Benchmark SSL

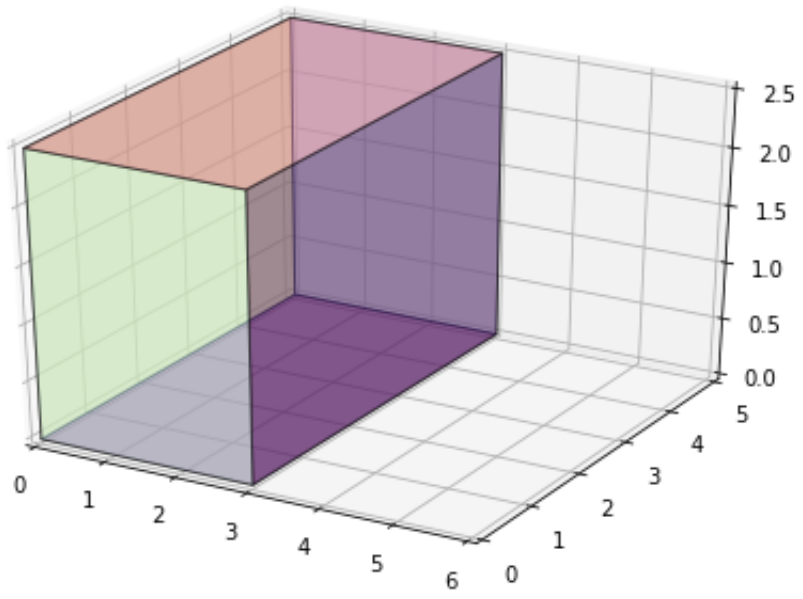
We design in what follows a benchmark to characterize and compare sound source localization algorithms. We use different setups to assess the performances of the algorithms in different situations.

I - Shoe Box Room

Shoe box room are rectangular parallelepiped. In this situation, source oclusion can't happen. Microphones are placed around a sphere of radius $10cm$ at the center of the room and the source is placed at the position $(1, 1, 1)$

Constants

- Room Dimensions : $3m * 5m * 2.5m$
- Max Order : 4
- Absorbtion factor : 0.2 (for all walls)
- Number of sources : 1



1. Setup n°1 : 5 microphones

Results

- *ML-TDOA* : on average the $L2$ Distance between ground truth position and recovered postion

is : $60cm$
- *SRP-PHAT* : on average the $L2$ Distance between ground truth position and recovered postion is: $2cm$

2. Setup n°2 : 50 microphones

Results

- *ML-TDOA* : on average the $L2$ Distance between ground truth position and recovered postion is: 40 cm
- *SRP-PHAT* : on average the $L2$ Distance between ground truth position and recovered postion is: 2 cm

3. Conlusion

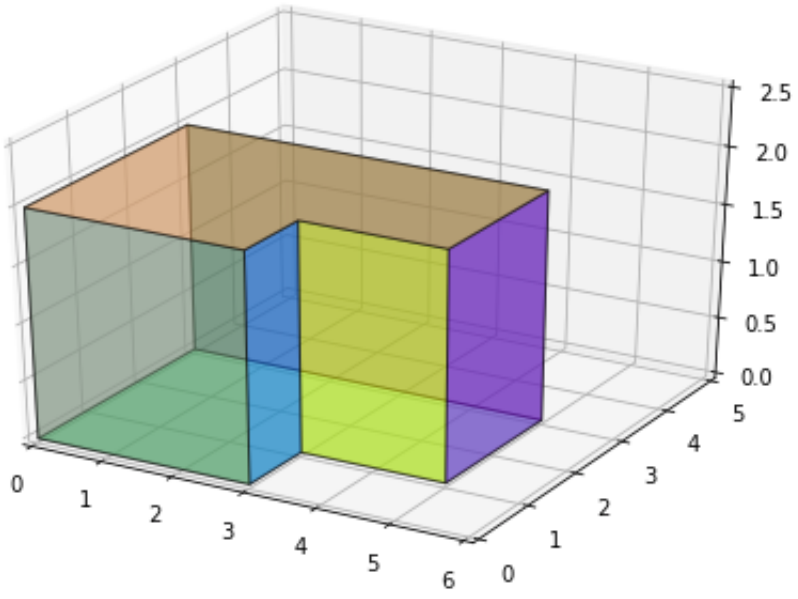
From these 2 setups we can observe that by increasing the number of microphones the accuracy of the *ML-TDOA* algorithm has increased. However, *SRP-PHAT* always performs better than *ML-TDOA*, this difference of performance can be explained by the fact that the beamforming-based approach (*SRP-PHAT*) is robust in adverse acoustic environments.

II - L-Shaped Room

In this context source oclusion could happen, and we will model it to see how both algorithm perform. Again, microphones are placed around a sphere of radius $10cm$ but this time at position $(4, 2, 1)$

Constants

- Surface dimensions : $(3m, 3m, 5m, 2m, 2m, 1m)$
- Elevation : $2m$
- Max Order = 4
- Absorbtion factor = 0.2 (for all walls)
- Number of Sources: 1
- Number of Microphones: 5



1. Setup n°1 : no occlusion

The source is placed at : $(1, 2, 1)$

Results

- *ML-TDOA* : on average the $L2$ Distance between ground truth position and recovered postion is : $121cm$
- *SRP-PHAT* : on average the $L2$ Distance between ground truth position and recovered postion is : $45cm$

2. Setup n°2 : occlusion

The source is placed at : $(2.5, 0.5, 1)$

Results

- *ML-TDOA* : on average the $L2$ Distance between ground truth position and recovered postion is : $260cm$
- *SRP-PHAT* : on average the $L2$ Distance between ground truth position and recovered postion is : $57cm$

3. Conlusion

Occlusion decreases the performance of both algorithms. SRP-PHAT's performance has decreased the most but it is still more robust than ML-TDOA.