Master 2 mathematics -
Mathematical Modeling for signal and Image Processing.

# Matching Pursuit : A Greedy Algorithm for Hearing the Shape of a Room


Directors of internship:

M.DELEFORGE Antoine M.FRANCK Emmanuel.

Trainee since 1st April 2021

CHAHDI Khaoula


1 September 2021

# Contents

# Remerciement

The realisation of this internship was possible thanks to the help of several people to whom I would like to express my gratitude.

First of all, I would like to express my gratitude to my internship supervisor, Mr DELEFORGE Antoine, for his guidance and the knowledge he was able to pass on to me, as well as for his patience, availability and, above all, his judicious advice, as well as to my co-supervisor, Mr FRANCK Emmanuel, who guided me in my work and for the help he provided. I also thank him for his availability and the quality of his advice.

I would also like to thank Professor PROVENZI Edoardo for his advice, his patience and his availability.

I would like to thank all the staff of the Cerema and IRMA laboratories who, in one way or another, enabled me to work in the best possible conditions. I would like to express my sincere thanks to all the people who, through their words, their writings, their advice and their criticisms, guided my reflections and agreed to meet me and answer my questions during my internship.

To all of them, I offer my thanks, respect and gratitude.

# List of Figures

# List of Tables

# 1 Introduction

*Hearing the shape of a room* might sound like a literary poetic phrase but it is a scientific question that takes a large interest from scientists in the field of signal processing and architectural acoustics. Many animals like bats and dolphins and some birds are already using the sound or more precisely the echo of a sound for localisation and reconstructions of the surrounding, and some humans also use sound to hear a part of the shape of their way while moving.

So why reconstructing the geometry of a room from sound is getting such interest? If we are able to find the solution to this problem, the method will have many applications. For example in architectural acoustics, it would be easier to make a diagnosis of a room just using a sound source and a microphone array. It can be also useful to improve the simulation of rooms for 3D virtual and augmented reality reconstruction. It can also help getting the form of cavities that are not easily accessible, as in the case of historical monuments. That was the scientific motivation, but it is not only about developing new scientific methods for curiosity. If there is a method that meets the need, we can also save money that we lose on diagnostics that use expensive materials and are costly. So there is an economic benefit as well.

The question of using the sound to recover the form of an object remained unusual till Mark Kac decided to explore the possibility of estimating the shape of the outline of a drum from the sound it produces in his famous 1966 article "Can One Hear the Shape of a Drum ?" [7] which is a two-dimensional problem. The case of a three-dimensional room has been the subject of more recent works, but remains largely open to this day.

There are some works that have tried to address the issue of recovering the room geometry from a sound, as in the article "Acoustic echoes reveal room shape" Dokmanić et al. [6]. The authors use a specific method to answer this question, based on acoustic echoes and the *image source model* [2]. Using 5 microphones in their experiments, they tried to analyse the peaks of so-called *room impulse responses*, a type of acoustic measurement that will be introduced in details in the next section. For a given measurement, three steps are required to get the final result. First, peaks are extracted from it, which in this case was done manually. Then, the right association between peaks and walls must be found, a problem referred to as *echo labelling* which is combinatorially hard. Finally, triangulation is used to infer the wall positions. The authors obtained satisfying results when testing this approach to hear the shape of the Lausanne cathedral. However, it is still not fully automated and requires three independent steps that may be suboptimal, each error in a step leading to subsequent errors in the next ones. In addition, this approach strongly relies on the availability of broadband, high frequency signals in order to extract sharp peaks. A number of subsequent other articles used a simular 3 step approach, with the same limitations [14, 9, 3, 8].

To reconstruct the geometry of a room from an emitted sound and a microphone array, the choice of the microphone impacts also the quality of the 3D information in the signal. Hence, several researchers studied the correct shape and positions of microphones on the array, as detailed in the article of Park and Rafaely [11]

For our experiments, we will use room impulse responses obtained from a single point source and a compact spherical microphone array. Sound waves from a point source propagate as a sphere, they have a celerity and an energy that dissipates during the travel. The sound interacts with the surfaces of the surrounding, it can be transmitted, absorbed or reflected depending on the material.

Our method is based on the image source model. Hence we will use the reflections of the sound wave, also known as *echoes*. In many situations, echoes are seen as disturbances hindering sound understanding. However, the information they contain can help localization. For example, the times arrival of the reflected sound gives information about the distance travelled. In this work, we are interested only in the early reflections because they can be easily separated while the late reflections mix up in a phenomenon commonly known as *reverberation*.

The algorithm we use is an extension of Matching Pursuit [10], a greedy algorithm well known in the signal processing literature. It has notably been used for denoising, the resolution of linear inverse problems, and source separation. In this work, we will cast the problem of hearing the shape of a room as a unified inverse problem in which we are given acoustic measurements and we need to know the wall positions that match those measurements.

In Section 2.1 we introduce the necessary background in acoustics and signal processing, and present the proposed mathematical formulation of the problem at hand in the form of a single optimization problem. In Section 2.2 we present the proposed extension of the Matching Pursuit algorithm for the considered problem. In Section 3 we present extensive quantitative and qualitative results obtained with our method using a room acoustic simulator. Finally, Section 4 presents perspectives and leads for future work.

# 2 Background and Method

## 2.1 Problem statement

Sound waves propagate through any physical elastic medium at a speed c related to the medium. In room acoustics the sound is traveling within the air with a speed close to 343 meters per second at 20°C and thus the pressure field $u$ generated by this wave obeys the homogeneous wave equation :

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = 0. \tag{1}$$

In the presence of a perfect impulse sound emitted at time $t = 0$ by a point source placed at $\mathbf{r_0^{(S)}}$ in a free field (no walls) the resulting pressure field obeys the following inhomogeneous wave equation:

$$\frac{1}{c^2}\frac{\partial^2 u(\mathbf{r},t)}{\partial t^2} - \nabla^2 u(\mathbf{r},t) = a_0\delta(t)\delta(\mathbf{r}-\mathbf{r}_0^{(S)}). \tag{2}$$

Here, $a_0$ is a constant corresponding to an arbitrary mass flow rate at the source in $\mathrm{kg}\cdot\mathrm{s}^{-1}$. The value of this constant can be set to 1 without loss of generality, as pressures will only be considered up to a global normalising factor in this study.

The problem addressed in this work is the reconstruction of a convex polyhedral room from an emitted impulse sound propagating inside. Microphones are placed at $\mathbf{r}_1^{(M)}, ...$ $\mathbf{r}_M^{(M)} \in \mathbb{R}^3$. The sound source is placed in $\mathbf{r}_0^{(S)} \in \mathbb{R}^3$. In this situation, the pressure field verifies the wave equation with an additional rigid boundary condition of the form $\mathbf{n}\cdot\nabla u + b\frac{\partial u}{\partial t} = 0$ where $\mathbf{n}$ is the unit normal vector outward of the boundary and $b$ is a non-negative function defined on the boundary.

Such conditions do not lead to analytical solutions in the general case. To simplify the problem, we are going to use the image source model introduced in Allen and Berkley [2], which applies to the case of rigid, perfectly reflecting surfaces at the boundaries. Such surfaces induce specular reflections only, which is analogous to the light reflected by a mirror, *i.e.*, the incidence angle is equal to the reflection angle as in the Snell–Descartes's law. The reflected ray or the *echo* then seem to be emitted from a hidden *image source* behind the wall. The position of the wall is hence the bisector between the real source in the room and the image source. Then, the first order image sources are the six reflections from the walls as shown in figure 1.

To obtain a higher order of reflections, we can consider every single image source as the initial source and look for its symmetries. In the following we consider a *shoebox* (or cuboid) room which is a simple model of convex rooms. Indeed, most rooms and offices are rectangular and in the case of rigid surfaces the image solution rapidly approaches an exact solution of the wave equations. Nevertheless, the proposed methodology is general and could be applied to any convex room.

In this simplified case, the pressure field verifies the following wave equation :

$$\frac{1}{c^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = \sum_{k=0}^{K} a_k\delta\left(\mathbf{r}-\mathbf{r}_k^{(s)}\right)\delta(t) \tag{3}$$

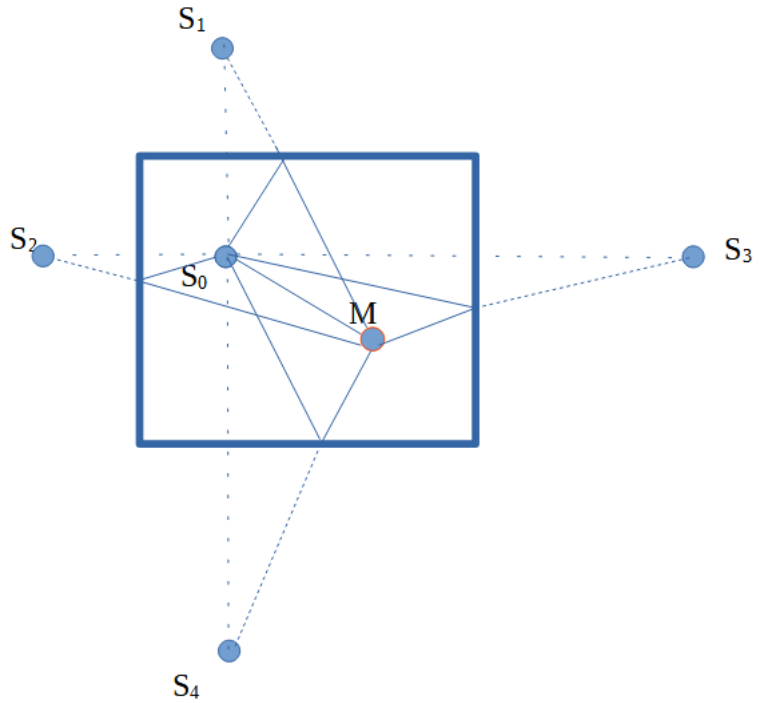


Figure 1: First order image sources.

where:

- $a_k$ depends on the absorption of the walls.
- The wall is the bisector of $(\mathbf{r}_0, \mathbf{r}_k)$ k=1,...,6.

