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

Part IV: Random Microphone Deployment

Sharon Gannot[1] and Alexander Bertrand[2]

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

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

EUSIPCO 2013, Marrakesh, Morocco

Special Thanks

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

Blind Sampling Rate Offset Estimation and Compensation

[Markovich-Golan et al., 2012]

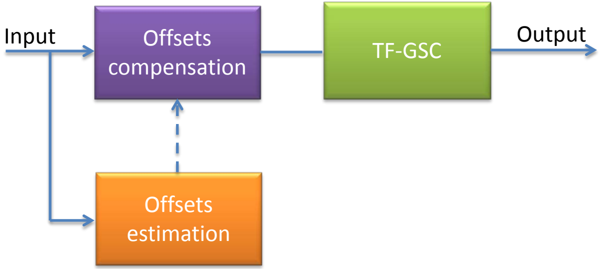
Scenario

- 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

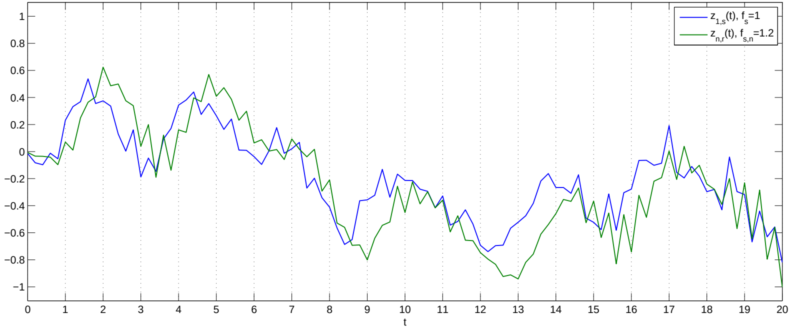


Synchronized TF-GSC

- 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 [Erup et al., 1993].
- Other beamforming sync. methods: [Wehr et al., 2004]; [Ono et al., 2009]; [Liu, 2008].

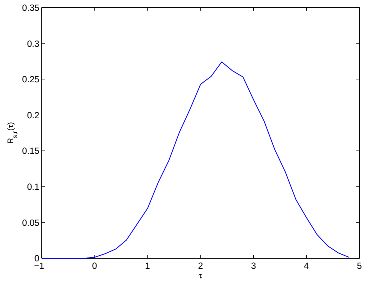
Continuous Microphone Signals

- Received noise component at microphone s, 1st node: $v_{1,s}(t)$.
- Received noise component at microphone r, nth node: $v_{n,r}(t)$.
- Noise only time-segment.

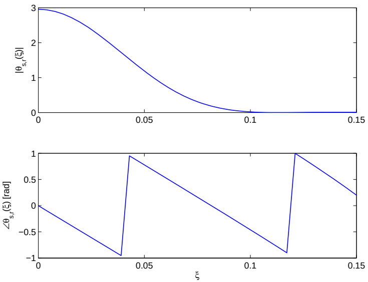


Statistics of Noise Components $v_{1,s}(t)$ and $v_{n,r}(t)$

- Cross-covariance: $R_{s,r}(\tau) = \mathrm{E}\{v_{1,s}(t)\, v_{n,r}(t-\tau)\}$.
- Cross-spectrum: $\theta_{s,r}(\xi) = \int_{-\infty}^{\infty} R_{s,r}(\tau) \exp(-\mathrm{j}\xi\tau)\, \mathrm{d}\tau$.



