Focusing and frequency smoothing for arbitrary arrays with application to speaker localization

Hanan Beit-On, Vladimir Tourbabin, Boaz Rafaely

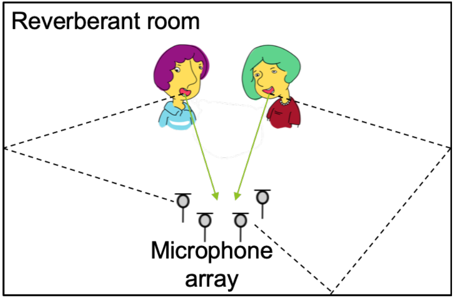
IEEE ICASSP June 2021

Ben-Gurion University of the Negev

facebook Reality Labs

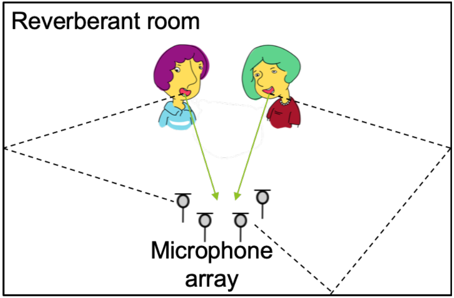
H. Beit-On and B. Rafaely, "Focusing and frequency smoothing for arbitrary arrays with application to speaker localization," IEEE/ACM TASLP, 2020.

Microphone arrays in reverberant environments



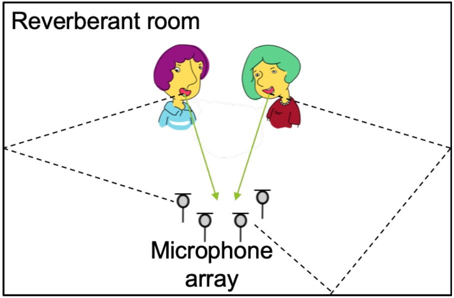
- Basic tasks:

Microphone arrays in reverberant environments



- Basic tasks:
  - Localization of sound-sources\reflections

Microphone arrays in reverberant environments



- Basic tasks:
  - Localization of sound-sources\reflections
  - Spatial filtering

Methods employing frequency smoothing

- Frequency smoothing that decorrelate coherent reflections in reverberant environment is useful for various applications:

Methods employing frequency smoothing

- Frequency smoothing that decorrelate coherent reflections in reverberant environment is useful for various applications:
  - O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," IEEE/ACM TASLP, vol. 22, no. 10, pp. 1494–1505, 2014.

Methods employing frequency smoothing

- Frequency smoothing that decorrelate coherent reflections in reverberant environment is useful for various applications:
  - O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," IEEE/ACM TASLP, vol. 22, no. 10, pp. 1494–1505, 2014.
  - T. Shlomo and B. Rafaely, "Blind localization of early room reflections using phase aligned spatial correlation," IEEE Transactions on Signal Processing, vol. 69, pp. 1213–1225, 2021.

Methods employing frequency smoothing

- Frequency smoothing that decorrelate coherent reflections in reverberant environment is useful for various applications:
  - O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," IEEE/ACM TASLP, vol. 22, no. 10, pp. 1494–1505, 2014.
  - T. Shlomo and B. Rafaely, "Blind localization of early room reflections using phase aligned spatial correlation," IEEE Transactions on Signal Processing, vol. 69, pp. 1213–1225, 2021.
  - M. Kalantari, "An approach for solving signal cancellation problem in spherical microphone array," Journal of Electrical and Computer Engineering Innovations (JECEI), 2021.

