

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

---

Fakulta elektrotechnická  
Katedra teorie obvodů

Semestrální práce z předmětu SZS

Jan Ingerle  
xingerle@fel.cvut.cz

---

29. června 2001

## Název:

Ekvalizace akustického kanálu metodou SIMO

## Úvod:

Moderní telekomunikace používají k přenosu řeči telefonním kanálem metody kódování řeči, jejichž chování silně závisí na šumových vlastnostech pozadí. Aby řeč byla přenášena telefonním kanálem spolehlivě, je třeba v co největší míře potlačit šum pozadí a tedy zvýraznit užitečný signál — řeč, aniž by utrpěl ztrátu na kvalitě.

Pro řešení tohoto problému byly a stále jsou vyvíjeny metody, které lze podle základního přístupu dělit na

- jednobandové metody
- vícebandové metody.

Zatímco jednobandové metody zvýrazňování řeči jsou závislé na relativně striktních požadavcích na vlastosti kanálu, zpracovávaného signálu a jeho předzpracování (např. na detektoru řeč/pauza), pracují vícebandové metody na principu prostorové selektivity. Tento přístup umožňuje připustit obecnější vlastnosti signálu a vynechat předzpracování signálu. Důležitou výhodou je fakt, že metody založené na tomto přístupu mohou vykazovat prostorovou selektivitu aniž by docházelo ke zkreslení zpracovávaného signálu, což je jeden z požadavků dalšího úspěšného zpracování. Z těchto důvodů se vícebandové metody dají považovat za perspektivní.

I přes zmíněné volnější požadavky na vlastnosti zpracovávaných signálů u vícebandových metod je třeba definovat charakter šumového pozadí, které chceme těmito metodami potlačovat. V tomto bodě se jednotlivé vícebandové metody dělí podle typu šumu, jenž jsou schopny zpracovávat. Jako vhodné kritérium dělení šumů pro tyto účely se jeví koherenční funkce  $\Gamma(f)$ :

$$\Gamma(f) = \frac{\Phi_{ij}(f)}{\sqrt{\Phi_i(f)\Phi_j(f)}}, \quad (1)$$

kde  $\Phi_i(f)$  je spektrální výkonová hustota v  $i$ -tém kanálu a  $\Phi_{ij}(f)$  je vzájemná spektrální výkonová hustota kanálů  $i$  a  $j$ . Koherenční funkce vyjadřuje vzájemnou „podobnost“ složek signálu na jednotlivých frekvencích. Z tohoto hlediska se dá šum rozdělit do tří základních skupin: na šum koherentní ( $\Gamma(f) = 1$ ), nekoherentní ( $\Gamma(f) = 0$ ) a difuzní ( $\Gamma(f) \approx \frac{\sin(f)}{f}$ ).

Jedno z nejběžnějších prostředí, které přichází v úvahu pro aplikaci zvýrazňování řeči v šumu, jsou kanceláře. Tyto prostory jsou tvořeny místnostmi s dozvukem, kde dochází k odrazům a vícecestnému šíření signálu (řeči).

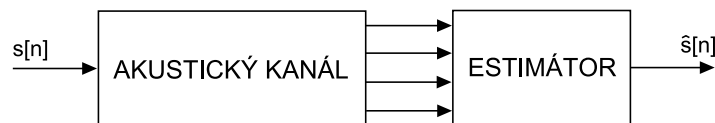
Na senzor (mikrofon) tak nepřichází pouze signál šířící se přímo od zdroje (mluvčího), ale i signál zpožděný a utlumený, šířící se odrazem od odrazových ploch (stěny, strop, vybavení). Tuto skutečnost lze vyjádřit zápisem:

$$x(t) = \sum_n \alpha_n(t) e^{-j\Theta_n(t)}, \quad (2)$$

kde  $x(t)$  je signál na mikrofonu a  $\alpha_n(t)$  a  $\Theta_n(t)$  je amplituda a fázový posun  $n$ -tého příspěvku. Díky akustickým vlastnostem reálných odrazových ploch v kancelářích, ovlivňující parametry  $\alpha_n(t)$  a  $\Theta_n(t)$ , dochází ke vzniku šumu difusního charakteru. Znamená to, že na nižších frekvencích je signál na mikrofonu silně korelován se signálem u zdroje a se zvyšující se frekvencí tato korelace klesá. Při řešení problému zvýrazňování řeči v těchto prostředích je třeba s tímto faktem počítat.

