Linear and Parametric Microphone Array Processing

Part III: Distributed Linear Spatial Processing

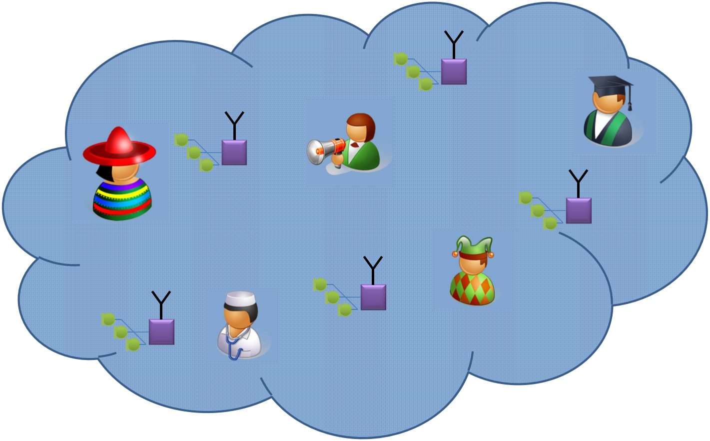
Emanuël A.P. Habets[1] and Sharon Gannot[2],

[1]International Audio Laboratories Erlangen, Germany

[2]Faculty of Engineering, Bar-Ilan University, Israel

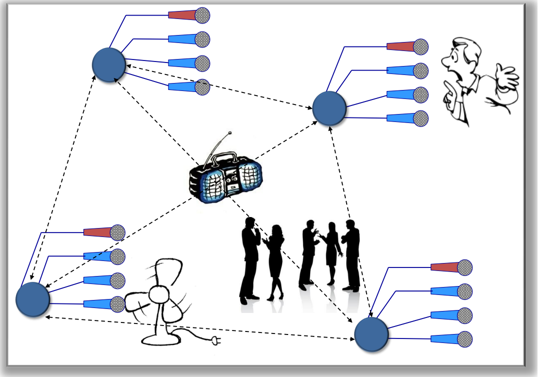
ICASSP 2013, Vancouver, Canada

Distributed Algorithms for Microphone Arrays



Outline

Wireless Acoustic Sensor Networks (WASNs)



Advantages

- Microphones can be placed randomly, avoiding tedious calibration.
- Using more microphones improves spatial resolution.
- High probability to find microphones close to a relevant sound source.
- Improved sound field sampling.

Challenges of Distributed Beamforming

Power

- Communication bandwidth
- Computational complexity

Arbitrary Constellation

- Ad hoc
- Dynamics
- Calibration

Communication

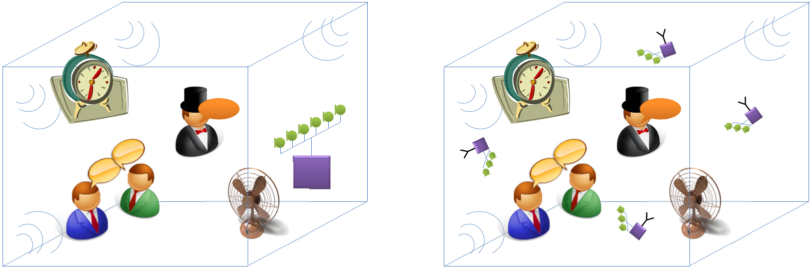
- Connectivity
- Protocol
- Capacity

Signal Processing

- Partial data
- Synchronization
- Dynamics
- Coherence

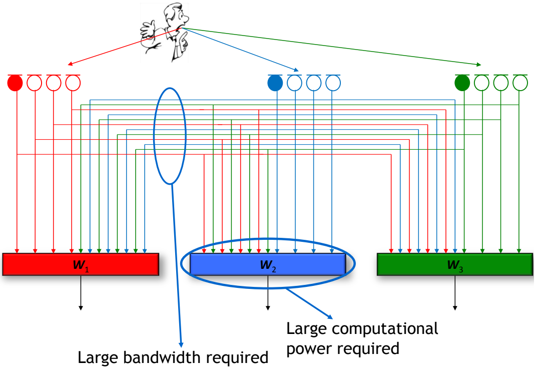
Goals

- Develop beamforming algorithms for distributed microphone constellation:
    - Ad hoc sensor networks.
    - Large volume (and many nodes).
- Robustness against randomly deployed microphones:
    - High fault percentage.
    - Arbitrary deployment of nodes.
- Applicability to Hearing Aids ([Doclo et al., 2009]; [Markovich-Golan et al., 2010]).



