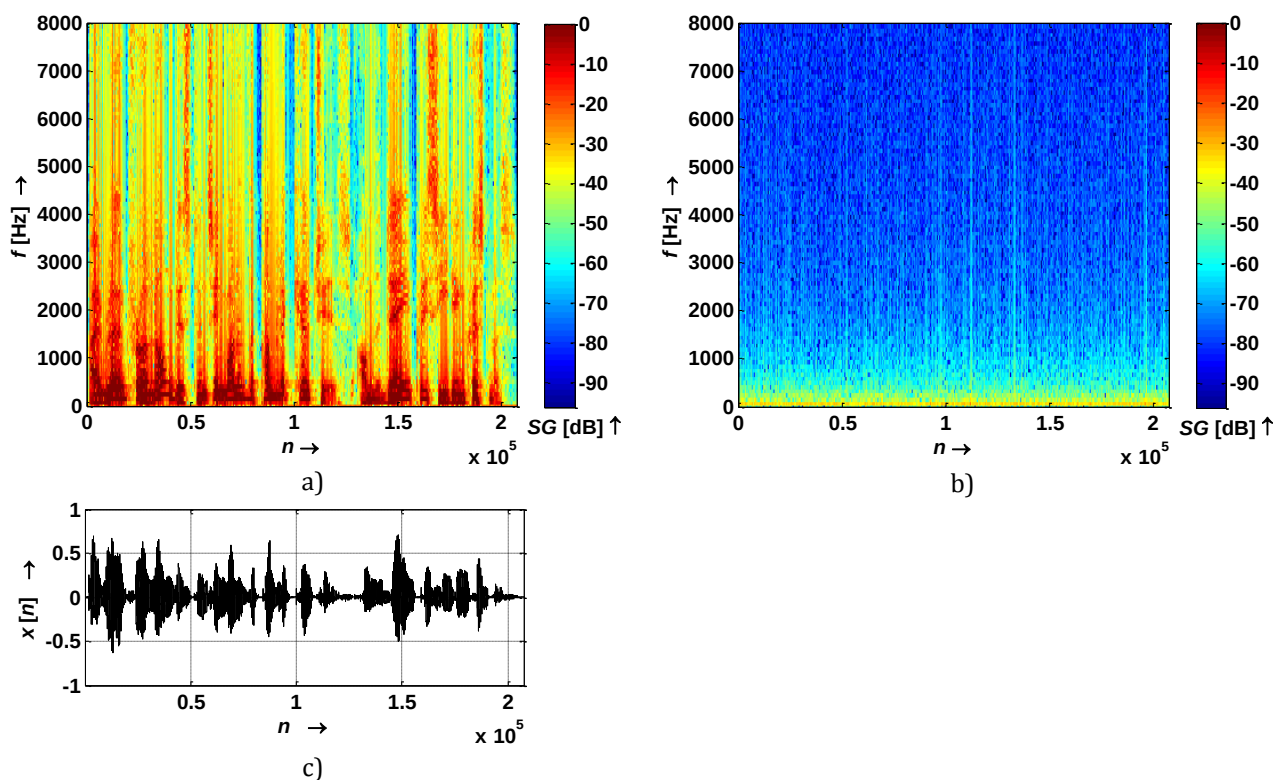
Veličina $SpecNMR(b)$ popisuje specifickou (ve smyslu výběrovou) průměrnou hodnotu odstupe šumu k masce (NMR, Noise to Mask Ratio). Tato veličina udává rozložení tohoto odstupe přes bárková pásma a to pro celý dlouhotrvající analyzovaný zvukový signál. Veličina $SpecNMR(b)$ je unikátní tím, že do výsledných 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.

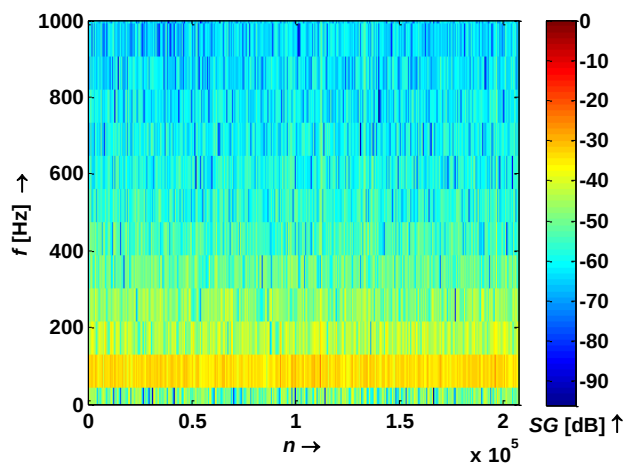
2.2.2 Demonstrační příklad

V tomto ukázkovém příkladu bude demonstrováno možné využití navrhované metody pro měření úrovně slyšitelného šumu pozadí ve zvukových systémech. Užitečný signál popředí je tvořen anglicky hovořícího mluvčím [28]. Spektrogram tohoto zvukového signálu je zobrazen na Obr. 2.2 a) a časový průběh na Obr. 2.2 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. Hodnota SNR je 40 dB.



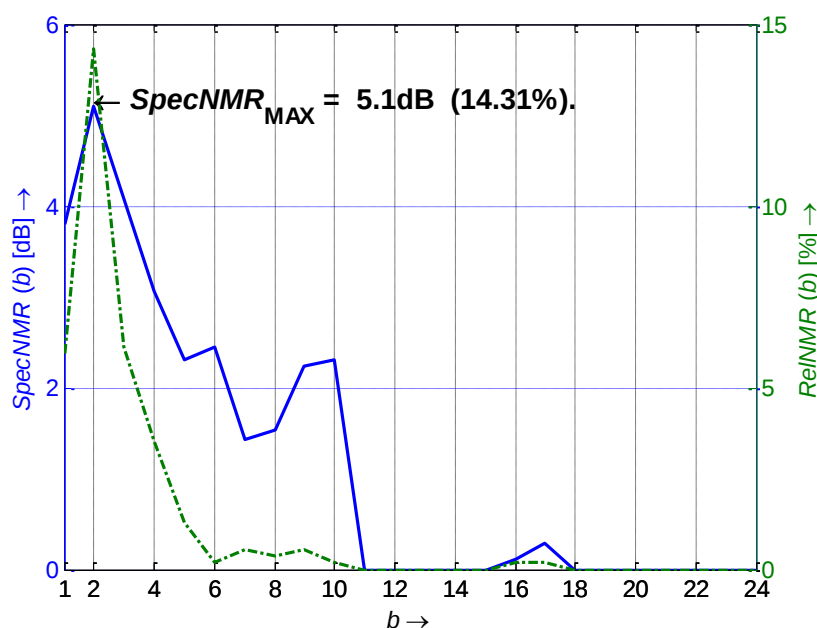
Obr. 2.2 Záznam řeči [28]: a) spektrogram, b) spektrogram šumu pozadí, šum audio pásky, c) časový průběh.

