# Compressive Sampling of Sound Fields Using Moving Microphones

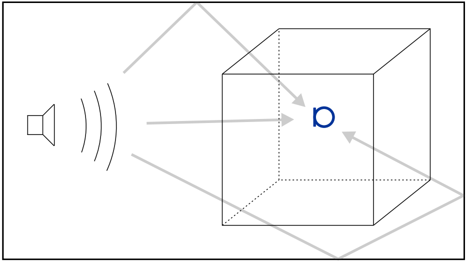
*Fabrice Katzberg, Radoslaw Mazur, Marco Maass, Philipp Koch, and Alfred Mertins*

Institute for Signal Processing
University of Lübeck

April 19, 2018

ICASSP 2018, Calgary

# Sound Transmission in Echoic Environments



Spatio-temporal room impulse response (RIR):

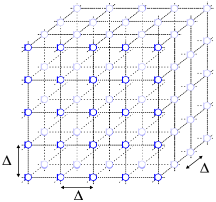
$$h(\boldsymbol{r}, t)$$

# Methods for Sound-Field Measurements

- Conventional approach

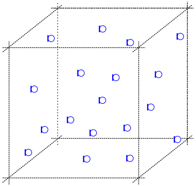
  Obey the spatial sampling theorem

$$\Delta \leq \Delta_{\max} \propto \frac{1}{f_{\max}}.$$

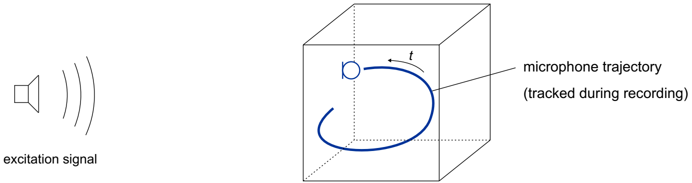



- Compressed sensing for static setups

  Use random microphone positions (Mignot et al. 2013, 2014).



# Proposed Dynamic Approach

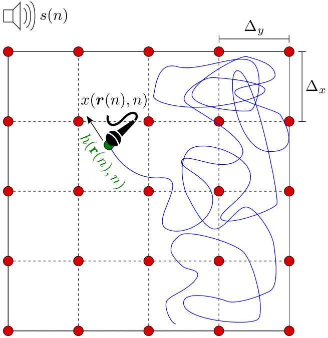



## Compressed-sensing (CS) formulation:

$$\min_{\boldsymbol{h}} \ \|\boldsymbol{x} - \boldsymbol{A}\boldsymbol{h}\|_2 \quad \text{subject to} \quad \|\boldsymbol{c}(\boldsymbol{h})\|_0 \leq K.$$

- $\boldsymbol{x}$: Measured signal
- $\boldsymbol{h}$: Sought impulse responses on a Cartesian grid
- $\boldsymbol{A}$: Matrix containing excitation signal and interpolation coefficients
- $\boldsymbol{c}(\boldsymbol{h})$: Sparse representation of $\boldsymbol{h}$

# Inverse Problem with Dynamic Measurements




$$x(\boldsymbol{r}(n), n) = s(n) * h(\boldsymbol{r}(n), n)$$
$$= \sum_{m=0}^{L-1} h(\boldsymbol{r}(n), m)\, s(n-m)$$
$$\approx \sum_{m=0}^{L-1} \sum_{u=1}^{N} h(\boldsymbol{g}_u, m)\varphi_n(\boldsymbol{g}_u)\, s(n-m)$$

Interpolation between virtual grid points $\boldsymbol{g}_u$

$M$ : Number of samples $x(\boldsymbol{r}(n), n)$
$L$ : Length of RIRs
$N$ : Number of grid RIRs
$U$ : Number of unknowns $(NL)$

# Structure of Measurement Model

$$x(\boldsymbol{r}(n), n) = \sum_{m=0}^{L-1} \sum_{u=1}^{N} \varphi_n(\boldsymbol{g}_u) s(n-m) \; h(\boldsymbol{g}_u, n)$$

$$\boldsymbol{x} = \boldsymbol{A}\,\boldsymbol{h} \qquad \boldsymbol{x} \in \mathbb{R}^M, \boldsymbol{A} \in \mathbb{R}^{M \times U}, \boldsymbol{h} \in \mathbb{R}^U$$

## Structure of sampling matrix:

$$\boldsymbol{A} = \left[\boldsymbol{\Phi}_1 \boldsymbol{S},\ \boldsymbol{\Phi}_2 \boldsymbol{S},\ \ldots,\ \boldsymbol{\Phi}_N \boldsymbol{S}\right]$$