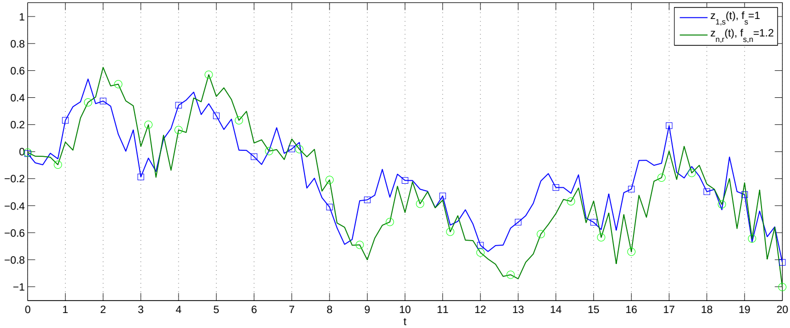
Cross-covariance



Cross-spectrum

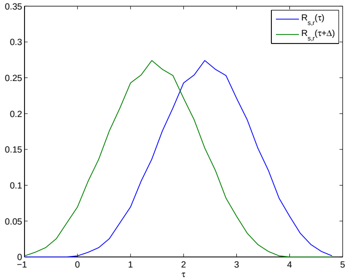
Sampled Microphone Signals

- $v_{1,s}[\ell] = v_{1,s}(\ell T_s)$.
- $v_{n,r}[\ell] = v_{n,r}(\ell T_{s,n})$.

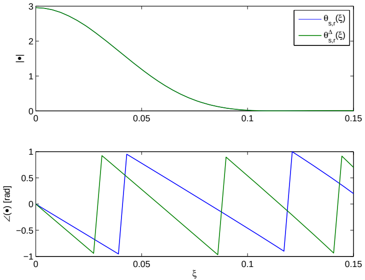


Statistics of Microphones $v_{1,s}(t)$ and $v_{n,r}(t-\Delta)$

- Cross-covariance: $R_{s,r}(\tau+\Delta)$.
- Cross-spectrum: $\theta^{\Delta}_{s,r}(\xi) = \exp(\mathrm{j}\xi\Delta)\,\theta_{s,r}(\xi)$.
- Time difference at the ℓth sample: $\Delta = \ell T_s - \ell T_{s,n} \approx \ell T_s \epsilon_n$ (using first-order Taylor series approximation).



Cross-covariance



Cross-spectrum

Statistics of Sampled Microphones $v_{1,s}[\ell]$ and $v_{n,r}[\ell]$

Cross-spectrum is band-limited by $\frac{f_s}{2}$

- Assume small offset.
- $\theta^{\ell}_{s,r}[k] = \theta^{\ell T_s \epsilon_n}_{s,r}\left(\frac{2\pi k f_s}{K}\right)$.
- Let, $\theta_{s,r}[k] = \theta_{s,r}\left(\frac{2\pi k f_s}{K}\right)$, then:

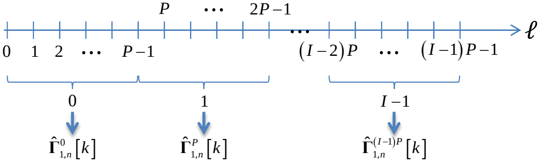
$$\theta^{\ell}_{s,r}[k] = \exp\left(\mathrm{j}\frac{2\pi k \ell \epsilon_n}{K}\right)\theta_{s,r}[k]$$

Coherence between microphones s and r at the ℓth sample

- Define $\gamma^{\ell}_{s,r}[k] = \frac{\theta^{\ell}_{s,r}[k]}{\sqrt{\theta_{s,s}[k]\theta_{r,r}[k]}}$ and $\gamma_{s,r}[k] = \frac{\theta_{s,r}[k]}{\sqrt{\theta_{s,s}[k]\theta_{r,r}[k]}}$; for $k = 0, 1, \ldots, K-1$.
- Then $\gamma^{\ell}_{s,r}[k] = \alpha^{\ell}_n \gamma_{s,r}[k]$ with $\alpha_n = \exp\left(\mathrm{j}\frac{2\pi k \epsilon_n}{K}\right)$.
- ϵ_n can be extracted from α_n.

Offset Estimation at the *n*th Node

- Given a noise-only time segment of P_s samples.
- Partition into I frames of P samples: $\ell = i \times P;\ i = 0, 1, \ldots, I-1$.
- estimate $M_1 \times M_n$ coherence matrix $\hat{\boldsymbol{\Gamma}}_{1,n}^{iP}[k]$ between microphones of the 1st and the *n*th nodes.



