

VĚDECKÉ SPISY VYSOKÉHO UČENÍ TECHNICKÉHO V BRNĚ

*Edice PhD Thesis, sv. 425*

*ISSN 1213-4198*

*thesis*  
**S**  
IS

*Ing. Petr Sysel*

**Jednokanálová metoda  
zvýraznění řeči  
s využitím waveletové transformace  
ve spektrální oblasti**

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ  
Fakulta elektrotechniky a komunikačních technologií  
Ústav telekomunikací

**Ing. Petr Sysel**

**JEDNOKANÁLOVÁ METODA ZVÝRAZNĚNÍ ŘEČI S VYUŽITÍM  
WAVELETOVÉ TRANSFORMACE VE SPEKTRÁLNÍ OBLASTI**

ONE-CHANNEL SPEECH ENHANCEMENT METHOD BASED ON  
WAVELET TRANSFORM IN SPECTRAL DOMAIN

ZKRÁCENÁ VERZE PH.D. THESIS

Obor: Teleinformatika  
Školitel: Prof. Ing. Zdeněk Smékal, CSc.  
Oponenti: Prof. Ing. Andrej Lúč, CSc.  
Ing. Robert Vích, DrSc.  
Datum obhajoby: 7.9.2007

## **KLÍČOVÁ SLOVA**

waveletová transformace, výkonová spektrální hustota, spektrální odečítání, zvýraznění řeči

## **KEYWORDS**

wavelet transform, power spectral density, spectral subtraction, speech enhancement

## **DISERTAČNÍ PRÁCE JE ULOŽENA:**

Ústav telekomunikací

Fakulta elektrotechniky a komunikačních technologií

Purkyňova 118

612 00 Brno

Czech Republic

© Petr Sysel, 2007

ISBN 978-80-214-3503-2

ISSN 1213-4198

# OBSAH

1	PŘEHLED SOUČASNÉHO STAVU PROBLEMATIKY . . . . .	5
1.1	Jednokanálové metody . . . . .	5
1.2	Vícekanálové metody . . . . .	6
2	CÍLE DOKTORSKÉ PRÁCE . . . . .	7
3	ČASOVĚ-KMITOČTOVÁ REPREZENTACE SIGNÁLU . . . . .	7
3.1	Rozdělení metod časově-kmitočtové analýzy . . . . .	7
3.2	Waveletová transformace . . . . .	8
3.2.1	<i>Spojité waveletové transformace</i> . . . . .	8
3.2.2	<i>Diskrétní waveletové transformace</i> . . . . .	10
3.3	Prahování koeficientů waveletové transformace . . . . .	10
3.4	Odhad velikosti prahů . . . . .	11
4	ODHAD VÝKONOVÉ SPEKTRÁLNÍ HUSTOTY RUŠENÍ . . . . .	12
4.1	Základní vlastnosti metod spektrální analýzy . . . . .	12
4.2	Klasické metody odhadu výkonové spektrální hustoty . . . . .	13
4.3	Parametrické metody odhadu . . . . .	15
5	METODA SPEKTRÁLNÍHO ODEČÍTÁNÍ . . . . .	16
5.1	Chyba rekonstrukce signálu metodou spektrálního odečítání . . . . .	17
5.2	Snížení rozptylu odhadu pomocí waveletové transformace . . . . .	18
5.3	Energetický detektor řečové aktivity . . . . .	19
6	ZÁVĚR . . . . .	20
	LITERATURA . . . . .	22
	CURRICULUM VITAE . . . . .	25
	ABSTRACT . . . . .	26



# 1 PŘEHLED SOUČASNÉHO STAVU PROBLEMATIKY

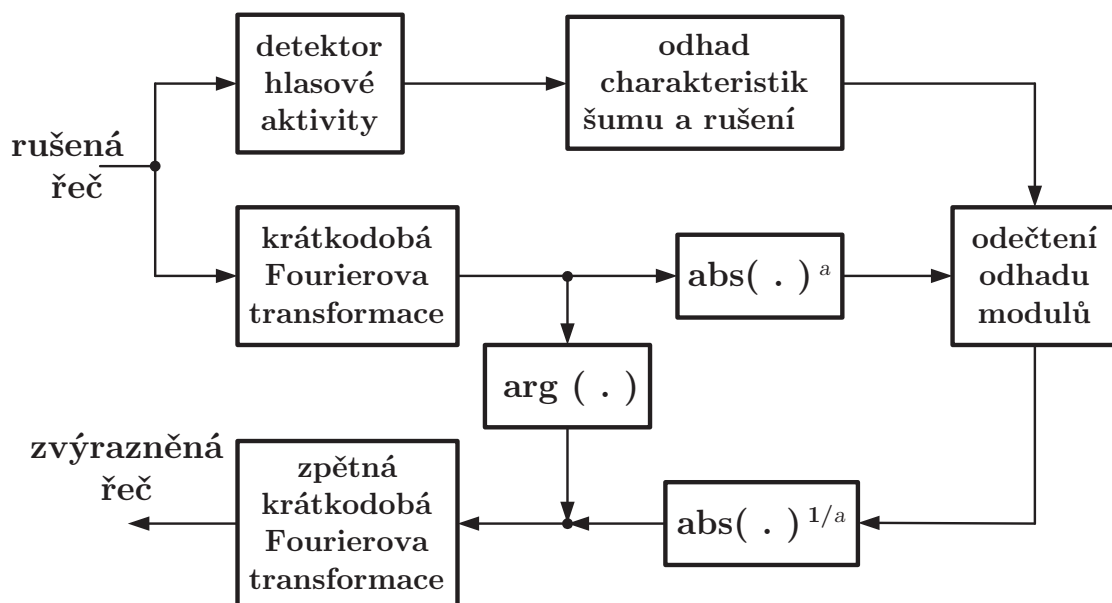
Vstupním prvkem do každého systému zpracování řeči je mikrofon, který v reálných akustických prostředích snímá i okolní hluk. Pokud není vstupní akustický signál ihned převeden do číslicové podoby, je dále degradován širokopásmovými šumy pasivních i aktivních prvků přenosového zařízení, vnějším elektromagnetickým polem, přeslechy a impulsním rušením různého původu, které se přes parazitní kapacitní a induktivní vazby velmi snadno přidávají k užitečnému signálu. V případě systémů pro dálkový přenos řeči je řečový signál dále degradován zkreslením způsobeným zdrojovým kóděrem a zdrojovým dekodérem. Důsledkem všech těchto zkreslení může být to, že na přijímací straně je řečový signál málo nebo zcela nesrozumitelný. Metody zvýrazňování řeči lze rozdělit podle mnoha hledisek: podle způsobu jakým metody modelují vytváření řeči, podle způsobu jakým modelují vnímání zvuku lidským sluchem, podle předpokládaného charakteru rušení, podle počtu použitých mikrofonů apod.

## 1.1 JEDNOKANÁLOVÉ METODY

Většina jednokanálových metod pro zlepšení srozumitelnosti řeči je založena na spektrálním odečítání [15]. Jejich největší nevýhodou je, že jsou schopny separovat (v tomto případě spektrálně odečíst) pouze hluk se stacionárním charakterem. Bohužel tento předpoklad je v praxi těžko splnitelný. Motory nemají konstantní otáčky, dešť neprší stejnou intenzitou, telefon nezvoní jediným tónem apod. Pokud předpokládáme přítomnost nestacionárního rušení, je nutné metodu spektrálního odečítání doplnit o detektor řečové aktivity (VAD) [10, 16] a odhad charakteristiky rušení provádět pouze v tichých úsecích mezi promluvami.

Podstatnou nevýhodou metody spektrálního odečítání je také přítomnost „zbytkového hudebního šumu“ v obnovené řeči. Díky rozptylu odhadu výkonové spektrální hustoty, který se mění podle použité metody odhadu, vzniknou po odečtení v modulovém spektru rekonstruovaného signálu úzkopásmové špičky. Úroveň těchto špiček není velká, maximálně dosahuje úrovně modulu šumu bez přítomnosti signálu, ale vyskytují se na náhodných kmitočtech. Po převedení do časové oblasti se špičky projeví jako součet několika harmonických signálů – tónů, které trvají po dobu jednoho rámce a jejichž kmitočet se náhodně mění. Jednou z možností jak snížit úroveň hudebního šumu je snížit rozptyl odhadu výkonové spektrální hustoty průměrováním periodogramů [2]. Další metodou pro snížení hladiny hudebního šumu je použití nelineárního spektrálního odečítání, kde odhad výkonové spektrální hustoty je při odečítání násoben kmitočtově závislou konstantou [18]. Některé metody také modifikují proměnné odečítací parametry podle maskovacích vlastností lidského sluchu [20].

Mezi další jednokanálové metody lze zařadit metodu adaptivního potlačení šumu a metodu filtrace adaptivním hřebenovým filtrem [8, 21], metoda tzv. RASTA filtrace



Obr. 1.1: Blokové schéma zobecněné metody spektrálního odečítání.

