Introduction to distributed speech enhancement algorithms for ad hoc microphone arrays and wireless acoustic sensor networks

Part III: GSC-based distributed speech enhancement in WASNs

<u>Sharon Gannot</u>¹ and Alexander Bertrand²

¹Faculty of Engineering, Bar-Ilan University, Israel

²KU Leuven, E.E. Department ESAT-STADIUS, Belgium



EUSIPCO 2013, Marrakesh, Morocco



This presentation is partly based on the Ph.D. dissertation by Shmulik Markovich-Golan



Outline

Outline

Goal

Develop recursive and distributed speech enhancement algorithms based on the generalized sidelobe canceller (GSC) structure.

Two Algorithms

- Distributed single constraint GSC (DS-GSC):
 - N (#nodes) broadcast channels.
 - 2 alternating GSC blocks per node.
- Distributed GSC (D-GSC):
 - *N* + *P* (#nodes+#constraints) broadcast channels.
 - 1 GSC block per node (with extended input).

LCMV & MVDR

[Er and Cantoni, 1983]; [Van Veen and Buckley, 1988]

LCMV Criterion and Solution

• Minimize noise power $\mathbf{w}^H \mathbf{\Phi}_{nn} \mathbf{w}$ Such that a linear constraint set is satisfied: $\mathbf{C}^H \mathbf{w} = \mathbf{g}$.

• Closed-form solution:
$$\mathbf{w}(\ell, k) = \mathbf{\Phi}_{nn}^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{\Phi}_{nn}^{-1} \mathbf{C})^{-1} \mathbf{g}.$$

MVDR Criterion and Solution

- One desired signal \Rightarrow Single constraint (P = 1).
- Minimize noise power $\mathbf{w}^H \mathbf{\Phi}_{nn} \mathbf{w}$ Such that $(\mathbf{h}^d)^H \mathbf{w} = 1$.

• Closed-form solution:
$$\mathbf{w}(\ell,k) = rac{\mathbf{\Phi}_{nn}^{-1}\mathbf{h}^d}{(\mathbf{h}^d)^H\mathbf{\Phi}_n^{-1}\mathbf{h}^d}.$$

The GSC Implementation Concepts

The Generalized Sidelobe Canceller Implementation

For Constrained Minimization [Griffiths and Jim, 1982]

Split the Beamformer

- $\mathbf{w} = \mathbf{q} \mathbf{w}_n$.
- Constraints Subspace:
 q ∈ Span{C}.
- Null Subspace: $\mathbf{w}_n \in \mathcal{N}\{\mathbf{C}\}$.
- $\mathbf{w}_n \triangleq \mathbf{Bf}$.
- **B**: $M \times (M P)$ matrix. Spans the Null Subspace.
- **f**: vector of M P filters.

• \Rightarrow w = q - Bf.



The Generalized Sidelobe Canceller Implementation

GSC Output

$$y = \mathbf{q}^H \mathbf{z} - \mathbf{f}^H \underbrace{\mathbf{B}^H \mathbf{z}}_{\mathbf{u}(\ell,k)}$$

Constraints Subspace ($\mathbf{q} \in \text{Span}\{C\}$):

$$\mathbf{q}(\ell,k) \triangleq \mathbf{C} (\mathbf{C}^{H}\mathbf{C})^{-1}\mathbf{g}$$

Null Subspace (columns of **B** span $\mathcal{N}\{C\}$):

$$\mathbf{B}(\ell,k) \triangleq \mathbf{I}_{\mathrm{M} \times \mathrm{M}} - \mathbf{C} \left(\mathbf{C}^{H} \mathbf{C} \right)^{-1} \mathbf{C}^{H}; \text{ (verify } \mathbf{B}^{H} \mathbf{C} = \mathbf{0})$$

Noise Cancelling Filters (orthogonality principle):

$$E\left\{\mathbf{u}\left(\mathbf{z}^{H}\mathbf{q}-\mathbf{u}^{H}\mathbf{f}\right)\right\} \Rightarrow \mathbf{f}(\ell,k) = \left(\mathbf{B}^{H}\mathbf{\Phi}_{zz}\mathbf{B}\right)^{-1}\mathbf{B}^{H}\mathbf{\Phi}_{zz}\mathbf{q}$$

S. Gannot (BIU) and A. Bertrand (KUL)

The GSC Structure [Griffiths and Jim, 1982]



GSC Blocks

- Fixed beamformer (FBF) satisfies the constraints (q).
- Blocking matrix (BM) generates M P unconstrained signals (**B**).
- Noise canceller (ANC) adaptively (LMS) suppresses the residual noise utilizing M P degrees of freedom (DoF) (f) [Widrow et al., 1975]; [Shynk, 1992].

