Li-chun Yang, Yun-tao Qian, 2014. Speech enhancement with a GSC-like structure employing sparse coding. *Journal of Zhejiang University-SCIENCE C (Computers & Electronics),* **15**(12):1154-1163. [doi:10.1631/jzus.C1400085]

Speech enhancement with a GSC-like structure employing sparse coding

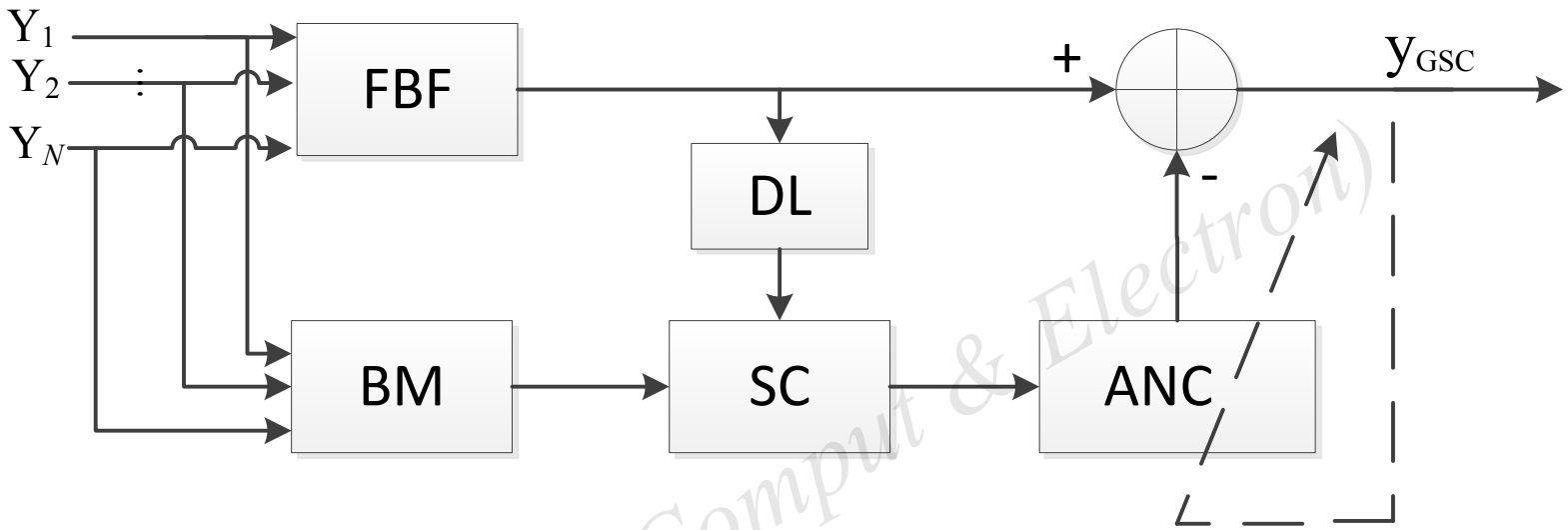
Key words: Generalized sidelobe canceller, Speech enhancement, Voice activity detection, Dictionary learning, Sparse coding

Contact: Li-chun YANG
E-mail: lichun_y@126.com
ORCID: http://orcid.org/0000-0003-1651-798X

Motivation

- Use a sparse coding algorithm to suppress the residual desired signal of a reference signal, so as to achieve a higher performance of interference suppression with less distortion for the generalized sidelobe canceller.
- Disadvantages of existing methods:
  - The error in time difference of arrival (TDOA) exists in the real acoustic environment
  - The adaptive blocking matrix cannot reduce the leakage in the real acoustic environment due to reverberation
  - The transfer function ratio estimation is not very accurate due to long impulse response

Structure of the proposed method



Compared with traditional GSC structure, the proposed structure includes a dictionary learning (DL) block and a sparse coding (SC) block.

The DL block is used to obtain the interferer dictionary.

The SC block is used to suppress the residual desired signal that leaks into the reference signal.

DL block

The training samples for DL are obtained in the segments of the desired signal inactivity.

For training samples $\boldsymbol{s}_i \in \mathbb{R}^m, i = 1,2,...,n$, dictionary matrix $\boldsymbol{D}$ and its coefficient vector $\boldsymbol{\beta}$ meet the following constraint:

$$\arg\min_{\boldsymbol{D},\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}\| \boldsymbol{D}\beta_i - \boldsymbol{s}_i \|_2^2 + \lambda' \| \beta_i \|_1\right)$$

$$\text{s.t.} \quad \forall j = 1,2,...,k, \quad \boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{d}_j \le 1$$

$$\text{where } m > n,\ \boldsymbol{D} \in \mathbb{R}^{n\times m},\ \ \boldsymbol{\beta} \in \mathbb{R}^{m\times 1}$$

$$\boldsymbol{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, ..., \boldsymbol{d}_m],\ \boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_m]$$

