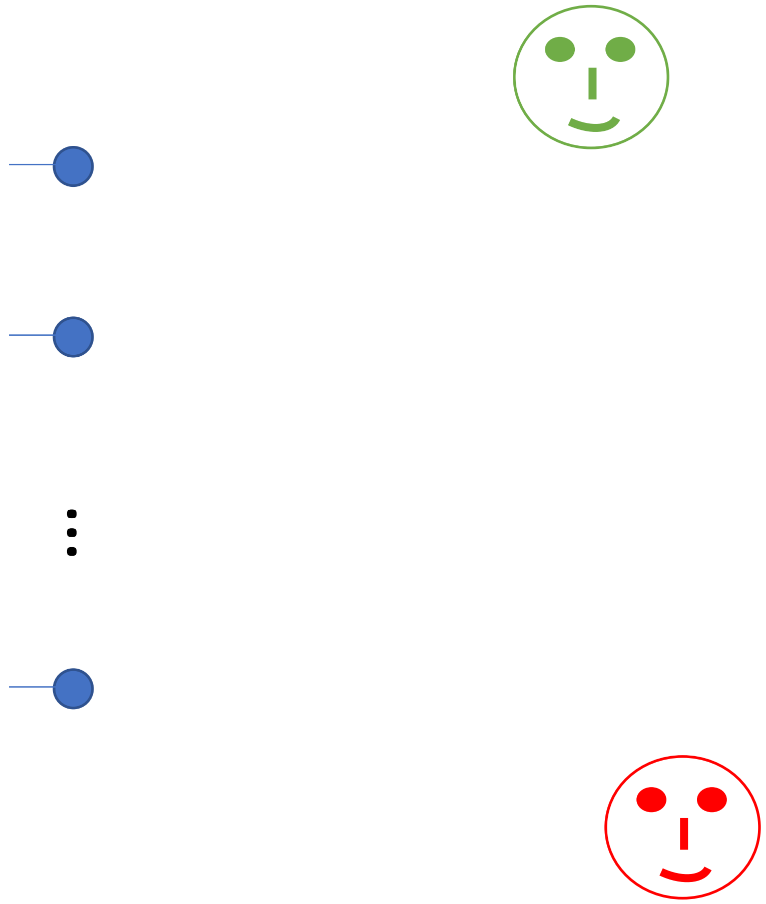
Multi-Microphone Noise Reduction and Source Separation

Andreas Koutrouvelis

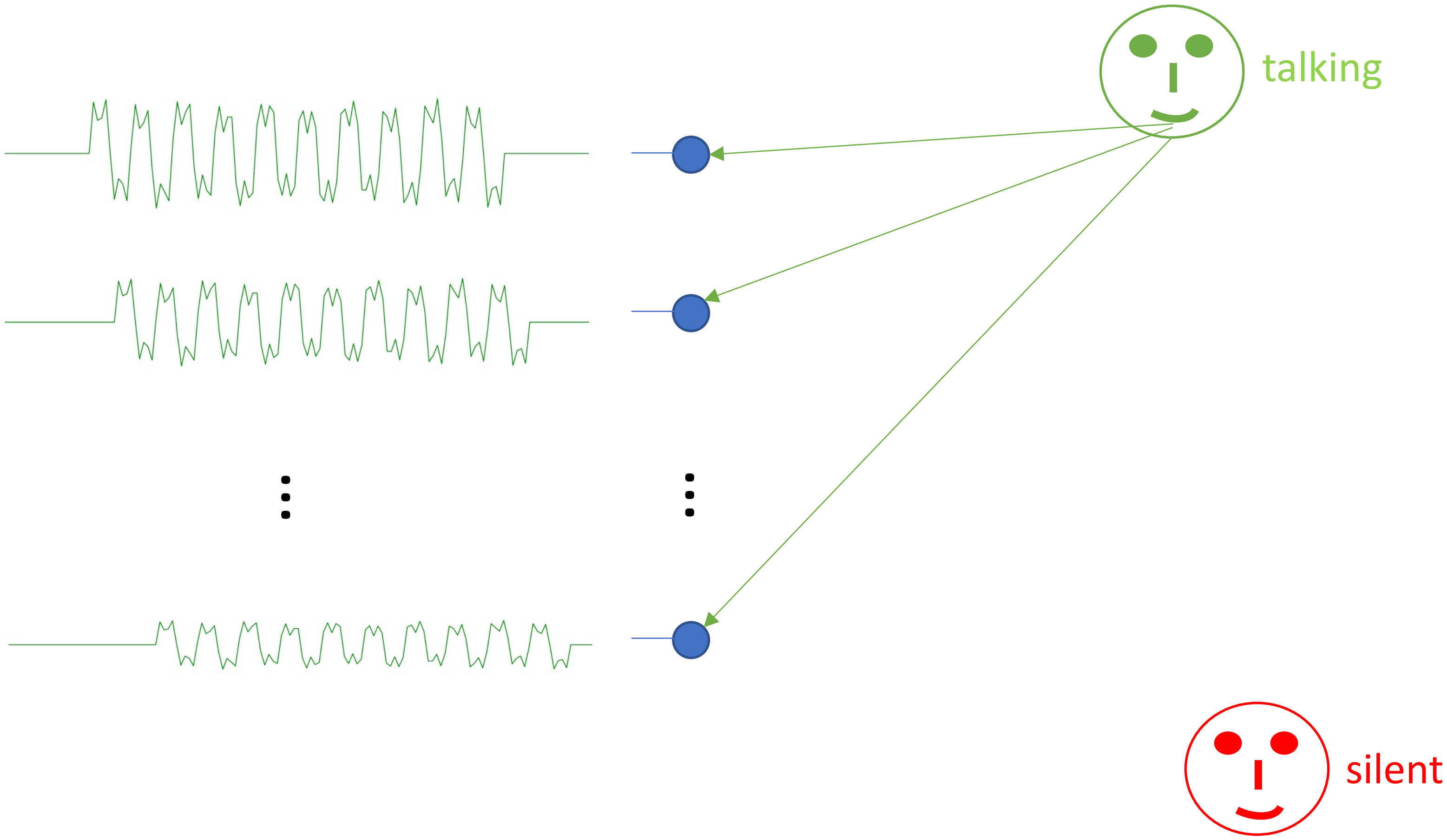
Outline

- Introduction
- Multi-microphone signal model
- Multi-microphone noise reduction
- Multi-microphone source separation
- Estimation of multi-microphone signal model parameters
- Experiments

