Statistical Room Acoustics in Acoustic Sensor Networks

Simon Doclo, Toby Christian Lawin-Ore

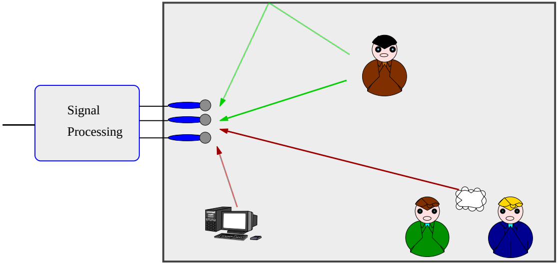
Signal Processing Group
University of Oldenburg, Germany

02.04.2012

Outline

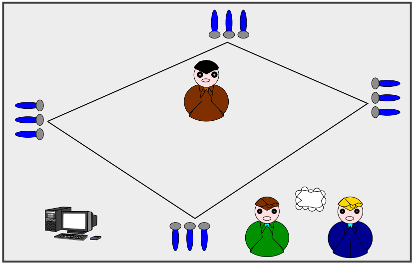
Acoustic Sensor Networks

- Signal acquisition in **adverse acoustic environments**
- "Traditional" microphone arrays:
  - Limited number of microphones (specific configuration)
  - Microphones possibly at large distance from desired source → background noise and reverberation



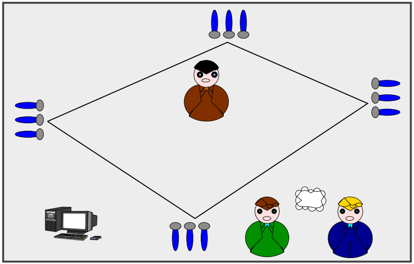
Acoustic Sensor Networks

- Signal acquisition in **adverse acoustic environments**
- Acoustic sensor networks:
  - Network of a large number of **spatially distributed** nodes, typically at unknown positions
  - More information about spatial sound field (microphones with higher SNR, direct-to-reverberant ratio)
  - Wired or wireless data transmission



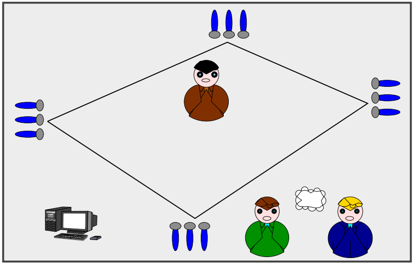
Acoustic Sensor Networks

- Signal acquisition in **adverse acoustic environments**
- Prototype applications:
  - Hearing aids using extra microphones (room, other HA, ...)
  - Video-conferencing using all microphones on laptops / room
  - Surveillance



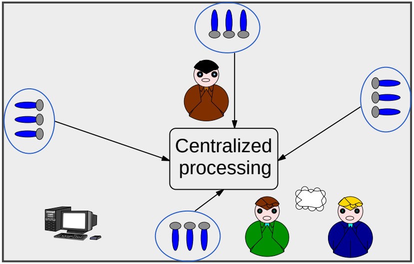
Acoustic Sensor Networks

- **Challenges**:
  - *Dynamic array configuration*: large number of microphones at unknown positions, dynamic subset selection



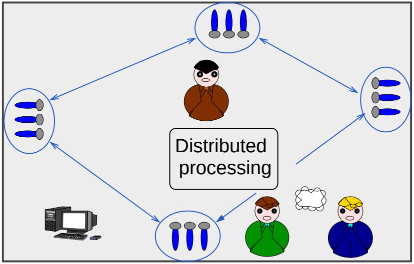
Acoustic Sensor Networks

- **Challenges**:
  - *Dynamic array configuration*: large number of microphones at unknown positions, dynamic subset selection
  - *Distributed and collaborative algorithms*: power and complexity constraints, effect of limited bandwidth and coding



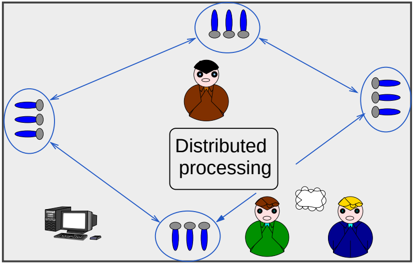
Acoustic Sensor Networks

- **Challenges**:
  - *Dynamic array configuration*: large number of microphones at unknown positions, dynamic subset selection
  - *Distributed and collaborative algorithms*: power and complexity constraints, effect of limited bandwidth and coding
  - *Calibration and synchronisation issues*



Acoustic Sensor Networks

- **Challenges**:
  - *Dynamic array configuration*: large number of microphones at unknown positions, dynamic subset selection
  - *Distributed and collaborative algorithms*: power and complexity constraints, effect of limited bandwidth and coding
  - *Calibration and synchronisation issues*



