Speech Enhancement Using Masking for Binaural Reproduction of Ambisonics Signals

Moti Lugasi and Boaz Rafaely

ICASSP June 2021

facebook Reality Labs

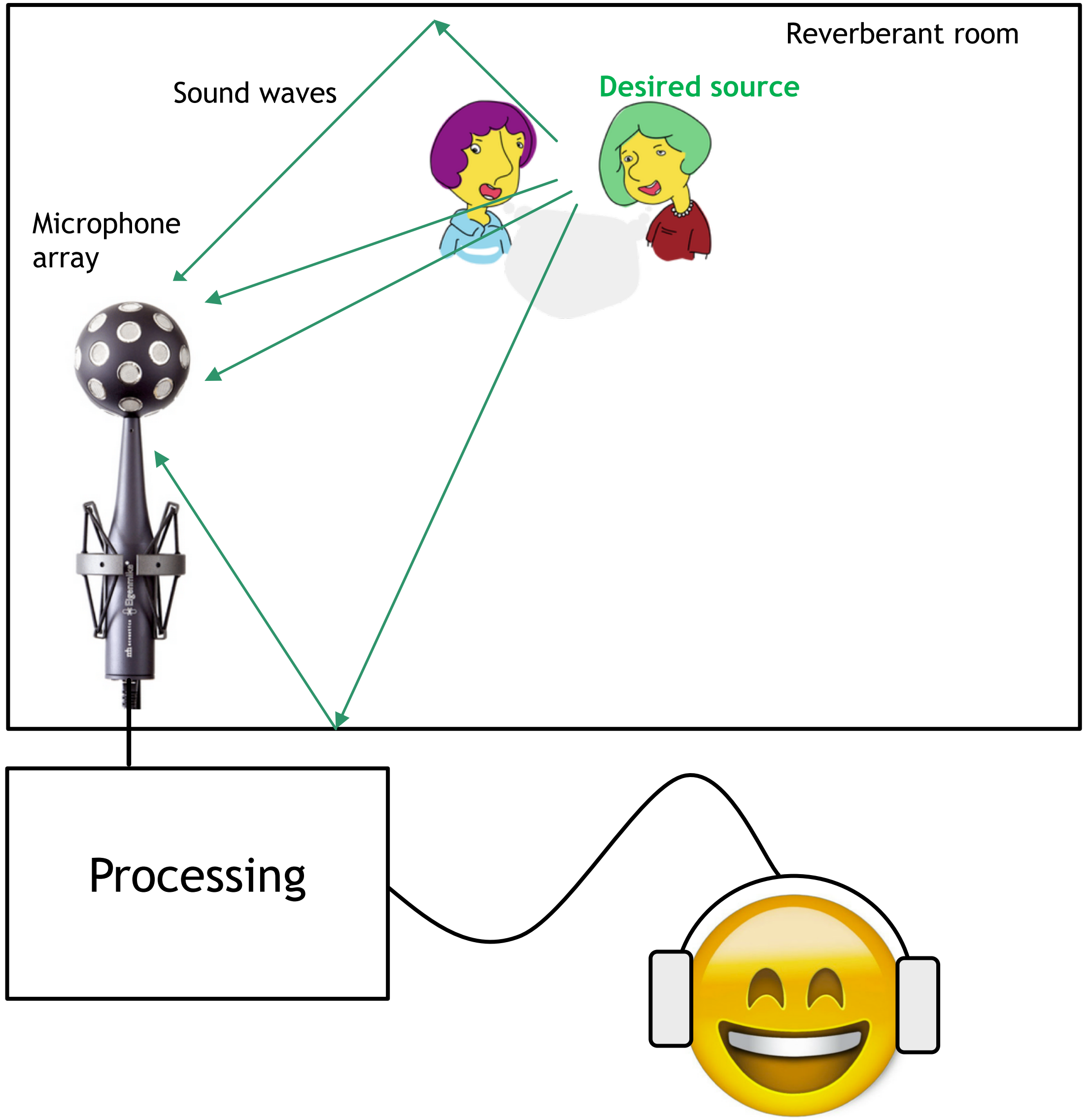
School of Electrical and Computer Engineering
Ben-Gurion University of the Negev

Lugasi, Moti, and Boaz Rafaely. "Speech Enhancement Using Masking for Binaural Reproduction of Ambisonics Signals." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020): 1767-1777.

