

**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

**FAKULTA ELEKTROTECHNIKY  
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

Ústav telekomunikací

**VYUŽITÍ WAVELETOVÉ TRANSFORMACE  
A MATEMATICKÉ STATISTIKY PRO SEPARACI  
SIGNÁLU A ŠUMU**

**UTILIZATION OF WAVELET TRANSFORM AND  
MATHEMATICAL STATISTICS FOR SEPARATING  
SIGNAL FROM NOISE**

*zkrácená verze Ph.D. Thesis*

**Mgr. Pavel Rajmic**

Obor: Teleinformatika

Školitel: Prof. Ing. Zdeněk Smékal, CSc.

Oponenti: Prof. Ing. Andrej Lúč, CSc.

Ing. Robert Vích, DrSc.

Doc. RNDr. Vítězslav Veselý, CSc.

Datum obhajoby: 7. 12. 2004

## **KLÍČOVÁ SLOVA**

waveletová transformace, potlačování šumu v signálu, statistická analýza rizikové funkce, segmentovaná waveletová transformace

## **KEYWORDS**

wavelet transform, signal denoising, exact risk analysis, segmented wavelet transform

Dizertační práce je uložena na oddělení vědy a výzkumu, Fakulta elektrotechniky a komunikačních technologií, Vysoké učení technické v Brně, Údolní 53, 602 00 Brno.

## OBSAH

ÚVOD .....	5
1 PŘEHLED SOUČASNÉ PROBLEMATIKY .....	5
1.1 Metody separace signálů od šumu a hluku .....	5
1.2 Waveletové metody zpracování signálu .....	7
2 CÍLE DIZERTAČNÍ PRÁCE .....	9
3 WAVELETOVÁ TRANSFORMACE A WAVELETOVÉ PAKETY .....	10
4 WAVELETOVÉ POTLAČOVÁNÍ ŠUMU .....	10
4.1 Statistická analýza prahovacích pravidel .....	12
4.2 Testování waveletových metod na signálech s aditivním šumem .....	15
5 NOVÁ METODA SEGMENTOVANÉ KONEČNÉ DISKRÉTNÍ WAVELETOVÉ TRANSFORMACE .....	17
5.1 Princip algoritmu segmentované waveletové transformace .....	19
ZÁVĚR .....	20
LITERATURA .....	22
SUMMARY OF THE THESIS .....	26



# ÚVOD

Waveletová neboli vlnková analýza je moderní disciplína v oblasti zpracování signálů, která má po krátkém, ale velmi prudkém a intenzivním vývoji velkou oblast uplatnění nejen v čistě technických oborech. Mezi nejvýznamnější aplikace waveletového zpracování signálů patří separace signálů od aditivního šumu z jejich směsí.

Dizertační práce se zabývá metodou potlačování gaussovského šumu v signálu tzv. prahováním waveletových koeficientů. Tento způsob separace užitečného a rušivého signálu vychází z metodologie *WaveShrink* [9]. Druhou podstatnou částí práce je kapitola pojednávající o nově vyvinuté metodě tzv. segmentované waveletové transformace, která umožňuje provádět waveletovou transformaci, a tedy v důsledku i tuto „waveletovou separaci“ kromě dalších výhod v *reálném čase*.

## 1 PŘEHLED SOUČASNÉ PROBLEMATIKY

Zhodnocení současného stavu problematiky sestává ze dvou částí. V části 1.1 jsou uvedeny metody, které využívají principů Fourierovy transformace, znalostí charakteristických vlastností řečových signálů, sofistikovaných statistických přístupů apod. Pozornost části 1.2 bude obrácena na waveletové metody zpracování signálů se zaměřením na používané metody separace deterministických signálů od šumu.

Šumem přitom zpravidla rozumíme náhodný proces, který má přesně definované statistické parametry (jako např. rozložení, střední hodnota, rozptyl, autokorelační funkce), přičemž tyto parametry nemusí být známé. Hlukem naopak rozumíme signál, který statistickým zákonitostem v tomto úzkém slova smyslu nepodléhá. Příkladem může být hluk motoru, vysavače, dopravního provozu ad.

### 1.1 METODY SEPARACE SIGNÁLŮ OD ŠUMU A HLUKU

Jednodušší separační techniky jsou v současné době nedílnou součástí každého komunikačního zařízení pracujícího v zarušeném prostředí. Většinou se jedná o jednoduché algoritmy s malou účinností vlastní separace. Ukazuje se, že nejobtížnější pro separaci jsou parazitní hluky, které kmitočtově spadají do pásma užitečné řeči. Takových zvuků je ale v praktickém provozu většina. Vhodný algoritmus by měl být proto schopen patřičně reagovat jak na širokopásmové šумы, tak impulsní hluky a trvalé rušící periodické signály.

Metody pro potlačení parazitního rušení v řečovém signálu lze rozdělit na *jednokanálové* a *vícekanálové*. V současné době jsou většinou v praxi používány jednoduché jednokanálové techniky založené na potlačení vyšších, případně nižších kmitočtů složek spektra vstupního signálu v závislosti na velikosti odstupu signálu od rušení (SNR). Je zřejmé, že klasickou číslicovou filtrací rušivých signálů sice docílíme potlačení hluku, ale v užitečném pásmu kmitočtů dále přetrvává, pouze je vlivem omezené šířky přenášeného pásma pro lidský sluch méně nápadný. V případě užití adaptivní lineární filtrace (například algoritmu typu LMS) vzniká nežádoucí přídavný efekt, který

způsobí, že výstupní signál ztrácí svou dynamiku, hlas ztrácí barvu a zní strojově.

Podstatně dokonalejší jednokanálovou metodou, užívanou firmou Motorola v mobilních telefonech pro jednotky „handsfree“, je technika RASTA, která využívá pásmovou filtraci časově proměnného spektra signálu, tzv. časových trajektorií [14]. Na základě kmitočtové analýzy časového průběhu jednotlivých spektrálních složek modulových spekter (časových trajektorií), je vypočteno spektrum časových trajektorií, které je zvykem nazývat modulačním spektrem. Aplikací vhodného číslicového filtru dochází k omezení modulačního spektra parazitního signálu, přitom modulační spektrum řeči musí zůstat nezměněno. Potlačení trvalých nebo naopak velmi rychle se v čase měnících (rychleji než řeč) parazitních signálů je dostatečné, ovšem pro většinu běžných hluků není separace příliš účinná. Jinými slovy, separace je velmi neefektivní v případě, kdy spektrum hluku má charakter spektra řečového signálu.

K robustnějším technikám potlačování rušení z řečových signálů patří metoda nazvaná mapování spektrogramu, vyvinutá na FEKT VUT v Brně [19]. Principem metody je adaptivní vytváření tzv. mapy (masky), kterou se následně násobí spektrogram řečového signálu, tím se signál „filtruje“, a poté se převede zpět do časové oblasti. Výsledky testování ukázaly, že metoda vykazuje excelentní separační vlastnosti v případě, kdy je možno optimálně stanovit velikost tzv. prahu pro dlouhé úseky signálu (až několik sekund). To ale v praxi není možné, neboť požadujeme možnost nasazení metody v reálném čase. Tehdy je nutné práh nastavovat velmi rychle a to často vede k odchýlkám od jeho optimální hodnoty, což se projevuje nežádoucími modulacemi řeči.

Pro rušení, které má charakter blízký stacionárnímu, je velmi vhodná metoda spektrálního odečítání [8] a její modifikace. V řečových pauzách je analyzován kmitočtový obsah rušení a v době řečové aktivity je toto rušení ve spektrální oblasti odečítáno. Ukázalo se, že se snižujícím se poměrem SNR drasticky klesá účinnost metody [32]. To vedlo ke vzniku různých modifikací této metody [18, 34]. Spektrální odečítání má vynikající separační vlastnosti, pokud hluk v pozadí má v čase neměnný charakter. Z jejího principu zároveň vyplývá, že pro impulzní rušení je naprosto nevhodná.

Odlišným přístupem k problematice separace je vývoj *vícekanálových* metod schopných pracovat v reálném čase. Tento typ metod má daleko větší nároky na výpočetní výkon než metody jednokanálové. Současně se zvyšujícím se výkonem dnešních signálových procesorů se vícekanálovým metodám dostává stále více pozornosti. Metody můžeme rozdělit do dvou skupin na metody pracující s exaktně determinovaným mikrofonním polem a na „slepé“ metody pracující přímo s několika směsmi, bez jakékoliv přídavné informace o poloze mikrofonů. V prvním případě problém řešíme deterministicky na základě souřadnic mikrofonů, řešení druhého případu nabízí matematická statistika – metoda ICA [15].

Výhodou vícekanálového přístupu je značné zlepšení výsledného SNR oproti jednokanálovým metodám, přičemž subjektivní hodnocení kvality je ještě výrazně lepší. Další výhodou je, že kvalita separace není podmíněna stacionárním charakterem rušení. Naopak nevýhodou je, že nevhodná konfigurace mikrofonního pole vede téměř

k nemožnosti separace signálů.

Nejnovější přístupy k problematice separace vycházejí z originální myšlenky tzv. resyntézy řeči. Algoritmy tohoto typu vycházejí z velmi dobré znalosti principu vzniku řeči v hlasovém traktu člověka. Nesnaží se v parazitním hluku nalézt řečový signál jako celek, ale detekují pouze základní parametry řečového signálu [33]. Z těchto parametrů pak, pomocí standardních technik syntézy řeči, rekonstruujeme užitečnou promluvu. Praktická realizace uvedeného principu je však zatím v počátcích.

Celkově se dá shrnout, že žádná ze separačních metod není zcela univerzální, každá má své kladné i záporné rysy. Nicméně u všech metod lze vysledovat, že potlačování hluku zkresluje také samotnou řeč. Tento jev se samozřejmě stává čím dál výraznějším se snižujícím se odstupem signálu od šumu (hluku). Nastavení parametrů separace (a tím její výsledek) je proto *vždy nutné volit jako kompromis mezi mírou potlačení hluku a zkreslením samotné řeči*. Stejně tvrzení platí přirozeně i o waveletových metodách separace (část 1.2.2).

## 1.2 WAVELETOVÉ METODY ZPRACOVÁNÍ SIGNÁLU

Zpracování signálů pomocí waveletové transformace je v posledních letech značně využívaná disciplína se stále se rozvíjející škálou aplikací při zpracování mnoha typů signálů. Jde o relativně nový přístup k analýze signálů, jehož základní principy byly položeny již před několika desítkami let, avšak jehož bouřlivý vývoj odstartoval v osmdesátých letech 20. století. Své uplatnění při analýze a syntéze signálů našla waveletová transformace zejména v oborech jako jsou telekomunikace, sonarová a radarová technika, seizmologie, meteorologie, lékařství. V těchto oborech se používá buď pro časově-kmitočtovou analýzu signálů, pro rekonstrukci a restauraci neúplných či silně zarušených signálů, nebo pro kompresi dat. Na tomto posledně jmenovaném poli dosahuje waveletová transformace vynikajících výsledků, zejména při kompresi digitálních obrazů [28]. Přehled méně obvyklých (avšak neméně důležitých) aplikací viz [26].

