

Zvýrazňování řeči pomocí vícekanálového zpracování

Václav Bolom, Pavel Sovka

Katedra teorie obvodů
Fakulta elektrotechnická
České vysoké učení technické v Praze
Technická 2, 166 27 Praha 6

Abstrakt

Úlohu zvýrazňování řeči přítomné ve směsi akustických signálů lze řešit pomocí vícekanálového zpracování. Příspěvek se zabývá použitím adaptivního algoritmu Generalised Sidelobe Canceller. Vlivem užitečného řečového signálu dochází při adaptaci vah k jejich rozvažování. V tomto příspěvku je diskutována možnost, jak rozvažování zabránit. Je navržena modifikace algoritmu s použitím detektoru řečové aktivity.

1 Úvod

Příspěvek se zabývá zvýrazňováním řeči v šumu. Tato úloha nachází uplatnění např. v mobilní komunikaci, pomůckách pro sluchově postižené, při sběru akustických dat určených k další analýze, rozpoznávání řeči apod.

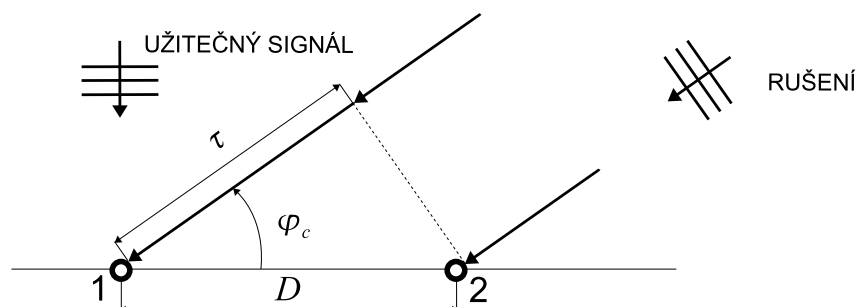
Ke zvýraznění řeči je použit algoritmus vícekanálového zpracování označovaný Generalised Sidelobe Canceller (GSC) [4]. Jeho součástí je adaptivní algoritmus měnící své váhy v průběhu výpočtu. V [1] byl diskutován vliv konvergenční konstanty adaptivního algoritmu. Ukázalo se, že užitečný signál prosakuje do algoritmu aktualizace vah. Je třeba volit kompromis mezi rychlostí konvergence a mírou rozvažování vah.

Podle [7], [5], [2] lze rozvažování vah užitečným signálem zabránit použitím detektoru řeči (Voice Activity Detector - VAD). Tento příspěvek řeší zařazení detektoru do algoritmu GSC.

Chování modifikované struktury bylo ověřeno pomocí simulace vícekanálového systému a též na naměřených datech z vícekanálové databáze AKOMORA [6].

2 Vstupní data

V simulacích jsou použita data generována podle modelu na obrázku 1. Mikrofony jsou rozmístěny v lineární řadě s konstantními rozestupy 5 cm. Data byla generována pro 2, 4, 6, a 8 mikrofonů. Předpokládáme, že jsou mikrofony v dostatečné vzdálenosti od zdrojů. Dopadající akustické vlny jsou považovány za rovinné. Užitečný řečový dopadá na mikrofonní řadu kolmo a rušení pod úhlem 10° .



Obrázek 1: Model signálu dopadajícího na řadu dvou mikrofonů.

Vzorky užitečného signálu jsou pro všechny mikrofony shodné. Díky úhlu ϕ_c vzniká mezi vzorky rušení zaznamenanými na jednotlivých mikrofonech zpoždění.

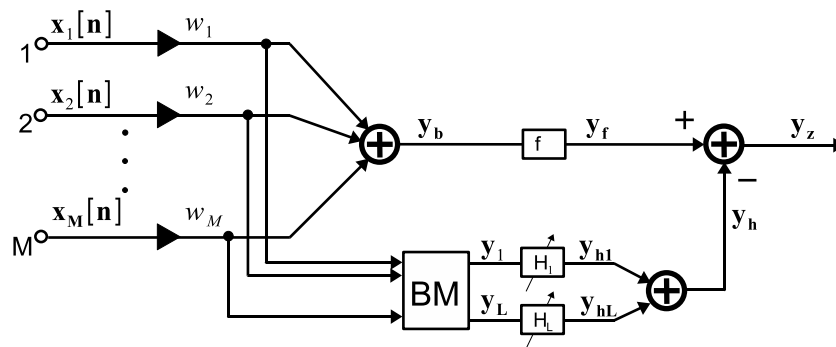
Zaznamenávaný signál musí být na mikrofonech dostatečně navzorkován, aby nedocházelo k nejednoznačnosti úhlu dopadu, prostorovému aliasingu [11], [8]. Signál nesmí být vzorkován rychleji, než je doba šíření vlnoplochy mezi sousedními mikrofony. Pro vzdálenost 5 cm stačí vzorkovací frekvence 8 kHz.