Šumem pozadí je šum audio pásky, viz spektrogram na Obr. 2.2 c). Většina výkonu tohoto šumu je koncentrována v pásmu v okolí 100 Hz, viz detailní spektrogram na Obr. 2.3.



Obr. 2.3 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. 2.4.



Obr. 2.4 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. 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. 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 $SpecNMR_{MAX}$.

Graf z Obr. 2.4 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. 2.2 c), a celková délka “tichých“ úseků v promluvě, Obr. 2.2 a-b).

2.3 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. V druhé fázi byly získány objektivní hodnoty prahů slyšitelnosti z testovaných metod, testované metody byly metoda PEAQ [25] a můj návrh metody z kapitoly 2.2. 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.

2.3.1 Porovnání výsledků pro hudební a řečové signály

V této kapitole budou porovnány dvě objektivní metody pro predikci prahu slyšitelnosti šumu pozadí.. Porovnávána bude metoda PEAQ a navržená metoda.. Testovacím kritériem pro toto porovnání bude střední kvadratická chyba MSE (mean squared error) definována

$$MSE(SNR_{OBJ}) = E \left[(SNR_{SBJ} - SNR_{OBJ})^2 \right], \quad (2.3)$$

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. 2.1 jsou zobrazeny hodnoty SNR_{SBJ} a v Tab. 2.2 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. 2.1 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. 2.2 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 popředejí 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.

2.3.2 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. Porovnávány budou metody PEAQ a nově navržená objektivní metoda. Testovacím kritériem pro toto porovnání bude střední kvadratická chyba MSE (2.3). Subjektivní prahy slyšitelnosti šumu pozadí SNR_{OBJ} získané z této 2. části provedeného experimentu jsou zobrazeny v Tab. 2.3.

Tab. 2.3 Tabulka s naměřenými subjektivními hodnotami SNR_{OBJ} prahu slyšitelnosti šumu pozadí pro 2. část experimentu.

Poslech				
	10 tónů	Růžový šum	1 kHz tón	Bílý šum
	SNR_{SBJ} [dB]	SNR_{SBJ} [dB]	SNR_{SBJ} [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. 2.4 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. 2.4 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 signály a ne na umělé zvukové signály [25]. 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 20dB. 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. Realizovaný poslechový experiment ale dokázal, že toto zkreslení, způsobené šumem na pozadí je nevnímatelné, 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.

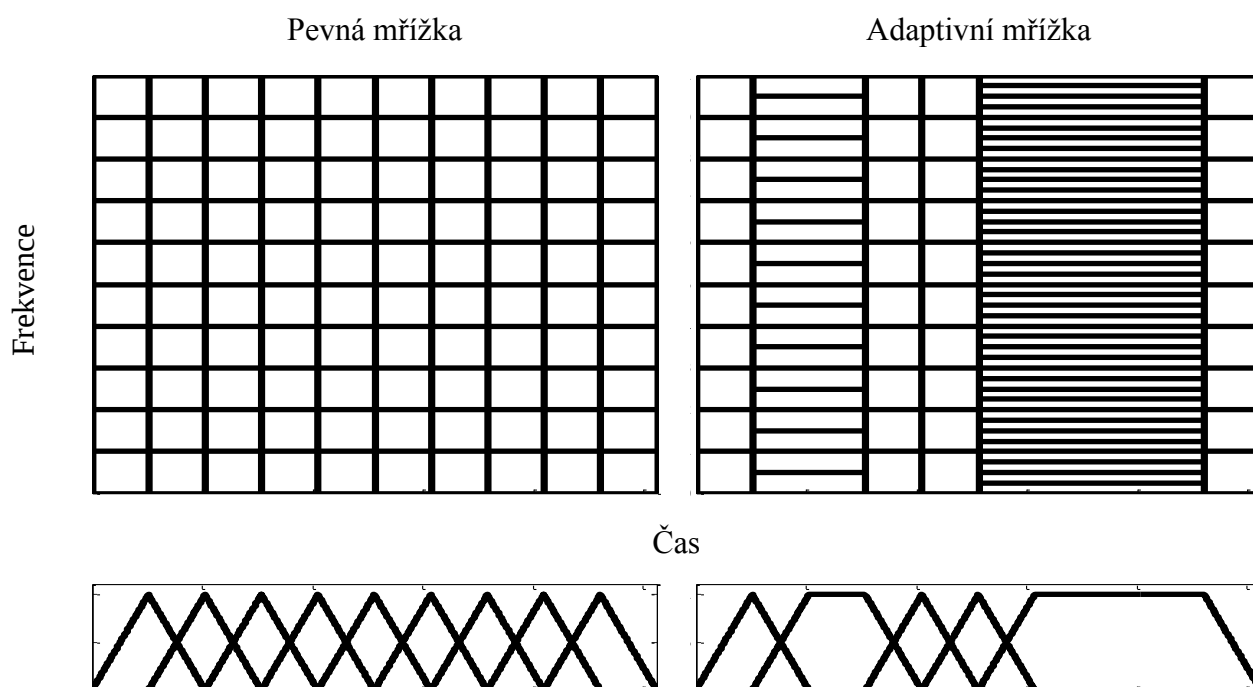
Naproti tomu odhad subjektivního vjemu slyšitelnosti šumu pozadí podle nově navržené metody dosahuje relativně přesných výsledků. V 11 z celkových 12 případů, byl proveden odhad s chybou maximálně 5 dB. V jednom případě se navržená metoda dopustila velmi špatného odhadu a to s chybou 15 dB. To bylo velmi 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 než pro případ metody PEAQ, viz $MSE_{NMR} = 37.5$ a $MSE_{PEAQ} = 126.6$.

3 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. Toto řešení v sobě kombinuje techniky adaptivní segmentace a techniky prahování.

3.1 Adaptivní segmentace

Myšlenkou adaptivní segmentace je rozdělit dlouho trvající signál na segmenty proměnné délky, které je možno pro účely následujícího zpracování považovat za “stacionární“. Na Obr. 3.1 jsou schématicky znázorněny rozdíly mezi pevnou a adaptivní segmentací z pohledu rozdílného časo-frekvenčního rozlišení [24]. 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 [19].