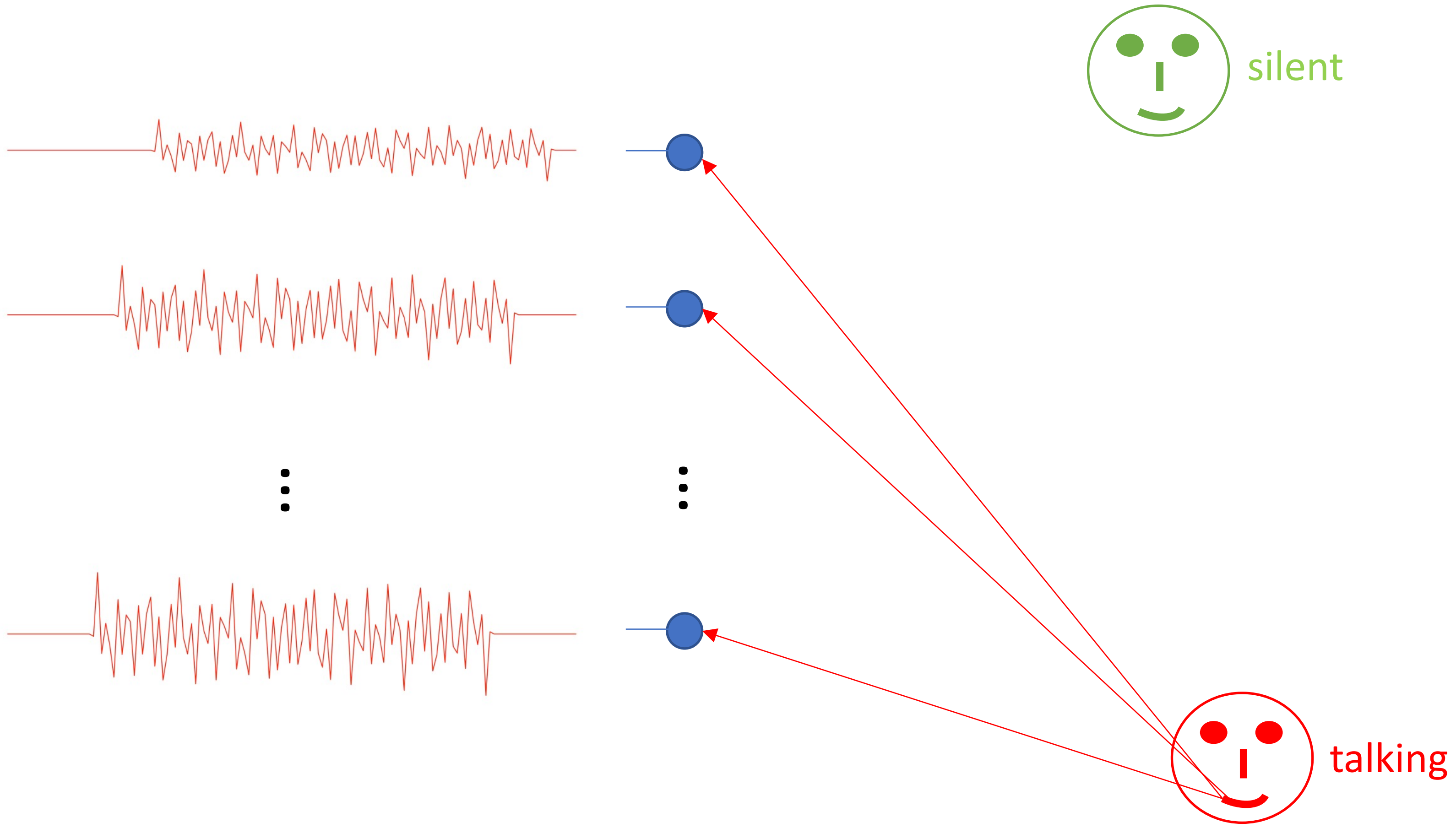
Introduction: a simple acoustic scene



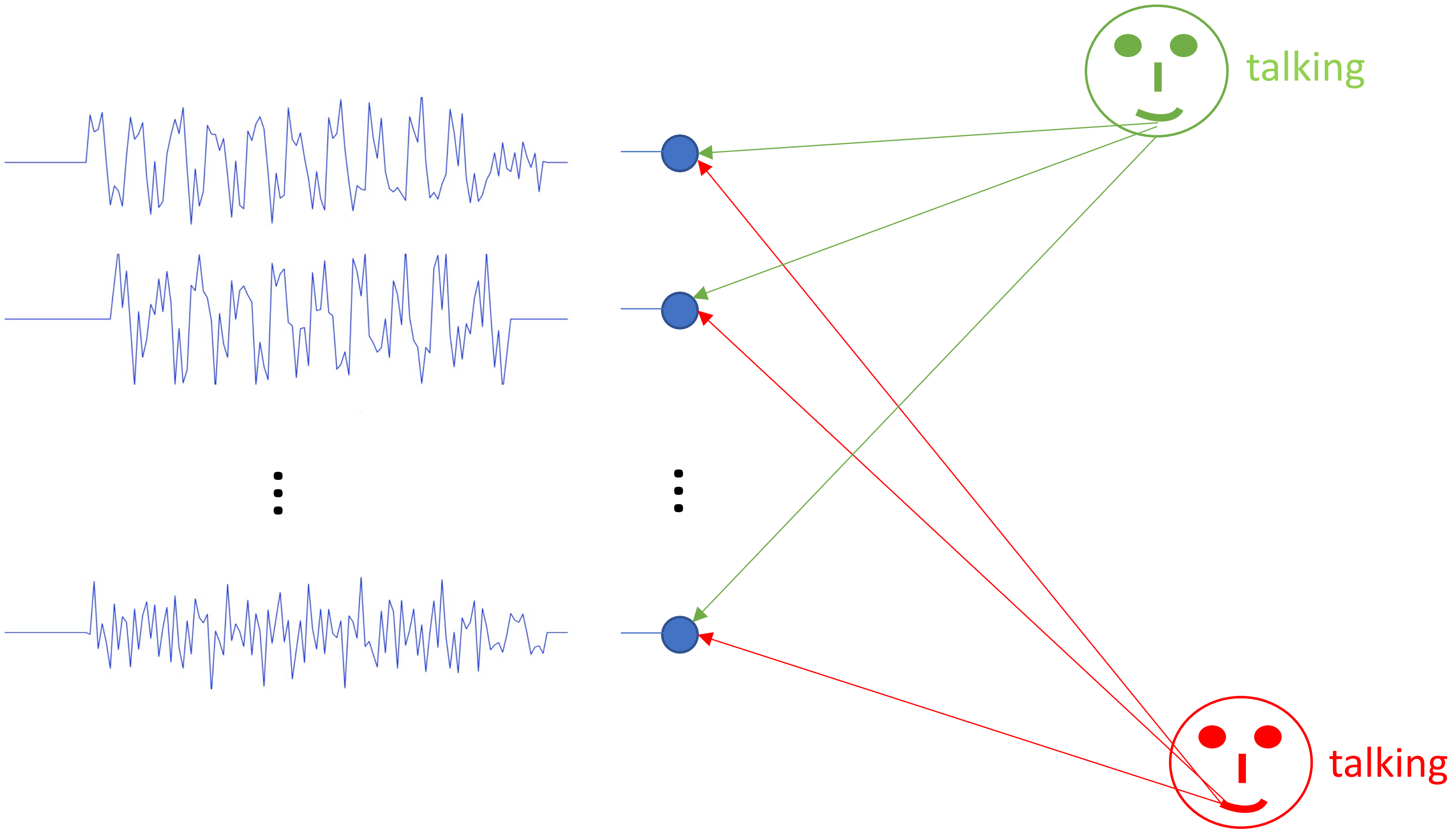
Introduction: a simple acoustic scene



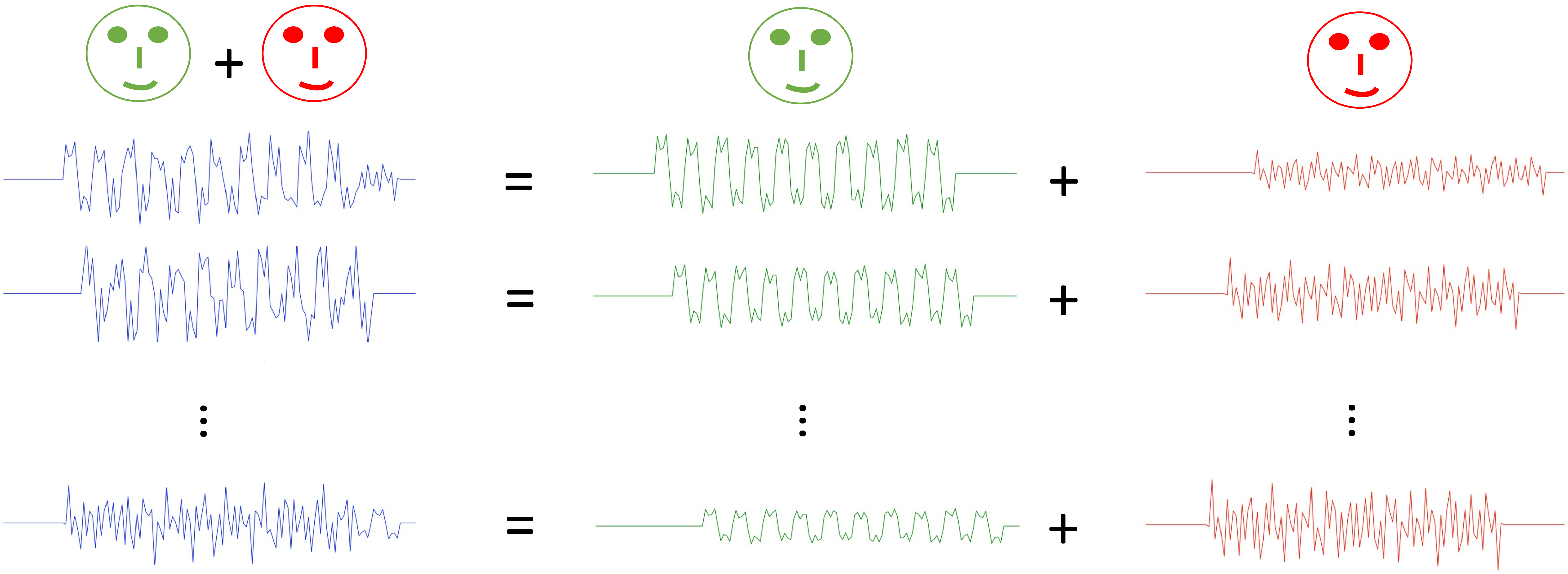
Introduction: a simple acoustic scene



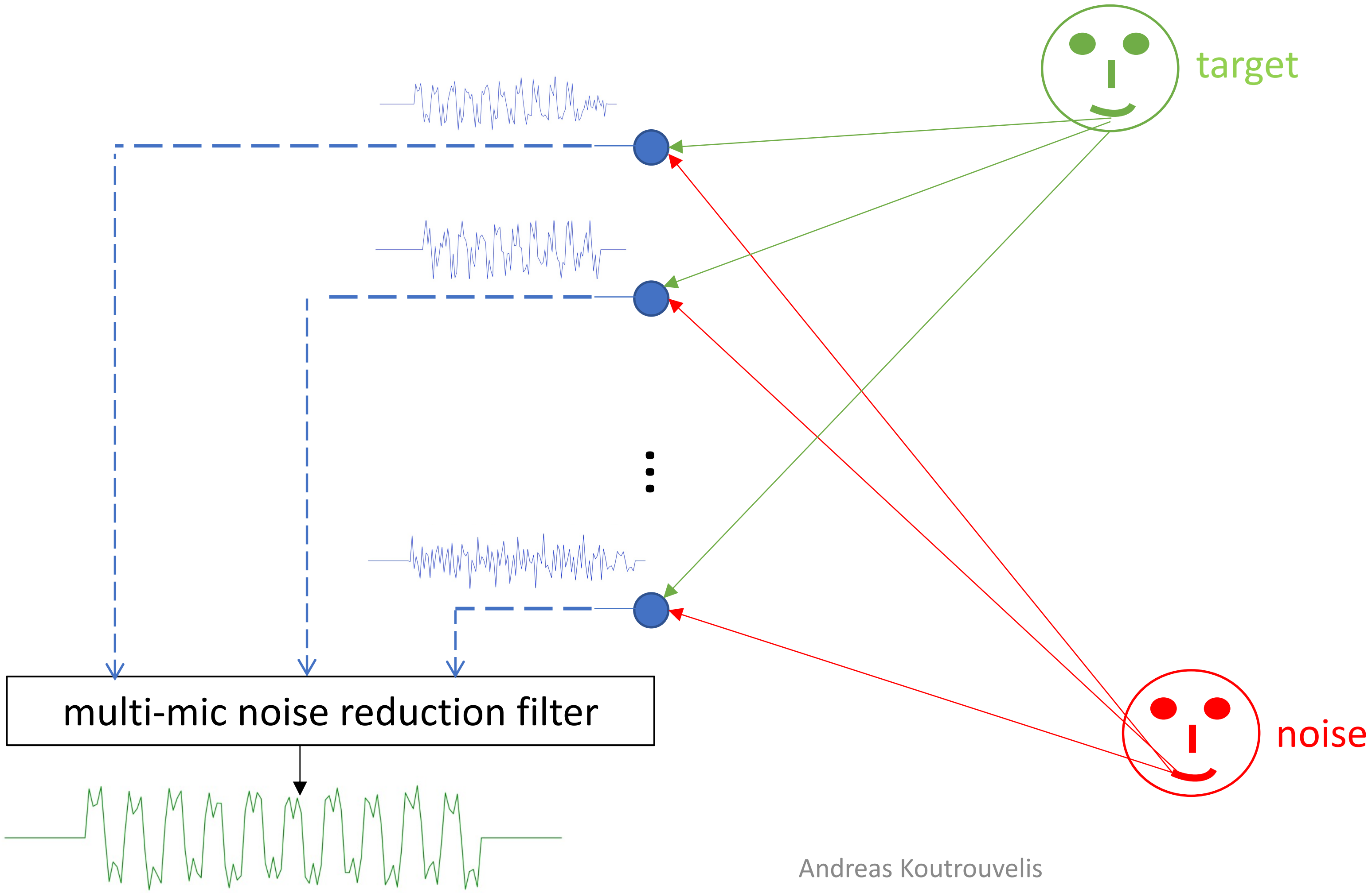
Introduction: a simple acoustic scene



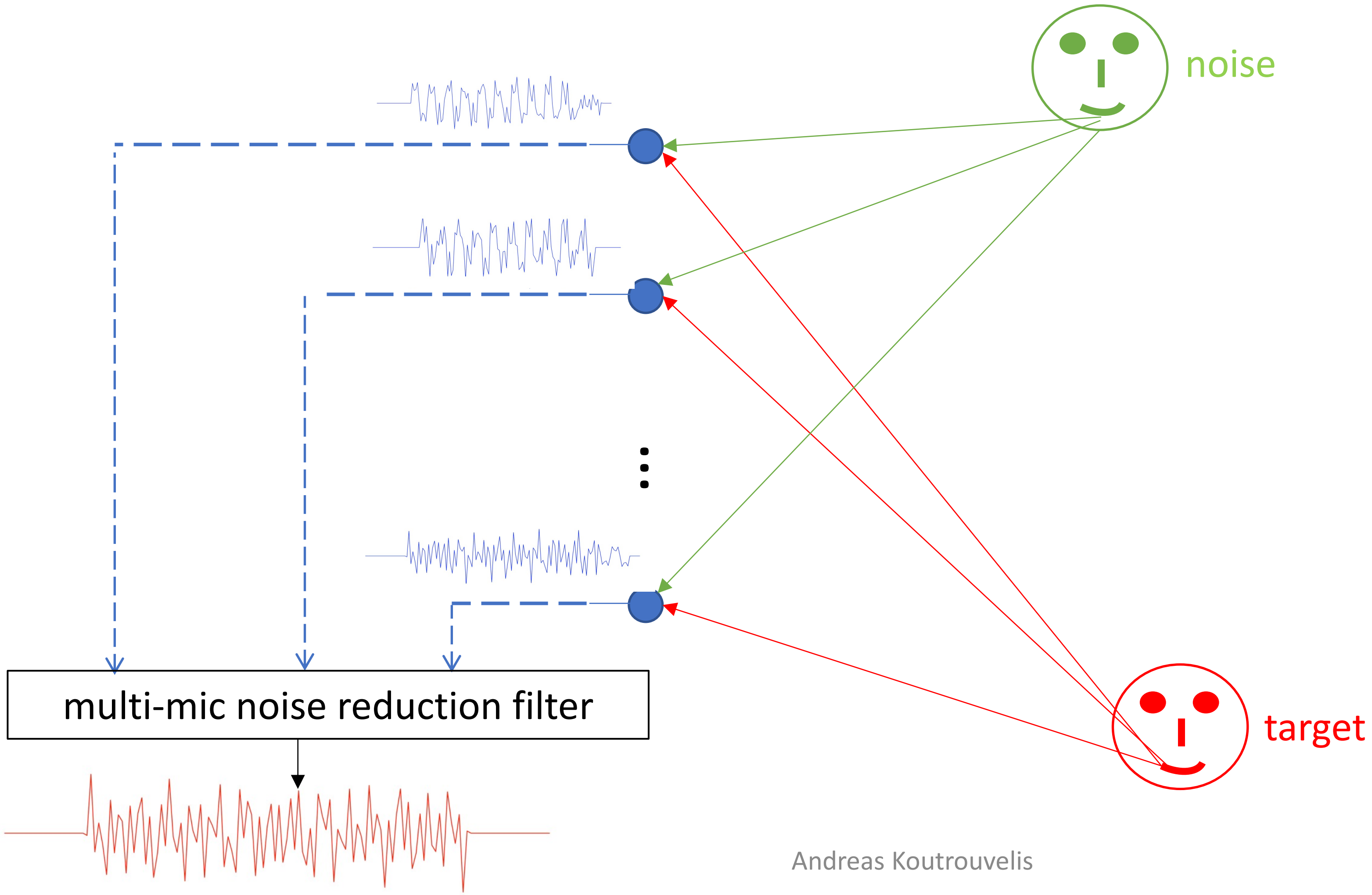
Introduction: a simple acoustic scene



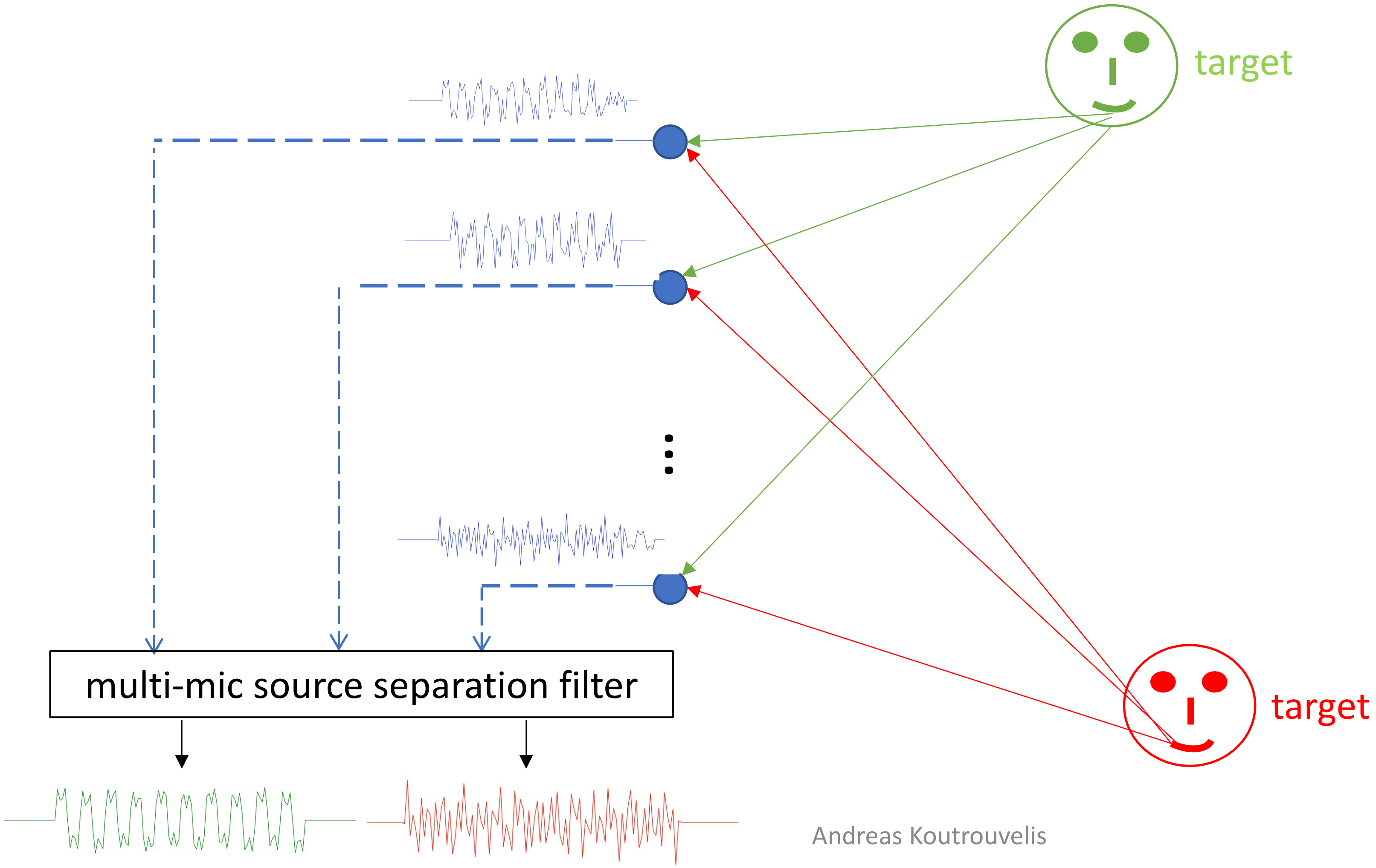
Introduction: multi-microphone noise reduction problem



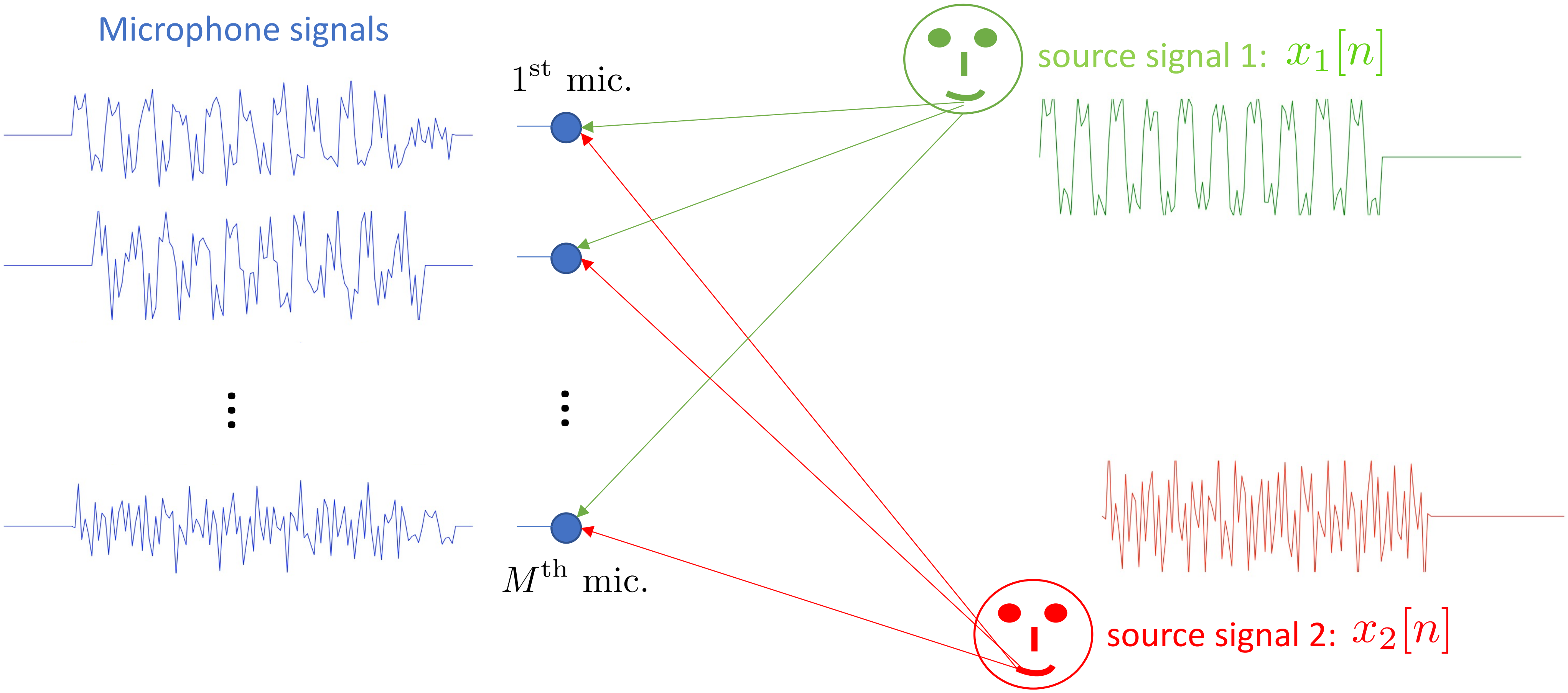
Introduction: multi-microphone noise reduction problem



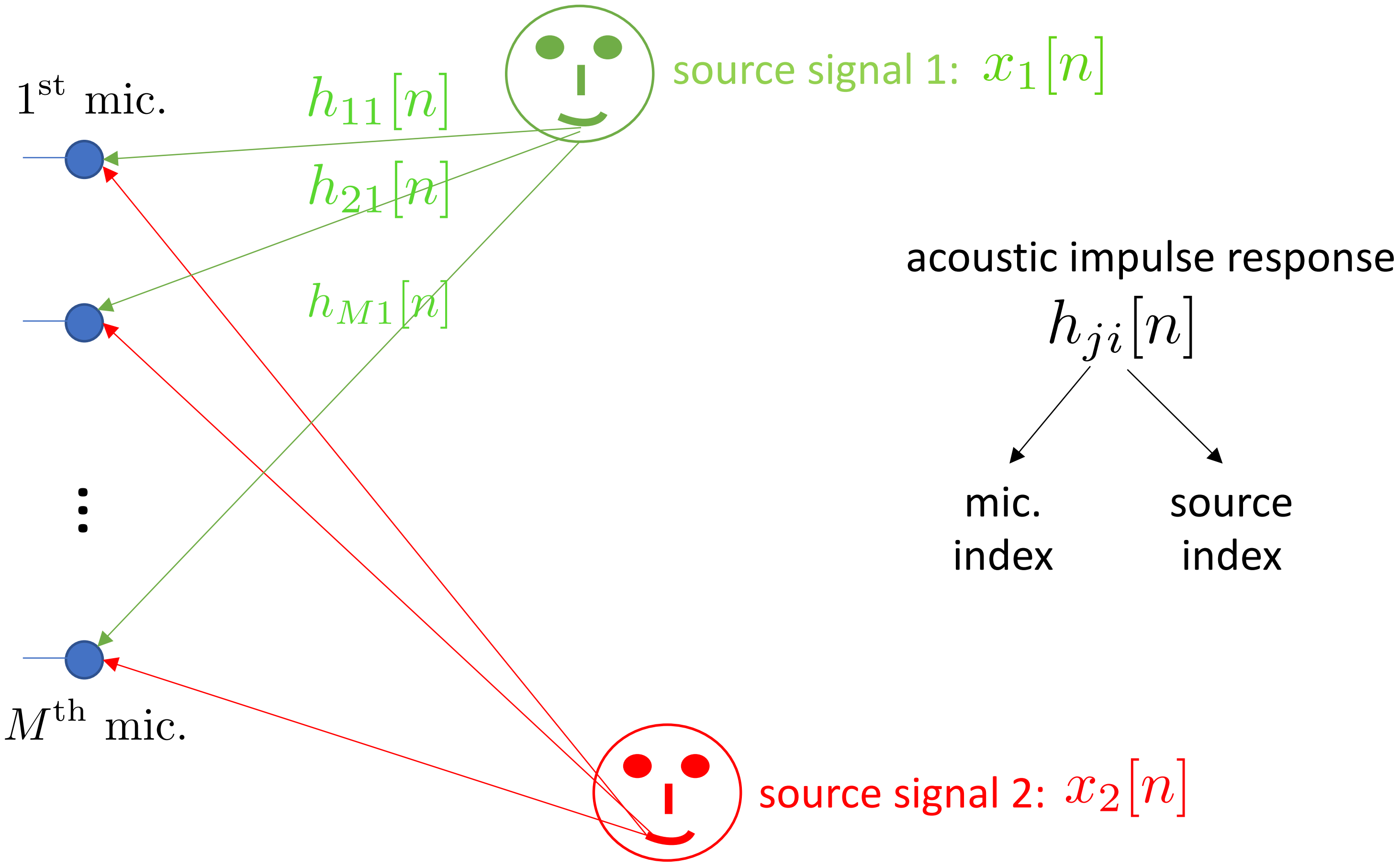
Introduction: source separation problem



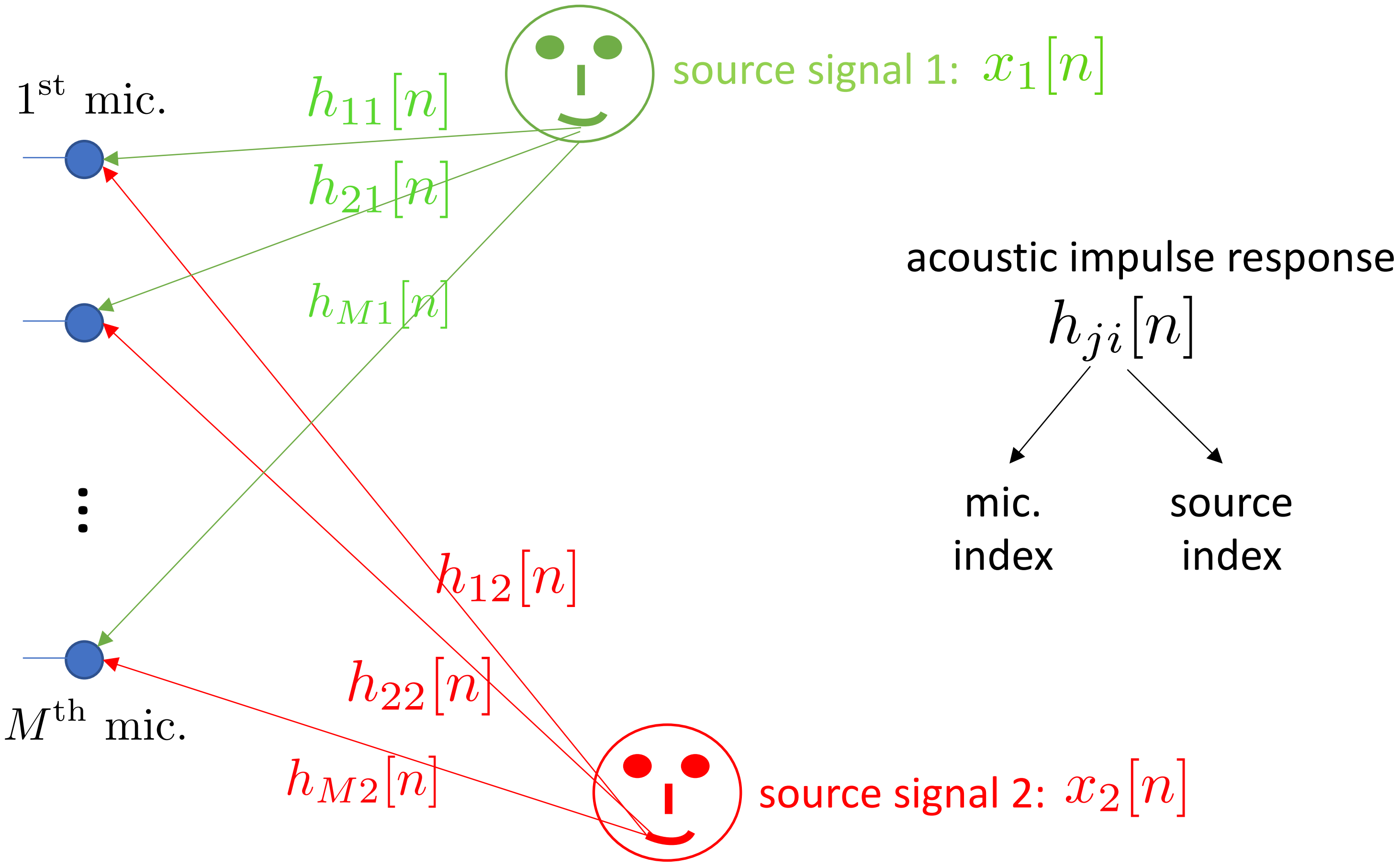
Multi-microphone signal model in time domain



