Joint source localization and dereverberation by sound field interpolation using sparse regularization

Niccolò Antonello[1,2], Enzo De Sena[3], Marc Moonen[1], Patrick A. Naylor[4], Toon van Waterschoot[1,2]

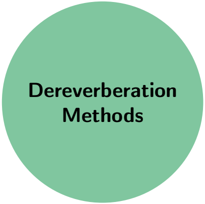
[1]ESAT–STADIUS, KU Leuven, BE
[2]e-Media Lab, KU Leuven, BE
[3]IoSR University of Surrey Guilford, UK
[4]Electronic Engineering, Imperial London College, UK

April 19, 2018, Calgary, Canada

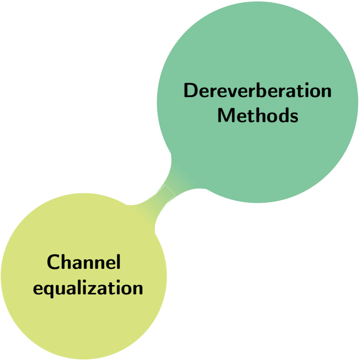
Outline

Outline

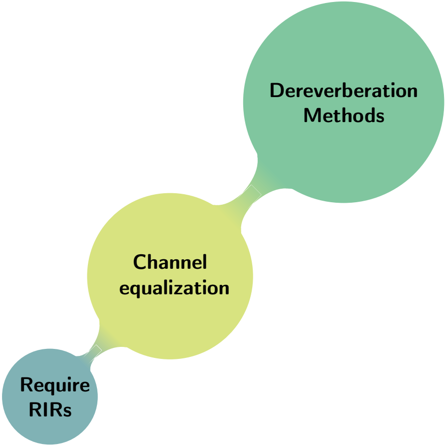
Reverberant environments reduce speech intelligibility