Objectives

- Performance of signal enhancement algorithms depends on **acoustical scenario**:
  - microphone configuration $\mathbf{P}_{mic}$
  - desired source position $\mathbf{p}_s$
  - noise position(s) $\mathbf{p}_n$ and SNR (*not for diffuse noise*)
  - room properties (e.g. T_{60}, dimensions, reflection coefficients)

Objectives

- Performance of signal enhancement algorithms depends on **acoustical scenario**:
  - microphone configuration $\mathbf{P}_{mic}$
  - desired source position $\mathbf{p}_s$
  - noise position(s) $\mathbf{p}_n$ and SNR (*not for diffuse noise*)
  - room properties (e.g. T_{60}, dimensions, reflection coefficients)

- **General objectives**:
  1. Microphone subset selection
     - For a *given* microphone configuration and room, compare performance of different subsets
     - Either for one specific source position or averaged over different source positions

Objectives

- Performance of signal enhancement algorithms depends on **acoustical scenario**:
  - microphone configuration $\mathbf{P}_{mic}$
  - desired source position $\mathbf{p}_s$
  - noise position(s) $\mathbf{p}_n$ and SNR (*not for diffuse noise*)
  - room properties (e.g. T_{60}, dimensions, reflection coefficients)
- **General objectives**:
  1. Microphone subset selection
     - For a *given* microphone configuration and room, compare performance of different subsets
     - Either for one specific source position or averaged over different source positions
  2. Optimisation of microphone positions
     - For a *given* room, optimise positions of distributed microphones (e.g. using average performance)

Objectives

- **Analyse performance of signal enhancement algorithms for different acoustical scenarios**
- Approaches:
  1. Using **measurements** (RIRs, noise coherence)
     - Most accurate
     - Time-consuming if large number of source positions and microphones configurations need to be compared

Objectives

- **Analyse performance of signal enhancement algorithms for different acoustical scenarios**
- Approaches:
  1. Using **measurements** (RIRs, noise coherence)
     - Most accurate
     - Time-consuming if large number of source positions and microphones configurations need to be compared
  2. Using **simulations** (RIRs, noise coherence)
     - E.g. using image model, room acoustics software
     - Numerical simulation

Objectives

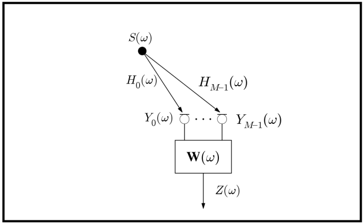
- **Analyse performance of signal enhancement algorithms for different acoustical scenarios**
- Approaches:
  1. Using **measurements** (RIRs, noise coherence)
     - Most accurate
     - Time-consuming if large number of source positions and microphones configurations need to be compared
  2. Using **simulations** (RIRs, noise coherence)
     - E.g. using image model, room acoustics software
     - Numerical simulation
  3. Derive (approximate) **analytical expressions** using **statistical room acoustics**, incorporating statistical properties of RIRs

Objectives

- **Analyse performance of signal enhancement algorithms for different acoustical scenarios**
- Approaches:
  1. Using **measurements** (RIRs, noise coherence)
     - Most accurate
     - Time-consuming if large number of source positions and microphones configurations need to be compared
  2. Using **simulations** (RIRs, noise coherence)
     - E.g. using image model, room acoustics software
     - Numerical simulation
  3. Derive (approximate) **analytical expressions** using **statistical room acoustics**, incorporating statistical properties of RIRs

This presentation: analytical expression for the spatially averaged output SNR of Multi-Channel Wiener Filter (MWF), *given* relative distance between source and microphones

Signal model and configuration



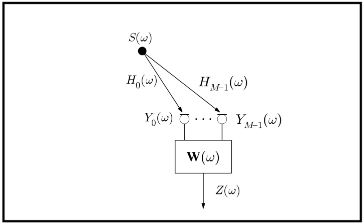
- Microphone signals in frequency-domain

$$\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega) = \mathbf{H}(\omega)S(\omega) + \mathbf{V}(\omega)$$

$$\mathbf{Y}(\omega) = [Y_0(\omega) \cdots Y_{M-1}(\omega)]^T, \quad \mathbf{H}(\omega) = [H_0(\omega) \cdots H_{M-1}(\omega)]^T$$

- Output signal: $Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega) = Z_x(\omega) + Z_v(\omega)$