Multi-microphone signal model in time domain



Multi-microphone signal model in time domain



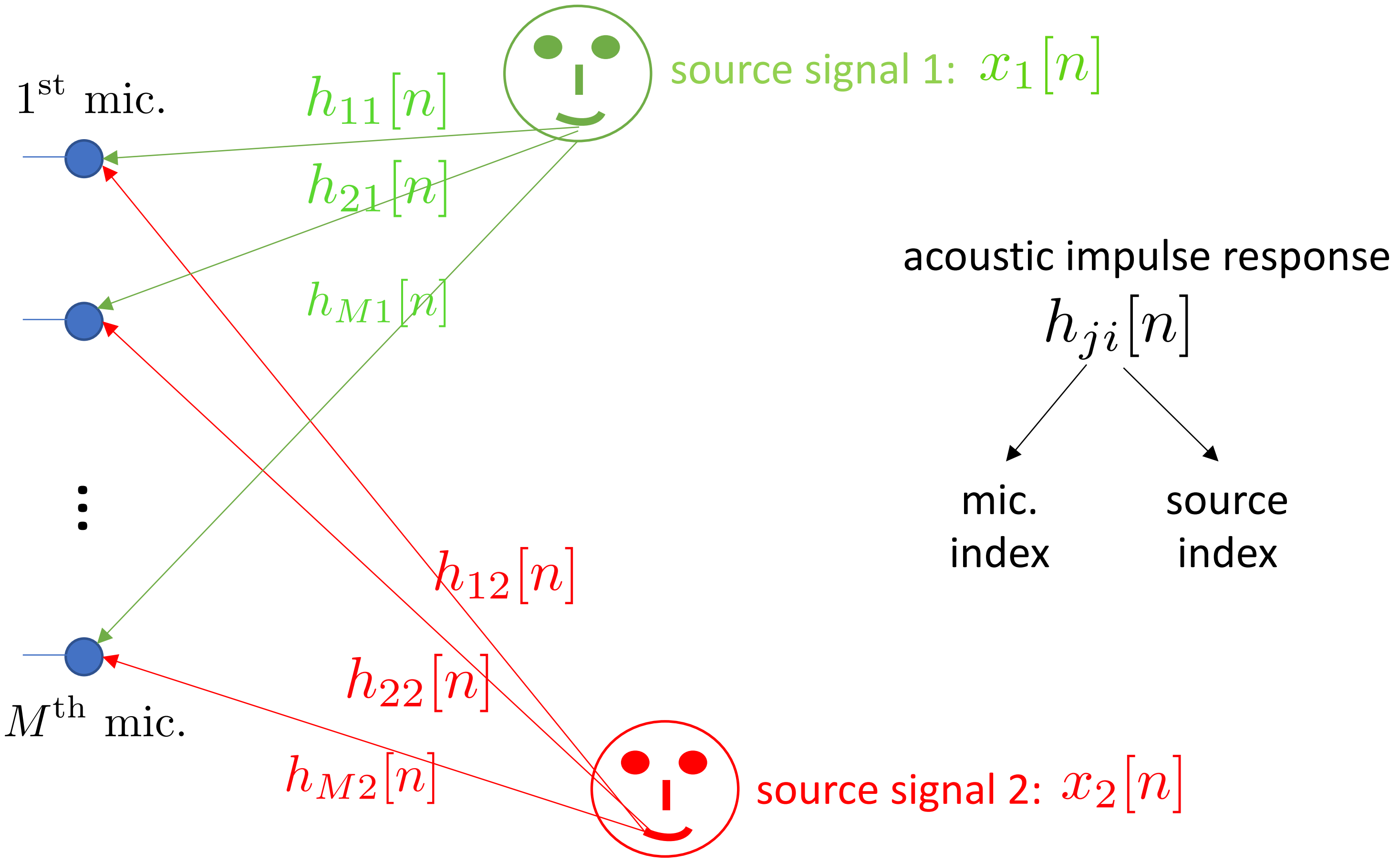
Multi-microphone signal model in time domain

$$z_1[n] = x_1[n] * h_{11}[n] + x_2[n] * h_{12}[n] + v_1[n]$$

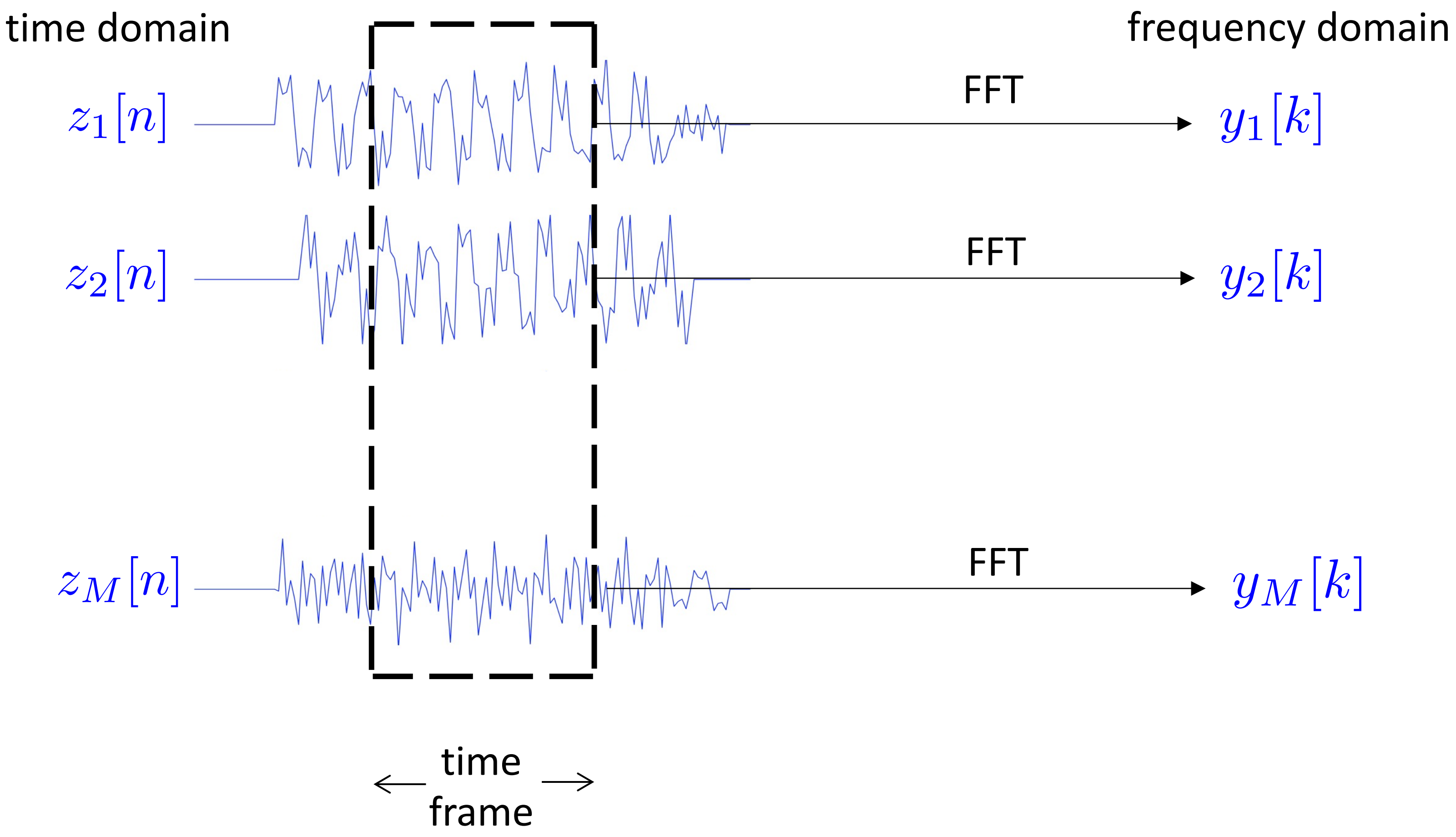
$$z_2[n] = x_1[n] * h_{21}[n] + x_2[n] * h_{22}[n] + v_2[n]$$

$$\vdots$$

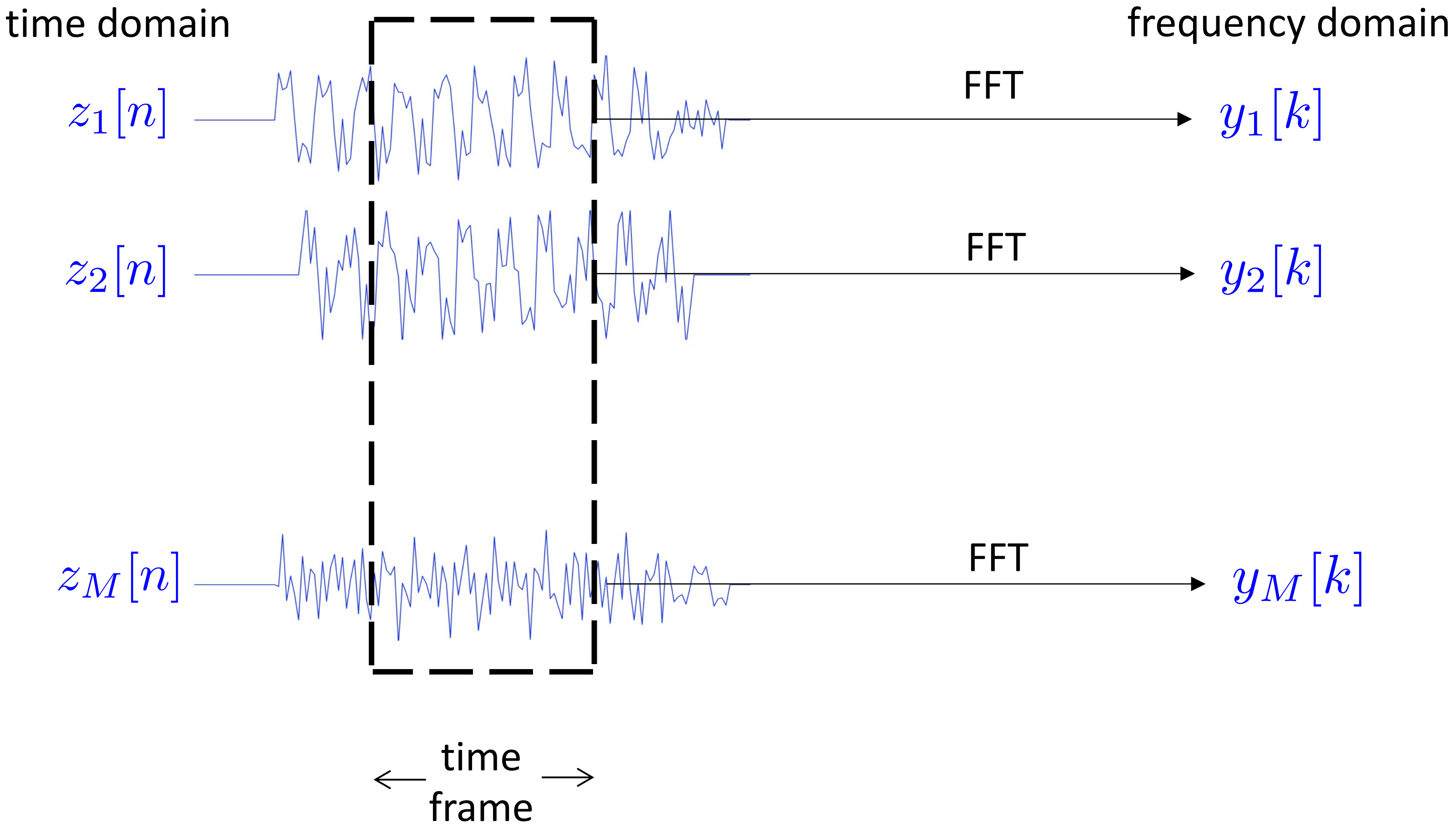
$$z_M[n] = x_1[n] * h_{M1}[n] + x_2[n] * h_{M2}[n] + \underbrace{v_M[n]}_{\text{mic. self noise}}$$



From time-domain to frequency domain



From time-domain to frequency domain



time domain frequency domain

$$x_1[n] \longrightarrow s_1[k]$$

$$x_2[n] \longrightarrow s_2[k]$$

$$h_{ji}[n] \longrightarrow a_{ji}[k]$$

$$v_j[n] \longrightarrow n_j[k]$$

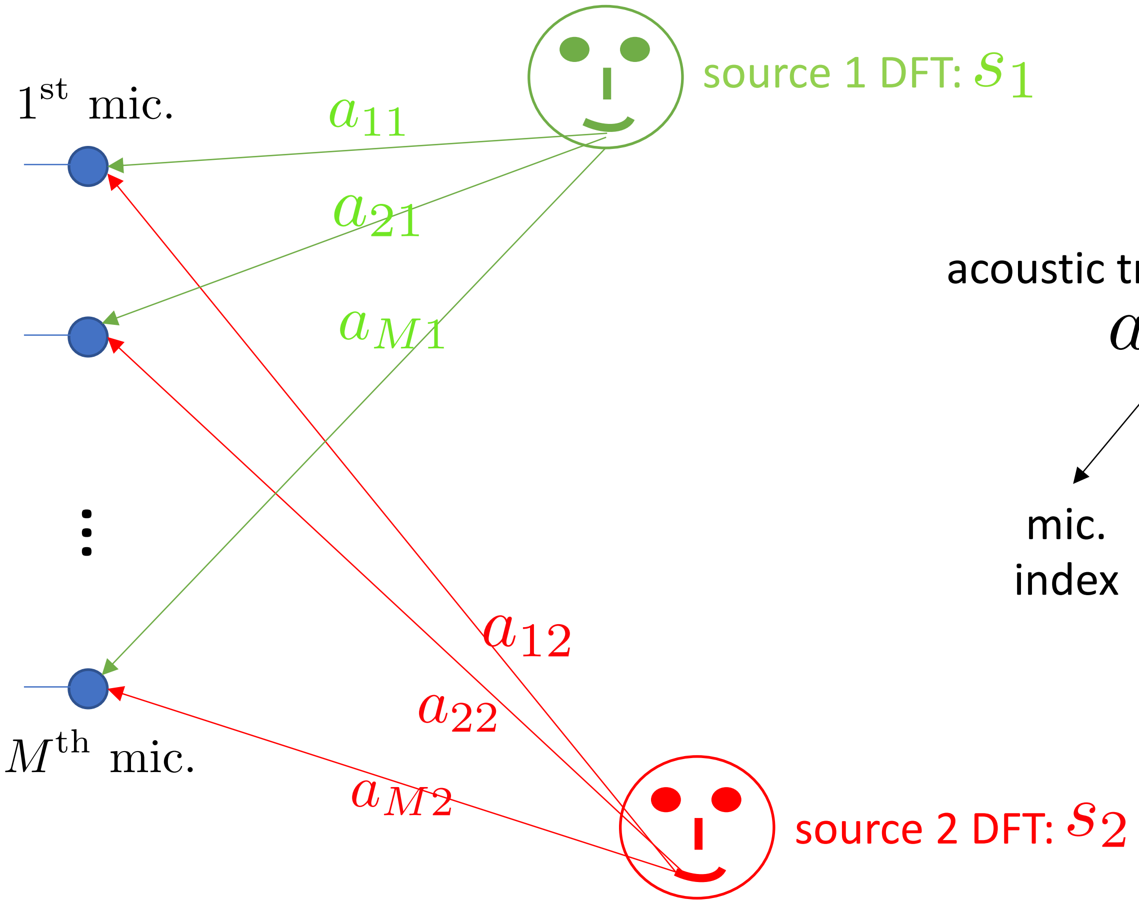
Multi-microphone signal model in frequency domain for a single frequency bin

$$y_1 = s_1 a_{11} + s_2 a_{12} + n_1$$

$$y_2 = s_1 a_{21} + s_2 a_{22} + n_2$$

$$\vdots$$

$$y_M = s_1 a_{M1} + s_2 a_{M2} + n_M$$



acoustic transfer function

a_{ji}

mic. index

source index

Multi-microphone signal model in frequency domain for a single frequency bin

$$y_1 = s_1 a_{11} + s_2 a_{12} + n_1$$

$$y_2 = s_1 a_{21} + s_2 a_{22} + n_2$$

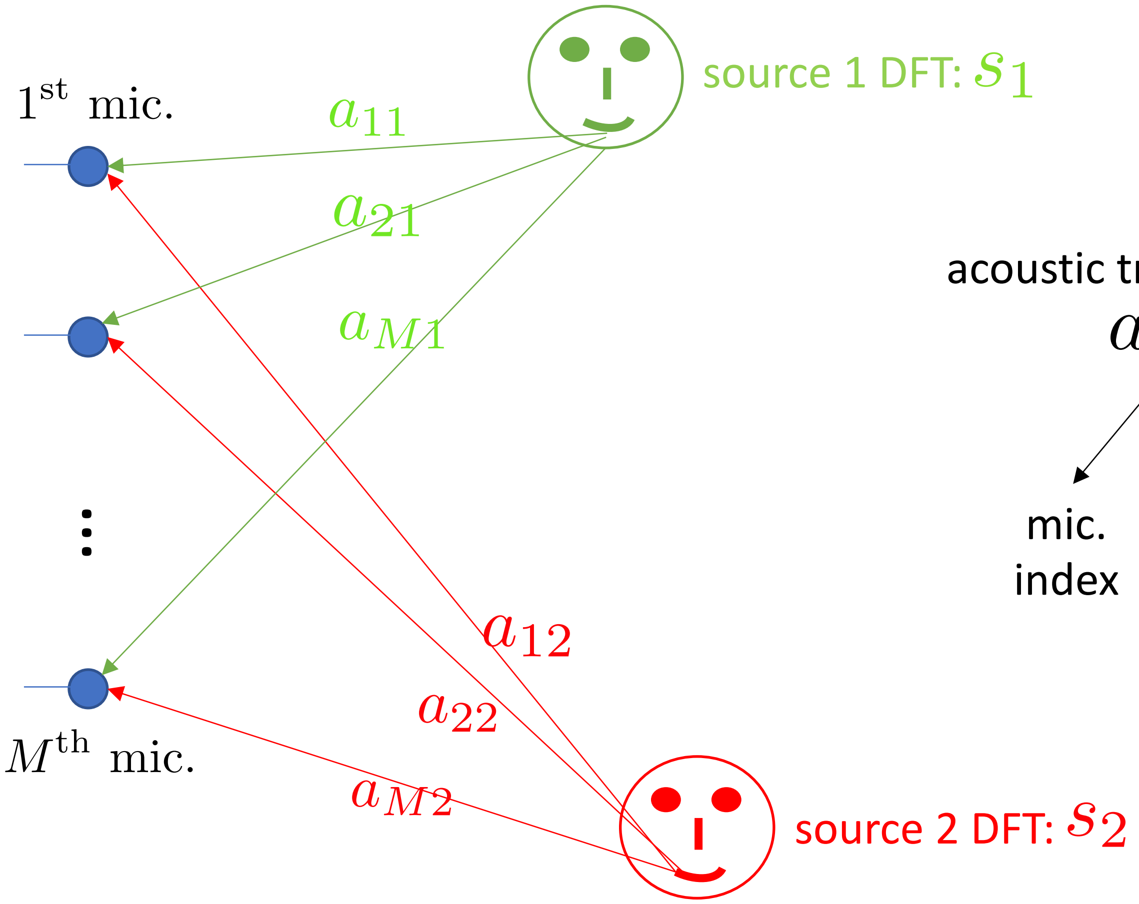
$$\vdots$$

$$y_M = s_1 a_{M1} + s_2 a_{M2} + n_M$$

Approximation!!!: DFT multiplication means circular convolution in time domain. If the FFT size is large enough the circular convolution (after computing the IFFT) becomes a good approximation of the linear convolution.