Reverberant environments reduce speech intelligibility

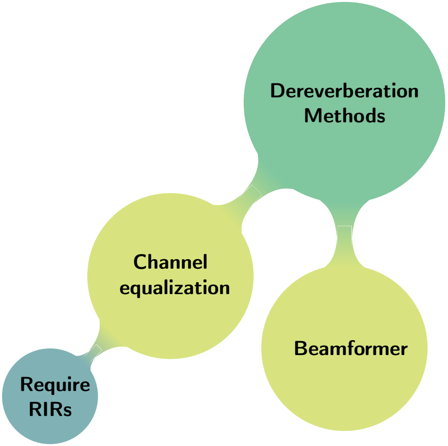


Reverberant environments reduce speech intelligibility

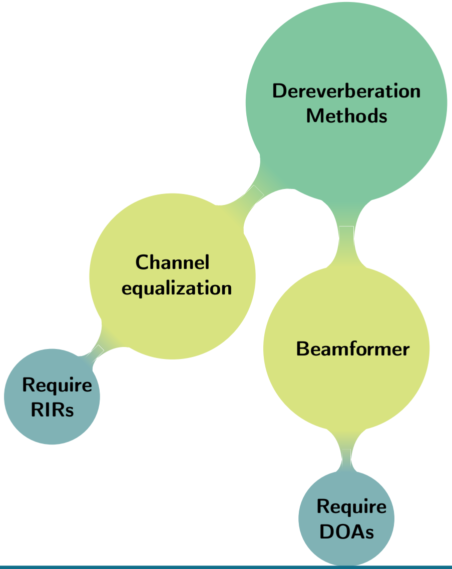


KU LEUVEN

Reverberant environments reduce speech intelligibility

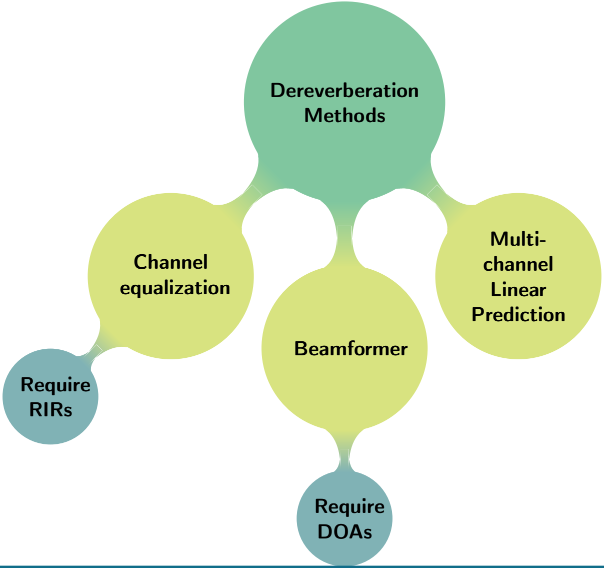


Reverberant environments reduce speech intelligibility



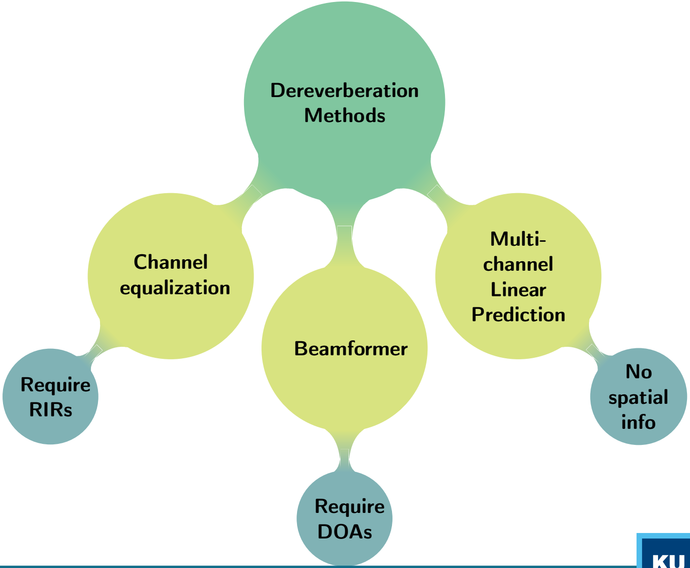
KU LEUVEN

Reverberant environments reduce speech intelligibility

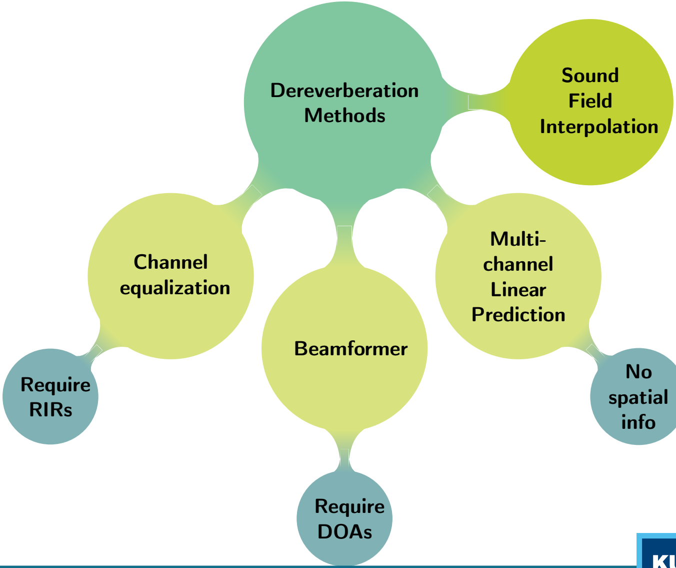


KU LEUVEN

Reverberant environments reduce speech intelligibility



Reverberant environments reduce speech intelligibility



KU LEUVEN

Novel method

- Localization and dereverberation formulated jointly

Novel method

- Localization and dereverberation formulated jointly
- Sound field recorded by mic array is **interpolated**

Novel method

- Localization and dereverberation formulated jointly
- Sound field recorded by mic array is **interpolated**
- Inverse problem:
  - Rely on **acoustic models**

Novel method

- Localization and dereverberation formulated jointly
- Sound field recorded by mic array is **interpolated**
- Inverse problem:
  - Rely on **acoustic models**
    - Plane Wave Decomposition Model (PWDM)
    - Time-domain Equivalent Source Model (TESM)

Novel method

- Localization and dereverberation formulated jointly
- Sound field recorded by mic array is **interpolated**
- Inverse problem:
  - Rely on **acoustic models**
    - Plane Wave Decomposition Model (PWDM)
    - Time-domain Equivalent Source Model (TESM)
  - Sparse prior
    - spatial sparsity
    - spatio-temporal sparsity
    - spatio-spectral sparsity