The image source model reveals that finding the positions of the 6 walls in a shoebox room is equivalent to the problem of localizing in 3D the emitting source and its 6 first-order image sources.

Note that here and in the remainder of this thesis, we will only consider first order reflections and neglect higher order ones. This gives $K = 6$ for a shoebox room. While this is an strong simplification, it is justified by the fact that higher order echoes are less energetic and occur later in recorded signals. In a real room, $K$ is in fact infinite.

The classical solution of the corresponding Helmholtz equation in the Fourier domain gives the solution:

$$u(\mathbf{r}, t) = \sum_{k=0}^{K} a_k \frac{\delta\left(t - \left\|\mathbf{r} - \mathbf{r}_k^{(s)}\right\| / c\right)}{4\pi \left\|\mathbf{r} - \mathbf{r}_k^{(s)}\right\|} \tag{4}$$

The recording of the response to a point source emitting a perfect impulse in a room by a microphone placed at $\mathbf{r}_m^{(M)}$ is called a *room impulse response* (RIR). RIRs are among the most commonly used measurements in room and building acoustics, and there exist a number of techniques and protocols to reliably estimate them, see *e.g.* [13]. An impulse response is the output signal of a linear time-invariant system (such as a room) that is

obtained from an temporal impulse at the input, i.e. a sudden and brief variation of the signal, modeled as a Dirac. It characterize the system in the sense that if we know the impulse response of a system $g$, the output of this system given any signal $v$ at its input will simply be the convolution $g * v$.

The image source method can be explicitly linked to room impulse response measurements based on equation (4). Each peak in the RIR represents either the direct path of the emitted sound or its echoes that corresponds to peaks with a delay and an attenuation. Hence the signal received by the microphone antenna is a constellation of all rays coming from $K$ image sources of all orders namely the direct path and all the echoes of the emitted sound.

In practice, microphone does not directly record the sound field, but a filtered and discretized version of the pressure signal at the microphone location:

$$x_{m,n} = (\phi * u(\mathbf{r}_m^{(M)}, .))(n/f_s)$$

where: $\phi(t) = \text{sinc}(t) = \frac{\sin(t)}{t}$ is the chosen filter in this case. This corresponds to an ideal low pass filter with cutoff frequency $\frac{f_s}{2}$.

Then, for a microphone $m$, the signal observed at $t = n$ is:

$$x_{m,n} = \sum_{k=0}^{K} \frac{a_k}{4\pi \|\mathbf{r}_m^{(M)} - \mathbf{r}_k^{(S)}\|} \text{sinc}\left(\frac{n}{f_s} - \frac{\|\mathbf{r}_m^{(M)} - \mathbf{r}_k^{(S)}\|}{c}\right) \stackrel{\text{def}}{=} \sum_{k=0}^{K} a_k h_{m,n}(\mathbf{r}_k^{(S)}) \tag{5}$$

We denote by $\mathbf{x} = [x_{m,n}]_{n,m} \in \mathbb{R}^{MN}$ the vector of all observations at all microphone $m = 1 \dots M$ and all discrete time samples $n = 1 \dots N$. Then:

$$\mathbf{x} = \sum_{k=0}^{K} a_k \mathbf{h}(\mathbf{r}_k^{(S)}) \tag{6}$$

where $\mathbf{h} : \mathbb{R}^3 \longrightarrow \mathbb{R}^{MN}$. In order to estimate the position of the source and its images from recorded impulse responses, and hence, *hear the shape of the room*, we aim to find a solution to the following optimisation problem:

$$\underset{\mathbf{r}_0,\dots,\mathbf{r}_K \in \mathbb{R}^3,\, a_0,\dots,a_K \in \mathbb{R}}{\text{argmin}} \left\| \mathbf{x} - \sum_{k=0}^{K} a_k \mathbf{h}(\mathbf{r}_k) \right\|_2^2. \tag{7}$$

The key difficulty is that $\mathbf{h}$ is a non-linear non-convex function, so it is hardly invertible, and the algorithms based on gradient descent are not applicable because of local minimum hindrance.

To alleviate this, we explored the possibility of solving this problem on a grid. Let $G = \left\{\mathbf{r}'_{\mathbf{p}}\right\}_{p=1}^{P} \subset \mathbb{R}^3$ be a grid of 3D points. The main idea is to find the K+1 positions on this grid that minimize our cost function (7) and hence correspond to the nearest positions to the image sources of the room. Then, the problem turns from finding $K$ positions in all of $\mathbb{R}^3$ which is an infinite space to the finite space $\mathcal{G}$ limited by a predefined number $P$ of points. Therefore, we will optimize the following problem:

$$\underset{\mathbf{r}_0,\ldots,\mathbf{r}_K \in G,\, a_0,\ldots,a_K \in \mathbb{R}}{\operatorname{argmin}} \left\| \mathbf{x} - \sum_{k=0}^{K} a_k \mathbf{h}(\mathbf{r}_k) \right\|_2^2 .$$

Let $\mathbf{D} = [h(\mathbf{r}'_1), \ldots, \mathbf{h}(\mathbf{r}'_P)] \in \mathbb{R}^{MN \times P}$, then the problem becomes equivalent to solving:

$$\underset{\mathbf{z} \in \mathbb{R}^P \; s.a \; \|\mathbf{z}\|_0 = K+1}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2,$$

where the $K+1$ non-zero coefficients of the sparse vector $\mathbf{z}$ are $\{a_0, \ldots, a_K\}$. In the following, the matrix $\mathbf{D}$ will be referred to as a *dictionnary* and its columns $\mathbf{d}_i$ as *atoms*, following the litterature on sparse estimation, *e.g.* [10].

To approximate the real positions, we need a large grid $G$ with a sufficient number of points to cover the volume of the room and the surrounding. If we choose 100 point for each direction of the grid, we can easily reach one million point. Choosing $K+1$ position from $P = 1,000,000$ seems to be an intractable combinatorial problem since there is $\begin{pmatrix} P \\ K+1 \end{pmatrix}$ possibilities. This appears like looking for a needle in a haystack. To tackle this task, we will make use of a greedy algorithm to find the sparse vector $\mathbf{z}$. We choose to proceed with the Matching Pursuit algorithm.

## 2.2 Matching Pursuit a greedy iterative algorithm

The strategy of Matching Pursuit is to estimate the coefficients $a_k \in \mathbb{R}$ and the corresponding indices $i_k \in [|1, P|]$ of dictionary columns one by one, without going back on each choice, which is why the algorithm is called "greedy". It was first introduced for signal processing by Mallat and Zhang [10]. It is an iterative algorithm that takes the signal $\mathbf{x}$ and a dictionary $\mathbf{D}$ as input, then selects the corresponding atoms of $\mathbf{D}$ to reconstruct $\mathbf{x}$. In other words, $\mathbf{x}$ is a weighted sum of a finite number of atoms from the dictionary $\mathbf{D}$. Atoms from $\mathbf{D}$ correspond to RIRs calculated at each position on the grid which are seen as a source position. Using a correlation function, the algorithm can select iteratively from $P$ points in $\mathcal{G}$ the list of $K+1$ coefficients or weights and indices for the corresponding atoms which are the output. The algorithm hence chooses the indices that contribute to the reconstruction of the received signal. In summary, we have $\|\mathbf{x} - \mathbf{D}\mathbf{z}\|_2^2 = \|\mathbf{x} - \sum_{k=0}^{K} a_k \mathbf{d}_{i_k}\|_2^2$, where $\mathbf{d}_{i_k} = \mathbf{h}(r_{i_k})$ with $\mathbf{r}_{i_k} \in \mathcal{G}$. This is illustrated in figure 2.

The proposed extension of Matching Pursuit for our problem is described in Algorithm 1 in the form of pseudo-code.