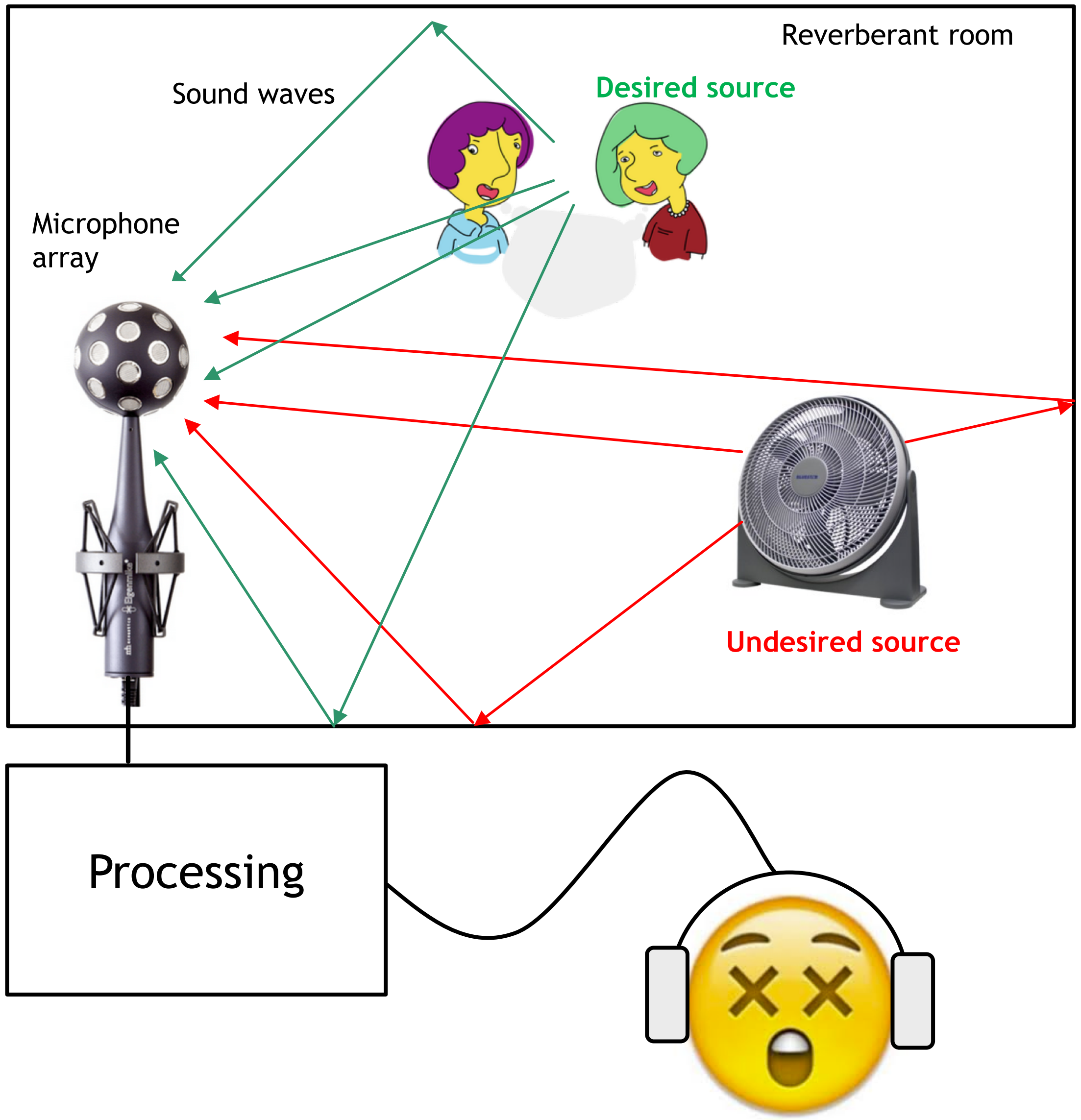
Contents

- The problem - noisy Ambisonics signals
- The aim - enhancement of these signals
- The method - masking the noise
- Objective analysis
- Conclusions

The problem – noisy Ambisonics signals



The problem – noisy Ambisonics signals



The research aims

The research aims

- Attenuate the undesired components

The research aims

- Attenuate the undesired components
- Preserve the desired components

The research aims

- Attenuate the undesired components
- Preserve the desired components
- Preserve the spatial cues of the acoustic scene

Current methods for speech enhancement

- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.