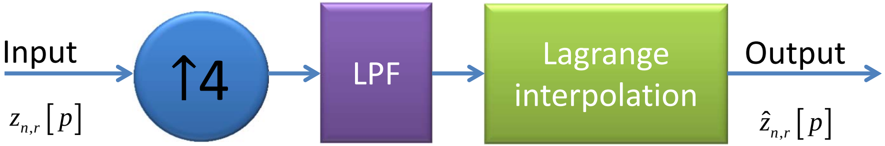
- $|\epsilon_n| < \epsilon_{\max} \Rightarrow$ no 2π ambiguity for $k \leq k_{\max} = \frac{K}{2P\epsilon_{\max}}$.
- Estimate the *n*th node sampling rate offset:
  - s, r pair: $\hat{\epsilon}_{n,s,r} = \text{avg}_k \left(\frac{K}{2\pi P k} \angle \underbrace{\underset{i}{\text{avg}}\, \hat{\gamma}_{s,r}^{iP}[k] / \gamma_{s,r}^{(i-1)P}[k]}_{\hat{\alpha}_n^P} \right)$.
  - Average all microphone pairs: $\hat{\epsilon}_n = \frac{1}{M_1 M_n} \sum_{s=1}^{M_1} \sum_{r=1}^{M_n} \hat{\epsilon}_{n,s,r}$.

Resampling with Lagrange Polynomials Interpolation

[Pawig et al., 2010],[Erup et al., 1993]

Resample $z_{n,r}(pT_{s,n})$ to $z_{n,r}(pT_s)$

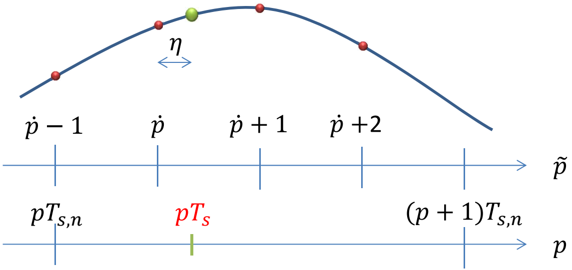
- Interpolate $z_{n,r}[p]$ by factor 4: $\tilde{z}_{n,r}[\tilde{p}]$.
- Denote: $\dot{p} = \lfloor \frac{4pT_s}{T_{s,n}} \rfloor = \lfloor 4p(1+\hat{\epsilon}_n) \rfloor$, the closest interpolated sample index from the left to time pT_s.
- The resampled value of $z_{n,r}(pT_s)$ is $\hat{z}_{n,r}[p] = \beta^p_{-1}\tilde{z}_{n,r}[\dot{p}-1] + \beta^p_0\tilde{z}_{n,r}[\dot{p}] + \beta^p_1\tilde{z}_{n,r}[\dot{p}+1] + \beta^p_2\tilde{z}_{n,r}[\dot{p}+2]$.



Resampling with Lagrange Polynomials Interpolation [Pawig et al., 2010],[Erup et al., 1993] II

Calculate Weights

- $\eta = 4p\left(1 + \hat{\epsilon}_n\right) - \dot{p}$.
- $\beta^p_{-1} = -\frac{\eta(\eta-1)(\eta-2)}{6}$.
- $\beta^p_0 = \frac{(\eta+1)(\eta-1)(\eta-2)}{2}$.
- $\beta^p_1 = -\frac{(\eta+1)\eta(\eta-2)}{2}$.
- $\beta^p_2 = \frac{(\eta+1)\eta(\eta-1)}{6}$.



Experimental Study

Q directional stationary interfering sources

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. - excess values

WASNs with Random Node Deployment

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

Scenarios

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

Questions

- How many microphones are required?
- What is the expected performance?
- Is there an optimal deployment? [Kodrasi et al., 2011]

Outline

- Array of randomly located microphones in a reverberant enclosure.
- Single desired speaker.
- Utilizing the statistical model of the ATFs, statistical models for the SIR and WNG are derived.
- The reliability of the SDW-MWF is computed for:
  - Multiple coherent noise sources.
  - Diffuse sound field.
- The reliability functions can be used to determine the number of microphones required to assure a desired performance level (with a controlled level of uncertainty).