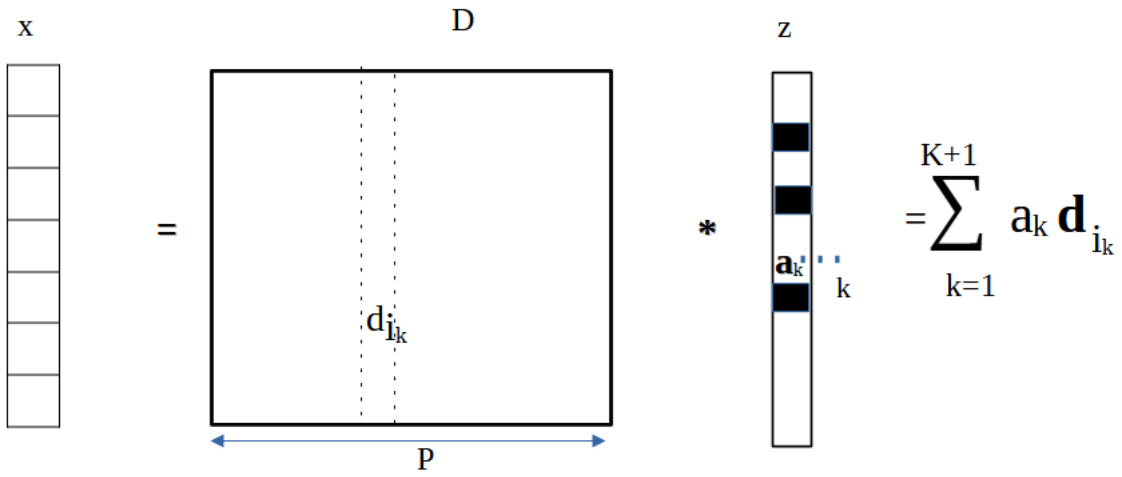


Figure 2: Reconstruction of $\mathbf{x}$ from atoms of $\mathbf{D}$

**Algorithm 1** Matching Pursuit

**Input:** $x$, $\mathbf{D} \in \mathbb{R}^{MN}$ $K$, $rad_{sep}$, $ang_{sep}$.
**Output:** $\{a_k\}_{k=0}^{K}, \{\mathbf{r}_{i_k}\}_{k=0}^{K}$
// Initialisation:
1: $\mathbf{g} \longleftarrow \mathbf{x}$
2: $\overline{\mathbf{D}} \longleftarrow \mathbf{D}$
3: **for** $k = 0, \ldots, K$ **do**
4: $\quad a_k, i_k \longleftarrow \underset{i \in [|1,P|]\,;\, a \in \mathbb{R}}{\operatorname{argmin}} \|\mathbf{g} - a\mathbf{h}(\mathbf{r}_i)\|_2^2$.
5: $\quad \mathbf{g} \longleftarrow \mathbf{g}$ - $a_k \mathbf{h}(\mathbf{r}_{i_k})$
6: $\quad \overline{\mathbf{D}} \longleftarrow \overline{\mathbf{D}} \setminus \{\mathbf{h}(\mathbf{r}) : \mathbf{r} \in \mathcal{G},\ gc_{dist}(\mathbf{r}, \mathbf{r}_{i_k}) < ang_{sep},\ |\|\mathbf{r}\|_2 - \|\mathbf{r}_{i_k}\|_2| \leq rad_{sep}\}$
7: **end for**
8: **return** $\{a_k\}_{k=0}^{K}, \{\mathbf{r}_{i_k}\}_{k=0}^{K}$

In **Line 4**, $i_k$ is the index $i$ maximizing the inner product $\dfrac{|\langle \mathbf{g}, \mathbf{h}(\mathbf{r}_i)\rangle|}{\|\mathbf{h}(\mathbf{r}_i)\|^2}$, and $a_k = \dfrac{\langle \mathbf{g}, \mathbf{h}(\mathbf{r}_{i_k})\rangle}{\|\mathbf{h}(\mathbf{r}_{i_k})\|^2}$.

Preliminary test of Matching Pursuit showed that the algorithm tended to estimate many positions clustered around the image sources with strongest amplitudes, and hence missed some of the image sources with lower amplitudes. This is a well known limitation of this type of on-grid method, see Denoyelle et al. [4]. To overcome this issue, we implemented a procedure iteratively removing points from the grid contained in a *safe zone* with a predefined angular width $ang_{sep}$ and radial width $rad_{sep}$ around each of the detected position. This can be seen in **Line 6**, of algorithm 1, where some columns are progressively removed from the searched $\overline{\mathbf{D}}$, initialized by $\mathbf{D}$ in **Line 2**. We do not only remove the position in $\mathcal{G}$ corresponding to the selected index $i_k$, but we remove all the points $\mathbf{r} \in \mathcal{G}$ around it that verify: $gc_{dist}(\mathbf{r}, \mathbf{r}_{i_k}) < ang_{sep}$ and $|\|\mathbf{r}\|_2 - \|\mathbf{r}_{i_k}\|_2| \leq rad_{sep}$ where $gc_{dist}$ is the great circle distance which is the shortest distance between two points on the surface of a sphere, and $\|\|_2$ is the euclidean norm.

We explored two types of grid. First we considered a Cartesian grid which is a grid with regular steps along the axis X,Y and Z with Cartesian coordinates (x,y,z), that did not lead to satisfying results in preliminary experiments. This can be explained by the fact that the angular resolution is related to the microphone array used while the radial resolution is related to the frequency of sampling of the microphones $f_s$ only. We concluded that we needed to control both types of resolution which is not possible with a Cartesian grid. Hence, a spherical grid seemed like a more natural choice. We considered angles uniformly spaced on the whole sphere, to avoid condensed points at the poles, using a uniform angular steps. We use the spherical coordinates system to fill in the grid's positions $(r, \theta, \phi)$. The radius $r$ varies from 1 to 15 meters with a certain radial resolution $\Delta r = rad_{step}$. We used an elevation angular step $ang_{step} = \Delta\phi$ where $\phi \in [0, \pi]$ is the elevation. The azimuth angular step $\Delta\theta$ is then a function of the angle $\phi$ and $ang_{step}$ such that: $\Delta\theta = ang_{step} \times \cos(\phi - \pi/2)$. In the next section, several radial and angular grid resolution $rad_{step}$ and $ang_{step}$ will be tested.

# 3 Experiments

## 3.1 Tools and hypothesis

The experiments of this internship are all implemented based on the room simulator Pyroomacoustics, which is a python package for audio room acoustics introduced in Scheibler et al. [12].

We did our experiments under simplifying hypothesis to explore first results, with the idea that if the results are encouraging we can add more realistic hypothesis in future works. First, we only considered shoebox rooms, i.e., cuboid rooms, since most of rooms and offices in the majority of contemporary buildings are rectangular. Hence we will have six surfaces to estimate. We choose frequency-independent absorption coefficients fixed to 0.1 for each of the 6 walls. Such absorption profiles are typical of highly reflexive materials such as tiles or concrete. Moreover, we only consider specular reflections while neglecting all the diffuse ones. As explained in the previous section, we use the image source method for simulation, and since the acoustic echoes of first order are sufficient to estimate the geometry of a room, we limit ourselves to the first order. So our aim is to find the position of the true source together with the six positions of the image sources given by the first order acoustic echoes using our Matching Pursuit algorithm.

For the microphone array, we used the geometry of the commercially available spherical array of 32 microphones embedded in a rigid sphere called eigenmike Acoustics [1] patented by mhacoustics, as shown in figure 3.

Figure 3: The Eigenmike used for simulating the RIRs

As we can see this microphone array is compact not bulky and small which make it

practical for measurement and displacements. Its radius, as cited in Acoustics [1], is about 4.2 cm. Hence, the diameter of the microphone antenna is 8.4 cm. However,in our experiments, we tested different scaling of the array to see if the size of the microphone array influences our results.

**Remark:** For the microphone array we gave only a file containing all the coordinates of its 32 microphones to the pyroomacoustics simulator. We did not do any hand manipulation with the real microphone antenna for this preliminary study.

## 3.2 Parameters

For each experiment we generated 100 room geometries width length, with and height respectively picked uniformly at random in $[4, 8] \times [4, 8] \times [3, 6]$ meters. This corresponds to dimensions of typically encountered rooms in buildings. For each room, the source and receiver are also positioned at randomly peaked positions while respecting some constraints. First, both of the positions must be far from all of the 6 walls at least by $1m$. Second, they must be distant from each other by at least $1m$. This is do avoid situations where some acoustic reflections would be highly prominent with respect to others, and also corresponds to standard measurement protocols in acoustics.

We fixed the the minimum radius $r_{min}$ and the maximum radius $r_{max}$ of our spherical grid $\mathcal{G}$ respectively at 1 and 15 m. Since the minimum distance between the microphone array and the sound source is 1 m that justifies our choice of $r_{min}$. Meanwhile, $r_{max}$ is the maximum radius that an image source can reach for the configuration of dimensions that we previously defined. So we want to detect all image sources situated between $r_{min}$ and $r_{max}$ on our grid $\mathcal{G}$

The only grid parameters we vary are the radial resolution $rad_{step}$ given by $\Delta r$, the step in the radial direction, and the angular resolution $ang_{step}$ for both angular directions as detailed in the last paragraph of section 2. We will also manipulate the microphone size and the frequency of sampling $f_s$.

### 3.2.1 Default parameters

For all the experiments, we define a set of default values for parameters $rad_{step}$, $ang_{step}$, Mic size $\times$ and $f_s$ as shown in Table 1.

| $rad_{step}$ | $ang_{step}$ | Mic size $\times$ | $f_s$ |
|---|---|---|---|
| 0.2 m | 5° | 5 | 4 kHz |

Table 1: Default parameters of the experiments

As we can see in Table 1, $rad_{step}$ is set to 0.2m, which means that we are looking for image sources in every 20 cm in the radial direction. $ang_{step}$ is set to 5°, which is equivalent to check positions within 5 degrees in both of the angular directions. For Mic size $\times$, we choose to multiply the microphone size by 5, yielding a diameter of 42cm, which remains relatively compact.

