

# Wavelet transformace v metodách zvýrazňování řeči

Petr Opršal<sup>1</sup>

<sup>1</sup>Katedra elektrických měření, FEI, VŠB – Technická Univerzita Ostrava, 17. listopadu 15,  
708 33, Ostrava-Poruba  
oprshal@tiscali.cz

**Abstrakt.** Wavelet transformace je moderní nástroj analýzy signálů. Nachází po boku Fourierovy transformace stále více uplatnění v metodách analýzy a zpracování nestacionárních signálů. Tento příspěvek se věnuje popisu použití Wavelet transformace pro redukci šumu v zašuměném signálu. Dále shrnuje potřebné znalosti pro praktickou aplikaci této transformace v systémech zvýrazňování řeči.

**Klíčová slova:** Zvýrazňování řeči, wavelet transformace

## 1 Úvod

Mluvená řeč je základním prostředkem komunikace člověka s okolím. S otázkou efektivního kódování řeči v hlučném prostředí a také spolehlivého rozpoznání řečového signálu, vznikla potřeba zvýraznění řeči s potlačením hluků a šumů pozadí.

Moderním prostředkem pro analýzu řečového signálu je wavelet transformace (dále WT). Při aplikaci krátkodobé Fourierovy transformace (dále STFT) na signál je možné frekvenci v čase stanovit pouze s přesností danou délkou časového okna. WT však poskytuje přesnou časovou informaci o změnách a nespojitostech signálu v různých rozlišovacích úrovních v závislosti na frekvenci. Pro analýzu signálů používá jiné než harmonické funkce.

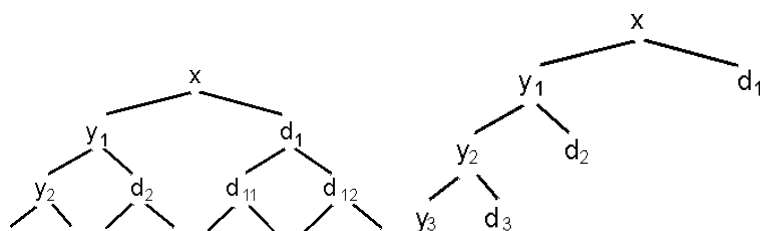
Při praktické aplikaci WT na řečový signál, se narazí na několik problémů, spojených s přípravou signálu a s vyhodnocením výsledku.

## 2 Diskrétní wavelet transformace

Pro analýzu s použitím výpočetní techniky je výhodné použít zjednodušené formy WT, označované jako diskrétní wavelet transformace. Používá se Mallatův pyramidový algoritmus výpočtu pomocí multirozkladu nebo paketového rozkladu [11], obr. 1., a dyadické dělení s ortonormální wavelet bází. Při výpočtu transformace je posouván vektor vstupního signálu oproti jádru transformace o dva prvky. Ve výsledku dochází k dvojnásobné kompresi a zároveň k dělení měřítka dvěma. Tento princip je známý také jako „rychlá“ WT. Podmínkou je, aby délka vstupního vektoru byla  $N = 2^x$  prvků, kde  $x$  je celé číslo. Pro každou hladinu  $m$  wavelet transformace se provede rozklad podle rovnice (1).

$$\mathbf{Y}_{D_{m+1}} = \begin{pmatrix} \mathbf{y}_{m+1} \\ \mathbf{d}_{m+1} \end{pmatrix} = \mathbf{M} \mathbf{y}_m \quad (1)$$

Kde  $\mathbf{Y}_{D_{m+1}}$  je výledný sloupcový vektor WT (m+1) –té hladiny rozkladu, obsahující komprimovaný signál  $\mathbf{y}_{m+1}$  a ortogonální doplněk  $\mathbf{d}_{m+1}$ .  $\mathbf{y}_{m+1}$  je sloupcový vektor vstupního signálu (m+1) – té hladiny komprimace. Vektor  $\mathbf{y}_m$  pro m=0 představuje vektor vstupního signálu  $\mathbf{x}\{n\}^T$ .  $\mathbf{d}_{m+1}$  je sloupcový vektor ortogonálních doplňků, který mi vyjadřuje jemnosti na m+1 hladině.  $\mathbf{M}$  je čtvercová matice vytvořená z scale a wavelet filtračních koeficientů o rozměrech  $N \times N$ . Matice je normovaná pro zjednodušení výpočtu inverzní transformace. m=0,1,2,... vyjadřuje hladinu rozkladu.



