Can one hear the shape of a room ?

A. Deleforge (Inria), C. Foy (UMRAE), Y. Privat (IRMA),
T. Sprunck (IRMA, Inria)

April 7, 2023

Outline

Problem formulation

Can one hear the shape of a room ?

Problem formulation

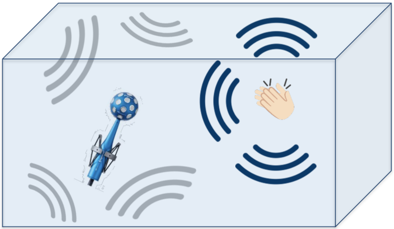
Can one hear the shape of a room ?

In other words, given:

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

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

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



Model

The pressure field $p(\mathbf{r}, t)$ in the room is solution to the following wave equation with boundary conditions:

$$\begin{cases} \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = a_0\delta\left(\mathbf{r} - \mathbf{r}^{\mathrm{src}}\right)\delta(t) & \mathbf{r} \in \Omega \\ \mathbf{n}(\mathbf{r}) \cdot \nabla p(\mathbf{r}, t) + \frac{1}{c}\frac{\partial}{\partial t}\beta(\mathbf{r}, t) * p(\mathbf{r}, t) = 0 & \mathbf{r} \in \partial\Omega \end{cases} \quad (1)$$

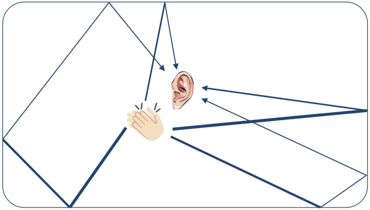
Model

The pressure field $p(\mathbf{r}, t)$ in the room is solution to the following wave equation with boundary conditions:

$$\begin{cases} \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = a_0\delta\left(\mathbf{r} - \mathbf{r}^{\text{src}}\right)\delta(t) & \mathbf{r} \in \Omega \\ \mathbf{n}(\mathbf{r}) \cdot \nabla p(\mathbf{r}, t) + \frac{1}{c}\frac{\partial}{\partial t}\beta(\mathbf{r}, t) * p(\mathbf{r}, t) = 0 & \mathbf{r} \in \partial\Omega \end{cases} \quad (1)$$

Simplification:

- polyhedric room geometry
- perfectly reflective (rigid) walls
- frequency independant walls



$\rightarrow$ the Image Source Model (ISM)