The frequency of sampling $f_s$ is fixed to 4000Hz. Since most microphones can record at 44.1 or 48 kHz, this is a relatively low frequency. However, many commonly encountered sound sources such as *e.g.* human voice are most energetic under 2 kHz. Since higher frequencies are not available in such signals, using higher sampling rates would be redundant. The $f_s$ parameter will test the ability of our method to work in challenging, low frequency regimes.

In practice, all the RIRs were cut to a fixed duration of $t_{max} = r_{max}/c = \frac{15}{343} \approx 43.75$ ms, where $c$ is the speed of sound. This corresponds to only keeping information from sources at a distance smaller than $r_{max} = 15$ m.

### 3.2.2 Evaluation metrics

**Metrics**

To evaluate the proposed method we used two different metrics defined as follows: the true positive rate

$$TPR = \frac{\text{The number of True Positives}}{\text{The number of true sources}},$$

and false positive rate

$$FPR = \frac{\text{The number of False Positives}}{\text{The number of estimated sources}},$$

where :

- The number of True positives is the number of true sources detected by the algorithm according to predefined thresholds,
- The number of true sources is the number of image sources corresponding to each wall which is 6 plus the real source which make a total of 7 true sources,
- The number of false positive is the number of sources selected by the algorithm that exceed the thresholds we fixed,
- The number of estimated sources is the total number of sources selected by the algorithm. In other words, it is the number of iteration $K + 1$ in Matching Pursuit.

**Thresholds**

The thresholds we fixed correspond to the precision with which we want to detect the sources. In this case we fixed a radial threshold $rad_{thresh}$ at $25cm$ and a angular threshold $ang_{thresh}$ at 10°. We will examine the influence of these thresholds on metrics later **How the sources are sorted**

After recovering the output results from Matching Pursuit and fixing the values of our thresholds, we use the great circle distance to calculate the angular distance between the real sources of the room and the estimated sources of Matching Pursuit. If the angular distance is below the $ang_{thresh}$ we fixed and if the radial distance is less than $rad_{thresh}$ we say that it is a true positive, if only one of the two conditions is not verified we say that the estimated source is a False positive.

As we do the calculations over 100 room in each experiment, we take the mean values of the metrics of over 100 rooms.

### 3.2.3 Experiment with defaults parameters

For the first experiment, we used the set of defaults parameters listed in table 1. The chosen radial and angular resolution yielded a grid of 115,659 points. The RIRs length corresponding to $t_{max}$ is about $5,600$ samples. This gives a dictionary of size $115,659 \times 5,600$ which can yield to computer memory issue without some careful implementation. Hence, in practice, we divided our dictionary into $Q$ sub-dictionaries (in this experiment $Q = 71$) of smaller size, which were independently screened at each iteration of Matching Pursuit.

| $rad_{step}$ | $ang_{step}$ | Mic size $\times$ | $f_s$ | $rad_{thresh}$ | $ang_{thresh}$ | TPR | FPR |
|---|---|---|---|---|---|---|---|
| 0.2m | 5° | $5 \times 8.2cm$ | 4 kHZ | 0.25 | 10° | 54% | 46 % |

Table 2: Results of the first experiment.

The results are showed in table 2. As we can see, for the radial resolution $rad_{step} = 0.2m$ and the angular resolution $ang_{step} = 5°$ and a microphone array's size amplified 5 times at the frequency of sampling $f_s = 4000Hz$, we obtain a TPR of 54% and a FPR of 46%. This means that for our predefined metric and thresholds set at 25 cm for $rad_{thresh}$ and 10° for $ang_{thresh}$, Matching Pursuit is able to recover 54% of sources with high precision. This is equivalent to $54\% \times 7 = 3.75 \approx 4$ sources out of 7 per room. Hence, the algorithm found the true source and 3 image sources which is the same as locating 3 walls, since a wall is the bisector between the real source and an image source. At first glance, these results may give the impression that the algorithm works only 50% of the time like for heads or tails, but in fact, finding 4 out of 7 sources within a grid of 115,659 point at a precision of 25cm and 10° is more akin to successfully finding a needle in a haystack.While not perfect, these results show the ability of the proposed technique to partially "hear the shape" of the room, with much higher precision than chance.

For the remainder of this section we will focus for conciseness on the TPR metric as it is easier to read and interpret.

**Remark** In practice, the computational time of the method, coded in python, was about 12 hours to pre-compute the dictionary and another 12 hour for running the Matching Pursuit algorithm on 100 rooms, using the CPU of a regular laptop and no specific optimization.

**The influence of thresholds and precision**

To see the impact of thresholds on the results, we fixed all the parameters as in the table 2 except for the angular threshold which was set to 30° as shown in table 3.

We can see that the TPR rate have increased to become 60% which is equivalent to find 4.2 sources of 7. Hence increasing $ang_{thresh}$ may increase the number of sources we find, and we can deduce the same thing for $rad_{thresh}$ as they play a symmetrical role. All depend on at which precision we want to evaluate our algorithm. For the rest of the

| $rad_{step}$ | $ang_{step}$ | Mic size × | $f_s$ | $rad_{thresh}$ | $ang_{thresh}$ | TPR | FPR |
|---|---|---|---|---|---|---|---|
| 0.2m | 5° | $5 \times 8.2cm$ | 4kHZ | 0.25 | 30° | 0.60% | 0.4 % |

Table 3: The fist experiment results after varying the angular threshold

experiments we fixed our thresholds at 25 cm for $rad_{thresh}$ and 10° for $ang_{thresh}$. In the following we are rather interested in the influence of the other parameters on the precision of our method.

## 3.3 Influence of algorithmic and measurement parameters

In this part, we will discuss the results of all experiments where we tested the influence of 4 different parameters on the performance of the method.

### 3.3.1 The influence of the frequency sampling on the method

In this experiments we vary the frequency of sampling $f_s$ from 2kHz to 8kHz while the other parameters are set to default values. We tested $f_s$ values as described in Table 4.

| $f_{s1}$ | $f_{s2}$ | $f_{s3}$ | $f_{s4}$ |
|---|---|---|---|
| 1000 Hz | 2000Hz | 4000 Hz | 8000 Hz |

Table 4: $f_s$ values tested with the other default parameters.

The results of those experiments are summarized in table 5.

| (TPR) $f_{s1}$ | (TPR) $f_{s2}$ | (TPR) $f_{s3}$ | (TPR) $f_{s4}$ |
|---|---|---|---|
| 0.3% | 61% | 54% | 32% |

Table 5: Influence of $f_s$ on the True Positive Rate.

Figure 4 shows the influence of varying the frequency of sampling from 2000 to 8000 Hz. We can see that at $f_s = 1000Hz$ the method failed dramatically and did not find any source position. Then we tried $f_s = 2000Hz$ which gives 61% of sources. Hence at this frequency the method seems to be the most efficient. As a last trial we calculated the TPR rate for $f_s$ = 8000Hz which gave 32% of sources. Interestingly, we can conclude that the intuitive hypothesis that increasing the frequency of sampling would give better results fails. We can rather say that there is an optimal frequency of sampling for every predefined grid $\mathcal{G}$. In fact, the frequency of sampling and the radial resolution are highly related. The sharpness of the peaks in the RIRs depends on the frequency of sampling $f_s$: the higher it is, the sharper the peaks. Considering the grid $\mathcal{G}$ if the peak falls directly on a point of the grid it will be easily detectable but if it is between two points it is less likely to be selected. To conclude, we can say that the frequency of sampling influences highly the results of our method and that $f_s = 2000Hz$ fits better our predefined grid $\mathcal{G}$.

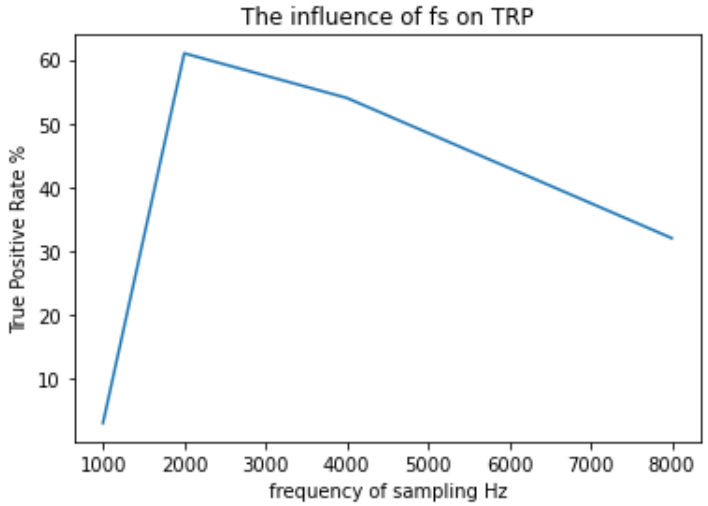


Figure 4: Influence of $f_s$ on the True Positive Rate.

### 3.3.2 The influence of the microphone size on the method

In this experiments we vary the microphone array size (Mic size $\times$) from $1 \times 8.4cm$ to $10 \times 8.4cm$ while the other parameters are set to default values. We tested Mic size values as described in Table 6.

| Mic size 1 | Mic size 2 | Mic size 3 | Mic size 4 |
|---|---|---|---|
| $1 \times 8.4cm$ | $2 \times 8.4cm$ | $5 \times 8.4cm$ | $10 \times 8.4cm$ |

Table 6: Mic size $\times$ values tested with the other default parameters.