Novel method

- Localization and dereverberation formulated jointly
- Sound field recorded by mic array is **interpolated**
- Inverse problem:
  - Rely on **acoustic models**
    - Plane Wave Decomposition Model (PWDM)
    - Time-domain Equivalent Source Model (TESM)
  - Sparse prior
    - spatial sparsity
    - spatio-temporal sparsity
    - spatio-spectral sparsity
- Large-scale nonsmooth optimization problem:
  - accelerated variant of Proximal Gradient (PG) algorithm
  - Weighted Overlap-Add (WOLA) procedure

Outline

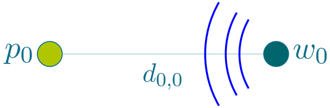
KU LEUVEN

Time-domain equivalent source model (TESM)

Time-domain spherical wave definition (Green's function):

$$p(t,\mathbf{x})|_{\mathbf{x}=\mathbf{x}_m} = \frac{1}{4\pi d_{l,m}}\delta\left(t - \tfrac{d_{l,m}}{c}\right) * w(t),$$

where $w(t)$ is a *weight signal* driving the *equivalent source*.



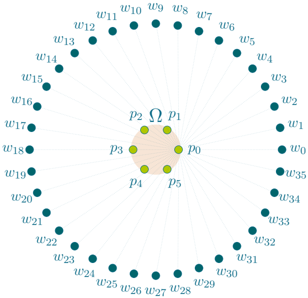
KU LEUVEN

Time-domain equivalent source model (TESM)

Using many equivalent sources and discretizing over time:

$$p(\mathbf{x},t)|_{\mathbf{x}=\mathbf{x}_m} = \sum_{l=0}^{N_w-1} \frac{1}{4\pi d_{l,m}} \delta\left(t - \frac{d_{l,m}}{c}\right) * w_l(t),\ \mathbf{x}_m \in \Omega,$$

any sound pressure inside source-free volume Ω can be modeled.



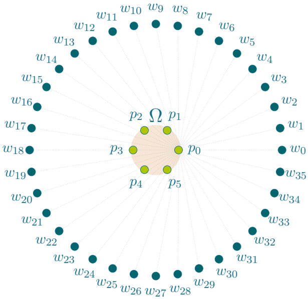
KU LEUVEN

Time-domain equivalent source model (TESM)

Generalizing for many mic positions:

$$\mathbf{P} = \mathsf{D}_t \mathbf{W}$$

where $\mathsf{D}_t : \mathbb{R}^{N_t \times N_w} \to \mathbb{R}^{N_t \times N_m}$ *dictionary* of *spherical waves*.

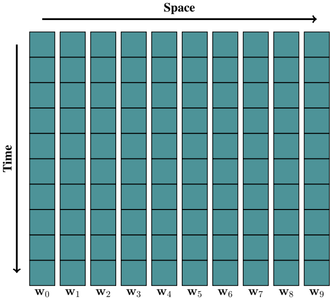


Time-domain equivalent source model (TESM)

Generalizing for many mic positions:

$$\mathbf{P} = \mathsf{D}_t \mathbf{W}$$

where $\mathsf{D}_t : \mathbb{R}^{N_t \times N_w} \rightarrow \mathbb{R}^{N_t \times N_m}$ *dictionary* of *spherical waves*.

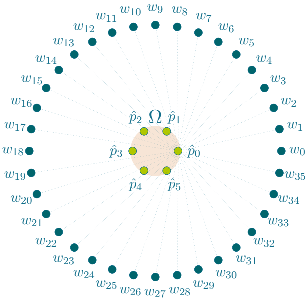


Plane wave decomposition model (PWDM)

Same using many plane waves (frequency domain):

$$\hat{p}(\mathbf{x}, f)|_{\mathbf{x}=\mathbf{x}_m} \approx \sum_{l=0}^{N_w-1} e^{-ik_f d_{l,m}} \hat{w}_l(f) \text{ for } \mathbf{x}_m \in \Omega,$$

any sound pressure inside source-free volume Ω can be modeled.