Methods employing frequency smoothing

- Frequency smoothing that decorrelate coherent reflections in reverberant environment is useful for various applications:
  - O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," IEEE/ACM TASLP, vol. 22, no. 10, pp. 1494–1505, 2014.
  - T. Shlomo and B. Rafaely, "Blind localization of early room reflections using phase aligned spatial correlation," IEEE Transactions on Signal Processing, vol. 69, pp. 1213–1225, 2021.
  - M. Kalantari, "An approach for solving signal cancellation problem in spherical microphone array," Journal of Electrical and Computer Engineering Innovations (JECEI), 2021.
- However, its application is limited to arrays with frequency-independent steering matrix (e.g. spherical)

Methods employing frequency smoothing

- Frequency smoothing that decorrelate coherent reflections in reverberant environment is useful for various applications:
  - O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," IEEE/ACM TASLP, vol. 22, no. 10, pp. 1494–1505, 2014.
  - T. Shlomo and B. Rafaely, "Blind localization of early room reflections using phase aligned spatial correlation," IEEE Transactions on Signal Processing, vol. 69, pp. 1213–1225, 2021.
  - M. Kalantari, "An approach for solving signal cancellation problem in spherical microphone array," Journal of Electrical and Computer Engineering Innovations (JECEI), 2021.
- However, its application is limited to arrays with frequency-independent steering matrix (e.g. spherical)
- We want to enable their application to arbitrary arrays

Focusing transformations

- Focusing process to removes steering matrix frequency dependence

Focusing transformations

- Focusing process to removes steering matrix frequency dependence
- A general signal model:

$$\mathbf{p}(f) = \mathbf{H}(f, \psi)\mathbf{s}(f) + \mathbf{n}(f)$$

Focusing transformations

- Focusing process to removes steering matrix frequency dependence
- A general signal model:

$$\mathbf{p}(f) = \mathbf{H}(f, \psi)\mathbf{s}(f) + \mathbf{n}(f)$$

- A focusing transformation $\mathbf{T}\left(f, f'\right)$ satisfies:

$$\mathbf{T}\left(f, f'\right)\mathbf{H}\left(f, \psi\right) = \mathbf{H}\left(f', \psi\right)$$

Focusing transformations

- Focusing process to removes steering matrix frequency dependence
- A general signal model:

$$\mathbf{p}(f) = \mathbf{H}(f,\psi)\mathbf{s}(f) + \mathbf{n}(f)$$

- A focusing transformation $\mathbf{T}\left(f, f'\right)$ satisfies:

$$\mathbf{T}\left(f, f'\right)\mathbf{H}\left(f,\psi\right) = \mathbf{H}\left(f',\psi\right)$$

- Transformed signal:

$$\begin{aligned} \mathbf{x}\left(f\right) &\triangleq \mathbf{T}\left(f, f'\right)\mathbf{p}\left(f\right) \\ &= \mathbf{H}\left(f',\psi\right)\mathbf{s}\left(f\right) + \tilde{\mathbf{n}}\left(f\right) \end{aligned}$$

Focusing transformations

- Focusing process to removes steering matrix frequency dependence
- A general signal model:

$$\mathbf{p}(f) = \mathbf{H}(f, \psi)\mathbf{s}(f) + \mathbf{n}(f)$$

- A focusing transformation $\mathbf{T}\left(f, f'\right)$ satisfies:

$$\mathbf{T}\left(f, f'\right)\mathbf{H}\left(f, \psi\right) = \mathbf{H}\left(f', \psi\right)$$

- Transformed signal:

$$\begin{aligned}\mathbf{x}\left(f\right) &\triangleq \mathbf{T}\left(f, f'\right)\mathbf{p}\left(f\right) \\ &= \mathbf{H}\left(f', \psi\right)\mathbf{s}\left(f\right) + \tilde{\mathbf{n}}\left(f\right)\end{aligned}$$

- Focusing matrices:

$$\mathbf{T}\left(f, f'\right) = \mathbf{H}\left(f', \psi\right)\mathbf{H}\left(f, \psi\right)^{\dagger}$$

Focusing transformations

- Focusing process to removes steering matrix frequency dependence
- A general signal model:

$$\mathbf{p}(f) = \mathbf{H}(f,\psi)\mathbf{s}(f) + \mathbf{n}(f)$$

- A focusing transformation $\mathbf{T}(f,f')$ satisfies:

$$\mathbf{T}\left(f,f'\right)\mathbf{H}\left(f,\psi\right) = \mathbf{H}\left(f',\psi\right)$$