The results of those experiments are summarized in Table 7.

| (TPR) Mic size 1 | (TPR) Mic size 2 | (TPR) Mic size 3 | (TPR) Mic size 4 |
|---|---|---|---|
| 0.1% | 1.4% | 54% | 55% |

Table 7: Results of varying the microphone array size .

Figure 5 shows the influence of varying the microphone array size from $1 \times 8.4cm$ to $10 \times 8.4cm$. We can see that for the Mic size 1 which is exactly the real size of the microphone array 8.4 cm, the method failed to select any point source. Then we considered a mic size two times bigger, *i.e.*, an antenna diameter of $16.8cm$. For this experiment the (TPR) rate is around 1.4%. Hence even for this mic size, the method failed. For the third experiment we took a microphone array five time bigger than the original one, the mic diameter is then equal to $42cm$ this time the (TPR) rate is about 54%. We can see clearly the big jump of the (TPR), so the microphone array size is also a critical parameter to check. After this satisfying result we tend to say that the larger the mic size, the better the precision will be. In this final experiment on mic size the diameter this time is 10 times the original version, so it is about $84cm$. The (TPR) of the method using this size of microphone array is approximately 55%. Comparing to the results of Mic size 3, we notice that there is not a large difference. Since the results are almost the same for Mic

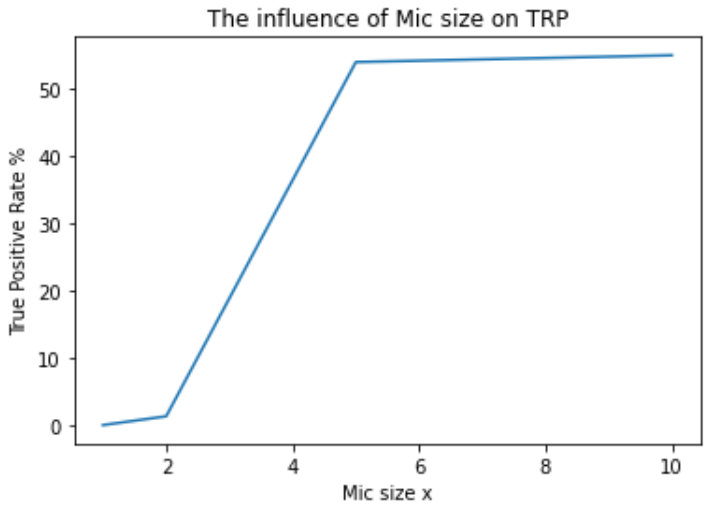


Figure 5: Results of varying the microphone array size.

size 4 and Mic size 3, it is more practical to choose Mic size 3 for our microphone array because it is two times smaller than Mic size 4 and it generates a good results and it is ergonomic.

To conclude, mic size is an important parameter that influences the precision of our method. This due to the fact that the microphone antenna is strongly linked to the angular resolution in sound source localization. In our experiments the microphone array we use is a compact rigid sphere of 8.2cm diameter and contains 32 microphones. That means that the microphones are so close that they cannot distinguish the times of arrival of sound sources for the considered frequency rangr . In other words, a too small diameter for the 32 microphones becomes equivalent to having a single microphone for our method, and for 3D localization we need more than one microphone. This explains why we obtain better results with a larger microphone array.

### 3.3.3 The influence of the radial grid resolution on the method

In this experiments we vary the radial resolution $rad_{step}$ of the grid from $0.1m$ to $1m$. The smaller the step, the denser is the grid and vice versa. We set the other parameters at default values. We tested radial resolution values as described in the following Table 8.

| $rad_{step1}$ | $rad_{step2}$ | $rad_{step3}$ | $rad_{step4}$ |
|---|---|---|---|
| 0.1 m | 0.2 m | 0.4 m | 1 m |

Table 8: $rad_{step}$ values tested in 4 experiments with the other parameters set to default values.

The results of those experiments are summarized in Table 9 below.

| (TPR) $rad_{step1}$ | (TPR) $rad_{step2}$ | (TPR) $rad_{step3}$ | (TPR) $rad_{step4}$ |
|---|---|---|---|
| 83% | 54% | 33% | 18% |

Table 9: Results of varying the radial resolution.

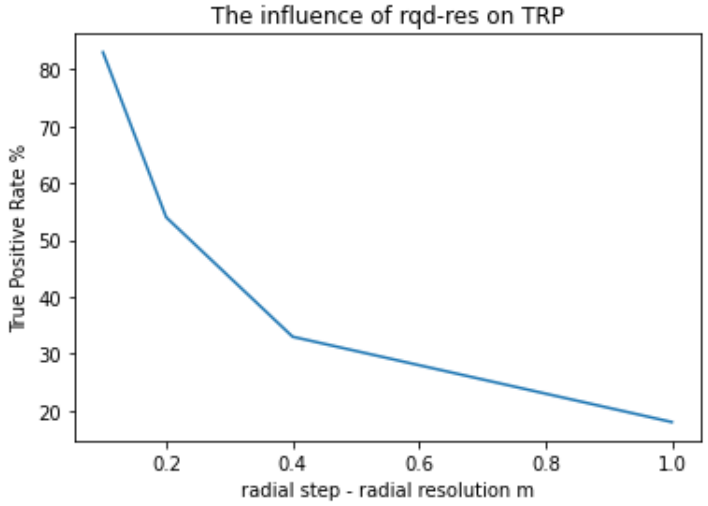


Figure 6: Results of varying the radial resolution.

Figure 6 shows the influence of varying the radial resolution from $0.1m$ to $1m$. We started with small steps in the radial direction, the smallest the step, the higher the resolution will be. This means that for the smallest step the grid $\mathcal{G}$ contains a higher number of points that the algorithm have to check. We can see in the curve that for the highest resolution we tested $rad_{step1} = 0.1$, the method succeeded to find around 83% of sources, this is equivalent to $0.83 \times 7 = 5,81 \approx 6$ sources out of 7 which is a very satisfying result. It means that we found 5 walls out of 6. With $rad_{step2} = 0.2m$ we obtained a TPR rate of 54% which means 3 walls of 7. Hence, it suggests that decreasing the radial resolution impacts the quality and the precision of our method. We can check this hypothesis by making more experiments and decreasing the radial resolution once again to $rad_{step3} = 0.4m$. The result of this experiment shows that at this radial resolution, the method can detect only 33% of sources, it is about $0.33 \times 7 = 2,31 \approx 2$ sources and only 1 wall, which supports our hypothesis. We did a final experiment on the $rad_{step}$ to have a clear conclusion, this time we fixed it to $rad_{step4} = 1m$. The number of true positive sources detected at this resolution is $0.18 \times 7 = 1.26 \approx 1$ source and no walls detected.

Overall, we see clearly that the radial resolution $rad_{step}$ is a critical parameter and to achieve our goal which detect the geometry of the room we need to work at high radial resolution. Because we have constraints of computer memory and time calculation we fixed our default parameter only at $rad_{step} = 0.2m$. Indeed, the higher the radial resolution, the higher number of points in the grid, which significantly increases computational time and memory costs.

### 3.3.4 The influence of the angular resolution on the method

In this experiments we vary the angular resolution $ang_{step}$ from 5° to 20°. We set the other parameters at default values. We tested angular resolution values as described in the following table 10.

| $ang_{step1}$ | $ang_{step2}$ | $ang_{step3}$ | $ang_{step4}$ |
|---|---|---|---|
| 5° | 10° | 15° | 20° |

Table 10: $ang_{step}$ values tested in 4 experiments with the other parameters set to default values.

The results of those experiments are summarized in the table angInflRes.

| (TPR) $ang_{step1}$ | (TPR) $ang_{step2}$ | (TPR) $ang_{step3}$ | (TPR) $ang_{step4}$ |
|---|---|---|---|
| 54% | 51% | 41% | 33% |

Table 11: Results of varying the angular resolution.

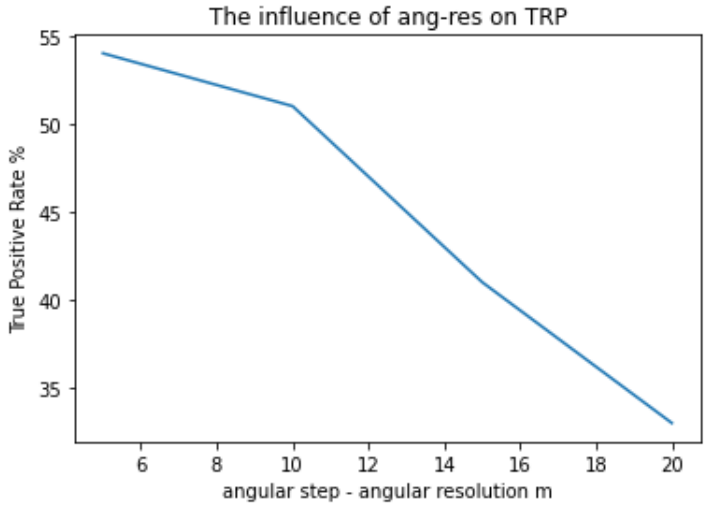


Figure 7: Results of varying the angular resolution.