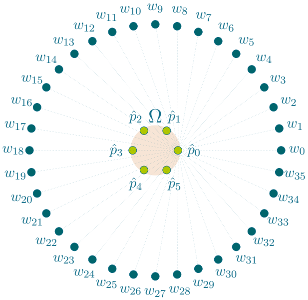


Plane wave decomposition model (PWDM)

Generalizing for many mic positions:

$$\hat{\mathbf{P}} = \mathsf{D}_p \hat{\mathbf{W}}$$

where $\mathsf{D}_p : \mathbb{C}^{N_f \times N_w} \to \mathbb{C}^{N_f \times N_m}$ *dictionary* of *plane waves*.



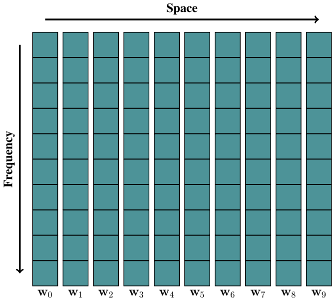
KU LEUVEN

Plane wave decomposition model (PWDM)

Generalizing for many mic positions:

$$\hat{\mathbf{P}} = \mathsf{D}_p \hat{\mathbf{W}}$$

where $\mathsf{D}_p : \mathbb{C}^{N_f \times N_w} \rightarrow \mathbb{C}^{N_f \times N_m}$ *dictionary* of *plane waves*.

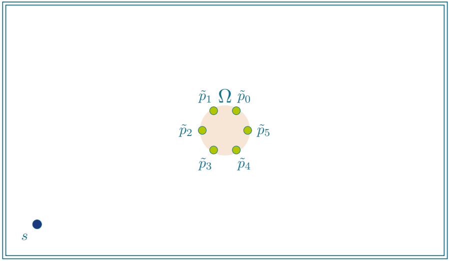


Outline

KU LEUVEN

The inverse problem

Sound field created by *far field* source is recorded by an array of N_m mics.



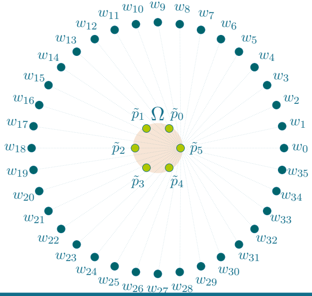
KU LEUVEN

The inverse problem

Optimal weight signals $\mathbf{W}^\star$ found by solving:

$$\mathbf{W}^\star = \underset{\mathbf{W}}{\operatorname{argmin}} \quad f(\mathbf{W}) = \frac{1}{2}\|\mathrm{D}\mathbf{W} - \tilde{\mathbf{P}}\|_F^2$$

heavily *ill-posed problem* $\rightarrow$ **over-fitting**



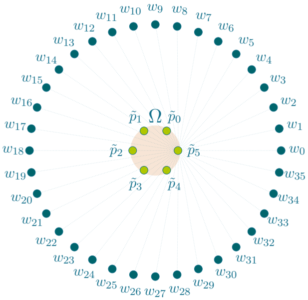
KU LEUVEN

The inverse problem

Optimal weight signals $\mathbf{W}^\star$ found by solving:

$$\mathbf{W}^\star = \underset{\mathbf{W}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathrm{D}\mathbf{W} - \tilde{\mathbf{P}}\|_F^2 + \lambda g(\mathbf{W})$$

regularize using a sparsity inducing *regularization term* g.



KU LEUVEN

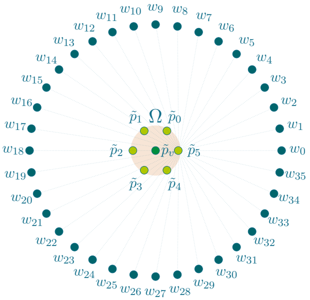
The inverse problem

Optimal weight signals $\mathbf{W}^\star$ found by solving:

$$\mathbf{W}^\star = \underset{\mathbf{W}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathrm{D}\mathbf{W} - \tilde{\mathbf{P}}\|_F^2 + \lambda g(\mathbf{W})$$

regularize using a sparsity inducing *regularization term* g.