Hlavní zásluhou waveletové analýzy je, že přinesla nový typ reprezentace signálů a rozšířila tak soubor nástrojů pro jejich zpracování. Ve srovnání s obvyklým fourierovským přístupem pozorujeme lepší, ale i horší vlastnosti. Použití waveletových technik tedy *není univerzální*, ale pro každý typ úlohy je třeba dobře zvážit jeho vhodnost.

Matematická teorie waveletové transformace se vyvíjela dlouhá léta vedle prvních praktických inženýrských aplikačních experimentů, aniž by se tyto oblasti střetly. Počátkem osmdesátých let dvacátého století upozornil na vzájemnou úzkou provázanost Stéphane Mallat a odstartoval tím „waveletovou revoluci“. Spoluprací mezi zmíněnými oblastmi se poté velmi intenzivním způsobem začala dotvářet jak sofistikovaná teorie, tak stále nové a nové aplikace. Budoucí vývoj v této oblasti nyní směřuje k systémům, které jsou po formální stránce zobecněním waveletů a jež jsou více přizpůsobeny konkrétním úlohám na úkor univerzálnosti jejich použití.



### 1.2.1 Využití waveletových metod při zpracování řeči

Waveletová transformace se uplatnila také v některých oblastech zpracování řečových signálů. Protože řeč má přibližně stacionární povahu pouze pro její velmi krátké úseky, fourierovský typ zpracování, který vychází z nekonečně dlouhých harmonických básových funkcí, není pro určité aplikace vhodný. Waveletová transformace, umožňující ze své podstaty reprezentativnější pohled na signál z krátkodobého hlediska, se pokusila standardní metody zpracování řeči vylepšit.

Publikace [20] ukazuje, že při *detekci základního tónu řeči*  $F_0$  má waveletová transformace několik výhod včetně odstranění problému dvojnásobení základního kmitočtu. V oblasti *segmentace řeči na fonémy* [27] waveletová transformace nepředčila tradiční metody, zejména z toho důvodu, že řečový signál je velmi proměnlivý a není možné všechny jeho aspekty zachytit pomocí pevné waveletové báze. Přínosem v oblasti *potlačování různých typů šumů* v pozadí řeči se stala metoda waveletového vyhlazování spektrogramu šumu [17]. Použití této metody podle výzkumné zprávy [25] přineslo zlepšení SNR v průměru o 3–5 dB.

### 1.2.2 Waveletové metody pro separaci signálu od šumu

Základ pro waveletové techniky separace byl položen v článku [9]. Zde byl uveden princip úpravy waveletového spektra po ortogonální transformaci za účelem odstranění šumu. Metodologie, kterou její tvůrci nazvali *Waveshrink*, vychází z vlastnosti transformace koncentrovat podstatnou část energie do několika málo waveletových koeficientů. V případě, že signál není pohlcen šumem, stačí vypustit „malé“ koeficienty, které reprezentují šumovou složku.

Na základě této metody byla postupně vystavěna různá zobecnění a vylepšení. Například Silverman a Johnstone [11] zobecnili tuto metodu i pro barevný (korelovaný) šum. Dále byly zkoumány přínosy přístupu „wavelet shrinkage“ [31], což je ještě obecnější přístup k úpravě koeficientů než prahování. Původně pro kompresi otisků prstů FBI byla určena nová metoda, kterou přinesli Wickerhauser a Coifman [5] a která se nazývá „wavelet packets“. Metoda se ukázala být vhodnou pro daleko širší škálu problémů. Stejní autoři zároveň ukázali, jak vybírat optimální waveletovou bázi, která se co nejvíce přimyká charakteru zpracovávaného signálu. Při zvolení takovéto báze je možné dosáhnout lepších výsledků i v *separačních úlohách*.

U waveletové separace signálu od šumu pomocí *prahování* je nutné zvolit tzv. prahovací pravidlo, které říká, jakým způsobem bude zacházeno s koeficienty, které nebyly vynulovány. V základní metodě existují dvě pravidla – tvrdé a měkké. Vzhledem k tomu, že obě tato základní pravidla kromě specifických výhod také trpí specifickými nedostatky, byla později uvedena ještě další pravidla, která se snaží příhodné vlastnosti obou typů kombinovat [31, 23].

Na prahovací pravidla můžeme nahlížet z perspektivy matematické statistiky a v tomto smyslu posuzovat jejich vlastnosti. Statistickou technikou „Exact Risk Analysis“ lze odvodit několik veličin, které charakterizují základní znaky prahovacích



pravidel [2, 12, 13]. V literatuře však chybí právě takový typ analýzy u tzv. hyperbolického prahování, což je jeden z nestandardních typů; zřejmě to je z důvodu, že je nutno analyticky vyjádřit velmi speciální integrály. Rovněž chybí širěji pojaté pojednání, které by se zabývalo srovnáním vlastností všech používaných pravidel v jednom celku. Statistické vlastnosti bývají v literatuře uváděny pro speciální případ jednotkového rozptylu šumu, tj.  $\sigma = 1$ . Důkaz, že tento předpoklad není na újmu obecnosti, a zejména jakým způsobem vlastnosti přepočítat na případ šumu s jiným rozptylem, v dostupné literatuře také není uveden.

Nutno poznamenat, že waveletové metody separace signálů jsou omezeny na případ, kdy aditivní šum má známé (nebo alespoň odhadnutelné) statistické vlastnosti. Čím více se šum „blíží“ bílému, tím vyšší účinnost metody zaznamenávají. Pro separaci dvou signálů deterministického charakteru (např. řeč a hluk motoru) metody selhávají.

### 1.2.3 Waveletová transformace realizovatelná v reálném čase

Za současného stavu vývoje problematiky není možné aplikovat waveletovou transformaci pro zpracování signálů v reálném čase. Autoři se zabývají teoretickými i praktickými aspekty waveletové transformace, avšak pouze pro zpracování signálu „off-line“<sup>1</sup>; v současném stavu poznání chybí sofistikovaný algoritmus, který by tento způsob zpracování umožnil.

Jestliže někteří autoři přistupují k sekvenčnímu waveletovému zpracování dat, tzn. po segmentech, např. [6, 24], volí jejich délku a překrytí ad hoc a kompenzují pak vzniklá zkreslení průměrováním či vážením výstupních segmentů. V dizertační práci je prezentován nově vyvinutý algoritmus, pomocí něhož se určí překryvy segmentů a způsob jejich zpracování tak, že *k žádnému zkreslení nedochází*. Vývoj takového algoritmu je úkol značně obtížný a vyžaduje hlubokou znalost algoritmu waveletové transformace, neboť do úvahy se musí vzít volitelné veličiny jako délka segmentu, délka waveletového filtru a v neposlední řadě počet úrovní dekompozice.

## 2 CÍLE DIZERTAČNÍ PRÁCE

Na základě rozboru současného stavu problematiky v částech 1.2.2 a 1.2.3 byly formulovány cíle této dizertační práce:

- Provést srovnání pěti druhů prahovacích pravidel na základě statistické analýzy rizikové funkce. K tomu je třeba nalézt vzorce střední hodnoty, rozptylu a rizikové funkce pro hyperbolické prahování, případně vyjádřit tyto statistiky alespoň numericky.

Uvést důkaz, že předpoklad jednotkového rozptylu  $\sigma = 1$  není na újmu obecnosti, a uvést z toho plynoucí výpočty statistických veličin pro případ  $\sigma \neq 1$  pomocí předchozího případu.

---

<sup>1</sup>tzn. předpokládáme-li znalost celého průběhu signálu předem

- Prozkoumat, zda a jak velký přínos má změna waveletové báze (technika „wavelet packets“, která se v minulosti ukázala přínosnou v oblasti efektivní komprese obrazových dat) pro řečové signály zatížené šumem. Tento přínos posoudit kvantitativně i kvalitativně.
- Vytvořit algoritmus, pomocí kterého bude možné waveletové zpracování diskrétního signálu po segmentech, tj. zejména v reálném čase, bez ztráty informace v něm obsažené. Nový algoritmus bude přitom využívat postupy používané ve standardní waveletové transformaci.

### 3 WAVELETOVÁ TRANSFORMACE A WAVELETOVÉ PAKETY

V dizertační práci byl rozveden koncept tzv. waveletových paketů (transformace typu wavelet packet, zkratka WP). Tento koncept tvoří zobecnění obvyklé waveletové transformace (zkratka WT) v tom smyslu, že při rozkladu signálu pomocí pyramidového algoritmu podrobujeme dekompozici nejen vektory obsahující detailní koeficienty, ale i všechny vektory s aproximačními koeficienty.

Tím dosáhneme úplné struktury stromového typu se zvolenou hloubkou dekompozice  $d$ . Informace obsažená v této množině waveletových koeficientů je přirozeně redundantní, čehož lze ale s úspěchem využít. Pomocí algoritmu „nejlepší báze“ [5] lze z waveletového paketu pro daný signál vybrat množinu koeficientů, které vystihují chování tohoto signálu lépe než koeficienty u WT. Toho se využívá např. v moderních metodách komprese číslicových obrazů a jak prokázaly výsledky dizertační práce (viz dále), tato metoda je přínosná i pro zpracování řečových signálů. Algoritmus nejlepší báze pracuje na základě ohodnocení stromové struktury pomocí některého druhu *entropie*.

### 4 WAVELETOVÉ POTLAČOVÁNÍ ŠUMU

Nežádoucí šum v signálu má obvykle aditivní charakter a proto je vhodné uvažovat model

$$\mathbf{X} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad (1)$$

kde  $\mathbf{X}$  je náhodný vektor (konečný diskrétní náhodný signál),  $\mathbf{f}$  je deterministický (čistý) signál a  $\boldsymbol{\varepsilon}$  je náhodný vektor s  $n$ -rozměrným normálním rozložením,  $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Waveletovou transformací vyjádřenou pomocí matice  $\mathbf{W}$  přejde vztah (4) s využitím linearit v

$$\mathbf{Y} = \mathbf{b} + \boldsymbol{\nu}, \quad (2)$$

kde  $\mathbf{Y} = \mathbf{W}\mathbf{X}$ ,  $\mathbf{b} = \mathbf{W}\mathbf{f}$ ,  $\boldsymbol{\nu} = \mathbf{W}\boldsymbol{\varepsilon}$ . Díky ortogonalitě transformace platí  $\boldsymbol{\nu} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ , tedy rozložení pravděpodobnosti šumu zůstalo stejné. Označme pozorování náhodných vektorů  $\mathbf{X}$ ,  $\boldsymbol{\varepsilon}$ ,  $\mathbf{Y}$ ,  $\boldsymbol{\nu}$ , pořadě symboly  $\mathbf{x}$ ,  $\mathbf{e}$ ,  $\mathbf{y}$ ,  $\mathbf{v}$ . Vektory  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{v}$  tedy

představují waveletové koeficienty příslušné vektorům  $\mathbf{x}$ ,  $\mathbf{f}$ ,  $\mathbf{e}$ .