Signal model and configuration



- Desired source at position $\mathbf{p}_s = [x_s \;\; y_s \;\; z_s]^T$
- Microphones at positions $\mathbf{p}_m = [x_m \;\; y_m \;\; z_m]^T, m = 0 \cdots M-1$
- Relative distance between source and microphones

$$\mathbf{D} = \begin{bmatrix} D_0 \\ \vdots \\ D_{M-1} \end{bmatrix} = \begin{bmatrix} \|\mathbf{p}_s - \mathbf{p}_0\| \\ \vdots \\ \|\mathbf{p}_s - \mathbf{p}_{M-1}\| \end{bmatrix}$$

Multi-channel Wiener Filter

- **Goal**: MMSE estimate of speech component X_{m_0}

$$\xi(\mathbf{W}) = \mathcal{E}\left\{\left|X_{m_0} - \mathbf{W}^H\mathbf{Y}\right|^2\right\}$$

Multi-channel Wiener Filter

- **Goal**: MMSE estimate of speech component X_{m_0}

$$\xi(\mathbf{W}) = \mathcal{E}\left\{\left|X_{m_0} - \mathbf{W}^H\mathbf{Y}\right|^2\right\}$$

- **Solution**:

$$\mathbf{W}_{m_0} = \mathbf{\Phi}_y^{-1}\mathbf{\Phi}_x\mathbf{e}_{m_0} \qquad \mathbf{e}_{m_0} = [0\cdots1\cdots0]^T$$

- Speech and noise correlation matrices:

$$\mathbf{\Phi}_y = \mathbf{\Phi}_x + \mathbf{\Phi}_v$$

$$\mathbf{\Phi}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\}, \quad \mathbf{\Phi}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\}$$

Multi-channel Wiener Filter

- **Goal**: MMSE estimate of speech component X_{m_0}

$$\xi(\mathbf{W}) = \mathcal{E}\left\{\left|X_{m_0} - \mathbf{W}^H\mathbf{Y}\right|^2\right\}$$

- **Solution**:

$$\mathbf{W}_{m_0} = \mathbf{\Phi}_y^{-1}\mathbf{\Phi}_x\mathbf{e}_{m_0} \qquad \mathbf{e}_{m_0} = [0\cdots1\cdots0]^T$$

- Speech and noise correlation matrices:

$$\mathbf{\Phi}_y = \mathbf{\Phi}_x + \mathbf{\Phi}_v$$

$$\mathbf{\Phi}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\}, \quad \mathbf{\Phi}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\}$$

- **Output SNR**:

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\mathcal{E}\{|Z_x|^2\}}{\mathcal{E}\{|Z_v|^2\}} = \frac{\mathbf{W}_{m_0}^H\mathbf{\Phi}_x\mathbf{W}_{m_0}}{\mathbf{W}_{m_0}^H\mathbf{\Phi}_v\mathbf{W}_{m_0}}$$

Multi-channel Wiener Filter

- For a **single speech source**: $\mathbf{\Phi}_x = \phi_s \mathbf{H}\mathbf{H}^H$, $\phi_s = \mathcal{E}\{|S|^2\}$

$$\mathbf{W}_{m_0} = \frac{\mathbf{\Phi}_v^{-1}\mathbf{H}}{\phi_s^{-1} + \Lambda} H_{m_0}^*, \qquad \mathrm{SNR}_{\mathrm{out}} = \phi_s \Lambda$$

$$\boxed{\Lambda = \mathbf{H}^H \mathbf{\Phi}_v^{-1} \mathbf{H}}$$

Multi-channel Wiener Filter

- For a **single speech source**: $\boldsymbol{\Phi}_x = \phi_s \mathbf{H}\mathbf{H}^H$, $\phi_s = \mathcal{E}\{|S|^2\}$

$$\mathbf{W}_{m_0} = \frac{\boldsymbol{\Phi}_v^{-1}\mathbf{H}}{\phi_s^{-1} + \Lambda} H_{m_0}^*, \qquad \mathrm{SNR}_{\mathrm{out}} = \phi_s \Lambda$$

$$\boxed{\Lambda = \mathbf{H}^H \boldsymbol{\Phi}_v^{-1} \mathbf{H}}$$

- **Homogeneous noise field**: $\boldsymbol{\Phi}_v = \phi_v \boldsymbol{\Gamma}_v$

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\phi_s}{\phi_v}\rho \qquad \boxed{\rho = \mathbf{H}^H \boldsymbol{\Gamma}_v^{-1} \mathbf{H}}$$

Multi-channel Wiener Filter

- For a **single speech source**: $\mathbf{\Phi}_x = \phi_s \mathbf{H}\mathbf{H}^H$, $\phi_s = \mathcal{E}\{|S|^2\}$

$$\mathbf{W}_{m_0} = \frac{\mathbf{\Phi}_v^{-1}\mathbf{H}}{\phi_s^{-1} + \Lambda} H_{m_0}^*, \qquad \mathrm{SNR_{out}} = \phi_s \Lambda$$