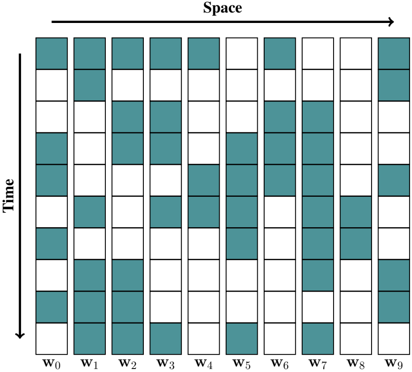
Use additional mic $\tilde{\mathbf{p}}_v$ for tuning regularization parameter λ.



KU LEUVEN

Spatio-temporal sparsity

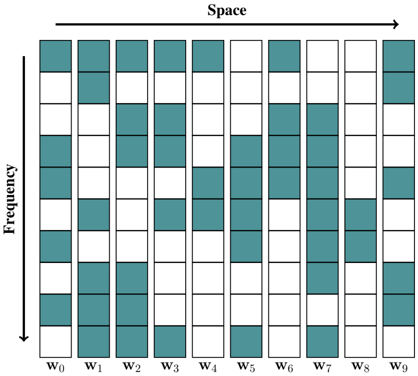
Model: TESM
Regularization: $g(\mathbf{W}) = \|\mathbf{W}\|_1$



Acoustic model and Sparsity combinations

Spatio-spectral sparsity

Model: PWDM
Regularization: $g(\mathbf{W}) = \|\mathbf{W}\|_1$

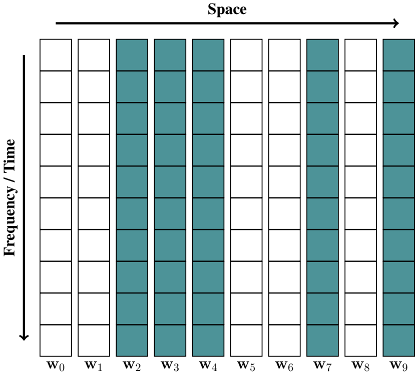


KU LEUVEN

Acoustic model and Sparsity combinations

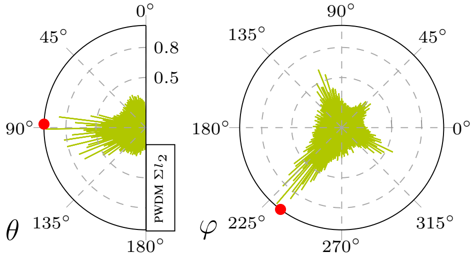
Spatial sparsity

Model: both PWDM and TESM
Regularization: $g(\mathbf{W}) = \sum_{l=0}^{N_w-1} \|\mathbf{w}_l\|_2$



Joint Dereverberation & Localization

After solving inverse problem obtain $\mathbf{W}^\star$

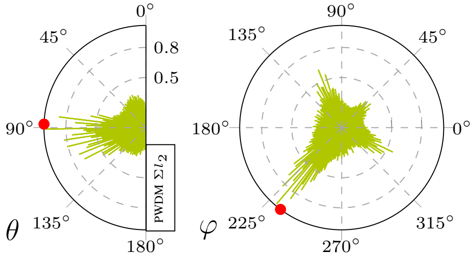


Plot energy of columns of $\mathbf{w}_l^\star$ as a function of azimuthal (φ) and polar (θ) angles.

KU LEUVEN

Joint Dereverberation & Localization

After solving inverse problem obtain $\mathbf{W}^\star$

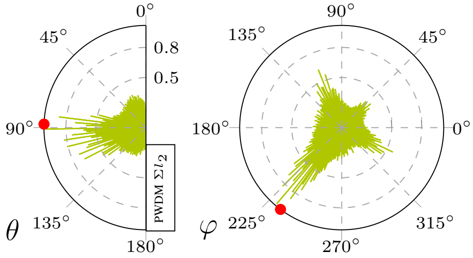


Localization

Find d-th column of $\mathbf{W}^\star$ with strongest energy.
This corresponds to equivalent source with specific DOA.