Multi-microphone signal model in frequency domain for a single frequency bin using matrices and vectors

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = s_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{M1} \end{bmatrix} + s_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix}$$



acoustic transfer function

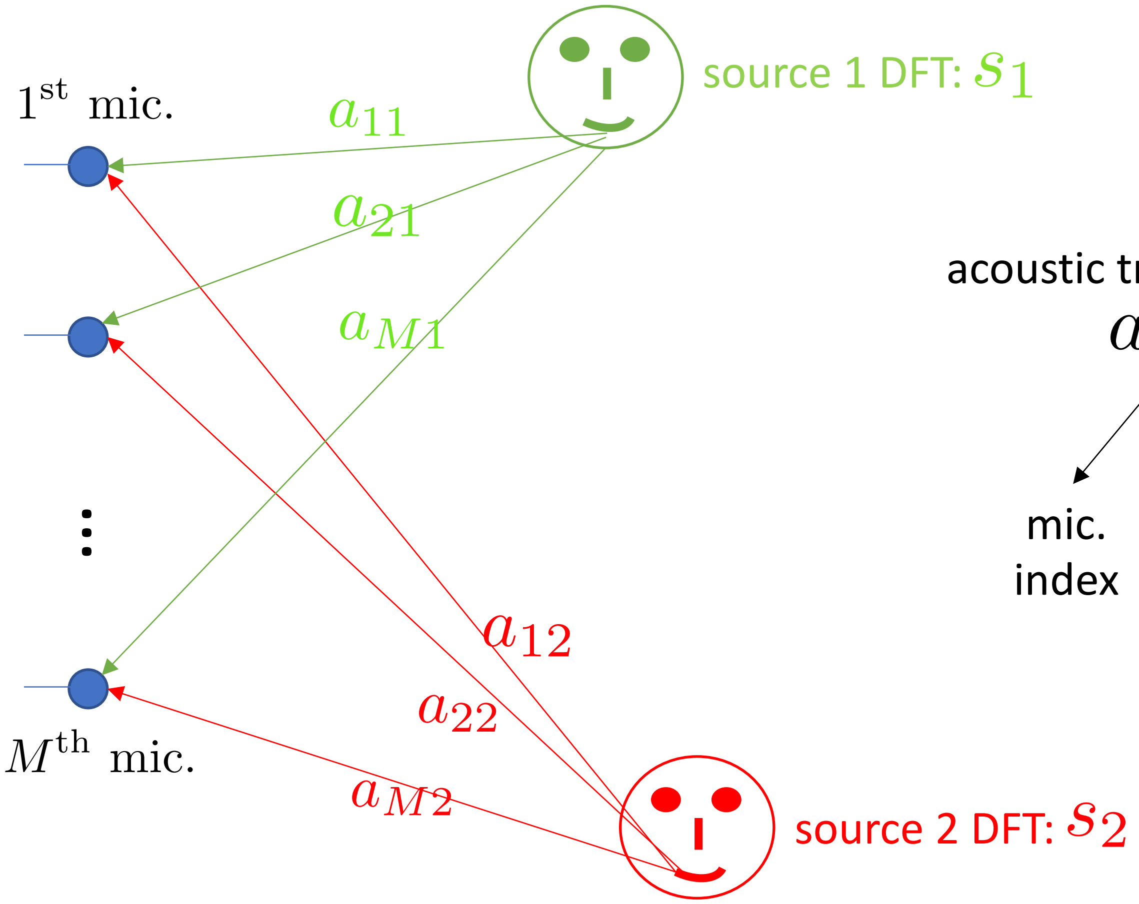
a_{ji}

mic. index

source index

Multi-microphone signal model in frequency domain for a single frequency bin using matrices and vectors

$$\underset{\boldsymbol{y}}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}} = s_1 \underset{\boldsymbol{a}_1}{\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{M1} \end{bmatrix}} + s_2 \underset{\boldsymbol{a}_2}{\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{M2} \end{bmatrix}} + \underset{\boldsymbol{n}}{\begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_M \end{bmatrix}}$$



acoustic transfer function

a_{ji}

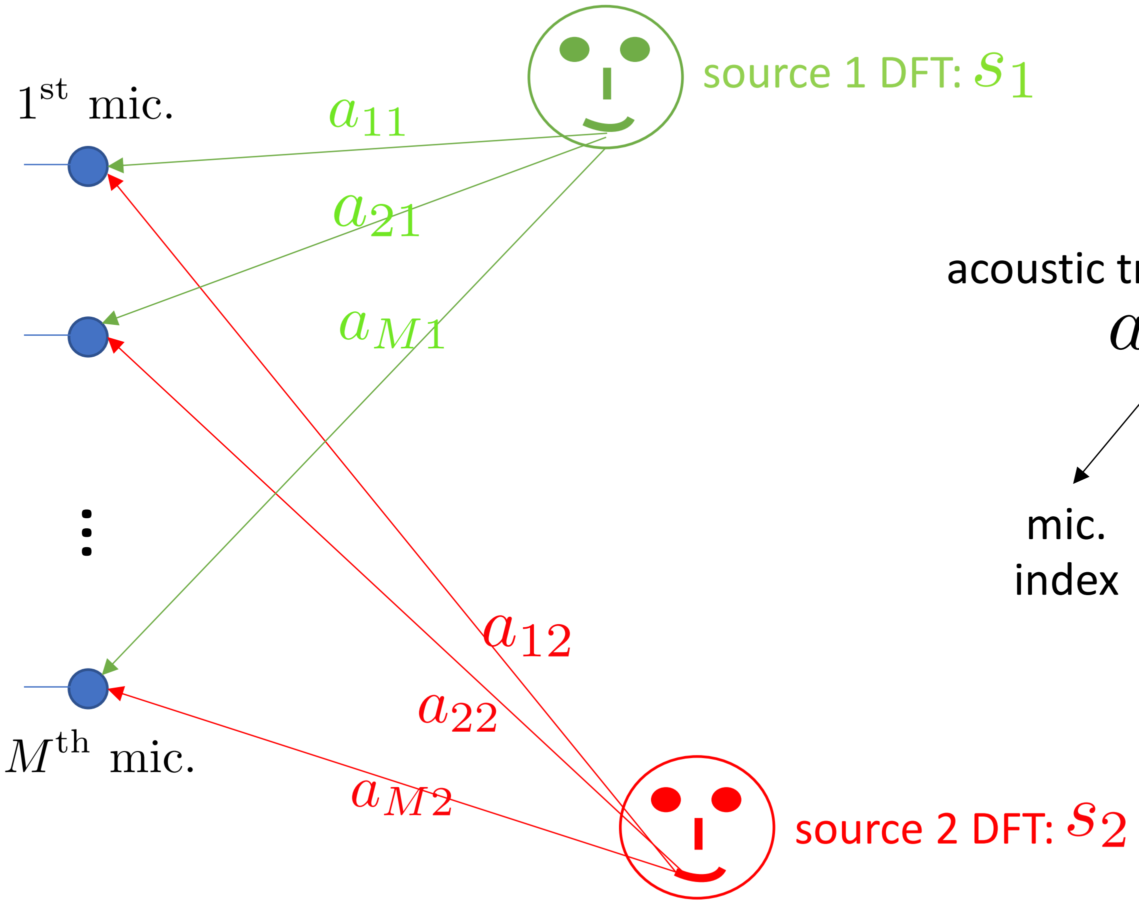
mic. index   source index

bold-face lower-case letters for vectors

Multi-microphone signal model in frequency domain for a single frequency bin using matrices and vectors

$$\boldsymbol{y} = \underbrace{\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2\end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \end{bmatrix}}_{\mathbf{s}} + \boldsymbol{n}$$

$M \times 1 \quad M \times 2 \quad 2 \times 1 \quad M \times 1$



acoustic transfer function

a_{ji}

mic. index

source index

bold-face lower-case letters for vectors

bold-face upper-case letters for matrices

Multi-microphone signal model in frequency domain with r sources

$$\mathbf{y} = \sum_{i=1}^{r} \mathbf{a}_i s_i + \mathbf{n} \rightarrow \mathbf{y} = \underbrace{\mathbf{A}}_{M \times r}\underbrace{\mathbf{s}}_{r \times 1} + \underbrace{\mathbf{n}}_{M \times 1}$$

Multi-microphone signal model in frequency domain with r sources

$$\mathbf{y} = \sum_{i=1}^{r} \mathbf{a}_i s_i + \mathbf{n} \rightarrow \mathbf{y} = \underbrace{\mathbf{A}}_{M \times r}\underbrace{\mathbf{s}}_{r \times 1} + \underbrace{\mathbf{n}}_{M \times 1}$$

$$\mathbf{P}_{\mathbf{y}} = \mathrm{E}\left[(\mathbf{y} - \mu_{\mathbf{y}})(\mathbf{y} - \mu_{\mathbf{y}})^H\right] \quad \text{cross-power spectral density matrix}$$

Multi-microphone signal model in frequency domain with r sources

$$\mathbf{y} = \sum_{i=1}^{r} \mathbf{a}_i s_i + \mathbf{n} \rightarrow \mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

$$M \times r \quad r \times 1 \quad M \times 1$$

$\mathbf{P}_{\mathbf{y}} = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right]$ cross-power spectral density matrix if $\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{0}$

Multi-microphone signal model in frequency domain with r sources

$$\mathbf{y} = \sum_{i=1}^{r} \mathbf{a}_i s_i + \mathbf{n} \rightarrow \mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

$$M \times r \quad r \times 1 \quad M \times 1$$

$$\mathbf{P}_{\mathbf{y}} = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right] = \sum_{i=1}^{r} \underbrace{\mathrm{E}\left[s_i s_i^*\right]}_{p_i} \mathbf{a}_i \mathbf{a}_i^H + \underbrace{\mathrm{E}\left[\mathbf{n}\mathbf{n}^H\right]}_{\mathbf{P}_{\mathbf{n}}} = \sum_{i=1}^{r} p_i \mathbf{a}_i \mathbf{a}_i^H + \mathbf{P}_{\mathbf{n}}$$

$$= \mathbf{A}\mathbf{P}\mathbf{A}^H + \mathbf{P}_{\mathbf{n}}$$

Multi-microphone signal model in frequency domain with r sources

$$\mathbf{y} = \sum_{i=1}^{r} \mathbf{a}_i s_i + \mathbf{n} \rightarrow \mathbf{y} = \underbrace{\mathbf{A}}_{M \times r}\underbrace{\mathbf{s}}_{r \times 1} + \underbrace{\mathbf{n}}_{M \times 1}$$

$$\mathbf{P_y} = \mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right] = \sum_{i=1}^{r} \underbrace{\mathrm{E}\left[s_i s_i^*\right]}_{p_i} \mathbf{a}_i \mathbf{a}_i^H + \underbrace{\mathrm{E}\left[\mathbf{n}\mathbf{n}^H\right]}_{\mathbf{P_n}} = \sum_{i=1}^{r} p_i \mathbf{a}_i \mathbf{a}_i^H + \mathbf{P_n}$$

$$= \mathbf{A}\mathbf{P}\mathbf{A}^H + \boxed{\mathbf{P_n}}$$

is a diagonal matrix because the microphone self noises are mutually uncorrelated.

Multi-microphone noise reduction

Multi-microphone noise reduction

- Typically, in multi-microphone noise reduction we have one target. That is,

$$\mathbf{y} = \underbrace{\mathbf{a}_1 s_1}_{\text{target}} + \underbrace{\sum_{i=2}^{r} \mathbf{a}_i s_i + \mathbf{n}}_{\mathbf{q}:\ \text{total noise}}$$

- The goal is to estimate the target s_1 from the noisy DFT coefficients $\mathbf{y}$. That is,

$$\hat{s}_1 = f(\mathbf{y}),$$

where $f(.)$ is the filter function, where in this presentation is linear. That is,

$$\hat{s}_1 = \mathbf{w}^H \mathbf{y},$$

where $\mathbf{w}$ is the filter vector.

- We want to find a $\mathbf{w}$ such that $\hat{s}_1 = \mathbf{w}^H \mathbf{y} = \underbrace{\mathbf{w}^H \mathbf{a}_1 s_1}_{\approx s_1} + \underbrace{\mathbf{w}^H \mathbf{q}}_{\approx 0},$

Multi-microphone noise reduction

- Typically, in multi-microphone noise reduction we have one target. That is,

$$\mathbf{y} = \underbrace{\mathbf{a}_1 s_1}_{\text{target}} + \underbrace{\sum_{i=2}^{r} \mathbf{a}_i s_i + \mathbf{n}}_{\mathbf{q}:\ \text{total noise}}$$

- The goal is to estimate the target s_1 from the noisy DFT coefficients $\mathbf{y}$. That is,

$$\hat{s}_1 = f(\mathbf{y}),$$

  where $f(.)$ is the filter function, where in this presentation is linear. That is,

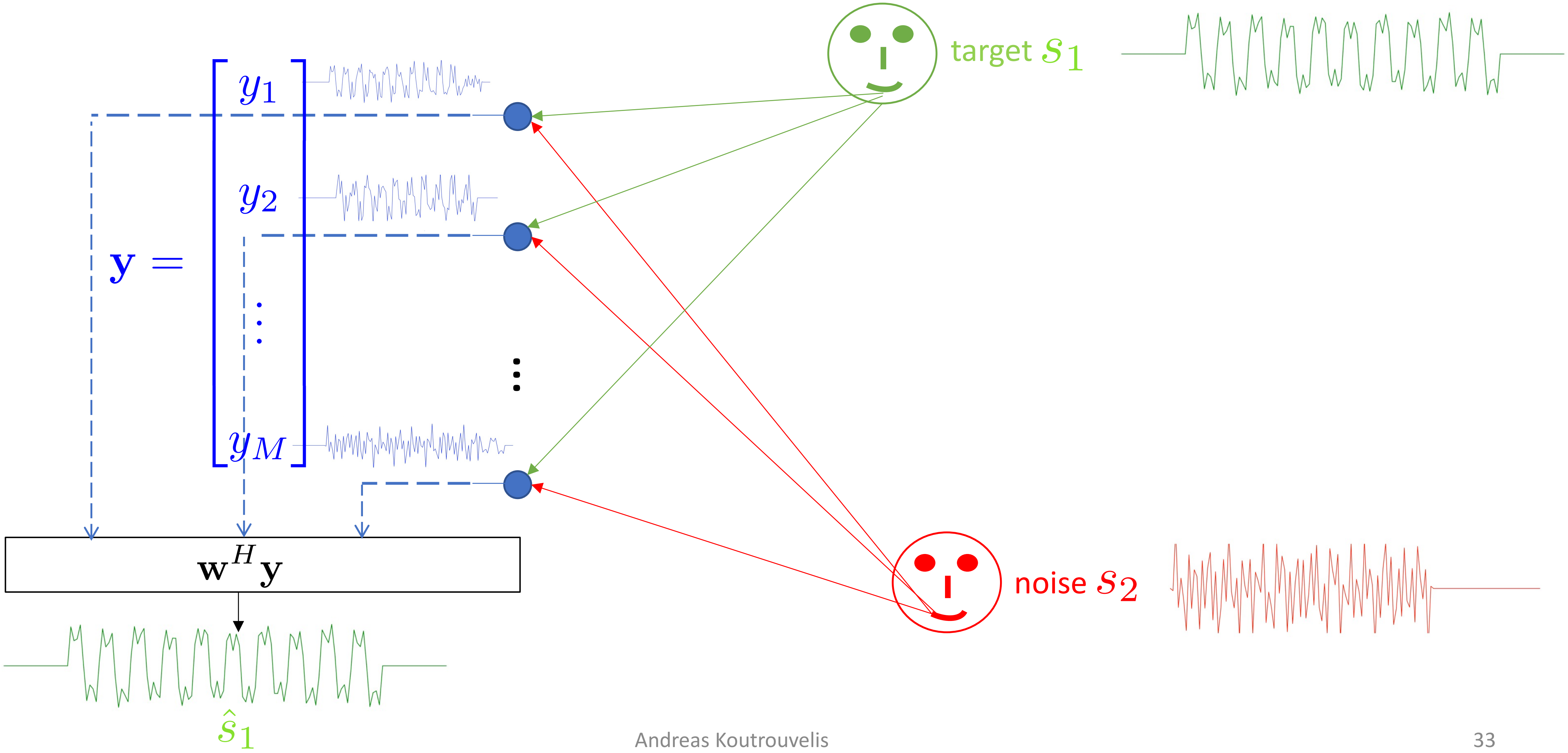
$$\hat{s}_1 = \mathbf{w}^H \mathbf{y},$$

  where $\mathbf{w}$ is the filter vector.

$$\mathbf{w}^H \mathbf{a}_1 = 1$$

- We want to find a $\mathbf{w}$ such that $\hat{s}_1 = \mathbf{w}^H \mathbf{y} = \underbrace{\mathbf{w}^H \mathbf{a}_1 s_1}_{\approx s_1} + \underbrace{\mathbf{w}^H \mathbf{q}}_{\approx 0},$

Multi-microphone noise reduction



Multi-microphone noise reduction

- Signal model in freq. domain: $\mathbf{y} = \underbrace{\mathbf{a}_1 s_1}_{\text{target}} + \underbrace{\sum_{i=2}^{r} \mathbf{a}_i s_i + \mathbf{n}}_{\mathbf{q}:\ \text{total noise}}$