Centralized Network-Wide MWF (Fully Connected)

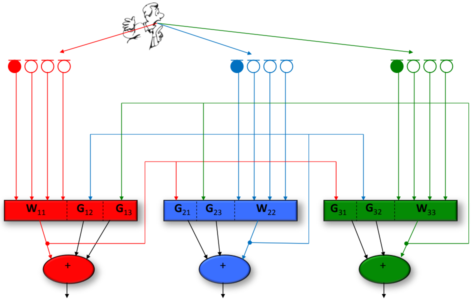
Courtesy of Dr. Alexander Bertrand, K.U.Leuven



- Each node has its own local reference microphone.
- Each node k solves network-wide MWF (here: 12-channel MWF).

Distributed Adaptive Node-specific Signal Estimation (DANSE)

Courtesy of Dr. Alexander Bertrand, K.U.Leuven



- Parameterized network-wide filter at node 1: $\mathbf{W}_1^T = \left[\mathbf{W}_{11} \; \mathbf{W}_{22}\mathbf{G}_{12} \; \ldots \; \mathbf{W}_{NN}\mathbf{G}_{1N} \right]$
- Node k adapts its filter coefficients $\mathbf{W}_{kk}$, $\mathbf{G}_{k1}, \ldots, \mathbf{G}_{kN}$ based on local MWF (here: 6-channel MWF).
- If single desired source: DANSE converges to centralized MWF

[Bertrand and Moonen, 2010a].

DANSE and Extensions

Courtesy of Dr. Alexander Bertrand, K.U.Leuven

- Can be generalized to Q desired sources [Bertrand and Moonen, 2010a].
- Small modification allows for simultaneous node updating [Bertrand and Moonen, 2010b].
- DANSE in networks with a tree topology (Tree-DANSE [Bertrand and Moonen, 2011]).
- LCMV-based DANSE (LC-DANSE [Bertrand and Moonen, 2012]).
- Robust DANSE (R-DANSE) for ill-conditioned scenarios (e.g., low-SNR nodes [Bertrand and Moonen, 2009]).
- Improved tracking using internal adaptive filters (this ICASSP [Szurley et al., 2013]).

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 $\mathbf{\Phi}_{zz}$.
  A cumbersome task in distributed networks.

Naïve GSC Implementation

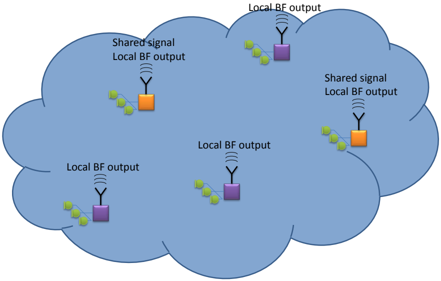
- Summation of local BFs: $y = \sum_{n=1}^{N} y_n$.
- Implement a local GSC at each node:
  - $M_n - P$ outputs of the BM at the nth node (might go negative!).
  - Total number of BM outputs: $\sum_{n=1}^{N}(M_n - P) = M - (N \times P)$.
  - $M - (N \times P) < (M - P) \Rightarrow$ degrees of freedom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

Distributed GSC

[Markovich-Golan et al., 2013a]

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 nth Node:

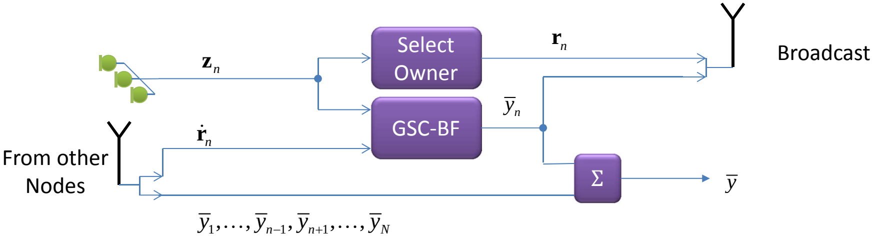
- A node n that receives the pth source with the highest SNR is declared its "owner".
- The shared signals broadcast by the nth node: $\mathbf{r}_n = \mathbf{D}_n^H \mathbf{z}_n$.
- $\mathbf{D}_n$: an $M_n \times P_n$ selection matrix.
- A shared signal (one component of $\mathbf{r}_n$) is responsible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

Extended Inputs at the nth Node

- $P - P_n$ shared signals (excluding self-owned signals): $\dot{\mathbf{r}}_n$.
- Total number of signals: $\bar{M}_n = M_n + P - P_n$.
- Signals: $\bar{\mathbf{z}}_n = \begin{bmatrix} \mathbf{z}_n^T & \dot{\mathbf{r}}_n^T \end{bmatrix}^T$.

DGSC at the *n*th Node

High Level Block-Diagram



Local & Global BF

- An $\bar{M}_n \times 1$ local GSC-BF at the nth node: $\bar{\mathbf{w}}_n$.
- Outputs of local GSC-BFs: $\bar{y}_n = \bar{\mathbf{w}}_n^H \bar{\mathbf{z}}_n; \ \forall n = 1, 2, \ldots, N$.
- Global BF: $\bar{\mathbf{w}} \triangleq \begin{bmatrix} \bar{\mathbf{w}}_1^T & \cdots & \bar{\mathbf{w}}_N^T \end{bmatrix}^T$.
- Global output (available at each node): $\bar{y} = \sum_{n=1}^N \bar{y}_n$.

Blocks of the DGSC at the nth Node

Fixed Beamformer (Local)

- $\hat{\mathbf{H}}_n$: the RTF relating the extended inputs and the shared signals.
- Build local FBF $\bar{\mathbf{q}}_n$ using only local RTFs.
- $\bar{\mathbf{q}}_n \triangleq \frac{1}{N}\hat{\mathbf{H}}_n \left(\hat{\mathbf{H}}_n^H \hat{\mathbf{H}}_n\right)^{-1} \mathbf{g} \Rightarrow \bar{\mathbf{H}}_n^H \bar{\mathbf{q}}_n = \mathbf{g}$.

Blocking Matrix (Block Diagonal)

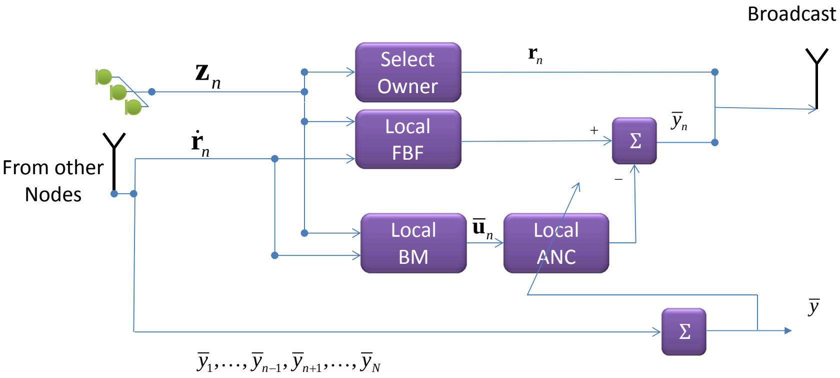
- $\bar{\mathbf{B}}_n$: $M_n \times (\bar{M}_n - P)$ BM.
- Noise references: $\bar{\mathbf{u}}_n = \bar{\mathbf{B}}_n^H \bar{\mathbf{z}}_n$
- $\sum_{n=1}^{N}(\bar{M}_n - P) = \sum_{n=1}^{N}(M_n - P_n) = M - P \Rightarrow$ DoF fully utilized.

