

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

Problém zvýrazňování řeči přítomné ve směsi akustických signálů lze řešit pomocí vícekanálového zpracování. V příspěvku je popsána implementace jedné z metod vícekanálového zpracování označované jako Generalised Sidelobe Canceller (GSC). Metoda je implementována pro zpracování ve frekvenční oblasti pomocí metody sčítání přesahů. V příspěvku jsou ukázány výsledky simulací provedených na umělých testovacích datech. Dále jsou navrženy metody, které vylepšují popisovaný algoritmus. V závěru příspěvku je uveden návrh měřicího pracoviště, na němž budou ověřovány výsledky simulací. Metod směrového příjmu lze využít např. v mobilní komunikaci, v pomůckách pro sluchově postižené, pro sběr dat pro analýzu řeči apod.

1 Úvod

Příspěvek se zabývá zvýrazňováním řeči v šumu. Je zde popsána implementace algoritmu vícekanálového zpracování. Popisovaný algoritmus je založen na metodě GSC (Generalized Sidelobe Canceller) [3]. K snímání zvuku ve vícekanálových systémech se používá mikrofonních polí. Popisovaný algoritmus je navržen pro lineární řadu dvanácti mikrofonů s konstantní roztečí 5 cm. V příspěvku budou ukázány výsledky simulací s umělými daty.

2 Vstupní data

V simulacích je uvažováno rozložení zdrojů zvukového signálu podle obrázku 1. Zdroje se nacházejí v dostatečné vzdálenosti od senzorů, takže příchozí signály tvoří rovinné vlny. Užitečný řečový signál dopadá kolmo na řadu a rušení pod úhlem φ_c .

Vzorky užitečného signálu jsou na všech mikrofonech shodné. V případě rušení dochází ke zpoždění τ mezi jednotlivými kanály. Zpoždění závisí na úhlu φ_c a vzdálenosti mezi mikrofony D . Je dáno vztahem

$$\tau = \frac{D}{c} \cos \varphi_c, \quad (1)$$

kde c je rychlost šíření zvuku. Další zdroje rušení nejsou v tomto modelu uvažovány.

Signál na i -tém senzoru je dán vztahem

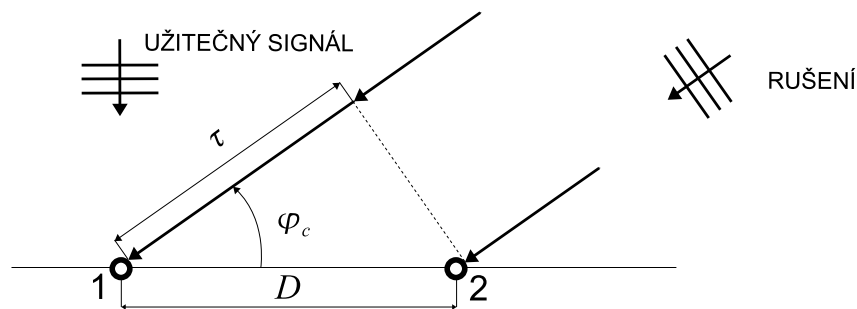
$$x_i[k] = s[k] + n_i[k], \quad (2)$$

kde $s[k]$ označuje užitečný signál a $n_i[k]$ rušení na i -tém senzoru.

2.1 Modelování vstupních dat

Za užitečný signál byla vybrána nahrávka promluvy pořízené v bezodrazové komoře. Tento signál je shodný pro všechny kanály.

Rušení je tvořeno šumovým signálem s normálním rozdělením. Signál byl nejprve zkopírován do všech kanálů. Poté byly kanály zpožděny. Zpoždění vzorků v jednotlivých kanálech odpovídá úhlu $\varphi_c = 30^\circ$ a rozteči $D = 5\text{cm}$. Vzhledem k neceločíselnému zpoždění vzorků bylo posunutí provedeno ve frekvenční oblasti [5]. Amplituda vstupních signálů je normována do intervalu $(-1, 1)$.