The Relative Transfer Function GSC (TF-GSC) [Gannot et al., 2001]

MVDR Implementation with RTF:

RTF: :

$$\tilde{\mathbf{h}}^{d}(\ell,k) \triangleq \frac{\mathbf{h}^{d}}{h_{1}^{d}} = \left[1 \ \frac{h_{2}^{d}}{h_{1}^{d}} \ \dots \ \frac{h_{M}^{d}}{h_{1}^{d}}\right]^{T}$$

Constraint:

$$\mathbf{w}^{H}(\ell,k) \widetilde{\mathbf{h}}^{d}(\ell,k) = 1.$$

Closed-form solution:

$$\mathbf{w}(\ell,k) = \frac{\mathbf{\Phi}_{nn}^{-1}\mathbf{h}^d}{(\mathbf{h}^d)^H \mathbf{\Phi}_n^{-1}\mathbf{h}^d}$$

Output signal:

$$y(\ell,k) = \underbrace{h_1^d s^d}_{\tilde{s}_1^d(\ell,k)} + ext{residual noise and interference signals}$$

S. Gannot (BIU) and A. Bertrand (KUL)

The GSC Implementation TF-GSC

The Transfer Function GSC utilizing RTF

[Gannot et al., 2001]

Fixed beamformer:

$$\mathbf{w}_0(\ell,k) = ilde{\mathbf{h}}^d / \| ilde{\mathbf{h}}^d \|^2$$

Blocking matrix:

$$\mathbf{B}(\ell,k) = \begin{bmatrix} -(\tilde{h}_2^d)^* & -(\tilde{h}_3^d)^* & \dots & -(\tilde{h}_M^d)^* \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \dots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

Distributed single constraint GSC (DS-GSC) [Markovich-Golan et al., 2012a]

Problem Statement

Goal

Develop a distributed version of the TF-GSC.

WASN structure

- N nodes, fully connected.
- M_n microphones in the *n*th node.
- Total $M = \sum_{n=1}^{N} M_n$ microphones.

Signals in the STFT domain

- Desired source: $s(\ell, k)$.
- Acoustic transfer function (ATF) of desired source: $h(\ell, k)$.
- Received interferences: $\mathbf{v}(\ell, k)$, covariance: $\mathbf{\Phi}_{vv}(\ell, k)$.
- Microphone signals: $z(\ell, k) = h(\ell, k)s(\ell, k) + v(\ell, k)$.

Local notation

*n*th node

- Microphone signals: $\mathbf{z}_n(\ell) = \mathbf{U}_n^H \mathbf{z}(\ell)$.
- **U**_n an $M \times M_n$ matrix which extracts the M_n elements of the *n*th node from an M dimensional vector.

п

•
$$\begin{bmatrix} z_{n,1}(\ell) \\ \vdots \\ z_{n,M_n}(\ell) \end{bmatrix} = \begin{bmatrix} 0 \cdots 0 & 1 & 0 \cdots 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 \cdots 0 & 1 & 0 \cdots 0 \\ \sum_{n'=1}^{n-1} M_{n'} & M_n & \sum_{n'=n+1}^{N} M_{n'} \end{bmatrix} \begin{bmatrix} z_1(\ell) \\ \vdots \\ z_M(\ell) \end{bmatrix}.$$

Goal

Enhance a filtered version of the desired signal $h_1(\ell, k)s(\ell, k)$.

S. Gannot (BIU) and A. Bertrand (KUL)

Distributed speech enhancement

N local BFs I

Applied Independently at each Node



S. Gannot (BIU) and A. Bertrand (KUL)

Wireless link

N local BFs II

Applied Independently at each Node



Global BF

A replica is Concurrently Applied in all Nodes

- $\dot{y}_g(\ell) = \dot{w}_g^H \dot{y}_l(\ell)$ with input $\dot{y}_l(\ell)$.
- Global RTF: $\dot{\mathbf{h}}_{g} = \frac{\dot{\mathbf{W}}_{l}^{H}\mathbf{h}}{(\dot{\mathbf{W}}_{l}^{H}\mathbf{h})_{1}} = \frac{\dot{\mathbf{W}}_{l}^{H}\mathbf{h}}{h_{1}}.$ Can be estimated from the global signals.
- $\dot{\mathbf{w}}_g = \dot{\mathbf{q}}_g \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g$ satisfying the constraint $\dot{\mathbf{h}}_g^H \dot{\mathbf{w}}_g = 1.$
- FBF: $\dot{\mathbf{q}}_g = \frac{\dot{\mathbf{h}}_g}{\|\dot{\mathbf{h}}_g\|^2}$.
- BM: $\dot{\mathbf{B}}_g$ from SVD $(\dot{\mathbf{h}}_g)$.
- Noise references: $\dot{\mathbf{u}}_{g}(\ell) = \dot{\mathbf{B}}_{g}^{H} \dot{\mathbf{y}}_{I}(\ell).$