Jako užitečný signál byl pro všechny kanály použit úsek řečového signálu. Rušení je reprezentováno šumem s normálním rozložením. Jednotlivé kanály rušení byly vzájemně zpožděny tak, aby byl simulován dopad pod požadovaným úhlem.

Vedle uměle generovaných dat byly použity nahrávky z vícekanálové databáze AKOMORA [6]. Jedná se o nahrávku řeči a nahrávku bílého šumu pořízené pomocí řady čtyř mikrofonů v bezodrazové komoře.

3 Použitý Algoritmus

Struktura algoritmu GSC je zachycena na obrázku 2. Algoritmus se skládá ze dvou částí. Horní větev je označována jako DAS (Delay And Sum beamformer), dolní jako ANC (Adaptive Noise Canceller).



Obrázek 2: Struktura GSC.

DAS provádí průměrování kanálů. Signál y_b je poté zpožděn, aby byla zajištěna kauzalita ANC. Filter f představuje okrajovou podmínku Frostova algoritmu [3]. (GSC je alternativní formulací Frostova algoritmu.)

ANC větev provádí odhad koherentního rušení [6] přítomného v signálu y_f . Správná funkce ANC je podmíněna odstraněním užitečného signálu ze vstupů. To zajišťuje blokovací matice BM. H_i představuje sadu adaptivních FIR filtrů. Jejich koeficienty jsou aktualizovány LMS algoritmem [12]:

$$H_{ij}[n + 1] = H_{ij}[n] - \mu y_{ij} y_z. \quad (1)$$

μ reprezentuje konvergenční konstantu LMS algoritmu. μ má vliv na rychlost adaptace vah. V [1] byl diskutován vliv tohoto parametru. Je-li jeho hodnota malá, konvergují váhy k optimální hodnotě pomalu. Je-li velká, dochází díky přítomnosti užitečného signálu k rozvažování vah. Pojem malé a velké hodnoty parametru μ souvisí s rozptylem vstupního signálu.

Pro vyřešení dilematu pomalá konvergence vs. rozvažování vah lze použít detektor řeči (VAD). Ten zastaví adaptaci vah při přítomnosti řečového signálu. K adaptaci pak dochází pouze v pauzách mezi řečí.

3.1 Detektor řeči

K experimentům popsaným v tomto příspěvku bylo použito jednoduchého detektoru řeči. Jeho řešení vychází z předpokladu, že řeč nese větší energii než pauza. K tomuto účelu je třeba vypočítat energetickou obálku E_{env} vstupního signálu x . Tu lze vypočítat např. pomocí filtru klouzavých průměrů [10]:

$$E_{env}[n] = \sum_{k=1}^N x^2[k], \quad (2)$$

kde N značí délku filtru. Řeč je detekována v případě, že energetická obálka E_{env} překročila práh THR_e . Pro přehlednost zavedme funkci řečová aktivita VAD :

$$VAD[k] = \begin{cases} 1, & E_{env} \geq THR_e \\ 0, & E_{env} < THR_e. \end{cases} \quad (3)$$