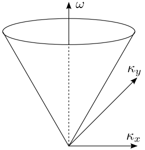
- Convolution matrix $\boldsymbol{S} \in \mathbb{R}^{M \times L}$
- Diagonal matrix $\boldsymbol{\Phi}_u \in \mathbb{R}^{M \times M}$ with weightings for $u$-th grid position
- $m$-th row of $\boldsymbol{A}$ is composed of the spatially weighted source signal

$$s_m(\boldsymbol{g}, n) = \varphi_{m-1}(\boldsymbol{g})\, s(m-1-n)$$

# Sparse Sound-Field Representation

- Under far-field assumptions, the sound-field spectrum ideally lives on the hypercone (Ajdler et al. 2006)

$$\kappa_x^2 + \kappa_y^2 + \kappa_z^2 = \frac{\omega^2}{c_0^2}.$$




$\omega$ : Angular frequency in time
$\kappa_{x,y,z}$ : Angular frequencies in space

$\Rightarrow$ Describe grid RIRs by 4D frequency representation $\boldsymbol{c} = \boldsymbol{\Psi}\boldsymbol{h}$, where

$$\boldsymbol{\Psi} = \boldsymbol{T}_Z \otimes \boldsymbol{T}_Y \otimes \boldsymbol{T}_X \otimes \boldsymbol{T}_L$$

is a unitary $U \times U$ matrix and $\boldsymbol{c} \in \mathbb{C}^U$ is a $K$-sparse vector.

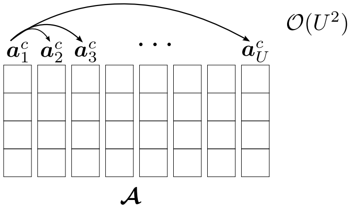
$\Rightarrow$ CS matrix: $\boldsymbol{\mathcal{A}} = \boldsymbol{A}\boldsymbol{\Psi}^{\mathrm{H}}$

# Coherence of Measurements

- For practical applications, the coherence of $\boldsymbol{\mathcal{A}}$ may be used to evaluate the CS problem:

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{1 \leq u \neq v \leq U} \frac{|\langle \boldsymbol{a}_u^c, \boldsymbol{a}_v^c \rangle|}{\|\boldsymbol{a}_u^c\|_2 \|\boldsymbol{a}_v^c\|_2},$$

where $\boldsymbol{a}_u^c$ denotes the $u$-th column of $\boldsymbol{\mathcal{A}}$.




- Theoretical error bounds for CS recovery improve with smaller coherence (Donoho et al. 2001, 2003).

# Structure of the CS Matrix: Rows

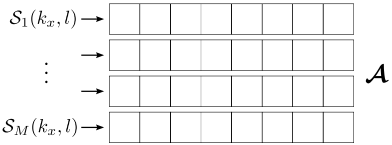
- For simplicity, let us consider the 2D case with $\boldsymbol{\Psi} = \boldsymbol{T}_X \otimes \boldsymbol{T}_L$.

$\Rightarrow$ $m$-th row of $\boldsymbol{A}$: $s_m(g_x, n) = \varphi_{m-1}(g_x)\, s(m-1-n)$.

$\Rightarrow$ For $\boldsymbol{\Psi}$ performing the 2D DFT on $h(g_x, n)$, the $m$-th row of $\boldsymbol{\mathcal{A}}$ is

$$\mathcal{S}_m(k_x, l) = \frac{1}{\sqrt{XL}} \sum_{g_x=0}^{X-1} \sum_{n=0}^{L-1} s_m(g_x, n)\, e^{-2\pi j \frac{l}{L} n} e^{-2\pi j \frac{k_x}{X} g_x},$$

where $k_x \in \{-\frac{X-1}{2}, \ldots, \frac{X-1}{2}\}$ and $l \in \{-\frac{L-1}{2}, \ldots, \frac{L-1}{2}\}$ are the sampled frequency variables for the space and time dimension.



# Structure of the CS Matrix: Columns

- Each column of $\boldsymbol{\mathcal{A}}$ comprises a specific frequency pair $(k'_x, l')$ of the sampled spectra:

$$\boldsymbol{a}^c_{(k'_x,l')} = \left[\mathcal{S}_1(k'_x, l'), \mathcal{S}_2(k'_x, l'), \ldots, \mathcal{S}_M(k'_x, l')\right]^T .$$

- Let us define the trajectory relative to the modeled grid in space:

$$D_x(n) = \frac{r_x(n) - r_0}{\Delta_x}.$$

- For spectrally flat excitation and interpolation, the movement of the microphone from point $r_x(n)$ to $r_x(n+m)$ ideally corresponds to recursive phase shifts in the discrete Fourier spectrum,

$$\mathcal{S}_{n+m}(k'_x, l') = e^{-2\pi j (D_x(m) - D_x(n)) \frac{k'_x}{X}}\, e^{-2\pi j m \frac{l'}{L}}\, \mathcal{S}_n(k'_x, l').$$