Current methods for speech enhancement

- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- **Significantly distort the desired sound field**

Current methods for speech enhancement

- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.

- **Significantly distort the desired sound field**

- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- Abend, U., & Rafaely, B. (2016, November). Spatio-spectral masking for spherical array beamforming. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)* (pp. 1-5). IEEE.

Current methods for speech enhancement

- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.

- **Significantly distort the desired sound field**

- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- Abend, U., & Rafaely, B. (2016, November). Spatio-spectral masking for spherical array beamforming. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)* (pp. 1-5). IEEE.

- **May preserve the entire sound field**

Current methods for speech enhancement

- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- **Significantly distort the desired sound field**
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- Abend, U., & Rafaely, B. (2016, November). Spatio-spectral masking for spherical array beamforming. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)* (pp. 1-5). IEEE.
- **May preserve the entire sound field**
- **Based on time-frequency masking**

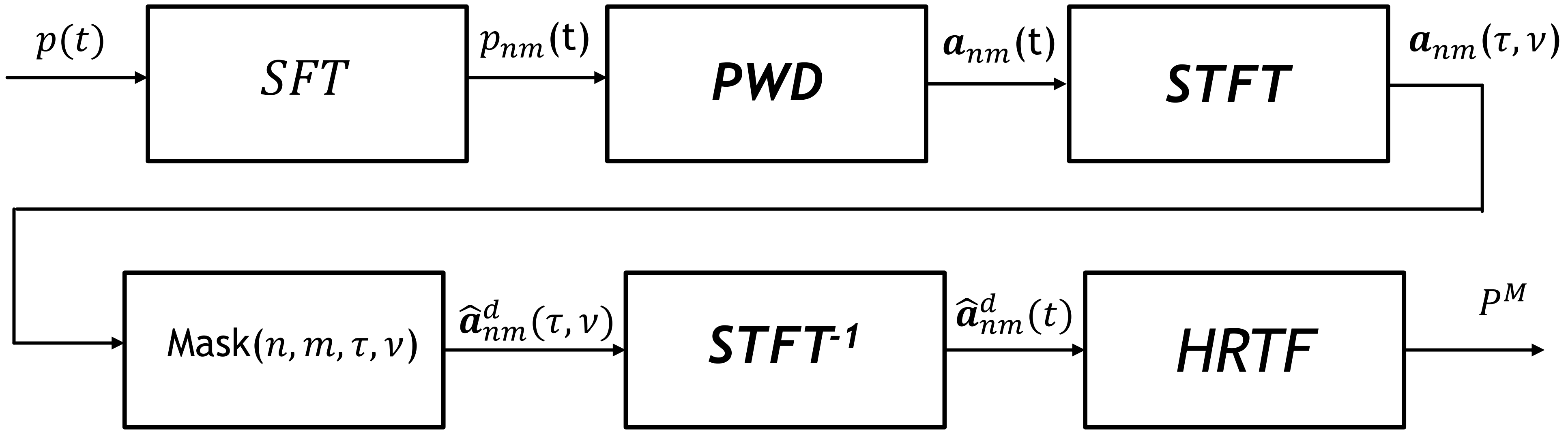
Current methods for speech enhancement

- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- **Significantly distort the desired sound field**
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- Abend, U., & Rafaely, B. (2016, November). Spatio-spectral masking for spherical array beamforming. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)* (pp. 1-5). IEEE.
- **May preserve the entire sound field**
- **Based on time-frequency masking**
- **Have not yet been extensively investigated**