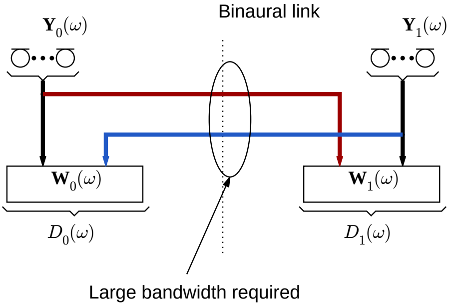
$$\boxed{\Lambda = \mathbf{H}^H \mathbf{\Phi}_v^{-1} \mathbf{H}}$$

- **Homogeneous noise field**: $\mathbf{\Phi}_v = \phi_v \mathbf{\Gamma}_v$

$$\mathrm{SNR_{out}} = \frac{\phi_s}{\phi_v}\rho \qquad \boxed{\rho = \mathbf{H}^H \mathbf{\Gamma}_v^{-1} \mathbf{H}}$$

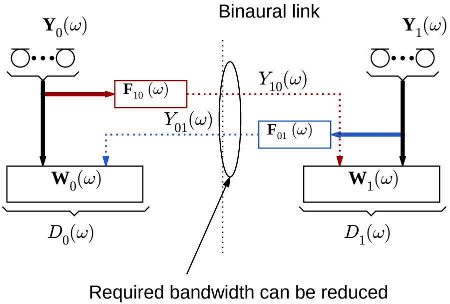
SNR improvement only depends on Acoustical Transfer Function vector $\mathbf{H}$ and noise coherence matrix $\mathbf{\Gamma}_v$

Intermezzo: Distributed MWF



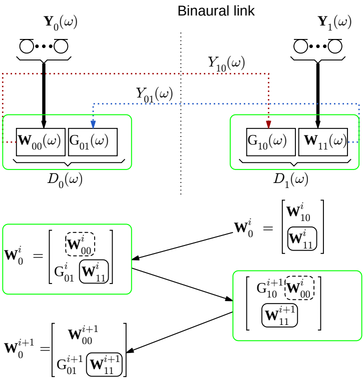
- All microphone signals are transmitted over a wireless link

Intermezzo: Distributed MWF



- All microphone signals are transmitted over a wireless link
- Reduce bandwidth requirement of wireless link by transmitting one signal
    - Iterative distributed binaural MWF scheme (DB-MWF)

Distributed MWF



- In each iteration $\mathbf{F}_{10}$ is equal to $\mathbf{W}_{00}$ from previous iteration, and $\mathbf{F}_{01}$ is equal to $\mathbf{W}_{11}$ from previous iteration
- **Converges to centralized MWF !**

Rate constraints

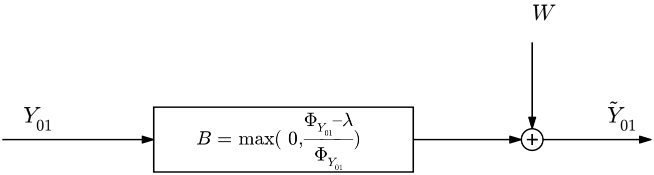
- Objective: investigate influence of link bandwidth on performance of binaural MWF algorithm
- The signal $\mathbf{Y}_{01}$ is encoded at finite bitrate R before transmission

$$
\begin{aligned}
R(\lambda) &= \frac{1}{4\pi} \int_{-\infty}^{\infty} \max(0, \log_2 \frac{\Phi_{\mathbf{Y}_{01}}}{\lambda}) d\omega \\
D(\lambda) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \min(\lambda, \Phi_{\mathbf{Y}_{01}}) d\omega
\end{aligned}
$$

- $\Phi_{\mathbf{Y}_{01}}$ is the PSD of $\mathbf{Y}_{01}$
- Parameter λ links the transmission rate to the distortion

Rate constraints

Upper bound on achievable performance can be calculated using forward channel representation



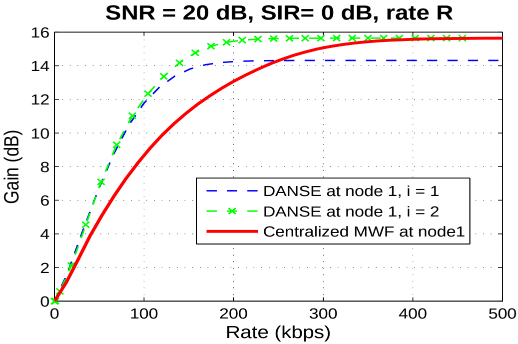
$$\Phi_W = \max(0, \lambda \frac{\Phi_{\mathbf{Y}_{01}} - \lambda}{\Phi_{\mathbf{Y}_{01}}}).$$