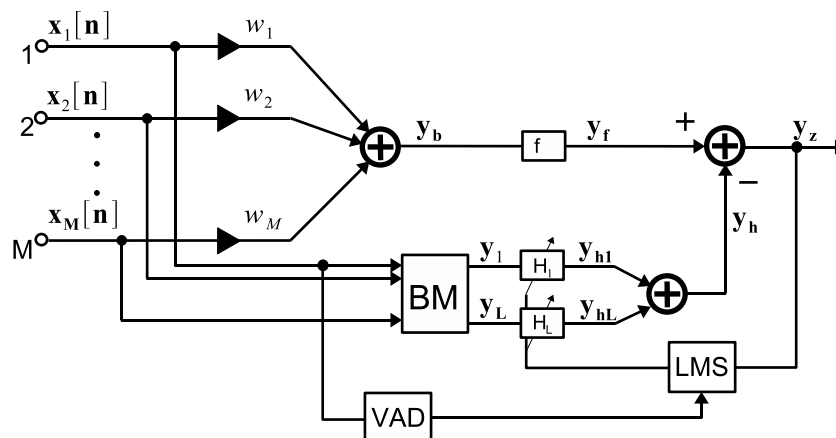
Délka filtru N a práh THR_e byly voleny experimentálně tak, aby byl řečový signál detekován.

3.2 Zařazení detektoru do GSC

Detektor řeči byl do struktury GSC zařazen podle obrázku 3. Vstupní signál je brán z prvního kanálu. Výstup ovlivňuje aktualizaci vah (rovnici 1) následujícím způsobem:

$$H_{ij}[n+1] = H_{ij}[n] - VAD[n]\mu y_{ij}y_z. \quad (4)$$

Při přítomnosti řečového signálu dojde k zastavení adaptace vah.



Obrázek 3: Zapojení VAD do GSC.

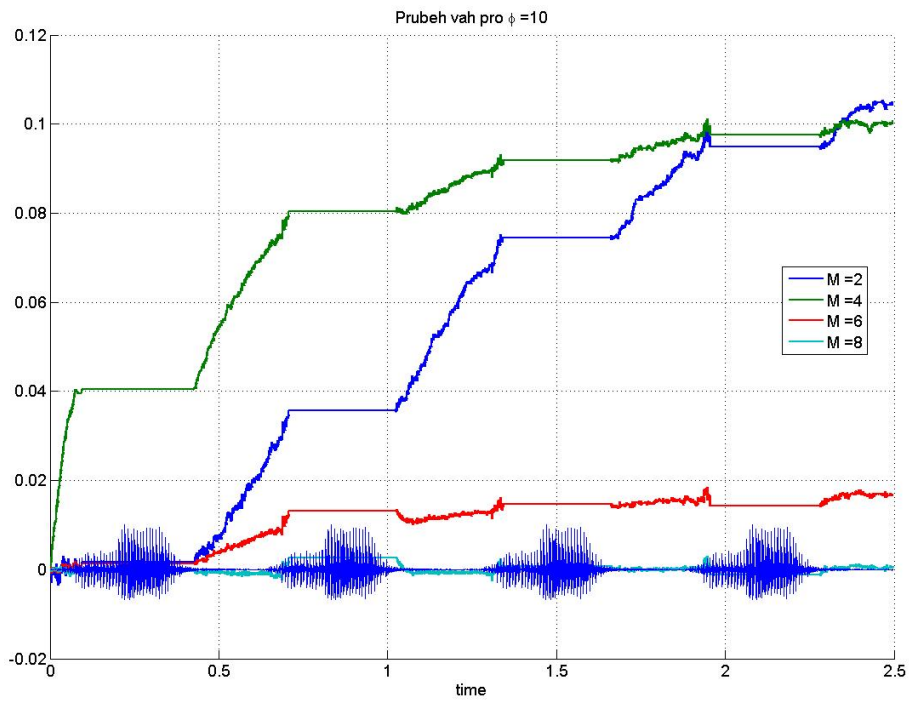
4 Výsledky

Hodnocení zvýraznění řeči bylo prováděno podle míry zvýraznění odstupů signálu od šumu (SNRE) [9]. SNRE je odhadováno z prvního vstupního kanálu a výstupu.