Obr. 3.1 Rozdíly v časo-frekvenčním rozlišení mezi pevnou a adaptivní segmentací [27].

V následujícím textu bude použit symbol N pro délku nejkratšího adaptivního segmentu a symbol L bude použit pro aktuální délku adaptivního segmentu. Překryv mezi sousedními segmenty bude mít vždy velikost $N/2$.

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

$$y_{\text{SEG}}(n, p, q) = \begin{cases} w_A(n, q) y\left(n + (p-1)\frac{N}{2}\right) & \text{pro } n = 0, 1, \dots, (q-1)\frac{N}{2} + N - 1 \\ 0 & \text{pro } n \neq 0, 1, \dots, (q-1)\frac{N}{2} + N - 1 \end{cases}, \quad (3.1)$$

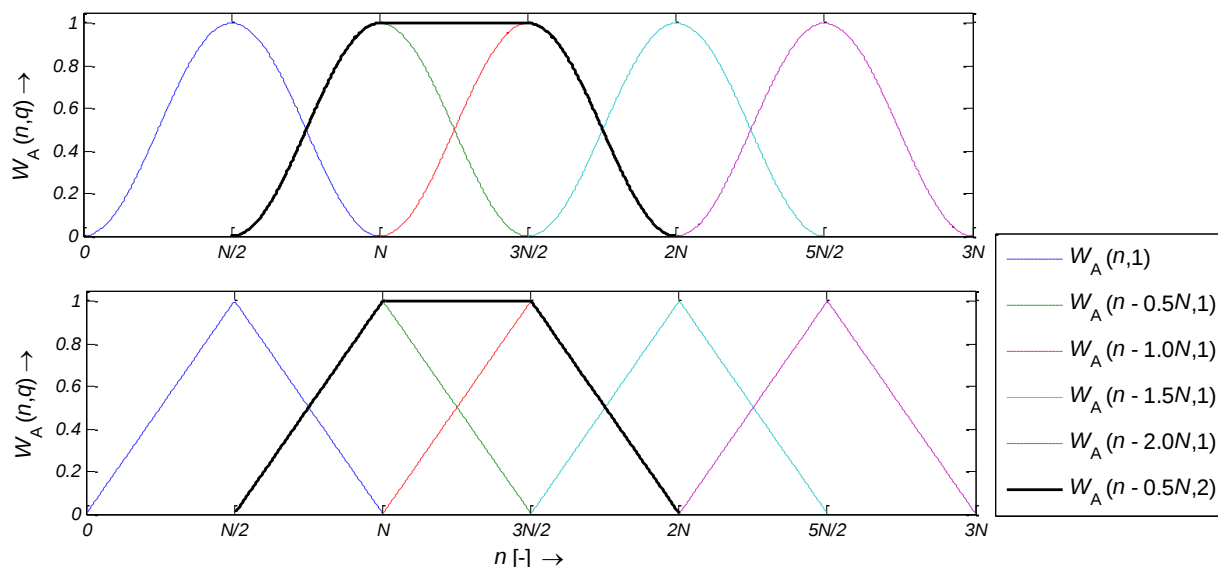
kde p je časový index toho segmentu, resp. posunutí toho segmentu vůči začátku $y(n)$ a symbol q vyjadřuje výslednou délku toho segmentu. Mezi sousedními segmenty $y_{\text{SEG}}(n, p, q)$ dochází k překryvu o délce $N/2$. Symbol $w_A(n, q)$ je použit pro adaptivní váhovací okno o délce q s počátkem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

$$w_A(n, q) = \sum_{l=0}^{q-1} w\left(n - l\frac{N}{2}\right) \quad (3.2)$$

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

$$w(n) = \begin{cases} 0.5 \left(1 - \cos\left(2\pi \frac{n}{N}\right)\right), & \text{pro } n = 0, \dots, N-1 \\ 0, & \text{pro } n \neq 0, \dots, N-1 \end{cases} \quad (3.3)$$

V případě jiného typu okna, než je trojúhelníkové okno, je význam parametru q stejný, viz Obr. 3.2 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. 3.2 Hannova (nahore) a trojúhelníková (dole) adaptivní váhovací okna.

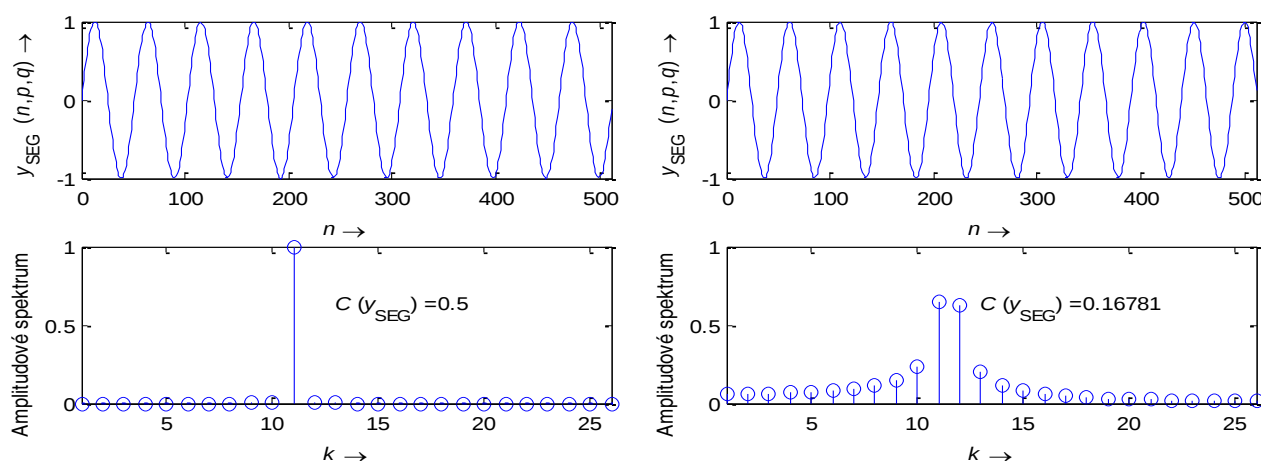
3.2 Algoritmus Maximum greedy selection

V této kapitole bude představeno vlastní řešení adaptivní segmentace, které jsem nazval Maximum greedy selection (Maximální nenasytný výběr).

3.2.1 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.