- The resulting MSE for SDW-MWF

$$\xi(R) \quad = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\Phi_{X_{0,1}} - \Phi^{H}_{\tilde{\mathbf{Y}}_0 X_{0,1}} \Phi^{-1}_{\tilde{\mathbf{Y}}_0} \Phi_{\tilde{\mathbf{Y}}_0 X_{0,1}} \right) d\omega$$

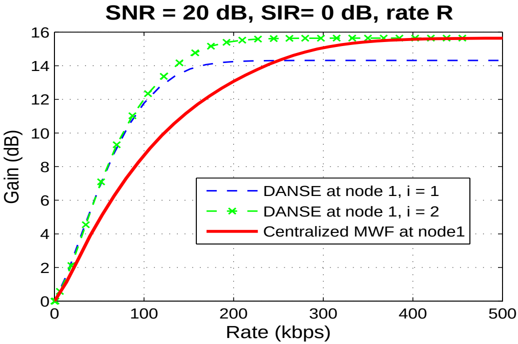
Rate constraints

- Effect on performance of DB-MWF-algorithm
  - Single signal is compressed/transmitted in each iteration
  - Spread iterations over subsequent frames



Rate constraints

- Effect on performance of DB-MWF-algorithm
  - Single signal is compressed/transmitted in each iteration
  - Spread iterations over subsequent frames



➪ DB-MWF-algorithm converges after $K = 2$ iterations, moreover achieving **highest performance gain**

Multi-channel Wiener Filter

- For a **single speech source**: $\mathbf{\Phi}_x = \phi_s \mathbf{H}\mathbf{H}^H$, $\phi_s = \mathcal{E}\{|S|^2\}$

$$\mathbf{W}_{m_0} = \frac{\mathbf{\Phi}_v^{-1}\mathbf{H}}{\phi_s^{-1} + \Lambda} H_{m_0}^*, \qquad \mathrm{SNR}_{\mathrm{out}} = \phi_s \Lambda$$

$$\boxed{\Lambda = \mathbf{H}^H \mathbf{\Phi}_v^{-1} \mathbf{H}}$$

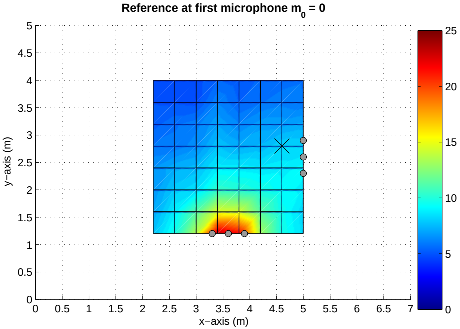
- **Homogeneous noise field**: $\mathbf{\Phi}_v = \phi_v \mathbf{\Gamma}_v$

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\phi_s}{\phi_v}\rho \qquad \boxed{\rho = \mathbf{H}^H \mathbf{\Gamma}_v^{-1} \mathbf{H}}$$

SNR improvement only depends on Acoustical Transfer Function vector $\mathbf{H}$ and noise coherence matrix $\mathbf{\Gamma}_v$

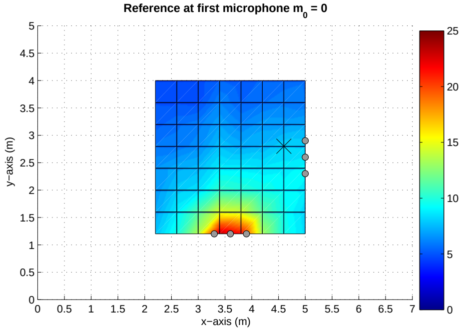
Multi-channel Wiener Filter

Output SNR of the MWF can be computed using measured or simulated ATFs and noise coherence matrices.



Multi-channel Wiener Filter

Output SNR of the MWF can be computed using measured or simulated ATFs and noise coherence matrices.



Objective: (approximate) analytical expression for the output SNR of the MWF using **statistical room acoustics**

Statistical properties of ATFs

- Decomposition of ATFs $\mathbf{H}$ in direct and reverberant component

$$\mathbf{H}(\theta) = \mathbf{H}_d(\theta) + \mathbf{H}_r(\theta)$$

- Stochastic variable $\theta = [\mathbf{p}_s,\ \mathbf{p}_0 \cdots \mathbf{p}_{M-1}]$

[1] M. R. Schroeder, "Frequency correlation function of frequency responses in room", *Journal of the Acoustical Society of America*, vol. 34, no.12, pp. 1819-1823, Dec. 1962