SC block

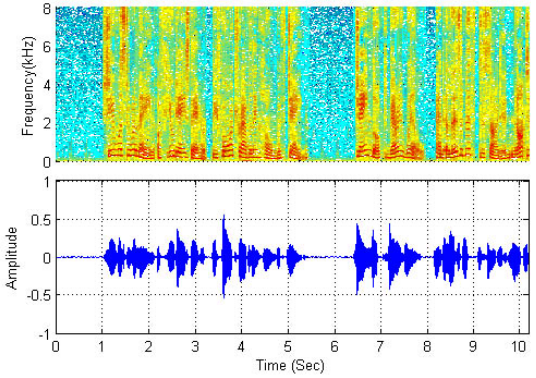
Suppose the frame length of a signal is *m* and define ***z*** as a vector with *m* samples of the reference signal. The coefficient vector ***w*** and the reference signal ***z*** in the interferer dictionary satisfy

$$\tilde{\boldsymbol{w}} = \underset{\boldsymbol{w}}{\operatorname{argmin}} \frac{1}{2} \left\| \boldsymbol{D}\boldsymbol{w} - \boldsymbol{z} \right\|_2^2 + \lambda \left\| \boldsymbol{w} \right\|_1$$

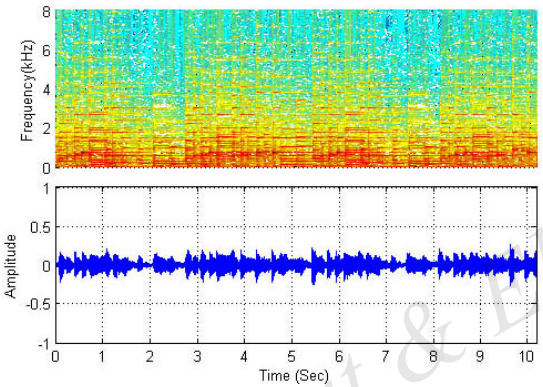
The output signal of sparse reconstruction $\tilde{\boldsymbol{z}}$ is

$$\tilde{\boldsymbol{z}} = \boldsymbol{D}\tilde{\boldsymbol{w}}$$

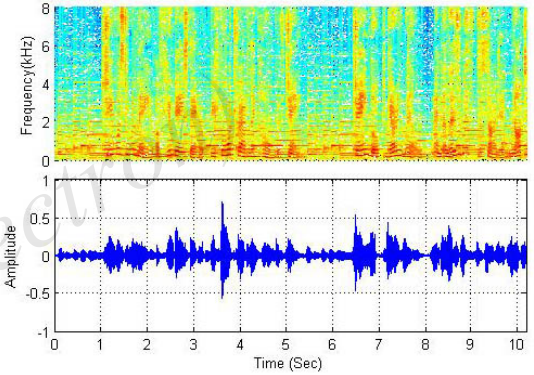
Simulation results



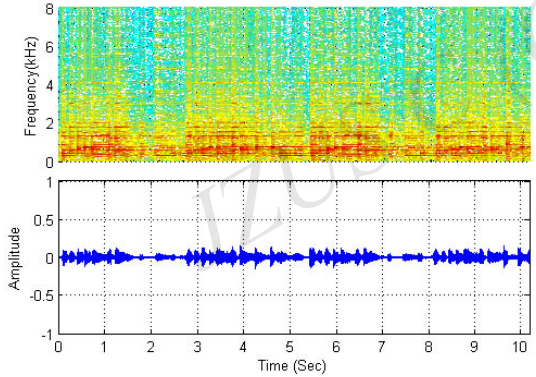
(a) Clean speech



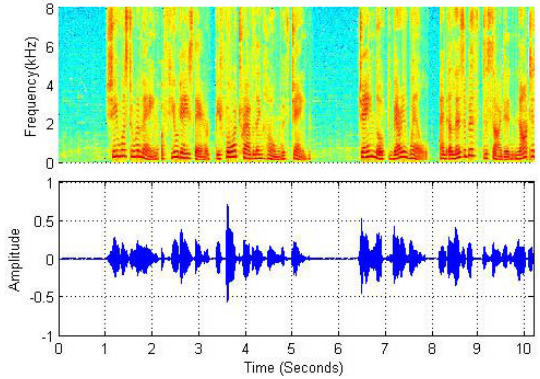
(b) Music



(c) Noisy signal



(d) Reference signal with sparse coding



(e) Speech enhancement with the proposed method

Conclusions

- The training samples for interferer dictionary learning come from the segments of the desired signal inactivity.
- Reference signal of interference with the sparse coding method has been studied for reducing the residual desired signal.
- GSC with improved reference signal can suppress the interfering signal effectively.