Adaptive Noise Canceler (Local)

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

DGSC at the *n*th Node

Low Level Block-Diagram



DGSC Summary I

Features

- Distributed processing for distributed constellation.
- It is shown [Markovich-Golan et al., 2013a] 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 $\bar{M}$ 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.

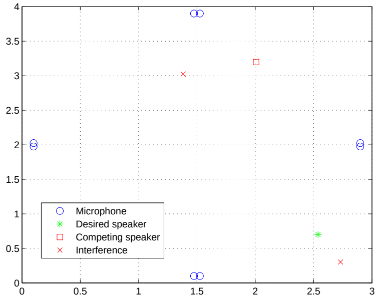
DGSC Summary II

Important Practical Considerations

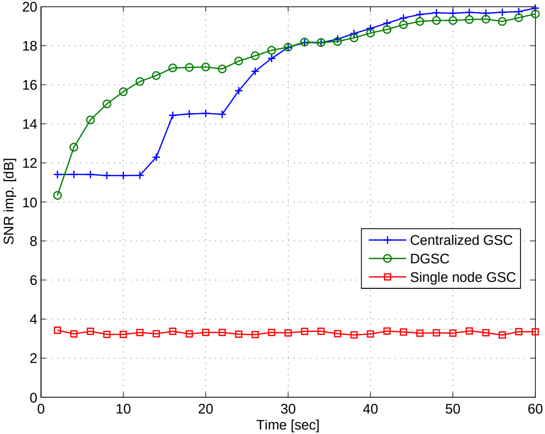
- Latency in the communication channel might require large buffering in each node.
- Owner selection is a cumbersome task if several speakers are concurrently active, since it is not clear how to identify each speaker.
- RTF can be very long for remote nodes.
- Number of nodes and constraints can dynamically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

Scenario

- $4\text{m} \times 4\text{m} \times 3\text{m}$ 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).

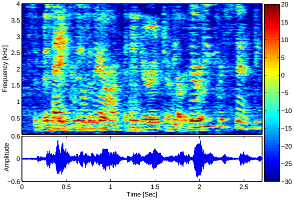


Convergence

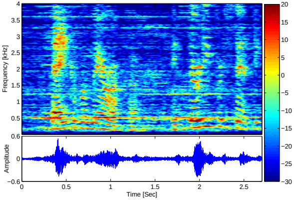


The convergence of the tested algorithms versus time.

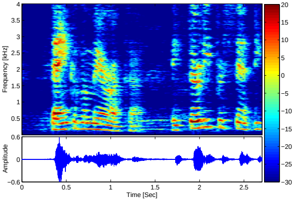
Speech Samples



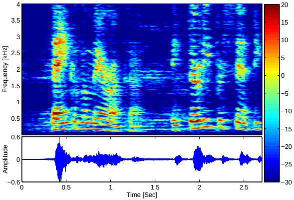
(a) Noisy at mic. #1



(b) Single node GSC



(c) Centralized GSC

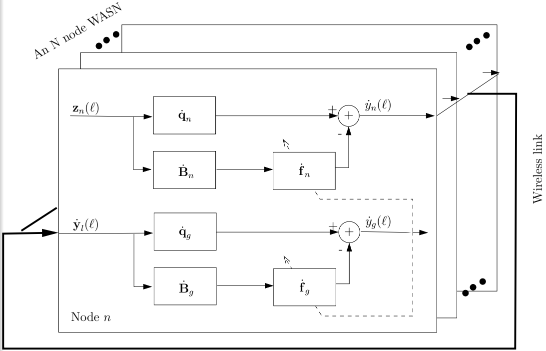


(d) Distributed GSC

Distributed single constraint GSC (DS-GSC)

[Markovich-Golan et al., 2012b]