- If the total noise $\mathbf{q}$ is stochastic with zero mean (i.e., $\mathrm{E}\left[\mathbf{q}\right] = \mathbf{0}$) and the target s_1 is deterministic, the variance of the the estimator $\hat{s}_1 = \mathbf{w}^H \mathbf{y}$, is given by

$$
\begin{aligned}
\mathrm{E}\left[(\hat{s}_1 - \mathrm{E}[\hat{s}_1])^2\right] &= \mathrm{E}\left[\left(\mathbf{w}^H \mathbf{y} - \mathrm{E}[\mathbf{w}^H \mathbf{y}]\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{w}^H (\mathbf{a}_1 s_1 + \mathbf{q}) - \mathrm{E}[\mathbf{w}^H (\mathbf{a}_1 s_1 + \mathbf{q})]\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{w}^H \mathbf{a}_1 s_1 + \mathbf{w}^H \mathbf{q} - \mathbf{w}^H \mathbf{a}_1 s_1 - \mathbf{w}^H \mathrm{E}[\mathbf{q}])\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{w}^H \mathbf{q}\right)^2\right] = \mathrm{E}\left[\mathbf{w}^H \mathbf{q}\mathbf{q}^H \mathbf{w}\right] = \mathbf{w}^H \underbrace{\mathrm{E}\left[\mathbf{q}\mathbf{q}^H\right]}_{\mathbf{P}_\mathbf{q}} \mathbf{w} \\
&= \mathbf{w}^H \mathbf{P}_\mathbf{q} \mathbf{w}
\end{aligned}
$$

Multi-microphone noise reduction

- Signal Model in freq. domain: $\mathbf{y} = \underbrace{\mathbf{a}_1 s_1}_{\text{target}} + \underbrace{\sum_{i=2}^{r} \mathbf{a}_i s_i + \mathbf{n}}_{\mathbf{q}:\ \text{total noise}}$

- If the total noise $\mathbf{q}$ is stochastic with zero mean (i.e., $\mathrm{E}\left[\mathbf{q}\right] = \mathbf{0}$) and the target s_1 is deterministic, the variance of the the estimator $\hat{s}_1 = \mathbf{w}^H \mathbf{y}$, is given by

$$
\begin{aligned}
\mathrm{E}\left[(\hat{s}_1 - \mathrm{E}[\hat{s}_1])^2\right] &= \mathrm{E}\left[\left(\mathbf{w}^H \mathbf{y} - \mathrm{E}[\mathbf{w}^H \mathbf{y}]\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{w}^H (\mathbf{a}_1 s_1 + \mathbf{q}) - \mathrm{E}[\mathbf{w}^H (\mathbf{a}_1 s_1 + \mathbf{q})]\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{w}^H \mathbf{a}_1 s_1 + \mathbf{w}^H \mathbf{q} - \mathbf{w}^H \mathbf{a}_1 s_1 - \mathbf{w}^H \mathrm{E}[\mathbf{q}]\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{w}^H \mathbf{q}\right)^2\right] = \mathrm{E}\left[\mathbf{w}^H \mathbf{q}\mathbf{q}^H \mathbf{w}\right] = \mathbf{w}^H \underbrace{\mathrm{E}\left[\mathbf{q}\mathbf{q}^H\right]}_{\mathbf{P}_\mathbf{q}} \mathbf{w} \\
&= \boxed{\mathbf{w}^H \mathbf{P}_\mathbf{q} \mathbf{w}} \rightarrow \text{filter output noise variance/power}
\end{aligned}
$$

$\mathbf{P}_\mathbf{q} = \mathrm{E}\left[\mathbf{q}\mathbf{q}^H\right]$

is the cross power spectral density matrix of the total noise

Multi-microphone noise reduction

- Multi-microphone noise reduction filters:
  - Spatial filters (DS, MVDR and LCMV)
  - Spatio-temporal filters (MWF)
- Spatial filters (also known as beamformers):
  - Aim at suppressing the noise, while at the same time leaving the target signal undistorted.
- Spatio-temporal filters:
  - Provide more aggressive noise suppression at the expense of some target signal distortions.

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

Estimation of s_1 with an unconstrained quadratic optimization problem also known as least squares.

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{y} - \mathbf{a}_1 s_1||_2^2$$

deterministic

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

Estimation of s_1 with an unconstrained quadratic optimization problem also known as least squares.

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{y} - \mathbf{a}_1 \overset{\text{deterministic}}{s_1}||_2^2 \rightarrow \hat{s}_1 = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1^H \mathbf{y}$$

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

Estimation of s_1 with an unconstrained quadratic optimization problem also known as least squares.

deterministic

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{y} - \mathbf{a}_1 s_1||_2^2 \rightarrow \hat{s}_1 = \boxed{\frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1^H} \mathbf{y}$$

$$\hat{\mathbf{w}}^H = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1^H$$: DS filter

Proof: $$\frac{\partial}{\partial \hat{s}_1^*} ||\mathbf{y} - \mathbf{a}_1 \hat{s}_1||_2^2 = 0 \rightarrow \frac{\partial}{\partial \hat{s}_1^*} (\mathbf{y} - \mathbf{a}_1 \hat{s}_1)^H (\mathbf{y} - \mathbf{a}_1 \hat{s}_1) = 0 \rightarrow$$

$$\rightarrow \frac{\partial}{\partial \hat{s}_1^*} \left[\mathbf{y}^H \mathbf{y} - \mathbf{y}^H \mathbf{a}_1 \hat{s}_1 - \hat{s}_1^* \mathbf{a}_1^H \mathbf{y} + \hat{s}_1^* \hat{s}_1 \mathbf{a}_1^H \mathbf{a}_1\right] = 0 \rightarrow -\mathbf{a}_1^H \mathbf{y} + \hat{s}_1 \mathbf{a}_1^H \mathbf{a}_1 = 0 \rightarrow \hat{s}_1 = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1^H \mathbf{y}$$

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

deterministic

Estimation of s_1 with an unconstrained quadratic optimization problem also known as least squares.

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{y} - \mathbf{a}_1 s_1||_2^2 \rightarrow \hat{s}_1 = \boxed{\frac{1}{\mathbf{a}_1^H \mathbf{a}_1}\mathbf{a}_1^H}\mathbf{y}$$

$$\hat{\mathbf{w}}^H = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1}\mathbf{a}_1^H$$: DS filter

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_1 = 1 \rightarrow \hat{\mathbf{w}} = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1}\mathbf{a}_1$$

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_1 = 1 \rightarrow \hat{\mathbf{w}} = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1$$

Proof: $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{w} + 2\Re\left\{\lambda \left(\mathbf{w}^H \mathbf{a}_1 - 1\right)\right\}$

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H\mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H\mathbf{a}_1 = 1 \rightarrow \hat{\mathbf{w}} = \frac{1}{\mathbf{a}_1^H\mathbf{a}_1}\mathbf{a}_1$$

Proof: $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H\mathbf{w} + 2\Re\left\{\lambda\left(\mathbf{w}^H\mathbf{a}_1 - 1\right)\right\}$

$$\frac{\partial}{\partial\hat{\mathbf{w}}^*}\left[\hat{\mathbf{w}}^H\hat{\mathbf{w}} + \lambda\left(1 - \hat{\mathbf{w}}^H\mathbf{a}_1\right) + \lambda^*\left(1 - \mathbf{a}_1^H\hat{\mathbf{w}}\right)\right] = 0 \rightarrow \hat{\mathbf{w}} - \lambda\mathbf{a}_1 = 0 \rightarrow \hat{\mathbf{w}} = \lambda\mathbf{a}_1$$

$$\left.\begin{aligned} \hat{\mathbf{w}} &= \lambda\mathbf{a}_1 \\ \hat{\mathbf{w}}^H\mathbf{a}_1 &= 1 \end{aligned}\right\} \lambda = \frac{1}{\mathbf{a}^H\mathbf{a}}$$

Multi-microphone noise reduction: delay and sum (DS) filter (Flanagan 1993)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_1 = 1 \rightarrow \hat{\mathbf{w}} = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1$$

Proof: $$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{w} + 2\Re\left\{\lambda\left(\mathbf{w}^H \mathbf{a}_1 - 1\right)\right\}$$

$$\hat{\mathbf{w}} = \frac{1}{\mathbf{a}_1^H \mathbf{a}_1} \mathbf{a}_1$$

$$\frac{\partial}{\partial \hat{\mathbf{w}}^*}\left[\hat{\mathbf{w}}^H \hat{\mathbf{w}} + \lambda\left(1 - \hat{\mathbf{w}}^H \mathbf{a}_1\right) + \lambda^*\left(1 - \mathbf{a}_1^H \hat{\mathbf{w}}\right)\right] = 0 \rightarrow \hat{\mathbf{w}} - \lambda \mathbf{a}_1 = 0 \rightarrow \hat{\mathbf{w}} = \lambda \mathbf{a}_1$$

$$\left.\begin{aligned} \hat{\mathbf{w}} &= \lambda \mathbf{a}_1 \\ \hat{\mathbf{w}}^H \mathbf{a}_1 &= 1 \end{aligned}\right\} \lambda = \frac{1}{\mathbf{a}^H \mathbf{a}}$$

Multi-microphone noise reduction: minimum variance distortionless response (MVDR) filter (Capon 1969)

Estimation of s_1 with an unconstrained quadratic optimization problem also known as weighted least squares.

deterministic

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{P}_{\mathbf{q}}^{-1/2}(\mathbf{y} - \mathbf{a}_1 s_1)||_2^2$$

Multi-microphone noise reduction: minimum variance distortionless response (MVDR) filter (Capon 1969)

Estimation of s_1 with an unconstrained quadratic optimization problem also known as weighted least squares.

deterministic

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{P}_{\mathbf{q}}^{-1/2}(\mathbf{y} - \mathbf{a}_1 s_1)||_2^2 \rightarrow \hat{s}_1 = \boxed{\frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1}} \mathbf{y}$$

$$\hat{\mathbf{w}}^H = \frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} : \text{MVDR filter}$$

Multi-microphone noise reduction: minimum variance distortionless response (MVDR) filter (Capon 1969)

Estimation of s_1 with an unconstrained quadratic optimization problem also known as weighted least squares.

deterministic

$$\hat{s}_1 = \arg\min_{s_1} ||\mathbf{P}_{\mathbf{q}}^{-1/2}(\mathbf{y} - \mathbf{a}_1 s_1)||_2^2 \rightarrow \hat{s}_1 = \boxed{\frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1}} \mathbf{y}$$

$$\hat{\mathbf{w}}^H = \frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} : \text{MVDR filter}$$

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{a}_1 = 1 \rightarrow \hat{\mathbf{w}} = \frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1$$

Multi-microphone noise reduction: linearly constrained minimum variance (LCMV) filter (Frost 1972)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H$$

Multi-microphone noise reduction: linearly constrained minimum variance (LCMV) filter (Frost 1972)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \boxed{\mathbf{w}^H \mathbf{\Lambda} = \mathbf{f}^H}$$

example

$$\mathbf{w}^H \mathbf{a}_1 = 1$$
$$\mathbf{w}^H \mathbf{a}_2 = 0$$

Multi-microphone noise reduction: linearly constrained minimum variance (LCMV) filter (Frost 1972)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \boxed{\mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H}$$

example

$$\boxed{\begin{aligned} \mathbf{w}^H \mathbf{a}_1 &= 1 \\ \mathbf{w}^H \mathbf{a}_2 &= 0 \end{aligned}} \rightarrow \boxed{\mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}}_{\underset{M \times 2}{\boldsymbol{\Lambda}}} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\underset{1 \times 2}{\mathbf{f}^H}}}$$

Multi-microphone noise reduction: linearly constrained minimum variance (LCMV) filter (Frost 1972)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \boxed{\mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H}$$

example

$$\boxed{\begin{aligned} \mathbf{w}^H \mathbf{a}_1 &= 1 \\ \mathbf{w}^H \mathbf{a}_2 &= 0 \end{aligned}}$$

$$\boxed{\mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}}_{\underset{M \times 2}{\boldsymbol{\Lambda}}} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\underset{1 \times 2}{\mathbf{f}^H}}}$$

M-2 degrees of freedom for noise reduction

Multi-microphone noise reduction: linearly constrained minimum variance (LCMV) filter (Frost 1972)

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \boxed{\mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H}$$

example

$$\boxed{\begin{aligned} \mathbf{w}^H \mathbf{a}_1 &= 1 \\ \mathbf{w}^H \mathbf{a}_2 &= 0 \end{aligned}}$$

$$\boxed{\mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}}_{\substack{\boldsymbol{\Lambda} \\ M\times 2}} = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{\substack{\mathbf{f}^H \\ 1\times 2}}}$$

M-2 degrees of freedom for noise reduction

More constraints means more control, but less total noise reduction!

Multi-microphone noise reduction: linearly constrained minimum variance (LCMV) filter (Frost 1972)

LCMV filter

Estimation of $\mathbf{w}$ with a constrained quadratic optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathbf{w}^H \mathbf{P}_{\mathbf{q}} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \boldsymbol{\Lambda} = \mathbf{f}^H \rightarrow \boxed{\hat{\mathbf{w}} = \mathbf{P}_{\mathbf{q}}^{-1} \boldsymbol{\Lambda} \left(\boldsymbol{\Lambda}^H \mathbf{P}_{\mathbf{q}}^{-1} \boldsymbol{\Lambda} \right)^{-1} \mathbf{f}}$$

- The MVDR filter is a special case of the LCMV filter which provides the minimum filter output noise power, but does not control how the noise sources will be suppressed.
- The LCMV optimization problem gives you more control to design the filter vector than the MVDR optimization problem.
- More constraints means more control, but less total noise reduction!

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

stochastic

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[|\mathbf{w}^H\mathbf{y} - s_1|^2\right]$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

stochastic

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[|\mathbf{w}^H\mathbf{y} - s_1|^2\right] \rightarrow \hat{\mathbf{w}} = p_1\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{a}$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

stochastic

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[|\mathbf{w}^H\mathbf{y} - s_1|^2\right] \rightarrow \hat{\mathbf{w}} = p_1\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{a}$$

Proof: $$\frac{\partial}{\partial\hat{\mathbf{w}}^*}\left\{\mathrm{E}\left[|\hat{\mathbf{w}}^H\mathbf{y} - s_1|^2\right]\right\} = 0 \rightarrow \frac{\partial}{\partial\hat{\mathbf{w}}^*}\left\{\mathrm{E}\left[(\hat{\mathbf{w}}^H\mathbf{y} - s_1)^*(\hat{\mathbf{w}}^H\mathbf{y} - s_1)\right]\right\} = 0 \rightarrow$$

$$\rightarrow \frac{\partial}{\partial\hat{\mathbf{w}}^*}\left\{\mathrm{E}\left[s_1^*s_1 - s_1^*\hat{\mathbf{w}}^H\mathbf{y} - s_1\mathbf{y}^H\hat{\mathbf{w}} + \hat{\mathbf{w}}^H\mathbf{y}\mathbf{y}^H\hat{\mathbf{w}}\right]\right\} = 0 \rightarrow$$

$$\rightarrow \frac{\partial}{\partial\hat{\mathbf{w}}^*}\left\{\mathrm{E}\left[s_1^*s_1\right] - \mathrm{E}\left[s_1^*\hat{\mathbf{w}}^H\mathbf{y}\right] - \mathrm{E}\left[s_1\mathbf{y}^H\hat{\mathbf{w}}\right] + \mathrm{E}\left[\hat{\mathbf{w}}^H\mathbf{y}\mathbf{y}^H\hat{\mathbf{w}}\right]\right\} = 0 \rightarrow$$

$$\rightarrow -\frac{\partial}{\partial\hat{\mathbf{w}}^*}\mathrm{E}\left[s_1^*\hat{\mathbf{w}}^H\mathbf{y}\right] + \frac{\partial}{\partial\hat{\mathbf{w}}^*}\mathrm{E}\left[\hat{\mathbf{w}}^H\mathbf{y}\mathbf{y}^H\hat{\mathbf{w}}\right] = 0 \rightarrow -\frac{\partial}{\partial\hat{\mathbf{w}}^*}\hat{\mathbf{w}}^H\mathrm{E}\left[s_1^*\mathbf{y}\right] + \frac{\partial}{\partial\hat{\mathbf{w}}^*}\hat{\mathbf{w}}^H\mathrm{E}\left[\mathbf{y}\mathbf{y}^H\right]\hat{\mathbf{w}} = 0 \rightarrow$$

$$\rightarrow \mathrm{E}\left[s_1^*\mathbf{y}\right] + \mathbf{P}_{\mathbf{y}}\hat{\mathbf{w}} = 0 \rightarrow \mathrm{E}\left[s_1^*s_1\mathbf{a}_1 + s_1^*\mathbf{n}\right] + \mathbf{P}_{\mathbf{y}}\hat{\mathbf{w}} = 0 \rightarrow \mathrm{E}\left[s_1^*s_1\right]\mathbf{a}_1 + \mathrm{E}\left[s_1^*\mathbf{n}\right] + \mathbf{P}_{\mathbf{y}}\hat{\mathbf{w}} = 0 \xrightarrow{\mathrm{E}[s_1^*\mathbf{n}]=0} \hat{\mathbf{w}} = p_1\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{a}_1$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[|\mathbf{w}^H\mathbf{y} - s_1|^2\right] \rightarrow \hat{\mathbf{w}} = p_1 \mathbf{P}_{\mathbf{y}}^{-1}\mathbf{a}$$

stochastic

Estimation of s_1 with an unconstrained MSE optimization problem.

$$\hat{s}_1 = \arg\min_{s_1} \mathrm{E}\left[|\hat{s}_1 - s_1|^2\right] = \mathrm{E}\left[s_1|\mathbf{y}\right] \xrightarrow[]{\substack{s_1 \sim N(0, p_1) \\ \mathbf{n} \sim N(\mathbf{0}, \mathbf{P}_{\mathbf{n}})}} \hat{s}_1 = \underbrace{\boxed{p_1 \mathbf{a}_1^H \mathbf{P}_{\mathbf{y}}^{-1}}}_{\hat{\mathbf{w}}^H}\mathbf{y}$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \underbrace{\mathrm{E}}_{\text{stochastic}}\left[\left|\mathbf{w}^H\mathbf{y} - s_1\right|^2\right] \rightarrow \hat{\mathbf{w}} = p_1 \mathbf{P}_{\mathbf{y}}^{-1}\mathbf{a}_1$$

$$= p_1\left(p_1\mathbf{a}_1\mathbf{a}_1^H + \mathbf{P}_{\mathbf{q}}\right)^{-1}\mathbf{a}_1$$

$$= \frac{p_1}{p_1 + \underbrace{\left(\mathbf{a}_1^H\mathbf{P}_{\mathbf{q}}^{-1}\mathbf{a}_1\right)^{-1}}_{\text{MVDR output noise power}}} \underbrace{\frac{1}{\mathbf{a}_1^H\mathbf{P}_{\mathbf{q}}^{-1}\mathbf{a}_1}\mathbf{P}_{\mathbf{q}}^{-1}\mathbf{a}_1}_{\text{MVDR filter}}$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Sherman Morrison formula

$$\mathbf{P}_{\mathbf{q}}^{-1} - \frac{p_1 \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1 \mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1}}{1 + p_1 \mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1}$$

stochastic

Estimation of $\mathbf{W}$ with an unconstrained LMSE optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[\left| \mathbf{w}^H \mathbf{y} - s_1 \right|^2 \right] \rightarrow \hat{\mathbf{w}} = p_1 \mathbf{P}_{\mathbf{y}}^{-1} \mathbf{a}_1$$

$$= p_1 \left(p_1 \mathbf{a}_1 \mathbf{a}_1^H + \mathbf{P}_{\mathbf{q}} \right)^{-1} \mathbf{a}_1$$

$$= \frac{p_1}{p_1 + \underbrace{\left(\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1 \right)^{-1}}_{\text{MVDR output noise power}}} \underbrace{\frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1}_{\text{MVDR filter}}$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \overset{\text{stochastic}}{\mathrm{E}} \left[\left| \mathbf{w}^H \mathbf{y} - s_1 \right|^2 \right] \rightarrow \hat{\mathbf{w}} = p_1 \mathbf{P}_{\mathbf{y}}^{-1} \mathbf{a}_1$$

$$= p_1 \left(p_1 \mathbf{a}_1 \mathbf{a}_1^H + \mathbf{P}_{\mathbf{q}} \right)^{-1} \mathbf{a}_1$$

$$= \underbrace{\boxed{\frac{p_1}{p_1 + \underbrace{\left(\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1 \right)^{-1}}_{\text{MVDR output noise power}}}}}_{\text{single channel Wiener filter (SWF)}} \underbrace{\frac{1}{\mathbf{a}_1^H \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1} \mathbf{P}_{\mathbf{q}}^{-1} \mathbf{a}_1}_{\text{MVDR filter}}$$

Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

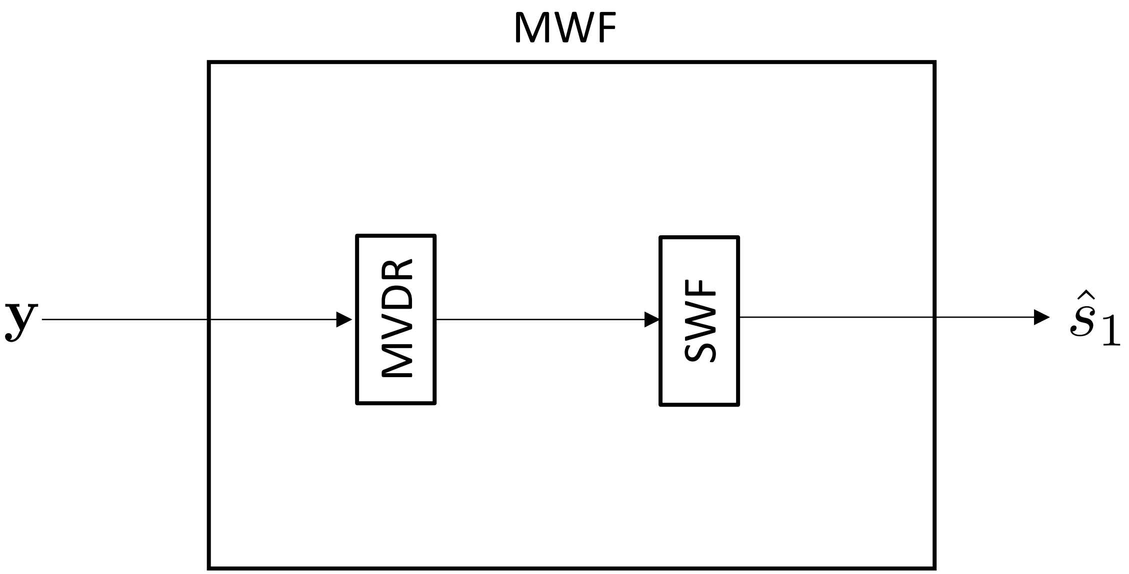
stochastic

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[\left|\mathbf{w}^H\mathbf{y} - s_1\right|^2\right] \rightarrow \hat{\mathbf{w}} = p_1 \mathbf{P}_{\mathbf{y}}^{-1}\mathbf{a}_1$$

$$= p_1\left(p_1\mathbf{a}_1\mathbf{a}_1^H + \mathbf{P}_{\mathbf{q}}\right)^{-1}\mathbf{a}_1$$

$$= \boxed{\frac{p_1}{p_1 + \underbrace{\left(\mathbf{a}_1^H\mathbf{P}_{\mathbf{q}}^{-1}\mathbf{a}_1\right)^{-1}}_{\text{MVDR output noise power}}}} \underbrace{\frac{1}{\mathbf{a}_1^H\mathbf{P}_{\mathbf{q}}^{-1}\mathbf{a}_1}\mathbf{P}_{\mathbf{q}}^{-1}\mathbf{a}_1}_{\text{MVDR filter}}$$

single channel Wiener filter (SWF)



Multi-microphone noise reduction: multi-channel Wiener filter (MWF) (Brandstein 2001)

Estimation of $\mathbf{w}$ with an unconstrained LMSE optimization problem.

stochastic

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \mathrm{E}\left[|\mathbf{w}^H \mathbf{y} - s_1|^2\right] \rightarrow \hat{\mathbf{w}} = p_1 \mathbf{P}_{\mathbf{y}}^{-1} \mathbf{a}_1$$

- Typically, the MWF filter achieves extra noise reduction compared to the MVDR filter.
- Unlike the MVDR filter, the MWF filter distorts the target signal. This is due to the fact that

$$\mathbf{w}^H \mathbf{a}_1 = p_1 \mathbf{a}_1^H \mathbf{P}_{\mathbf{y}}^{-1} \mathbf{a}_1 \neq 1$$

Multi-microphone source separation

Multi-microphone source separation

- Typically, in multi-microphone source separation all point sources are considered targets. That is,

$$\mathbf{y} = \underbrace{\sum_{i=1}^{r} \mathbf{a}_i s_i}_{\text{targets}} + \underbrace{\mathbf{n}}_{\text{noise}} \rightarrow \mathbf{y} = \underbrace{\mathbf{As}}_{\text{targets}} + \underbrace{\mathbf{n}}_{\text{noise}}$$

- The goal is to estimate the targets $\mathbf{s}$ from the noisy DFT coefficients $\mathbf{y}$. That is,

$$\hat{\mathbf{s}} = f(\mathbf{y}),$$

  where $f(.)$ is the filter function, where in this presentation is linear. That is,

$$\hat{\mathbf{s}} = \mathbf{W}^H \mathbf{y},$$

  where $\mathbf{W}$ is the filter matrix.

- We want to find a $\mathbf{W}$ such that $\hat{\mathbf{s}} = \underbrace{\mathbf{W}^H \mathbf{As}}_{\approx \mathbf{s}} + \underbrace{\mathbf{W}^H \mathbf{n}}_{\approx 0},$

Multi-microphone source separation

- Typically, in multi-microphone source separation all point sources are considered targets. That is,

$$\mathbf{y} = \underbrace{\sum_{i=1}^{r} \mathbf{a}_i s_i}_{\text{targets}} + \underbrace{\mathbf{n}}_{\text{noise}} \rightarrow \mathbf{y} = \underbrace{\mathbf{A}\mathbf{s}}_{\text{targets}} + \underbrace{\mathbf{n}}_{\text{noise}}$$

- The goal is to estimate the targets $\mathbf{s}$ from the noisy DFT coefficients $\mathbf{y}$. That is,

$$\hat{\mathbf{s}} = f(\mathbf{y}),$$

where $f(.)$ is the filter function, where in this presentation is linear. That is,

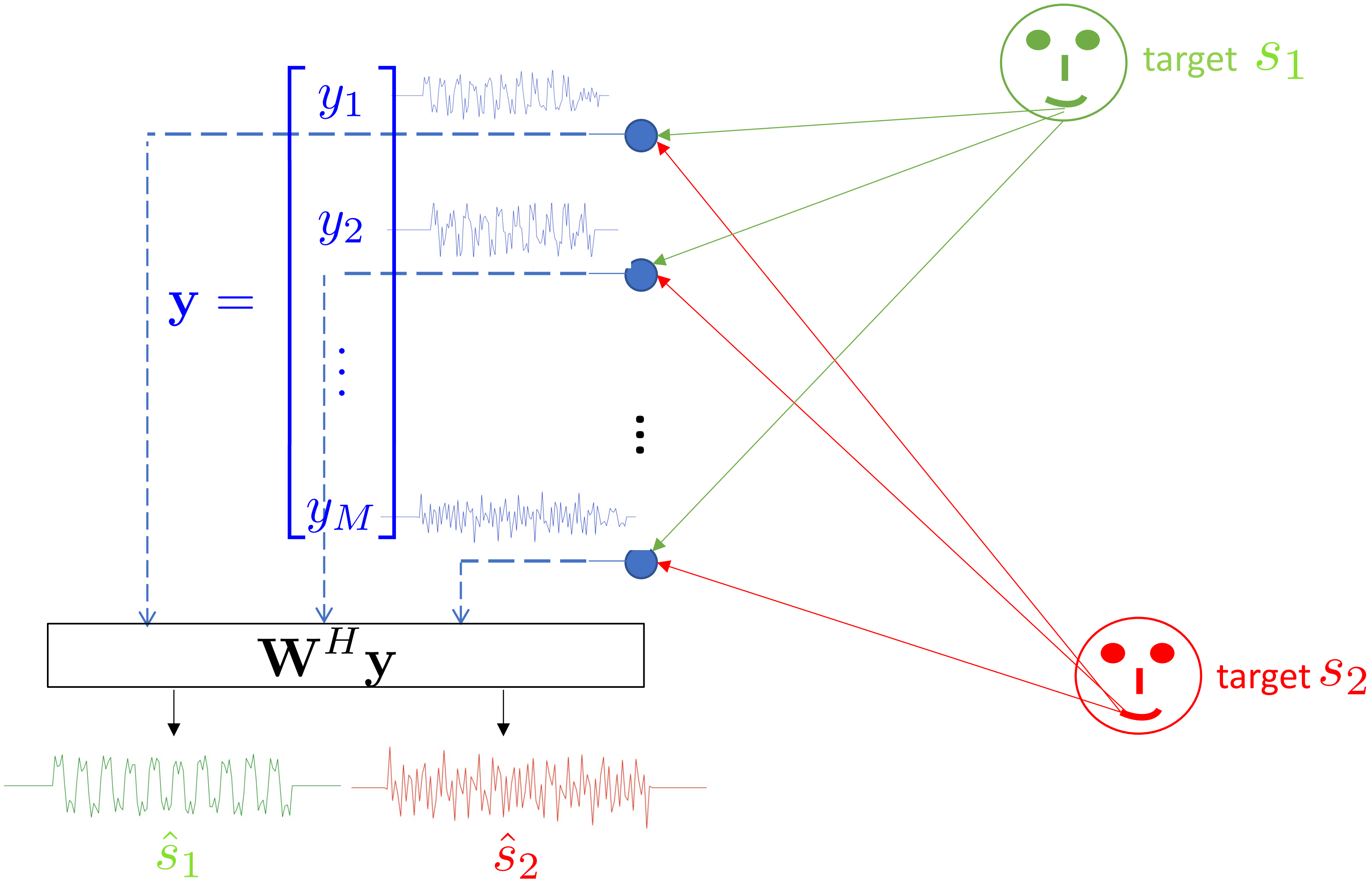
$$\hat{\mathbf{s}} = \mathbf{W}^H \mathbf{y},$$

$$\mathbf{W}^H \mathbf{A} = \mathbf{I}$$

where $\mathbf{W}$ is the filter matrix.

- We want to find a $\mathbf{W}$ such that $\hat{\mathbf{s}} = \underbrace{\mathbf{W}^H \mathbf{A}\mathbf{s}}_{\approx \mathbf{s}} + \underbrace{\mathbf{W}^H \mathbf{n}}_{\approx 0},$

Multi-microphone source separation



Multi-microphone source separation

deterministic

1st Method
Least squares

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{y} - \mathbf{A}\mathbf{s}||_2^2 \rightarrow \hat{\mathbf{s}} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{y}$$

deterministic

2nd Method
Weighted least squares

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{P}_{\mathbf{n}}^{-\frac{1}{2}}\left(\mathbf{y} - \mathbf{A}\mathbf{s}\right)||_2^2 \rightarrow \hat{\mathbf{s}} = \left(\mathbf{A}^H\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{y}$$

stochastic

3d Method
MMSE

$$\hat{\mathbf{s}} = \arg\min_{\hat{\mathbf{s}}} \mathrm{E}[||\hat{\mathbf{s}} - \mathbf{s}||_2^2] = \mathrm{E}\left[\mathbf{s}|\mathbf{y}\right] \xrightarrow[]{\substack{\mathbf{s} \sim N(\mathbf{0}, \mathbf{P}) \\ \mathbf{n} \sim N(\mathbf{0}, \mathbf{P}_{\mathbf{n}})}} \hat{\mathbf{s}} = \mathbf{P}\mathbf{A}^H\mathbf{P}_{\mathbf{y}}^{-1}\mathbf{y}$$

Multi-microphone source separation

$$\mathbf{W}^H\mathbf{A} = \mathbf{I}$$

1st Method
Least squares

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{y} - \underbrace{\mathbf{A}\mathbf{s}}_{\text{deterministic}}||_2^2 \rightarrow \hat{\mathbf{s}} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{y}$$

2nd Method
Weighted least squares

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{P}_{\mathbf{n}}^{-\frac{1}{2}}\left(\mathbf{y} - \underbrace{\mathbf{A}\mathbf{s}}_{\text{deterministic}}\right)||_2^2 \rightarrow \hat{\mathbf{s}} = \left(\mathbf{A}^H\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{P}_{\mathbf{n}}^{-1}\mathbf{y}$$

3d Method
MMSE

$$\hat{\mathbf{s}} = \arg\min_{\hat{\mathbf{s}}} \mathrm{E}[||\hat{\mathbf{s}} - \underbrace{\mathbf{s}}_{\text{stochastic}}||_2^2] = \mathrm{E}\left[\mathbf{s}|\mathbf{y}\right] \xrightarrow[\mathbf{n}\sim N(\mathbf{0},\mathbf{P}_{\mathbf{n}})]{\mathbf{s}\sim N(\mathbf{0},\mathbf{P})} \hat{\mathbf{s}} = \mathbf{P}\mathbf{A}^H\mathbf{P}_{\mathbf{y}}{}^{-1}\mathbf{y}$$

Estimation of multi-microphone signal model parameters

Maximum likelihood estimation of $\mathbf{P}_{\mathbf{y}}$

$$\hat{\mathbf{P}}_{\mathbf{y}} = \arg\min_{\mathbf{P}_{\mathbf{y}}} \; - \ln p\left(\{\mathbf{y}_1, \cdots, \mathbf{y}_N\} | \mathbf{P}_{\mathbf{y}}\right)$$

Maximum likelihood estimation of $\mathbf{P_y}$

$$\mathbf{y}_i,\ i = 1, \cdots, N, \quad \text{indipendent}$$

$$\begin{aligned}
\hat{\mathbf{P}}_{\mathbf{y}} &= \arg\min_{\mathbf{P}_{\mathbf{y}}} \ -\ln p\left(\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}|\mathbf{P}_{\mathbf{y}}\right) \\
&= \arg\min_{\mathbf{P}_{\mathbf{y}}} \ -\ln \prod_{i=1}^{N} p\left(\mathbf{y}_i|\mathbf{P}_{\mathbf{y}}\right) \\
&= \arg\min_{\mathbf{P}_{\mathbf{y}}} \ -\sum_{i=1}^{N} \ln p\left(\mathbf{y}_i|\mathbf{P}_{\mathbf{y}}\right)
\end{aligned}$$

Maximum likelihood estimation of $\mathbf{P_y}$

$$\text{iid } \mathbf{y}_i \sim \mathcal{N}(\mathbf{0},\, \mathbf{P_y})$$

$$p\left(\mathbf{y}_i|\mathbf{P_y}\right) = \frac{1}{|\pi \mathbf{P_y}|} \exp\left(-\mathbf{y}_i^H \mathbf{P_y}^{-1} \mathbf{y}_i\right)$$

$$\mathbf{y}_i,\ i = 1, \cdots, N, \quad \text{indipendent}$$

$$\begin{aligned}
\hat{\mathbf{P}}_\mathbf{y} &= \underset{\mathbf{P_y}}{\arg\min}\ -\ln p\left(\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}|\mathbf{P_y}\right) \\
&= \underset{\mathbf{P_y}}{\arg\min}\ -\ln \prod_{i=1}^{N} p\left(\mathbf{y}_i|\mathbf{P_y}\right) \\
&= \underset{\mathbf{P_y}}{\arg\min}\ -\sum_{i=1}^{N} \ln p\left(\mathbf{y}_i|\mathbf{P_y}\right) \\
&= \underset{\mathbf{P_y}}{\arg\min}\ -\sum_{i=1}^{N} \left[\ln(1) - \ln\left(\pi^M\right) - \ln\left(|\mathbf{P_y}|\right) + \ln\left(\exp\left(-\mathbf{y}_i^H \mathbf{P_y}^{-1} \mathbf{y}_i\right)\right)\right]
\end{aligned}$$

Maximum likelihood estimation of $\mathbf{P_y}$

$$\text{iid } \mathbf{y}_i \sim \mathcal{N}(\mathbf{0},\, \mathbf{P_y})$$

$$p\left(\mathbf{y}_i | \mathbf{P_y}\right) = \frac{1}{|\pi \mathbf{P_y}|} \exp\left(-\mathbf{y}_i^H \mathbf{P_y}^{-1} \mathbf{y}_i\right)$$

$$\mathbf{y}_i,\ i = 1, \cdots, N, \quad \text{indipendent}$$

$$\hat{\mathbf{P}}_\mathbf{y} = \arg\min_{\mathbf{P_y}} \ -\ln p\left(\{\mathbf{y}_1, \cdots, \mathbf{y}_N\} | \mathbf{P_y}\right)$$

$$= \arg\min_{\mathbf{P_y}} \ -\ln \prod_{i=1}^{N} p\left(\mathbf{y}_i | \mathbf{P_y}\right)$$

$$= \arg\min_{\mathbf{P_y}} \ -\sum_{i=1}^{N} \ln p\left(\mathbf{y}_i | \mathbf{P_y}\right)$$

$$= \arg\min_{\mathbf{P_y}} \ \sum_{i=1}^{N} \left[\ln\left(|\mathbf{P_y}|\right) + \mathbf{y}_i^H \mathbf{P_y}^{-1} \mathbf{y}_i\right]$$