### Definice problému:

Řešení problému tak jak byl nastíněn výše, lze definovat jako návrh vícecestupového estimátoru realizujícího odhad vstupního signálu  $\hat{s}[n]$  na výstupu diskrétní akustického soustavy — kanálu, modelujícího prostředí s dozvukem. Vstupním signálem  $s[n]$  je řeč. Problém je naznačen na obrázku 1.



Obrázek 1: Blokové schéma systému

### Řešení:

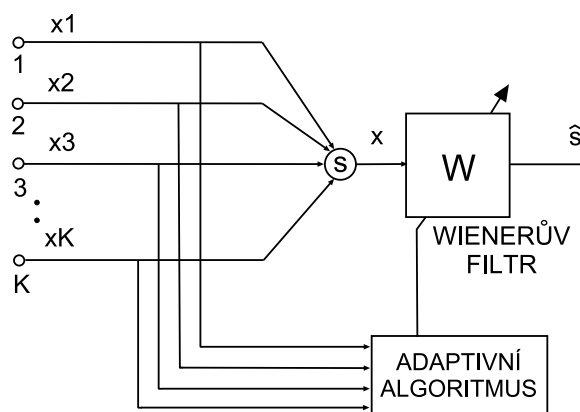
Problému, tak jak byl popsán, lze řešit několika způsoby. Podle toho, jak jednotlivé způsoby získávají charakteristiky vstupního signálu k následnému odhadu vlastností kanálu je lze rozdělit do následujících skupin:

- metody s referenční vstupní posloupností
- metody bez referenční vstupní posloupností (tzv. slepá ekvalizace)
- metody smíšené.

Metody s referenční posloupností (viz např. [2]) či smíšené (viz např. [3]) získávají informace o kanálu většinou ze známé — referenční posloupnosti, kterou vysílají do kanálu. Ze znalosti vstupní a naměřené posloupnosti pak odhadují vlastnosti příslušného kanálu.

Metody bez referenční posloupnosti většinou odhadují vlastnosti vstupního signálu pomocí vícekanálového zpracování výstupního signálu. V teorii ekvalizace pro telekomunikace byla těchto metod popsána řada (viz např. [7] nebo [9]). Většina z nich však vychází z určitých předpokladů na statistické charakteristiky vstupního signálu daných specifiky aplikace přenosu informace telekomunikačním kanálem. V této aplikaci dochází ke kódování informací do posloupností mající vlastnosti zjednodušující přenos. Při řešení problému, kde je vstupním signálem řeč však tyto algoritmy selhávají, neboť řeč je nestacionární náhodný signál jehož statistické charakteristiky se určují velice nesnadno. Existují však přístupy pomocí kterých lze danou problematiku, alespoň částečně, řešit.

Jeden ze přístupů, nazývaný „beamformer s adaptivní postfiltrací“, je na obrázku 2 (viz [10]).



Obrázek 2: Beamformer s adaptivní postfiltrací

Aby bylo možno použít v systém pro estimaci signálu Wienerův filtrem, je třeba zaručit stacionaritu signálů. Řeč však stacionární není. Běžným prostředkem pro řešení tohoto problému je segmentace řečových signálů (viz např. [8]): Rozdělíme-li podle určitých pravidel řečový signál na úseky dlouhé 10–20 ms, lze signál v těchto úsecích považovat za stacionární a jako se stacionárním s ním počítat. Rekonstrukcí provedenou podle daných pravidel pak získáme zpracovaný signál o původní délce.

Druhou podmínkou správné funkce systému z obrázku 2 je vzájemná nekorelovanost vzorků šumu na mikrofonech. Avšak, jak již bylo řečeno, v aplikaci zvýrazňování řeči v místnostech s dozvukem je charakter šumu difusní, což znamená na nízkých frekvencích korelovaný. Následující úvahou lze tento problém částečně řešit:

V [1] je odvozen vztah pro vzájemnou výkonovou hustotu signálů na dvou mikrofonech:

$$\Phi_{ij}(d, \lambda) = \Phi_{nn} \frac{\sin(2\pi \frac{d}{\lambda})}{2\pi \frac{d}{\lambda}}, \quad (3)$$

kde  $d$  je vzdálenost mikrofónů,  $\lambda$  vlnová délka a  $\Phi(d, \lambda)$  je výkonová spektrální hustota šumu. Vzhledem k (1),  $\Phi_{ii} = \Phi_{jj} = \Phi_{nn}$  a  $\lambda = \frac{c}{f}$  platí, že

$$\Gamma_{ij}(f) = \frac{\sin^2(2\frac{d}{c}f)}{2\frac{d}{c}f}. \quad (4)$$

Tato rovnice udává vztah mezi koherenční funkcí a vzdáleností mikrofónů. První nula koherenční funkce je na frekvenci  $f_0 = \frac{c}{2d}$ , pokles o 3dB na frekvenci  $f_{3dB} = \frac{c}{d}$ .