Cílem waveletového potlačování šumu je nalézt vhodný modifikační předpis  $\delta(\cdot)$  tak, aby odhad  $\hat{b}_i = \delta(y_i) \approx b_i$  byl v nějakém smyslu dobrým odhadem  $b_i$ . Koeficienty  $b_i$  totiž jsou waveletové koeficienty příslušící signálu *nezatíženého šumem*. Po úpravě waveletových koeficientů navrátíme pomocí matice  $\mathbf{W}^T = \mathbf{W}^{-1}$  signál do výchozí domény. Tím získáme signál s potlačenou šumovou složkou. Tuto v principu jednoduchou metodologii nazvali její autoři *WaveShrink* [9].

Popsaný přístup je analogií klasické fourierovské filtrace, kde podobně upravujeme Fourierovy koeficienty signálu. Avšak u waveletové analýzy je příspěvek každého koeficientu pouze lokální (neboť wavelet je tlumený směrem k  $\pm\infty$ ), takže waveletová reprezentace dovoluje tímto způsobem konstruovat *lokálně adaptivní filtry*, což je vynikající rys ve srovnání s fourierovskými filtry, kde efekt každého koeficientu je obecně globální [29].

Existuje několik technik, jak konstruovat odhady  $\hat{b}_i = \delta(y_i)$ . Mezi nejpoužívanější metody patří *prahovací techniky* [31]. Jejich princip je následující: hodnoty  $y_i$ , které jsou v absolutní hodnotě menší než zvolený práh  $\lambda > 0$ , jsou vynulovány a ostatní jsou ještě dále upraveny. Je zřejmé, že prahování je nelineární operací na datech. Podle způsobu úprav koeficientů  $y_i$  rozlišujeme těchto pět druhů prahování:

1. **Tvrdé prahování**, obrázek 1(a):

$$\hat{b}_i = \delta^h(y_i, \lambda), \text{ kde}$$

$$\delta^h(x, \lambda) = \begin{cases} 0 & \text{pro } |x| \leq \lambda \\ x & \text{pro } |x| > \lambda. \end{cases}$$

2. **Měkké prahování**, obrázek 1(b):

$$\hat{b}_i = \delta^s(y_i, \lambda), \text{ kde}$$

$$\delta^s(x, \lambda) = \text{sgn}(x) \max\{0, |x| - \lambda\}.$$

3. **Poloměkké prahování**, obrázek 1(c):

Toto pravidlo závisí na dvou parametrech  $0 < \lambda_1 < \lambda_2$ ;  $\hat{b}_i = \delta^{ss}(y_i, \lambda_1, \lambda_2)$ , kde

$$\delta^{ss}(x, \lambda_1, \lambda_2) = \begin{cases} 0 & \text{pro } |x| \leq \lambda_1 \\ \text{sgn}(x) \frac{\lambda_2(|x| - \lambda_1)}{\lambda_2 - \lambda_1} & \text{pro } \lambda_1 < |x| \leq \lambda_2. \\ x & \text{pro } |x| > \lambda_2. \end{cases}$$

4. **Nezáporná garota<sup>2</sup>**, obrázek 1(d):

$$\hat{b}_i = \delta^{\text{nng}}(y_i, \lambda), \text{ kde}$$

$$\delta^{\text{nng}}(x, \lambda) = \begin{cases} 0 & \text{pro } |x| \leq \lambda \\ x - \frac{\lambda^2}{x} & \text{pro } |x| > \lambda. \end{cases}$$

---

<sup>2</sup>garota – železné škrtidlo, z anglického non-negative garrotte

## 5. Hyperbolické prahování, obrázek 1(e):

$\hat{b}_i = \delta^{\text{hy}}(y_i, \lambda)$ , kde

$$\delta^{\text{hy}}(x, \lambda) = \begin{cases} 0 & \text{pro } |x| < \lambda \\ \text{sgn}(x)\sqrt{x^2 - \lambda^2} & \text{pro } |x| \geq \lambda. \end{cases}$$

### 4.1 STATISTICKÁ ANALÝZA PRAHOVACÍCH PRAVIDEL

Vlastnosti odhadů posuzujeme pomocí statistické techniky Exact Risk Analysis – analýza rizika [31]. Vycházíme z modelu (5) a uvažujeme *jedinou* náhodnou veličinu  $Y_i$ , kterou pro jednoduchost budeme dále značit pouze  $Y$ . Bez újmy na obecnosti můžeme nadále předpokládat jednotkový rozptyl šumu, tedy  $Y \sim N(b_i, 1)$ . Místo  $\delta(Y, \lambda)$  budeme dále pro jednoduchost psát pouze  $\delta(Y)$ .

Prahovací pravidla  $\delta^{\text{h}}$ ,  $\delta^{\text{s}}$ ,  $\delta^{\text{ss}}$ ,  $\delta^{\text{nng}}$ ,  $\delta^{\text{hy}}$  posuzujeme z hlediska:

1. **střední hodnoty**  $M_\lambda(b) = E[\delta(Y)]$ ,
2. **rozptylu**  $V_\lambda(b) = \text{var}[\delta(Y)]$ ,
3. **rizika**  $R_\lambda(b) = E[\delta(Y) - b]^2$ .

pro pevně zvolenou hodnotu prahu  $\lambda > 0$ . Nejdůležitější informací přitom obsahuje hodnota rizika, neboť udává odchylku od „správné“ hodnoty  $b$ , nikoliv od její střední hodnoty  $M_\lambda(b)$ .

Označme příslušnost výše uvedených statistik k prahovacím pravidlům horním indexem, tj. např. rozptyl měkkého prahování bude označen  $V_\lambda^{\text{s}}(b)$ . V dizertační práci jsou odvozeny vzorce pro všechna prahovací pravidla. Na základě těchto vzorců je provedeno vzájemné srovnání statistických vlastností všech pravidel.

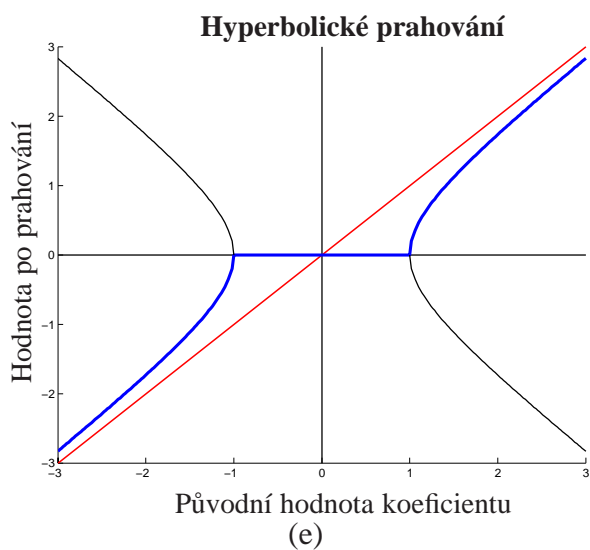
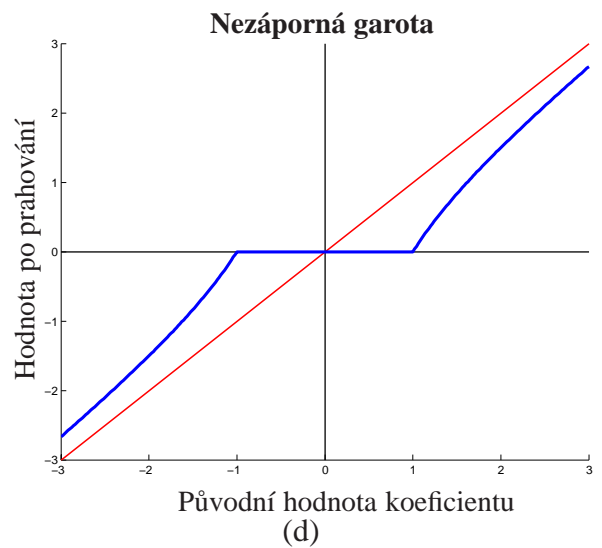
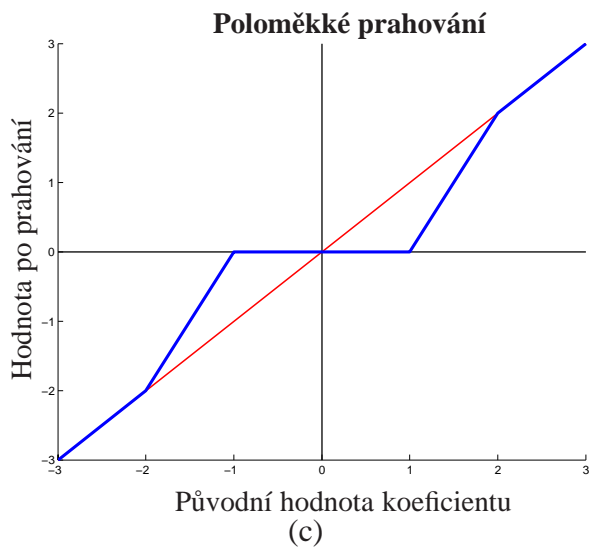
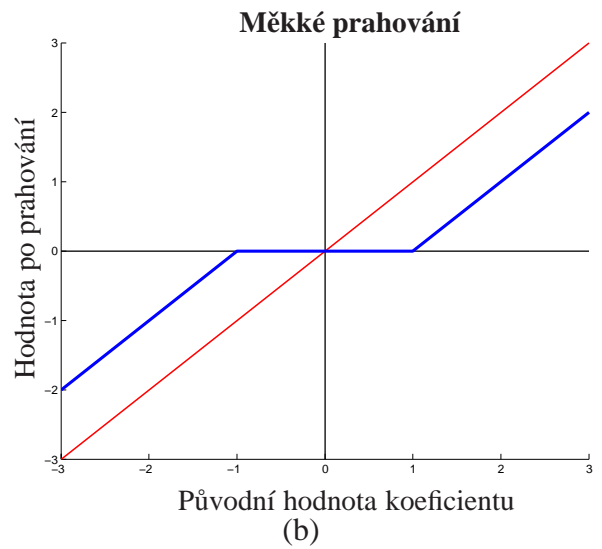
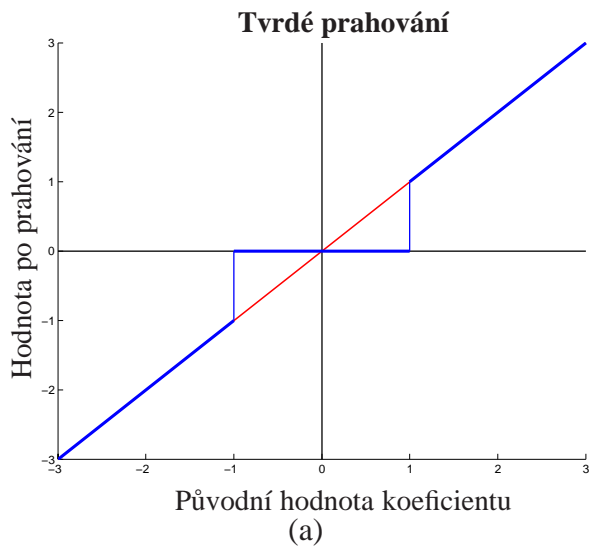
Pro účely této analýzy bylo nutné odvodit vzorce střední hodnoty, rozptylu a rizika pro hyperbolické prahování, které se v současné odborné literatuře nevyskytují. Tyto vzorce se nám podařilo odvodit, avšak v konečné fázi nestandardní integrál, který obsahují, se nepodařilo vyjádřit analyticky, a to ani jako nekonečnou řadu. Proto jsme pro praktické výpočty použili obvyklých metod numerické integrace.