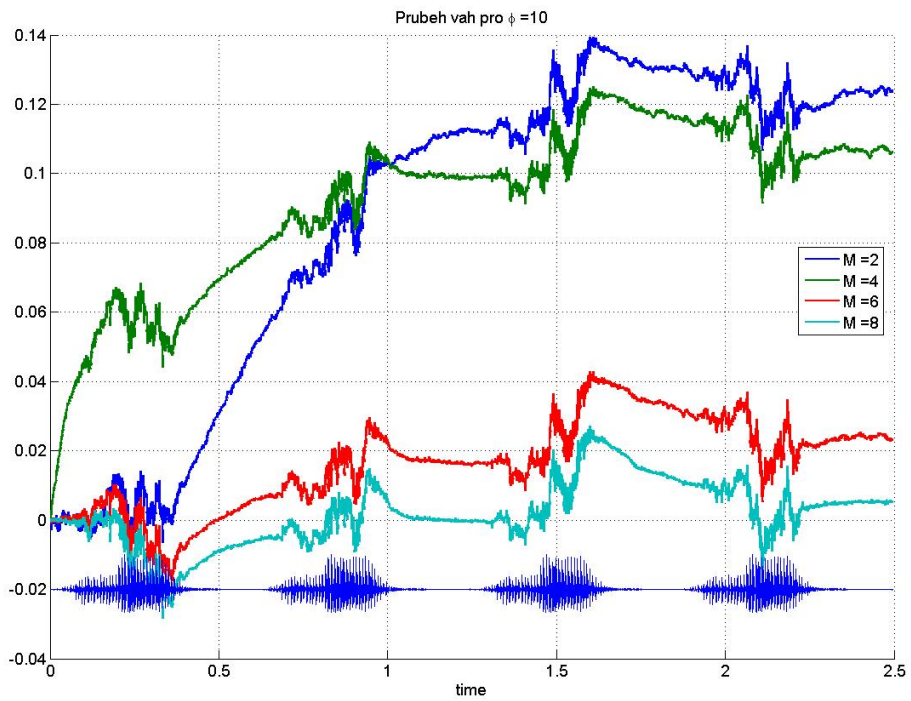
Cílem experimentů bylo posoudit vliv detektoru řeči na SNRE. V experimentu s umělými daty byly provedeny výpočty pro 2, 4, 64 a 8 kanálů. Konvergenční konstanta μ nabývala hodnot 0.005, 0,05, 0.2 a 0.5. Adaptivní filtry H_i měly délku 7 vzorků.

Výsledné hodnoty SNRE (v dB) pro GSC a modifikovaný GSC s detektorem řeči ukazují tabulky 1 a 2. M značí počet kanálů.

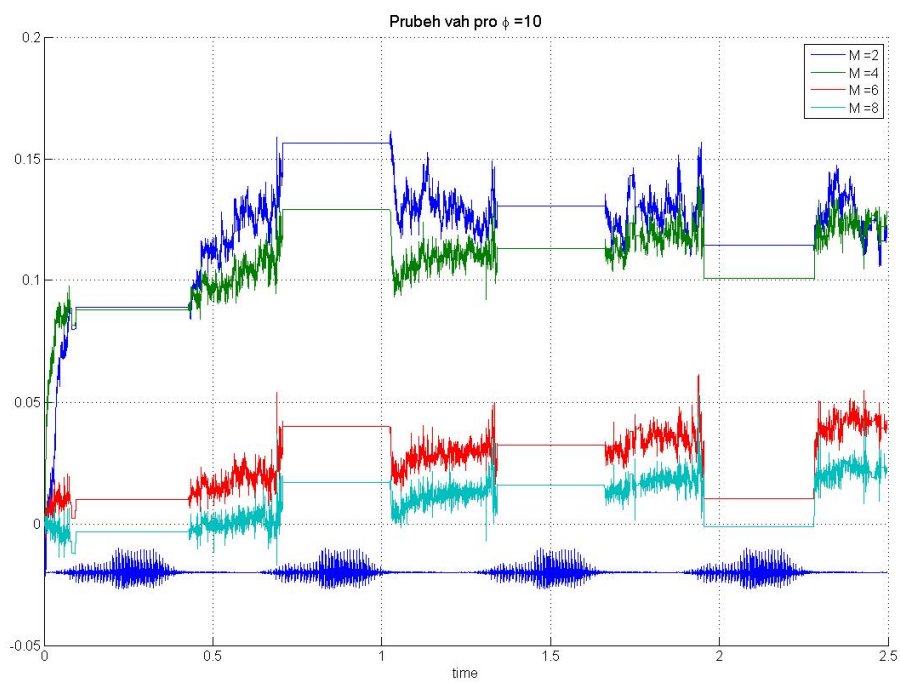
Vliv detektoru řeči demonstrují následující obrázky. Průběhy vah pro malou konvergenční konstantu $\mu = 0,02$ ukazují obrázky 4 a 5. Vliv velké hodnoty $\mu = 0,2$ ukazují obrázky 6 a 7.