- Two stage filtering:
  - Local filtering
  - Global filtering
- N transmission channels
- Alternating local and global filter updates
  - Iterative version
  - Time-recursive version
- Converges to the centralized TF-GSC



Blind Sampling Rate Offset Estimation and Compensation

[Markovich-Golan et al., 2012a]

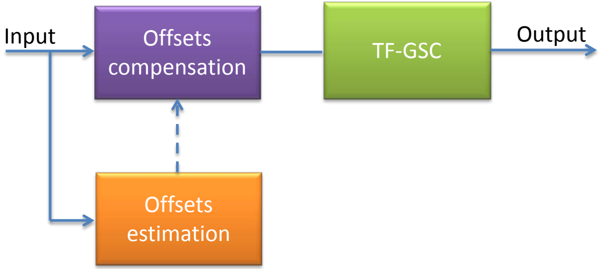
Scenario

- Fully connected N nodes network with M_n microphones at the nth node.
- Nominal sampling rate f_s.
- Sampling rate $f_{s,n} = (1 + \epsilon_n) f_s$, sampling period $T_{s,n}$ with Sampling rate offset ϵ_n.

TF-GSC [Gannot et al., 2001] with Sampling Rate Offsets

- RTF is constantly changing: signal distortion.
- ANC is constantly updating: increased noise level.
- Microphone signals are less coherent: degraded performance.

Block diagram of synchronized TF-GSC



Synchronized TF-GSC

- Sampling rate estimation: based on the phase drift of the coherence between microphones in stationary noise-only segments (in coherent frequency bands).
- Resampling with Lagrange polynomials interpolation.
- Other beamforming sync. methods: [Wehr et al., 2004]; [Ono et al., 2009]

Results

TF-GSC Algorithms

W.o. offsets; Conventional TF-GSC; Synchronized TF-GSC

Criteria

Signal to Distortion ratio (SDR); Signal to Noise (SNR)

	Without offset		With offset			
	Conventional		Conventional		Synchronized	
Q	SDR	SNR	Ex. Dist.	Ex. Noise	Ex. Dist.	Ex. Noise
1	15.0	34.3	11.2	7.7	0.0	0.0
2	14.9	27.5	11.2	4.9	0.1	0.0
3	14.6	24.5	11.5	3.4	0.4	0.1
4	14.7	23.5	11.9	2.9	0.8	0.2

Values in dB, Ex. stands for excess values

WASNs with Random Node Deployment

[Markovich-Golan et al., 2011]; [Markovich-Golan et al., 2013b]; general reading [Lo, 1964]

Scenarios

- Ad hoc sensor networks.
- Large volume (and many nodes).
- High fault percentage.
- Arbitrary deployment of nodes.

Questions

- How many nodes are required?
- What is the expected performance?
- Is there an optimal deployment?

[Kodrasi et al., 2011]

Random Beampattern using WASN

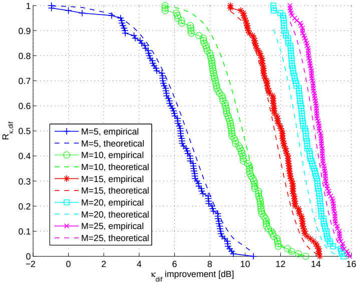
Scenario

- Desired speaker.
- Coherent or diffused noise fields.
- Microphones are randomly positioned.
- Reverberant enclosure.

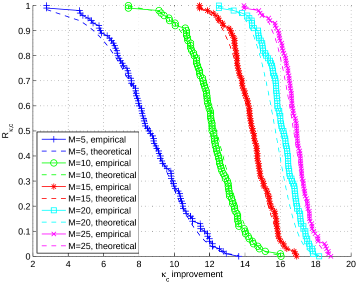
Derivation

- ATFs are complex random variables.
- Derive a model for ATFs (based on the work of Schröder [Schroeder, 1962]; [Schroeder, 1987]).
- Consider two BF performance criteria: SNR, White noise gain
- These criteria become random variables.
  - Analyze the statistics.
  - Derive reliability functions: The probability that the criteria exceed a pre-defined level.