časových trajektorií jednotlivých spektrálních složek modulu spektrogramu [7]. Mezi nové metody patří i metoda mapování spektrogramu vyvinutá na Ústavu telekomunikací [13].

## 1.2 VÍCEKANÁLOVÉ METODY

Nevýhodou těchto metod je nutnost použít více mikrofonů, náročnost jejich optimálního umístění v daném prostoru a vysoká výpočetní náročnost. Počet mikrofonů musí být větší nebo roven počtu nezávislých snímaných zdrojů signálu.

Mezi tyto metody patří metoda řízení směru maximálního příjmu mikrofonního pole [6]. Takzvaná *beamforming* technika využívá vlastnosti šíření zvuku, kdy signály dopadají na senzory s různými fázemi. Nevýhodou této techniky je nutnost použití velkého počtu mikrofonů. Výhodou je skutečnost, že řečový signál není degradován, protože se jedná o lineární operaci se signálem.

Problematiku obnovení časových průběhů akustických signálů nezávislých zdrojů z jejich lineární kombinace řeší i tzv. Blind Source Separation (BSS) technika [25, 5], která je odvozena z techniky Independent Component Analysis (ICA) [9]. Výhodou je, že problematika rekonstrukce je řešena čistě matematickými metodami [9], takže nejsou potřebné žádné informace o charakteru signálů nebo rušení. Pro jednoznačné obnovení nezávislých zdrojů je nutné splnit podmínku, že počet mikrofonů je větší nebo roven počtu nezávislých zdrojů. Dále je nutné v reálné situaci uvažovat konečnou rychlost šíření zvuku a možnost vzniku ozvěn.

## 2 CÍLE DOKTORSKÉ PRÁCE

Cílem doktorské práce je analyzovat možnosti použití waveletové transformace [3, 19, 24] pro zpracování řečových signálů. Za hluk pozadí bude považován aditivní širokopásmový šum (příkladem může být bílý Gaussův šum, hluk na ulici či křižovatce, hluk strojů v tovární hale apod.). Dále bude předpokládán pouze jeden přenosový kanál, vícekanálové metody tedy nebudou uvažovány. Za těchto předpokladů lze techniku prahování waveletových koeficientů využít dvěma způsoby:

- aplikace v časové oblasti,
- aplikace ve spektrální oblasti.

Aplikace techniky prahování v *časové oblasti* je dostatečně prozkoumána a je popsána v literatuře [4, 12, 24, 19]. Doktorská práce tedy bude zaměřena na aplikaci techniky prahování v *spektrální oblasti*. Jak totiž vyplývá z některých prací, je možné prahování waveletových koeficientů aplikovat např. na periodogram šumového signálu a tak vyhladit odhad výkonové spektrální hustoty šumu [22, 11].

Cílem disertační práce je:

- Porovnat metody odhadu charakteristik rušení a šumu z hlediska přesnosti a spolehlivosti. Analyzovat možnosti použití waveletové transformace pro vyhlazení odhadu charakteristik šumu a porovnat ji s běžně používanými metodami.
- Na základě těchto analýz navrhnout novou metodu jednokanálové separace řeči ze šumového pozadí s využitím waveletové transformace a vyhodnotit ji pro různé typy rušení a šumu z hlediska zlepšení poměru signálu od šumu.

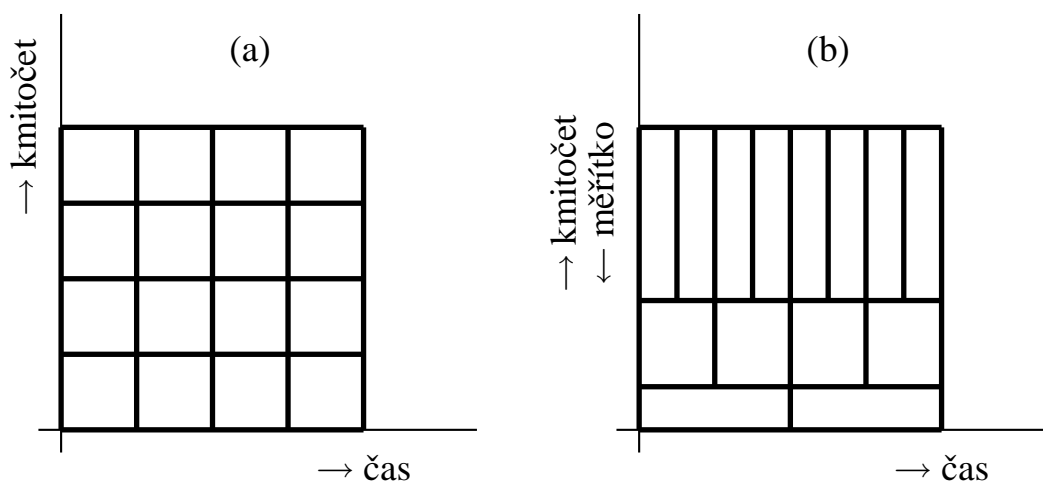
## 3 ČASOVĚ-KMITOČTOVÁ REPREZENTACE SIGNÁLU

### 3.1 ROZDĚLENÍ METOD ČASOVĚ-KMITOČTOVÉ ANALÝZY

Popis řečového signálu je vzhledem k jeho komplexnosti, různorodosti a s ohledem na jeho nestacionaritu velmi náročný. Proto se používá především časově-kmitočtová reprezentace signálu. Nejčastěji je založena na spojitě nebo diskrétní Fourierově transformaci a jejích modifikacích [17].

V prostém spektrogramu se ukazuje nevýhoda diskrétní Fourierovy transformace, jejíž absolutní rozlišovací schopnost je konstantní (viz obr. 3.1a). Délka okna tedy musí být určena potřebnou kmitočtovou rozlišovací schopností na straně nízkých kmitočtů a omezuje tak časové rozlišení. V praxi by bylo však zpravidla užitečnější, kdyby konstantní byla relativní kmitočtová rozlišovací schopnost, tzn. kdyby rozlišovací schopnost byla poměrná v oblasti nízkých i vysokých kmitočtů.





Obr. 3.1: Konstatní (a) a proměnná (b) rozlišovací schopnost časově-kmitočtové analýzy.

Naopak *waveletové transformace* již ve svém principu využívá několikanásobného rozlišení. Bázovými funkcemi waveletové transformace jsou časově omezené průběhy (plynule narůstajícího a doznívajícího „waveletu“), které jsou generované časovou dilatací a translací z jediného vzoru (mateřské funkce), takže podle požadavku násobného rozlišení s prodloužením délky okna klesá kmitočet vnitřních oscilací. Se změnou měřítka se mění i časová a kmitočtová rozlišovací schopnost a celý časově-měřítkový prostor má podobu znázorněnou na obrázku 3.1b.

## 3.2 WAVELETOVÁ TRANSFORMACE

Podstatným rozdílem mezi waveletovou transformací a ostatními transformacemi (Fourierovou, kosinovou apod.) je, že ostatní transformace používají báze funkce o nekonečné délce, zatímco báze funkce waveletové transformace jsou nenulové pouze na konečném časovém intervalu nebo velmi rychle klesají směrem k nekonečnu. Přitom báze funkce pokrývají po částech celý časově-měřítkový prostor analyzovaného signálu, takže nedochází ke ztrátě informace a je možné provést zpětnou transformaci. [3, 12, 19, 24]

### 3.2.1 Spojitá waveletová transformace

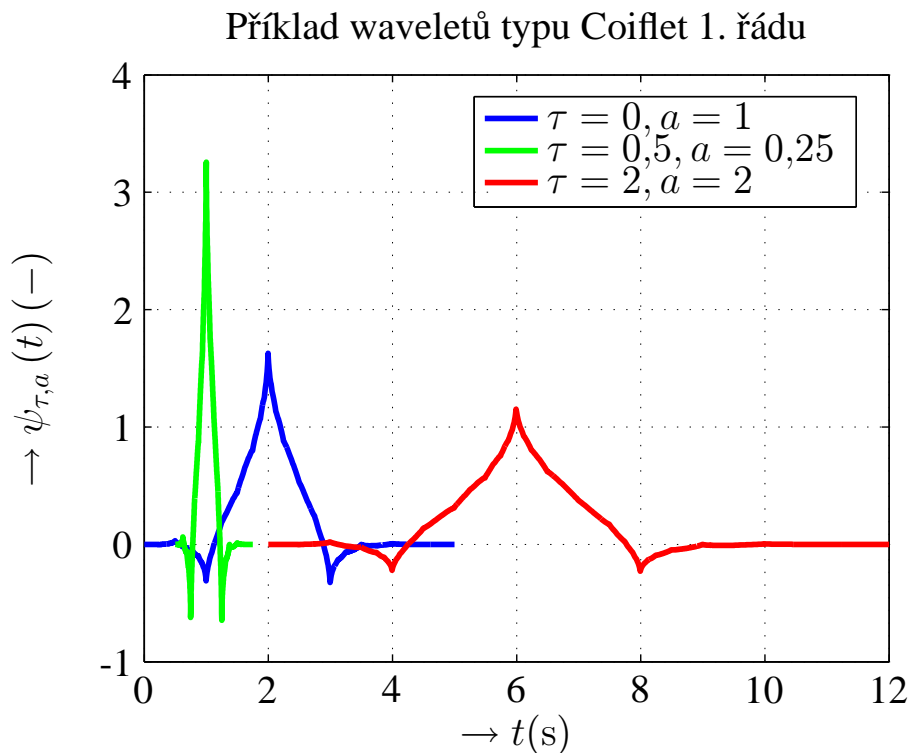
Spojitá waveletová transformace je podobně jako Fourierova transformace definována skalárním součinem v  $L^2$  prostoru analyzovaného signálu  $f(t)$  s báze funkcí  $\psi_{\tau,a}(t)$  podle vztahu