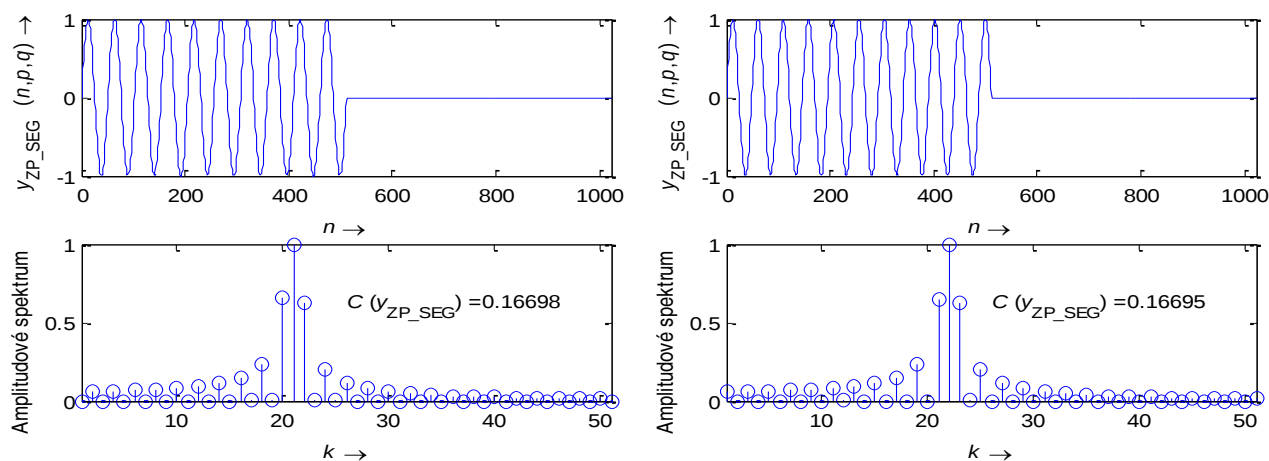
Na Obr. 3.3 jsou zobrazeny časové průběhy dvou segmentů $y_{SEG}(n,p,q)$ o délce 500 vzorků. U prvního segmentu (sloupec vlevo) nedochází k spektrálnímu prosakování. Lze vidět, že jeho amplitudové spektrum je vyjádřeno jediným spektrálním koeficientem ($k = 11$). Hodnota koeficientu lokální spektrální špičatosti $C(y_{SEG})$ 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 $C(y_{SEG})$ 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 vliv spektrálního prosakování je u každého ze segmentů jinak veliký.



Obr. 3.3 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 (3.1) 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.

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. 3.3, jen po doplnění nulovými vzorky, je zobrazena na Obr. 3.4. Lze vidět, že rozdíly v hodnotách koeficientů C pro tyto nulou doplněné segmenty $y_{ZP_SEG}(n,p,q)$ jsou velmi malé.



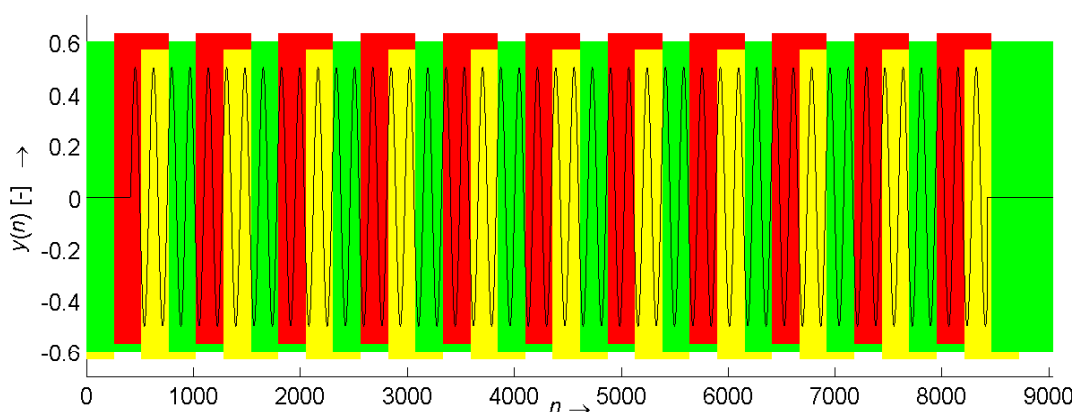
Obr. 3.4 Snížení vlivu jevu spektrálního prosakování na koeficientu C pomocí techniky doplnění nulami.

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

$$y_{ZP_SEG}(n,p,q) = \sum_{n=0}^{2L} (y_{SEG}(n,p,q) + o(n,2L)). \quad (3.4)$$

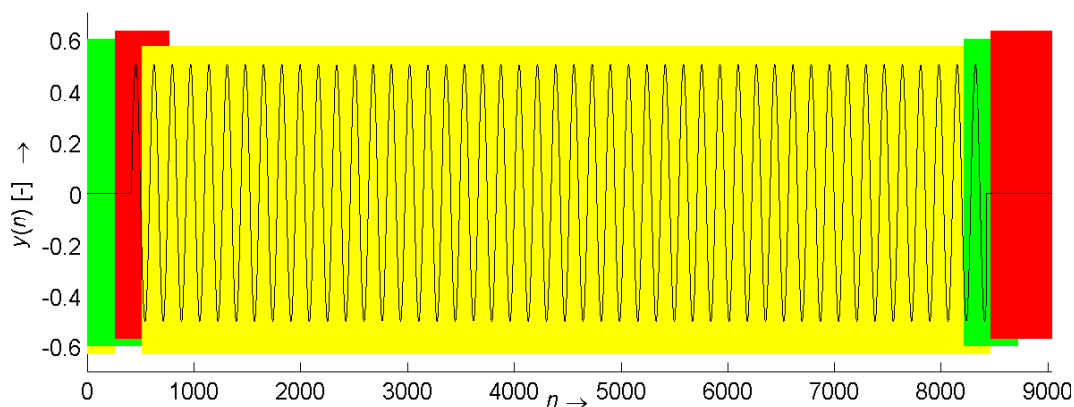
kde $o(n,2L)$ je nulový vektor délky $2L$ a $y_{ZP_SEG}(n,p,q)$ má proto také délku $2L$.

Na Obr. 3.5 je zobrazen příklad adaptivní segmentace podle původního algoritmu Greedy selection, tj. bez doplnění nulovými vzorky. Segmentován je zde stacionární (harmonický) signál s kmitočtem 46,875 Hz při vzorkovacím kmitočtu $f_s = 8$ 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. 3.5 Adaptivní segmentace podle původního algoritmu Greedy selection, tj. bez doplnění nulovými vzorky.