- Transformed signal:

$$\begin{aligned}\mathbf{x}\left(f\right) &\triangleq \mathbf{T}\left(f,f'\right)\mathbf{p}\left(f\right)\\ &= \mathbf{H}\left(f',\psi\right)\mathbf{s}\left(f\right) + \tilde{\mathbf{n}}\left(f\right)\end{aligned}$$

- Focusing matrices:

$$\mathbf{T}\left(f,f'\right) = \mathbf{H}\left(f',\psi\right)\mathbf{H}\left(f,\psi\right)^{\dagger}$$

× Computation of $\mathbf{T}\left(f',f\right)$ requires the directions of all sources ψ

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

- Updated system model:

$$\mathbf{p}(f) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)\,\mathbf{s}(f) + \mathbf{n}(f)$$

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

- Updated system model:

$$\mathbf{p}(f) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)\,\mathbf{s}(f) + \mathbf{n}(f)$$

- Focusing transformation should satisfy:

$$\mathbf{T}\left(f,f'\right)\mathbf{H_{nm}}(f) = \mathbf{H_{nm}}\left(f'\right)$$

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

- Updated system model:

$$\mathbf{p}(f) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)\,\mathbf{s}(f) + \mathbf{n}(f)$$

- Focusing transformation should satisfy:

$$\mathbf{T}\left(f,f'\right)\mathbf{H_{nm}}(f) = \mathbf{H_{nm}}\left(f'\right)$$

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f,f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}(f)^{\dagger}$$

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

- Updated system model:

$$\mathbf{p}(f) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)\,\mathbf{s}(f) + \mathbf{n}(f)$$

- Focusing transformation should satisfy:

$$\mathbf{T}\left(f,f'\right)\mathbf{H_{nm}}(f) = \mathbf{H_{nm}}\left(f'\right)$$

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f,f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}(f)^{\dagger}$$

- In this work:

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

- Updated system model:

$$\mathbf{p}(f) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)\,\mathbf{s}(f) + \mathbf{n}(f)$$

- Focusing transformation should satisfy:

$$\mathbf{T}\left(f,f'\right)\mathbf{H_{nm}}(f) = \mathbf{H_{nm}}\left(f'\right)$$

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f,f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}(f)^{\dagger}$$

- In this work:
  - Is direction-independent focusing effective?

Direction-independent focusing

- Representing the steering matrix using spherical harmonics:

$$\mathbf{H}(f,\psi) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)$$

- Updated system model:

$$\mathbf{p}(f) = \mathbf{H_{nm}}(f)\,\mathbf{Y}(\psi)\,\mathbf{s}(f) + \mathbf{n}(f)$$

- Focusing transformation should satisfy:

$$\mathbf{T}\left(f,f'\right)\mathbf{H_{nm}}(f) = \mathbf{H_{nm}}\left(f'\right)$$

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f,f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}(f)^{\dagger}$$

- In this work:
  - Is direction-independent focusing effective?
  - What affects its performance?

Factors affecting focusing error

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f,f'\right)=\mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}\left(f\right)^{\dagger}$$

Factors affecting focusing error

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f, f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}\left(f\right)^{\dagger}$$

1. **Truncation error:** $\mathbf{T}(f, f')$ is computed using a truncated $\mathbf{H_{nm}}(f)$

Factors affecting focusing error

- Direction-independent focusing matrix:

$$\mathbf{T}\left(f, f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}\left(f\right)^{\dagger}$$

1. **Truncation error:** $\mathbf{T}\left(f, f'\right)$ is computed using a truncated $\mathbf{H_{nm}}\left(f\right)$
2. **Matrix inversion error:** more coefficients than mics. $(N+1)^2 > I$

Factors affecting focusing error

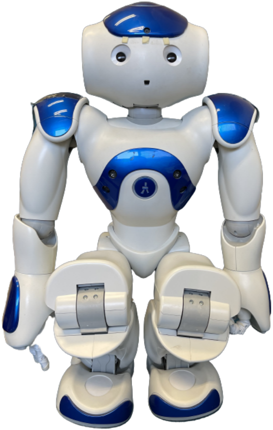
- Direction-independent focusing matrix:

$$\mathbf{T}\left(f, f'\right) = \mathbf{H_{nm}}\left(f'\right)\mathbf{H_{nm}}\left(f\right)^{\dagger}$$