$$WT \{f(t)\} = C_{\tau,a}(t) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{\tau,a}(t)} dt, \quad (3.1)$$

kde  $C_{\tau,a}(t)$  jsou koeficienty spojité waveletové transformace a bázové funkce  $\psi_{\tau,a}(t)$  jsou odvozeny z jediné mateřské funkce [3, 12]

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right). \quad (3.2)$$

Parametr  $\tau$  představuje posunutí mateřské funkce v čase a parametr  $a$  představuje měřítko roztažení nebo zkrácení mateřské funkce. Podělení hodnotou  $\sqrt{a}$  zajišťuje zachování energie waveletu při roztažení.



Obr. 3.2: Příklad waveletů odvozených z mateřské funkce typu Coiflet 1. řádu.

Změna měřítka má vliv také na kmitočtové vlastnosti waveletu. Podle věty o změně měřítka, platné pro Fourierovu transformaci, platí

$$\mathcal{F}\left\{\psi\left(\frac{t}{a}\right)\right\} = a\Psi(a\omega), \quad (3.3)$$

kde spektrum mateřské funkce

$$\Psi(\omega) = \mathcal{F}\{\psi(t)\}.$$

Jestliže je wavelet roztažen v čase v poměru měřítka  $a$ , pak Fourierovo spektrum waveletu je podél kmitočtové osy stlačeno ve stejném poměru měřítka  $a$ . Wavelety pro malá měřítka tak mají krátké trvání, ale široké spektrum. Poskytují tedy velké časové rozlišení, ale malé kmitočtové rozlišení. Naopak wavelety pro velká měřítka mají dlouhé trvání, ale úzké spektrum. Poskytují malé časové rozlišení, ale velké kmitočtové rozlišení.

### 3.2.2 Diskrétní waveletová transformace

Pojem *diskrétní waveletová transformace* označuje waveletovou transformaci, která provádí výpočet pouze pro diskrétní hodnoty parametrů měřítka  $a$  a posunutí  $\tau$ . Nejčastěji se používá tzv. *dyadická diskrétní waveletová transformace* DWT, při které jsou hodnoty měřítka a posunutí násobkem mocniny dvou

$$\begin{aligned} a &= 2^j, \\ \tau &= m 2^j = m a, \end{aligned}$$

kde  $m, j$  jsou celá čísla. Parametr  $j$  se označuje jako úroveň.

Pokud je navíc uvažován diskrétní analyzovaný signál  $x[n]$ , dyadická diskrétní waveletová transformace přejde do dyadické diskrétní waveletové transformace s diskrétním časem. Zkráceně se tato transformace označuje jako *waveletová transformace s diskrétním časem* (DTWT) a definiční integrál je nahrazen sumou [3, 12, 24]

$$C_{j,m} = \sum_{n=0}^{N-1} x[n] \psi_{j,m}[n], \quad (3.4)$$

kde  $x[n]$  jsou vzorky analyzovaného signálu,  $\psi_{j,m}[n]$  jsou vzorky posunuté verze diskrétního waveletu, jež je v měřítku  $2^j$ ,  $N$  je počet vzorků signálu.

*Inverzní waveletová transformace s diskrétním časem* (IDTWT) je lineární kombinací použitých waveletů

$$x[n] = \frac{1}{N} \sum_{j=0}^{\log_2 N - 1} \sum_{k=0}^{\lfloor \frac{N}{2^{j+1}} \rfloor} C_{j,m} \psi_{j,m}[n]. \quad (3.5)$$

### 3.3 PRAHOVÁNÍ KOEFICIENTŮ WAVELETOVÉ TRANSFORMACE

Při rekonstrukci zkresleného signálu s použitím waveletové transformace se provádí rozložení signálu do detailních a aproximačních koeficientů, modifikace koeficientů nejčastěji prahováním a signál je rekonstruován zpětnou waveletovou transformací (viz obr. 3.3). Prahování je nelineární operace, která přiřazuje malým hodnotám vzorků signálu menší váhu nebo je nuluje. Existuje více prahovacích funkcí, to je předpisů jak tuto váhu přidělit v závislosti na stanovené hodnotě prahu  $\rho$  [3, 12, 19, 24].

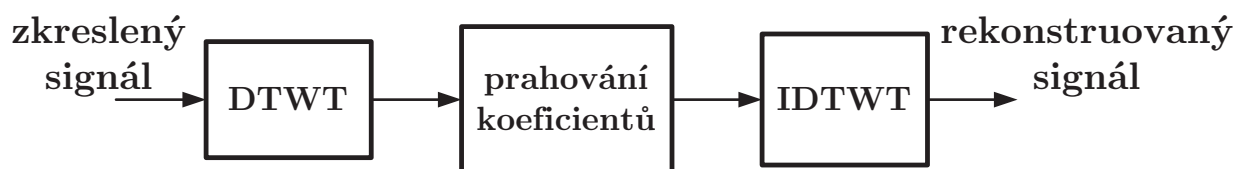
Mezi běžně používaná prahovací pravidla patří:

#### 1. Tvrdé prahování

$$T_H(C_{j,m}, \rho) = \begin{cases} 0 & \text{pro } |C_{j,m}| \leq \rho \\ C_{j,m} & \text{pro } |C_{j,m}| > \rho \end{cases}. \quad (3.6)$$

#### 2. Měkké prahování

$$T_S(C_{j,m}, \rho) = \text{sgn}(C_{j,m}) \max\{0, |C_{j,m}| - \rho\}. \quad (3.7)$$



Obr. 3.3: Blokové schéma rekonstrukce signálu pomocí waveletové transformace.

Dále existuje poloměkké prahování, prahování nezápornou garotou, hyperbolické prahování a další.

### 3.4 ODHAD VELIKOSTI PRAHŮ

Před provedením prahování koeficientů  $C_{j,m}$  je nutné určit optimální hodnotu prahu  $\rho$ , která zajistí největší potlačení šumu a rušení při nejmenším zkreslení rekonstruovaného signálu. Výpočet optimální hodnoty prahu je značně složitý, vyžaduje znalost statistických vlastností koeficientů užitečného signálu i rušení, a v praxi se proto většinou nahrazuje odhadem optimální hodnoty. Při odhadu příliš velkého prahu dojde k odstranění složek užitečného signálu a k jeho zkreslení, při odhadu příliš malého prahu není ze signálu odstraněna většina rušení a šumu.

Mezi nejjednodušší odhad optimální hodnoty prahu patří univerzální práh prvního typu, který byl odvozen pro odstranění šumu s Gaussovým rozdělením. Pro odhad prahu  $\rho$  je použit vztah [4, 19]

$$\rho = \sigma_n \sqrt{2 \log N}, \quad (3.8)$$

kde  $\sigma_n$  značí směrodatnou odchylku šumu a  $N$  je délka signálu. Vztah (3.8) byl odvozen pro aditivní bílý šum s Gaussovým rozdělením.

Kromě těchto jednoduchých metod existují i iterační postupy, které se snaží minimalizovat účelovou funkci vyjadřující odhad chyby rekonstrukce. Mezi tyto metody patří *metoda zobecněného křížového ověření*. Za optimální je považován stav, kdy je vnitřní chyba minimalizovaná podle [19]

$$\beta_{\text{GCV}}(\rho) = \frac{1}{N} \frac{\|C_{j,m} - C_{j,m}^{\text{th}}\|^2}{\left(\frac{N_0}{N}\right)^2}, \quad (3.9)$$

kde  $C_{j,m}$  je vektor koeficientů diskrétní waveletové transformace v úrovni  $j$  a  $C_{j,m}^{\text{th}}$  je vektor koeficientů po provedení prahování.