Na obrázku 2 je vyobrazeno srovnání rizika všech pěti zmíněných pravidel.

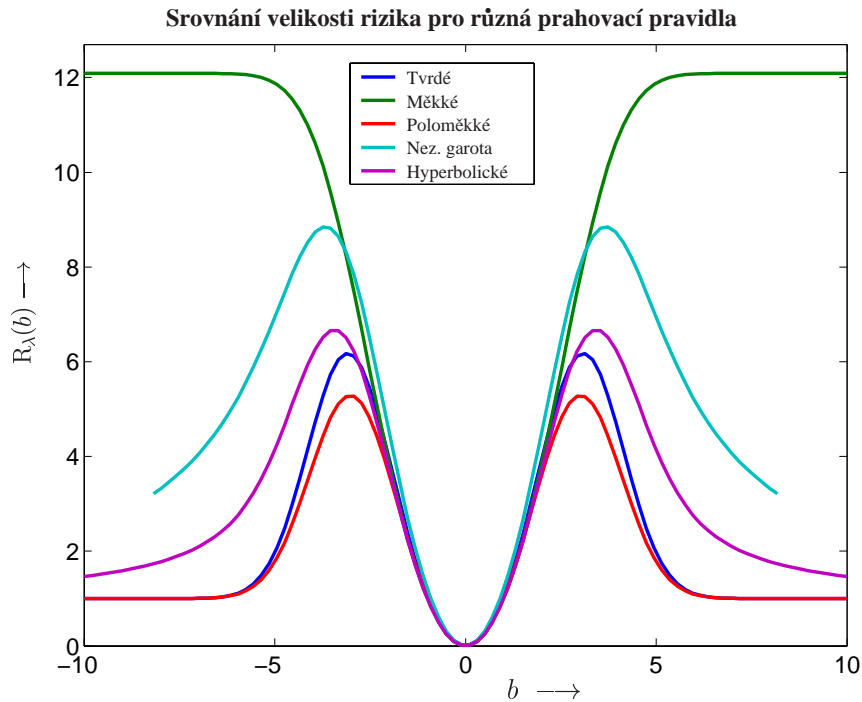
Statistická analýza umožnila vyvodit následující důsledky [2]:

**Důsledek 1** Pro pevné  $\lambda \in \mathbb{R}^+$  má měkké prahování menší rozptyl než tvrdé prahování pro všechna  $b$ , tj.  $V_\lambda^{\text{s}}(b) \leq V_\lambda^{\text{h}}(b)$ .

**Důsledek 2** Pro pevné  $\lambda \in \mathbb{R}^+$  má měkké prahování mnohem větší výchylku než tvrdé prahování pro  $|b|$  dostatečně velké. Také platí, že pro  $b \rightarrow \pm\infty$ :  $R_\lambda^{\text{s}}(b) \rightarrow 1 + \lambda^2$ , zatímco  $R_\lambda^{\text{h}}(b) \rightarrow 1$ . To znamená, že průměrná kvadratická odchylka od správné hodnoty je pro rostoucí  $b$  vyšší u měkkého prahování, a to tím větší, čím větší je zvoleno  $\lambda$ . Riziko tvrdého prahování v limitě na hodnotě prahu nezávisí. Celkově to znamená, že k detailním waveletovým koeficientům, které představují významné rychlé změny v signálu, a mají tedy relativně velké  $b$ , se tvrdé prahování chová



**Obr. 1: Grafy prahovacích pravidel. Pro všechna prahovací pravidla zvoleno  $\lambda = 1$ , kromě poloměkkého prahování, kde  $\lambda_1 = 1$  a  $\lambda_2 = 2$ .**



**Obr. 2: Srovnání rizika pěti prahovacích pravidel. Zde  $\lambda = 3,33$ ,  $\lambda_1 = 2,86$ ,  $\lambda_2 = 3,72$ . Je zřetelné, že stejnoměrně nejmenšího rizika dosahuje poloměkké prahování, avšak jen v případě optimální volby  $\lambda_1, \lambda_2$ .**

obežretně – ponechá jim svou původní velikost, kdežto měkké prahování jejich velikost sníží. Tato vlastnost se odráží v tom, že *měkké prahování má tendenci „přehladit“ rychlé změny, skoky a špičky v signálu.*

**Důsledek 3** Pro pevné  $\lambda \in \mathbb{R}^+$  má tvrdé prahování maximální výchylku, rozptyl a riziko pro hodnoty  $b \approx \pm\lambda$ . Měkké prahování má maximální výchylku, rozptyl a riziko pro velké hodnoty  $|b|$ .

**Důsledek 4** Pro pevné  $\lambda \in \mathbb{R}^+$  má měkké prahování menší riziko v počátku  $b = 0$ . Obě statistiky  $R_\lambda^s(0)$  a  $R_\lambda^h(0)$  jsou monotónně klesající pro rostoucí  $\lambda$ .

Ostatní uvedené typy prahování tvoří kompromis mezi tvrdým a měkkým prahováním. Gao a Bruce [13] ukázali, že pro dané  $\lambda$  existují  $\lambda_1 < \lambda < \lambda_2$  tak, že pro všechna  $b_i$  platí  $E[\delta^{ss}(y_i, \lambda_1, \lambda_2) - b_i]^2 < E[\delta^h(y_i, \lambda) - b_i]^2$ , tedy že poloměkké prahování  $\delta^{ss}$  má stejnoměrně menší riziko než  $\delta^h$ . Tato skutečnost je vidět na obrázku 2, kde byla zvolena vhodná  $\lambda_1, \lambda_2$ . Volbou parametrů  $\lambda_1, \lambda_2$  můžeme  $\delta^{ss}$  podle potřeby více „přiblížit“ vlastnostem tvrdého nebo měkkého prahování. Poloměkké prahování má výhodu v tom, že překonává citlivost tvrdého prahování na malé změny ve vstupních datech, a zároveň zamezuje vzniku velké výchylky pro velká  $|b|$ , jak se tomu děje u měkkého prahování. Nutnost volby dvou prahových hodnot je však velkou nevýhodou tohoto prahování. Postup určení optimálních prahů je algoritmicky i výpočetně velmi náročný [13], je potřeba použít dvojrozměrnou Newtonovu iterační metodu.

Zbývá dvě prahovací pravidla, nezáporná garota a hyperbolické prahování, se snaží s tímto hendikepem vyrovnat, a to tak, že napodobují tvar poloměkkého prahování



funkcí s jedním volitelným parametrem – prahem. Statistické vlastnosti poloměkkého prahování jsou přitom v asymptotickém smyslu zachovány. Gao [12] ukázal, že nezáporná garota je lepší než tvrdé i měkké prahování ve smyslu velikosti rizika a také citlivosti na malé změny ve výchozích datech. Také hyperbolické prahování kombinuje vlastnosti tvrdého a měkkého prahování, a z obrázku 2 je zřejmé, že z hlediska rizika je úspěšnější než nezáporná garota.

Následující tvrzení, jehož důkaz je uveden v dizertační práci, umožňuje všechny předchozí úvahy rozšířit na náhodnou veličinu s obecným normálním rozložením a zároveň dává návod, jak v takovém případě spočítat střední hodnotu, rozptyl a riziko.

**Tvrzení:** Pokud  $Y \sim N(b, \sigma^2)$ , pak  $\frac{Y}{\sigma} \sim N(b/\sigma, 1)$  a platí

1.  $\delta(Y, \lambda) \equiv \sigma \delta\left(\frac{Y}{\sigma}, \frac{\lambda}{\sigma}\right)$ ,
2.  $M_\lambda(b) = \sigma M_{\lambda/\sigma}\left(\frac{b}{\sigma}\right)$ ,
3.  $V_\lambda(b) = \sigma^2 V_{\lambda/\sigma}\left(\frac{b}{\sigma}\right)$ ,
4.  $R_\lambda(b) = \sigma^2 R_{\lambda/\sigma}\left(\frac{b}{\sigma}\right)$ .