1. **Truncation error:** $\mathbf{T}(f, f')$ is computed using a truncated $\mathbf{H_{nm}}(f)$
2. **Matrix inversion error:** more coefficients than mics. $(N+1)^2 > I$
3. **Spatial aliasing:** insufficient samples to compute $\mathbf{H_{nm}}(f)$, $L < (N+1)^2$

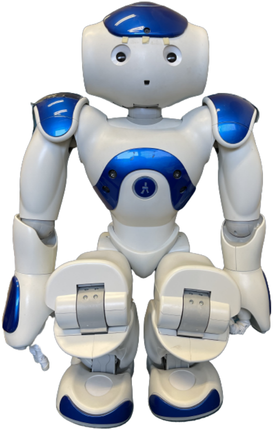
Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head



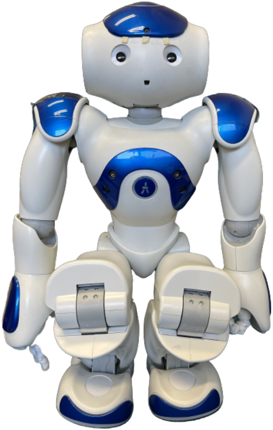
Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$



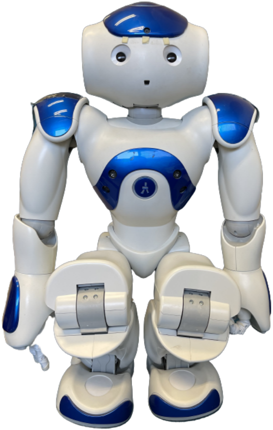
Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\mathrm{cm}$
- Steering function samples from $L = 240$ directions



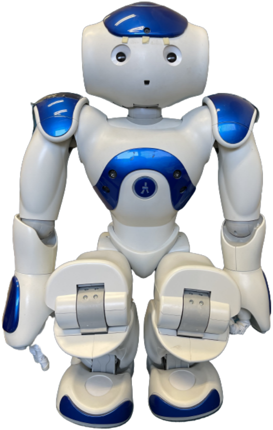
Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$
- Steering function samples from $L = 240$ directions
- Sampling frequency $Fs = 10\,\text{kHz}$



Focusing analysis for robot head

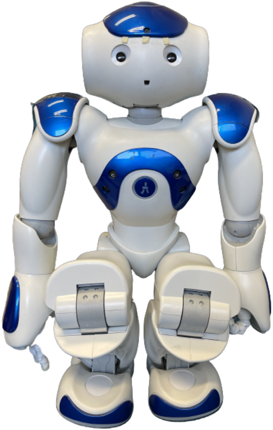
- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$
- Steering function samples from $L = 240$ directions
- Sampling frequency $Fs = 10\,\text{kHz}$
- Spherical harmonics order $N \leq 6$



Focusing analysis for robot head

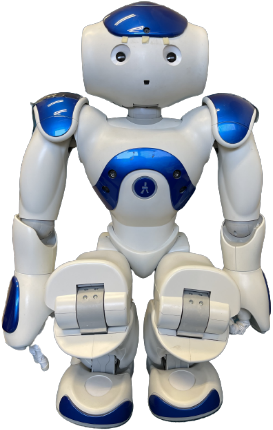
- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$
- Steering function samples from $L = 240$ directions
- Sampling frequency $Fs = 10\,\text{kHz}$
- Spherical harmonics order $N \leq 6$

- For focusing matrix of order N:



Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$
- Steering function samples from $L = 240$ directions
- Sampling frequency $Fs = 10\,\text{kHz}$
- Spherical harmonics order $N \leq 6$