Current methods for speech enhancement

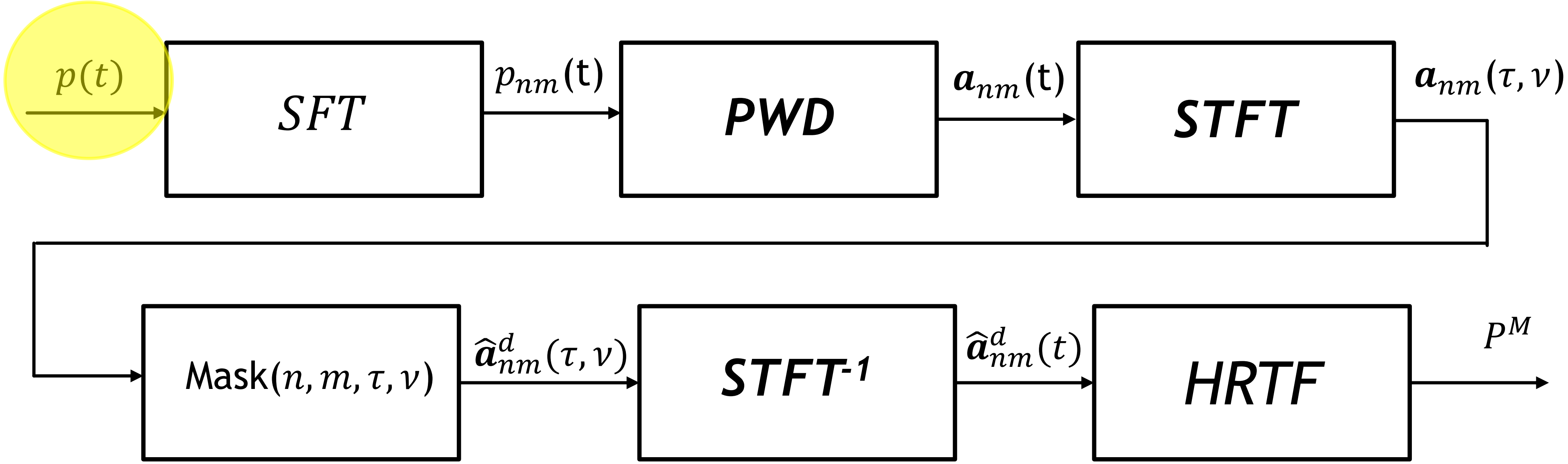
- Shabtai, N. R., & Rafaely, B. (2013). Generalized spherical array beamforming for binaural speech reproduction. *IEEE/ACM transactions on audio, speech, and language processing*, *22*(1), 238-247.
- Sun, H., Yan, S., & Svensson, U. P. (2011). Optimal higher order ambisonics encoding with predefined constraints. *IEEE transactions on audio, speech, and language processing*, *20*(3), 742-754.
- Borrelli, C., Canclini, A., Antonacci, F., Sarti, A., & Tubaro, S. (2018, September). A denoising methodology for higher order ambisonics recordings. In *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)* (pp. 451-455). IEEE.
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- **Significantly distort the desired sound field**
- Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.
- Abend, U., & Rafaely, B. (2016, November). Spatio-spectral masking for spherical array beamforming. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)* (pp. 1-5). IEEE.
- **May preserve the entire sound field**
- **Based on time-frequency masking**
- **Have not yet been extensively investigated**

Wiener mask in the SH domain - TFSH



Abend, U., & Rafaely, B. (2016, November). Spatio-spectral masking for spherical array beamforming. In *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)* (pp. 1-5). IEEE.

