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):

$$h(r, t)$$
Methods for Sound-Field Measurements

- Conventional approach

  Obey the spatial sampling theorem

  \[ \Delta \leq \Delta_{\text{max}} \propto \frac{1}{f_{\text{max}}}. \]

- Compressed sensing for static setups

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

Compressed-sensing (CS) formulation:

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

- $x$: Measured signal
- $h$: Sought impulse responses on a Cartesian grid
- $A$: Matrix containing excitation signal and interpolation coefficients
- $c(h)$: Sparse representation of $h$
Inverse Problem with Dynamic Measurements

\[ x(r(n), n) = s(n) \ast h(r(n), n) \]

\[ = \sum_{m=0}^{L-1} h(r(n), m) s(n - m) \]

\[ \approx \sum_{m=0}^{L-1} \sum_{u=1}^{N} h(g_u, m) \phi_n(g_u) s(n - m) \]

Interpolation between virtual grid points \( g_u \)

\( M \): Number of samples \( x(r(n), n) \)
\( L \): Length of RIRs
\( N \): Number of grid RIRs
\( U \): Number of unknowns \( (NL) \)
Structure of Measurement Model

\[ x(r(n), n) = \sum_{m=0}^{L-1} \sum_{u=1}^{N} \varphi_n(g_u) s(n - m) h(g_u, n) \]

\[ x = \begin{bmatrix} A \end{bmatrix} h \quad x \in \mathbb{R}^M, A \in \mathbb{R}^{M \times U}, h \in \mathbb{R}^U \]

Structure of sampling matrix:

\[ A = [\Phi_1 S, \Phi_2 S, \ldots, \Phi_N S] \]

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

\[ s_m(g, n) = \varphi_{m-1}(g) s(m - 1 - n) \]
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

⇒ Describe grid RIRs by 4D frequency representation \(c = \Psi h\), where

\[
\Psi = T_Z \otimes T_Y \otimes T_X \otimes T_L
\]

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

⇒ CS matrix: \(\mathbf{A} = \mathbf{A} \Psi^H\)
Coherence of Measurements

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

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

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

Structure of the CS Matrix: Rows

- For simplicity, let us consider the 2D case with $\Psi = T_X \otimes T_L$.

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

⇒ For $\Psi$ performing the 2D DFT on $h(g_x, n)$, the $m$-th row of $A$ is

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

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 $\mathbf{A}$ comprises a specific frequency pair $(k'_{x}, l')$ of the sampled spectra:

$$\mathbf{a}^c_{(k'_{x}, l')} = [S_1(k'_{x}, l'), S_2(k'_{x}, l'), \ldots, S_M(k'_{x}, l')]^T.$$ 

- 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,

$$S_{n+m}(k'_{x}, l') = e^{-2\pi j(D_x(m) - D_x(n)) \frac{k'_{x}}{X}} e^{-2\pi jm \frac{l'}{L}} S_n(k'_{x}, l').$$
Fast Coherence Analysis 2D

The coherence of $\mathcal{A}$ is

$$
\mu(\mathcal{A}) = \max_{(k_x', l') \neq (k_x'', l'')} \frac{\left| \langle a_{(k_x', l')}, a_{(k_x'', l'')} \rangle \right|}{\| a_{(k_x', l')} \|_2 \| a_{(k_x'', l'')} \|_2}
$$

$$
= \max_{(\Delta k_x, \Delta l) \neq (0,0)} \frac{1}{M} \left| \sum_{n=0}^{M-1} e^{-2\pi j \frac{D_x(n)}{X} \Delta k_x} e^{-2\pi j \frac{n}{L} \Delta l} \right|
$$

where

$$
\Delta k_x = k_x' - k_x'', \; \Delta k_x \in \{-(X-1), \ldots, X-1\},
$$

$$
\Delta l = l' - l'', \; \Delta l \in \{-(L-1), \ldots, L-1\},
$$

are the differences of the discrete frequency variables

$k_x', k_x'' \in \{-\frac{X-1}{2}, \ldots, \frac{X-1}{2}\}$ and $l', l'' \in \{-\frac{L-1}{2}, \ldots, \frac{L-1}{2}\}$. 
Fast Coherence Analysis 4D

Defining $r_D(n) = [D_x(n), D_y(n), D_z(n)]^T$, $d = [\Delta k_x, \Delta k_y, \Delta k_z]^T$, and

$$\mathcal{X}(r_D(n), d) = e^{-2\pi j \left( \frac{D_x(n)}{x} \Delta k_x + \frac{D_y(n)}{y} \Delta k_y + \frac{D_z(n)}{z} \Delta k_z \right)}$$

the coherence of the 4D sampling problem is

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

with $(d, \Delta l) \neq (0, 0)$.

$\rightarrow$ Calculating coherence is reduced from a problem in $O(U^2)$ to $O(U)$.

$\rightarrow$ Coherence only depends on the grid related trajectory $r_D(n)$.

$\Rightarrow$ 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 $f_s = 8$ kHz.
- Length of RIRs is $L = 511$, ROI is a $5 \times 5$ grid with $\Delta = 0.02$ m, design of extended $7 \times 7$ grid
- SNR = 40 dB
- Lagrange interpolator of order three and Fourier representations
- Quality measure for sound-field recovery:

$$MNSM = \frac{1}{N} \sum_{u=1}^{N} \frac{\| \hat{h}_u - h_u \|^2}{\| h_u \|^2}$$
Results

Random sampling

Lissajous sampling

\[ \mu(A) = 0.55 \]

\[ \mu(A) = 0.39 \]
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.