Notations I

Signals

- Let $s_d(\ell, k)$ be a desired speaker signal located at $\mathbf{r}_d$.
- Microphone signals:

$$\mathbf{z}(\ell, k) \triangleq \mathbf{h}_d(\ell, k)\, s_d(\ell, k) + \mathbf{v}(\ell, k).$$

- Microphone signals PSD:

$$\mathbf{\Phi}_{zz}(\ell, k) \triangleq \mathrm{E}\{\mathbf{z}(\ell, k)\mathbf{z}^H(\ell, k)\} = \sigma_d^2(\ell, k)\mathbf{h}_d(\ell, k)\mathbf{h}_d^H(\ell, k) + \mathbf{\Phi}_{vv}(\ell, k)$$

- Noise PSD:

$$\mathbf{\Phi}_{vv}(\ell, k) \triangleq \mathrm{E}\{\mathbf{v}(\ell, k)\mathbf{v}^H(\ell, k)\}.$$

Notations II

Room Constellation

- Room volume and surface area:

$$V \triangleq D_x \times D_y \times D_z$$
$$A \triangleq 2\left(D_x \times D_y + D_x \times D_z + D_y \times D_z\right)$$

- Reverberation time: T_{60}.
- M microphones randomly deployed with a uniform distribution at coordinates $\mathbf{r}^m \triangleq \left[\, r_x^m \;\; r_y^m \;\; r_z^m \,\right]^T$; $m = 1, \ldots, M$.

Criterion

SDW-MWF

$$\mathbf{w} \triangleq \underset{\mathbf{w}'}{\operatorname{argmin}} \left|1 - \left(\left(\mathbf{w}'\right)^H \mathbf{h}_d\right)\right|^2 \sigma_d^2 + \mu \left(\mathbf{w}'\right)^H \mathbf{\Phi}_{vv} \mathbf{w}' = \frac{\mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d}{\mathbf{h}_d^H \mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d + \frac{\mu}{\sigma_d^2}}$$

SINR and WNG are Random Variables

Signal to Interference and Noise (SINR):

$$\kappa \triangleq \frac{\sigma_d^2 |\mathbf{w}^H \mathbf{h}_d|^2}{\mathbf{w}^H \mathbf{\Phi}_{vv} \mathbf{w}} = \sigma_d^2 \mathbf{h}_d^H \mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d$$

White noise gain (WNG):

$$\xi \triangleq \frac{|\mathbf{w}^H \mathbf{h}_d|^2}{\|\mathbf{w}\|^2} = \frac{\left(\mathbf{h}_d^H \mathbf{\Phi}_{vv}^{-1} \mathbf{h}_d\right)^2}{\mathbf{h}_d^H \mathbf{\Phi}_{vv}^{-2} \mathbf{h}_d}$$

Statistical ATF Modelling

ATF relating a coherent source at $\mathbf{r}_d$, and the mth microphone at $\mathbf{r}^m$

$$h \triangleq \bar{h} + \hat{h}$$

- $\bar{h}$ the direct arrival.
- $\hat{h}$ the reverberant component.
- The direct arrival and the reverberant tail assumed uncorrelated.

Reverberant Tail Model [Schroeder, 1987],[Kuttruff, 2000]

Under the Assumptions:

- The signal wavelength is much smaller than the room dimensions.
- The microphones and sources are at least half wavelength away from the walls.
- The signal frequency is above the Schroeder frequency, $f_{\mathrm{Schroeder}} \triangleq 2000\sqrt{\frac{T_{60}}{V}}$ (typically few hundred Hz).