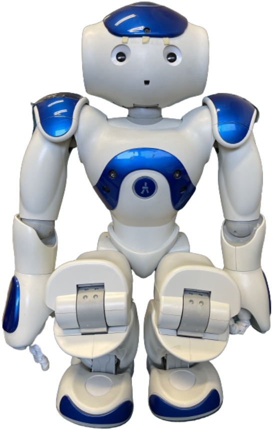


- For focusing matrix of order N:

1. Truncation error: small for $6 \leq N$

Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$
- Steering function samples from $L = 240$ directions
- Sampling frequency $Fs = 10\,\text{kHz}$
- Spherical harmonics order $N \leq 6$

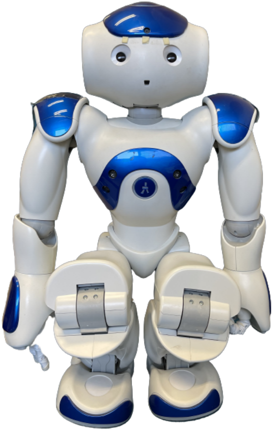


- For focusing matrix of order N:

1. Truncation error: small for $6 \leq N$
2. Inversion error: small for $N = 1, 2$, such that $(N+1)^2 < I$

Focusing analysis for robot head

- $I = 12$ mics. mounted on Nao robot head
- Head radius $r \approx 6\,\text{cm}$
- Steering function samples from $L = 240$ directions
- Sampling frequency $Fs = 10\,\text{kHz}$
- Spherical harmonics order $N \leq 6$



- For focusing matrix of order N:

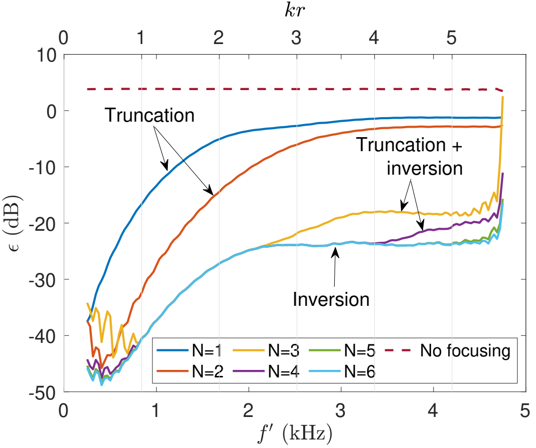
1. Truncation error: small for $6 \leq N$
2. Inversion error: small for $N = 1, 2$, such that $(N+1)^2 < I$
3. Spatial aliasing: small because $(N+1)^2 \leq 49 < L$

Focusing performance for robot head

- Focusing error: $$\epsilon\left(f'\right) = \frac{1}{J_f}\sum_{f}\frac{\|\mathbf{T}(f,f')\mathbf{H_{nm}}(f) - \mathbf{H_{nm}}(f')\|_F}{\|\mathbf{H_{nm}}(f)\|_F}$$

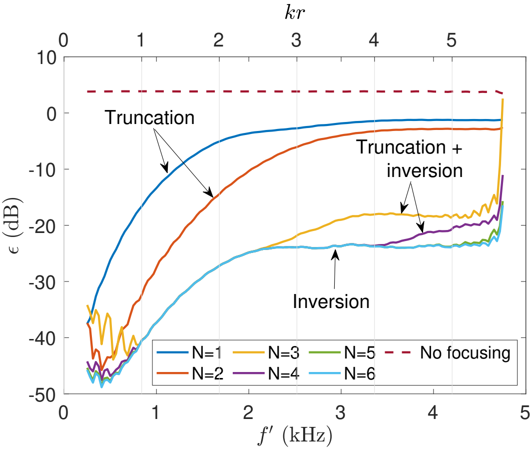
Focusing performance for robot head

- Focusing error: $\epsilon\left(f'\right) = \frac{1}{J_f}\sum_{f}\frac{\|\mathbf{T}(f,f')\mathbf{H_{nm}}(f)-\mathbf{H_{nm}}(f')\|_F}{\|\mathbf{H_{nm}}(f)\|_F}$