Figure 7 shows the influence of varying the angular resolution from 5° to 20° this parameter plays a similar role to the radial resolution. As we can see, the curve is decreasing which means that the true positive rate is decreasing while reducing the angular resolution as in the case of the radial resolution curve. For the fist experiment the angular resolution is fixed to $ang_step1 = 5°$, the method selected 54% of sources. We then reduced the angular resolution to $ang_{step2} = 10°$ which gave 51% of sources, which is not a significant difference from the previous result. Let up keep reducing the $ang_{step}$ and this time fix it to 15°. The (TPR) rate then drops to about 41%. This is equivalent to finding approximately 3 sources or 2 walls. In the last experiment, we fixed $ang_{step}$ to 20°. The result of the (TPR) is around 33%, which is equivalent to finding 1 wall on average.

We can clearly see that our method is also sensitive to the angular resolution paramete. Hence, it must be chosen carefully as a compromise between computational efficiency and precision.

## 3.4 Qualitative results

In this part we will finally show some qualitative results for a specific room, whose geometry is showed in Figure 8.

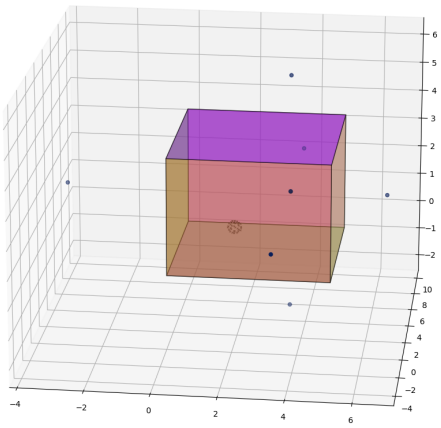

Figure 8: Representation of a room of dimension $5 \times 7 \times 4m^3$. The microphone array is fixed at (2, 1, 1.5) and is represented by the small sphere in the center of the reference frame, while the sound source is placed at (3.5, 4, 2). The blue points are the image sources for this room.

For this experiment, we use the default parameters described in the table 1. We can see visually the true and the estimated sources of this room in Figure 9.

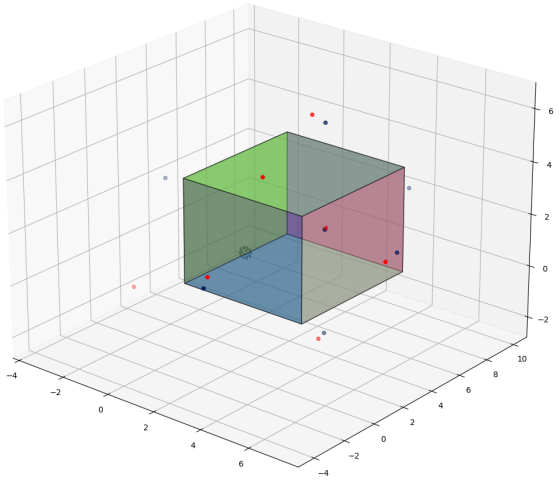

Figure 9: The true image sources of the room are in blue color while the estimated ones are red

To see visually how close the estimations were, for each sphere of our spherical grid, we display as a colormap the score of the Matching Pusruit algorithm during the first iteration, *i.e.*, $\frac{|\langle \mathbf{x}, \mathbf{h}(\mathbf{r}_i) \rangle|}{\|\mathbf{h}(\mathbf{r}_i)\|^2}$. In the following we will show only the colormaps of the spheres that contains important information.

In the figure 10 above, we see that in the 13th sphere

which corresponds to a radius of 3.4 m, the bright zone is the area where the scores are higher than the rest, the red point here is the real source and the magenta cross shape represents the estimated source using Matching Pursuit. It appears that the estimated point source corresponds to the real source in that case.

Let us move to the next interesting sphere, which is the 20th one, situated at a radius of $4.8m$ and showed in Figure 11. As already mentioned, the brightest zone is the area with higher calculated scores. We can see that the estimated source in magenta cross shape is closely fitting the red real source, hence the method succeeded to detect this source. Then since we have the real source and an image source we can affirm that we localized a wall. The brightness of points in this sphere is less than the previous one. This may be because the intensity of the signal of that the peak of the RIRs is closer to the grid but did not reach it, and also due to the attenuation of sound as a function of distance and wall absorption.

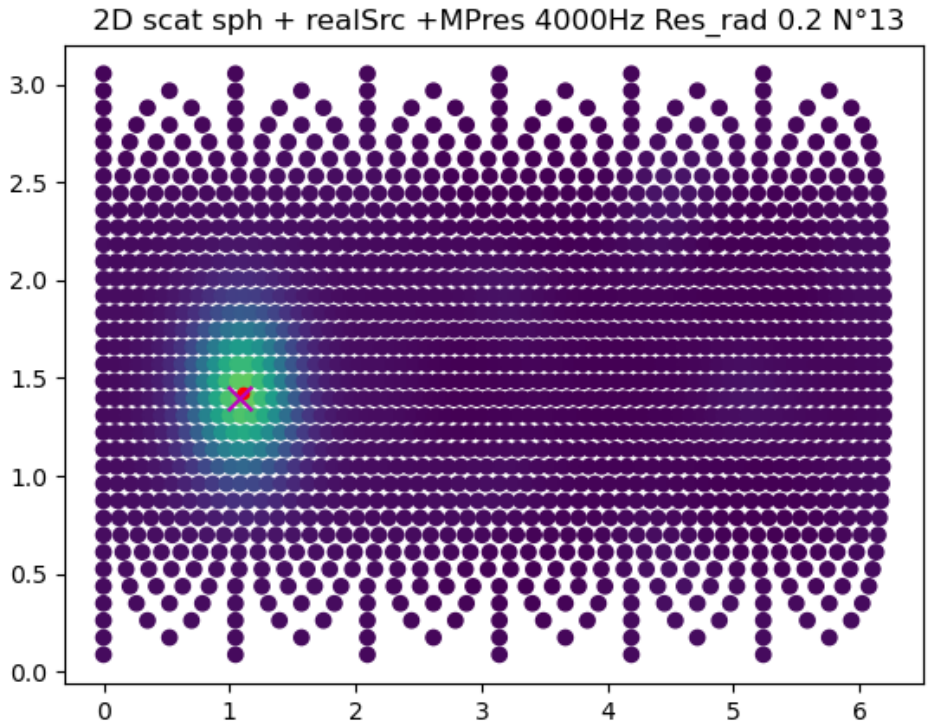


Figure 10: The colormap score of the first interesting sphere where the real sound source is detected.

Now let's see the sphere number 22 situated at the radius $5.2m$ and displayed in Figure 12.

For this sphere, the method also found a point source close to the real source, hence we can count 2 walls detected for the moment.

The following sphere is the 23th one situated just next to the previous sphere at a radius of $5.4m$ and showed in the figure 13 below:

The figure 13 above is a little bit different. The effect of artefacts is clearly visible on the colormap score of the points of this sphere. We can observe many bight zones which suggests the existence of important information although there is only one real source on the sphere. Seeing the estimated point source in magenta cross shape close to the red real source, we could think that it is a true positive source. But even if they are on the same sphere and so close, with the angular threshold we fixed at 10° we see that it is not close enough. This estimated point source is hence a false positive. The artefact effect may be caused by an interference with two or more peaks, leading the algorithm to detect the echoes of the previous image source, since it is situated on a neighbour sphere. There are several hypothesis that could explain this phenomenon, and further theoretical investigation would be needed to understand it better. It is likely that the presence of many sources at small distances from this sphere can cause such effect. There exists also another point source in the following sphere as we will see in the next figure 14, which

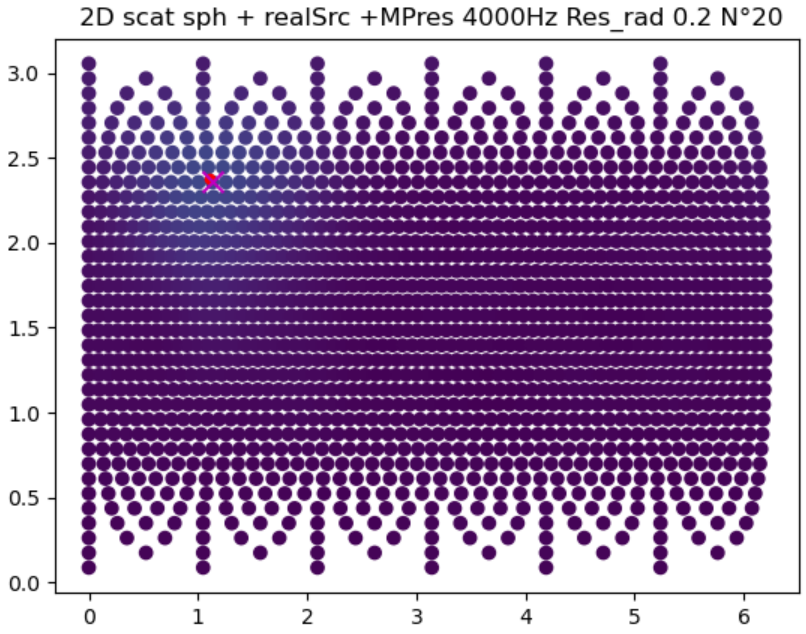


Figure 11: The colormap score of the 2nd interesting sphere at radius 4.8 m.