Joint Dereverberation & Localization

After solving inverse problem obtain $\mathbf{W}^{\star}$



Dereverberation

Dereverberated signal readily available: $\mathbf{w}_d^{\star}$
Reverberation is spatially distributed among equivalent sources

Outline

KU LEUVEN

Optimization algorithm

Inverse problem: nonsmooth cost function

- f is smooth
- g nonsmooth

KU LEUVEN

Optimization algorithm

Proximal Gradient (PG) algorithm

- can deal with nonsmooth terms
- cheap iterations
- suitable for large scale problems
- can be accelerated using quasi Newton methods†

† N. Antonello, L. Stella, P. Patrinos and T. van Waterschoot, "Proximal gradient algorithms: applications in signal processing", arXiv:1803.01621, 2018.

`github.com/kul-forbes/StructuredOptimization.jl`

Optimization algorithm

Proximal Gradient (PG) algorithm

- can deal with nonsmooth terms
- cheap iterations
- suitable for large scale problems
- can be accelerated using quasi Newton methods[†]

Optimization problem may still be intractable:

- if $N_w = 500$ $N_t = 16000$ (2 s $F_s = 8$ kHz) $\rightarrow 8 \cdot 10^6$ optimization variables

KU LEUVEN

Optimization algorithm

Proximal Gradient (PG) algorithm

- can deal with nonsmooth terms
- cheap iterations
- suitable for large scale problems
- can be accelerated using quasi Newton methods†

Optimization problem may still be intractable:

- if $N_w = 500$ $N_t = 16000$ (2 s $F_s = 8$ kHz) $\rightarrow 8 \cdot 10^6$ optimization variables

Weighted Overlap-Add procedure:

- split $\tilde{\mathbf{P}}$ into overlapping frames of $N_\tau = 512 \rightarrow 256 \cdot 10^3$ optimization variables *sub-problem*

KU LEUVEN

Outline

KU LEUVEN

Set-up

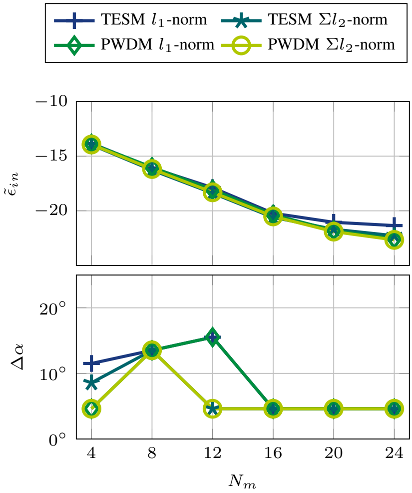
- RIRs simulated with Randomized Image Method[1]
  - $T_{60} = 1$ s
  - $[L_x, L_y, L_z] = [7.34, 8.09, 2.87]$ m
- Mic signals created with male speech signal (sensor noise SNR $= 40$dB)
- Spherical microphone array with N_m
- $N_w = 500$ equivalent sources on a Fibonacci Lattice radius 2.9 m

[1] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the Modeling of Rectangular Geometries in Room Acoustic Simulations", IEEE Transactions of Audio, Speech Language Processing, vol. 21, no. 4, 2015.

github.com/nantonel/RIM.jl

KU LEUVEN

Interpolation error & DOA estimate

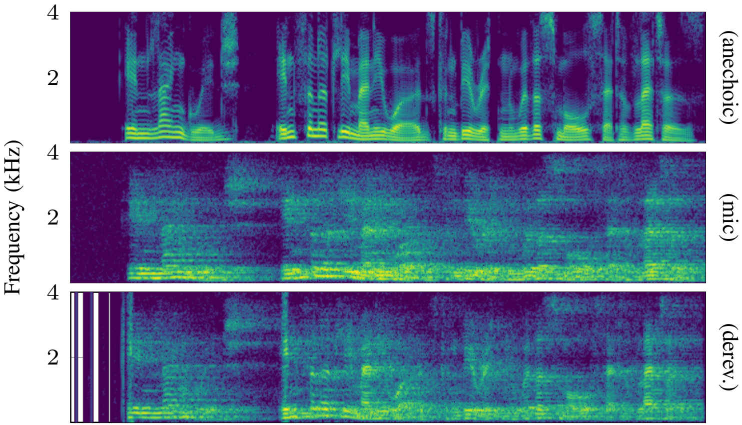


- Average interpolation error $\bar{\epsilon}_{in}$
  - Similar performances
  - Spatio-temporal sparsity slightly worse
- DOA estimate
  - good performance even with few mics
  - minimum angular distance of $4.5°$ due to finite number of directions