Focusing performance for robot head

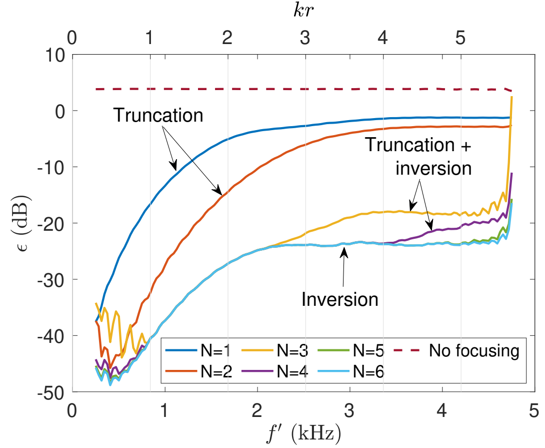
- Focusing error: $\epsilon(f') = \frac{1}{J_f} \sum_f \frac{\|\mathbf{T}(f,f')\mathbf{H}_{\mathbf{nm}}(f) - \mathbf{H}_{\mathbf{nm}}(f')\|_F}{\|\mathbf{H}_{\mathbf{nm}}(f)\|_F}$



- For $N = 1, 2$ matrix inversion error is small, since $l = 12$

Focusing performance for robot head

- Focusing error: $$\epsilon\left(f'\right) = \frac{1}{J_f}\sum_f \frac{\|\mathbf{T}(f,f')\mathbf{H}_{\mathbf{nm}}(f) - \mathbf{H}_{\mathbf{nm}}(f')\|_F}{\|\mathbf{H}_{\mathbf{nm}}(f)\|_F}$$



- For $N = 1, 2$ matrix inversion error is small, since $I = 12$
- For $N = 6$ truncation error is negligible

Speaker localization using LOCATA challenge data

- Real world recordings of speakers in a room using Nao head array

Speaker localization using LOCATA challenge data

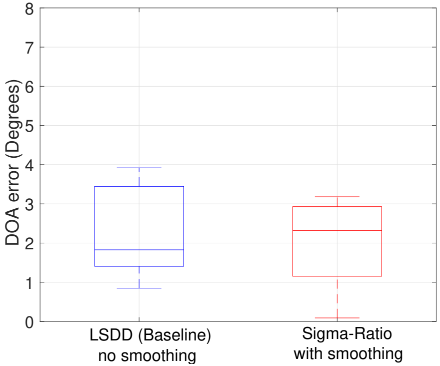
- Real world recordings of speakers in a room using Nao head array
- 2 direct-path dominance (DPD) test based methods were compared

Speaker localization using LOCATA challenge data

- Real world recordings of speakers in a room using Nao head array
- 2 direct-path dominance (DPD) test based methods were compared
- Only one incorporate frequency smoothing and focusing

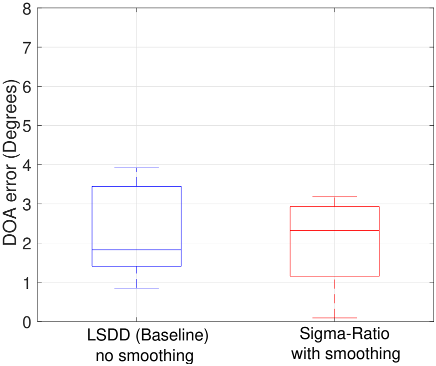
Speaker localization using LOCATA challenge data

- Real world recordings of speakers in a room using Nao head array
- 2 direct-path dominance (DPD) test based methods were compared
- Only one incorporate frequency smoothing and focusing



Speaker localization using LOCATA challenge data

- Real world recordings of speakers in a room using Nao head array
- 2 direct-path dominance (DPD) test based methods were compared
- Only one incorporate frequency smoothing and focusing



- Focusing and frequency smoothing perform well

Conclusions

- The factors affecting focusing error were formulated

Conclusions

- The factors affecting focusing error were formulated
- Focusing and frequency smoothing perform well for speaker localization

Conclusions

- The factors affecting focusing error were formulated
- Focusing and frequency smoothing perform well for speaker localization
- This implies that focusing and frequency smoothing can be applied for other applications as well