## 4 ODHAD VÝKONOVÉ SPEKTRÁLNÍ HUSTOTY RUŠENÍ

Pomocí Wienerova-Chinčina teoremu [14, 17] lze odvodit, že výkonová spektrální hustota stacionárního náhodného procesu je diskrétní Fourierovou transformací autokorelační posloupnosti  $\gamma_{xx}[m]$

$$G_{xx}[k] = \sum_{m=-\infty}^{\infty} \gamma_{xx}[m] e^{-j2\pi k m}, \quad (4.1)$$

kde  $\gamma_{xx}[m]$  je autokorelační funkce stacionárního náhodného procesu vypočtená podle vztahu

$$\gamma_{xx}[m] = \sum_{n=-\infty}^{\infty} x[n+m]\overline{x[n]}. \quad (4.2)$$

Za stacionární je označen náhodný proces takový, že jeho střední hodnota a rozptyl nezávisí na čase  $n$  a autokorelační funkce  $\gamma_{xx}[m]$  závisí pouze na vzájemném posunutí  $m$ . Dochází ke ztrátě informace o fázi signálu a vztah mezi časovým průběhem náhodného signálu a jeho spektrální výkonovou hustotou není jednoznačný. K další ztrátě informace dochází proto, že statistické charakteristiky signálu, a tedy i autokorelační posloupnost, lze prakticky pouze odhadovat.

Metody pro odhad výkonové spektrální hustoty lze rozdělit do dvou skupin [17]:

- klasické metody – metody založené na použití diskrétní Fourierovy transformace,
- parametrické metody – metody založené na popisu signálu souborem parametrů.

### 4.1 ZÁKLADNÍ VLASTNOSTI METOD SPEKTRÁLNÍ ANALÝZY

Velké množství metod pro odhad výkonové spektrální hustoty poskytuje možnost podle účelu zvolit tu nejvhodnější. Mezi základní vlastnosti metod patří *kmitočtové rozlišení*, *schopnost detekce signálu v šumu* a *schopnost věrného určení tvaru spektra při použití nízkého počtu parametrů*. Při odhadu výkonové spektrální hustoty je nutné vzít v úvahu i náhodný charakter signálu a tím náhodný charakter odhadu výkonové spektrální hustoty. Metody pak lze hodnotit i podle:

- vychýlení odhadu,
- rozptylu odhadu,
- konzistentnosti (statistické stability) odhadu.

Vychýlení odhadu (systematická chyba)  $B\{\hat{\Theta}\}$  je definována jako rozdíl mezi skutečnou teoretickou hodnotou sledovaného statistického parametru  $\Theta$  a střední hodnotou jeho odhadu  $E\{\hat{\Theta}\}$

$$B\{\hat{\Theta}\} = \Theta - E\{\hat{\Theta}\} \quad (4.3)$$

Odhad je nestranný, je-li jeho střední hodnota rovna teoretické hodnotě odhadovanému parametru

$$E \left\{ \hat{\Theta} \right\} = \Theta \quad (4.4)$$

Konzistentní (statisticky stabilní) odhad je takový, který konverguje se vzrůstajícím počtem realizací  $M$  k odhadovanému parametru podle pravděpodobnosti

$$\lim_{M \rightarrow \infty} P \left\{ \left| \hat{\Theta} - \Theta \right| \geq \zeta \right\} = 0, \quad (4.5)$$

kde  $\zeta$  je libovolné pevné kladné číslo.

Rozptyl odhadu je definován jako střední hodnota kvadrátu rozdílu mezi hodnotou odhadu  $\hat{\Theta}$  a střední hodnotou odhadu  $E \left\{ \hat{\Theta} \right\}$

$$D \left\{ \hat{\Theta} \right\} = E \left\{ \left( \hat{\Theta} - E \left\{ \hat{\Theta} \right\} \right)^2 \right\} \quad (4.6)$$

Přesnost odhadu je možné definovat pomocí normalizované střední kvadratické chyby

$$\epsilon_c^2 = \epsilon_b^2 + \epsilon_r^2, \quad (4.7)$$

kde  $\epsilon_b^2$  je normalizovaná kvadratická chyba vychýlením a  $\epsilon_r^2$  je normalizovaná kvadratická chyba rozptylem.

## 4.2 KLASICKÉ METODY ODHADU VÝKONOVÉ SPEKTRÁLNÍ HUSTOTY

Klasické metody odhadu výkonové spektrální hustoty využívají pro výpočet diskrétní Fourierovu transformaci. Autokorelační posloupnost lze odhadnout podle vztahu [14]

$$\begin{aligned} r_{xx} [m] &= \frac{1}{N} \sum_{n=0}^{N-m-1} x [n+m] \overline{x [n]}, m = 0, 1, \dots, N-1 \\ r_{xx} [m] &= \frac{1}{N} \sum_{n=|m|}^{N-1} x [n+m] \overline{x [n]}, m = -1, -2, \dots, 1-N \end{aligned} \quad (4.8)$$

kde  $N$  je délka signálu. Použitý odhad je sice vychýlený, ale asymptoticky se blíží autokorelační funkci

$$\lim_{N \rightarrow \infty} r_{xx} [m] = \gamma_{xx} [m]. \quad (4.9)$$

Navíc rozptyl odhadu s rostoucím  $N \rightarrow \infty$  klesá k nule, takže se jedná o konzistentní odhad [14].

Dosazením (4.8) do (4.1) dostaneme odhad výkonové spektrální hustoty ergodickeho náhodného procesu

$$P_{xx}[k] = \sum_{m=-(N-1)}^{N-1} r_{xx}[m] e^{-j2\pi km} = \frac{1}{N} \mathbf{X}[k] \overline{\mathbf{X}[k]} \quad (4.10)$$

Takto definovaný odhad výkonové spektrální hustoty se označuje jako *periodogram*  $P_{xx}[k]$ . Periodogram je opět vychýleným odhadem, ale asymptoticky nevychýleným odhadem

$$\lim_{N \rightarrow \infty} E \{P_{xx}[k]\} = G_{xx}[k]. \quad (4.11)$$

Pro rozptyl periodogramu je možné odvodit [14]

$$D \{P_{xx}[k]\} = G_{xx}^2[k] \left[ 1 + \left( \frac{\sin 2\pi \frac{k}{2N-1} N}{N \sin 2\pi \frac{k}{2N-1}} \right)^2 \right]. \quad (4.12)$$

Je zřejmé, že pro  $N \rightarrow \infty$  se rozptyl odhadu blíží kvadrátu výkonové spektrální hustoty

$$\lim_{N \rightarrow \infty} D \{P_{xx}[k]\} = G_{xx}^2[k]. \quad (4.13)$$

Odhad pomocí periodogramu je tedy nekonzistentní odhad. Navíc jeho rozptyl dosahuje kvadrátu hodnoty výkonové spektrální hustoty a relativní chyba rozptylem dosahuje hodnoty 100 %.

Jednou z metod vyhlazení odhadu výkonové spektrální hustoty je výpočet střední hodnoty několika periodogramů, které jsou získány z několika dostatečně dlouhých realizací náhodného procesu. [14, 17] Pokud máme k dispozici pouze jednu realizaci náhodného procesu, lze dílčí realizace získat jejím rozdělením na segmenty. Tím ovšem klesne dosažitelné kmitočtové rozlišení. Z každého segmentu je vypočten periodogram

$$P_{xx}^{(i)}[k] = \frac{1}{M} \left| \sum_{n=0}^{M-1} x_i[n] e^{-j2\pi kn} \right|^2.$$

Odhad výkonové spektrální hustoty je potom roven střední hodnotě dílčích periodogramů

$$P_{xx}^B[k] = \frac{1}{K} \sum_{i=1}^{K-1} P_{xx}^{(i)}[k]. \quad (4.14)$$

Tato metoda je označována jako Bartlettova metoda [1]. Opět platí, že odhad výkonové spektrální hustoty pomocí Bartlettovy metody je asymptoticky nevychýlený. Rozptyl odhadu však bude  $K$ krát menší [14]

$$D \{P_{xx}^B[k]\} = \frac{1}{K} G_{xx}^2[k] \left[ 1 + \left( \frac{\sin 2\pi \frac{k}{2N-1} N}{N \sin 2\pi \frac{k}{2N-1}} \right)^2 \right]. \quad (4.15)$$

Další modifikaci průměrování periodogramů navrhl Welch [23]. Princip je stejný jako u Bartlettovy metody, ale Welch povolil překrývání segmentů a jednotlivé segmenty jsou váhovány zvoleným oknem. Lze tak snížit rozptyl odhadu při zachování dostatečného kmitočtového rozlišení odhadu.

### 4.3 PARAMETRICKÉ METODY ODHADU

Parametrické metody se oproti klasickým metodám odhadu snaží na základě krátké realizace sledovaného náhodného procesu odhadnout parametry systému, který náhodný proces generuje. Na základě modelu je pak možné odhadnout vzorky náhodného procesu mimo zaznamenaný úsek a tedy i vzorky autokorelační funkce pro  $|m| \geq N$ . Výhodou je, že poskytují větší kmitočtové rozlišení než klasické metody a dosažitelné kmitočtové rozlišení je téměř nezávislé na délce realizace náhodného procesu.