Z předchozího vyplývá, že pokud chceme v popsáném akustickém systému dosáhnout na jednotlivých mikrofonech vzájemně korelovaných vzorků šumu, musíme volit vzdálenost mezi mikrofony  $d < \frac{c}{2f_{min}}$ , kde  $f_{min}$  je předpokládaná minimální frekvence rušení.

Pokud chceme dosáhnout vzájemně nekorelovaných vzorků, je nezbytné volit  $d > \frac{c}{2f_{min}}$ . Horní hranice  $d$  pak vychází s kvaziperiodicity řeči: maximální zpoždění mezi mikrofony může být nejvýše polovinu základní periody zpracovávané řeči. Tedy:

$$d < \frac{\lambda_{pitch}}{2}. \quad (5)$$

Aby byly splněny obě předcházející podmínky, musí platit:

$$d = \frac{c_{sound}}{2f_{min}} = \frac{c_{sound}}{2f_{pitch}}. \quad (6)$$

Vzdálenost mezi mikrofony tedy ovlivňuje vzájemné korelace mezi jednotlivými vzorky šumu na mikrofonech. Pokud při návrhu systému dodržíme podmínku (6), můžeme považovat vzorky za vzájemně nekorelované.

Jsou-li splněny výše uvedené požadavky, můžeme přistoupit k návrhu Wienerova filtru. Při jeho odvození uvažujeme systém diskrétní v čase. Pro uvažovaný problém lze signál na vstupech  $x_i$  zapsat jako:

$$x_i[m] = s[m - \tau_i] + n_i[m], \quad (7)$$

kde  $x_i[m]$  je signál na  $i$ -tém mikrofónu,  $s[m]$  je signál zdroje, tedy odhadovaná řeč,  $\tau_i$  je zpoždění signálu na  $i$ -tém mikrofónu a  $n_i[m]$  je šum na příslušném mikrofónu. Dále je možné napsat, že

$$x[m] = \sum_{i=1}^K x_i[m]. \quad (8)$$

Filtrační rovnice pro Wienerův filtr s konečnou impulsovou odezvou je:

$$\hat{s}[m] = \sum_{j=0}^M w[j]x[m-j], \quad (9)$$

kde  $w$  je vektor koeficientů filtru,  $M$  je řád filtru,  $x[m]$  vstupní a  $\hat{s}[m]$  výstupní signál filtru.

Za chybu odhadu označíme rozdíl odhadovaného signálu a výstupu filtru

$$e[m] = s[m] - \hat{s}[m] = s[m] - \sum_{j=0}^M w[j]x[m-j] \quad (10)$$

a jako kritérium odhadu zvolíme střední kvadratickou hodnotu této chyby:

$$\mu = E[e^2[m]] = E[(s[m] - \hat{s}[m])^2], \quad (11)$$

gradient této funkce je

$$\nabla\mu = \frac{\partial E[e^2[m]]}{\partial w[l]} = 2E \left[ e[m] \frac{\partial e[m]}{\partial w[l]} \right] = E[e[m]x[m-l]], \quad (12)$$

kde  $k = 0, 1, \dots, M$ . Minimum kritéria získáme anulováním gradientu, čímž získáme normální rovnice:

$$E \left[ \left( s[m] - \sum_{j=0}^M w[j]x[m-j] \right) x[m-l] \right] = 0, \quad (13)$$

což lze zapsat, s využitím linearit operatoru střední hodnoty, jako:

$$\sum_{j=0}^M w[j]E[x[m-j]x[m-l]] = E[s[m]x[m-l]], \quad k = 0, 1, \dots, M. \quad (14)$$

Protože výraz  $E[x[m]x[m-l]]$  je autokorelační funkce  $R_{xx}(k)$  jedné realizace náhodného procesu, lze rovnici (14) přepsat na tzv. Wiener-Hopfovou rovnici:

$$\mathbf{R}_{xx}\mathbf{w} = \mathbf{R}_{sx}, \quad (15)$$

kde  $\mathbf{R}_{xx}$  je matice autokorelačních koeficientů signálu  $x$ ,  $\mathbf{w}$  je vektor hledaných koeficientů filtru a  $\mathbf{R}_{sx}$  je vektor korelačních koeficientů  $s$  a  $x$ .