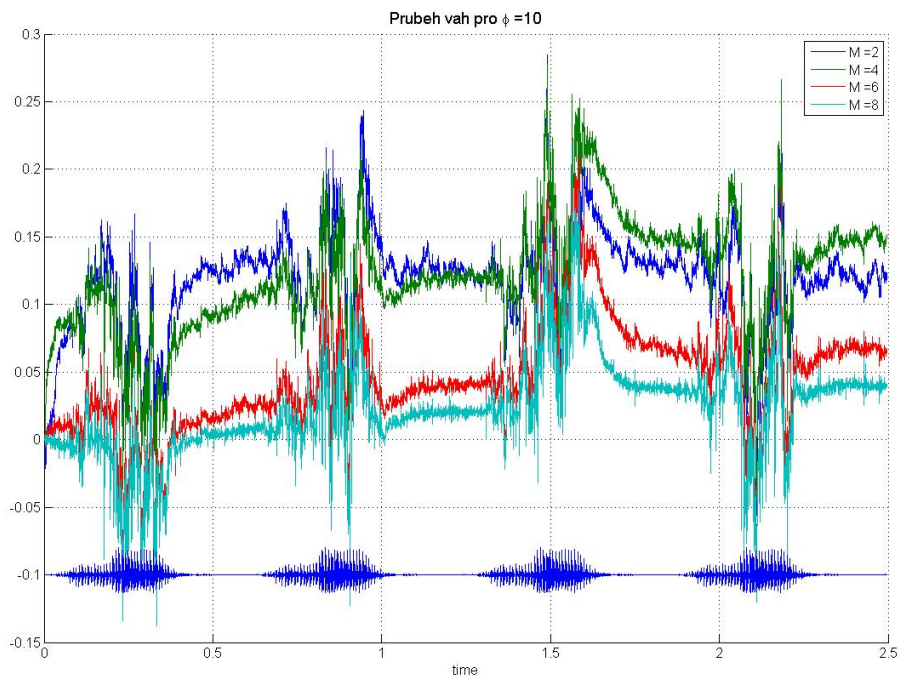
Obrázek 4: Průběhy vah s VAD $\mu = 0,02$.



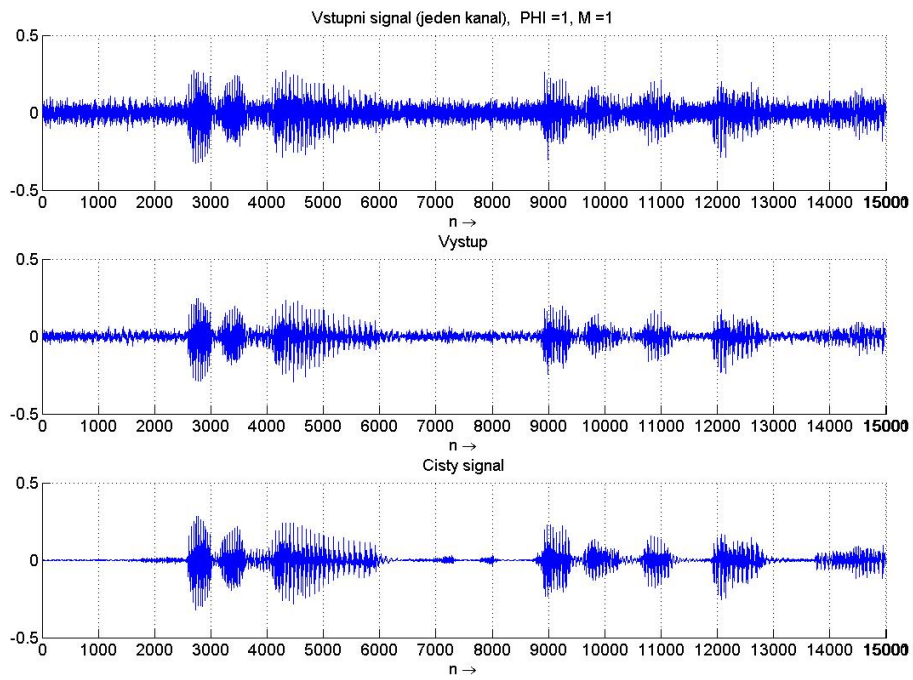
Obrázek 5: Průběhy vah bez VAD $\mu = 0,02$.