Statistical properties of ATFs

- Decomposition of ATFs $\mathbf{H}$ in direct and reverberant component

$$\mathbf{H}(\theta) = \mathbf{H}_d(\theta) + \mathbf{H}_r(\theta)$$

- Stochastic variable $\theta = [\mathbf{p}_s,\ \mathbf{p}_0 \cdots \mathbf{p}_{M-1}]$
- Statistical properties of ATFs under specific conditions[1]
  - room dimensions large relative to wavelength of signals
  - frequencies above Schröder frequency $f_g = 2000\sqrt{T_{60}/V}$ (V = volume of room, T_{60} = reverberation time)
  - microphones and source at least half a wavelength away from walls
- spatial expectation $\mathcal{E}_\theta\{\cdot\}$ = ensemble average over all realizations of θ

[1] M. R. Schroeder, "Frequency correlation function of frequency responses in room", *Journal of the Acoustical Society of America*, vol. 34, no.12, pp. 1819-1823, Dec. 1962

Statistical properties of ATFs

- Statistical properties of ATFs (fixed relative distance $\mathbf{D}$)
  - A1 Direct path component independent of realization of θ

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})|D_m\} = H_{m,d} = \frac{e^{-j\frac{\omega}{c}D_m}}{4\pi D_m} \; \forall m$$

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})H^*_{n,d}(\boldsymbol{\theta})|D_m, D_n\} = H_{m,d}H^*_{n,d} = \frac{e^{j\frac{\omega}{c}(D_n - D_m)}}{(4\pi)^2 D_m D_n} \quad \forall m, n$$

Statistical properties of ATFs

- Statistical properties of ATFs (fixed relative distance $\mathbf{D}$)
  - A1 Direct path component independent of realization of θ

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})|D_m\} = H_{m,d} = \frac{e^{-j\frac{\omega}{c}D_m}}{4\pi D_m} \; \forall m$$

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})H^*_{n,d}(\boldsymbol{\theta})|D_m, D_n\} = H_{m,d}H^*_{n,d} = \frac{e^{j\frac{\omega}{c}(D_n-D_m)}}{(4\pi)^2 D_m D_n} \quad \forall m, n$$

  - A3 Spatial expectation of spectrum of reverberant component

$$\mathcal{E}_{\boldsymbol{\theta}}\{|H_{m,r}(\boldsymbol{\theta})|^2\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A} \qquad (\bar{\alpha} = \frac{0.161V}{AT_{60}},\ A = \text{total surface area})$$

  - A4 Spatially expected correlation between reverberant components

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,r}(\boldsymbol{\theta})H^*_{n,r}(\boldsymbol{\theta})\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}{\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|} \quad \forall m, n$$

Statistical properties of ATFs

- Statistical properties of ATFs (fixed relative distance $\mathbf{D}$)

A1 Direct path component independent of realization of θ

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})|D_m\} = H_{m,d} = \frac{e^{-j\frac{\omega}{c}D_m}}{4\pi D_m} \ \forall m$$

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})H^*_{n,d}(\boldsymbol{\theta})|D_m, D_n\} = H_{m,d}H^*_{n,d} = \frac{e^{j\frac{\omega}{c}(D_n - D_m)}}{(4\pi)^2 D_m D_n} \quad \forall m, n$$

A3 Spatial expectation of spectrum of reverberant component

$$\mathcal{E}_{\boldsymbol{\theta}}\{|H_{m,r}(\boldsymbol{\theta})|^2\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A} \quad (\bar{\alpha} = \frac{0.161V}{AT_{60}}, \ A = \text{total surface area})$$

A4 Spatially expected correlation between reverberant components

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,r}(\boldsymbol{\theta})H^*_{n,r}(\boldsymbol{\theta})\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\frac{\omega}{c}\|\mathbf{p}_m - \mathbf{p}_n\|}{\frac{\omega}{c}\|\mathbf{p}_m - \mathbf{p}_n\|} \quad \forall m, n$$

A5 Direct and reverberant components are uncorrelated

$$\mathcal{E}_{\theta}\{H_{m,d}(\theta)H^*_{n,r}(\theta)|D_m, D_n\} = 0, \quad \forall m, n$$

Statistical properties of ATFs

- Statistical properties of ATFs (fixed relative distance $\mathbf{D}$)

A1 Direct path component independent of realization of θ

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})|D_m\} = H_{m,d} = \frac{e^{-j\frac{\omega}{c}D_m}}{4\pi D_m} \quad \forall m$$

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})H^*_{n,d}(\boldsymbol{\theta})|D_m, D_n\} = H_{m,d}H^*_{n,d} = \frac{e^{j\frac{\omega}{c}(D_n - D_m)}}{(4\pi)^2 D_m D_n} \quad \forall m, n$$

A3 Spatial expectation of spectrum of reverberant component

$$\mathcal{E}_{\boldsymbol{\theta}}\{|H_{m,r}(\boldsymbol{\theta})|^2\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A} \qquad \left(\bar{\alpha} = \frac{0.161V}{AT_{60}},\ A = \text{total surface area}\right)$$