The Tail Statistics:

$$\hat{h} \sim \mathcal{CN}\left(0, \hat{\alpha}\right)$$

with $\hat{\alpha} \triangleq \frac{1-\varepsilon}{\pi\varepsilon A}$ and $\varepsilon \triangleq \frac{0.161V}{AT_{60}}$, the exponential decay rate of the RIR tail.

The Direct Arrival (Spherical Wave Propagation)

The Direct Arrival Model

$$\bar{h} \triangleq \bar{a} \exp\left(j\bar{\phi}\right)$$

where:

$$\bar{a} = \begin{cases} 1 & ; \mathbf{r}_d \leq \frac{1}{4\pi} \\ \frac{1}{4\pi\|\mathbf{r}_d - \mathbf{r}^m\|} & ; \frac{1}{4\pi} < \mathbf{r}_d \end{cases}$$

$$\bar{\phi} = \frac{2\pi\|\mathbf{r}_d - \mathbf{r}^m\|}{\lambda_k}$$

Second-Order Statistics of single ATF

Arbitrary Sensor Location within $\bar{r}$

Under the Assumptions:

- For $\bar{r} \gg r_c$, where $r_c \triangleq \sqrt{\frac{V}{100\pi T_{60}}}$ is the critical distance, the direct path is negligible [Kuttruff, 2000].
- For $\bar{r} \gg \lambda_k$ multiple 2π phase cycles are repeated while sound wave propagates.
- $\bar{r}$ arbitrarily chosen (the results are not sensitive to the exact value).

Approximations:

- $\mathrm{E}\left\{\bar{h}\right\} \approx 0 \Rightarrow \mathrm{E}\left\{h\right\} = 0$.
- $\mathrm{E}\left\{|h|^2\right\} \triangleq \alpha = \frac{4\pi\bar{r}^3}{3V}\bar{\alpha} + \hat{\alpha}$ with $\bar{\alpha} = \frac{6\pi\bar{r}-1}{32\pi^3\bar{r}^3}$.
- The ATFs $h_m;\ m = 1,\ dots, M$ relating the source and randomly deployed microphones are i.i.d. (for large sphere and few microphones).

Covariance of ATFs Relating 2 Sources and Randomly Located Microphone

ATFs covariance:

$$\mathrm{E}\left\{h_1 h_2^*\right\} = \mathrm{E}\left\{\bar{h}_1 \bar{h}_2^*\right\} + \mathrm{E}\left\{\hat{h}_1 \hat{h}_2^*\right\}$$

- Reverberant tail is diffused [Jacobsen and Roisin, 2000]:

$$\mathrm{E}\left\{\hat{h}_1 \hat{h}_2^*\right\} = \hat{\alpha}\,\mathrm{sinc}\left(\frac{2\pi\|\mathbf{r}_1 - \mathbf{r}_2\|}{\lambda_k}\right)$$

- Assuming $\|\mathbf{r}_1 - \mathbf{r}_2\| \gg \lambda_k$:
  - $\mathrm{E}\left\{\hat{h}_1 \hat{h}_2^*\right\} \approx 0$, since the sinc is decaying.
  - $\mathrm{E}\left\{\bar{h}_1 \bar{h}_2^*\right\} \approx 0$, since multiple 2π phase cycles are repeated while sound wave propagates.

Reliability Measures

SIR Reliability

The reliability of an SIR level of κ_0 is defined as the probability that the output SIR will exceed κ_0:

$$R_\kappa(\kappa_0) \triangleq \Pr(\kappa \geq \kappa_0).$$

White Noise Gain (WNG) Reliability

The reliability of a WNG level of ξ_0 is defined as the probability that the WNG will exceed ξ_0:

$$R_\xi(\xi_0) \triangleq \Pr(\xi \geq \xi_0).$$

P Directional Noise Sources

For high INR and $P \ll M$

$$R_{\kappa,c}(\kappa_0) = 1 - F_{\eta,c}\left(\frac{2}{\alpha}\frac{\sigma_u^2}{\sigma_d^2}\kappa_0\right)$$
$$R_{\xi,c}(\xi_0) = 1 - F_{\eta,c}\left(\frac{2}{\alpha}\xi_0\right).$$

