SPARSE RECOVERY BEAMFORMING AND UPSCALING IN THE RAY SPACE

Shiduo Yu[(1)], Craig Jin [(1)], Fabio Antonacci [(2)], Augusto Sarti [(2)]

[(1)] University of Sydney, Australia [(2)] Politecnico di Milano, Italy

1. Introduction

Motivation: This work continues the application of Sparse Recovery to enhance Ray Space Analysis using a single linear array. We previously proved Sparse Recovery can increase the SNR of Ray Space Transform. However, the image resolution is restricted by the number of microphones. In this work, we use Upscaling to achieve the higher resolution of the Ray Space image.

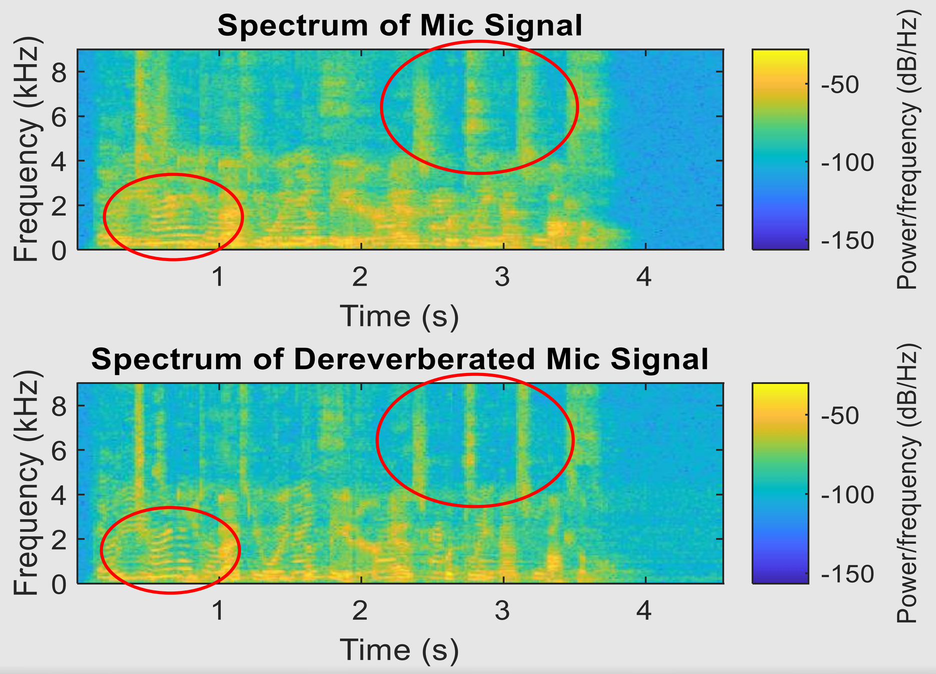
Aim: Develop an algorithm which takes advantage of both the Sparse Recovery and the Ray Space to improve source localization and separation.

2. Algorithm (5 Steps)

Step 1 : Dereverberation

- Two dereverberation methods are applied serially to make the sound field more sparse:

a) Linear Prediction Dereverberation
b) Direct-ambient signal separation

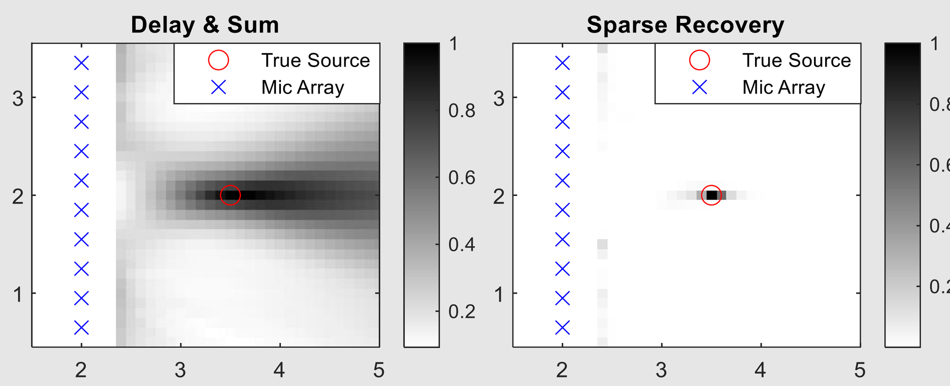


Step 2 : Sparse Recovery & Upscaling

- Upscaling applies Sparse Recovery sound field analysis to synthetize a denser array of microphone signals.