Obrázek 1: Model signálů dopadajících na řadu tří mikrofonů.

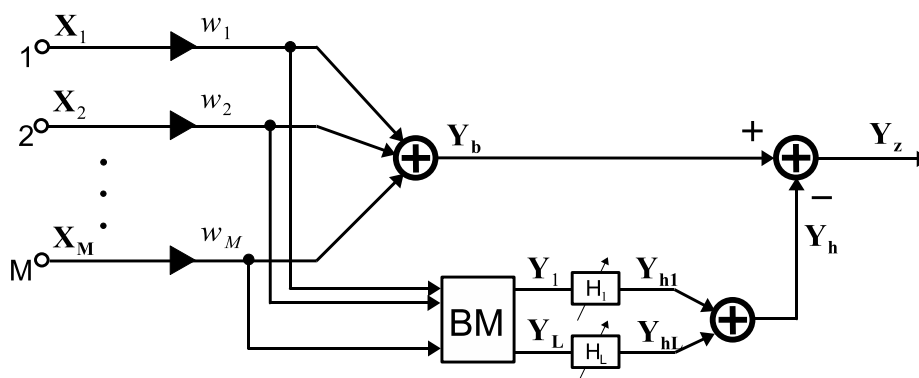
3 Algoritmus GSC

Struktura algoritmu GSC [3] je zachycena na obrázku 2. Tento algoritmus je ekvivalentní adaptivnímu beamformeru [2]. Na vstupu jsou spektra $\mathbf{X}_i(e^{j\omega T})$ vstupních signálů $x_i[k]$. Systém se skládá ze dvou větví. Horní větev je tvořena konvenčním beamformerem. Vstupy jsou průměrovány. Váhy w_i mají hodnotu $\frac{1}{M}$. M představuje počet mikrofonů. Výstupem konvenčního beamformeru je $\mathbf{Y}_b(e^{j\omega T})$. Ke zvýraznění užitečného signálu dojde díky sečtení všech kanálů.

Spodní větev provádí adaptivní potlačování šumu (ANC - Adaptive Noise Cancelation). Cílem této větve je odhadnout rušení přítomné v $\mathbf{Y}_b(e^{j\omega T})$. Podmínka správné činnosti je oddělení užitečného signálu ze všech kanálů. K tomu slouží blokovácí matice BM. Nejčastěji je používána matice s M řádky a M - 1 sloupcem. Provádí rozdíl sousedních kanálů a má tvar

$$BM = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}. \quad (3)$$

Váhy H_i jsou řízeny LMS algoritmem [6] tak, aby $\mathbf{Y}_h(e^{j\omega T})$ byl odhadem rušení přítomného v $\mathbf{Y}_b(e^{j\omega T})$ ve smyslu nejmenších čtverců.



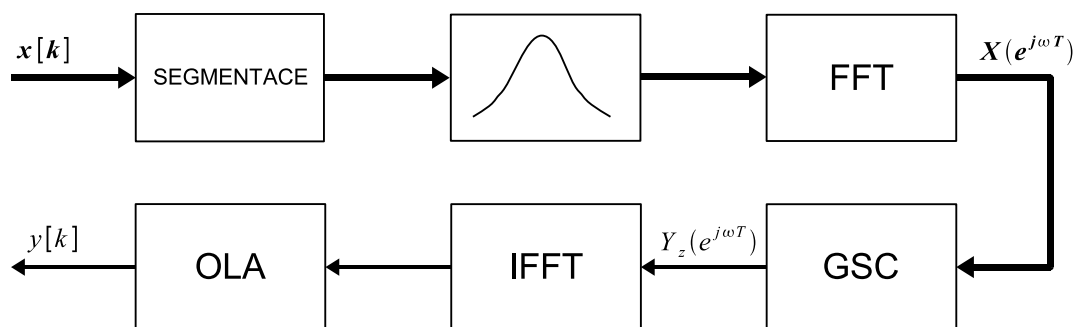
Obrázek 2: Blokové schéma GSC.