Diskrétní stacionární náhodný proces  $x[n]$  s neznámou výkonovou spektrální hustotou  $G_{xx}[k]$  je možné transformovat na bílý šum  $w[n]$ , když je posloupnost  $x[n]$  filtrována lineárním číslicovým filtrem s přenosovou funkcí  $1/H(z)$  inverzní k přenosové funkci (4.18). Tento filtr se poté nazývá *bělicí filtr*. Pokud se podaří najít přenosovou funkci bělicího filtru, lze zjistit výkonovou spektrální hustotu původního náhodného procesu podle vztahu

$$G_{xx}(e^{j2\pi f}) = \sigma_w^2 H(e^{j2\pi f}) \overline{H(e^{j2\pi f})} = \sigma_w^2 |H(e^{j2\pi f})|^2, \quad (4.16)$$

kde

$$\sigma_w^2 = e^{v[0]}, \quad (4.17)$$

$$H(z) = \exp\left(\sum_{m=1}^{\infty} v[m] z^{-m}\right) = \frac{\sum_{i=0}^r b_i z^{-i}}{1 + \sum_{j=0}^s a_j z^{-j}}, \quad |z| > r_1, r_1 < 1, \quad (4.18)$$

je přenosová funkce systému, kterým je generován sledovaný náhodný proces. Koeficienty  $v[m]$  jsou kepsrální koeficienty autokorelační funkce  $\gamma_{xx}[m]$ .

Podle tvaru přenosové funkce  $H(z)$  je možné rozlišit tři možné případy [17]. V prvním případě má přenosová funkce pouze nulové body. Tento případ se označuje jako plovoucí průměr MA – Moving Average. V druhém případě má přenosová funkce pouze póly a tento případ je označován jako autoregresivní model AR – AutoRegressive. Tento model je používán nejčastěji, protože úloha hledání parametrů bělicího filtru je lineární a model poskytuje dostatečné kmitočtové rozlišení a malý rozptyl odhadu výkonové spektrální hustoty. Koeficienty modelu lze získat pomocí iteračního Levinsonova-Durbinova algoritmu [14, 17]. Poslední možnost vznikne kombinací obou výše zmíněných, přenosová funkce má nulové body i póly a model je označován jako ARMA – Auto Regressive Moving Average.



Nedílnou součástí parametrických metod je určení řádu  $s$  modelu. Pokud řád modelu zvolíme příliš nízký, získáme extrémně vyhlazený odhad výkonové spektrální hustoty, který nemusí odpovídat teoretické výkonové spektrální hustotě. Malý řád modelu totiž znamená malý počet parametrů modelu a důsledkem je nepřesný popis signálu. Pokud zvolíme řád příliš vysoký, může být popis zkreslen vlivem zvětšování rozptylu odhadu, neboť pro určení koeficientů potřebujeme více hodnot autokorelační posloupnosti. Ukazatelem přesnosti modelu je minimální střední kvadratická chyba  $E^{(i)}$ , která klesá se vzrůstajícím řádem modelu. Zvyšování řádu je možné zastavit v okamžiku, kdy pokles je dostatečně malý. Bylo navrženo několik kritérií, kdy zvyšování řádu zastavit. [14, 17] U krátkých signálů je možné řád modelu určit podle empirické rovnice [17]

$$0,05N \leq s \leq 0,2N, \quad (4.19)$$

kde  $N$  je délka signálu a  $s$  je optimální řád modelu. Krátkým signálem se myslí signál takové délky, že lze dosáhnout minimálního kmitočtového rozlišení  $\Delta f$ , tedy  $N = \frac{f_{vz}}{\Delta f}$ .

## 5 METODA SPEKTRÁLNÍHO ODEČÍTÁNÍ

Předpokládejme, že vstupní signál  $y[n]$  je aditivní směsí užitečného signálu  $x[n]$  a šumu  $\nu[n]$

$$y[n] = x[n] + \nu[n]. \quad (5.1)$$

Vzhledem k linearitě diskrétní Fourierovy transformace je výkonové spektrum takto vzniklé směsi rovno

$$|Y[k]|^2 = |X[k]|^2 + |V[k]|^2 + \overline{X[k]}V[k] + X[k]\overline{V[k]}. \quad (5.2)$$

Druhý člen  $|V[k]|^2$  je výkonová spektrální hustota šumu  $G_{\nu\nu}[k]$  a třetí a čtvrtý člen  $\overline{X[k]}V[k] + X[k]\overline{V[k]}$  představují vzájemnou výkonovou spektrální hustotu. V případě, že užitečný signál  $x[n]$  a šum  $\nu[n]$  jsou nekorelované, jsou hodnoty vzájemné výkonové spektrální hustoty rovny nule. V takovém případě lze odhad výkonového spektra užitečného signálu  $|\hat{X}[k]|^2$  získat odečtením odhadu výkonové spektrální hustoty šumu

$$|\hat{X}[k]|^2 = |Y[k]|^2 - \hat{G}_{\nu\nu}[k] \quad (5.3)$$

Rovnici (5.3) lze také přepsat do tvaru

$$|\hat{X}[k]|^2 = |Y[k]|^2 \left( 1 - \frac{\hat{G}_{\nu\nu}[k]}{|Y[k]|^2} \right) = |Y[k]|^2 |H[k]|^2, \quad (5.4)$$

kde  $H[k]$  je kmitočtová charakteristika ekvivalentního filtru.

## 5.1 CHYBA REKONSTRUKCE SIGNÁLU METODOU SPEKTRÁLNÍHO ODEČÍTÁNÍ

Ve skutečnosti ovšem přesné hodnoty výkonové spektrální hustoty rušení  $G_{\nu\nu}[k]$  a výkonového spektra zarušeného signálu  $|\hat{Y}[k]|^2$  nelze určit. Lze získat pouze odhady

$$|\hat{Y}[k]|^2 = |Y[k]|^2 + \delta_Y[k], \quad (5.5)$$

$$\hat{G}_{\nu\nu}[k] = G_{\nu\nu}[k] + \delta_\nu[k], \quad (5.6)$$

kde  $\delta_Y[k]$  a  $\delta_\nu[k]$  jsou nezávislé nekorelované náhodné procesy s nulovou střední hodnotou (pro nevychýlené odhady) a s rozptylem rovným rozptylu odhadu. Rovnici spektrálního odečítání (5.3) pak lze zapsat ve tvaru

$$\begin{aligned} |\hat{X}[k]|^2 &= |Y[k]|^2 + \delta_Y[k] - (G_{\nu\nu}[k] + \delta_\nu[k]) = \\ &= |Y[k]|^2 - G_{\nu\nu}[k] + \delta_Y[k] - \delta_\nu[k] = \\ &= |X[k]|^2 + \delta_Y[k] - \delta_\nu[k]. \end{aligned} \quad (5.7)$$

Chyba rekonstrukce pak bude rovna

$$\begin{aligned} \epsilon &= E \left\{ \left( |X[k]|^2 + \delta_Y[k] - \delta_\nu[k] - |X[k]|^2 \right)^2 \right\} = \\ &= E \left\{ (\delta_Y[k] - \delta_\nu[k])^2 \right\} = \\ &= E \{ \delta_Y[k]^2 \} + E \{ \delta_\nu[k]^2 \} - 2E \{ \delta_Y[k] \delta_\nu[k] \}. \end{aligned} \quad (5.8)$$

Vzhledem k nulovým středním hodnotám náhodných procesů  $\delta_Y[k]$  a  $\delta_\nu[k]$  jsou první dva členy rovny rozptylům odhadů. Odhady výkonového spektra vstupního signálu  $|\hat{Y}[k]|^2$  a výkonové spektrální hustoty  $\hat{G}_{\nu\nu}[k]$  jsou určovány z různých segmentů signálu a proto je lze považovat za nezávislé. Za tohoto předpokladu jsou nezávislé i náhodné procesy  $\delta_Y[k]$  a  $\delta_\nu[k]$ . Potom střední hodnota součinu je dána součinem středních hodnot a s ohledem na nulovou střední hodnotu procesů  $\delta_Y[k]$  a  $\delta_\nu[k]$  bude třetí člen nulový. Potom chyba rekonstrukce je rovna součtu rozptylu odhadů

$$\epsilon = \sigma_Y^2 + \sigma_\nu^2. \quad (5.9)$$

Jak bylo ukázáno v části 4.2, normalizovaná kvadratická chyba rozptylem periodogramu je rovna 1 a rozptyl periodogramu dosahuje hodnoty výkonové spektrální hustoty. Proto se odhad pomocí periodogramu nepoužívá, ale je nutné použít některý z vyhlazených odhadů jako je Bartlettova nebo Welchova metoda nebo exponenciální průměrování

$$G_{\nu\nu}^{(l)}[k] = \lambda G_{\nu\nu}^{(l-1)}[k] + (1 - \lambda) P_{\nu\nu}[k] \quad (5.10)$$

kde  $G_{\nu\nu}^{(l-1)}[k]$  je předchozí odhad výkonové spektrální hustoty,  $P_{\nu\nu}[k]$  je nový odhad pomocí periodogramu současného segmentu,  $G_{\nu\nu}^{(l)}[k]$  je aktualizovaný odhad výkonové spektrální hustoty a  $\lambda \in \langle 0, 1 \rangle$  je zvolená konstanta.