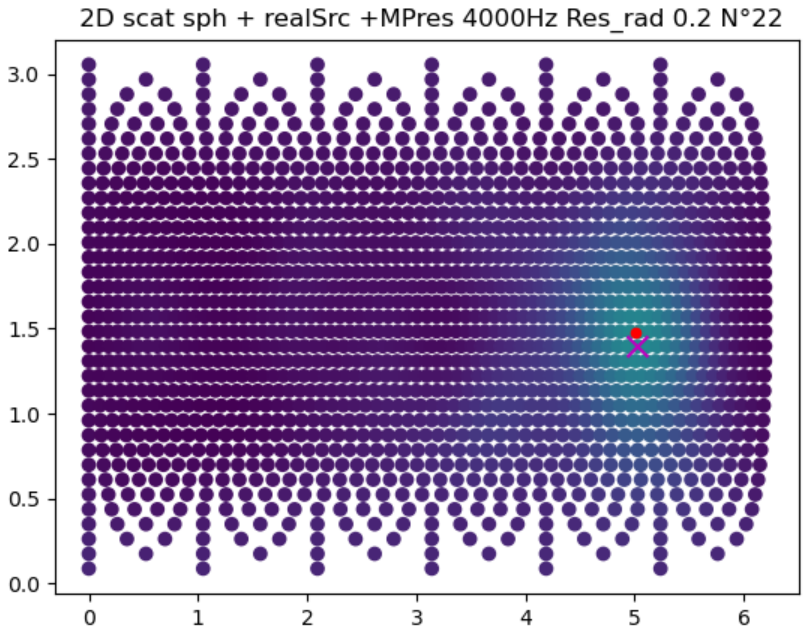


Figure 12: The colormap score of the 3rd interesting sphere at radius 5.2 m.

strengthens the hypothesis.

The sphere we mentioned is the 24th one, situated at the radius 5.6m. We can observe the power of the estimated source from the brightness of scores around it, and we see

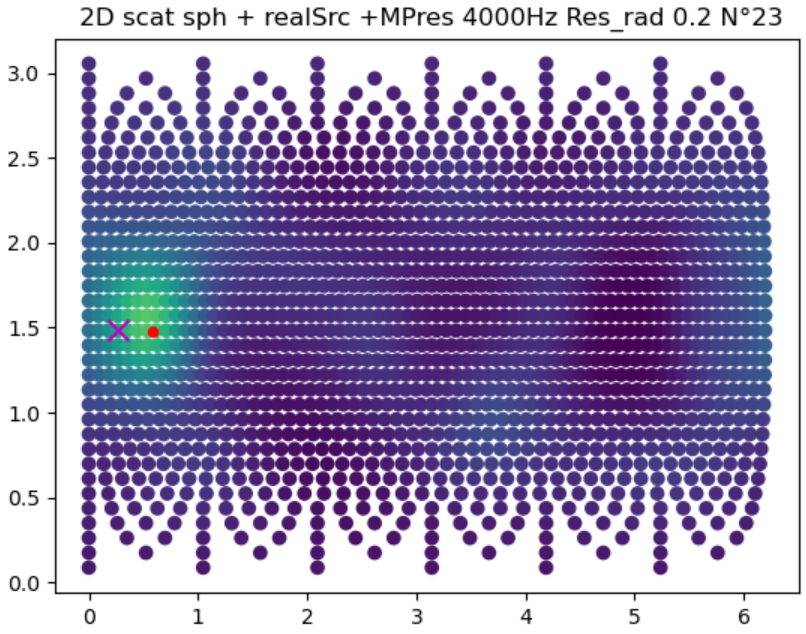


Figure 13: The colormap score of the 4th interesting sphere at radius 5.4 m.

clearly that it is matched to a real source, hence we located a 3rd wall. The colormap of this sphere is displayed in the figure 14 below:

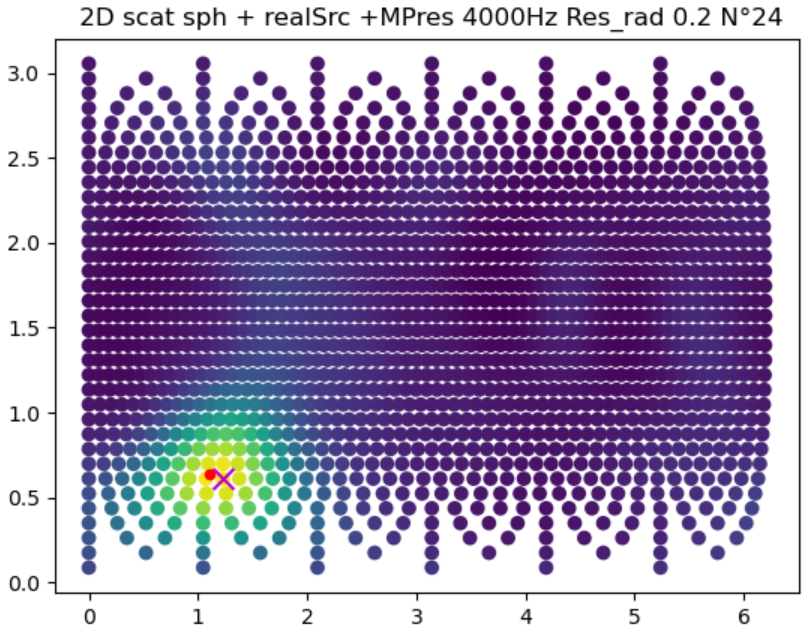


Figure 14: The colormap score of the 5th interesting sphere at radius 5.6 m.

In the following we will see a particular case of colormap score as showed in the next figure 15 beneath :

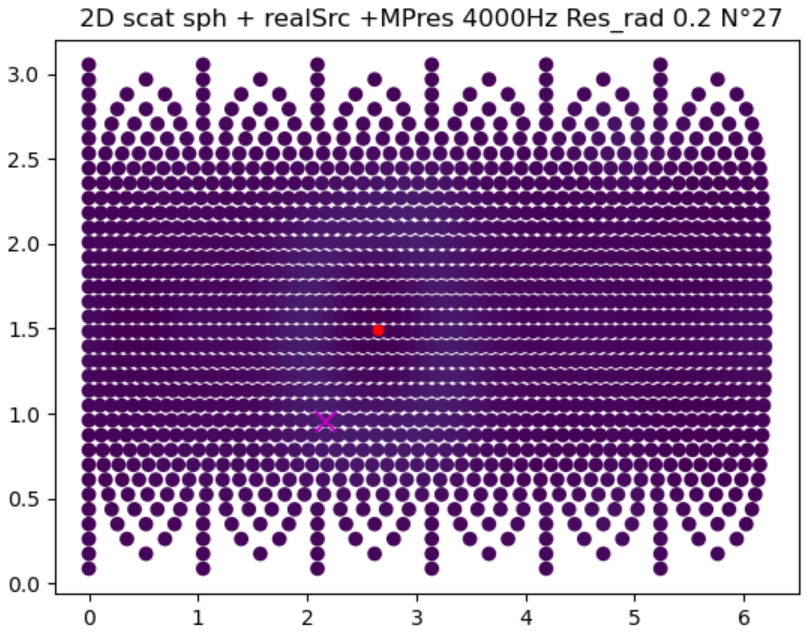


Figure 15: The colormap score of the 6th interesting sphere at radius 6.2 m.

In the figure above we see a ring of bright scores around the real red source on this sphere

and that the selected source using our method is situated on a point of that ring. Clearly the algorithm failed to the detect the real source since the angular distance is largely above the threshold. The ring phenomenon could be due to the source detected in the previous sphere, since its energy was clearly strong. The explication that may justify this case is that the echo of the previous source disturbed the signal of this one, causing the angular precision to fail.

The penultimate colormap is for the sphere number 28, situated at the radius 6.8 m and showed in the figure below:

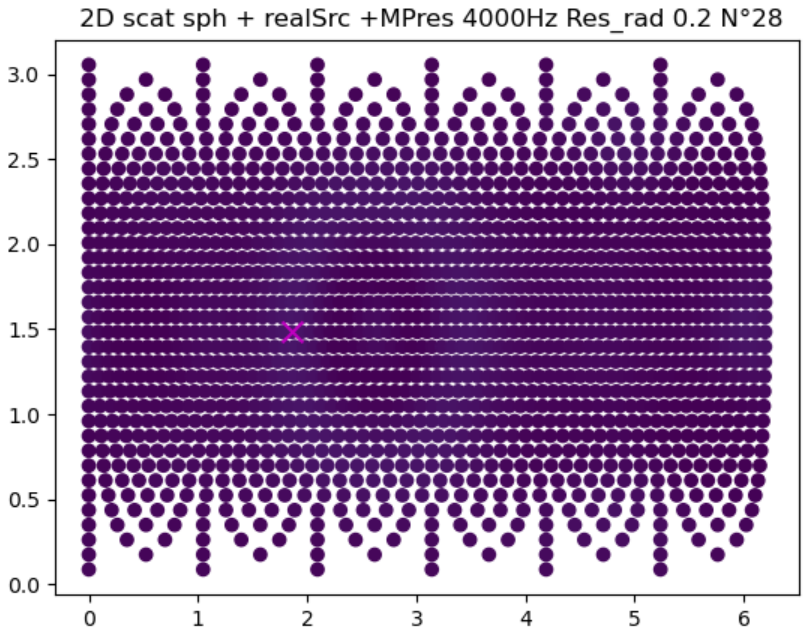


Figure 16: The colormap score of the 7th interesting sphere at radius 6.4 m.