4 Implementace

Zpracování probíhá ve frekvenční oblasti. V popisovaném algoritmu je využito metody sčítání přesahů (OLA - Overlap And Add) [5]. Blokové schéma celého algoritmu je na obrázku 3. Při segmentaci je využito Hammingova okna a 50% přesahu.

Zpracování ve frekvenční oblasti má několik výhod. Při adaptivní filtraci je využito symetrií ve spektru. Stačí tedy zpracovávat pouze polovinu spektra. Další výhodou je možnost rozšíření

metody o další zpracování ve frekvenční oblasti. Toho je využito v případě, kdy užitečný signál nedopadá kolmo na řadu. Pro natáčení dopadající vlnoplochy se použije stejné metody jako při generování rušivého signálu.



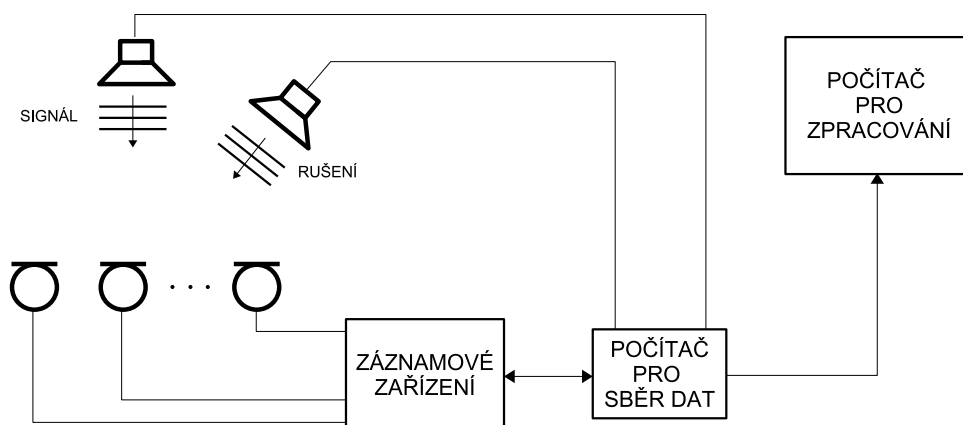
Obrázek 3: Blokové schéma algoritmu využívajícího metody sčítání přesahů.

4.1 Modifikace algoritmu

Metodu GSC lze rozšířit tak, aby potlačovala i nekoherentní rušení. Nekoherentní rušení má na každém senzoru jiný průběh. Bývá způsobeno např. šumem elektrických soustav. Potlačení nekoherentního rušení lze docílit zařazením Wienerova filtru do horní větve za konvenční beamformer [1] [7]. Wienerův filtr pracuje ve frekvenční oblasti. Filtr je řízen vstupními spektry $\mathbf{X}_i(e^{j\omega T})$.