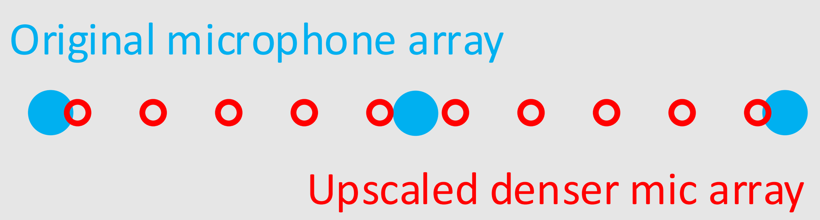
The Sparse Recovery technique uses the sparsity of the sound field to estimate the sound field, $\mathbf{S}(f)$,

$$\hat{\mathbf{S}}(f) = \underset{\mathbf{S}(f)}{\operatorname{argmin}}\{\|\mathbf{S}(f)\|_{0,2}: \mathbf{X}(f) = \mathbf{A}(f)\mathbf{S}(f)\}.$$



By multiplying an upscaled array manifold matrix, $\mathbf{A}_u(f)$, we can now synthetize the upscaled array signals, $\mathbf{X}_u(f)$, as

$$\mathbf{X}_u(f) = \mathbf{A}_u(f)\hat{\mathbf{S}}(f) = \mathbf{A}_u(f)\mathbf{D}(f)\mathbf{X}(f).$$



Step 3 : Ray Space Analysis Integrated with Sparse Recovery beamforming

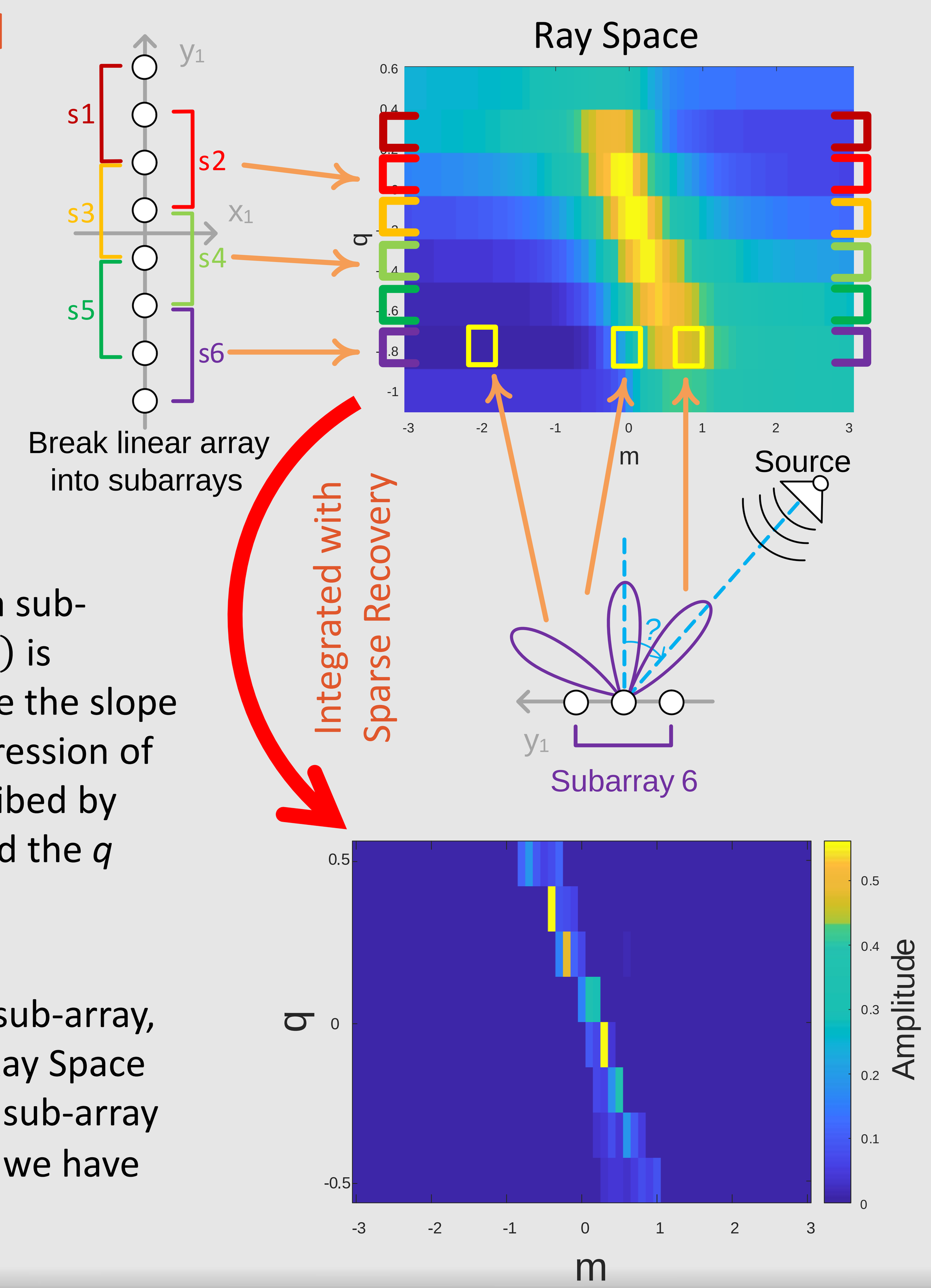
- Use Sparse Recovery to raise the SNR of Ray Space Transform.