Image sources

Each reflection path is associated to an image source

Image sources

Each reflection path is associated to an image source

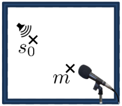


Image sources

Each reflection path is associated to an image source

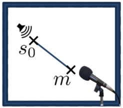


Image sources

Each reflection path is associated to an image source

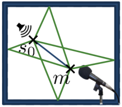


Image sources

Each reflection path is associated to an image source

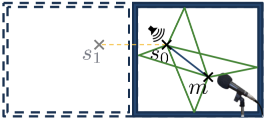


Image sources

Each reflection path is associated to an image source

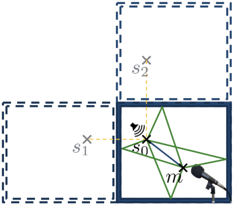


Image sources

Each reflection path is associated to an image source

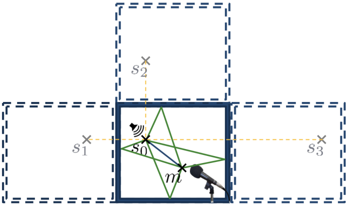


Image sources

Each reflection path is associated to an image source

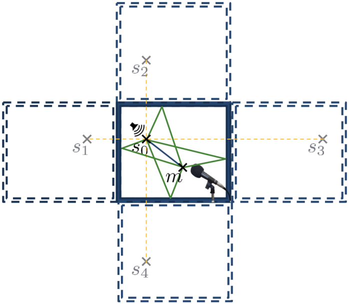


Image sources

Each reflection path is associated to an image source

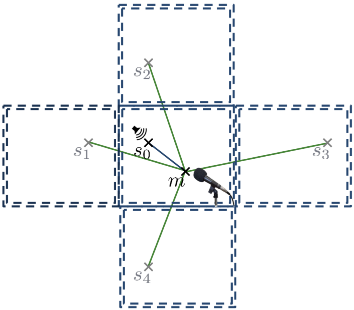


Image sources

Each reflection path is associated to an image source

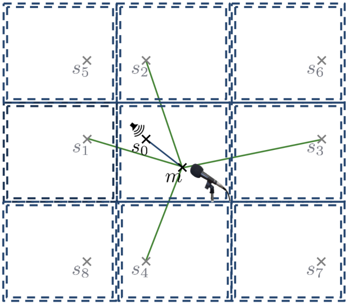


Image sources

Each reflection path is associated to an image source

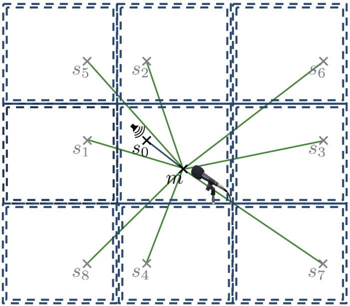
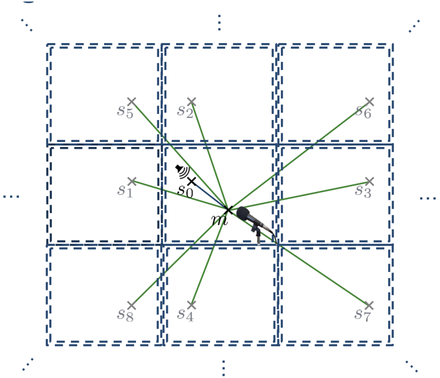


Image sources

Each reflection path is associated to an image source



The first order sources are reflections of the source with respect to the walls

$\longrightarrow$ knowing the position of the source and the first order image sources yields the room geometry

Image sources model (ISM)

Perfectly specular reflections translate in Neumann boundary conditions:

$$\begin{cases} \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = a_0\delta\left(\mathbf{r}-\mathbf{r}^{\mathrm{src}}\right)\delta(t) & \mathbf{r}\in\Omega \\ \mathbf{n}(\mathbf{r})\cdot\nabla p(\mathbf{r},t) = 0 & \mathbf{r}\in\partial\Omega \end{cases} \tag{2}$$

Image sources model (ISM)

Perfectly specular reflections translate in Neumann boundary conditions:

$$\begin{cases} \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = a_0 \delta\left(\mathbf{r} - \mathbf{r}^{\text{src}}\right)\delta(t) & \mathbf{r} \in \Omega \\ \mathbf{n}(\mathbf{r}) \cdot \nabla p(\mathbf{r},t) = 0 & \mathbf{r} \in \partial\Omega \end{cases} \tag{2}$$

When considering a cuboid room, the solution writes:

$$p(\mathbf{r},t) = A \sum_{k=0}^{+\infty} \frac{\delta(\|\mathbf{r} - \mathbf{r}_k^{\text{src}}\| - ct)}{\|\mathbf{r} - \mathbf{r}_k^{\text{src}}\|} \tag{3}$$

where the $\mathbf{r}_k^{\text{src}}$ are the locations of the **image sources** and A is a constant.

Image sources model (ISM)

Perfectly specular reflections translate in Neumann boundary conditions:

$$\begin{cases} \frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = a_0 \delta\left(\mathbf{r} - \mathbf{r}^{\text{src}}\right)\delta(t) & \mathbf{r} \in \Omega \\ \mathbf{n}(\mathbf{r}) \cdot \nabla p(\mathbf{r},t) = 0 & \mathbf{r} \in \partial\Omega \end{cases} \tag{2}$$

When considering a cuboid room, the solution writes:

$$p(\mathbf{r},t) = A \sum_{k=0}^{+\infty} \frac{\delta(\|\mathbf{r} - \mathbf{r}_k^{\text{src}}\| - ct)}{\|\mathbf{r} - \mathbf{r}_k^{\text{src}}\|} \tag{3}$$

where the $\mathbf{r}_k^{\text{src}}$ are the locations of the **image sources** and A is a constant.

p is then solution to the free field wave equation:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2} p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \sum_{k=0}^{+\infty} \delta\left(\mathbf{r} - \mathbf{r}_k^{\text{src}}\right)\delta(t) \tag{4}$$

Image sources model (ISM)

Extension: consider the free-field wave equation:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \sum_{k=0}^{K} a_k \delta\left(\mathbf{r} - \mathbf{r}_k^{\mathsf{src}}\right)\delta(t) \tag{4}$$

The general solution of (4) is given by:

$$p(\mathbf{r},t) = \sum_{k=0}^{K} a_k \frac{\delta\left(t - \left\|\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right\|_2 / c\right)}{\left\|\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right\|}. \tag{5}$$