Pro řešení této rovnice je nutné hodnoty korelačních koeficientů nahradit jejich odhady ze signálů na mikrofonech. Hodnotu  $R_{xx}[l] = E[x_i[m]x_i[m+l]]$  lze odhadnout přímo ze signálů na jednotlivých mikrofonech. V případě, že

je splněna podmínka nekorelovanosti vzorků šumu a nezávislosti šumu na signálu (podmínka (6)), lze psát:

$$E[x_i[m]x_j[m+l]] = E[(s[m] + n_i[m])(s[m+l] + n_j[m+l])] = R_{ss}[l], \quad (16)$$

$$R_{sx} = E[s[m]x_i[m+l]] = E[s[m](s[m+l] + n_i[m+l])] = R_{ss}[l], \quad (17)$$

$R_{sx}$  lze pak odhadnout ze vzájemné korelace signálů na různých mikrofonech. Vlastní odhad můžeme provést pomocí Wiener-Chinčiny věty ze spektrálních výkonových hustot signálů a frekvenční charakteristiku Wienerova filtru pak určit z těchto hodnot následovně:

$$A[k] = \frac{1}{M} \sum_{i=1}^M |X_i[k]|^2, \quad (18)$$

$$C[k] = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M X_i[k]X_j^*[k], \quad (19)$$

$$W[k] = \frac{\frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M X_i[k]X_j^*[k]}{\frac{1}{M} \sum_{i=1}^M |X_i[k]|^2}, \quad (20)$$

kde  $A[k]$  je odhad spektrální výkonové hustoty signálu  $x[m]$ ,  $C[k]$  je odhad vzájemné spektrální výkonové hustoty mezi signály jednotlivých kanálů,  $X_i$  je Fourierův obraz signálu  $x_i$ ,  $M$  je počet kanálů systému.

Výsledný systém podle obrázku 2 tedy odhaduje vstupní signál potlačováním šumu nezávislého na užitečném signálu s nekorelovanými vzorky tak, že vstupní vzorky jednotlivých kanálů sečte a získaný signál přefiltruje filtrem, jehož váhy jsou pro každý segment nastaveny podle předpisu (20).

### Simulace:

Pro simulaci systému bylo použito prostředí MATLAB. Simulace byla provedena na směsi bílého šumu a reálné čisté řeči vzorkované frekvencí  $f_s = 8000Hz$  a na řeči snímané v odrazové místnosti vzorkované frekvencí  $f_s = 22000Hz$ . Počet mikrofonů byl  $M = 4cm$  a jejich vzdálenost  $d_{mic} = 30cm$ . Při segmentaci byla použita délka segmentů 512 vzorků. Počet vah Wienerova filtru  $M = 41$ .

Průběhy signálů první simulace (vstupní signál je směs řeči a bílého šumu) je na obrázku 3. V části a) je zachycen součet vstupních signálů  $x[m]$  tak, jak je ho možné pozorovat za sumátorem. Část b) znázorňuje výstupní

signál systému  $\hat{s}[m]$ . V poslední části obrázku je zachycen časový vývoj vah Wienerova filtru  $w[n]$ . Z obrázků je vidět, že na tomto signálu systém pracuje podle našich představ — šum obsažený v řeči je částečně eliminován.

Průběh druhé simulace (vstupní signál je řeč snímaná v místnosti s dozvukem) je na obrázku 4. Znázorněné signály jsou stejné jako v předchozím případě: vstupní signál  $x[m]$ ,  $\hat{s}[m]$  a  $w[n]$ . Na průběhu této simulace je vidět, že pro reálnou situaci systém nepracuje zcela správně a že kromě částečného potlačení šumu, nižšího než v předchozím případě, dochází i k poškození užitečného signálu.