Derivation of the Iterative Version I



Derivation of the Iterative Version II

$$\dot{\mathbf{v}}(\ell) = \dot{\mathbf{w}}_{g}^{H} \dot{\mathbf{W}}_{l}^{H} \mathbf{v}(\ell) = \dot{\mathbf{w}}_{l}^{H} \dot{\mathbf{W}}_{g}^{H} \mathbf{v}(\ell)$$

$$\begin{split} \gamma = & E\{|\dot{\mathbf{v}}|^2\} \\ = & \dot{\mathbf{w}}_I^H \dot{\mathbf{W}}_g^H \Phi_{vv} \dot{\mathbf{W}}_g \dot{\mathbf{w}}_I = \left(\dot{\mathbf{q}}_I - \dot{\mathbf{B}}_I \dot{\mathbf{f}}_I\right)^H \dot{\mathbf{W}}_g^H \Phi_{vv} \dot{\mathbf{W}}_g \left(\dot{\mathbf{q}}_I - \dot{\mathbf{B}}_I \dot{\mathbf{f}}_I\right) \\ = & \dot{\mathbf{w}}_g^H \dot{\mathbf{W}}_I^H \Phi_{vv} \dot{\mathbf{W}}_I \dot{\mathbf{w}}_g = \left(\dot{\mathbf{q}}_g - \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g\right)^H \dot{\mathbf{W}}_I^H \Phi_{vv} \dot{\mathbf{W}}_I \left(\dot{\mathbf{q}}_g - \dot{\mathbf{B}}_g \dot{\mathbf{f}}_g\right) \end{split}$$

Derivation of the Iterative Version III

Alternately optimizing local and global filters

• Local filters update:

$$\dot{\mathbf{f}}_{g}^{(i+1)} = \left(\dot{\mathbf{B}}_{g}^{H}\left(\dot{\mathbf{W}}_{l}^{(i)}\right)^{H} \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_{l}^{(i)} \dot{\mathbf{B}}_{g}\right)^{-1} \cdot \dot{\mathbf{B}}_{g}^{H}\left(\dot{\mathbf{W}}_{l}^{(i)}\right)^{H} \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_{l}^{(i)} \dot{\mathbf{q}}_{g}.$$

• Global filters update:

$$\dot{\mathbf{f}}_{l}^{(i+1)} = \left(\dot{\mathbf{B}}_{l}^{H}\left(\dot{\mathbf{W}}_{g}^{(i)}\right)^{H} \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_{g}^{(i)} \dot{\mathbf{B}}_{l}\right)^{-1} \cdot \dot{\mathbf{B}}_{l}^{H}\left(\dot{\mathbf{W}}_{g}^{(i)}\right)^{H} \mathbf{\Phi}_{vv} \dot{\mathbf{W}}_{g}^{(i)} \dot{\mathbf{q}}_{l}.$$

Derivation of the Time-recursive Version

Stochastic Gradient Minimization

- Replace iteration index i with time index ℓ .
- Local step (applied at each node L_u times): $\dot{\mathbf{f}}_n(\ell+1) = \dot{\mathbf{f}}_n(\ell) + \frac{\mu}{\dot{\lambda}_{u,l}^n(\ell)} \left(\dot{w}_{g,n}^* \dot{\mathbf{u}}_n(\ell) \right) \dot{y}_g^*(\ell)$
- Global step (applied at the replica of each node L_u times): $\dot{\mathbf{f}}_g(\ell+1) = \dot{\mathbf{f}}_g(\ell) + \frac{\mu}{\dot{\lambda}_{u,g}(\ell)} \dot{\mathbf{u}}_g(\ell) \dot{y}_g^*(\ell).$
- Power normalization factors: $\dot{\lambda}_{u,l}^{n}(\ell)$, $\dot{\lambda}_{u,g}(\ell)$.