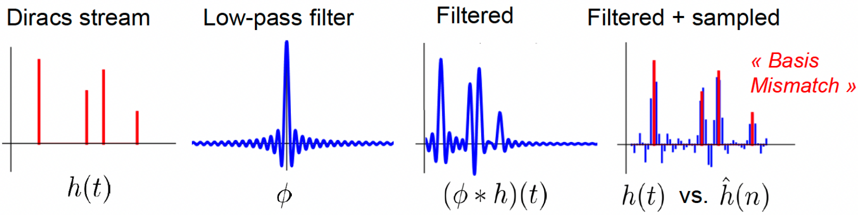
- an absorption factor $\alpha_i, 1 \leq i \leq 7$ is associated to each wall
- the amplitude a for an order n image source is given by :

$$a = \prod_{l=0}^{n} \sqrt{1 - \alpha_{n_l}} \tag{6}$$

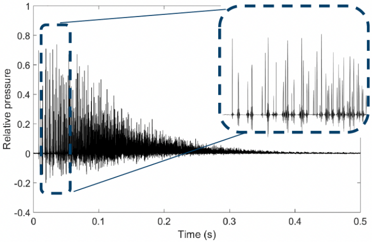
where n_l is the index of the wall corresponding to the l-th reflection

Jont B Allen and David A Berkley, *Image method for efficiently simulating small-room acoustics*, The Journal of the Acoustical Society of America, vol. 65, no. 4, pp. 943–950, 1979

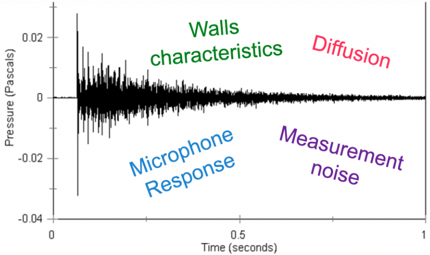
Measured Room Impulse Response signal



Full synthetic RIR (specular only)



Full measured RIR



Audio reconstruction

Note: The RIR embeds the acoustical properties of the room. Given a solution G (Green function) of the wave equation with impulse source term:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \delta\left(\mathbf{r} - \mathbf{r}^{\mathrm{src}}\right)\delta(t) \qquad (7)$$

$t \mapsto G(\mathbf{r}, t)$ corresponds to the RIR observed at position $\mathbf{r}$

Audio reconstruction

Note: The RIR embeds the acoustical properties of the room. Given a solution G (Green function) of the wave equation with impulse source term:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \delta\left(\mathbf{r} - \mathbf{r}^{\mathrm{src}}\right)\delta(t) \tag{7}$$

$t \mapsto G(\mathbf{r},t)$ corresponds to the RIR observed at position $\mathbf{r}$
A person speaking at position $\mathbf{r}^{\mathrm{src}}$ can be modelled by the equation:

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \delta\left(\mathbf{r} - \mathbf{r}_k^{\mathrm{src}}\right)h(t) \tag{8}$$

The solution of this last equation is given by a time convolution with G :

$$p(r,t) = \left(G(r,.) * h(.)\right)(t) \tag{9}$$

$\rightarrow$ we can simulate how a sound would play in a room by convolving with a RIR No Reverb abs0.1 abs0.01

Audio measurements

In practice a RIR is measured by real microphones:

- time convolution with a filter κ
- time discretization by a sampling frequency f_s
- recorded at different locations by an array of microphones (multichannel RIR)

Audio measurements

In practice a RIR is measured by real microphones:

- time convolution with a filter κ
- time discretization by a sampling frequency f_s
- recorded at different locations by an array of microphones (multichannel RIR)

The idealized multichannel RIR is given by:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \sum_{k=0}^{K} a_k \frac{\kappa\left(n/f_s - \|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}_k^{\mathsf{src}}\|/c\right)}{\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}_k^{\mathsf{src}}\|} \tag{10}$$

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

Outline

General formulation

- we want to reconstruct a **3D** 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) : $\mathbf{x} = \Gamma\psi = \int_r \gamma(r) d\psi(r) \in R^{N \times M}$

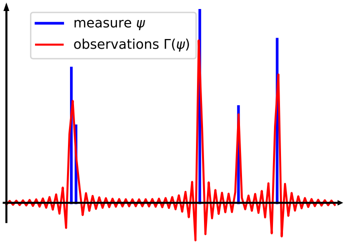


Figure: Example of 1D measure and its observation

General formulation

- we want to reconstruct a **3D** 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) : $\mathbf{x} = \Gamma\psi = \int_r \gamma(r) d\psi(r) \in R^{N\times M}$