The Ray Space technique breaks a long linear array into short subarrays, with each subarray providing a slightly different viewpoint. The Ray Space image explains how much energy each sub-array can see from different viewpoints. A Ray Space image is shown on the right figure, where each row represents the steering angle energy of a sub-array. It can be seen that a sound source at (x_0, y_0) is mapped into a straight line: $y_0 = mx_0 + q$, where the slope $m = \tan\theta$, q is the intersection. A simplified expression of the Ray Space Transform, $\mathbf{R}(m,q,f)$, can be described by Delay and Sum Beamformer weights, $\mathbf{W}(m,f)$, and the q sub-array signals, $\mathbf{X}_{\text{sub}}(q,f)$, as

$$\mathbf{R}(m,q,f) = \mathbf{W}(m,f)\mathbf{X}_{\text{sub}}(q,f).$$

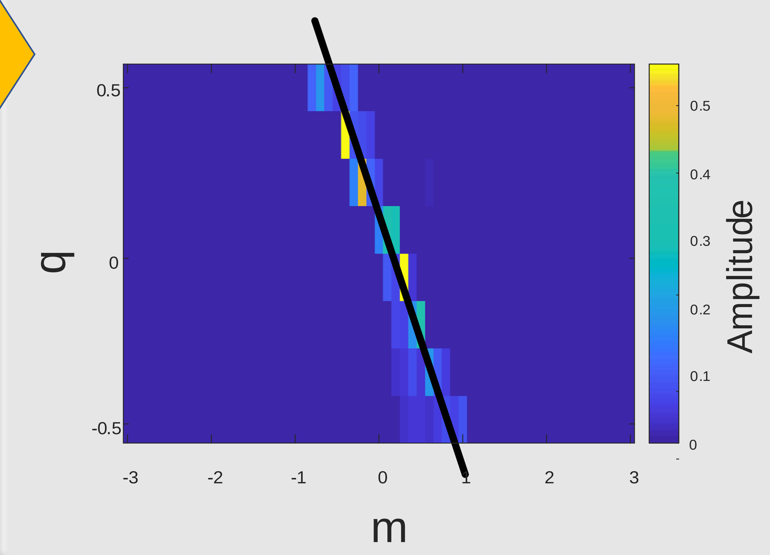
To take advantage of sound field sparsity for each sub-array, we integrate Sparse Recovery Beamforming with Ray Space Transform. We again use Sparse Recovery on each sub-array to estimate the source signals, $\hat{\mathbf{S}}_{\text{sub}}(m,q,f)$, then, we have