Algorithm summary (at the *n*th node)

Initialization

- Estimate local RTF **h**_n.
- Construct local FBF, $\dot{\mathbf{q}}_n$, and BM $\dot{\mathbf{B}}_n$.
- Estimate global RTF $\dot{\mathbf{h}}_g$.
- Construct global FBF, $\dot{\mathbf{q}}_g$, and BM $\dot{\mathbf{B}}_g$.

Perform repeatedly

- Broadcast output of local beamformer.
- Alternately, update local and global ANCs $\dot{\mathbf{f}}_n$ and $\dot{\mathbf{f}}_g$.
 - L_u local filter updates.
 - L_u global filter updates.
- Converges to the centralized TF-GSC.

Distributed LCMV

Formulation

- N nodes with M_n microphones.
- $\sum_{n=1}^{N} M_n = M.$
- $\mathbf{z} \triangleq \begin{bmatrix} \mathbf{z}_1^T \cdots \mathbf{z}_N^T \end{bmatrix}^T$.
- Closed-form LCMV necessitates the inversion of Φ_{zz}.
 A cumbersome task in distributed networks.

Naïve GSC Implementation

- Summation of local BFs: $y = \sum_{n=1}^{N} y_n$.
- Implement a local GSC at each node:
 - $M_n P$ outputs of the BM at the *n*th node (might go negative!).
 - Total number of BM outputs: $\sum_{n=1}^{N} (M_n P) = M (N \times P).$
 - $M (N \times P) < (M P) \Rightarrow$ degrees of freedom (DoF) lost
 - \Rightarrow incomplete minimization \Rightarrow performance degradation.

Distributed GSC

[Markovich-Golan et al., 2013]

Overview

- Introduce *P* shared signals:
 - Broadcast by a subset of the nodes.
 - Retrieve degrees of freedom.
- Extended inputs at each node:
 - Local microphones plus shared signals.
 - Purely local FBF, BM, ANC.
- DGSC adaptively converges to the centralized solution.



Total of N + P broadcast channels.

Nodes Connectivity

Sources "Owned" by the *n*th Node:

- A node *n* that receives the *p*th source with the highest SNR is declared its "owner".
- The shared signals broadcast by the *n*th node: $\mathbf{r}_n = \mathbf{D}_n^H \mathbf{z}_n$.
- \mathbf{D}_n : an $M_n \times P_n$ selection matrix.
- A shared signal (one component of \mathbf{r}_n) is responsible for only one source.
- Shared signals serve as a reference for RTF estimation in each node.

Extended Inputs at the *n*th Node

- $P P_n$ shared signals (excluding self-owned signals): $\dot{\mathbf{r}}_n$.
- Total number of signals: $\overline{M}_n = M_n + P P_n$.

• Signals:
$$\mathbf{\bar{z}}_n = \begin{bmatrix} \mathbf{z}_n^T & \dot{\mathbf{r}}_n^T \end{bmatrix}^T$$
.

Algorithm

DGSC at the *n*th Node

High Level Block-Diagram



Local & Global BF

- An $\overline{M}_n \times 1$ local GSC-BF at the *n*th node: $\overline{\mathbf{w}}_n$.
- Outputs of local GSC-BFs: $\bar{y}_n = \bar{\mathbf{w}}_n^H \bar{\mathbf{z}}_n$; $\forall n = 1, 2, \dots, N$.
- Global BF: $\mathbf{\bar{w}} \triangleq \begin{bmatrix} \mathbf{\bar{w}}_1^T \cdots \mathbf{\bar{w}}_N^T \end{bmatrix}^T$.

• Global output (available at each node): $\bar{y} = \sum_{n=1}^{N} \bar{y}_n$.

Blocks of the DGSC at the *n*th Node

Fixed Beamformer (Local)

- $\hat{\mathbf{H}}_n$: the RTF relating the extended inputs and the shared signals.
- Build local FBF $\mathbf{\bar{q}}_n$ using only local RTFs.

•
$$\mathbf{\bar{q}}_n \triangleq \frac{1}{N} \mathbf{\hat{H}}_n \left(\mathbf{\hat{H}}_n^H \mathbf{\hat{H}}_n \right)^{-1} \mathbf{g} \Rightarrow \mathbf{\bar{H}}_n^H \mathbf{\bar{q}}_n = \mathbf{g}.$$

Blocking Matrix (Block Diagonal)

•
$$\mathbf{\bar{B}}_n$$
: $\bar{M}_n \times (\bar{M}_n - P)$ BM.