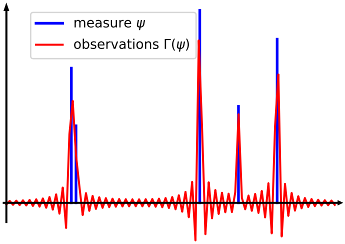


Figure: Example of 1D measure and its observation

Idea: consider a relaxed optimization problem over the entire space of Radon measures of $(\mathbb{R}^3)$:

$$\min_{\psi\in\mathcal{M}(\mathbb{R}^3)} \underbrace{\frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2}_{\text{data compliance}} + \underbrace{\lambda\|\psi\|_{\text{TV}}}_{\text{regularization}} \qquad \text{(BLASSO)}$$

Vincent Duval and Gabriel Peyré, *Exact Support Recovery for Sparse Spikes Deconvolution*, Foundations of Computational Mathematics, vol. 15, no. 5, pp. 1315-1355, 2015

Room Response

How can we express the RIR as the image of an operator Γ ?

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta(t) \tag{11}$$

Room Response

How can we express the RIR as the image of an operator Γ ?

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta(t) \tag{11}$$

For a given source distribution ψ, the solution to the wave equation (11) is given by a convolution with a Green function:

$$p(\mathbf{r},t) = \int_{r'} \frac{\delta\left(t - \left\|\mathbf{r}-\mathbf{r}'\right\|_2 / c\right)}{4\pi \left\|\mathbf{r}-\mathbf{r}'\right\|_2}\psi(\mathbf{r}')d\mathbf{r}'. \tag{12}$$

Room Response

How can we express the RIR as the image of an operator Γ ?

$$\frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\mathbf{r},t) - \Delta p(\mathbf{r},t) = \psi(\mathbf{r})\delta(t) \tag{11}$$

For a given source distribution ψ, the solution to the wave equation (11) is given by a convolution with a Green function:

$$p(\mathbf{r},t) = \int_{r'} \frac{\delta\left(t - \left\|\mathbf{r}-\mathbf{r}'\right\|_2 / c\right)}{4\pi\left\|\mathbf{r}-\mathbf{r}'\right\|_2}\psi(\mathbf{r}')d\mathbf{r}'. \tag{12}$$

The multi-channel response $\mathbf{x}$ measured by the microphones is:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2 / c\right)}{4\pi\left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2}\psi(\mathbf{r})\, d\mathbf{r} \tag{13}$$

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

Room operator

The multi-channel response measured by the microphone $\mathbf{x}$ is:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2 / c\right)}{4\pi \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2} \psi(\mathbf{r})\, d\mathbf{r} \tag{13}$$

We define the linear operator

$$\begin{array}{rccl} \Gamma: & \mathcal{M}(\mathbb{R}^3) & \longrightarrow & \mathbb{R}^{N\times M} \\ & \psi & \mapsto & \mathbf{x} = \left(\int_{r\in\mathbb{R}^3} \frac{\kappa(n_j/f_s - \|r - r_{m_j}^{\mathrm{mic}}\|_2/c)}{4\pi\|r - r_{m_j}^{\mathrm{mic}}\|_2} d\psi(r) \right)_{1\le j\le N\times M} \end{array} \tag{14}$$

Room operator

The multi-channel response measured by the microphone $\mathbf{x}$ is:

$$x_{m,n} := (\kappa * p(\mathbf{r}_m^{mic}, \cdot))(n/f_s) = \int_{\mathbf{r}\in\mathbb{R}^3} \frac{\kappa\left(n/f_s - \left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2/c\right)}{4\pi\left\|\mathbf{r}_m^{\mathrm{mic}} - \mathbf{r}\right\|_2}\psi(\mathbf{r})\,d\mathbf{r} \tag{13}$$

We define the linear operator

$$\begin{array}{rccl} \Gamma: & \mathcal{M}(\mathbb{R}^3) & \longrightarrow & \mathbb{R}^{N\times M} \\ & \psi & \mapsto & \mathbf{x} = \left(\int_{r\in\mathbb{R}^3} \frac{\kappa(n_j/f_s - \|r - r_{m_j}^{\mathrm{mic}}\|_2/c)}{4\pi\|r - r_{m_j}^{\mathrm{mic}}\|_2} d\psi(r)\right)_{1\le j\le N\times M} \end{array} \tag{14}$$

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.