- σ_u^2 - sensor noise variance and σ_d^2 - desired source varaince.
- $\eta_c \sim \chi^2(2(M-P))$ Chi-square RV with $2(M-P)$ degrees of freedom.
- $F_{\eta,c}(\eta_0) \triangleq \Pr(\eta_c \leq \eta_0) = \frac{\gamma_f\left(M-P,\frac{\eta_0}{2}\right)}{\Gamma_f(M-P)}$ is the respective CDF.
- Γ_f is the Gamma function.
- γ_f is the lower incomplete Gamma function.

Diffused Noise Source

Noise Field

$$\Phi_{vv}\left(m, m'\right) = \sigma_{\text{dif}}^2 \operatorname{sinc}\left(\frac{2\pi \|\mathbf{r}_m - \mathbf{r}_{m'}\|}{\lambda_k}\right) \approx \sigma_{\text{dif}}^2 \mathbf{I}.$$

- σ_{dif}^2 variance of the diffuse field.
- For enclosures larger than λ_k.

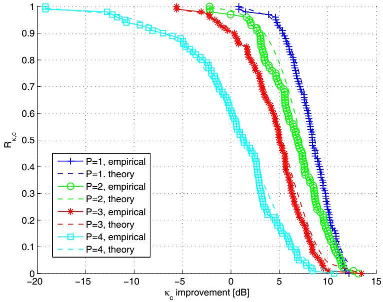
Reliability

$$R_{\kappa,\text{dif}}\left(\kappa_0\right) = 1 - F_{\eta,\text{dif}}\left(\frac{2}{\alpha}\frac{\sigma_{\text{dif}}^2}{\sigma_d^2}\kappa_0\right)$$

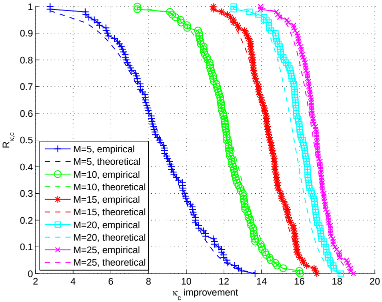
$$R_{\xi,\text{dif}}\left(\xi_0\right) = 1 - F_{\eta,\text{dif}}\left(\frac{2}{\alpha}\xi_0\right)$$

- $\eta_{\text{dif}} \sim \chi^2\left(2M\right)$ Chi-square RV with $2M$ degrees of freedom.
- $F_{\eta,\text{dif}}\left(\eta_0\right) \triangleq \Pr\left(\eta_{\text{dif}} \leq \eta_0\right) = \frac{\gamma_f\left(M, \frac{\eta_0}{2}\right)}{\Gamma_f(M)}$ is the respective CDF.

SIR Reliability



(a) $M = 5$



(b) $P = 1$

$\mathrm{SINR}_{\mathrm{out}} - \mathrm{SNR}_{\mathrm{in}}$ for coherent noise sources. $T_{60} = 0.4$sec, room dimensions $4 \times 4 \times 3$m. Similar trends for diffused noise field.

References and Further Reading I

Erup, L., Gardner, F., and Harris, R. (1993).
Interpolation in digital modems, II- implementation and performance.
IEEE Transactions on Communications, 41(6):998–1008.

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.

Jacobsen, F. and Roisin, T. (2000).
The cohrerence of reverberant sound fields.
Journal of the Acoustical Society of America, 108(1):204–210.

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.

Kuttruff, H. (2000).
Room acoustics.
Taylor & Francis.

Liu, Z. (2008).
Sound source separation with distributed microphone arrays in the presence of clock synchronization errors.
In *The International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, WA, USA.

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

References and Further Reading II

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. (2012).
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.

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

Pawig, M., Enzner, G., and Vary, P. (2010).
Adaptive sampling rate correction for acoustic echo control in voice-over-ip.
IEEE Transactions on Signal Processing, 58(1):189 –199.

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

References and Further Reading III

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.