Pokud dojde k doplnění nulovými vzorky podle (3.4), tak se sníží vliv spektrálního prosakování na hodnoty koeficientu C a adaptivní segmentace bude pro tento příklad fungovat správně, viz Obr. 3.6.



Obr. 3.6 Adaptivní segmentace s doplněním nulovými vzorky.

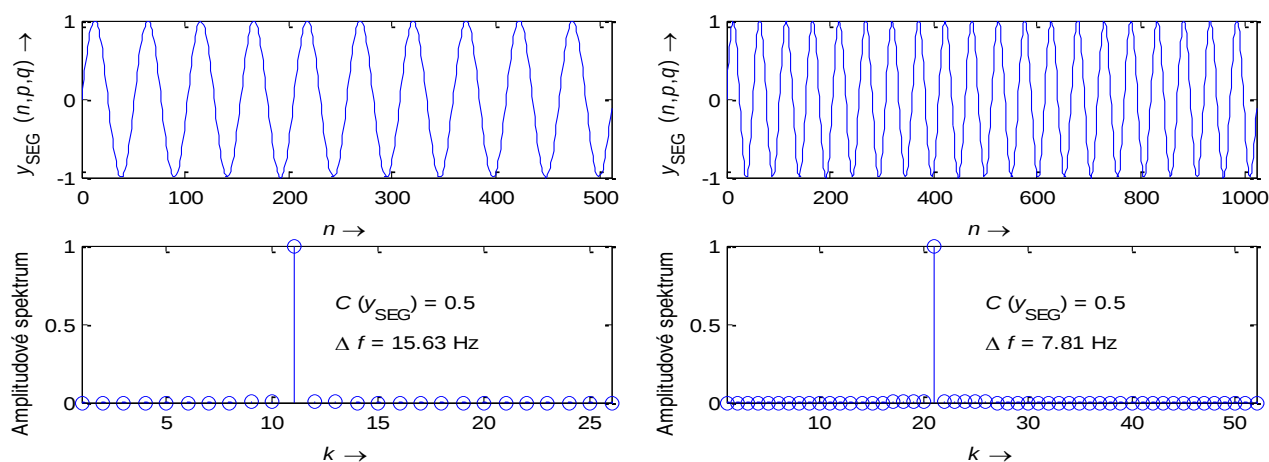
3.2.2 Započítání vlivu kmitočtového rozlišení

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

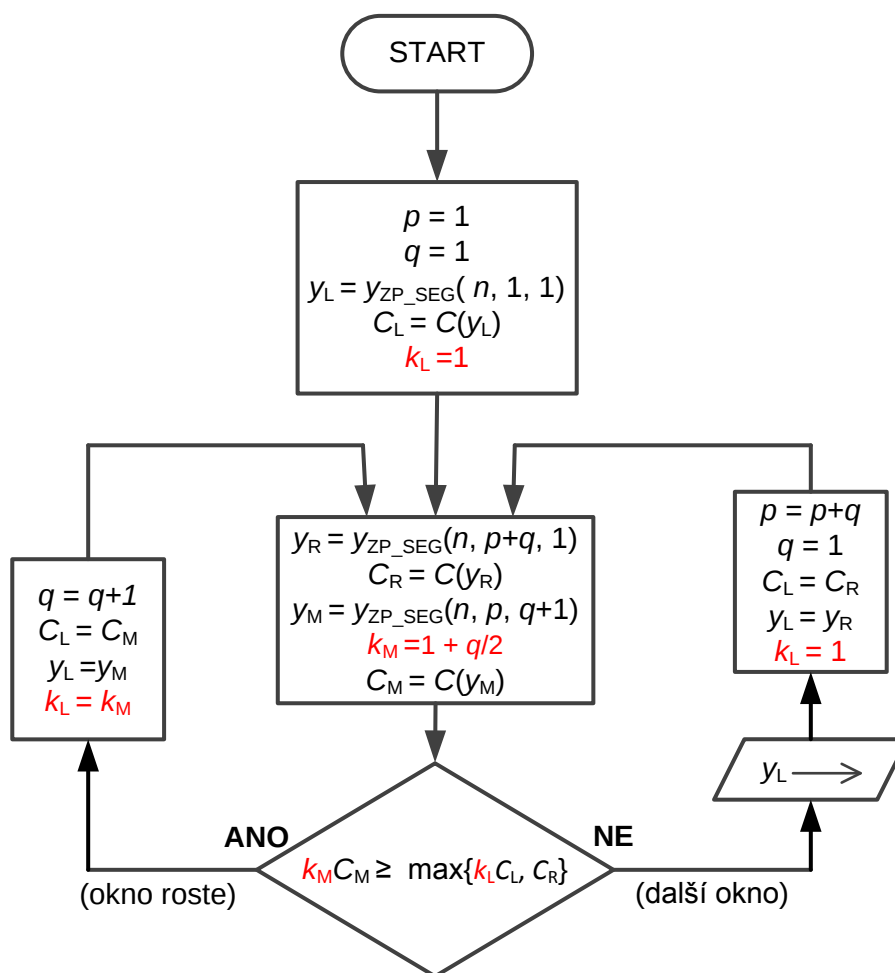
DFT do svých rozhodovacích procesů. Tuto modifikaci jsem pojmenoval jako *Maximum greedy selection* (maximální nenasytný výběr).

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. 3.7 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 $f_s = 8$ 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. 3.7 Časové průběhy a amplitudové spektra náhodně vybraného segmentu o délce $L = 512$ vzorků (vlevo), a $L = 1024$ vzorků (vpravo) při $f_s = 8$ kHz.