Room operator

Issue: the kernel $\gamma_{m,n} : r \mapsto \frac{\kappa(n_j/f_s - \|r - r_{m_j}^{\text{mic}}\|_2/c)}{4\pi\|r - r_{m_j}^{\text{mic}}\|_2}$ presents singularities at each microphone location

Room operator

Issue: the kernel $\gamma_{m,n} : r \mapsto \frac{\kappa(n_j/f_s - \|r - r^{\text{mic}}_{m_j}\|_2/c)}{4\pi\|r - r^{\text{mic}}_{m_j}\|_2}$ presents singularities at each microphone location

$\rightarrow$ we perform a surgery around each microphone

$\rightarrow$ Γ is restricted to $\mathcal{M}(\mathbb{R}^3_\varepsilon)$ where $\mathbb{R}^3_\varepsilon = \mathbb{R}^3 \setminus \cup_m B(\mathbf{r}^{\text{mic}}_m, \varepsilon)$

Room operator

Issue: the kernel $\gamma_{m,n} : r \mapsto \frac{\kappa(n_j/f_s - \|r - r^{\text{mic}}_{m_j}\|_2/c)}{4\pi\|r - r^{\text{mic}}_{m_j}\|_2}$ presents singularities at each microphone location

$\rightarrow$ we perform a surgery around each microphone
$\rightarrow$ Γ is restricted to $\mathcal{M}(\mathbb{R}^3_\varepsilon)$ where $\mathbb{R}^3_\varepsilon = \mathbb{R}^3 \setminus \cup_m B(\mathbf{r}^{\text{mic}}_m, \varepsilon)$

Note : the BLASSO optimization problem remains convex

$$\min_{\psi \in \mathcal{M}(\mathbb{R}^3_\varepsilon)} \underbrace{\frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2}_{\text{data compliance}} + \underbrace{\lambda\|\psi\|_{\text{TV}}}_{\text{regularization}} \tag{15}$$

Outline

The Frank-Wolfe algorithm

Let $\mathcal{D}$ denote a weakly compact convex set of a Banach space and $f : \mathcal{D} \to \mathbb{R}$, a convex, differentiable real-valued function. The Frank–Wolfe algorithm solves numerically the optimization problem

$$\inf_{x \in \mathcal{D}} f(x)$$

Algorithm

- Initialization: Let x_0 denote any point in $\mathcal{D}$.
- Step 1: direction-finding. Find s_k solving

$$\inf_{s \in \mathcal{D}} Df(x_k) \cdot s$$

  Interpretation: minimize the linear approximation of the problem given by the first-order Taylor approximation of f around x_k.
- Step 2: step size determination. find α minimizing $\alpha \mapsto f(x_k + \alpha(s_k - x_k))$ subject to $0 \leq \alpha \leq 1$.
- Step 3: update. Set $x_{k+1} \leftarrow x_k + \alpha(s_k - x_k)$ and $k \leftarrow k + 1$ and go to Step 1.

The Sliding-Frank-Wolfe algorithm

Application of Frank-Wolfe to

$$\min_{\psi \in \mathcal{M}(\mathbb{R}^3_\varepsilon)} T_\lambda(\psi) \coloneqq \frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2 + \lambda\|\psi\|_{\mathsf{TV}}. \tag{16}$$

The Sliding-Frank-Wolfe algorithm

Application of Frank-Wolfe to

$$\min_{\psi \in \mathcal{M}(\mathbb{R}^3_\varepsilon)} T_\lambda(\psi) \coloneqq \frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2 + \lambda\|\psi\|_{\mathsf{TV}}. \tag{16}$$

Consider instead

$$\min_{(t,\psi) \in C} T_\lambda(t,\psi) \coloneqq \frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2 + t\lambda \tag{17}$$

where $C = \{(t,\psi) \in \mathbb{R}_+ \times \mathcal{M}(\mathbb{R}^3_\varepsilon),\ \|\psi\|_{\mathsf{TV}} \leq t \leq \frac{\|\mathbf{x}\|_2}{2\lambda}\}$

The Sliding-Frank-Wolfe algorithm

Application of Frank-Wolfe to

$$\min_{\psi \in \mathcal{M}(\mathbb{R}^3_\varepsilon)} T_\lambda(\psi) := \frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2 + \lambda\|\psi\|_{\mathrm{TV}}. \tag{16}$$

Consider instead

$$\min_{(t,\psi) \in C} T_\lambda(t,\psi) := \frac{1}{2}\|\mathbf{x} - \Gamma\psi\|_2^2 + t\lambda \tag{17}$$

where $C = \{(t,\psi) \in \mathbb{R}_+ \times \mathcal{M}(\mathbb{R}^3_\varepsilon),\ \|\psi\|_{\mathrm{TV}} \leq t \leq \frac{\|\mathbf{x}\|_2}{2\lambda}\}$
Main idea : construct a new Dirac at each iteration based on the residual, by maximizing :

$$\eta_\lambda^k : \mathbf{r} \mapsto \frac{1}{\lambda}\Gamma^*(\mathbf{x} - \Gamma\psi^k)(\mathbf{r}) \tag{18}$$