# Fast Coherence Analysis 2D

The coherence of $\boldsymbol{\mathcal{A}}$ is

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{(k_x',l')\neq(k_x'',l'')} \frac{|\langle \boldsymbol{a}^c_{(k_x',l')}, \boldsymbol{a}^c_{(k_x'',l'')}\rangle|}{\|\boldsymbol{a}^c_{(k_x',l')}\|_2 \|\boldsymbol{a}^c_{(k_x'',l'')}\|_2}$$
$$= \max_{(\Delta k_x,\Delta l)\neq(0,0)} \frac{1}{M} \left| \sum_{n=0}^{M-1} e^{-2\pi j \frac{D_x(n)}{X}\Delta k_x} \, e^{-2\pi j \frac{n}{L}\Delta l} \right|,$$

where

$$\Delta k_x = k_x' - k_x'',\ \Delta k_x \in \{-(X-1),\dots,X-1\},$$
$$\Delta l = l' - l'',\ \Delta l \in \{-(L-1),\dots,L-1\},$$

are the differences of the discrete frequency variables $k_x', k_x'' \in \{-\frac{X-1}{2},\dots,\frac{X-1}{2}\}$ and $l', l'' \in \{-\frac{L-1}{2},\dots,\frac{L-1}{2}\}$.

# Fast Coherence Analysis 4D

Defining $\boldsymbol{r}_D(n) = [D_x(n), D_y(n), D_z(n)]^T$, $\boldsymbol{d} = [\Delta k_x, \Delta k_y, \Delta k_z]^T$, and

$$\mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d}) = e^{-2\pi j\left(\frac{D_x(n)}{X}\Delta k_x + \frac{D_y(n)}{Y}\Delta k_y + \frac{D_z(n)}{Z}\Delta k_z\right)},$$

the coherence of the 4D sampling problem is

$$\mu(\boldsymbol{\mathcal{A}}) = \max_{(\boldsymbol{d},\Delta l)} \frac{1}{M}\left|\sum_{n=0}^{M-1} \mathcal{X}(\boldsymbol{r}_D(n), \boldsymbol{d})\, e^{-2\pi j \frac{n}{L}\Delta l}\right|$$ with $(\boldsymbol{d}, \Delta l) \neq (\boldsymbol{0}, 0)$.

- → Calculating coherence is reduced from a problem in $\mathcal{O}(U^2)$ to $\mathcal{O}(U)$.
- → Coherence only depends on the grid related trajectory $\boldsymbol{r}_D(n)$.

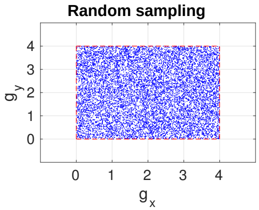
⇒ Efficient tool for finding optimal trajectories for sought grids, alternatively, for modeling optimal grids for given measurements.

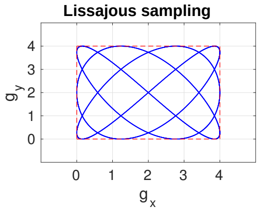
# Experiments

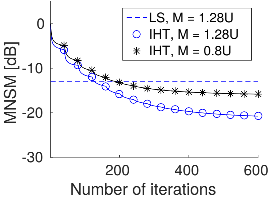
- We simulated the sound field inside an office sized room by using the image source method with $f_s = 8\,\mathrm{kHz}$.
- Length of RIRs is $L = 511$, ROI is a $5 \times 5$ grid with $\Delta = 0.02\,\mathrm{m}$, design of extended $7 \times 7$ grid
- SNR $= 40\,\mathrm{dB}$
- Lagrange interpolator of order three and Fourier representations
- Quality measure for sound-field recovery:

$$\text{MNSM} = \frac{1}{N} \sum_{u=1}^{N} \frac{\|\boldsymbol{h}_u^{\text{true}} - \hat{\boldsymbol{h}}_u\|_2^2}{\|\boldsymbol{h}_u^{\text{true}}\|_2^2}$$

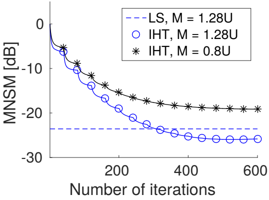
# Results




$$\mu(\boldsymbol{\mathcal{A}}) = 0.55$$




$$\mu(\boldsymbol{\mathcal{A}}) = 0.39$$

# Conclusions

- CS framework for sound-field recovery using moving microphones.
- Linear system by using source signal and microphone positions.
- CS solution allows for robust recovery in the underdetermined case.
- Straightforward analysis of CS matrix for Fourier representations.
- Fast coherence analysis for spectrally flat excitation/interpolation.

Thank you for your attention.