$\rightarrow$ more mics, better results

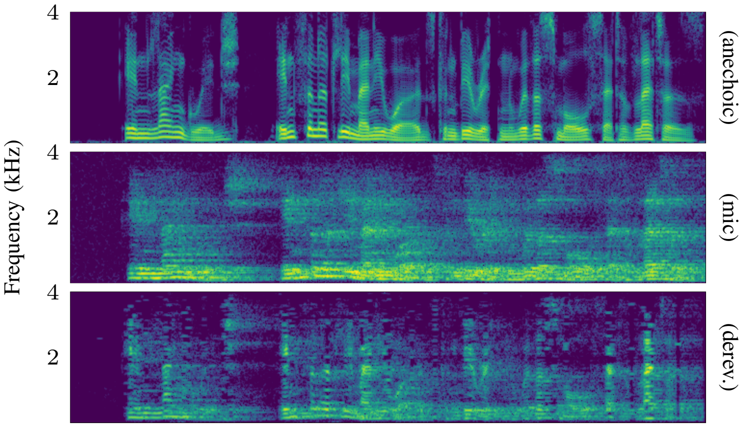
KU LEUVEN

Dereverberation - Spatio-temporal sparsity $N_m = 16$



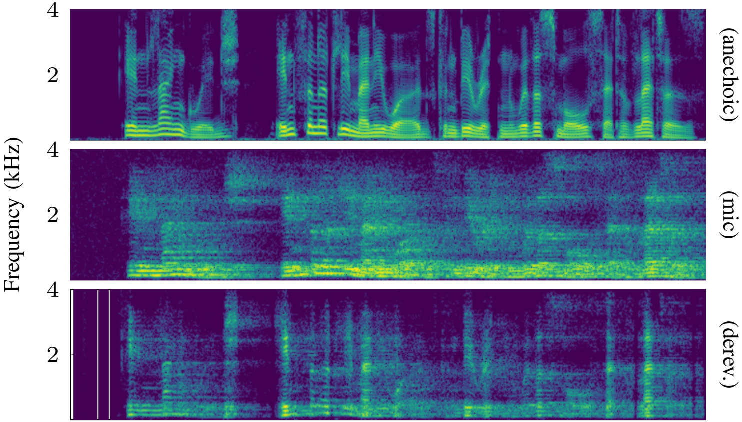
KU LEUVEN

Dereverberation - Spatio-spectral sparsity $N_m = 16$



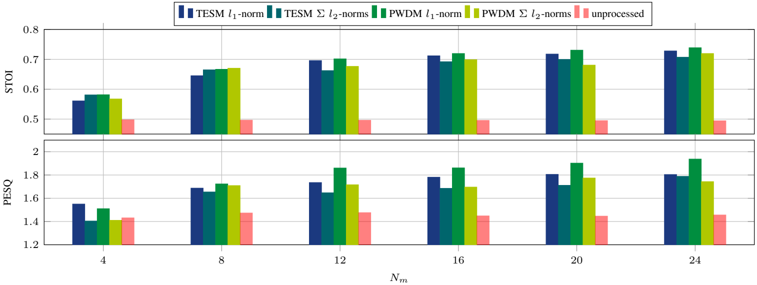
KU LEUVEN

Dereverberation - Spatial sparsity sparsity $N_m = 16$



KU LEUVEN

Objective measures



- Speech intelligibility (STOI) & Speech Quality (PESQ) scores
- In line with interpolation error and DOA estimates
- Better performance with spatio-spectral sparsity
- Informal listening test: worse results are with spatio-temporal sparsity
- Sound samples

KU LEUVEN

Outline

KU LEUVEN

Conclusions

- Novel method for joint source localization and dereverberation
  - Inverse problem rely on acoustic model and sparse prior
    - Spatio-temporal sparsity (TESM + l_1-norm)
    - Spatio-spectral sparsity (PWDM + l_1-norm)
    - Spatial sparsity (Σl_2-norm)
  - Solve with accelerated PG algorithm & WOLA procedure
- Simulation results
  - good DOA estimates with few mics
  - Better dereverberation as more mics are used