Q. Denoyelle, V. Duval, G. Peyré and E. Soubies, *The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy*, Inverse Problems, vol. 36, no 1, p. 014001, 2019

The Sliding-Frank-Wolfe algorithm

- Initialization: let $\psi^0 = 0$ the null measure.
- k-th iteration: let $\psi^k(\mathbf{r}^k) = \sum_i^{N_k} a_i^k \delta_{\mathbf{r}_i^k}$ ($\mathbf{r}_i^k$ pairwise distinct);
  - Step 1 (spike-finding): Find $\mathbf{r}_*^k$ solving the optimization problem

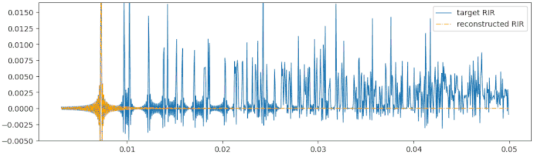
$$\sup_{\mathbf{r}\in\mathbb{R}^3_\varepsilon} |\eta_\lambda^k(\mathbf{r})|, \quad \eta_\lambda^k(\mathbf{r}) = \frac{1}{\lambda}\Gamma^*(\mathbf{x} - \Gamma\psi^k)(\mathbf{r}).$$

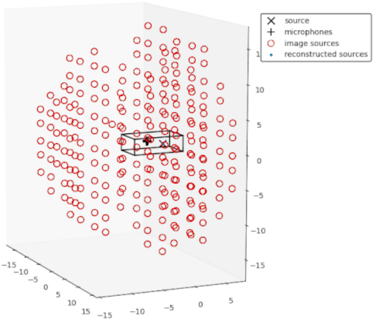
    If $\|\eta_\lambda^k(\mathbf{r})\|_2 \leq 1$, we are done. Else:
  - Step 2 (amplitude optimization): find $\psi^{k+1/2} = \sum_i^{N_k} a_i^{k+1/2}\delta(\mathbf{r} - \mathbf{r}_i^k) + a_{N_k+1}^{k+1/2}\delta(\mathbf{r} - \mathbf{r}_*^k)$ solving

$$\inf_{a_i^{k+1/2}\geq 0} \frac{1}{2}\|\mathbf{x} - \Gamma\psi^{k+1/2}\|_2^2 + \lambda\|\psi^{k+1/2}\|_{TV}.$$

  - Step 3 (sliding): find ψ^{k+1} minimizing locally the criterion wrt $(a, \mathbf{r})$ using as initial point $(a^{k+1/2}, \mathbf{r}^{k+1/2})$.
  - Eventually remove zero amplitudes Dirac masses from ψ^{k+1}.

Numerical results





Iteration 1

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

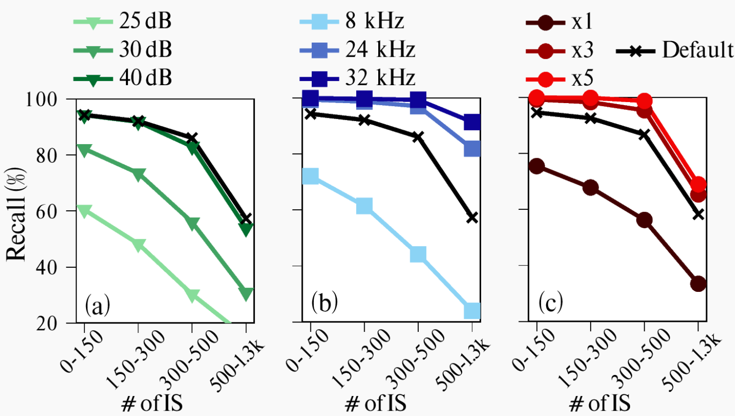


Figure: Recall over a room dataset for varying noise ratios (PSNR), sampling frequencies and spherical microphone array diameter

Default parameters: noiseless, $f_s = 16$kHz, $d = 16.8$cm

T. Sprunck, A. Deleforge, Y. Privat, and C. Foy. *Gridless 3D Recovery of Image Sources from Room Impulse Responses*, IEEE Signal Processing Letters, vol. 29, pp. 2427-2431 2022

Outline

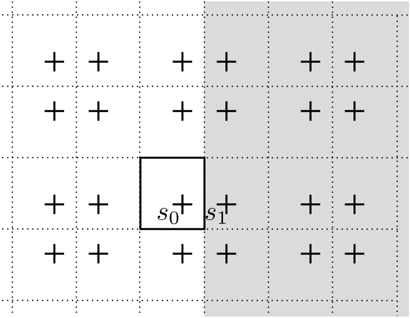
Order 1 method and downsides

Known data: coordinates of the microphones $\mathbf{r}_m^{\mathrm{mic}}$ and the image sources $\mathbf{r}_m^{\mathrm{src}}$ **in the referential of the antenna**.