## 4.2 TESTOVÁNÍ WAVELETOVÝCH METOD NA SIGNÁLECH S ADITIVNÍM ŠUMEM

Míra úspěšnosti waveletového potlačování šumu byla testována a srovnávána na řečovém signálu. Šlo o promluvu mužského mluvčího „*vysyp všechny pytle s pšenici*“. Tento signál, vzorkovaný s kmitočtem 16 kHz, byl uměle zarušen gaussovským šumem s různými rozptyly tak, že k dispozici byly testovací signály se SNR 15 dB, 10 dB, 5 dB, 0 dB a  $-5$  dB. Po separaci řeči a šumu (viz níže) byl odhadnut výstupní SNR vztahem

$$\widehat{\text{SNR}} [\text{dB}] = 10 \log_{10} \frac{\|\mathbf{y}\|}{\|\mathbf{y} - k\hat{\mathbf{y}}\|}, \quad (3)$$

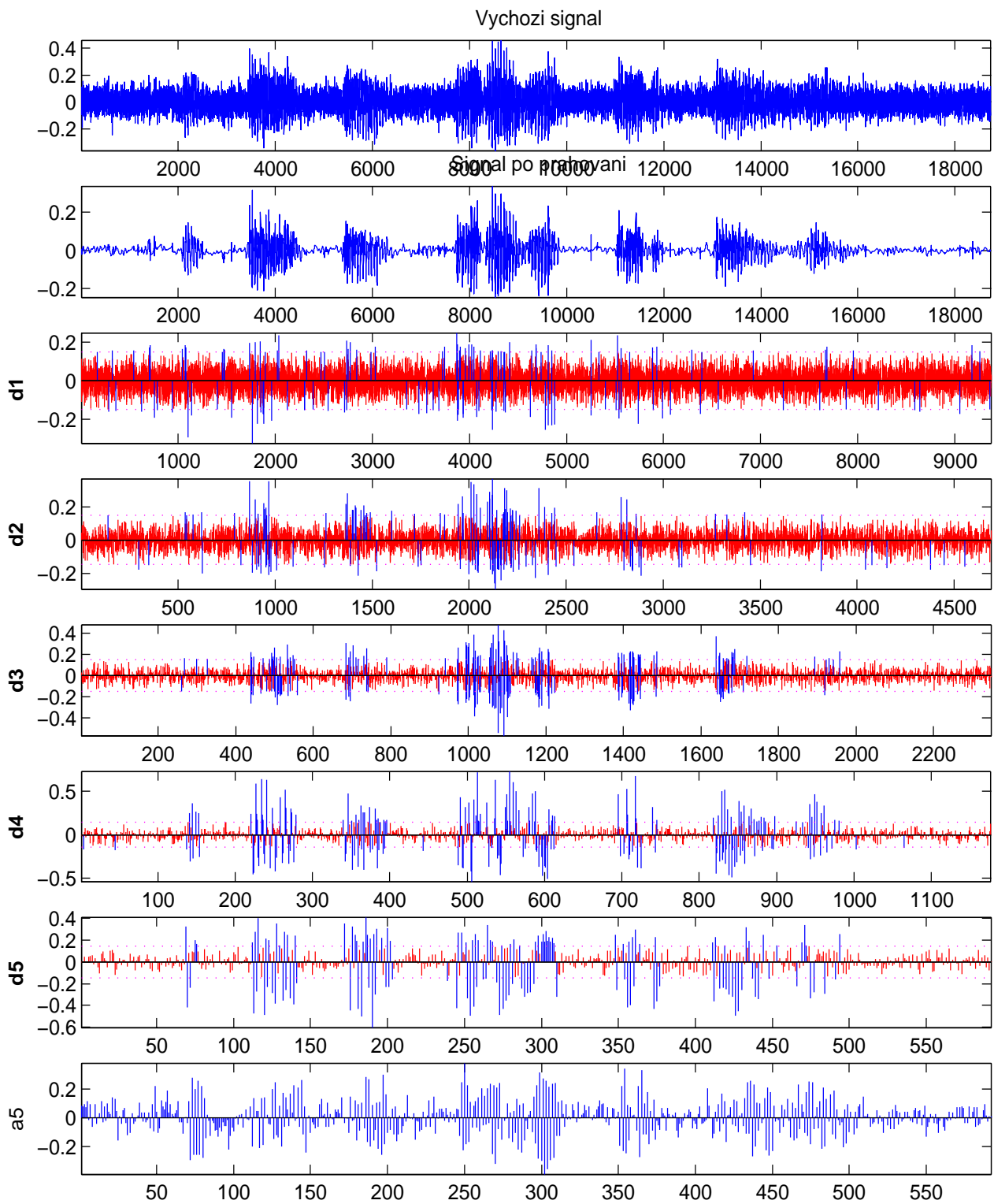
kde  $\hat{\mathbf{y}}$  je signál s potlačenou šumovou složkou a  $k = \frac{\|\mathbf{y}\|}{\|\hat{\mathbf{y}}\|}$  je kompenzační konstanta.

Při všech analýzách byl použit wavelet Daubechies řádu 8 a hloubka dekompozice (a také hloubka prahování) byla zvolena  $d = 5$ . Testování probíhalo v systému MATLAB pomocí balíku ThreshLab (vytvořeného v rámci práce). Příklad grafického znázornění prahování je na obr. 3.

Pro každý z pěti testovacích SNR byla nejprve stanovena prahová hodnota  $\lambda$ , přičemž výchozím odhadem byla tzv. univerzální hodnota prahu  $\lambda_{\text{univ}}$  [9]. Použitá prahová hodnota byla nakonec zvolena subjektivně. Pro tuto hodnotu bylo provedeno tvrdé, měkké, hyperbolické prahování a nezáporná garota, a to:

- [WP] detailních koeficientů klasické waveletové transformace do hloubky  $d = 5$ ,
- [WP] koeficientů odpovídajících nejlepší bázi ve waveletovém paketu hloubky  $d = 5$ . Nejlepší báze byla vyhledána pomocí tzv. prahové entropie s prahem  $\lambda$ .





**Obr. 3: Příklad prahování zarušeného signálu. Shora dolů jsou vykresleny: původní zarušený signál, separovaný signál a prahované waveletové koeficienty.**

Výsledky testování jsou obsaženy v tabulce 1. V tabulce lze nalézt kromě objektivního parametru výstupního  $\widehat{\text{SNR}}$  také subjektivní zhodnocení kvality separace (stanovení pořadí úspěšnosti). Kritérii subjektivního posuzování byly stejnou měrou *příjemnost poslechu* (množství rušivých artefaktů v signálu) a *srozumitelnost* odrušené nahrávky.

Celkově lze k výsledkům říci:

Nezaznamenali jsme velký rozdíl v parametru  $\widehat{\text{SNR}}$  mezi metodami WT a WP, avšak rozdíl mezi nimi v subjektivním vnímání je podstatný: objevují-li se v odrušeném signálu nežádoucí artefakty, pak u WT se většinou jedná o osamocené „lupance“ nebo jejich rychlý sled, na rozdíl od WP, kde dochází k rozostření šumu a k efektu „bublání“ na pozadí, což je lidskému sluchu mnohem příjemnější na poslech. Rovněž u WP pozorujeme lepší srozumitelnost.

Nejméně úspěšné pro zpracování *řečového* signálu je měkké prahování, a to jak z objektivního, tak subjektivního hlediska. Nejhorší výsledky  $\widehat{\text{SNR}}$  jsou paradoxem, neboť jak bylo řečeno výše, právě měkké prahování nejlépe potlačuje šum. Toto tvrzení ale zároveň musíme potvrdit ze subjektivního pohledu – skutečně je tomu tak, avšak u řečového signálu tím zanikají neznělé souhlásky jako zejména *s, š, c, č, z, ž*, které jsou právě pro srozumitelnost řeči nejpodstatnější.

V subjektivním hodnocení dosáhlo obecně nejlepších výsledků hyperbolické prahovací pravidlo, i když se snižujícím se vstupním SNR mu výrazně konkuruje tvrdé prahování. To lze vysvětlit tím, že směrem k malým hodnotám SNR přirozeně klesá srozumitelnost odrušených signálů a právě tvrdé prahování i přes své nepříjemné artefakty srozumitelnost zachovává nejdéle. Ve chvíli, kdy kvalita poslechu je již pro všechny typy prahování velmi nízká, dáváme přednost tvrdému, které dává nejsrozumitelnější výsledek.

Ukázalo se, že nastavení hodnoty prahu je vždy záležitostí kompromisu mezi mírou potlačení šumu a zachováním co nejpřirozenějšího charakteru řeči, tak jak je tomu u všech metod jednokanálové separace. Výsledný poměr  $\widehat{\text{SNR}}$  lze umělým snižováním prahu zpravidla ještě zlepšovat, avšak subjektivní ohodnocení rychle klesá.

## 5 NOVÁ METODA SEGMENTOVANÉ KONEČNÉ DISKRÉTNÍ WAVELETOVÉ TRANSFORMACE

V praktických aplikacích je často třeba signál zpracovávat v „reálném čase“, tzn. s minimálním zpožděním. Předem neznámý signál přichází na vstup nějakého systému po úsecích – na sobě nezávislých segmentech – které musejí být efektivně zpracovány a modifikované poslány na výstup systému. Typickým příkladem je např. zpracování akustických signálů, zejména řečových signálů v telekomunikacích.

Nová metoda nazvaná segmentovaná konečná waveletová transformace (SegWT), která tento typ zpracování umožňuje, má velké potencionální využití také v případě, že je třeba zpracovat dlouhý signál (ne nutně v reálném čase), ale přitom není k dis-

	SNR = 15 dB		SNR = 10 dB		SNR = 5 dB		SNR = 0 dB									
	WT	WP	WT	WP	WT	WP	WT	WP								
Prahování	$\widehat{\text{SNR}}$ [dB]	Pořadí subjekt. hodnocení	$\widehat{\text{SNR}}$ [dB]	Pořadí subjekt. hodnocení	$\widehat{\text{SNR}}$ [dB]	Pořadí subjekt. hodnocení	$\widehat{\text{SNR}}$ [dB]	Pořadí subjekt. hodnocení								
tvrdé	26,4	3	26,3	3	17,5	2	17,6	3	9,2	3	9,2	1	2,1	1	2,1	1
měkké	22,3	3	22,6	4	15,2	4	15,4	4	8,1	4	8,2	4	1,6	4	1,7	1
nezáp. garota	24,6	2	24,7	2	16,5	2	16,7	1	8,7	1	8,8	1	1,9	2	1,9	1
hyperbolické	25,6	1	25,6	1	17,0	1	17,2	1	9,0	1	9,0	1	2,0	2	2,0	1

Tab. 1: Výsledky testování na zarušených signálech s různým vstupním SNR.

pozici dostatečná výpočetní síla nebo paměťový prostor. Tehdy je možné touto novou metodou signál ekvivalentně zpracovat po částech a tím ušetřit výpočetní výkon i paměťový prostor. V tomto smyslu SegWT koresponduje s algoritmy přičtení přesahu (overlap-add) a úschovy přesahu (overlap-save) [10] u fourierovské lineární filtrace. Dalším možným využitím algoritmu SegWT je okamžitá vizualizace signálu pomocí zobrazení, které nazýváme „waveletogram“. Waveletogram zobrazuje graficky hodnoty waveletových koeficientů. Je to technika příbuzná vykreslování spektrogramu v reálném čase. U waveletové transformace máme však výhodu, že signál nemusíme váhovat okny, čímž dochází ke zkreslení kmitočtové informace, jak je tomu právě u spektrogramu. Waveletogram vytvořený pomocí SegWT je navíc zcela *nezávislý na zvolené délce segmentu*.