$$\mathbf{R}(m,q,f) = \hat{\mathbf{S}}_{\text{sub}}(m,q,f).$$



Step 4 : Sub-array Based Source Localization

Weighted linear regression is used to fit a line to the putative source lines in the ray space image to determine the source locations. The weights employed in the regression analysis are proportional to the image pixel intensity.



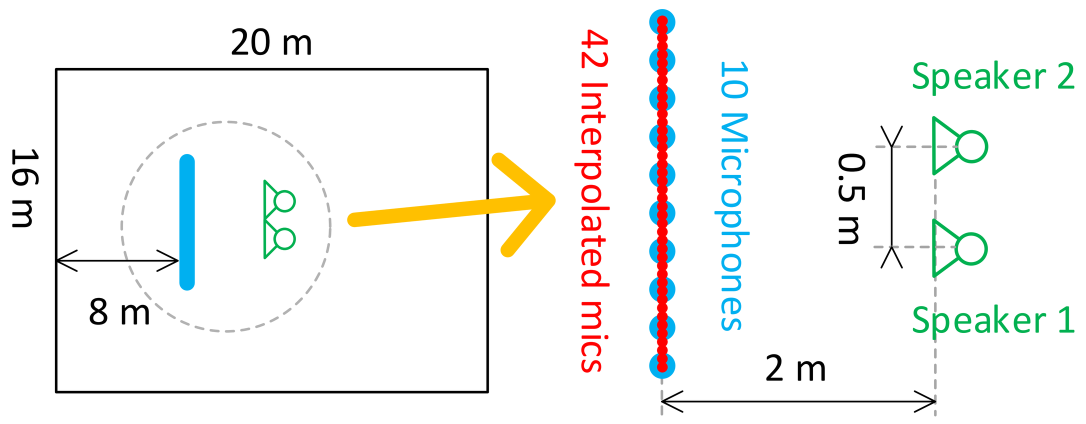
Step 5 : Sub-array Based Source Separation

After determining the source locations, $\boldsymbol{z}_n$, the n-th individual source signals, $\hat{\mathbf{S}}_{source}(n,f)$, are then estimated as a weighted summation of the sub-array signals, $\hat{\mathbf{S}}_{\text{sub}}(\boldsymbol{z}_n,q,f)$:

$$\hat{\mathbf{S}}_{source}(n,f) = \sum_q w_q \hat{\mathbf{S}}_{\text{sub}}(\boldsymbol{z}_n,q,f)\,,$$

where $w_q = \frac{\|\hat{\mathbf{S}}_{\text{sub}}(\boldsymbol{z}_n,q,f)\|_2^2}{\sum_q \|\hat{\mathbf{S}}_{\text{sub}}(\boldsymbol{z}_n,q,f)\|_2^2}$.

3. Simulation Configuration



- MCRoomSim is used to simulate a 20 m x 16 m x 5 m room.
- There is one linear microphone array 1.8 m in length with 10 microphones. (20 cm inter-sensor spacing with an aliasing frequency of 3430 Hz)
- The microphone array is 8 m away from the west wall at a 1.6 m height.
- Two speech sources with a 0.5 m separation are positioned 2 m away from the center of the array.
- Sources are in the near-field of the array because the ray space assumes multiple viewpoints.
- The critical distance is 4.2 m. The reverberation time T60 is 0.74 s.
- Simulations were run for 10 pairs of speaker signals in the Archimedes dataset.