Order 1 method and downsides

Known data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_m^{\text{src}}$ **in the referential of the antenna**.

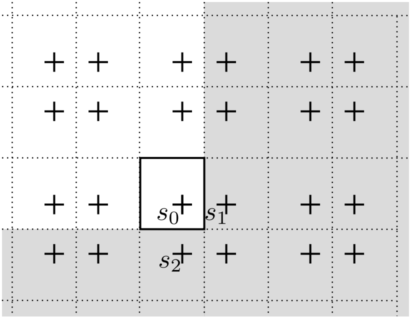
Naive algorithm: identify the source and order 1 sources



Order 1 method and downsides

Known data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_m^{\text{src}}$ **in the referential of the antenna**.

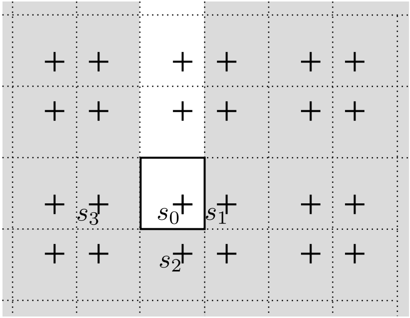
Naive algorithm: identify the source and order 1 sources



Order 1 method and downsides

Known data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_m^{\text{src}}$ **in the referential of the antenna**.

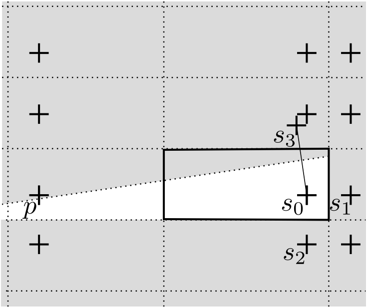
Naive algorithm: identify the source and order 1 sources



Order 1 method and downsides

Known data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_m^{\text{src}}$ **in the referential of the antenna**.

Issues: False positives and imprecise reconstruction



Three steps method

Input data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_k^{\text{src}}$ **in the referential of the antenna**.

Three steps method

Input data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_k^{\text{src}}$ **in the referential of the antenna**.

Idea: use the full reconstructed image sources cloud
$\rightarrow$ exploit the grid structure of the image sources
$\rightarrow$ increased robustness to inaccuracies/false positives

Three steps method

Input data: coordinates of the microphones $\mathbf{r}_m^{\text{mic}}$ and the image sources $\mathbf{r}_k^{\text{src}}$ **in the referential of the antenna**.

Idea: use the full reconstructed image sources cloud
→ exploit the grid structure of the image sources
→ increased robustness to inaccuracies/false positives

Step 1: find the orientation of the room in the antenna referential

Step 2: project the image sources on the three room axes (normals to the walls)

Step 3: apply a 1D clustering algorithm on each axis to find the source/walls distances on each axis

Step 1: finding the orientation

We want to find the three axes of the room using the IS

Step 1: finding the orientation

We want to find the three axes of the room using the IS
$\longrightarrow$ we search for the vectors that are orthogonal to a maximum of vectors built from reconstructed IS

Step 1: finding the orientation

We want to find the three axes of the room using the IS
$\longrightarrow$ we search for the vectors that are orthogonal to a maximum of vectors built from reconstructed IS

$$\max_{\|u\|=1} L(u) := \sum_{i,j} f(u, \mathbf{r}_i^{\mathrm{src}} - \mathbf{r}_j^{\mathrm{src}}) \tag{19}$$

$$\forall u, v \in \mathbb{R}^3, \quad f(u, v) = \begin{cases} 1 & \text{if } u \perp v \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Step 1: finding the orientation

We want to find the three axes of the room using the IS
$\longrightarrow$ we search for the vectors that are orthogonal to a maximum of vectors built from reconstructed IS

$$\max_{\|u\|=1} L(u) := \sum_{i,j} f(u, \mathbf{r}_i^{\text{src}} - \mathbf{r}_j^{\text{src}}) \tag{19}$$