Úkolem segmentované konečné diskrétní waveletové transformace je tedy zpracovat signál po jednotlivých segmentech, a to tak, abychom po postupném zpracování všech segmentů obdrželi stejný výsledek (tj. stejné waveletové koeficienty), jako kdybychom signál znali celý předem a zpracovali ho pomocí běžně používaného algoritmu DTWT. Vývoj takového algoritmu vyžaduje detailní znalosti výchozího algoritmu DTWT. Nový algoritmus SegWT musí navíc brát do úvahy vztahy mezi zvolenou hloubkou waveletové dekompozice  $d$ , délkou waveletového filtru  $m$  a délkou segmentu signálu  $s$ .

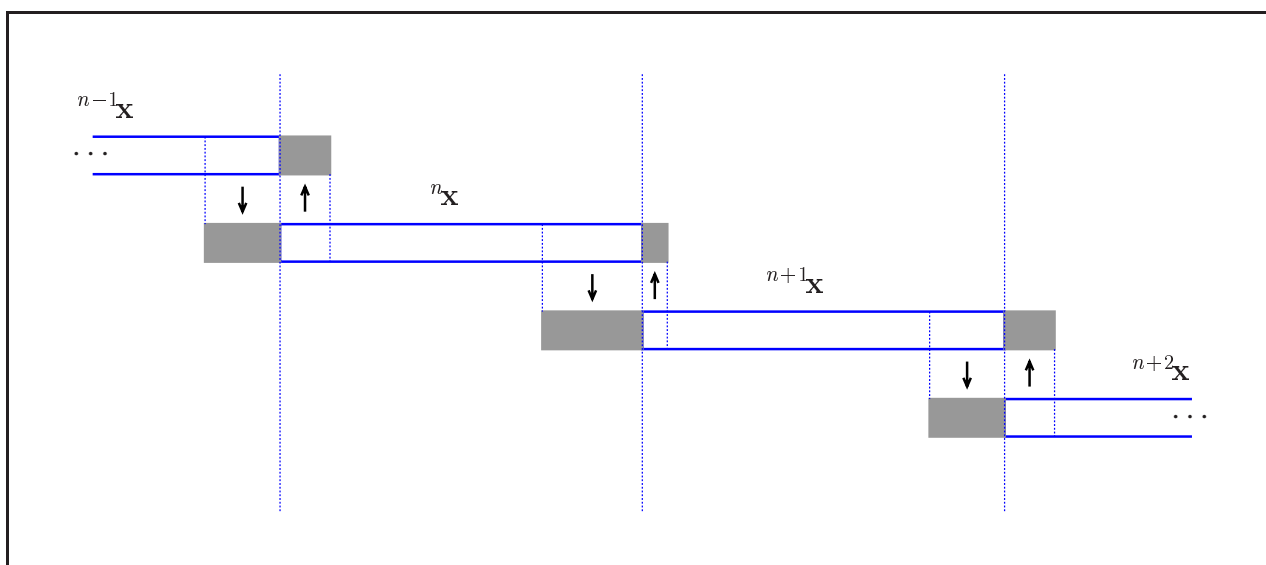
## 5.1 PRINCIP ALGORITMU SEGMENTOVANÉ WAVELETOVÉ TRANSFORMACE

Pro účely vysvětlené výše je nejprve nutno jednotlivé vstupní segmenty před provedením modifikované transformace upravit. Konkrétně jde o to, že segmenty délky  $s$ , které označujeme  $^1x, ^2x, \dots$  musíme vhodným způsobem prodloužit, aby mezi nimi nastalo optimální překrytí. Rovněž musíme samostatně řešit zacházení s okrajovými segmenty (tj. s prvním a posledním).

Lze ukázat, že délka překrytí segmentů musí splňovat přísná pravidla. Dokonce při obecné volbě  $s$  platí, že navzájem různé segmenty mají obecně různé délky jim příslušných pravých a levých prodloužení. Schéma segmentace a prodloužování segmentů je znázorněno na obr. 4. Po optimálním překrytí jednotlivých segmentů se z nich vypočítá *modifikovaným* způsobem waveletová transformace. Tato modifikace spočívá v tom, že vynecháváme určité kroky prováděné v klasickém algoritmu DTWT. Navíc je nutné zvláštním postupem zpracovat první a poslední segment. Vypočtené waveletových koeficienty ze všech segmentů nakonec jednoduchým způsobem spojíme dohromady.

Co se týká zpoždění, algoritmus SegWT pracuje tak, že ve chvíli, kdy je znám segment číslo  $n$  a segment číslo  $n + 1$ , je možno vypočítat waveletové koeficienty příslušející segmentu  $n$ . To znamená, že zpoždění je rovno  $s$  vzorkům plus čas potřebný na výpočet waveletových koeficientů z aktuálního segmentu. Ve speciálním případě, kdy  $s$  je dělitelné číslem  $2^d$ , však dokonce platí, že zpoždění algoritmu je determinováno *pouze délkou výpočtu!*

Algoritmus je možno téměř bez úprav využít i pro transformaci typu wavelet packet.



Obr. 4: Grafické znázornění segmentace a prodlužování segmentů. Každý segment je prodloužen zleva o určitý počet posledních vzorků předchozího segmentu a zprava o určitý počet prvních vzorků segmentu následujícího.

## ZÁVĚR

V dizertační práci byla prezentována metoda potlačování šumu v signálu pomocí waveletové transformace a některé její modifikace za účelem zlepšení separačních vlastností. Rovněž byla uvedena nová metoda tzv. segmentované waveletové transformace.

Nejvýznamnější technikou ve waveletovém potlačování šumu je tzv. prahování. Byla provedena detailní analýza pěti tzv. prahovacích pravidel na základě statistické techniky *Exact Risk Analysis*. Srovnáním dvou hlavních typů prahování, *tvrdého* a *měkkého*, jsme vyvodili závěry, že tvrdé prahování hůře potlačuje šum, ale lépe zachovává rychlé změny v signálu (skoky a špičky), zatímco měkké prahování je velmi úspěšné při potlačování šumu, ale má tendenci „přehladit“ skoky a špičky. Více než jen kompromis tvoří *poloměkké prahování*, které má při správné volbě dvou parametrů nejmenší riziko chyby. Ovšem dva parametry metody zároveň tvoří velkou nevýhodu oproti jiným metodám, které pracují pouze s parametrem jedním. Další dvě uvedená prahovací pravidla, *nezáporná garota* a *hyperbolické prahování*, se snaží tuto diskvalifikaci zmírnit. Pracují pouze s jedním parametrem, avšak ve statistickém smyslu nedosahují takové kvality jako poloměkké prahování.

Na základě testování uvedených waveletových metod na *řečových signálech* se ukázalo, že waveletové potlačování šumu vykazuje při správně nastavených parametrech dobré výsledky, zejména při hodnotách SNR nad 0 dB. V objektivním i subjektivním srovnání přitom metoda prahování nejlepší báze waveletového paketu (WP) předčila metodu prahování v klasické waveletové transformaci (WT), ovšem za cenu zvýšení výpočetní náročnosti. Nahrávky separované pomocí WP obsahují mnohem méně rušivých vjemů a jsou srozumitelnější. V rozsahu SNR od  $\infty$  přibližně do 3 dB se jako

nejvhodnější pro řečový signál jeví hyperbolické prahovací pravidlo. Pro nižší odstup signálu od šumu je pak již vhodnější použít tvrdé prahování.

Praktickým testováním se potvrdily závěry teoretické analýzy: měkké prahování skutečně nejlépe potlačuje šum, ale na úkor zachování detailů. Těmito detaily v řečovém signálu jsou však bohužel neznělé souhlásky *s, š, c, č, z, ž*, podstatné pro zachování srozumitelnosti. Z tohoto důvodu je měkké prahování nevhodné pro řečový signál.

Poslední částí dizertační práce je věnována uvedení nové metody tzv. *segmentované waveletové transformace*, která umožňuje zpracovávat signál po segmentech zvolené délky, a tudíž umožňuje jakýkoliv typ waveletového zpracování signálů *v reálném čase*, což za dosavadního stavu problematiky nebylo možné. Metoda má velký význam pro budoucí vývoj v širokém spektru aplikací, jako např. pro separaci řeči a šumu v telekomunikacích, zpracování hudebních signálů pomocí systému plug-in modulů nebo pro modulace waveletového typu, které se experimentálně začínají používat v systémech XDSL a SDR.

Součástí dizertační práce je rovněž programový balík ThreshLab pro MATLAB. Tento balík obsahuje funkce pro zpracování signálů pomocí všech typů prahování, pro numerickou i grafickou analýzu prahovacích pravidel, pro provedení a zobrazení MR-analýzy apod.

V budoucím výzkumu bude vhodné se zaměřit na metody obecného návrhu waveletových bazí [35], na tzv. frejmy (frames) [4], biortogonální systémy [26], nebo dokonce na reprezentaci signálu v neúplných systémech [30]. Předmětem zájmu by mohlo být také přenesení waveletových metod separace na směsi signálu se šumem, který již nemá gaussovské rozdělení (barevné šумы apod.). Rovněž hodláme optimalizovat algoritmus SegWT, jelikož jeho současná verze obsahuje duplicitu výpočtů na přechodech jednotlivých segmentů. Cestou v separaci *řečového signálu* a šumu by mohlo být adaptivní prahování [1], které by umělo přizpůsobit velikost prahu a typ prahování na základě určitých indikátorů jako je například znělost / neznělost aktuálního fonému.