### Závěr:

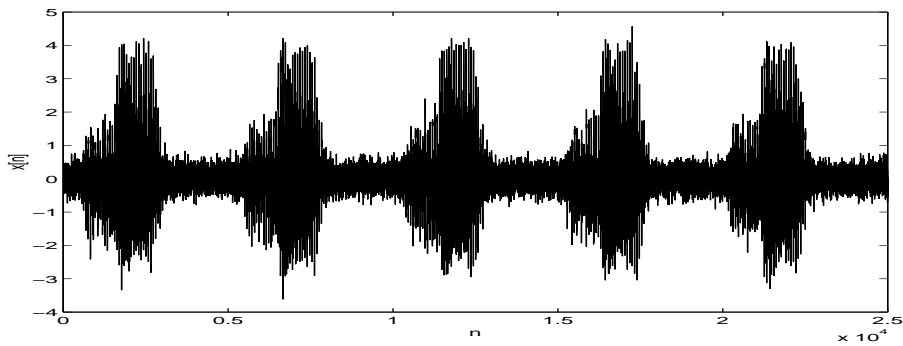
V práci byl odvozen algoritmus pro odhad signálu na vstupu akustického kanálu pomocí vícekanálové metody (metody SIMO). Z tohoto odhadu lze dalším výpočtem jednoduše odhadnout vlastnosti akustického kanálu, což však v dané aplikaci není podstatné. Dále byla odvozena podmínka pro vzdálenost mezi mikrofony tak, aby byla zachována relativní nezávislost šumu a užitečného signálu a nekorelovanost šumu. Odvozená podmínka je pouze kompromis mezi zmíněnými požadavky a požadavky pro zachování vlastností systému umožňující další zpracování signálu. Z tohoto důvodu — nedodržení předpokladů pro odvození filtru, nelze předpokládat absolutní funkčnost algoritmu. Zmíněná fakta dokládají simulace, ze kterých je zřejmé, že pro reálný signál dochází k poškození užitečného signálu. Tyto vlastnosti, omezující použití systému, skýtají široké pole pro další vývoj, který je směřován ke konstrukci co možná nejuniverzálnějšího systému.

### Literatura

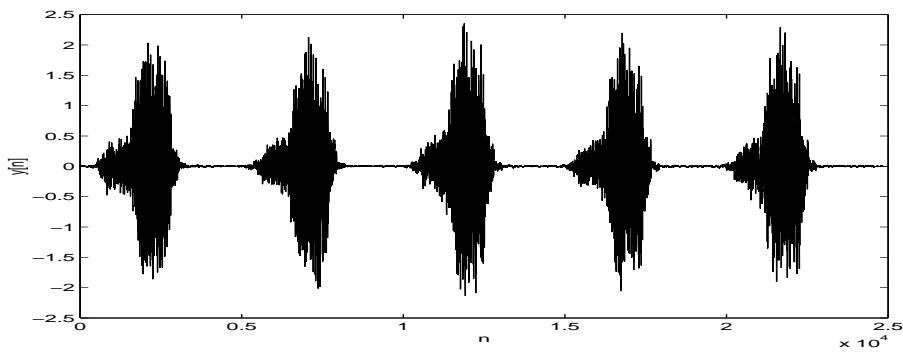
- [1] Cron, B.F., Sherman, C.H.: *Spatial-correlation function for various noise models*, Journal of Acoustic Soc. of America, vol. 34, no. 11, 1962
- [2] Davis, L.M., Collings, I.B., Evans, R.: *Identification of Time-varying Linear Channels*, IEEE Transactions, pp.3921, 1997
- [3] Gorokhov, A., Loubaton, Ph. : *Semi-Blind Seccon Order Identification of Convolutive Cannels*, IEEE Transactions, pp.3905, 1997
- [4] Hrdina, Z.: *Statistická radiotechnika*, Vydavatelství ČVUT, 1996
- [5] Kay, M.K.: *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall Int., London, 1993
- [6] Marro, C., Mahieux, Y., Simmer, K.U.: *Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfiltering*, IEEE Trans. on Speech and Audio Process., vol.6. no.3, 1998
- [7] Proakis, J.G.: *Digital Communications*, McGraw-Hill, New York, 1995



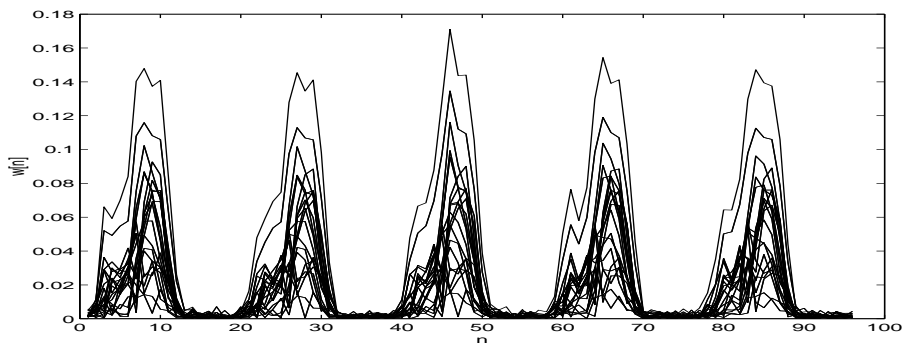
- [8] Uhlř, J., Sovka, P.: *Číslicové zpracování signálu*, Vyd. ČVUT, Praha, 1996
- [9] Yang, H.H.: *On-line Blind Equalization via On-line Blind Separation*, Signal Processing, vol. 68, 1998
- [10] Zelinski, R.: *A Micorphone Array with Adaptive Post-Filtering for Noise Reduction in Reverberant Rooms*, ICASSP'88 Proceedings, New York, 1988
- [11] *Summary of Activities during the 1st year of the SWAN project*,  
<http://www.mclink.it/com/swan/addtweb10.htm>



