Gridless 3D Recovery of Image Sources from Room Impulse Responses

T. Sprunck [1,2], Y. Privat [2], C. Foy [3], A. Deleforge [1]

[1]Inria, Nancy, France [2]IRMA, Strasbourg, France

[3]UMRAE, Strasbourg, France

Outline

Problem formulation

Can one hear the shape of a room ?

Problem formulation

Can one hear the shape of a room ?
More precisely, given:

→ an initial **sound impulse** (Dirac in time and 3D space)

→ discrete-time, multichannel, low-pass measurements of the **room impulse response** (RIR)

can we reconstruct the positions of the walls, floor and ceiling ?

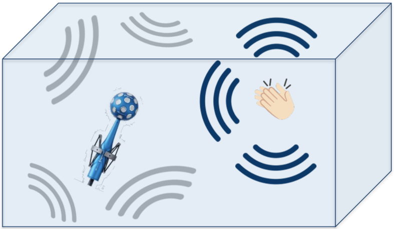
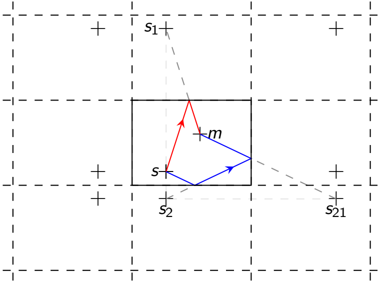
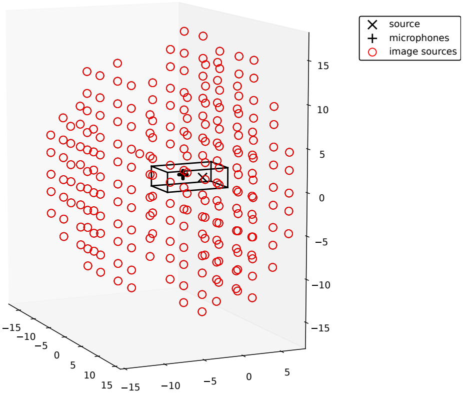


Image Sources



(a) First and second order IS



(b) 3D view of an IS point cloud

- the image sources contain the information about the **acoustic properties** of the room
- in particular, room geometry is given by the locations of the source and the **first order** image sources

Model

Using the **image source model** [2], the approximated pressure field $p(\mathbf{r}, t)$ is solution to an inhomogeneous free-field wave equation given by :

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = \sum_{k=0}^{\infty} a_k \delta\left(\mathbf{r} - \mathbf{r}_k^{\text{src}}\right)\delta(t) \qquad (1)$$

Model

Using the **image source model** [2], the approximated pressure field $p(\mathbf{r}, t)$ is solution to an inhomogeneous free-field wave equation given by :

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = \sum_{k=0}^{\infty} a_k \delta\left(\mathbf{r} - \mathbf{r}_k^{\text{src}}\right)\delta(t) \tag{1}$$

The general solution of 1 is given by:

$$p(\mathbf{r}, t) = \sum_{k=0}^{\infty} a_k \frac{\delta\left(t - \left\|\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right\|_2 / c\right)}{4\pi \left\|\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right\|_2}. \tag{2}$$

Model

Using the **image source model** [2], the approximated pressure field $p(\mathbf{r}, t)$ is solution to an inhomogeneous free-field wave equation given by :

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = \sum_{k=0}^{\infty} a_k \delta\left(\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right)\delta(t) \tag{1}$$

The general solution of 1 is given by:

$$p(\mathbf{r}, t) = \sum_{k=0}^{\infty} a_k \frac{\delta\left(t - \left\|\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right\|_2 / c\right)}{4\pi\left\|\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right\|_2}. \tag{2}$$

Hypotheses:

1. rectangular cuboid rooms
2. frequency-independent walls, floor and ceiling
3. omnidirectional sources and receivers
4. specular reflections
5. one point source emitting a perfect impulse at $t = 0$
6. fixed source and receiver responses: ideal low-pass filters

Outline

Super-resolution : general formulation

Super-resolution : general formulation

- we want to reconstruct a d-dimensional measure $\psi = \sum_i a_i \delta_{\mathbf{r}_i}$
- we only have access to a vector of observations via a linear operator Γ (with kernel γ) : $\mathbf{x} = \Gamma(\psi) = \int_r \gamma(r) d\psi(r) \in \mathbb{R}^{N_{obs}}$