**Obr. 1.** Tvorba pyramidální struktury wavelet transformace, a rozdíl mezi a) paketovým rozkladem a b) multirozkladem.

Rozklad podle vzorce (1) se provádí postupně na jednotlivých hladinách v pyramidové struktuře. Počet hladin je určen podle požadovaného rozlišení frekvence až do maximální hladiny, vypočtené podle Shannonovy míry neurčitosti [3,11]. S doplňky je možné pracovat dále jako se signálem. Tím vytvořím rovnoměrnou síť z hlediska frekvence. Rozklad lze tedy chápat jako aplikaci filtrů.

## 2.1 Vytvoření báze WT

Matice báze wavelet transformace  $\mathbf{M}$  je ortogonální a je tvořena maticí scale filtračních koeficientů  $\mathbf{H}$  a maticí wavelet filtračních koeficientů  $\mathbf{G}$ . Má několik nenulových prvků s následující pravidelnou strukturou (2), zde uvedenou pro  $N=4$  filtračních koeficientů. Naprosto identicky se tvoří pro jiný počet koeficientů  $N$ . Řád matice je totožný s počtem prvků vektoru vstupního signálu  $\mathbf{y}_m$ . Z vyobrazení matice  $\mathbf{M}$  je zřejmé, že počítačová realizace FWT je nenáročná na využití procesoru, i na obsazení paměti.

$$\mathbf{M} = \begin{pmatrix} \mathbf{H} \\ \mathbf{G} \end{pmatrix} = \begin{pmatrix} h_0 & h_1 & h_2 & h_3 & \dots & 0 & 0 \\ 0 & 0 & h_0 & h_1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ h_2 & h_3 & 0 & 0 & \dots & h_0 & h_1 \\ g_0 & g_1 & g_2 & g_3 & \dots & 0 & 0 \\ 0 & 0 & g_0 & g_1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_2 & g_3 & 0 & 0 & \dots & g_0 & g_1 \end{pmatrix} \quad (2)$$

Samotná tvorba diskretní ortogonální báze pro FWT není jednoduchá. Základem je definice a získání scale (měřítkové) funkce  $\varphi(t)$  [3,6,11]. Scale filtrační koeficienty se pak dostanou multirozkladem z rovnice (3) na konkrétní hladině  $m$ .

$$\varphi_{m,n}(t) = \frac{1}{\sqrt{2^m}} \varphi(2^m t - n) \quad h_0 = \varphi_{m,0}, h_n = \varphi_{m,n} \quad (3)$$

Scale filtrační koeficienty tedy představují funkci měřítka. Pomocí nich se tvoří komprimovaný signál a určují celkový trend vstupních dat.

V matici  $H$  se s výhodou používají ortonormální scale koeficienty I. Daubechies s  $2N$  nenulovými prvky, označované db $2N$ . Jsou publikované pro 2,4, ... až 20 nenulových hodnot. Symbolem db1 se označuje Haarova báze, která má 2 koeficienty, db2 značí Daubechies bázi se 4 koeficienty, atd. Daubechies wavelety nejsou symetrické [6].

Pomocí wavelet filtračních koeficientů  $g_k$  (nejsou totožné s wavelet funkcí), se tvoří ortogonální doplněk ke komprimovanému signálu, který obsahuje jemnosti na jednotlivých hladinách. Wavelet filtrační koeficienty se tvoří lineární kombinací scale filtračních koeficientů. Například  $g_k = (-1)^k h_{N-1-k}$ .

## 2.2 Zpětná wavelet transformace

Rekonstrukci signálu z doplňků lze provést zpětnou wavelet transformací (IWT), rovnice (4), nebo také interpolací se zanedbáním některých doplňků.

$$\mathbf{y}_{m-1} = \mathbf{M}^T(\mathbf{y}_m, \mathbf{d}_m) \quad (4)$$

Výpočet musí být vykonán ve všech hladinách v pořadí opačném dle platné pyramidální struktury, obr.1.