Maximum likelihood estimation of $\mathbf{P}_{\mathbf{y}}$

$$\text{iid } \mathbf{y}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{\mathbf{y}})$$

$$\mathbf{y}_i,\ i = 1, \cdots, N, \quad \text{indipendent}$$

$$p\left(\mathbf{y}_i | \mathbf{P}_{\mathbf{y}}\right) = \frac{1}{|\pi \mathbf{P}_{\mathbf{y}}|} \exp\left(-\mathbf{y}_i^H \mathbf{P}_{\mathbf{y}}^{-1} \mathbf{y}_i\right)$$

$$\frac{\partial}{\partial \hat{\mathbf{P}}_{\mathbf{y}}} \sum_{i=1}^{N} \left[\ln\left(|\hat{\mathbf{P}}_{\mathbf{y}}|\right) + \mathbf{y}_i^H \hat{\mathbf{P}}_{\mathbf{y}}^{-1} \mathbf{y}_i\right] = \mathbf{0} \rightarrow \sum_{i=1}^{N} \left[\frac{\partial}{\partial \hat{\mathbf{P}}_{\mathbf{y}}} \ln\left(|\hat{\mathbf{P}}_{\mathbf{y}}|\right) + \frac{\partial}{\partial \hat{\mathbf{P}}_{\mathbf{y}}} \mathbf{y}_i^H \hat{\mathbf{P}}_{\mathbf{y}}^{-1} \mathbf{y}_i\right] = \mathbf{0} \rightarrow$$

$$\rightarrow \sum_{i=1}^{N} \left[\hat{\mathbf{P}}_{\mathbf{y}}^{-1} - \hat{\mathbf{P}}_{\mathbf{y}}^{-1} \mathbf{y}_i \mathbf{y}_i^H \hat{\mathbf{P}}_{\mathbf{y}}^{-1}\right] = \mathbf{0} \rightarrow \sum_{i=1}^{N} \left[\hat{\mathbf{P}}_{\mathbf{y}}^{-1} - \hat{\mathbf{P}}_{\mathbf{y}}^{-1} \mathbf{y}_i \mathbf{y}_i^H \hat{\mathbf{P}}_{\mathbf{y}}^{-1}\right] \hat{\mathbf{P}}_{\mathbf{y}} = \mathbf{0} \hat{\mathbf{P}}_{\mathbf{y}} \rightarrow$$

$$\rightarrow \sum_{i=1}^{N} \left[\mathbf{I} - \hat{\mathbf{P}}_{\mathbf{y}}^{-1} \mathbf{y}_i \mathbf{y}_i^H\right] = \mathbf{0} \rightarrow \hat{\mathbf{P}}_{\mathbf{y}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i \mathbf{y}_i^H$$

Estimation of multi-microphone signal model parameters $\hat{r}, \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$

$$\boxed{\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{r}, \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}}$$

$$\downarrow$$

$$\hat{\mathbf{S}}$$

Estimation of multi-microphone signal model parameters $\hat{r}, \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$

$$\hat{\mathbf{P}}_{\mathbf{y}}$$

exploratory factor analysis (EFA)

$$\hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$$

where $\mathbf{\Phi} = \mathbf{A}\mathbf{P}\mathbf{A}^H$

Estimation of multi-microphone signal model parameters $\hat{r}, \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$

$$\hat{\mathbf{P}}_{\mathbf{y}}$$

exploratory factor analysis (EFA)

$$\hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$$

where $\mathbf{\Phi} = \mathbf{A}\mathbf{P}\mathbf{A}^H$

$$\downarrow$$

$$\hat{r} = \mathrm{rank}\left(\hat{\mathbf{\Phi}}\right)$$

Estimation of multi-microphone signal model parameters $\hat{r}, \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$

$$
\begin{tikzcd}
 & \hat{\mathbf{P}}_{\mathbf{y}} \arrow[dl, "\text{exploratory factor analysis (EFA)}"'] \arrow[dr, "\text{confirmatory factor analysis (CFA)}"] & \\
\hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} \arrow[rr, "\hat{r}"] & & \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}
\end{tikzcd}
$$

where $\mathbf{\Phi} = \mathbf{A}\mathbf{P}\mathbf{A}^H$

$$\downarrow$$

$$\hat{r} = \mathrm{rank}\left(\hat{\mathbf{\Phi}}\right)$$

Exploratory Factor Analysis (EFA) $\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{\boldsymbol{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$

Minimum rank factor analysis (Lederman 1940, Ten Berge et al. 1991):

$$\begin{aligned}\hat{\boldsymbol{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} = \arg \min_{\boldsymbol{\Phi}, \mathbf{P}_{\mathbf{n}}} \operatorname{rank}(\boldsymbol{\Phi}) \\ \text{s.t. } & \hat{\mathbf{P}}_{\mathbf{y}} = \boldsymbol{\Phi} + \mathbf{P}_{\mathbf{n}} \\ & \boldsymbol{\Phi} \succeq 0 \\ & \mathbf{P}_{\mathbf{n}} = \operatorname{Diag}(p_{n1}, p_{n2}, \cdots, p_{n2}) \succeq 0.\end{aligned}$$

This is a non-convex optimization problem.

Exploratory Factor Analysis (EFA) $\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{\boldsymbol{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$

eigenvalues

$\boldsymbol{\Lambda} = \mathrm{Diag}\left(\lambda_1, \cdots, \lambda_r\right)$

$\lambda_i \geq 0$

eigenvectors

$e_i^H e_j = 0$

$\boldsymbol{\Phi} = \mathbf{A}\mathbf{P}\mathbf{A}^H = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H$

Minimum rank factor analysis (Lederman 1940, Ten Berge et al. 1991):

$$\hat{\boldsymbol{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} = \underset{\boldsymbol{\Phi}, \mathbf{P}_{\mathbf{n}}}{\arg\min}\ \mathrm{rank}\left(\boldsymbol{\Phi}\right)$$

$$\text{s.t. } \hat{\mathbf{P}}_{\mathbf{y}} = \boldsymbol{\Phi} + \mathbf{P}_{\mathbf{n}}$$

$$\boldsymbol{\Phi} \succeq 0$$

$$\mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{n2}) \succeq 0.$$

This is a non-convex optimization problem.

Exploratory Factor Analysis (EFA) $\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$

eigenvalues

$\mathbf{\Lambda} = \mathrm{Diag}\left(\lambda_1, \cdots, \lambda_r\right)$

$\lambda_i \geq 0$

eigenvectors

$e_i^H e_j = 0$

Minimum rank factor analysis (Lederman 1940, Ten Berge et al. 1991):

$$\mathbf{\Phi} = \mathbf{A}\mathbf{P}\mathbf{A}^H = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H$$

$$\begin{aligned}\hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} = \arg\min_{\mathbf{\Phi}, \mathbf{P}_{\mathbf{n}}} & \ \mathrm{rank}\left(\mathbf{\Phi}\right) \\ \text{s.t.} & \ \hat{\mathbf{P}}_{\mathbf{y}} = \mathbf{\Phi} + \mathbf{P}_{\mathbf{n}} \\ & \ \mathbf{\Phi} \succeq 0 \\ & \ \mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{n2}) \succeq 0.\end{aligned}$$

$$\mathrm{rank}\left(\mathbf{\Phi}\right) = ||\boldsymbol{\lambda}||_0 \qquad \boldsymbol{\lambda} = \begin{bmatrix}\lambda_1 \\ \vdots \\ \lambda r\end{bmatrix}$$

This is a non-convex optimization problem.

Exploratory Factor Analysis (EFA) $\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$

eigenvalues

$$\mathbf{\Lambda} = \mathrm{Diag}\left(\lambda_1, \cdots, \lambda_r\right)$$

$$\lambda_i \geq 0$$

eigenvectors

$$e_i^H e_j = 0$$

Minimum trace factor analysis (Lederman 1940, Della Riccia et al. 1982):

$$\mathbf{\Phi} = \mathbf{A}\mathbf{P}\mathbf{A}^H = \mathbf{E}\mathbf{\Lambda}\mathbf{E}^H$$

$$\begin{aligned} \hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} = \arg \min_{\mathbf{\Phi}, \mathbf{P}_{\mathbf{n}}} \ & \mathrm{trace}\left(\mathbf{\Phi}\right) \\ \text{s.t.} \ & \hat{\mathbf{P}}_{\mathbf{y}} = \mathbf{\Phi} + \mathbf{P}_{\mathbf{n}} \\ & \mathbf{\Phi} \succeq 0 \\ & \mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{n2}) \succeq 0. \end{aligned}$$

This is a convex optimization problem. The global optimum will be obtained for sure.

Exploratory Factor Analysis (EFA) $\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{\boldsymbol{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$

eigenvalues

$\boldsymbol{\Lambda} = \mathrm{Diag}\left(\lambda_1, \cdots, \lambda_r\right)$

$\lambda_i \geq 0$

eigenvectors

$e_i^H e_j = 0$

Minimum trace factor analysis (Lederman 1940, Della Riccia et al. 1982):

$$\boldsymbol{\Phi} = \mathbf{APA}^H = \mathbf{E\Lambda E}^H$$

$$\hat{\boldsymbol{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} = \arg\min_{\boldsymbol{\Phi}, \mathbf{P}_{\mathbf{n}}} \mathrm{trace}\left(\boldsymbol{\Phi}\right)$$

$$\mathrm{trace}\left(\boldsymbol{\Phi}\right) = \sum_{i=1}^{M} \lambda_i = ||\boldsymbol{\lambda}||_1 \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda r \end{bmatrix}$$

$$\text{s.t. } \hat{\mathbf{P}}_{\mathbf{y}} = \boldsymbol{\Phi} + \mathbf{P}_{\mathbf{n}}$$

$$\boldsymbol{\Phi} \succeq 0$$

$$\mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{n2}) \succeq 0.$$

This is a convex optimization problem. The global optimum will be obtained for sure.

Exploratory Factor Analysis (EFA) $\hat{\mathbf{P}}_{\mathbf{y}} \xrightarrow{\text{estimate}} \hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}}$

Minimum trace factor analysis (Lederman 1940, Della Riccia et al. 1982):

$$\hat{\mathbf{\Phi}}, \hat{\mathbf{P}}_{\mathbf{n}} = \arg\min_{\mathbf{\Phi}, \mathbf{P}_{\mathbf{n}}} \operatorname{trace}(\mathbf{\Phi})$$

$$\text{s.t. } \hat{\mathbf{P}}_{\mathbf{y}} = \mathbf{\Phi} + \mathbf{P}_{\mathbf{n}}$$

$$\mathbf{\Phi} \succeq 0$$

$$\mathbf{P}_{\mathbf{n}} = \operatorname{Diag}(p_{n1}, p_{n2}, \cdots, p_{n2}) \succeq 0.$$

$\hat{r} = \operatorname{rank}\left(\hat{\mathbf{\Phi}}\right)$: is not a good estimate of number of sources

Scree Test (Cattell 1966) $\hat{\boldsymbol{\Phi}} \xrightarrow{\text{estimate}} \hat{r}$

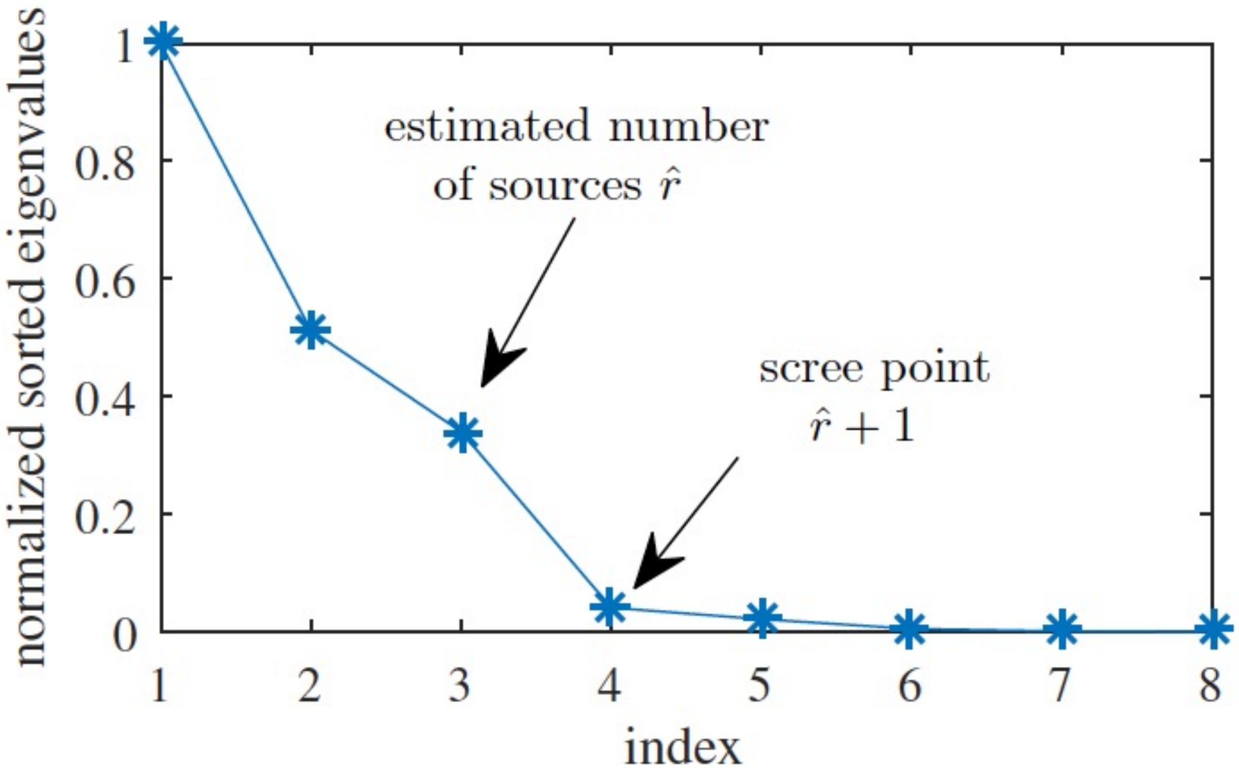


Figure: Normalized eigenvalues of $\hat{\boldsymbol{\Phi}}$ in descending order. Source: Koutrouvelis et al. 2019

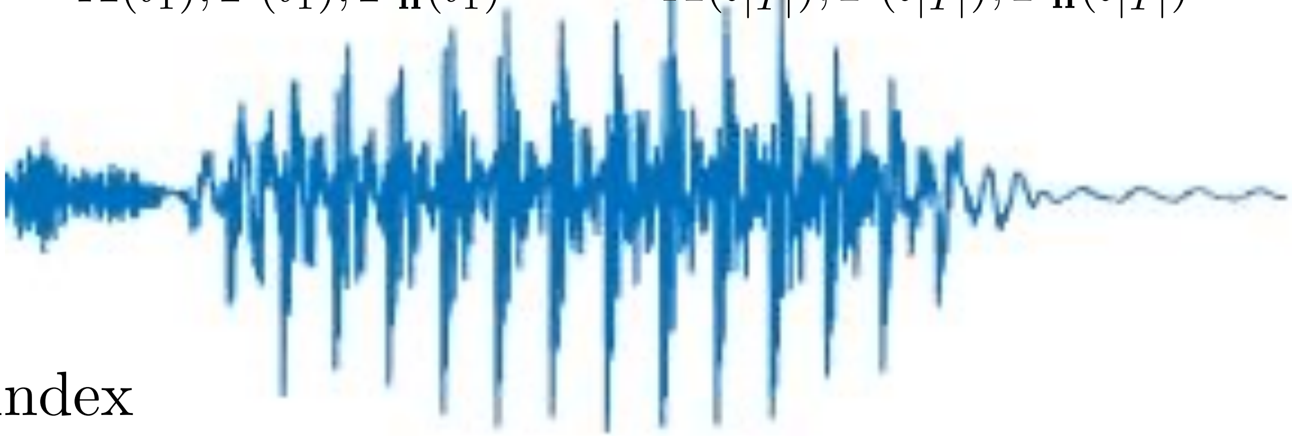
Confirmatory Factor Analysis (CFA) $\hat{\mathbf{P}}_{\mathbf{y}}, \hat{r} \xrightarrow{\text{estimate}} \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$

$\longleftarrow$ time segment $\longrightarrow$

$\longleftarrow$ time frame $\longrightarrow$ $\cdots$ $\longleftarrow$ time frame $\longrightarrow$

$\hat{\mathbf{P}}_{\mathbf{y}}(t_1), \hat{r}(t_1)$ $\qquad$ $\hat{\mathbf{P}}_{\mathbf{y}}(t_{|T|}), \hat{r}(t_{|T|})$

$\hat{\mathbf{A}}(t_1), \hat{\mathbf{P}}(t_1), \hat{\mathbf{P}}_{\mathbf{n}}(t_1)$ $\qquad$ $\hat{\mathbf{A}}(t_{|T|}), \hat{\mathbf{P}}(t_{|T|}), \hat{\mathbf{P}}_{\mathbf{n}}(t_{|T|})$



t : time frame index

$\mathcal{T}$: set of all time frames within the time segment, $t \in \mathcal{T}$

Confirmatory Factor Analysis (CFA) $\hat{\mathbf{P}}_{\mathbf{y}}, \hat{r} \xrightarrow{\text{estimate}} \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$

$\longleftarrow$ time segment $\longrightarrow$

$\longleftarrow$ time frame $\longrightarrow$ $\cdots$ $\longleftarrow$ time frame $\longrightarrow$

$\hat{\mathbf{P}}_{\mathbf{y}}(t_1), \hat{r}$ $\qquad$ $\hat{\mathbf{P}}_{\mathbf{y}}(t_{|T|}), \hat{r}$

$\hat{\mathbf{A}}, \hat{\mathbf{P}}(t_1), \hat{\mathbf{P}}_{\mathbf{n}}$ $\qquad$ $\hat{\mathbf{A}}, \hat{\mathbf{P}}(t_{|T|}), \hat{\mathbf{P}}_{\mathbf{n}}$

reduction of #unknowns:

$\mathbf{A}, \mathbf{P}_{\mathbf{n}}$ are the same $\forall t \in \mathcal{T}$

t : time frame index

$\mathcal{T}$: set of all time frames within the time segment, $t \in \mathcal{T}$

Confirmatory Factor Analysis (CFA) $\hat{\mathbf{P}}_{\mathbf{y}}, \hat{r} \xrightarrow{\text{estimate}} \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$ (Jöreskog 1969, Parra et al. 2000, Koutrouvelis et al. 2019)

$$\hat{\mathbf{A}}, \{\hat{\mathbf{P}}(t) : t \in \mathcal{T}\}, \hat{\mathbf{P}}_{\mathbf{n}} = \underset{\substack{\mathbf{A},\mathbf{P}_{\mathbf{n}},\\ \{\mathbf{P}(t):t\in\mathcal{T}\}}}{\arg\min} \sum_{\forall \tau \in \mathcal{T}} F(\hat{\mathbf{P}}_{\mathbf{y}}(\tau), \mathbf{P}_{\mathbf{y}}(\tau))$$

$$\begin{aligned} \text{s.t. } & \mathbf{P}_{\mathbf{y}}(t) = \mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{P}_{\mathbf{n}}, \ \forall t \in \mathcal{T} \\ & \mathbf{P}(t) = \mathrm{Diag}(p_1(t), p_2(t), \cdots, p_M(t)) \succeq 0 \\ & \mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{nM}) \succeq 0 \end{aligned}$$

$$F(\hat{\mathbf{P}}_{\mathbf{y}}, \mathbf{P}_{\mathbf{y}}) = \begin{cases} \text{(ML): } \log|\mathbf{P}_{\mathbf{y}}| + \mathrm{tr}\left(\hat{\mathbf{P}}_{\mathbf{y}}\mathbf{P}_{\mathbf{y}}^{-1}\right), \\ \text{(LS): } \frac{1}{2}||\mathbf{P}_{\mathbf{y}} - \hat{\mathbf{P}}_{\mathbf{y}}||_F^2, \\ \text{(GLS): } \frac{1}{2}||\hat{\mathbf{P}}_{\mathbf{y}}^{-\frac{1}{2}}(\mathbf{P}_{\mathbf{y}} - \hat{\mathbf{P}}_{\mathbf{y}})\hat{\mathbf{P}}_{\mathbf{y}}^{-\frac{1}{2}}||_F^2. \end{cases}$$