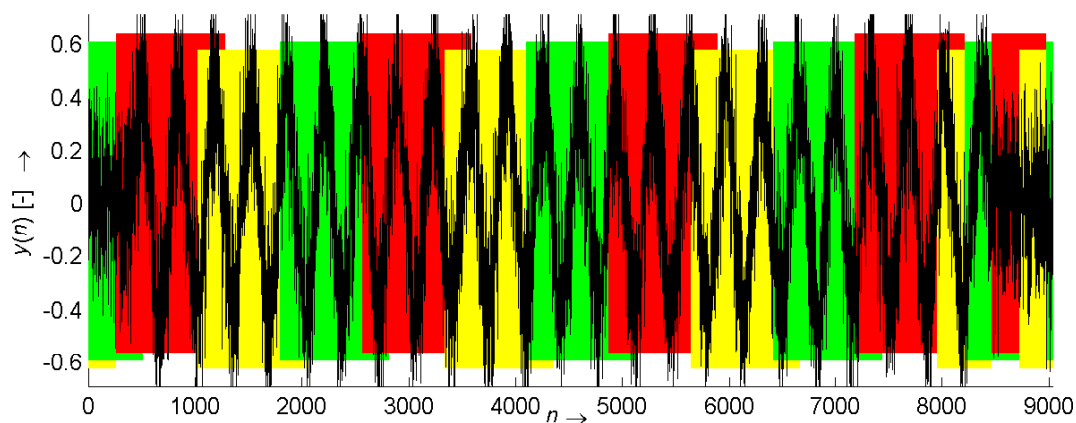
Výchozí úvaha při návrhu algoritmu *Maximum greedy selection* byla taková, že vliv rozdílných délek L segmentů y_L , y_L a y_M by se měl promítnout do rozhodovacích procesů o sloučení do výsledného segmentu y_M . A to z důvodu, že i když neroste hodnota koeficientu lokální spektrální špičatosti C , tak 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. 3.8. Modifikace jsou zobrazeny červeným písmem. Dalším rozdílem oproti původnímu algoritmu *Greedy selection* je ten, že hodnoty koeficientů C_L , C_L a C_M jsou vypočteny ze segmentů doplněných nulovými vzorky.



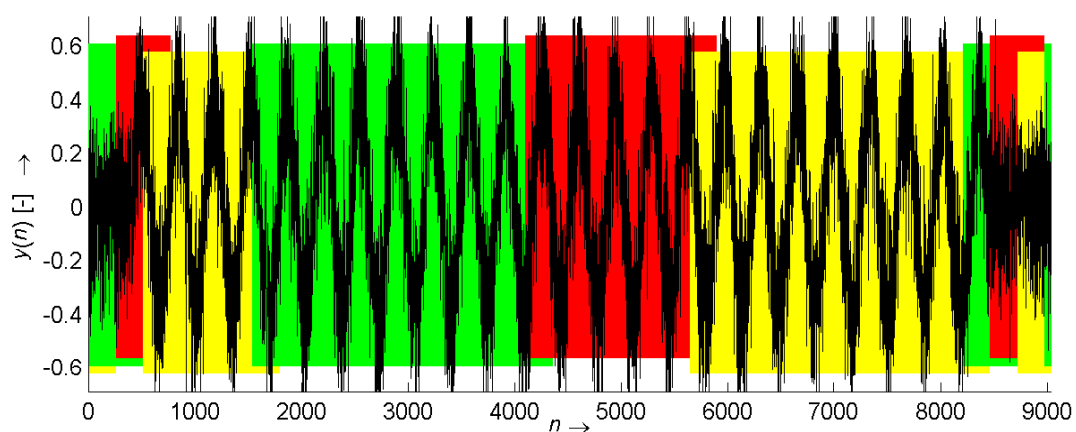
Obr. 3.8 Algoritmus Maximum greedy selection. Symbol $y_L \rightarrow$ značí uložení y_L do výstupní paměti pro další zpracování. Červeným písmem jsou vyznačeny změny oproti původní variantě algoritmu Greedy selection.

Na následujících obrázcích Obr. 3.9 až Obr. 3.11 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. 3.9 je zobrazena adaptivní segmentace podle algoritmu Greedy selection bez doplnění nulovými vzorky. Tedy ve své původní variantě. Na následujícím obrázku, Obr. 3.10, 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. 3.11, 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.

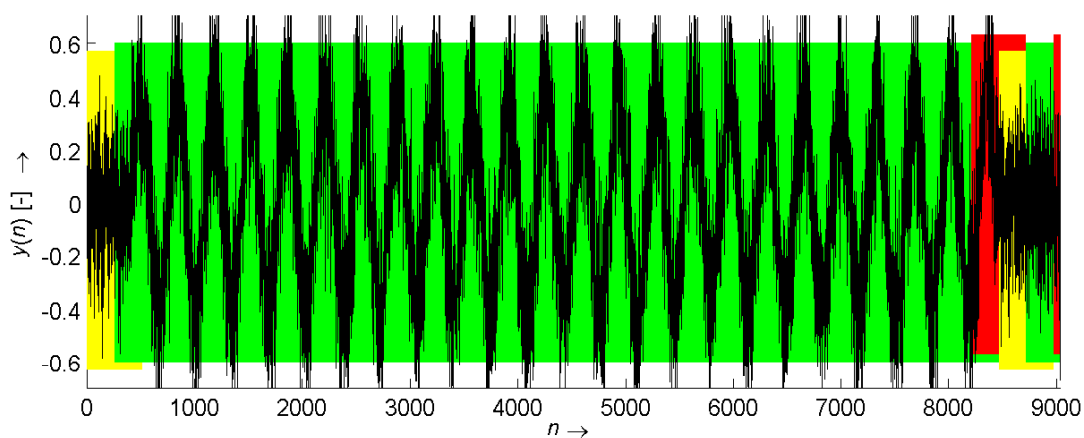
Společnými parametry pro všechny tři příklady z Obr. 3.9 až Obr. 3.11 jsou: kmitočet harmonického signálu $f = 46.875$ Hz, vzorkovací kmitočet $f_s = 8$ kHz, $SNR = 3$ dB, $N = 512$. Dále bylo použito Hannovo okno (3.3) s překryvem $N/2 = 216$ vzorků.