A4 Spatially expected correlation between reverberant components

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,r}(\boldsymbol{\theta})H^*_{n,r}(\boldsymbol{\theta})\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\frac{\omega}{c}\|\mathbf{p}_m - \mathbf{p}_n\|}{\frac{\omega}{c}\|\mathbf{p}_m - \mathbf{p}_n\|} \quad \forall m, n$$

A5 Direct and reverberant components are uncorrelated

$$\mathcal{E}_{\theta}\{H_{m,d}(\theta)H^*_{n,r}(\theta)|D_m, D_n\} = 0, \quad \forall m, n$$

A6 $\mathcal{E}_{\theta}\{|H_m(\theta)|^2\,|D_m\} = \frac{1}{(4\pi D_m)^2} + \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}$

Output SNR using statistical properties ATFs

- For each realization of θ

$$\rho(\theta) = \mathbf{H}_d^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\theta)+\mathbf{H}_d^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\theta)+\mathbf{H}_r^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\theta)+\mathbf{H}_r^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\theta)$$

Output SNR using statistical properties ATFs

- For each realization of θ

$$\rho(\theta) = \mathbf{H}_d^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\theta) + \mathbf{H}_d^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\theta) + \mathbf{H}_r^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\theta) + \mathbf{H}_r^H(\theta)\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\theta)$$

$$\rho(\theta) = \sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn}\Big(H_{d,m}^*(\theta)H_{d,n}(\theta) + H_{d,m}^*(\theta)H_{r,n}(\theta) + H_{r,m}^*(\theta)H_{d,n}(\theta) + H_{r,m}^*(\theta)H_{r,n}(\theta)\Big)$$

with

$$\mathbf{H}_i^H\mathbf{\Gamma}_v^{-1}\mathbf{H}_j = \sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn} H_{i,m}^* H_{j,n}$$

- $\breve{\gamma}_{mn}$ coefficients of the matrix $\mathbf{\Gamma}_v^{-1}$

Output SNR using statistical properties ATFs

- Spatial expectation of $\mathrm{SNR}_{\mathrm{out}}$ given **D**

$$\mathcal{E}_\theta\{\rho(\theta)|\mathbf{D}\} = \sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn}\Big(\mathcal{E}_\theta\{H^*_{d,m}(\theta)H_{d,n}(\theta)|\mathbf{D}\} + \mathcal{E}_\theta\{H^*_{d,m}(\theta)H_{r,n}(\theta)|\mathbf{D}\}$$
$$+\mathcal{E}_\theta\{H^*_{r,m}(\theta)H_{d,n}(\theta)|\mathbf{D}\} + \mathcal{E}_\theta\{H^*_{r,m}(\theta)H_{r,n}(\theta)\}\Big)$$

Output SNR using statistical properties ATFs

- Spatial expectation of $\mathrm{SNR}_{\mathrm{out}}$ given **D**

$$\mathcal{E}_\theta\{\rho(\theta)|\mathbf{D}\} = \sum_{m=1}^{M}\sum_{n=1}^{M}\breve{\gamma}_{mn}\Big(\mathcal{E}_\theta\{H^*_{d,m}(\theta)H_{d,n}(\theta)|\mathbf{D}\} + \cancel{\mathcal{E}_\theta\{H^*_{d,m}(\theta)H_{r,n}(\theta)|\mathbf{D}\}}$$

$$+\cancel{\mathcal{E}_\theta\{H^*_{r,m}(\theta)H_{d,n}(\theta)|\mathbf{D}\}} + \mathcal{E}_\theta\{H^*_{r,m}(\theta)H_{r,n}(\theta)\}\Big)$$

$$\mathcal{E}_\theta\{\mathrm{SNR}_{\mathrm{out}}(\theta)|\mathbf{D}\} = \frac{\phi_s}{\phi_v}\sum_{m=1}^{M}\sum_{n=1}^{M}\breve{\gamma}_{mn}\left(\frac{e^{j\frac{\omega}{c}(D_n-D_m)}}{(4\pi)^2 D_m D_n} + \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}{\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}\right)$$

- Depends on room properties A and $\bar{\alpha}$, relative distance **D** and noise coherence $\mathbf{\Gamma}_v$

Experimental results

Objective: Compare the *analytically* computed spatial expectation $\mathcal{E}_\theta\{\mathrm{SNR}_{\mathrm{out}}(\theta)|\mathbf{D}\}$ with the *numerically* computed spatially averaged output SNR

$$\overline{\mathrm{SNR}}_{\mathrm{out}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{SNR}_{out}(\tilde{\boldsymbol{\theta}}_i)$$

where N is the total number of realizations (translations, rotations) and $\tilde{\boldsymbol{\theta}}_i, i = 1 \cdots N$ corresponds to a single realization of $\boldsymbol{\theta}$