(Mulaik 2009)

Confirmatory Factor Analysis (CFA) $\hat{\mathbf{P}}_{\mathbf{y}}, \hat{r} \xrightarrow{\text{estimate}} \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$ (Jöreskog 1969, Parra et al. 2000, Koutrouvelis et al. 2019)

$$\hat{\mathbf{A}}, \{\hat{\mathbf{P}}(t) : t \in \mathcal{T}\}, \hat{\mathbf{P}}_{\mathbf{n}} = \underset{\substack{\mathbf{A},\mathbf{P}_{\mathbf{n}}, \\ \{\mathbf{P}(t): t \in \mathcal{T}\}}}{\arg\min} \sum_{\forall \tau \in \mathcal{T}} F(\hat{\mathbf{P}}_{\mathbf{y}}(\tau), \mathbf{P}_{\mathbf{y}}(\tau))$$

$$\begin{aligned} \text{s.t. } \mathbf{P}_{\mathbf{y}}(t) &= \mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{P}_{\mathbf{n}}, \ \forall t \in \mathcal{T} \\ \mathbf{P}(t) &= \mathrm{Diag}(p_1(t), p_2(t), \cdots, p_M(t)) \succeq 0 \\ \mathbf{P}_{\mathbf{n}} &= \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{nM}) \succeq 0 \end{aligned}$$

$$\hat{\mathbf{A}} \approx \mathbf{A}\mathbf{T}^{-1}, \hat{\mathbf{P}}(t) \approx \mathbf{T}\mathbf{P}(t)\mathbf{T}^H$$

$$\mathbf{T} = \mathbf{\Pi}\mathbf{S}$$

$$\text{permutation, scaling}$$

Confirmatory Factor Analysis (CFA) $\hat{\mathbf{P}}_{\mathbf{y}}, \hat{r} \xrightarrow{\text{estimate}} \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$ (Jöreskog 1969, Parra et al. 2000, Koutrouvelis et al. 2019)

$$\hat{\mathbf{A}}, \{\hat{\mathbf{P}}(t) : t \in \mathcal{T}\}, \hat{\mathbf{P}}_{\mathbf{n}} = \underset{\substack{\mathbf{A},\mathbf{P}_{\mathbf{n}}, \\ \{\mathbf{P}(t):t\in\mathcal{T}\}}}{\arg\min} \sum_{\forall \tau \in \mathcal{T}} F(\hat{\mathbf{P}}_{\mathbf{y}}(\tau), \mathbf{P}_{\mathbf{y}}(\tau))$$

$$\begin{aligned} \text{s.t.}\ & \mathbf{P}_{\mathbf{y}}(t) = \mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{P}_{\mathbf{n}},\ \forall t \in \mathcal{T} \\ & \mathbf{P}(t) = \mathrm{Diag}(p_1(t), p_2(t), \cdots, p_M(t)) \succeq 0 \\ & \mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{nM}) \succeq 0 \\ & a_{\rho i} = 1,\ \text{for } i = 1, \cdots, r. \end{aligned}$$

We select a reference microphone ρ

$$\hat{\mathbf{A}} \approx \mathbf{A}\mathbf{T}^{-1}, \hat{\mathbf{P}}(t) \approx \mathbf{T}\mathbf{P}(t)\mathbf{T}^H$$

$$\mathbf{T} = \color{red}{\mathbf{\Pi}}$$

$$\color{red}{\text{permutation}}$$

Confirmatory Factor Analysis (CFA) $\hat{\mathbf{P}}_{\mathbf{y}}, \hat{r} \xrightarrow{\text{estimate}} \hat{\mathbf{A}}, \hat{\mathbf{P}}, \hat{\mathbf{P}}_{\mathbf{n}}$ (Jöreskog 1969, Parra et al. 2000, Koutrouvelis et al. 2019)

$$\hat{\mathbf{A}}, \{\hat{\mathbf{P}}(t) : t \in \mathcal{T}\}, \hat{\mathbf{P}}_{\mathbf{n}} = \underset{\substack{\mathbf{A},\mathbf{P}_{\mathbf{n}},\\ \{\mathbf{P}(t):t\in\mathcal{T}\}}}{\arg\min} \sum_{\forall \tau \in \mathcal{T}} F(\hat{\mathbf{P}}_{\mathbf{y}}(\tau), \mathbf{P}_{\mathbf{y}}(\tau))$$

$$\begin{aligned} \text{s.t. } & \mathbf{P}_{\mathbf{y}}(t) = \mathbf{A}\mathbf{P}(t)\mathbf{A}^H + \mathbf{P}_{\mathbf{n}}, \ \forall t \in \mathcal{T} \\ & \mathbf{P}(t) = \mathrm{Diag}(p_1(t), p_2(t), \cdots, p_M(t)) \succeq 0 \\ & \mathbf{P}_{\mathbf{n}} = \mathrm{Diag}(p_{n1}, p_{n2}, \cdots, p_{nM}) \succeq 0 \\ & a_{\rho i} = 1, \ \text{for } i = 1, \cdots, r. \end{aligned}$$

Relative acoustic transfer functions

$$\hat{\mathbf{A}} \approx \mathbf{A}\mathbf{T}^{-1}, \hat{\mathbf{P}}(t) \approx \mathbf{T}\mathbf{P}(t)\mathbf{T}^H$$

$$\mathbf{T} = \mathbf{\Pi}$$

permutation

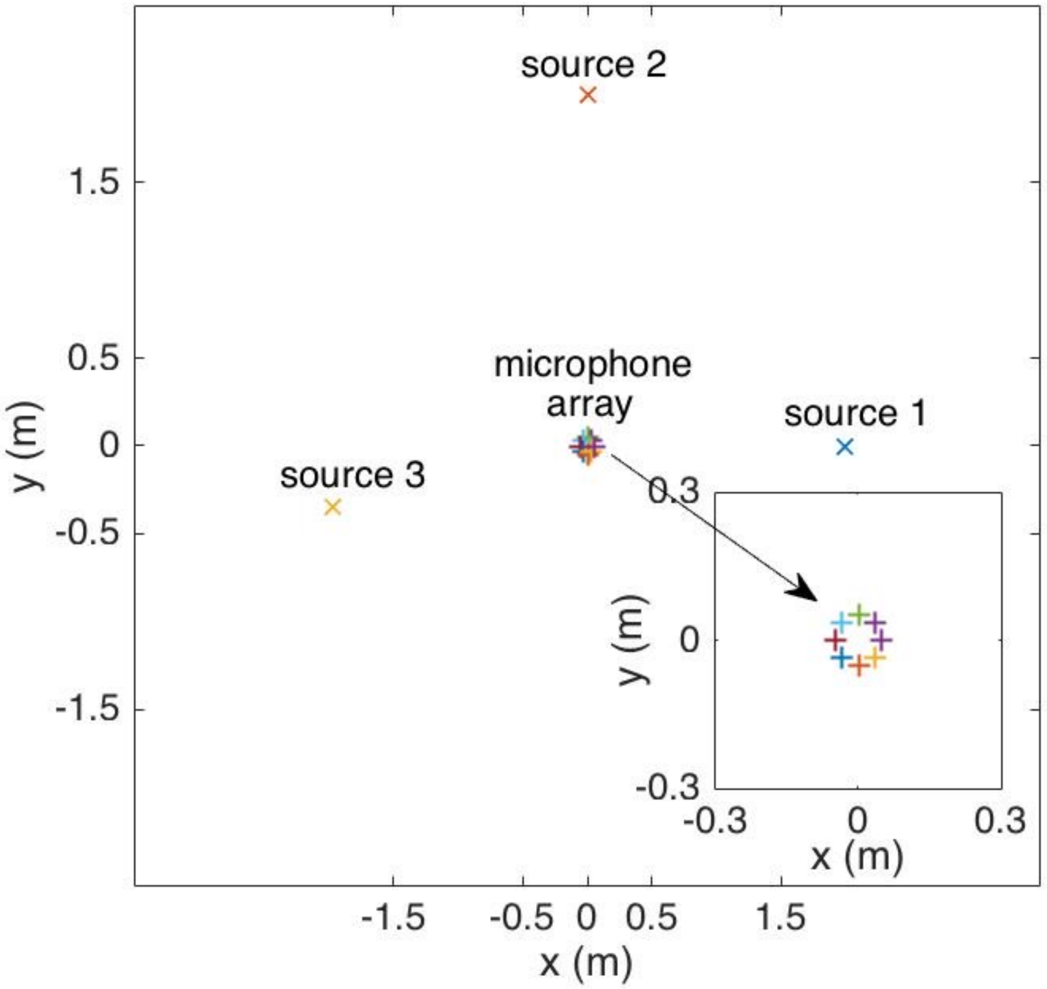
Permutation ambiguities

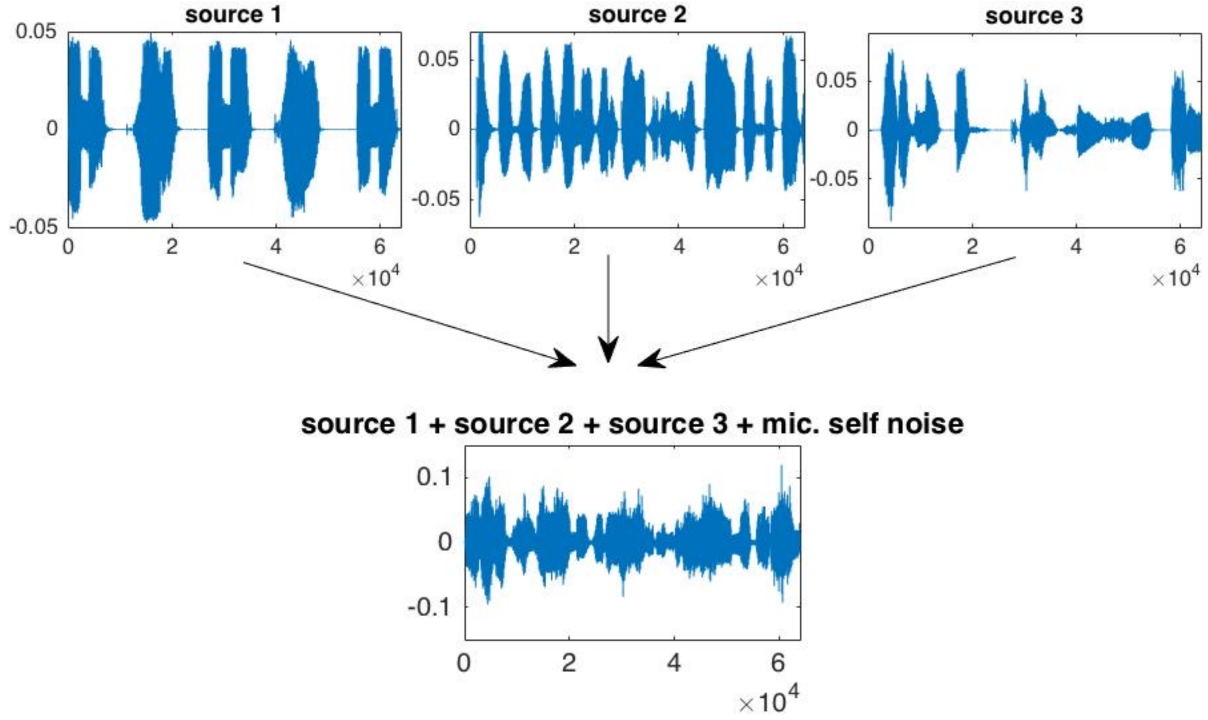
Every time-frequency bin may have a different permutation ambiguity (i.e., a different $\mathbf{\Pi}(t, k)$). There are several methods on how to solve the permutation ambiguities over time-frequency bins, e.g.,:

- R.Mukai, H. Sawada, S. Araki, and S. Makino, “Frequency-domain blind source separation of many speech signals using near-field and far-field models,” EURASIP J. Appl. Signal Process., vol. 2006, no. 1, pp. 1–13, 2006.

- D. Nion, K. Mokios, N. D. Sidiropoulos, and A. Potamianos, “Batch and adaptive parafac-based blind separation of convolutive speech mixtures,” IEEE Trans. Audio, Speech, Lang. Process., vol. 18, no. 6, pp. 1193–1207, Aug. 2010.

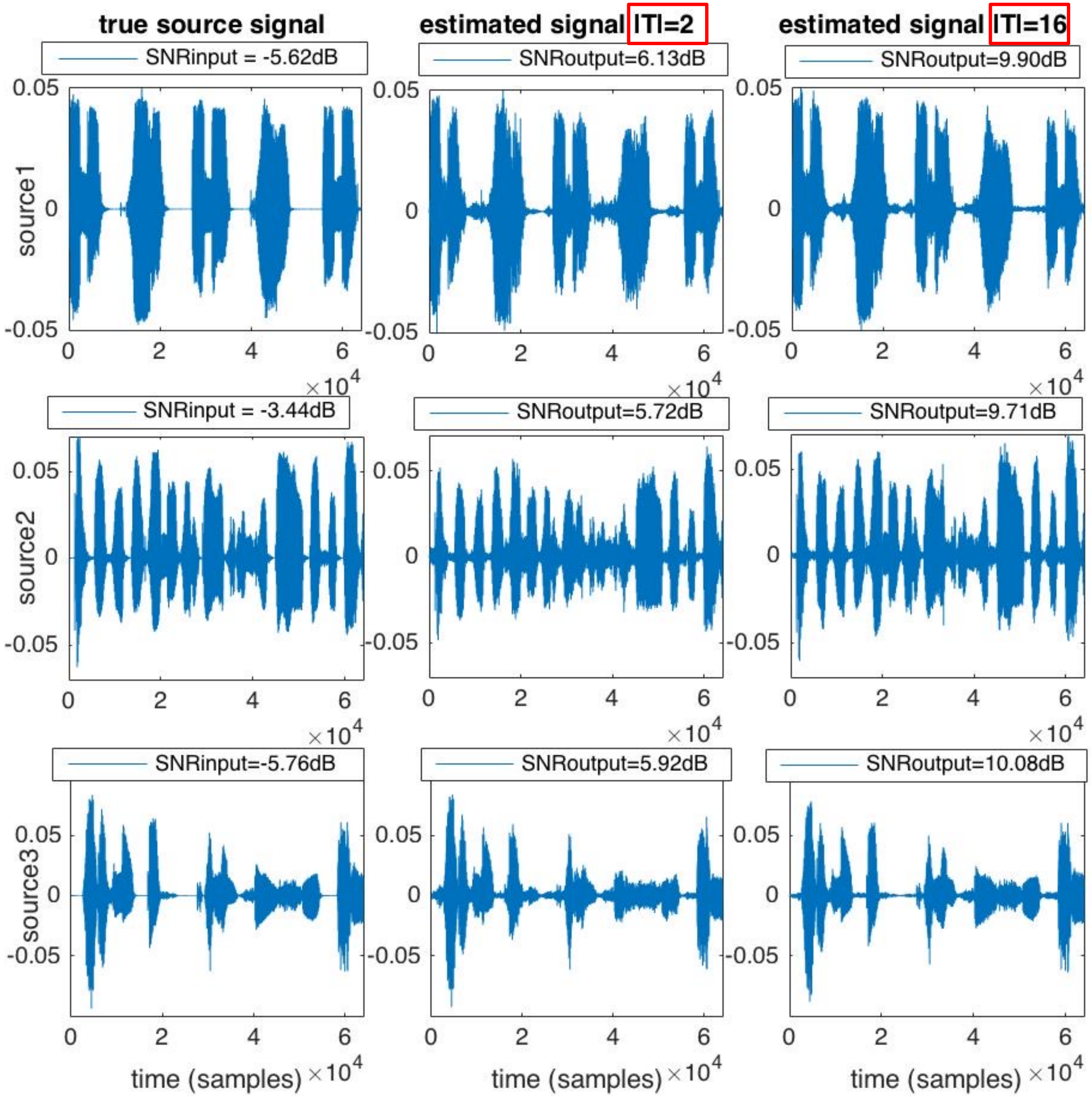
Experiments





$M = 8$ # microphones

Experiments



Typically, an increase in $|\mathcal{T}|$ implies a better performance

Estimation using EFA + CFA + MMSE

References

- J.L.Flanagan,A.C.Surendran,andE.E.Jan,“Spatially selective sound capture for speech and audio processing,” *ELSEVIER Speech Commun.*, vol. 13, no. 1-2, pp. 207–222, Oct. 1993.
- J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proc. of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- O. L. Frost III, “An algorithm for linearly constrained adaptive array processing,” *Proc. of the IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.
- M. Brandstein and D. Ward (Eds.), *Microphone Arrays: Signal Pro- cessing Techniques and Applications*. New York, NY, USA: Springer, 2001.
- L. Parra and C. Spence, “Convolutive blind separation of non-stationary sources,” IEEE Trans. Audio, Speech, Language Process., vol. 8, no. 3, pp. 320–327, 2000.
- D. N. Lawley and A. E. Maxwell, Factor Analysis as a Statistical Method. London Butterworths, 1963.
- K. G. Jöreskog, “A general approach to confirmatory maximum likelihood factor analysis,” Psychometrika, vol. 34, no. 2, pp. 183–202, 1969.
- K. G. Jöreskog, “Simultaneous factor analysis in several populations,” Psychometrika, vol. 36, no. 4, pp. 409–426, 1971.
- S. A. Mulaik, Foundations of factor analysis. CRC press, 2009.
- K. G. Jöreskog, “Factor analysis by generalized least squares,” Psychometrika, vol. 37, no. 3, pp. 243–260, 1972.
- K. G. Jöreskog and D. N. Lawley, “New methods in maximum likelihood factor analysis,” British J. Math. Statist. Psycol., vol. 21, pp. 85–96, 1968.
- A. M. Sardarabadi and A. J. van der Veen, “Complex factor analysis and extensions,” *IEEE Trans. on Signal Process.*, vol. 66, no. 4, pp. 954–967, 2018.
- A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, “Robust joint estimation of multi-microphone signal model parameters,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 27, no. 7, pp. 1136–1150, 2019.
- A.I. Koutrouvelis, R.C. Hendriks, R. Heusdens and J. Jensen, “Estimation of Sensor Array Signal Model Parameters Using Factor Analysis,” EUSIPCO 2019.
- J. M. Ten Berge and H. A. Kiers, “A numerical approach to the approximate and the exact minimum rank of a covariance matrix,” *Psychometrika*, vol. 56, no. 2, pp. 309–315, 1991.
- G. Della Riccia and A. Shapiro, “Minimum rank and minimum trace of covariance matrices,” *Psychometrika*, vol. 47, no. 4, pp. 443–448, 1982.
- W. Ledermann, “On a problem concerning matrices with variable diagonal elements,” *Proc. Roy. Soc. Endinburgh*, vol. 60, pp. 1–17, 1940.
- R. B. Cattell, “The scree test for the number of factors,” *Multivariate behavioral research*, vol. 1, no. 2, pp. 245–276, 1966.