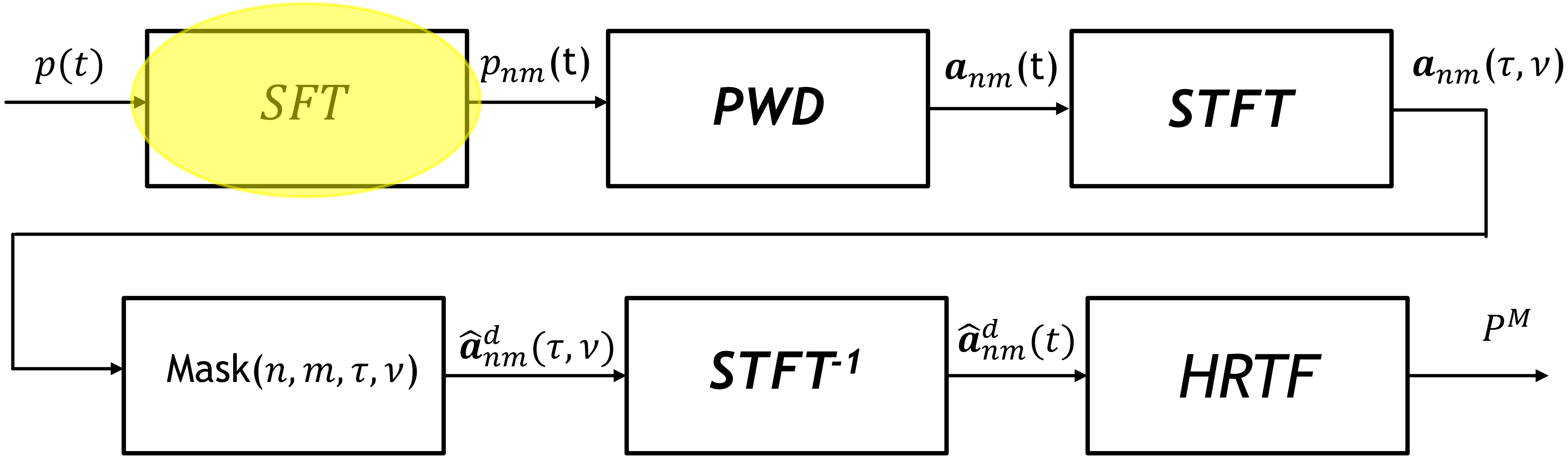
Wiener mask in the SH domain - TFSH



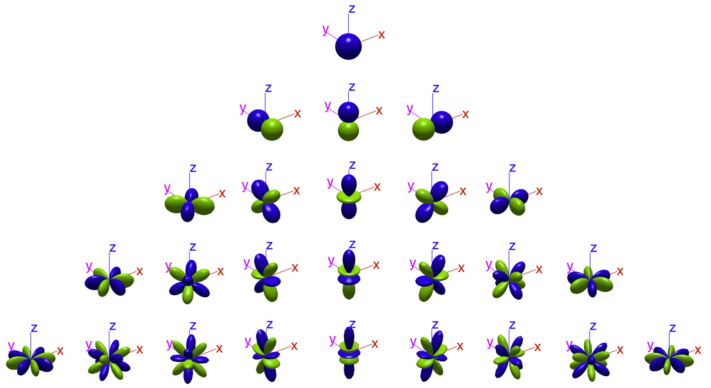
Microphone
input signals

$$p(t) = p_d(t) + p_u(t)$$

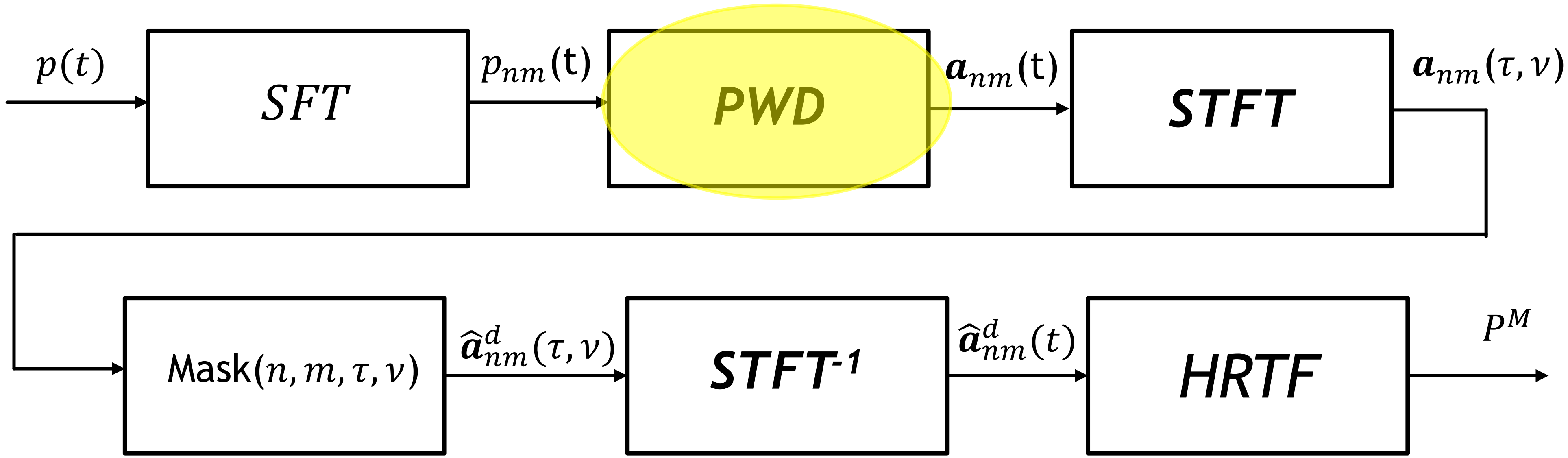