4.2 Návrh měřicího pracoviště

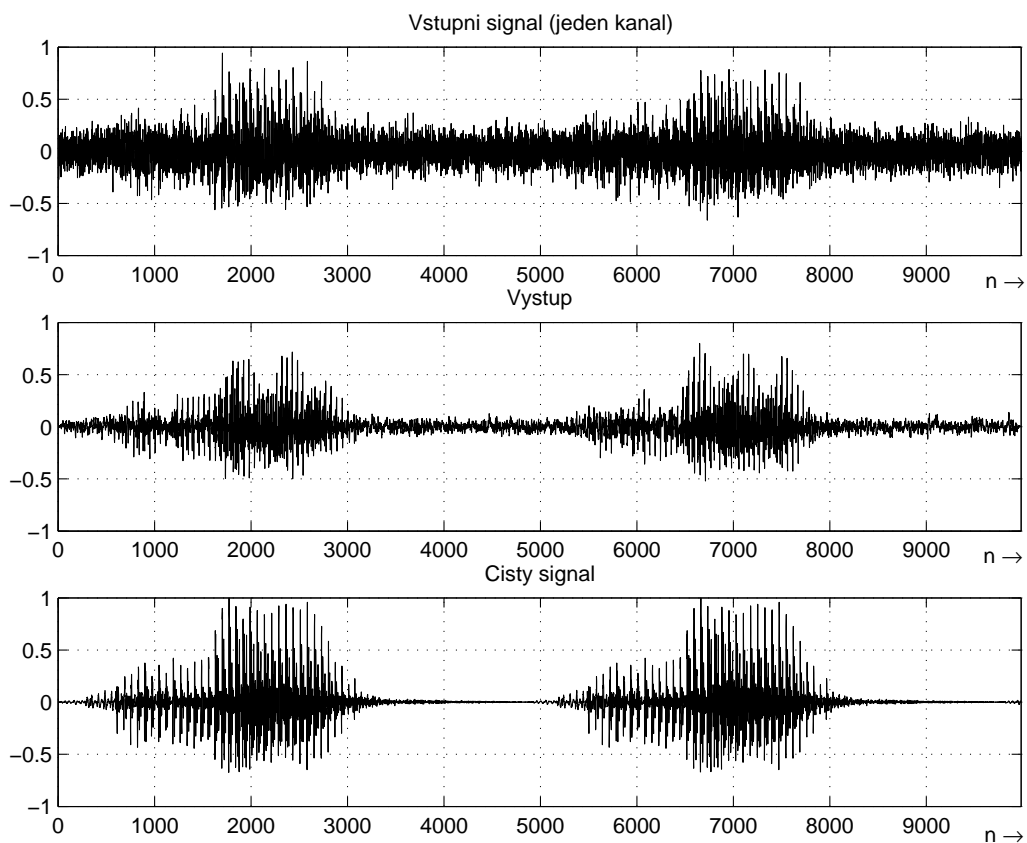
Dalším cílem ve výzkumu algoritmů směrového příjmu je ověření simulací v praxi. Naše pracoviště je vybaveno zařízením, které snímá zvuk až z dvanácti mikrofonů zároveň. Návrh pracoviště je schématicky zachycen na obrázku 4. Dalšími součástmi jsou reproduktory a počítač pro sběr dat a vysílání testovacích signálů. Zpracování dat bude probíhat na dalším počítači. Experimenty budou provedeny v bezodrazové komoře.



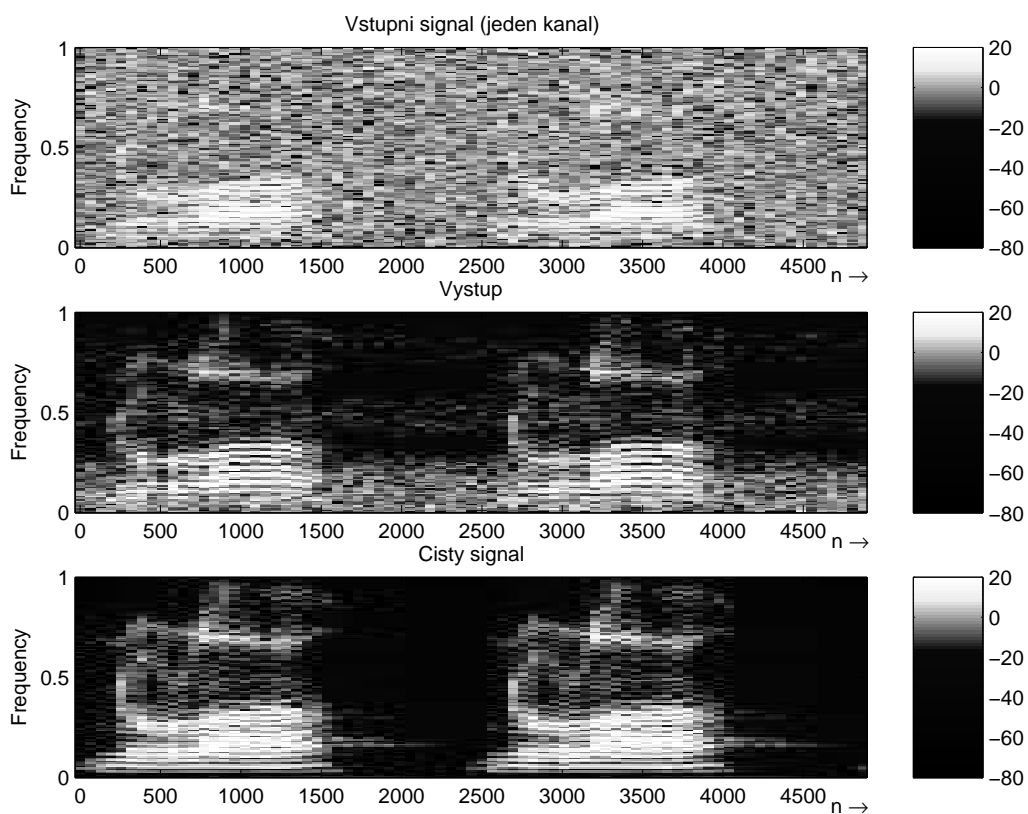
Obrázek 4: Návrh měřicího pracoviště.