## 5.2 SNÍŽENÍ ROZPTYLU ODHADU POMOCÍ WAVELETOVÉ TRANSFORMACE

Pro snížení rozptylu odhadu výkonové spektrální hustoty šumu je možné použít i waveletovou transformaci. Náhodný proces  $\delta_\nu[k]$  v rovnici (5.6) představuje rušení, které je z odhadu možné odstranit prahováním koeficientů waveletové transformace, stejně jako u jakéhokoliv jiného signálu.

Předpokládejme, že rušení  $\nu[n]$  je Gaussův náhodný proces. Protože Fourierova transformace je lineární, tak reálná a imaginární složka Fourierovy transformace budou mít také Gausovo rozdělení. Periodogram  $P_{xx}[k]$  bude mít rozdělení  $\chi_n^2$  se dvěma stupni volnosti  $n = 2$  a stejné rozdělení bude mít i náhodný proces  $\delta_\nu[k]$ . Rozptyl náhodného procesu  $\delta_\nu[k]$  je přitom roven výkonové spektrální hustotě  $G_{xx}[k]$ . Logaritmus periodogramu bude roven

$$\ln P_{xx}[k] = \ln G_{xx}[k] + \epsilon_c + \gamma, \quad (5.11)$$

kde  $\epsilon_c$  je náhodný proces zkreslující odhad výkonové spektrální hustoty a  $\gamma \approx 0,57721$  je Euler-Mascheroniho konstanta [19].

Náhodný proces  $\epsilon_c$ , který způsobuje rozptyl periodogramu, je možné odstranit pomocí prahování koeficientů waveletové transformace. Transformace je lineární a proto budou koeficienty představovat součet koeficientů reprezentující hledanou výkonovou spektrální hustotu  $g_{j,m}$  a koeficienty reprezentující šum  $e_{j,m}$

$$C_{j,m} = g_{j,m} + e_{j,m}. \quad (5.12)$$

Protože jednotlivé složky náhodného procesu  $\epsilon_c$  jsou nezávislé a transformace je ortogonální, budou koeficienty  $e_{j,m}$  také nezávislé. Zároveň však je jejich rozdělení pravděpodobnosti nezávislé na posunutí  $m$ , ale je závislé na úrovni  $j$ .

Pro určení prahu je možné použít některou z metod uvedených v části 3.4. Vzhledem k předpokládanému rozdělení rušícího signálu  $\epsilon[k]$  je výhodnější použít optimální stanovení prahu v závislosti na úrovni navržené pro odstranění šumu s rozdělením  $\ln \chi_2^2$  [19]. Pokud je úroveň velká, pak je práh roven

$$\rho_j = \alpha_j \ln N, \quad (5.13)$$

kde konstanty  $\alpha_j$  jsou uvedeny v tabulce 5.1 a  $N$  je délka dat. Pokud je úroveň malá  $j \ll J - 1$ , kde  $J = \lfloor \log_2 N \rfloor$  pak se práh určí podle vztahu

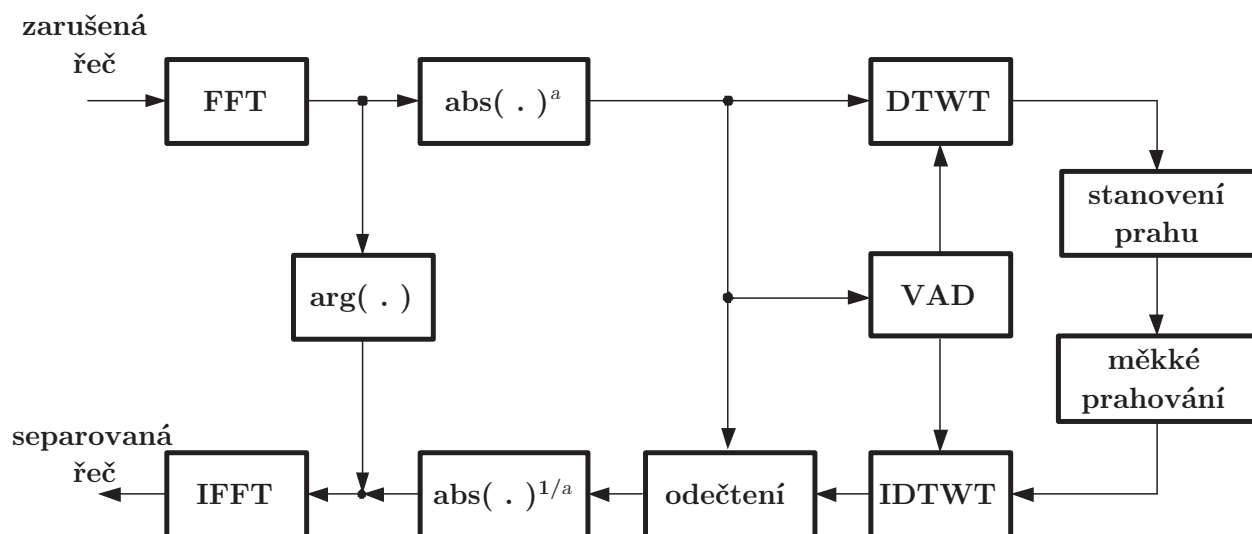
$$\rho_j = \frac{\pi}{\sqrt{3}} \sqrt{\ln N}. \quad (5.14)$$

Po určení prahu je provedeno některé prahování uvedené v části 3.3. Největšího vyhlazení odhadu a snížení rozptylu je dosahováno při použití měkkého prahování podle (3.7). Po prahování je získán vyhlazený odhad výkonové spektrální hustoty šumu  $\hat{G}_{xx}[k]$  pomocí zpětné waveletové transformace s diskretním časem.

Tab. 5.1: Hodnoty konstant  $\alpha_j$  pro určení prahu koeficientů waveletové transformace.

úroveň $j$	$\alpha_j$	úroveň $j$	$\alpha_j$
$J - 1$	1.29	$J - 6$	0.54
$J - 2$	1.09	$J - 7$	0.46
$J - 3$	0.92	$J - 8$	0.39
$J - 4$	0.77	$J - 9$	0.32
$J - 5$	0.66	$J - 10$	0.27

Blokové schéma modifikované metody spektrálního odečítání, které využívá pro odhad výkonové spektrální hustoty rušení prahování koeficientů waveletové transformace periodogramu je zobrazena na obrázku 5.1.



Obr. 5.1: Blokové schéma modifikované metody spektrálního odečítání.

### 5.3 ENERGETICKÝ DETEKTOR ŘEČOVÉ AKTIVITY

Při testování byl použit jednoduchý energetický detektor založený na sledování krátkodobé energie jednotlivých segmentů řeči vypočtené podle vztahu

$$E[k] = \frac{1}{N} \sum_{n=kN}^{(k+1)N-1} x^2[n], \quad (5.15)$$

kde  $N$  je délka segmentu,  $k$  je index segmentu a  $E[k]$  je krátkodobá energie signálu  $x[n]$ . Tato energie je porovnávána s dlouhodobou střední hodnotou energie šumového pozadí, která je průběžně aktualizována exponenciálním průměrováním. V segmentech, které jsou označeny jako řečové, je dlouhodobá střední hodnota energie šumu  $E_\nu[k]$  aktualizována podle vztahu

$$E_\nu[k] = \lambda_1 E_\nu[k-1] + (1 - \lambda_1) E[k], \quad (5.16)$$

kde  $E_\nu[k-1]$  je dlouhodobá střední hodnota určená v předchozím segmentu a  $\lambda_1$  je konstanta exponenciálního průměrování  $\lambda_1 \in \langle 0,85; 0,95 \rangle$ . V segmentech, které jsou označeny jako šumové, probíhá aktualizace dlouhodobé střední hodnoty energie šumu podle vztahu

$$E_\nu[k] = \lambda_2 E_\nu[k-1] + (1 - \lambda_2) E[k]. \quad (5.17)$$

Vhodné hodnoty konstanty jsou  $\lambda_2 \in \langle 0,98; 0,999 \rangle$ . Uvedené hodnoty platí pro délku segmentu přibližně 20 – 40 ms.