Reliability I



(a) Diffuse Noise, SIR=30dB, residual interference dominant



(b) Coherent Noise, SIR=0dB, residual noise dominnat

$\mathrm{SINR}_{\mathrm{out}} - \mathrm{SNR}_{\mathrm{in}}$, $T_{60} = 0.4$sec, Room dimensions $4 \times 4 \times 3$m.

Special Thanks

1. Shmulik Markovich-Golan
2. Dr. Alexander Bertrand
3. Prof. Simon Doclo
4. Prof. Walter Kellermann

References and Further Reading I

Bertrand, A. and Moonen, M. (2009).
Robust distributed noise reduction in hearing aids with external acoustic sensor nodes.
EURASIP Journal on Advances in Signal Processing.

Bertrand, A. and Moonen, M. (2010a).
Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: sequential node updating.
IEEE Transactions on Signal Processing, 58:5277–5291.

Bertrand, A. and Moonen, M. (2010b).
Distributed adaptive node-specific signal estimation in fully connected sensor networks – part II: simultaneous & asynchronous node updating.
IEEE Transactions on Signal Processing, 58:5292–5306.

Bertrand, A. and Moonen, M. (2011).
Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology.
IEEE Transactions on Signal Processing, 59(5):2196–2210.

Bertrand, A. and Moonen, M. (2012).
Distributed node-specific LCMV beamforming in wireless sensor networks.
IEEE Transactions on Signal Processing, 60:233–246.

Doclo, S., Moonen, M., Van den Bogaert, T., and Wouters, J. (2009).
Reduced-bandwidth and distributed mwf-based noise reduction algorithms for binaural hearing aids.
IEEE Transactions on Audio, Speech, and Language Processing, 17(1):38–51.

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.

References and Further Reading II

Kodrasi, I., Rohdenburg, T., and Doclo, S. (2011).
Microphone position optimization for planar superdirective beamforming.
In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 109–112, Prague, Czech Republic.

Lo, Y. (1964).
A mathematical theory of antenna arrays with randomly spaced elements.
IEEE Transactions on Antennas and Propagation, 12(3):257–268.

Markovich-Golan, S., Gannot, S., and Cohen, I. (2010).
A reduced bandwidth binaural MVDR beamformer.
In *The International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel-Aviv, Israel.
Best student paper award.

Markovich-Golan, S., Gannot, S., and Cohen, I. (2011).
Performance analysis of a randomly spaced wireless microphone array.
In *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 121–124, Prague, Czech Republic.

Markovich-Golan, S., Gannot, S., and Cohen, I. (2012a).
Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming.
In *The International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany.
Final list for best student paper award.

References and Further Reading III

Markovich-Golan, S., Gannot, S., and Cohen, I. (2012b).
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. (2012c).
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. (2013a).
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.

Markovich-Golan, S., Gannot, S., and Cohen, I. (2013b).
Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure.
IEEE Trans. Audio, Speech and Language Processing.
Accepted for publication.

Ono, N., Kohno, H., Ito, N., and Sagayama, S. (2009).
Blind alignment of asynchronously recorded signals for distributed microphone array.
In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 161–164, New Paltz, NY, USA.

Schroeder, M. R. (1962).
Frequency-correlation functions of frequency responses in rooms.
The Journal of the Acoustical Society of America, 34:1819.

References and Further Reading IV

Schroeder, M. R. (1987).
Statistical parameters of the frequency response curves of large rooms.
Journal of the Audio Engineering Society, 35(5):299–306.

Szurley, J., Bertrand, A., and Moonen, M. (2013).
Improved tracking performance for distributed node-specific signal enhancement in wireless acoustic sensor networks.
In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada.

Wehr, S., Kozintsev, I., Lienhart, R., and Kellermann, W. (2004).
Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation.
In *IEEE Sixth International Symposium on Multimedia Software Engineering*, pages 18–25.