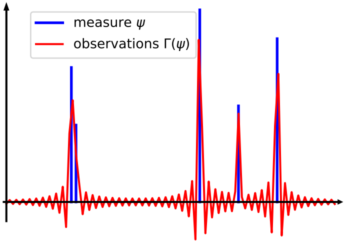


Figure: Example of 1D measure and its observation

Super-resolution : general formulation

- we want to reconstruct a d-dimensional measure $\psi = \sum_i a_i \delta_{\mathbf{r}_i}$
- we only have access to a vector of observations via a linear operator Γ (with kernel γ) : $\mathbf{x} = \Gamma(\psi) = \int_r \gamma(r) d\psi(r) \in \mathbb{R}^{N_{obs}}$

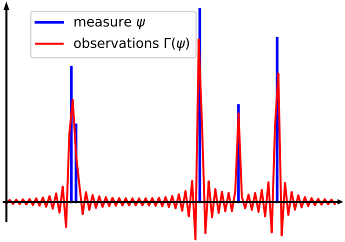


Figure: Example of 1D measure and its observation

Idea: consider a **relaxed** optimization problem over the entire space of **Radon measures** [3] of $\mathbb{R}^d$:

$$\min_{\psi \in \mathcal{M}(\mathbb{R}^d)} \underbrace{\frac{1}{2}\|\mathbf{x} - \Gamma(\psi)\|_2^2}_{\text{data compliance}} + \underbrace{\lambda \|\psi\|_{TV}}_{\text{regularization}} \qquad \text{(BLASSO)}$$

Room Response

We start by relaxing the source distribution in space:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \sum_{k=0}^{\infty} a_k \delta\left(\mathbf{r} - \mathbf{r}_k^{\text{src}}\right)\delta(t) \qquad (1)$$

Room Response

We start by relaxing the source distribution in space:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta(t) \tag{3}$$

Room Response

We start by relaxing the source distribution in space:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta(t) \tag{3}$$

For a given source distribution ψ, the solution to the wave equation (3) is given by

$$p(\mathbf{r},t) = \int_{r'} \frac{\delta\left(t - \|\mathbf{r}-\mathbf{r}'\|_2 / c\right)}{4\pi \|\mathbf{r}-\mathbf{r}'\|_2}\psi(\mathbf{r}')d\mathbf{r}'. \tag{4}$$

Room Response

We start by relaxing the source distribution in space:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta(t) \tag{3}$$

For a given source distribution ψ, the solution to the wave equation (3) is given by

$$p(\mathbf{r},t) = \int_{r'} \frac{\delta\left(t - \|\mathbf{r}-\mathbf{r}'\|_2 / c\right)}{4\pi\|\mathbf{r}-\mathbf{r}'\|_2}\psi(\mathbf{r}')d\mathbf{r}'. \tag{4}$$

The **multi-channel response x** measured by the microphones is:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2 / c\right)}{4\pi\left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2}\psi(\mathbf{r})\, d\mathbf{r} \tag{5}$$

where $\kappa : t \mapsto \mathrm{sinc}(\pi f_s t)$ is the ideal low-pass filter at the microphones frequency of sampling.

The **multi-channel response** measured by the microphone **x** is:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2 / c\right)}{4\pi \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2} \psi(\mathbf{r})\, d\mathbf{r} \quad (5)$$

We define the linear operator

$$\Gamma : \begin{array}{ccc} \mathcal{M}(\mathbb{R}^3) & \longrightarrow & \mathbb{R}^{N\times M} \\ \psi & \mapsto & \mathbf{x} = \left(\int_{r\in\mathbb{R}^3} \frac{\kappa(n_j/f_s - \|r - r_{m_j}^{\text{mic}}\|_2/c)}{4\pi\|r - r_{m_j}^{\text{mic}}\|_2} d\psi(r)\right)_{1\leq j\leq N\times M} \end{array} \quad (6)$$

The **multi-channel response** measured by the microphone **x** is:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2/c\right)}{4\pi\left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2}\psi(\mathbf{r})\,d\mathbf{r} \tag{5}$$