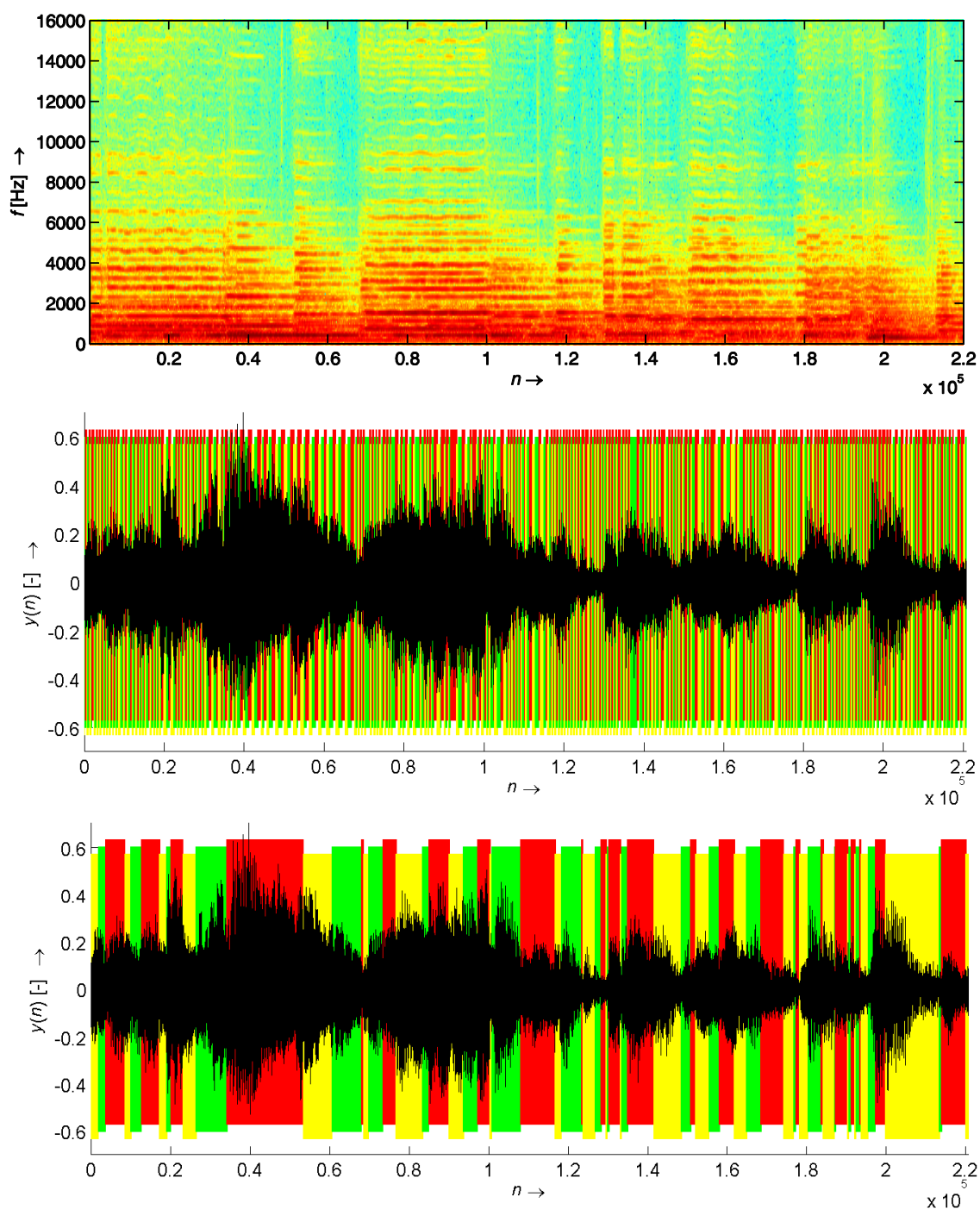
Obr. 3.9 Greedy selection , adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, bez doplnění nulovými vzorky.



Obr. 3.10 Greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, s doplněním nulovými vzorky.



Obr. 3.11 Maximum greedy selection, adaptivní segmentace harmonického signálu zarušeného aditivním bílým šumem, s doplněním nulovými vzorky.



Obr. 3.12 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. 3.12 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).

3.3 Poslechový experiment

3.3.1 Popis poslechového experimentu

Cílem tohoto experimentu je zjistit, zda navržený systém pro potlačení šumu na pozadí s adaptivní segmentací dosahuje lepších výsledků, než stávající systémy. Všechny testované systémy využívají techniky prahování koeficientů DFT. 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 (3.1) 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 (3.1) s hodnotami p a q nastavenými podle algoritmu Greedy selection.. 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 (3.1) s hodnotami p a q nastavenými podle algoritmu Maximum greedy selection. 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 šumu pozadí. Délka maximální délka adaptivního okna L není nijak omezena. Ve vztahu (3.2) bylo použito Hannovo okno (3.3).

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. Rozsah hodnot SNR je 10, 20 a 30 dB.

Vliv zkreslení užitečného signálu na výslednou subjektivní kvalitu je velmi významný. Často bývá významnější než vliv přítomnosti šumu pozadí na tuto kvalitu [18]. 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][32].

3.3.2 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 [26].

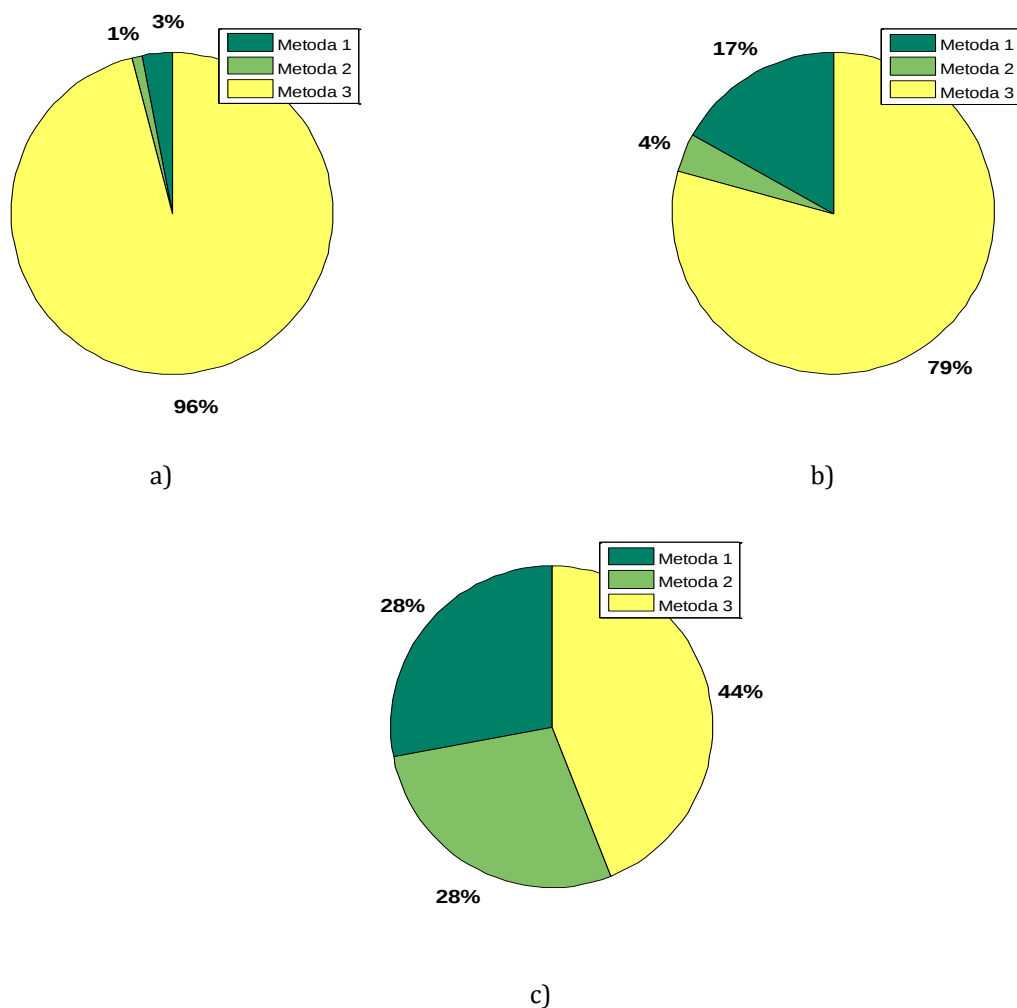
3.3.3 Výsledky provedeného poslechového experimentu