## 2.3 Úpravy doplňků

Doplňky, vzniklé po aplikaci WT na konkrétní hladině, nabývají kladných i záporných hodnot. Jejich vhodnou modifikací se získá na výstupu IWT, obr. 2., zvýrazněný řečový signál s redukováným šumem.

Pro jednotlivé koeficienty doplňku platí rovnice  $y = s + n$ . Kde  $y$  vyjadřuje zašuměný signál,  $s$  čistý signál a  $n$  šum [1,9,10]. Koeficienty se nejčastěji upravují prahováním podle rovnice (5).

$$y_{\text{prah}} = \begin{cases} z & , |y| > p \\ 0 & , |y| < p \end{cases} \quad (5)$$

Kde  $y_{\text{prah}}$  vyjadřuje koeficient doplňku zašuměného signálu po aplikaci prahování. Při vhodně zvoleném prahu  $p$  může být tato hodnota rovna odhadu čistého signálu

$y_{\text{prah}} = s_{\text{odh}}$ . Pro tvrdé prahování se používá  $z = y$ . Pro měkké prahování se stanovuje například  $z = \text{sign}(y)(|y| - p)$ . Samotná hodnota prahu  $p$  může být navržena rovnicí (6).

$$p = \sigma \sqrt{2 \log(N)} \quad (6)$$

Kde  $N$  je počet vzorků doplňku  $y_m$ .

Prahování se aplikuje pro jednotlivé doplňky s různými  $\sigma$ , podle typu nežádoucího šumu v signálu. U bílého šumu je pro všechny doplňky konstantní. Pro nepravidelně se měnící úroveň šumu se musí použít některý z adaptabilních algoritmů.

### 3 Aplikace WT pro metody zvýrazňování řeči

WT se v zpracování signálů často používá jako prostředek k filtraci nebo kompresi. Pro potřebu zvýraznění řeči se WT aplikuje pro rozklad do konkrétní hladiny, vhodné pro úpravy doplňků. Ty po výpočtu IWT poskytnou zvýrazněný řečový signál. Na obr. 2. je zobrazeno blokové schéma systému.



**Obr. 2.** Blokové schéma postupu výpočtů s použitím WT pro metody zvýraznění řeči.

Řečový signál je často snímán jedním mikrofonom. Při snímání dvěma a více mikrofony mohou využít vzájemných vazeb a prostorové informace [7].

Pro definování délky vstupního vektoru  $y_0$  je potřeba kvalitních detektorů řečové aktivity k určení začátku a konce promluvy. Ty slouží také k definování řečové pauzy pro odečtení hladiny samotného šumu. Problematika detektorů řečové aktivity je velmi rozsáhlá [např.13], mnohdy na jeho kvalitě závisí úspěch konkrétní metody.

U konkrétní aplikace WT dále vyvstanou tyto otázky:

#### 3.1 Promluva v reálném čase

Při aplikaci WT na krátký řečový povel, se na vstupu systému, obr. 2, počítá pouze s jediným vektorem  $y_0$ . Při delší promluvě v reálném čase je potřeba signál rozdělit do krátkých časových úseků  $y_{0,t}$ . Je výhodné použít metodu sčítání přesahů, která byla navržena pro rozdělení signálu na okna konstantní délky pro aplikaci STFT. Po úpravě spektra a zpětné transformaci se signál zpětně rekonstruuje [12]. Při použití této metody pro WT lze úseky řeči rozčlenit do časových rámců v trvání jednotek vteřin, buď konstantních, nebo v závislosti na řečových pauzách.

### 3.2 Metody zvýrazňování řeči

Jedna z nejstarších a nejjednodušších metod zvýrazňování řeči a redukce šumů se nazývá spektrální odečítání [např.12]. Z hodnoty spektrální hustoty zašuměného signálu je odečten odhad spektra šumu. Je vhodná pro stacionární šумы nekorelované s řečovým signálem, kdy se odhad spektrální hustoty šumu získá v řečových pauzách. Tato metoda upravená pro WT využívá měkkého prahování. WT v porovnání s STFT dosahuje v této metodě nižší úroveň residuálních šumů.

Šum v řečovém signálu často mění své parametry. Proto je potřebné měřit šum nejenom v době řečové pauzy. Jedna z metod je založena na pořizování statistiky výkonu šumového pozadí a jejím odečítání od zašuměného řečového signálu [5].