Detekce je provedena podle jednoduchého pravidla

$$E[k] < E_\nu[k] + \rho_N \Rightarrow f_{\text{VAD}}[k] = 0, \quad (5.18)$$

$$E[k] > E_\nu[k] + \rho_S \Rightarrow f_{\text{VAD}}[k] = 1, \quad (5.19)$$

kde  $f_{\text{VAD}}[k]$  je příznak řečové aktivity,  $\rho_N \in \langle 0,1; 0,4 \rangle$  a  $\rho_S \in \langle 0,5; 0,8 \rangle$  jsou konstanty pro zavedení hystereze v rozhodování, aby nedocházelo k náhodným přeskokům mezi řečovým a šumovým segmentem. Pokud neplatí ani jedna z uvedených podmínek, je segment označen podle typu předchozího segmentu  $f_{\text{VAD}}[k] = f_{\text{VAD}}[k-1]$ . Pro zvýšení spolehlivosti je někdy výstup detektoru ještě zpracován mediánovým filtrem.

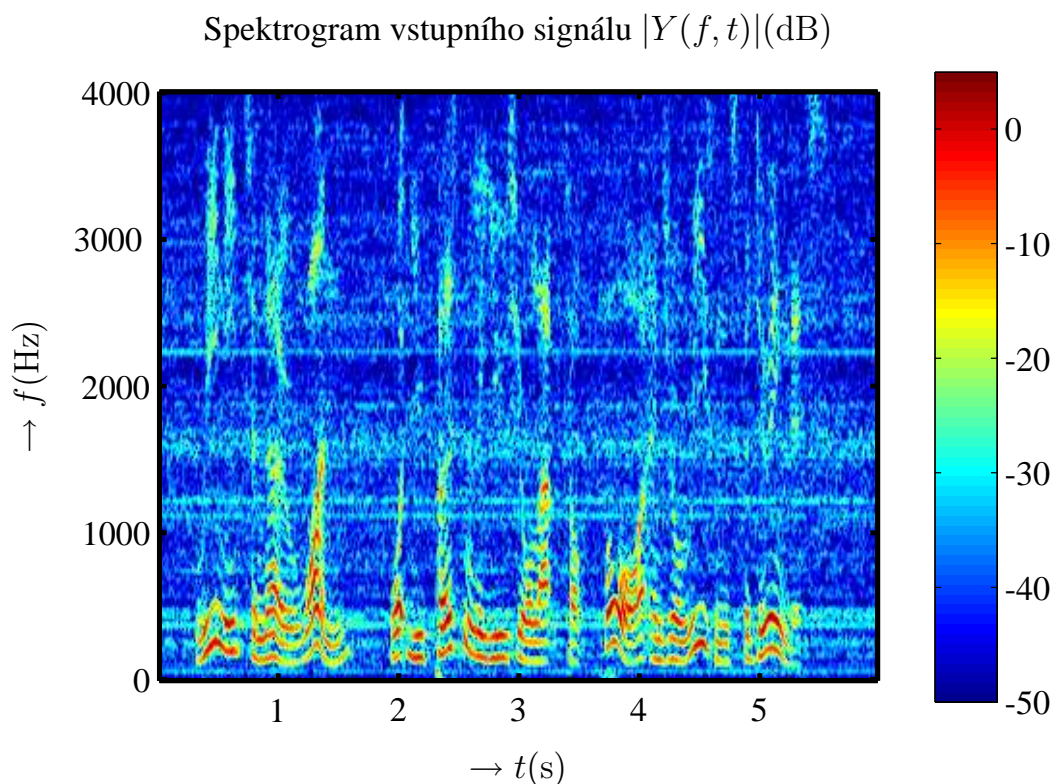
## 6 ZÁVĚR

Metoda spektrálního odečítání používající pro odhad výkonové spektrální hustoty exponenciální průměrování, Bartlettovu metodu, AR model nebo prahování waveletových koeficientů byla testována na 58 záznamech délky přibližně 20 s řečového signálu pořízených dvěma mluvčími v různých reálných prostředích s různým typem rušení a odstupem signálu od šumu. Pro porovnání metod byl sledován poměr signálu od šumu rekonstruovaného řečového signálu. Z výsledků testování shrnutých v tab. 6.1 lze konstatovat, že všechny metody dosahovaly přibližně stejného průměrného zlepšení poměru signálu od šumu. Metody se však výrazně lišily v rozptylu hodnot zlepšení poměru signálu od šumu. Největší rozptyl vykazovala podle očekávání metoda využívající exponenciální průměrování. Při použití Bartlettovy metody rozptyl mírně poklesl, významnější pokles však nastal až při využití AR modelu. Nejmenšího rozptylu pak dosahovala navržená metoda prahování koeficientů waveletové transformace.

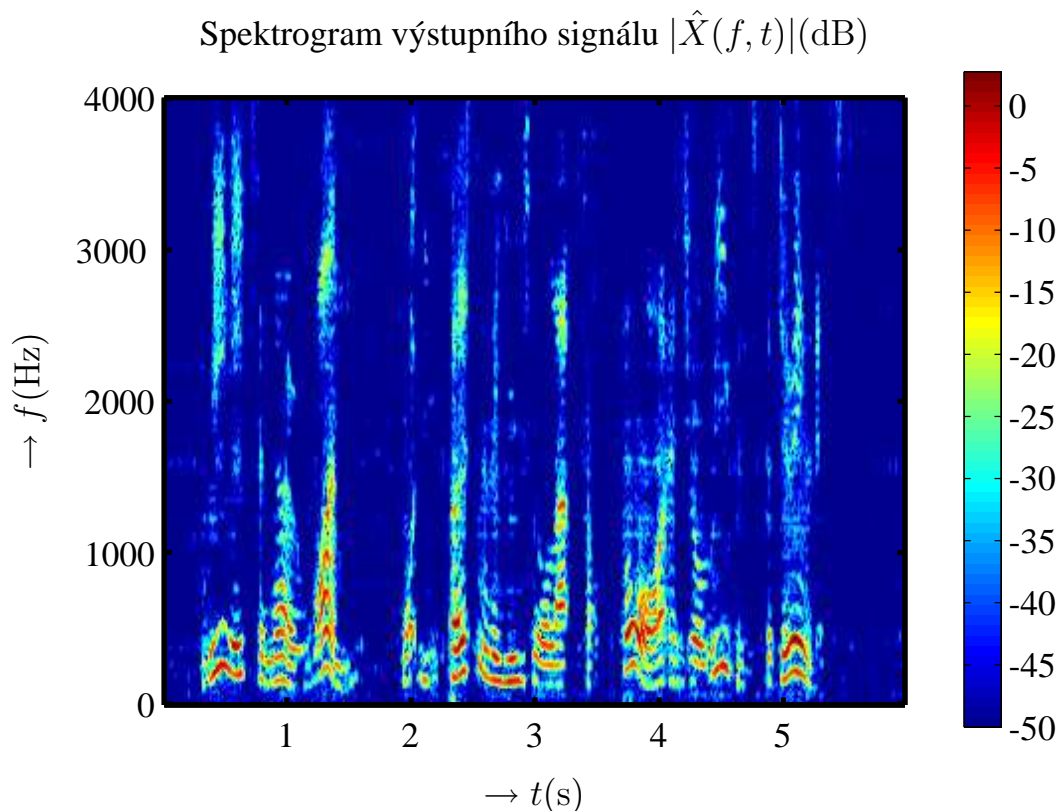
Tab. 6.1: Vyhodnocení metody spektrálního odečítání s různými metodami odhadu výkonové spektrální hustoty.

Metoda odhadu PSD	Zlepšení poměru signálu od šumu $\Delta\kappa$ (dB)			
	Minimum	Maximum	Průměr	Rozptyl
exponenciální průměrování	2,17	21,67	8,65	17,97
Bartlettova metoda	2,72	21,58	9,26	16,53
prahování koeficientů DTWT	3,52	20,10	8,68	10,19
AR model $s = 20$	3,69	23,87	10,02	11,69

Nejvhodnější byla navržená metoda pro rušení, které má náhodný charakter nebo které kromě harmonických složek obsahují velký podíl šumu. Mezi tyto rušení patří např. tekoucí voda (děšť, sprcha, splav), hluk vysavače, mixéru nebo vrtačky, kde kromě harmonických složek způsobených otáčením rotoru je přítomen náhodný signál způsobený třením proudu vzduchu nebo náplní mixéru. Lze také konstatovat, že metoda ponechává ve výstupním signálu méně hudebního šumu než při odhadu výkonové spektrální hustoty rušení pomocí Bartlettovy metody. Ve spektrogramu výstupního signálu (viz 6.2) je méně šumu typu pepř a sůl. Rušení, které případně v signálu zůstává, má charakter déle trvajících tónů, které však mají nižší úroveň než ve vstupním signálu.



Obr. 6.1: Spektrogram řečového signálu zarušeného hlukem vysavače.



