Notes for the meeting on February 19th, 2016.

The algorithm application basically consists of:

1. Obtaining the impulse responses of an array (from simulation or recordings).

2. Dividing the "stack" of IRs in windows of a given size, determined from the estimated "echo density" that depends on the room volume (but not on the absorption).

3. For each of those windows:
    1. Comprised of three regions:
        1. Direct sound
        2. Early reflections (single reflection per window)
        3. Diffuse reverberation (several reflections per window)
    2. After the "mixing time" (separation between early and diffuse) we consider only the reflection with most energy (so we'll get still only one reflection out of each window, although several may be contained).
    3. We use a method of our choice to estimate the impinging direction / distance to source (the paper considers plane waves but we should try to do it with spherical waves).

4. A map of image sources (with location and amplitude) is reconstructed from the information extracted from all windows

Limitations or improvements we can try to apply:

- The simple array used in their paper can probably be improved. We can study different array geometries, evaluating their performance, to find a "best array geometry" for the case of evaluating the room.

- Defining the time window. They use a constant value of 1ms, just because their work is based on the previous paper "Spatial Impulse Response Rendering". They did it for comparison purposes. We could further research the window size, not only to smaller values as they suggest, but also to a not constant values but depending on the moment in time. As it's obvious, the density of reflections increases as the time goes on. Therefore, we could study finding a method to define a rule for a time window that gets smaller as time advances.
- Solving the distance and impinging wave of each window.

    - They're considering plane waves and solving the problem using least-squares algorithm (minimization problem, solution given by the normal equations, page 8 from "Lecture notes in adaptive filters" by Jesper Kjær Nielsen). It has a good thing: the estimator is efficient (minimum variance achieved) but it has the drawback that it considers plane waves.

    - This will not be completely realistic on small spaces so we can consider spherical waves and use, for example, the Capon algorithm [http://kom.aau.dk/~jdomin14/capon/] but with a generalized Vandermonde vector to consider spherical waves [http://kom.aau.dk/~jdomin14/bartlett/]. Please, suggest other algorithms if you think this one is not suitable. Using this method carries one drawback, since it's iterative as opposed to what they use. On the other hand, we eliminate the approximation of plane waves. If we're trying to find information about the enclosure, it doesn't need to be a real-time system, so it shouldn't be a problem anyway.

Our extended application:

The paper's purpose is to show how the auralizations obtained by their method compare to the real scenarios. In our case we want to extract information about the room / enclosure. Hence, finding the map of image sources is only the first step for us. Once we achieve good results in estimating the map of image sources, we have to investigate methods to guess characteristics from the room from it.

Steps in the project I see:

A. Define and program a simulation system capable of producing image sources of a given room geometry.
B. Study and select the array geometry best suited for our specific purpose.
C. Study the time window size, perhaps as a function of time / other parameters.
D. Evaluate and select the method for estimating the impinging direction / TDOA (beamforming).
E. Find what characteristics of the enclosure can be found from knowing the map of image sources.

Step A, simulating the room:

I'm implementing the simulation programs to generate the synthetic impulse responses of a source and mic on a room. There are some differences with their approach:

- They used the mirrored sources method in their paper (page 26, second paragraph). They suggest using ray-tracing to better approximate the effect of diffuse reflections. We will make the simulations with ray tracing.
- They consider directly a 3D space when simulating the mirrored sources. As a first step, I've implemented the simulation working in a 2D plane. This means that we will have less reflections in the synthetic IR, but should be good enough to check other calculations and ideas. I will implement the 3D version if we consider it necessary.
- They consider an average absorption coefficient, while I have defined one for each segment of the boundaries. Their approach is probably good enough for auralization purposes, but since we want to estimate characteristics about the room, we probably need a more precise description of the reflections in our simulated IRs.

Parts that I'm still working on:

- Including the effect of air absorption.
- Defining the absorption coefficient as a set for narrow bands, instead of as it is now, a single broadband coefficient for each boundary surface. This is important for us, since we won't be able to say anything about the frequency characteristics of our final results otherwise.