4. Numerical Experiment

The average localization error for the original 10 mic array :

- Delay and Sum – 22.0 cm
- MPDR – 17.9 cm
- Sparse Recovery –11.5 cm

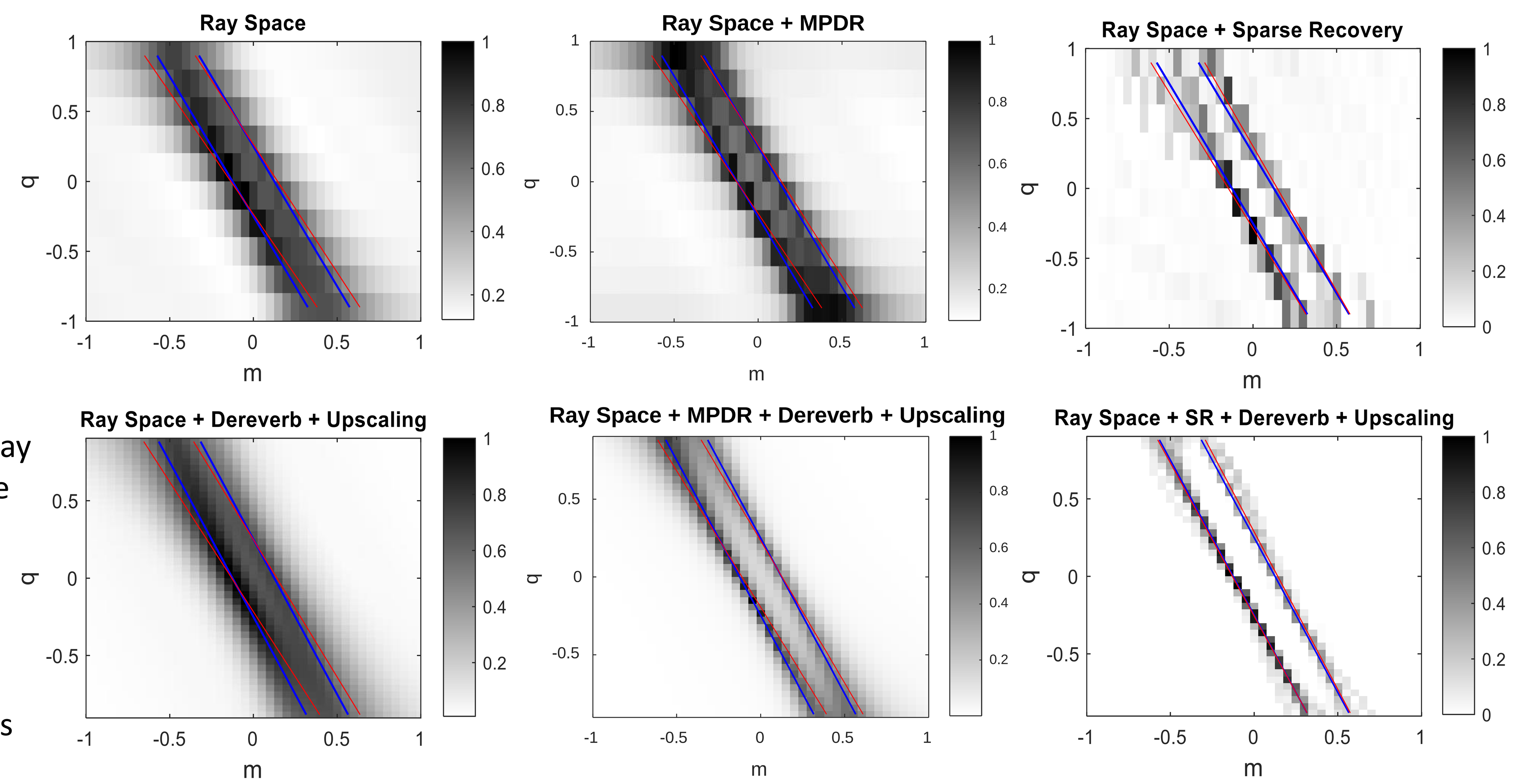
The average localization error for the upscaled array with 42 microphones:

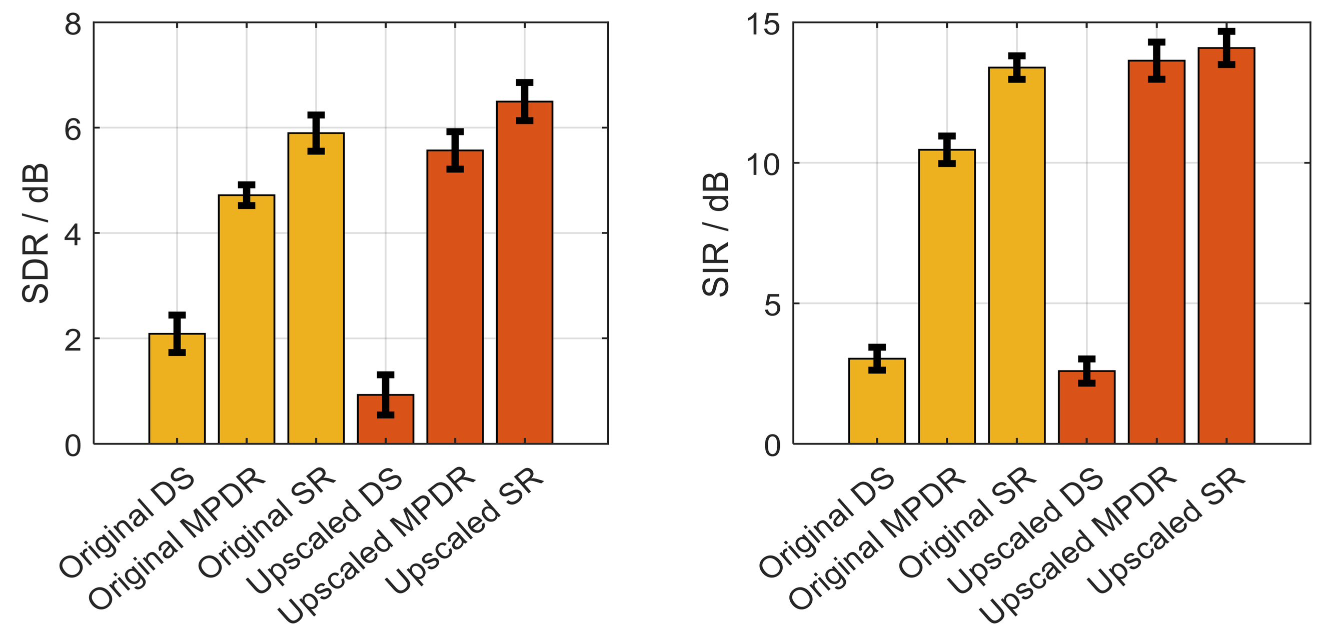
- Delay and Sum – 28.5 cm
- MPDR – 21.3 cm
- Sparse Recovery – 9.0 cm.

The ability of the upscaling algorithm and the weighted sub-array beamforming to separate the two sources was evaluated by the signal-to-interference ratio and the signal-to-distortion ratio.

The upscaling algorithm resulted in improved performance for both the weighted MPDR and the weighted SR beamforming, and worse performance for the weighted DS beamforming.

The performance with the MPDR and SR beamforming methods are significantly better than the DS beamforming method.





5. Conclusion

In this work, we consider sparse recovery beamforming and upscaling in the ray space. The context for this work is a preliminary study and development of a source localization and source separation algorithm for multiple and distributed arrays.

We have especially chosen the ray space paradigm because of its unified geometric framework for multiple arrays and viewpoints. By integrating the sparse recovery and ray space methods, we propose a source localization and separation algorithm that easily extends to multiple and distributed arrays, is adaptive (sparse recovery is an adaptive algorithm), accommodates both near-field and far-field analysis, and enables image processing of the ray space to obtain geometrical insights. Significantly, the upscaling seems to yield genuine improvements. This is likely an important consideration in the design of distributed arrays to cover large venues and areas. Future work will examine multiple arrays and empirical measurements.