Tabulku s výsledným preferenčním hodnocením lze nalézt v Tab. 3.1. V jednotlivých sloupcích jsou vypsány 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 nejkvalitnější 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. 3.1 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. 3.13a-c) jsou přehledně v koláčových grafech graficky vyneseny procentuální preferenční hodnocení pro rozdílné hodnoty SNR z Tab. 3.1.



Obr. 3.13 Koláčkové 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. 3.2. V této tabulce jsou navíc oproti Tab. 3.1 uvedeny preferenční hodnocení pro jednotlivé kombinace užitečného signálu a šumu pozadí včetně dělení na měření v testu a retestu.

Tab. 3.2 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

4 ZÁVĚR

Disertační práce se zabývala objektivním měřením a potlačováním šumu na pozadí v hudebních signálech. Po úvodní analýze současného stavu této problematiky následoval popis vlastního návrhu objektivní metody pro tento typ měření. Tato metoda se zaměřuje na velmi malé zarušení šumem na pozadí, tedy na ty případy, kdy šum na pozadí je na hranici vnímatelnosti posluchači. Šum na pozadí je zde slyšitelný jen v některých částech zvukového signálu, např. v tichých pasážích, na vyšších frekvencích a podobně. Dále tato metoda poskytuje prostředky k lokalizaci slyšitelného šumu pozadí v časo-frekvenční oblasti a k objektivnímu stanovení jeho velikosti. Protože se však tato metoda nezaměřuje pouze na hudební signály, ale také na různé umělé typy zvukových signálů, jako je například bílý šum, harmonický signál a směs harmonických signálů a podobně. Tak lze tuto metodou dále využít k odhadu slyšitelnosti jednoho umělého zvukového signálu na pozadí jiného, opět umělého zvukového signálu.

Přínos této navržené metody spočívá především v možnosti objektivně měřit slyšitelný šum na pozadí přesnějším způsobem, než který dosahují v současnosti používané metody. Dále je možno provádět měření slyšitelnosti i pro případ 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 část disertační práce se zabývala návrhem systému na potlačení šumu pozadí v hudebních signálech. Tento systém měl být navržen s ohledem na psychoakustické aspekty vnímání hudebních signálů posluchači. Navržený systém je charakteristický zejména tím, že k segmentaci dlouho-trvajícího signálu do kratších segmentů využívá tzv. adaptivní segmentační schéma. Tedy typ segmentace, kdy vstupní dlouhotrvající signál je segmentován do segmentů proměnlivé délky s cílem nalézt takové úseky, které lze pro následné zpracování považovat za stacionární úseky, v původně nestacionárním zvukovém signálu.

Vlastní přínos při návrhu tohoto systému na potlačení šumu na pozadí spočíval především v modifikaci algoritmu pro adaptivní segmentaci *Greedy selection*. Byly zavedeny dvě úpravy tohoto algoritmu s cílem zlepšit jeho vlastnosti. Provedenými poslechovými testy bylo dokázáno, že uvedené modifikace algoritmu *Greedy selection* zvyšují poslechovou kvalitu těchto výstupních signálů. Posluchači tyto signály v průměru více preferovali než ostatní testované signály.

Všechny stanovené cíle z kapitoly 1 považuji tedy za splněné. Dílčí výsledky práce byly v dostatečné míře prezentovány vědecké veřejnosti v publikacích [1] - [13].

5 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] ČSN 01 1600 *Akustika – Terminologie*. Český normalizační institut, Praha 2003. 80 s.
- [15] ČSN ISO 532 (011602), *Metoda pro výpočet hlasitosti*, Český normalizační institut, Praha, březen 1994. 24 s.
- [16] FASTL, H., ZWICKER E., *Psychoacoustic: Facts and Models*. 3rd ed. Berlin-Heidelberg-New York: Springer, 2006. 462 s. ISBN 978-3-540-23159-2.
- [17] 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.
- [18] 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.
- [19] 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
- [20] 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.
- [21] JAN, J.; *Číslicová filtrace, analýza a restaurace signálů*. 2nd edition. Brno, Vuitium, 2002. 427 s. ISBN: 80-214-1558-4
- [22] JANSEN, M.: *Noise reduction by wavelet thresholding*. New York : Springer, 2001. 191 s. ISBN 0-387-95244-6.
- [23] PETITCOLAS, F.; *MPEQ for Matlab*, software, poslední aktualizace prosinec 2012, <http://www.petitcolas.net/fabien/software/mpeg/>
- [24] PROAKIS, J. G., MANOLAKIS D. G., *Digital Signal Processing*. 4th ed. New Jersey: Parson Education, 2007. 1084 s. ISBN 0-13-187374-1.
- [25] Recommendation ITU-R BS.1387-1 (PEAQ). *Method for objective measurements of perceived audio quality*, 2001.
- [26] RUDROY, 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
- [27] RUDROY, 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
- [28] *Sound Quality Assessment Material recordings for subjective tests*, Users' handbook for the EBU SQAM CD, Geneva, 2008, [www: http://tech.ebu.ch/publications/sqamcd](http://tech.ebu.ch/publications/sqamcd)
- [29] SYROVÝ, V. *Hudební akustika*. Praha : Akademie múzických umění v Praze, 2003. 427 s. ISBN 80-7331-901-2.
- [30] SYROVÝ, V. *Hudební zvuk*. Praha : Akademie múzických umění v Praze, 2009. 303 s. ISBN 80-7331-161-2.
- [31] VASEGHI, S. V.: *Advanced Digital Signal Processing and Noise Reduction*. 2nd edition. New York : Wiley, 2000. 473 s. ISBN 0-471-62692-9.
- [32] 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

Významné v práci necitované publikace autora

- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.

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:

- **TAČR 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

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.