5 Výsledky

Výstupem systému je jednorozměrný signál zvýrazněné řeči. Obrázek 5 ukazuje časové průběhy vstupního signálu (jednoho kanálu), výstupního signálu a čistého nezašuměného signálu. Tytéž signály zachycuje obrázek 6 na spektrogramech.



Obrázek 5: Časové průběhy vstupních a výstupních dat.



Obrázek 6: Spektrogramy vstupních a výstupních dat.

6 Závěr

Z grafů 5 a 6 je patrné, že popsaná metoda provádí zvýraznění zašuměné řeči. Kvalitu zvýraznění lze posuzovat podle různých kritérií [4]. Jedním z nich je subjektivní posouzení na základě poslechu. Kvalita zvýraznění závisí na počtu použitých kanálů. Pro ilustraci byl zvolen jejich velký počet.

Popisovaná metoda je vhodná pro potlačení koherentního rušení. Jeho vzorky se objevují ve všech kanálech s časovým posunem. Lepšího zvýraznění lze dosáhnout tím, že je vážený součet vstupů nahrazen vhodnými FIR filtry [3]. Popsaná metoda však pracuje rychleji a má i menší paměťové nároky. Rychlost zpracování může být z jedním z kritérií pro výběr metody pro zpracování v reálném čase.

Zvýrazňování řeči pomocí vícekanálového zpracování nalezne uplatnění v mobilní komunikaci, pomůckách pro sluchově postižené, sběru dat pro analýzu řeči apod. Rychlost zpracování je důležitá zejména v prvních dvou případech.

7 Poděkování

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

Reference

- [1] Fischer S , Simmer KU. Beamforming microphone arrays for speech acquisition in noisy environments. *Speech communication*, 20:215–227, 1996.
- [2] Frost OL. An algorithm for linearly constrained adaptive array processing. In *IEEE*, volume 60, pages 926–934, August 1972.
- [3] Griffiths LJ, Charles WJ. An alternative approach to linearly constrained adaptive beamforming. *IEEE Transaction on Antennas and Propagation*, AP-30(1), January 1982.
- [4] Ingerle J. *Metody zvýrazňování řečového signálu kombinující směrový příjem a postfiltraci*. Disertační práce, FEL ČVUT, Prague, February 2003.
- [5] Sovka P ,Pollák P. *Vybrané metody číslicového zpracování signálů*. Vydavatelství ČVUT, Praha, 2003. Skriptum FEL ČVUT.
- [6] Widrow B, Stearns SD. *Adaptive Signal Processing*. Prentice-Hall, 1985.
- [7] Zelinski R. A microphone array with adaptive post-filtering for noise reduction in reverberant rooms. In *International Conference of Acoustic Speech Signal Processing*, pages 2578–2581, New York, 1988.

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