Wiener mask in the SH domain - TFSH



Spherical
Fourier
Transform

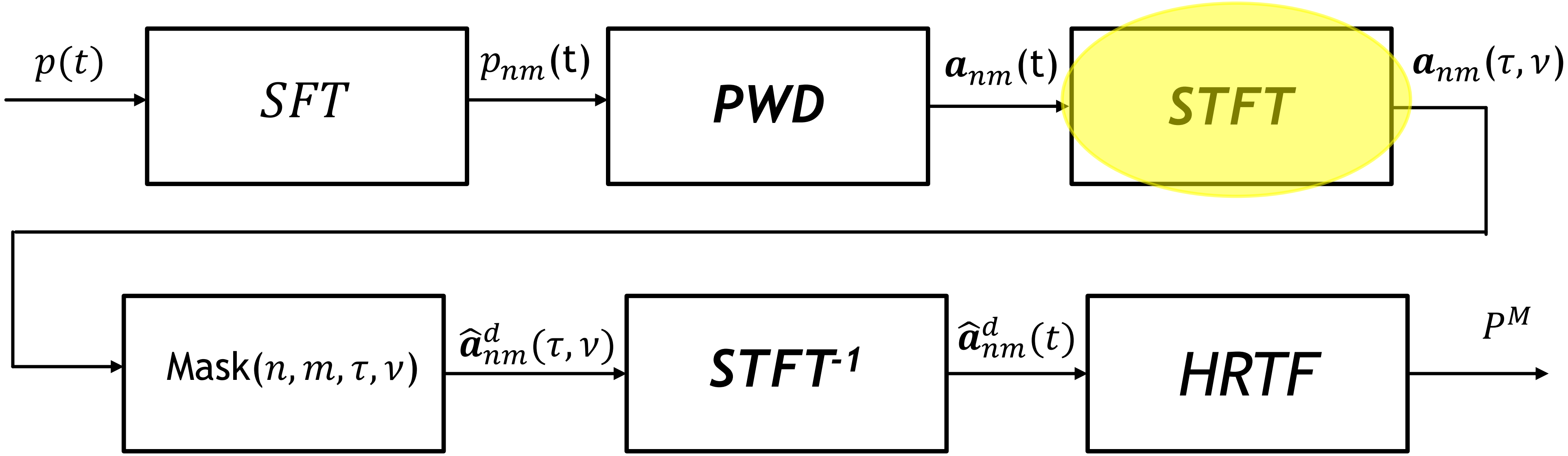


Wiener mask in the SH domain - TFSH



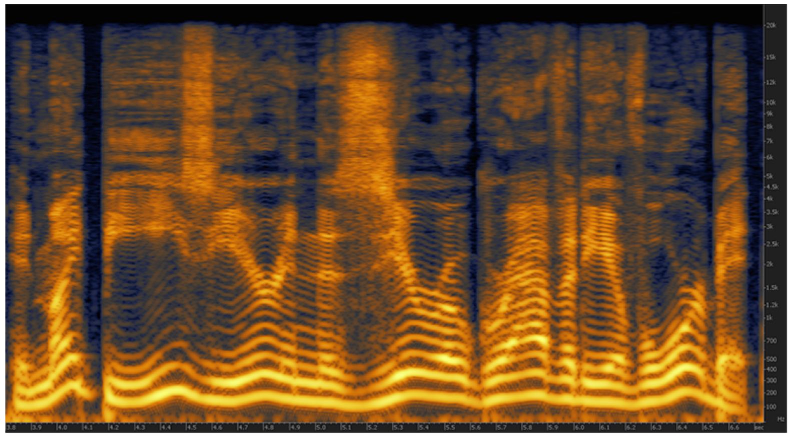
Plane Wave
Decomposition

Wiener mask in the SH domain - TFSH