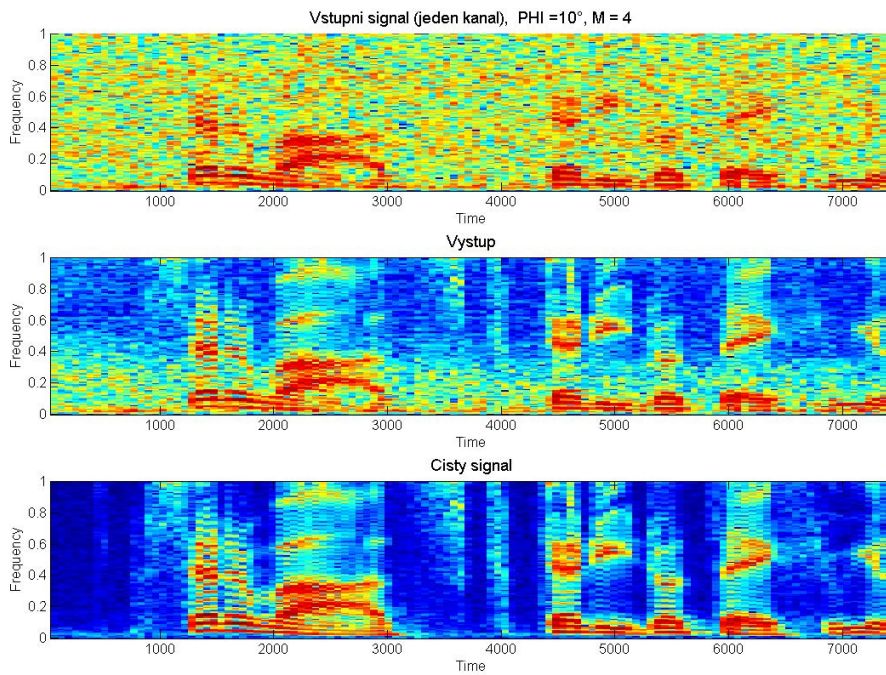
Obrázek 6: Průběhy vah s VAD $\mu = 0, 2$.



Obrázek 7: Průběhy vah bez VAD $\mu = 0, 2$.



Obrázek 8: Časový průběh signálů, $\mu = 0,9$.



Obrázek 9: Spektrogramy, $\mu = 0,9$.

M	2	4	6	8
$\mu = 0,005$	6,5	8,7	9,7	10,5
$\mu = 0,02$	8,1	9,5	10,2	10,7
$\mu = 0,05$	8,5	9,4	9,8	10
$\mu = 0,2$	8,4	7,9	6,6	4,9
$\mu = 0,5$	7,5	3	-41	-125

Tabulka 1: SNRE [dB] pro GSC.

M	2	4	6	8
$\mu = 0,005$	5,2	8	9,1	10
$\mu = 0,02$	7,2	9,2	10	10,7
$\mu = 0,05$	8,2	9,6	10,3	10,8
$\mu = 0,2$	8,8	9,3	9,2	9
$\mu = 0,5$	8,5	7,4	-2,1	-48

Tabulka 2: SNRE [dB] pro GSC s detektorem řeči.

Experimenty s naměřenými daty byly provedeny pouze pro 4 mikrofony. Právě tolik jich bylo použito pro záznam. Při zpracování byla konvergenční konstanta rovněž nastavována pro hodnoty 0.005, 0.02, 0.05, 0.2, 0.5 a 0.9. Výsledky ukazuje tabulka 3. Obrázky 8 a 9 ukazují časové průběhy a spektrogramy analyzovaných dat. Obrázky zachycují vstupní, výstupní a čistý signál.

μ	VAD	bez VAD
0,005	5	5,1
0,02	5,4	5,7
0,05	5,9	6,4
0,2	7,1	7,1
0,5	7,7	6,6
0,9	7,9	5,5

Tabulka 3: SNRE [dB] pro naměřená data.