In this sphere there is no real source, even though the algorithm selected a point source on this sphere. From the colormap score we can see again the ring form of the higher scores on this sphere. The only explanation of the selection problem is the higher signal intensity of the sources behind this sphere, as we mentioned before. Hence, this estimation is drastically missed out.

Finally the last sphere we display is the sphere number 42, localized at a radius of 9.2m. We see that there is a slightly bright area of score with the real red source situated in its center. We can not observe any existing estimated source in this sphere. It can be because of the dissipation of the sound energy while moving away from the sound source that the signal emitted by this image source was not strong enough for a peak to be detected. It is still just a hypothesis since we already saw that there are many parameters influencing the precision of our method. The figure corresponding to this sphere is just below:

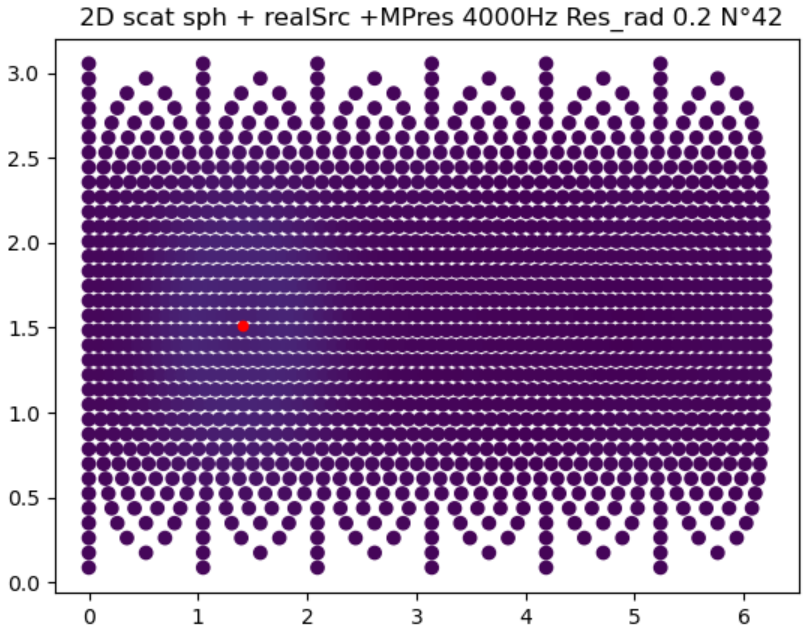


Figure 17: The colormap score of the 8th interesting sphere at radius 9.2 m.

# 4 Conclusion and perspectives

In this thesis we studied the problem of recovering the 3D geometry of a convex polyhedral room from room impulse response measurements with a microphone array, and we proposed a new method to solve it. It is an on-grid method based on the image source model introduced by Allen and Berkley [2]. It is about separating the peaks of the room response impulses that correspond to early sound reflections which are the analogue of the room's image sources. What is new about our method, is that we do not need any hand manipulation to select peaks, we adopted the Matching Pursuit algorithm to do it automatically. What characterizes this algorithm is that it is greedy and sparsen which helps solving combinatorial problems, such as in this case choosing 7 point sources from 115,659 points in the space.

For the experiments, we used a point-like single sound which is omni-directional and perfectly synchronized with the receiver. The receiver we used is a compact spherical microphone array of 32 microphones distributed on the rigid sphere of the microphone antenna and it is commercially available and named eigenmike. We fixed the higher radius that our grid can reach at 15m, that corresponds to the maximum distance where we can check the presence of the point sources. Then, thanks to the Pyroomacoustics simulator we generated 100 room to make a quantitative study of our method using default parameters and then we started varying them to see the influence of each one on the true positive rate of estimated points. We thus evaluated the precision of our method at fixed thresholds, they are the precision at which we want to select our points.

After running our tests with different values of the frequency of sampling $f_s$, the microphone array size Mic size, and both the radial and angular grid resolutions, we have seen that each parameter influence the precision of the method. Hence, based on the previous results, we can conclude that each grid have an optimal frequency of sampling that makes it more efficient. We can see also that microphone size is an important parameter since the method failed to find any point source for a diameter less than 42cm which is five time greater that the original diameter. We can also conclude that there is no need to take 10 times the diameter of the original antenna, i.e 84cm, because the difference in accuracy it gives is not too remarkable compared to the results obtained with the Mic size= $42cm$. The big influence we noticed also is when we varied the resolutions of the grid $\mathcal{G}$.

Since the sound propagates in a spherical way, it has a radial resolution and a spherical resolution and we must manipulate both radial steps and the angular steps to get an appropriate result. In fact, in the beginning of the experiments we noticed that having one step parameter in the case of Cartesian grid was not adapted to the problem we are trying to resolve and rapidly switched to a spherical grid to control both radial resolution and angular resolution. The previous results show that both resolutions have an important influence on the method, since the true positive rates decreased by lowering the resolution. The radial resolution parameter still have the biggest impact. With higher resolution, we succeeded in getting 83% of sources which is about the same as finding 5 walls out of 6 on average.

It seems easy to say that we can simply increase the resolution to get better results, but we must remember that we are limited in terms of memory and computing time because

increasing the resolution also means increasing the number of points on the grid that the algorithm have to scan and calculate. For the memory storage, this is binding.

Future work will need to check the robustness of the method to noise, because signals in real life contain noise and artefacts from measuring devices and the interaction with the surrounding. It also remains to use the RIRs with echoes of higher order. Indeed in the real situations the signal is a combination of echoes of all orders and there is not only the specular reflections, but also diffuse reflections depending on the type of surface absorption.

In the literature, there are some promising future directions to improve the results we get with our method, include improving matching pursuit using some of its extension such as Orthogonal Matching Pursuit. In this extension, we also select the $K+1$ atoms one-by-one but at each new atom selection, we recalculate the projection of $\mathbf{x}$ on the vector space generated by the selected vectors. Orthogonal Matching Pursuit can quickly converge to a zero error if $\mathbf{x}$ is a linear combination of $\mathbf{d}_k$, which is not the case for the basic Matching Pursuit.

As we saw that the grid resolution is a critical parameter, then a multiresolution approach could be a promising avenue. Multiresolution algorithms are an adaptive techniques based on a hierarchical decomposition of the signal that can help reducing dramatically computations and memory costs in the estimation. In the beginning of the calculations there is no need to start with a high resolution, the algorithm starts by selecting the likely area that contains the image source and increases the resolution only in that area and so on. The only risk is that this method introduces dependencies between data at different grid levels and it is then difficult to manage their locality.

Another promising avenue to avoid the dependency to the grid resolution is to use an off-grid method, in other words, working with super-resolution approaches, such as the Sliding-Frank Wolfe algorithm. It is also a sparse algorithm, also known as conditional gradient, that uses non-convex optimization steps as we have a non convex cost function. It has for instance been used for blind RIR estimation in the article of Di Carlo et al. [5].

# References


[1] Acoustics, M. (2013). em32 eigenmike® microphone array release notes. *MH Acoustics: Summit, NJ, USA*.

[2] Allen, J. B. and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 65(4):943–950.

[3] Crocco, M., Trucco, A., Murino, V., and Del Bue, A. (2014). Towards fully uncalibrated room reconstruction with sound. In *2014 22nd European Signal Processing Conference (EUSIPCO)*, pages 910–914. IEEE.

[4] Denoyelle, Q., Duval, V., Peyré, G., and Soubies, E. (2019). The sliding frank–wolfe algorithm and its application to super-resolution microscopy. *Inverse Problems*, 36(1):014001.

[5] Di Carlo, D., Elvira, C., Deleforge, A., Bertin, N., and Gribonval, R. (2020). Blaster: An off-grid method for blind and regularized acoustic echoes retrieval. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 156–160. IEEE.

[6] Dokmanić, I., Parhizkar, R., Walther, A., Lu, Y. M., and Vetterli, M. (2013). Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110(30):12186–12191.

[7] Kac, M. (1966). Can one hear the shape of a drum? *The american mathematical monthly*, 73(4P2):1–23.

[8] Lovedee-Turner, M. and Murphy, D. (2019). Three-dimensional reflector localisation and room geometry estimation using a spherical microphone array. *The Journal of the Acoustical Society of America*, 146(5):3339–3352.

[9] Mabande, E., Kowalczyk, K., Sun, H., and Kellermann, W. (2013). Room geometry inference based on spherical microphone array eigenbeam processing. *The Journal of the Acoustical Society of America*, 134(4):2773–2789.

[10] Mallat, S. G. and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, 41(12):3397–3415.

[11] Park, M. and Rafaely, B. (2005). Sound-field analysis by plane-wave decomposition using spherical microphone array. *The Journal of the Acoustical Society of America*, 118(5):3094–3103.

[12] Scheibler, R., Bezzam, E., and Dokmanić, I. (2018). Pyroomacoustics: A python package for audio room simulation and array processing algorithms. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 351–355. IEEE.

[13] Stan, G.-B., Embrechts, J.-J., and Archambeau, D. (2002). Comparison of different impulse response measurement techniques. *Journal of the Audio engineering society*, 50(4):249–262.

[14] Sun, H., Mabande, E., Kowalczyk, K., and Kellermann, W. (2012). Localization of distinct reflections in rooms using spherical microphone array eigenbeam processing. *The Journal of the Acoustical Society of America*, 131(4):2828–2840.