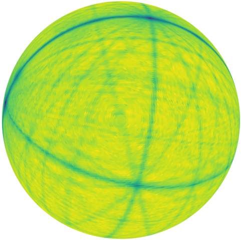
$$\forall u, v \in \mathbb{R}^3, \quad f(u, v) = \begin{cases} 1 & \text{if } u \perp v \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

In practice, we use the following approximation:

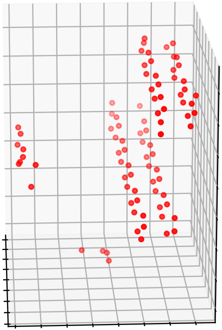
$$\forall u, v \in \mathbb{R}^3, \quad f_\sigma(u, v) = e^{-\frac{1}{2\sigma^2}\left(\frac{u.v}{\|u\|\|v\|}\right)^2} \tag{21}$$

Step 1: finding the orientation

- maximize f_σ over the unit sphere to find a first basis vector e_1
- consider the normal plane to e_1 and maximize the same function over the unit circle to find a second basis vector e_2
- get the last vector $e_3 = e_1 \times e_2$



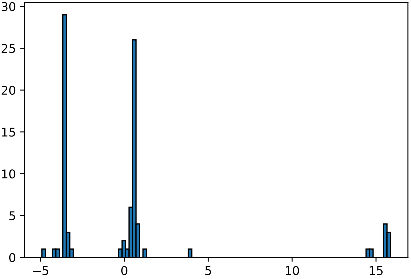
(a) Cost function values



(b) Image sources

Step 2 and 3: clustering projection and deducing dimensions

- compute the clusters of the projection on each e_i to get the room dimensions
- identify the central cluster (cluster containing the source) on each axis
- each dimension is the half of the distance between the two clusters surrounding the central cluster



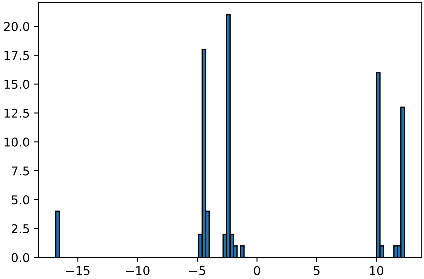


Figure: Projections on e_1, e_2

Conclusion and perspectives

The proposed method offers significant advantages for room geometry reconstruction :

- gridless, direct estimation of continuous 3D IS positions from discrete RIRs
- high precision recovery of low order image sources
- robustness to noise
- does not require prior information on the room properties

Conclusion and perspectives

The proposed method offers significant advantages for room geometry reconstruction :

- gridless, direct estimation of continuous 3D IS positions from discrete RIRs
- high precision recovery of low order image sources
- robustness to noise
- does not require prior information on the room properties

Some of the areas that remain to explore in this model:

- joint estimation of the source-microphone response κ
- taking into account the source and microphone directivities
- application to real data
- generalization to non-rectangular room shapes

Conclusion and perspectives

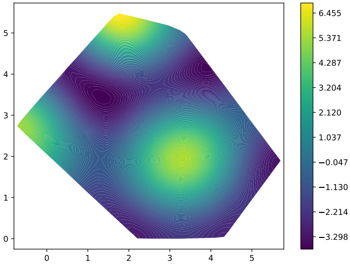
Another approach to handle non-rectangular geometries:

$$\inf_{\Omega \in \mathcal{O}_\ell^{\mathrm{adm}}} \sum_{m=1}^{M} \sum_{n=1}^{N} (\widetilde{p}_\Omega(k_n, \mathbf{r}_m^{\mathrm{mic}}) - \widetilde{x}_{m,n})^2 \, dt \tag{22}$$

where $\tilde{p}_\Omega$ is subject to the Helmholtz equation:

$$\begin{cases} \Delta \widetilde{p}(\mathbf{r}) + k_n^2 \widetilde{p}(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}^{\mathrm{src}}) & \mathbf{r} \in \Omega \\ \mathbf{n}(\mathbf{r}) \cdot \nabla \widetilde{p}(\mathbf{r}, t) = 0 & \mathbf{r} \in \partial\Omega \end{cases} \tag{23}$$

and $\mathcal{O}_\ell^{\mathrm{adm}} =$ set of polygons with at most ℓ vertices.



Thank you for your attention.

T. Sprunck, A. Deleforge, Y. Privat, and C. Foy. *Gridless 3D Recovery of Image Sources from Room Impulse Responses*, IEEE Signal Processing Letters, vol. 29, pp. 2427-2431 2022

A. Deleforge, C. Foy, Y. Privat, and T. Sprunck. *Can one hear the shape of a room? (article in preparation)*,