Pro adaptivní filtraci šumu během řeči se používají dvě skupiny algoritmů. První je založena na teorii Wienerovy fitrace [2] a druhá na teorii Kalmanovy fitrace. Jejich společnou vlastností jsou průběžné změny parametrů systému podle změn vstupu i výstupu [4].

Všeobecně nevýhodou všech těchto metod na bázi STFT je vysoká hladina residuálních šumů ve výsledku. Aplikací WT se míra těchto šumů snižuje.

### 3.3 Hodnotící kritéria zvýraznění řeči

Úspěšnost zvýraznění řečového signálu se popisuje hodnotícími kritérii. Standardní objektivní kritérium pro vyjádření úrovně šumu v signálu je SNR, které ovšem není optimální pro nestacionární řečový signál [8]. Důležitým objektivním kritériem systému, obr.4, je zlepšení SNR, rovnice (7).

$$\text{SNRE} = \text{SNR}_{\text{out}} - \text{SNR}_{\text{in}} = 10 \log \frac{P_Y - P_S}{P_{\hat{Y}} - P_S} \quad (7)$$

Kde  $P_Y$  vyjadřuje výkon vstupního signálu,  $P_{\hat{Y}}$  výkon výstupního signálu, za předpokladu, že výkon čisté řeči  $P_S$  není průchodem systémem zkreslen.

Za subjektivní hodnocení se považují poslechové testy skupiny posluchačů. Zde universální kritéria neexistují. Řeč je možno posuzovat z několika pohledů (např. srozumitelnost, přirozenost, ..).

## 4 Závěry

Prvním úkolem článku byla praktická realizace diskrétní WT pro potřeby zvýraznění řeči. Druhým úkolem bylo podání stručného přehledu o problematice, týkající se aplikace WT do systémů zvýrazňování řeči a kritéria vyhodnocení jejich úspěšnosti.

## Reference

1. Bahoura M., Rouat J. A new approach for wavelet speech enhancement. *Eurospeech 2001*. Aalborg, 2001.
2. Cohen I. Enhancement of speech using Bark-Scaled wavelet packet decomposition. *Eurospeech 2001*. Aalborg, 2001.
3. Častová N. Integrovní transformace. Syllaby pro doktorandské studium, VŠB-TU Ostrava, 2002.
4. Malenovský V. Adaptivní filtrace zašuměných řečových signálů. *Elektrorevue*. 2002. <http://www.elektrorevue.cz/clanky/02063/>.
5. Martin R. Spectral subtraction based on minimum statistics. *EUSPICO-94*. Edinburgh, 1994.
6. Missiti M., Missiti Y., Oppenheim G., Poggi J.M. *Wavelet Toolbox for use with Matlab*. User's Guide, The MathWorks, 1996.
7. Hung P.Q. Sovka P. Vícestupové metody redukce šumů v řeči. *Akustické listy* 8(3)-2002. ČsAS.
8. Pollák P. Metody odhadu odstupu signálu od šumu v řečovém signálu. *Akustické listy* 7(3)-2001. ČsAS.
9. Seok J.W., Bae K.S. Speech Enhancement with reduction of noise components in the wavelet domain. *IEEE Signal processing*. Albuquerque, 1997.
10. Sheikhzadeh H., Abutalebi H.R. An improved wavelet-based speech enhancement system. *Eurospeech 2001*. Aalborg, 2001.
11. Smutný J. Transformace wavelet a její využití při zpracování signálů, *Automatizace* 10/1998. ISSN 0005-125X
12. Sovka P., Pollák P. *Vybrané metody číslicového zpracování signálů*. ČVUT, 2001, Praha. ISBN 80-01-02416-4
13. Sovka P., Pollák P. The study of speech /pause detectors for speech enhancement methods. *Eurospeech 95*, Madrid, 1995

### Annotation:

*Wavelet transformation in speech enhancement methods.*

The wavelet transformation is modern signals analyzing tool. Wavelet transformation finds alongside FFT more and more applications in the nonstationary signals analysis. The paper is devoted to the description of wavelet transformation used for noise reduction in signal contaminated by additive noise. This paper describes the necessary knowledge for practical application wavelet transformation in speech enhancement systems.