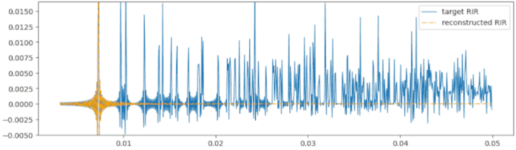
We define the linear operator

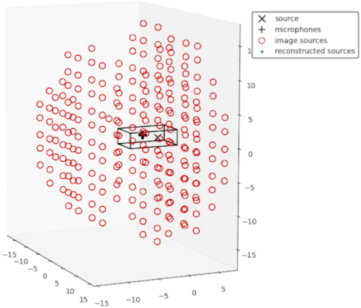
$$\Gamma : \begin{array}{ccc} \mathcal{M}(\mathbb{R}^3) & \longrightarrow & \mathbb{R}^{N\times M} \\ \psi & \mapsto & \mathbf{x} = \left(\int_{r\in\mathbb{R}^3} \frac{\kappa(n_j/f_s - \|r - r_{m_j}^{\mathrm{mic}}\|_2/c)}{4\pi\|r - r_{m_j}^{\mathrm{mic}}\|_2} d\psi(r)\right)_{1\le j\le N\times M} \end{array} \tag{6}$$

In particular, if $\psi = \sum_{k=0}^{K} a_k \delta_{\mathbf{r}_k^{\mathrm{src}}}$ (the measure defined by the image sources), $\mathbf{x} = \Gamma(\psi)$ is the multichannel RIR.

Outline

Adapted Sliding-Frank-Wolfe algorithm [1]





Iteration 1

Algorithm:

- find a new source by maximizing $\mathbf{r} \mapsto \frac{1}{\lambda}\Gamma^*(\mathbf{x} - \Gamma\psi^k)(\mathbf{r})$
- optimize over the amplitudes a_k

Last step:

local non-convex optimization of the cost function with regards to the amplitudes and positions $(a_k, \mathbf{r}_k)_k$

Experiments

Simulated experimental setup:

- compact spherical array of 32 microphones (scaled eigenmike with radius 4.2*cm*, 8.4*cm*, etc.)
- random room sizes ($2 \times 2 \times 2$m$\rightarrow 10 \times 10 \times 5$m), random sources and microphone locations
- synthetic noisy RIR cut at 50ms for the observations
- study the impact of the sampling frequency, the noise, the array radius

Numerical results

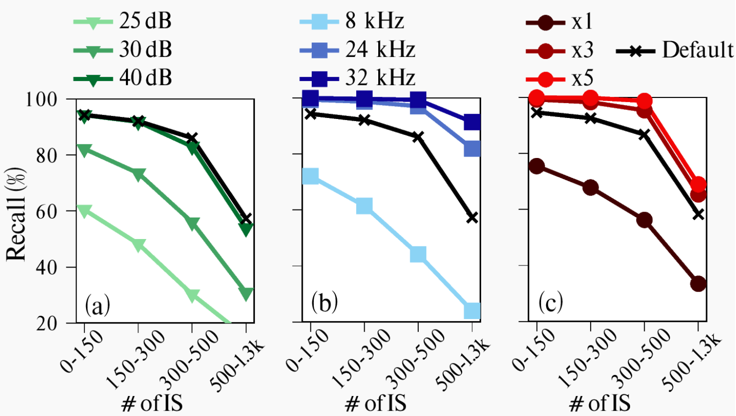


Figure: Recall over a room dataset for varying noise ratios (PSNR), sampling frequencies and spherical microphone array diameter
Default parameters: noiseless, $f_s = 16\text{kHz}$, $d = 16.8\text{cm}$

Conclusion

The proposed method offers significant advantages for room geometry reconstruction :

- **gridless**, direct estimation of continuous 3D source positions from discrete RIRs
- high precision recovery of low order image sources
- robustness to noise
- requires no prior information on the room properties

Conclusion

The proposed method offers significant advantages for room geometry reconstruction :

- **gridless**, direct estimation of continuous 3D source positions from discrete RIRs
- high precision recovery of low order image sources
- robustness to noise
- requires no prior information on the room properties

Some of the areas that remain to explore :

- estimating room parameters
- generalization to non-rectangular room shapes
- application to real data
- joint estimation of the source-microphone response $\kappa \ldots$

[1] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1976.

[2] V. Duval and G. Peyré, "Exact Support Recovery for Sparse Spikes Deconvolution," *Foundations of Computational Mathematics*, vol. 15, no. 5, pp. 1315–1355, 2015.

[3] Q. Denoyelle, V. Duval, G. Peyré, and E. Soubies, "The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy," *Inverse Problems*, 2019.