Obr. 6.2: Spektrogram výstupního signálu po zpracování metodou spektrálního odečítání s odhadem výkonové spektrální hustoty pomocí prahování waveletové transformace.

Metoda odhadu výkonové spektrální hustoty pomocí prahování waveletových koeficientů není vázána pouze na metodu spektrálního odečítání. Metodu je možné také použít i v metodě Wienerovy filtrace nebo v dalších metodách, které používají odhad výkonové spektrální hustoty rušení. Cílem dalšího výzkumu bude využití navržené metody odhadu i v metodě mapování spektrogramu, kdy přesnější odhad výkonové spektrální hustoty může vést k lepší identifikaci a lokalizaci složek řeči od složek rušení. Námětem pro další výzkum je také nalezení metody vyhlazení odhadu výkonové spektrální hustoty pomocí transformací typu frame, které by mohly poskytnout ještě přesnější odhady než doposud používané metody.

## LITERATURA

- [1] BARTLETT, M. S. Smoothing periodograms from time series with continuous spectra. *Nature*. May 1948, vol. 164, pp. 686–687. London.
- [2] BOLL, S. F. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transaction on Acoustics, Speech, Signal Processing*. April 1979, vol. ASSP-27, pp. 113–120.

- [3] DAUBECHIES, I. *Ten Lectures on Wavelets*. 7. edition. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics, 2002. 357 p. ISBN 0-89871-271-2.
- [4] DONOHO, D. L. – JOHNSTONE, I. M. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 1994, vol. 3, no. 81, pp. 425–455.
- [5] EKSLER, V. Evaluation of blind separated signals using speech recognition system. In *Proceedings of the International Conference on Computer as a tool EUROCON 2005*. Belgrade, Serbia and Montenegro: 2005. pp. 1650–1653. ISBN 1-4244-0050-3.
- [6] FISHER, S. – SIMMER, K. U. Beamforming microphone arrays for speech acquisition in noisy environments. *Speech Communication*. 1996, no. 3-4, pp. 215–229.
- [7] HERMAN, H. – WAN, W. A. – AVENDANO, C. Speech enhancement based on temporal processing. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* May 1995. pp. 405–408.
- [8] JOHN R. DELLER, J. – HANSEN, J. H. L. – PROAKIS, J. G. *Discrete-Time Processing of Speech Signals*. Reprint edition. New York: John Wiley, 2000. 936 p. ISBN 0-7803-5386-2.
- [9] LEE, T. W. *Independent Component Analysis*. First edition. Boston: Springer, October 1998. 248 p. ISBN 0-792-38261-7.
- [10] MALAH, D. – COX, R. – ACCARDI, A. Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments. In *Proceedings of the 24<sup>th</sup> IEEE ICASSP'99*. Phoenix: AS CR, March 1999. pp. 789–792. ISBN 808626915.
- [11] MOULIN, P. Wavelet thresholding techniques for power spectrum estimation. *IEEE Transactions on Signal Processing*. November 1994, vol. 42, no. 11, pp. 3126–3136.
- [12] PERCIVAL, D. B. – WALDEN, A. T. *Wavelet Methods for Time Series Analysis*. First edition. New York: Cambridge University Press, 2000. 620 p. ISBN 0-521-64068-7.
- [13] PLŠEK, M. The basic spectrogram mapping method of noisy speech signal. In *Proceedings of the Research in Telecommunication Technology RTT'2003*. Bratislava: 2003. pp. 1–4. ISBN 80-227-1934-X.
- [14] PROAKIS, J. G. – MANOLAKIS, D. G. *Digital Signal Processing Principles, Algorithms, and Applications*. Third edition. New Jersey: Prentice Hall, 1996. 1016 p. ISBN 0-13-373762-4.



- [15] SOVKA, P. – POLLAK, P. – KYBIC, J. Extended spectral subtraction. In *Proc. of European Conference on Signal Processing and Communication*. Trieste: September 1996.
- [16] STEJSKAL, V. Empty pause detection in noisy and clean speech conditions. In *Proceedings of the 16<sup>th</sup> Czech-German Workshop on Speech Processing*. Praha: AS CR, 2006. pp. 125–254. ISBN 808626915.
- [17] UHLÍŘ, J. – SOVKA, P. *Číslicové zpracování signálů*. 1 vydání. Praha: Vydavatelství ČVUT, 1995. 313 s. ISBN 80-01-01303-0.
- [18] VASEGHI, S. V. *Advanced Digital Signal Processing and Noise Reduction*. Second edition. New York: John Wiley & Sons, 2000. 456 p. ISBN 0-471-62692-9.
- [19] VIDA KOVIC, B. *Statistical Modeling by Wavelets*. First edition. New Jersey: Wiley, 1999. 408 p. Probability and Statistics. ISBN 0-471-29365-2.
- [20] VIRAG, N. Single channel speech enhancement based on masking properties of the human auditory system. *IEEE Transaction on Speech and Audio Processing*. March 1999, vol. 7, no. 2.
- [21] VONDRA, M. – VÍCH, R. Comb filtering in speech enhancement. In *Proceedings of 16<sup>th</sup> Czech-German Workshop SPEECH PROCESSING*. Praha: IREE AS CR, 2006. pp. 73–77. ISBN 80-86269-15-9.
- [22] WALDEN, A. T. – PERCIVAL, D. B. – MCCOY, E. J. Spectrum estimation by wavelet thresholding of multitaper estimators. *IEEE Transactions on Signal Processing*. 1998, vol. 46, no. 12, pp. 3153–3165.
- [23] WELCH, P. D. The use of fast fourier transform for the estimation of power spektra: a metod based on time averaging over short modified periodograms. *IEEE Trans. Audio nad Electroacoustics*. June 1967, vol. AU-15, pp. 70–73.
- [24] WICKERHAUSER, M. V. *Adapted Wavelet Analysis from Theory to Software*. First edition. Wellesley: AK Peters, July 1994. 504 p. ISBN 1-56881-041-5.
- [25] YEN, K. C. – ZHAO, Y. Adaptive co-channel speech separation and recognition. *IEEE Transactions on Speech and Audio Processing*. March 1999, vol. 07, no. 02, pp. 138–152.

## CURRICULUM VITAE

Jméno: **Petr Sysel**  
Narozen: 27. 1. 1976 v Prostějově  
Kontakt: sysel@feec.vutbr.cz

### Vzdělání

1990-1994 Střední průmyslová škola Prostějov, obor Zařízení silnoproudé elektrotechniky, zaměření Elektrické stroje a přístroje, zakončeno maturitní zkouškou  
1994-1999 Vysoké učení technické v Brně, Fakulta elektrotechniky a informatiky, obor Elektronika a sdělovací technika, magisterské studium zakončené státní zkouškou,  
1999-2007 Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, obor Teleinformatika, postgraduální doktorské studium

### Praxe

2002-dosud zaměstnanec Ústavu telekomunikací, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně

### Účast na nejvýznamnějších projektech

GA102/00/1084 RTD technologie hláskové separace zamaskované v šumu  
FD-K/040 Aplikace digitální separace řeči v komunikačních technologiích  
OC277.002 Non-linear methods of speech enhancement  
GA102/04/1097 Zvýrazňování řečového signálu zamaskovaného v šumu  
1ET301710509 Sofistikované potlačovače hluku a poruch při přenosu řečových signálů pro pevné a mobilní sítě nové generace  
FT-TA2/072 Výzkum a aplikace metod časově-frekvenční analýzy pro logopedii

### Pedagogické aktivity

Podíl na výuce kurzů Číslicové zpracování signálů, Číslicové filtry, Signálové procesory, Multimediální a grafické procesory. Vedení bakalářských a diplomových prací z oblasti číslicového zpracování signálů.

### Další aktivity

2001 člen redakční rady internetového časopisu Elektrovue  
2001-dosud člen české sekce mezinárodní organizace Audio Engineering Society  
2005 Členství ve výboru konference *Audio Technologies and Processing ATP 2005*. Praha: ČVUT, 24. 5. 2005.  
<http://aes.cz/atp2005/cz/comitee.htm>

## **ABSTRACT**

This doctoral thesis deals with methods of enhancing speech embedded in noise. In the first part the speech enhancement methods, which are currently used, are summarized and methods of time-frequency analysis are introduced, namely the short time Fourier transform and the wavelet transform. In the next part, methods of power spectral density estimation (periodogram, Bartlett method, AR models) are discussed and their characteristics are examined and compared (estimation bias, estimation variance). Based on this study a new method of power spectral density estimation via thresholding the wavelet coefficients of the periodogram is proposed that exhibits a smaller estimation variance than the other methods. Finally the enhancement of noisy speech records from an actual environment by spectral subtraction using various methods of power spectral density estimation is realized and the result is evaluated. The proposed method is suitable for noise that has a markedly random character such as shower, mixer noise, vacuum-cleaner noise, etc.