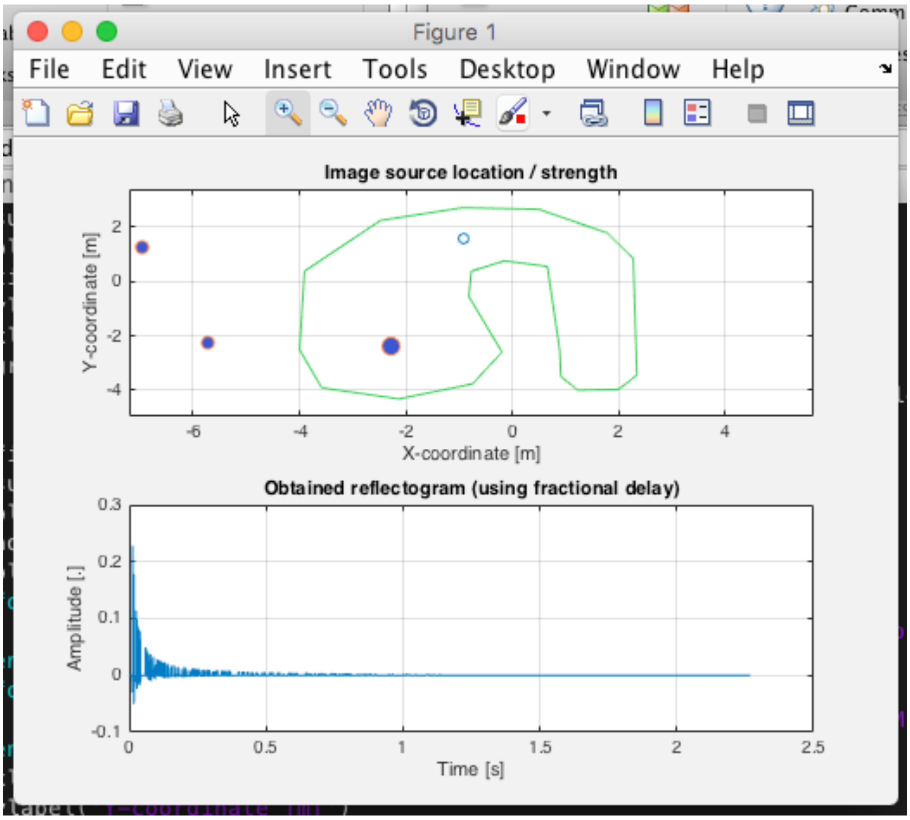
Conclusions extracted so far:

- The synthetic IR needs a well estimated tail to sound realistic. This means considering very high order reflections. Not a big deal, but it requires longer computational time.
- In order for the tail to be convincing, the air absorption is very important, as far reflections need to lose a lot of the "brightness" for us to perceive them realistic.
- The simulation parameters selection matters most. How to scan the space with rays is a whole another subject, and although we won't have time to deeply research ray-tracing methods (not the main topic of our thesis anyway), with a bit of trial and error (and using an "interactive

mode" of the simulation program) one can get a set of parameters that work well for a certain room geometry.

To-dos I can think of right now:

- We need to define what our objective room will be, so we can start to introduce its geometry and absorption characteristics in all simulations we are going to perform. They used a standardized shoe-box room, but I suggest we go for something a bit more interesting, as finding out the placement of the boundaries on a rectangular room doesn't seem "a big deal". For example, just to illustrate what I mean, I've been conducting simulations with (among others) a room that looks like:



- I suggest we pick up a room / place we can also perform measurements in, so we can validate the simulations at some point if we have time. *When we do so, we'll need to implement a new method for IR measurement, such as the sine sweep. This is probably not a priority now.*
- Study the different beamforming methods we can apply, discuss their capabilities for our application.
- Once we decide which beamformer(s) we want to use, evaluate the array geometries to achieve the best possible results.

The simulation programs I'm developing are hosted in GitHub.com: https://github.com/ignaciodsimon/room_ir_spatial_decomposition. So far I got:

- capture utility: to insert any *weird* room geometry, asking for the coefficients of each boundary and microphone / source positions. It generates a "simulation setup" with all necessary data to conduct a simulation.
- simulation program: uses the "simulation setup" and performs a ray-tracing scanning of the defined space. The output is redundant:
    - a list of the found image sources with location and amplitude

- a reflectogram from the previous information that can be used to generate IRs by including the responses of the loudspeaker and microphone.
- a listening utility, that processes an external audio with the results from the simulation to “hear the room”.