a) Součet vstupních signálů  $x_i$

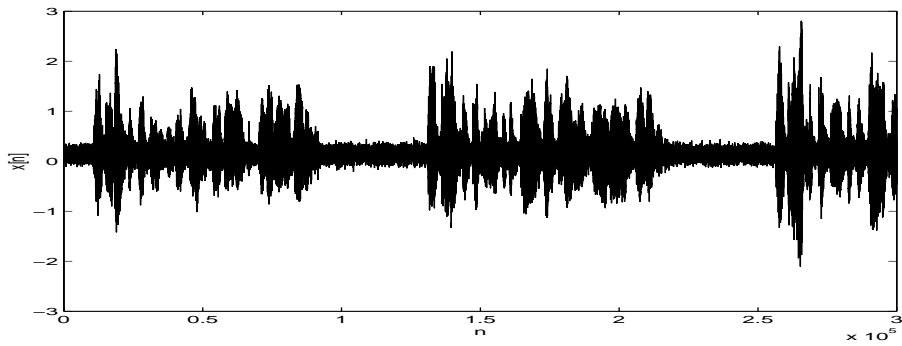


b) Výstupní signál  $\hat{s}$

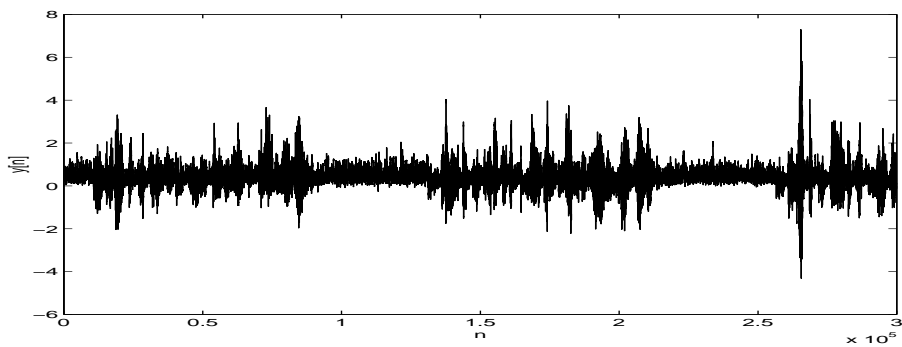


c) Časový průběh vah filtru

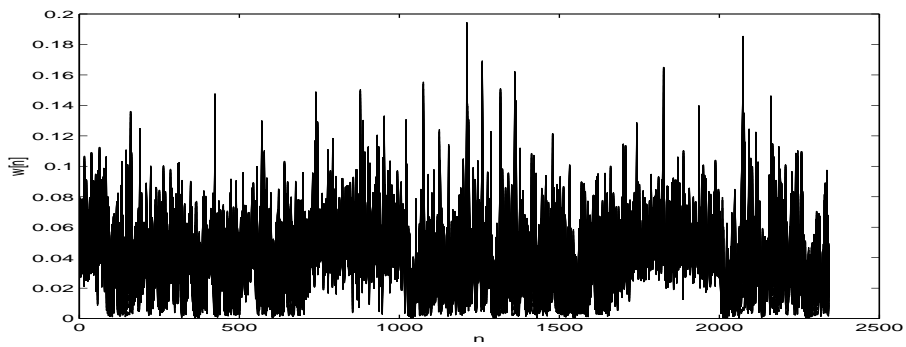
Obrázek 3: Průběhy signálů simulace s bílým šumem



a) Součet vstupních signálů  $x_i$



b) Výstupní signál  $\hat{s}$



c) Časový průběh vah filtru

Obrázek 4: Průběhy signálů simulace s reálným signálem