• Noise references: $\mathbf{\bar{u}}_n = \mathbf{\bar{B}}_n^H \mathbf{\bar{z}}_n$

•
$$\sum_{n=1}^{N} (\bar{M}_n - P) = \sum_{n=1}^{N} (M_n - P_n) = M - P \Rightarrow \text{DoF fully utilized}$$
.

Adaptive Noise Canceler (Local)

- Least Mean Squares: $\overline{\mathbf{f}}_n(\ell) = \overline{\mathbf{f}}_n(\ell-1) + \mu \frac{\overline{\mathbf{u}}_n(\ell)\overline{y}^*(\ell)}{\overline{P}_{u,n}(\ell)}$.
- Power normalization $\bar{P}_{u,n}(\ell)$.

DGSC at the *n*th Node

Low Level Block-Diagram



DGSC Features

- Distributed processing for distributed constellation.
- It is shown [Markovich-Golan et al., 2013] that the distributed and centralized LCMV implementations identifies.
- Proof is based on: constraint set is a subspace of the *M*-dimensional linear space. Extending the linear space dimensions to \overline{M} does not alter the sub-space.
- Local input signals selection (quasi-) fixed:
 - Original inputs.
 - Shared signals selected by the system.
 - Hence RTF estimation valid until the acoustics changes.
- The DGSC sequentially converges to the centralized solution using local ANC updates.
- Different from the LC-DANSE [Bertrand and Moonen, 2012].

Algorithm

Important Practical Considerations

- Latency in the communication channel might require large buffering in each node.
- Owner selection is a cumbersome task if several speakers are concurrently active, since it is not clear how to identify each speaker.
- RTF can be very long for remote nodes.
- Number of nodes and constraints can dynamically change (see [Markovich-Golan et al., 2012b] for possible cure).

Scenario

- $4m \times 4m \times 3m$ room.
- Reverberation time $T_{60} = 300 \text{ms.}$
- N=4 nodes.
- $M_n = 2$ microphones $\forall n$.
- Desired and competing speaker with the same level.
- 2 point source Gaussian noises, 13dB lower than the speech signals.
- Sensors noise.
- 90 Monte-Carlo experiments (sources' positions).



Convergence



The convergence of the tested algorithms versus time.

S. Gannot (BIU) and A. Bertrand (KUL)

Distributed speech enhancement

EUSIPCO 2013 29 / 32

Speech Samples



(a) Noisy at mic. #1

(b) Single node GSC



S. Gannot (BIU) and A. Bertrand (KUL)

Distributed speech enhancement

Bibliography

References and Further Reading I



Bertrand, A. and Moonen, M. (2012).

Distributed node-specific LCMV beamforming in wireless sensor networks. *IEEE Transactions on Signal Processing*, 60:233–246.



Er, M. and Cantoni, A. (1983).

Derivative constraints for broad-band element space antenna array processors. IEEE Transactions on Acoustics, Speech and Signal Processing, 31(6):1378–1393.



Gannot, S., Burshtein, D., and Weinstein, E. (2001).

Signal enhancement using beamforming and nonstationarity with applications to speech. *IEEE Transactions on Signal Processing*, 49(8):1614–1626.



Griffiths, L. J. and Jim, C. W. (1982).

An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. on Antennas and Propagation*, 30(1):27–34.



Markovich-Golan, S., Gannot, S., and Cohen, I. (2012a). Distributed GSC beamforming using the relative transfer function. In *The European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania. Invited paper.



Markovich-Golan, S., Gannot, S., and Cohen, I. (2012b).

Low-complexity addition or removal of sensors/constraints in LCMV beamformers. *IEEE Transactions on Signal Processing*, 60(3):1205–1214.



Markovich-Golan, S., Gannot, S., and Cohen, I. (2013).

Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks. *IEEE Transactions on Audio, Speech, and Language Processing,* 21(2):343–356. Bibliography

References and Further Reading II



Shynk, J. (1992).

Frequency-domain and multirate and adaptive filtering. *IEEE Signal Processing Magazine*, 9(1):14–37.



Van Veen, B. D. and Buckley, K. M. (1988).

Beamforming: A versatile approach to spatial filtering. IEEE Acoustics, Speech and Signal Proc. magazine, pages 4–24.

Widrow, B., Jr., J. G., McCool, J., Kaunitz, J., Williams, C., Hearn, R., Zeider, J., Jr., E. D., and Goodlin, R. (1975). Adaptive noise cancelling: Principals and applications. *Proceeding of the IEEE*, 63(12):1692–1716.