Experimental results

Objective: Compare the *analytically* computed spatial expectation $\mathcal{E}_\theta\{\mathrm{SNR}_{\mathrm{out}}(\theta)|\mathbf{D}\}$ with the *numerically* computed spatially averaged output SNR

$$\overline{\mathrm{SNR}}_{\mathrm{out}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{SNR}_{out}(\tilde{\boldsymbol{\theta}}_i)$$

where N is the total number of realizations (translations, rotations) and $\tilde{\boldsymbol{\theta}}_i, i = 1 \cdots N$ corresponds to a single realization of $\boldsymbol{\theta}$

Simulation and experimental setup

- Room dimensions 7 m $\times$ 5 m $\times$ 3.5 m and $T_{60} = 250$ ms
- $M = 3$ equally spaced microphones ($d = 4$ cm)
- Source located at endfire of the microphone array such that $\mathbf{D} = [1.36\ \ 1.40\ \ 1.44]^T$

Experimental results

Simulation and experimental setup

- Sampling frequency $f_s = 16\text{kHz}$
- Room impulse responses simulated using image model ($L = 4096$) [2]

[2] E. A. P. Habets, "Room impulse response (RIR) generator", *Available: http://home.tiscali.nl/ehabets/rir generator.html*, Oct. 2008

Experimental results

Simulation and experimental setup

- Sampling frequency $f_s = 16\text{kHz}$
- Room impulse responses simulated using image model ($L = 4096$) [2]
- Diffuse noise coherence matrix $\mathbf{\Gamma}_v$ theoretically computed

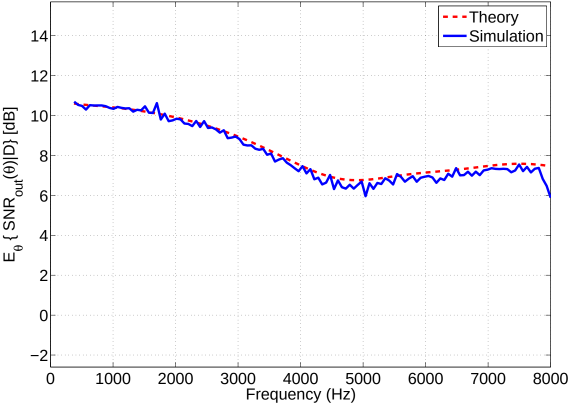
$$\gamma_{mn}(\omega) = \frac{\sin \frac{\omega}{c} \left\|\mathbf{p}_m - \mathbf{p}_n\right\|}{\frac{\omega}{c} \left\|\mathbf{p}_m - \mathbf{p}_n\right\|}$$

- Frequency-flat a-priori input SNR $\frac{\phi_s}{\phi_v}$

[2] E. A. P. Habets, "Room impulse response (RIR) generator", *Available: http://home.tiscali.nl/ehabets/rir generator.html*, Oct. 2008

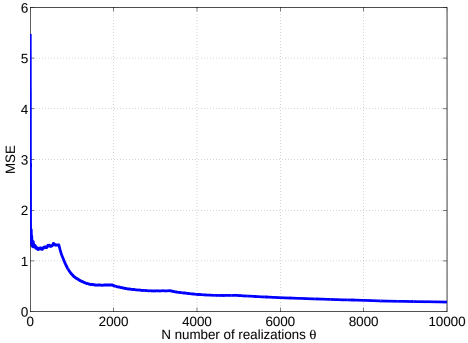
Experimental results

- Monte Carlo simulation using 10000 realizations



Experimental results

- $\text{MSE} = \sum_{\omega} \left| \mathcal{E}_{\boldsymbol{\theta}}\{\text{SNR}_{\text{out}}(\boldsymbol{\theta})|\mathbf{D}\} - \overline{\text{SNR}}_{\text{out}} \right|^2$ as a function of number of realizations θ



The larger the number of realizations, the smaller the MSE between the analytical expression and simulations

Conclusion and future work

- **Analytical expression** for spatial expectation of output SNR of MWF, depending on room properties, relative distance and noise coherence
- Simulation results show that analytical expression is **close to numerical simulations** (for large number of Monte Carlo realizations)

Conclusion and future work

- **Analytical expression** for spatial expectation of output SNR of MWF, depending on room properties, relative distance and noise coherence
- Simulation results show that analytical expression is **close to numerical simulations** (for large number of Monte Carlo realizations)
- Future work
  - For fixed microphone configuration, compute output SNR averaged over different source positions ⇒ optimal subset selection
  - Using average performance ⇒ optimize positions of distributed microphones