## Literatura

- [1] ABRAMOVICH, F., BENJAMINI, Y.: Adaptive thresholding of wavelet coefficients. *Computational Statistics & Data Analysis*, 22, 351–361, 1996.
- [2] BRUCE, A. G., GAO, H.-Y.: Understanding WaveShrink: Variance and Bias Estimation. *Research Report 36*, StatSci Division of MathSoft, Inc., 1996.
- [3] CHEN, S. S., DONOHO, D. L., SAUNDERS, M. A.: Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* 20 (1998), no. 1, 33–61, reprinted in *SIAM Review*, 43 (2001), no. 1, pp. 129–159.
- [4] CHRISTENSEN, O.: *An Introduction to Frames nad Riesz Bases*. Birkhäuser, Boston-Basel-Berlin, 2003.
- [5] COIFMAN, R.R., WICKERHAUSER, M.V.: Entropy-based Algorithms for best basis selection. *IEEE Trans. on Inf. Theory*, vol. 38, 2, pp. 713-718. 1992.
- [6] DARLINGTON, D., DAUDET, L., SANDLER, M.: Digital Audio Effects in the Wavelet Domain. In *Proc. of the 5th Int. Conf. on Digital Audio Effects (DAFX-02)*, Hamburg, 2002.
- [7] DAVIS, G.M.: *Noise reduction in speech applications*. Davis, G.M., editor, CRC Press LLC, 2002. ISBN 0-8493-0949-2
- [8] DELLER, J.R., HANSEN, J.H.L., PROAKIS, J.G: *Discrete-Time Processing of Speech Signals (IEEE Press Classic Reissue)*. Wiley-IEEE Press; Reprint edition, 1999. ISBN 0780353862
- [9] DONOHO, D.L., JOHNSTONE, I.M.: Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81**(3), pp. 425–455, 1994.
- [10] JAN, J.: *Číslíková filtrace, analýza a restaurace signálů*. Druhé vydání, Vysoké učení technické v Brně, 2002.
- [11] JOHNSTONE, I.M., SILVERMAN, B.W.: *Wavelet threshold estimators for data with correlated noise*. Research report, Stanford University and Bristol University, 1994.
- [12] GAO, H-Y.: Wavelet shrinkage denoising using the non-negative garrote. *J. Comput. Graph. Statist.*, 7 (4), pp. 469–488, 1998.
- [13] GAO, H-Y., BRUCE, A.G.: WaveShrink with firm shrinkage. *Research Report 39*, StatSci Division of MathSoft, Inc., 1996.
- [14] HERMANŠKY, H., MORGAN, N.: RASTA Processing of Speech. *IEEE Trans. Speech and Audio Processing*, Vol. 2, No. 4, pp. 578–589, 1994.
- [15] HYVARINEN, A., KARHUNEN, J., OJA, E.: *Independent Component Analysis*. John Wiley & Sons, 2001. ISBN 0-471-40540-X
- [16] MALLAT, S.: *A Wavelet Tour of Signal Processing*. 2<sup>nd</sup> edition, Academic Press, 1999. ISBN : 0-12-466606-X
- [17] MOULIN, P.: Wavelet Thresholding Techniques for Power Spectrum Estimation. *IEEE Transactions on Signal Processing*, vol. 42, no. 11, November 1994.
- [18] PORUBA, J.: Speech Enhancement Based on Nonlinear Spectral Subtraction. In *Proc. 4th Int. Caracas Conference on Devices, Circuits and Systems*, Aruba, Dutch Caribbean, 2002, pp. T031-1 – T031-4. ISBN 0-7803-7381-2
- [19] PTÁČEK, L., VRBA, K., SMĚKAL, Z., a kol.: Metoda mapování spektrogramu. Výzkumná zpráva projektu MPO ČR FD–K/125, 2002.



- [20] QIU, L., YANG, H., KOH, S.N.: A Fundamental Frequency Detector of Speech Signals Based on Short Time Fourier Transform. In *Proc. Tencon '94*, vol. 1, pp. 526–530, IEEE 1994.
- [21] RAJMIC, P.: Wavelet Spectrum Thresholding Rules. In *Proceedings of the Fourth International Conference on Recent Advantages in Soft Computing*, Nottingham Trent University, Nottingham, 2002. ISBN 1-84233-0764.
- [22] RAJMIC, P.: Statistical Properties of Wavelet Spectrum Thresholding Rules. In *Proceedings of 48. Internationales Wissenschaftliches Kolloquium*, Technische Universität Ilmenau, 2003. ISSN 1619-4098
- [23] RAJMIC, P.: Exact Risk Analysis of Wavelet Spectrum Thresholding Rules. In *Proceedings of 10th IEEE International Conference on Electronics, Circuits and Systems*, Sharjah, United Arab Emirates. University of Sharjah, United Arab Emirates, 2003.
- [24] RAJMIC, P., KLIMEK, J.: Removing Crackle from an LP Record via Wavelet Analysis. In *Proc. of the 7th Conference on Digital Audio Effects (DAFx'04)*, Naples, Italy, Federico II University of Napoli, 2004. In print.
- [25] SMÉKAL, Z., HERMAN, I., KÉPESI, M., NAGY, Z., PLŠEK, M., RAJMIC, P., SYSEL, P., ŠKORPIL, V., VONDRA, M., VRBA, K.: *RTD technologie hláskové separace zamaskované v šumu. Závěrečná výzkumná zpráva projektu GAČR č. 102/00/1084*, FEKT VUT v Brně, Brno, 2002.
- [26] SWELDENS, W.: The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets. *Applied and Computational Harmonic Analysis*, 3, 186–200, 1996.
- [27] SYSEL, P.: Speech segmentation into phonemes by wavelet transform. In *Proc. Telecommunications and Signal Processing – TSP'2000*, pp. 126–129, Brno University of Technology, Brno, 2000.
- [28] UNSER, M., BLU, T.: Mathematical Properties of the JPEG2000 Wavelet Filters. *IEEE Trans. Image Processing*, vol. 12, no. 9, pp. 1080–1090, September 2003.
- [29] VESELÝ, V.: Wavelety a jejich použití při filtraci dat. *Proceedings of the Summer School ROBUST'96* (J. Antoch and G. Dohnal, eds.), pp. 241–272, JČMF, 1997.
- [30] VESELÝ, V.: Hilbert-space Techniques For Spectral Representation in Terms of Overcomplete Bases. In *Proceedings of the Summer school DATASTAT 01*, Czech Republic. Masarykova univerzita v Brně, 2002.
- [31] VIDA KOVIC, B.: *Statistical Modeling by Wavelets (Wiley Series in Probability and Statistics)*. John Wiley & Sons, New York, 1999.
- [32] VRBA, K., SMÉKAL, Z., PORUBA, J., RAJMIC, P.: *Vývoj alternativních metod separace řeči. Dílčí výzkumná zpráva projektu MPO č. FD-K/125*, FEKT VUT v Brně, 2002.
- [33] VRBA, K., SMÉKAL, Z., PORUBA, J., RAJMIC, P., PLŠEK, M., ŠTEFÍČEK, R.: *Možnosti alternativních metod separace řeči. Dílčí výzkumná zpráva projektu MPO č. FD-K/125*, FEKT VUT v Brně, 2003.
- [34] VIRAG, N.: Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System. *IEEE Transaction on Speech and Audio Processing*, vol. 7, no. 2, Piscataway, NJ, 1999.
- [35] ZHANG, J.-K., DAVIDSON, T.N., WONG, K.M.: Efficient Design of Orthonormal Wavelet Bases for Signal Representation. *IEEE Transactions on Signal Processing*, 52(7), pp. 1983-1996, July 2004.

# CURRICULUM VITÆ

Jméno: **Mgr. Pavel Rajmic**

Narozen: 17.5.1978 v Uherském Hradišti

Kontakt: rajmic@feec.vutbr.cz

## Vzdělání

- 1992 – 1996 Gymnázium Uherské Hradiště  
1996 – 2001 Masarykova univerzita v Brně, Přírodovědecká fakulta,  
magisterské studium oboru Matematika-ekonomie  
2001 – 2004 Vysoké učení technické v Brně, Fakulta elektrotechniky a komuni-  
kačních technologií, doktorské studium oboru Teleinformatika

## Praxe

- 2000 – 2001 Masarykova univerzita v Brně, Fakulta sociálních studií,  
Centrum výzkumu vývoje osobnosti a etnicity, zaměstnán jako  
statistik

## Účast na řešení projektů

V rámci doktorského studia jsem se podílel na řešení těchto projektů:

- FRVŠ 1823/2002/G1 Statistické prahování waveletového spektra při extrakci  
signálu ze šumu  
FRVŠ 2206/2003/G1 Využití matematické statistiky ke konstrukci efektivních  
waveletových bazí  
FRVŠ 1963/2002/G1 Zlepšení kvality a srozumitelnosti řeči v mobilních  
komunikacích  
GAČR 102/00/1084 RTD technologie hláskové separace zamaskované v hluku  
GAČR 102/04/1097 Zvýrazňování řečového signálu zamaskovaného v šumu  
MPO ČR FD-K/125 Aplikace digitální separace řeči v komunik. technologiích  
MPO ČR FD-K/040 Příprava služby "digitální operátor": Nové metody  
potlačování rušivých signálů  
COST OC277 Mezinárodní projekt "Non-linear methods of speech  
enhancement"  
MŠMT LP0052 Prezentace výsledků výzkumu na veletrzích INVEX  
MSM 262200022 Výzkum mikroelektronických systémů a technologií  
MSM 262200011 Výzkum elektronických komunikačních systémů  
a technologií

## **Pedagogické aktivity**

Po celou dobu doktorského studia jsem se podílel na výuce předmětů Elektronické praktikum, Číslicové filtry, Signály a soustavy, Číslicové zpracování signálů. Vedl jsem pět ročníkových projektů a jednu diplomovou práci.

## **Další aktivity**

V roce 2003 jsem absolvoval tříměsíční odbornou stáž na KHBO Oostende (Belgie) v rámci programu Socrates–Erasmus. Svou práci pravidelně publikuji na zahraničních i tuzemských konferencích.

## SUMMARY OF THE THESIS

The dissertation thesis considers a method for suppressing Gaussian (normally distributed) noise in signals utilizing the so-called thresholding of wavelet coefficients, with particular focus on speech signals. This type of separation of effective and noisy signals starts from the so-called WaveShrink methodology. The first part of the thesis considers several modifications of this methodology with the aim of improving its performance, and, also, in particular the statistical analysis and comparison of five thresholding rules.

The second part is devoted to the newly-developed method of so-called segmented wavelet transform, which allows us to perform the wavelet transform, and thus also the “wavelet-type separation” in real time, in addition to other benefits.

## WAVELET-TYPE DENOISING

The undesirable noise in a signal is usually of additive character and hence it is convenient to consider the model

$$\mathbf{X} = \mathbf{f} + \boldsymbol{\varepsilon}, \quad (4)$$

where  $\mathbf{X}$  is the random vector (finite discrete random signal),  $\mathbf{f}$  is the “clean” signal, and  $\boldsymbol{\varepsilon}$  is the random vector with multivariate normal distribution,  $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Applying the wavelet transform represented by an orthogonal matrix  $\mathbf{W}$  to (4) results in

$$\mathbf{Y} = \mathbf{b} + \boldsymbol{\nu} \quad (5)$$

with the utilization of linearity;  $\mathbf{Y} = \mathbf{W}\mathbf{X}$ ,  $\mathbf{b} = \mathbf{W}\mathbf{f}$ ,  $\boldsymbol{\nu} = \mathbf{W}\boldsymbol{\varepsilon}$ . Due to the orthogonality of the transform there is  $\boldsymbol{\nu} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ . Let the observations of random vectors  $\mathbf{X}$ ,  $\boldsymbol{\varepsilon}$ ,  $\mathbf{Y}$ ,  $\boldsymbol{\nu}$  be denoted by  $\mathbf{x}$ ,  $\mathbf{e}$ ,  $\mathbf{y}$ ,  $\mathbf{v}$ , respectively. Hence the vectors  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{v}$  contain the wavelet coefficients corresponding to the vectors  $\mathbf{x}$ ,  $\mathbf{f}$ ,  $\mathbf{e}$ .