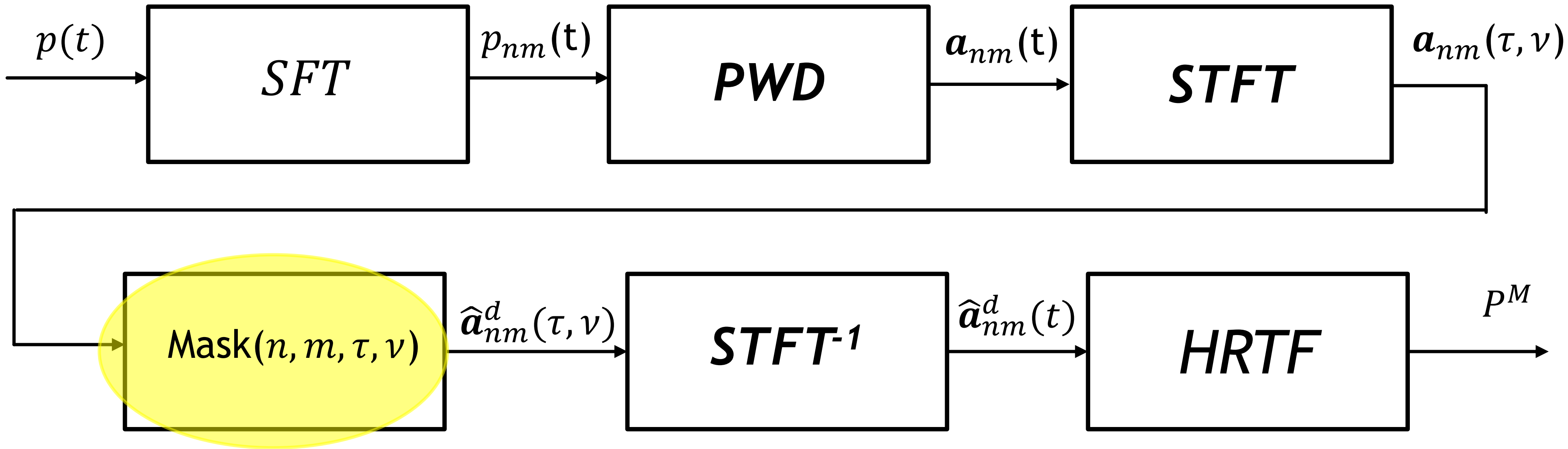


Short Time Fourier Transform

$$\boldsymbol{a}_{nm}(\tau,\nu) = \boldsymbol{a}^{\mathrm{d}}_{\mathrm{nm}}(\tau,\nu) + \boldsymbol{a}^{\mathrm{u}}_{\mathrm{nm}}(\tau,\nu)$$