5 Závěr

Tabulky v oddílu 4 ukazují, že popsané algoritmy provádějí zvýraznění řeči. Tabulky naznačují, pro jaké hodnoty μ má smysl VAD. Při malých hodnotách (0,005, 0,02) je detektor mírně na obtíž. Rychlost adaptace vah je natolik pomalá, že ji přítomnost užitečného signálu příliš neovlivní. VAD naopak jejich konvergenci ještě zpomalí. Tento jev také demonstrují obrázky 4 a 5.

Použití VAD má smysl pro velká μ , má-li být zajištěna rychlá konvergence. Tento efekt je patrný zejména pro $\mu = 0,2$, obrázky 7 a 6. Z tabulek 1 a 2 je zřejmé, že μ nelze zvyšovat do nekonečna. Pro řadu šesti a osmi mikrofonů a $\mu = 0,5$ je adaptivní algoritmus nestabilní a ke zýrazňování řeči nedochází.

Tabulka 3 a obrázky 8 a 9 ukazují, že algoritmy fungují nejen pro umělá data, ale i pro reálné signály. Výsledné SNRE nelze mezi umělými a reálnými daty porovnávat. Jako vstupy byly použity úplně jiné signály.

Experimenty ukázaly, že pro zrychlení adaptace vah má smysl zařazení VAD do struktury GSC. Největším úskalím řešení právě VAD. V popsaných experimentech byl použit jednoduchý VAD, jehož parametry byly nastavovány empiricky. V praxi by bylo třeba, aby byly tyto parametry nezávislé na vstupních datech.

6 Poděkování

Tato práce je podporována výzkumným záměrem Transdisciplinární výzkum v biomedicínském inženýrství 2 č. MSM6840770012 a grantem Modelování biologických a řečových signálů GD 102/03/H085.

Reference

- [1] Václav Bolom. Zvýrazňování řeči pomocí vícekanálového zpracování. In *Poster 2007 [CD-ROM]*, number 1, pages 16 – 23, Praha, 2007. CTU, Faculty of Electrical Engineering.
- [2] Markus Buck, Tim Haulick, and Hans-Jörg Pfeleiderer. Self-calibrating microphone arrays for speech signal acquisition: A systematic approach. *Signal Processing*, (86):1230–1238, 2006.
- [3] Otis Lamont Frost. An algorithm for linearly constrained adaptive array processing. In *IEEE*, volume 60, pages 926–934, August 1972.
- [4] LLOYD J Griffiths and W Jim Charles. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transaction on Antenas and Propagation*, AP-30(1), January 1982.
- [5] Jwu-Sheng Hu, Wei-Han Liu, and Chieh-Cheng Cheng. Robust beamforming of microphone array using h infinity adaptive filtering technique. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E89-A(3):708–715, 2006.
- [6] Jan Ingerle. *Methods of speech signal enhancement combining beamforming and postfiltration*. PhD thesis, FEL ČVUT, Prague, February 2003.
- [7] Jean-Baptiste Maj, Liesbeth Royackers, Jan Wouters, and Marc Moonen. Comparison of adaptive noise reduction algorithms in dual microphone hearing aids. *Speech Communication* 48 (2006), (48):957–970, 2006.
- [8] Prabhakar S Naidu. *Sensor array signal processing*. CRC Press LCC, 2001.
- [9] Pavel Sovka and Petr Pollák. *Vybrané metody číslicového zpracování signálů*. Vydavatelství ČVUT, Praha, 2003.
- [10] Jan Uhlíř and Pavel Sovka. *Číslicové zpracování signálů*. Vydavatelství ČVUT, Praha, 2002.
- [11] Barry D van Veen and Kevin M Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, pages 4–24, April 1988.
- [12] Bernard Widrow and S D Stearns. *Adaptive Signal Processing*. Prentice-Hall, 1985.

Václav Bolom

Katedra teorie obvodů, Fakulta Elektrotechnická, České vysoké učení technické v Praze, Technická 2, 166 27 Praha 6, tel: 224 355 869, email: bolomv1@fel.cvut.cz

Pavel Sovka

Katedra teorie obvodů, Fakulta Elektrotechnická, České vysoké učení technické v Praze, Technická 2, 166 27 Praha 6, tel: 224 352 291, email: sovka@fel.cvut.cz