The goal of wavelet denoising is to find an appropriate modification rule  $\delta(\cdot)$  such that  $\hat{b}_i = \delta(y_i) \approx b_i$  is a good estimator of  $b_i$ . After modifying the wavelet coefficients the signal is returned back to its original domain by means of matrix  $\mathbf{W}^\top$ . In this way we get a signal with the noisy part suppressed. This methodology was proposed by Donoho and Johnstone [9] and called WaveShrink.

Several techniques have been proposed how to construct the estimator  $\hat{b}_i$ . One of the most frequently used techniques is the so-called thresholding [31]. Its principle is as follows:  $y_i$ 's, which are absolutely smaller than a selected threshold  $\lambda > 0$ , are set to zero and the others are further processed.

The wavelet transform has an excellent property in that it concentrates the significant part of energy contained in vector  $\mathbf{f}$  into just a few wavelet coefficients  $b_i$ . This together with  $\boldsymbol{\nu} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$  implies and explains the principle of wavelet denoising. If the threshold  $\lambda$  is set properly, the thresholding will not affect the  $y_i$ 's, which primarily represent the deterministic signal, and contrarywise it will affect the “small” coefficients  $y_i$ , representing the additive normal noise.

According to the method used in modifying the  $y_i$ 's we distinguish different thresholding rules: *hard thresholding*  $\hat{b}_i = \delta^h(y_i, \lambda)$ , *soft thresholding*  $\hat{b}_i = \delta^s(y_i, \lambda)$ , and – depending on two thresholds – *semisoft thresholding*  $\hat{b}_i = \delta^{ss}(y_i, \lambda_1, \lambda_2)$ , then also *non-negative garrotte*  $\hat{b}_i = \delta^{nng}(y_i, \lambda)$  and *hyperbolic thresholding*  $\hat{b}_i = \delta^{hy}(y_i, \lambda)$ .

## EXACT RISK ANALYSIS OF THE THRESHOLDING RULES

We start from model (5) and consider a single random variable  $Y_i$ , which will be denoted just  $Y$  for simplicity. Without any loss of generality we also assume unit variance  $\sigma^2 = 1$ , i.e.  $Y \sim N(b_i, 1)$ . We judge the thresholding rules  $\delta^h, \delta^s, \delta^{ss}, \delta^{nng}, \delta^{hy}$  from the viewpoint of their:

1. **mean**  $M_\lambda(b) = E[\delta(Y)]$ ,
2. **variance**  $V_\lambda(b) = \text{var}[\delta(Y)] = E[\delta(Y) - M_\lambda(b)]^2$ ,
3. **risk**  $R_\lambda(b) = E[\delta(Y) - b]^2$ .

for a fixed threshold  $\lambda > 0$ . For our purposes the risk carries the most important information, for it represents the bias from the “correct” value  $b$ , not from its mean  $M_\lambda(b)$ .

Let us use a superscript to denote that the above statistics pertain to the thresholding rules, e.g. variance of the soft thresholding will be denoted  $V_\lambda^s(b)$ . From the formulas we can derive the following [2, 22, 23]:

- For a fixed  $\lambda \in \mathbb{R}^+$  soft thresholding has a lower variance than the hard one has, for every  $b$ , i.e.  $V_\lambda^s(b) \leq V_\lambda^h(b)$ .
- For a fixed  $\lambda \in \mathbb{R}^+$  the soft thresholding has a much bigger bias than the hard one has, for  $|b|$  sufficiently big. It also holds  $R_\lambda^s(b) \rightarrow 1 + \lambda^2$  for  $b \rightarrow \pm\infty$ , while  $R_\lambda^h(b) \rightarrow 1$ .
- For a fixed  $\lambda \in \mathbb{R}^+$  the hard rule has maximal bias, variance and risk for values  $b \approx \pm\lambda$ . The soft rule has its maximal bias, variance and risk for large values of  $|b|$ .

These items imply that the soft rule tends to “oversmooth” the abrupt changes and peaks in signal but at the same time, it suppresses noise very effectively. On the other hand, the hard rule preserves the significant details in signal, but it is less effective in suppressing the noise.

Further we can claim that the other rules represent a compromise between hard and soft thresholding. Gao and Bruce [13] showed that for a given  $\lambda$  there exist  $\lambda_1 < \lambda < \lambda_2$  such that  $E[\delta^{ss}(y_i, \lambda_1, \lambda_2) - b_i]^2 < E[\delta^h(y_i, \lambda) - b_i]^2$  for all  $b_i$ , that is,  $\delta^{ss}$  produces, uniformly in  $b_i$ , a smaller risk than  $\delta^h$  does. In contrast to its best risk performance this rule has a severe drawback because of the need to select two thresholds instead of just one. The method of computing optimal thresholds [13] is also algorithmically and computationally demanding.

The other two rules, non-negative garrotte and hyperbolic thresholding, try to cope with this disadvantage. They imitate the shape of the semisoft thresholding function

by other functions, dependent on just a single threshold. The statistical properties of the semisoft thresholding are asymptotically preserved. Gao [12] demonstrated that the non-negative garrotte has advantages over both hard and soft thresholding rules in terms of risk and sensitivity to small perturbations in the data.

## TESTING THE THRESHOLDING METHODS

A Czech speech signal was used for testing and comparing the successfulness of the wavelet-type denoising methods. Gaussian noise with different variances was added to the signal in order to achieve input testing signals with SNR 15 dB, 10 dB, 5 dB, 0 dB and  $-5$  dB. After the separation the output signal-to-noise ratio  $\widehat{\text{SNR}}$  was estimated.

In all the cases the wavelet Daubechies 8 and the depth of decomposition (and thresholding)  $d = 5$  were used. A small toolbox, called ThreshLab, to be used with MATLAB, and created as a part of the thesis, was used for the testing procedure. All five thresholding rules were applied with the same threshold value to the signal, in two “modes”:

- [WP] thresholding the detail coefficients in classical wavelet transform,
- [WP] thresholding all the coefficients representing the best basis in the wavelet packet of depth  $d$ . The best basis had been found using the so-called threshold entropy with threshold  $\lambda$ .

Apart from the objective parameter  $\widehat{\text{SNR}}$ , also the subjective quality of separation was judged. The following two criteria were equally important in this process: *pleasantness of the listening* (amount of disturbing artifacts) and *understandability* of the denoised recording.

We can summarize the results of the testing as follows:

Only small differences between the WT and WP methods were registered in terms of  $\widehat{\text{SNR}}$ , but from the subjective point of view the difference is significant. The prospective disturbing artifacts in WT are mostly of crackling character, either isolated crackles or a sequence of them, while in the WP method, we recorded kind of the noise being blurred resulting in “gurgitation” in the background, which is more pleasant for listening. We also gain a higher degree of understandability in the WP method.

The soft thresholding rule was found to be the worst one for speech signal processing. This rule suppresses noise better than the other do, which confirmed the conclusion of the theoretical part, but, together with the noise the unvoiced consonants like  $s$ ,  $\check{s}$ ,  $c$ ,  $\check{c}$ ,  $z$ ,  $\check{z}$  also vanish, which are most important for speech understandability.

From the subjective point of view, the best results had been achieved by the hyperbolic thresholding, even though with input SNR decreasing the hard thresholding is beginning to compete with it.

It has emerged that the choice of the threshold value is always a matter of compromise between the amount of noise suppressed and the preservation of the natural character of speech. Such a problem is well-known in all one-channel separation methods. The output  $\widehat{\text{SNR}}$  can be artificially improved, but the subjective evaluation drops quickly.



## NEWLY DEVELOPED METHOD OF SEGMENTED WAVELET TRANSFORM

It is often necessary to process the signal in “real time” in practical applications. A signal, which is not known beforehand, comes to the input of a system segment-by-segment, which have to be effectively processed and sent to the output. A typical example would be the processing of acoustic signals, in particular the speech signals.

The new method of segmented wavelet transform (SegWT), allowing this type of signal processing, can be potentially utilized also in case we need to process a very long signal (not necessarily in real time), but there is insufficient computational power or memory capacity for it. Then it is possible to process the signal part-by-part by the new method.

The goal of the SegWT algorithm is to process the signal segmentwise so that after subsequent processing of all of them we have the same result (the same wavelet coefficients) as if we knew the whole signal and processed it with the common wavelet transform algorithm, DTWT.

For the purposes of the new method it is first necessary to modify the incoming segments. Specifically, every segment must be extended in an appropriate way. The subsequent segments must overlap optimally. We have to cope with the problematic processing of the border segments (first and last) as well. It can be shown that the length of the overlapping parts must comply with strict rules. Yet, generally these lengths can differ from segment to segment. After the segments have been overlapped optimally, we compute a modified version of their wavelet transform. The modification consists in omitting certain steps in the DWTW algorithm. Finally, we join the respective sets of wavelet coefficients in a simple way and we have the result.

The SegWT algorithm is generally delayed one segment length plus the time necessary for the computation. In special case, when the segments' length  $s$  is divisible by  $2^d$ , the delay is equal to just the *computation duration*!

## CONCLUSION

The dissertation thesis consists of two main parts. In the first one, a method for wavelet denoising of signals and its several modifications have been presented. The currently used thresholding techniques have been given, including their statistical properties, and these properties have been compared. A comparison of two principal types of thresholding, i.e. hard and soft thresholding, has led to the conclusion that hard thresholding yields poorer denoising but better maintains sudden changes in the signal (jumps and peaks). Soft thresholding is very successful at denoising but has a tendency to “oversmooth” jumps and peaks. Semisoft thresholding is more than a compromise — with a correct choice of two parameters it has the least risk of “error”. However, the two parameters of this method constitute a great disadvantage compared to other methods that work with only one parameter. Other methods given in the text try to reduce this disqualification. They work with only one parameter but in the statistical



sense they do not achieve such a quality as semisoft thresholding does.

On the basis of the testing performed on speech signals, wavelet denoising has been shown to achieve good results when the parameters are set correctly. From the subjective point of view, the method utilizing wavelet packets and the best basis outdid the method utilizing the classical wavelet transform. However, the first mentioned method is much more demanding computationally. For the purpose of denoising signals with a higher input SNR the hyperbolic thresholding rule turned out to be the optimal choice. For the signals with input SNR close to 0 dB, it is better to use the hard thresholding.

The second part of the thesis is devoted to the newly-developed method of so-called segmented wavelet transform, which allows us to perform the wavelet transform, and thus also the “wavelet-type separation” in real time. It has been shown that if the incoming signal segments are overlapped optimally and consecutively subjected to a modified version of the wavelet transform, the same result is achieved as if we processed the signal as a whole.