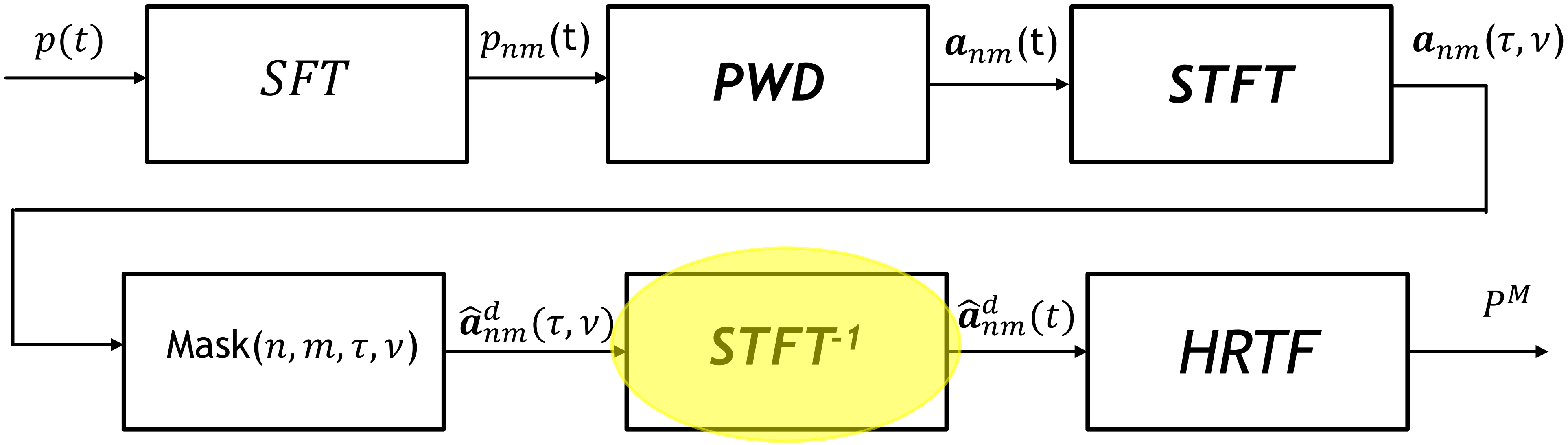
Wiener mask in the SH domain - TFSH



Wiener masking

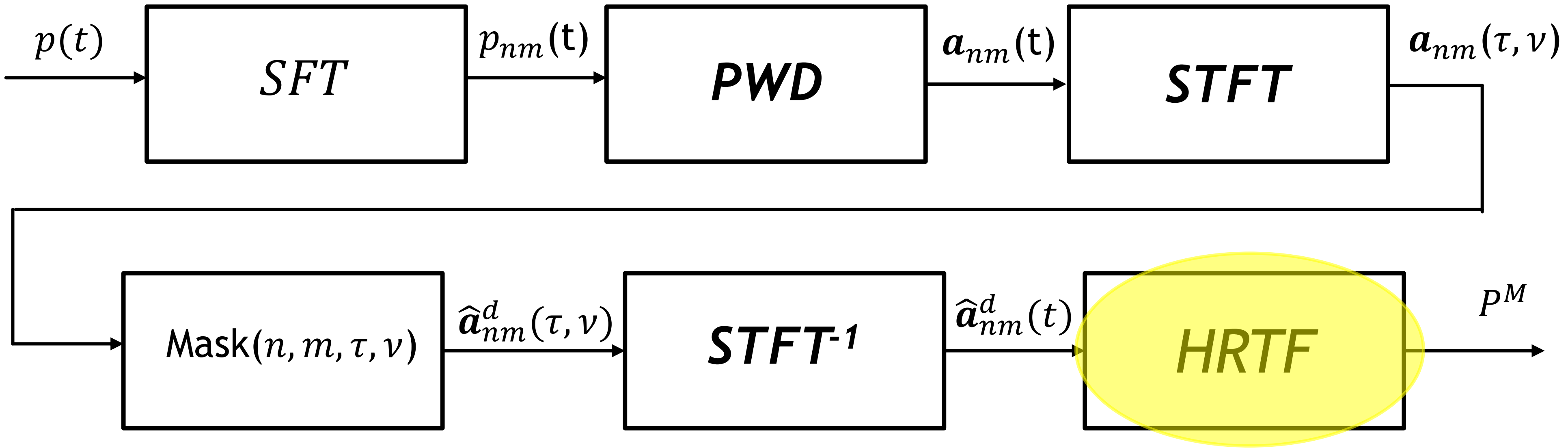
$$Mask(n, m, \tau, \nu) = \frac{SNR(n, m, \tau, \nu)}{SNR(n, m, \tau, \nu) + 1}$$

Wiener mask in the SH domain - TFSH



Inverse Short Time Fourier Transform

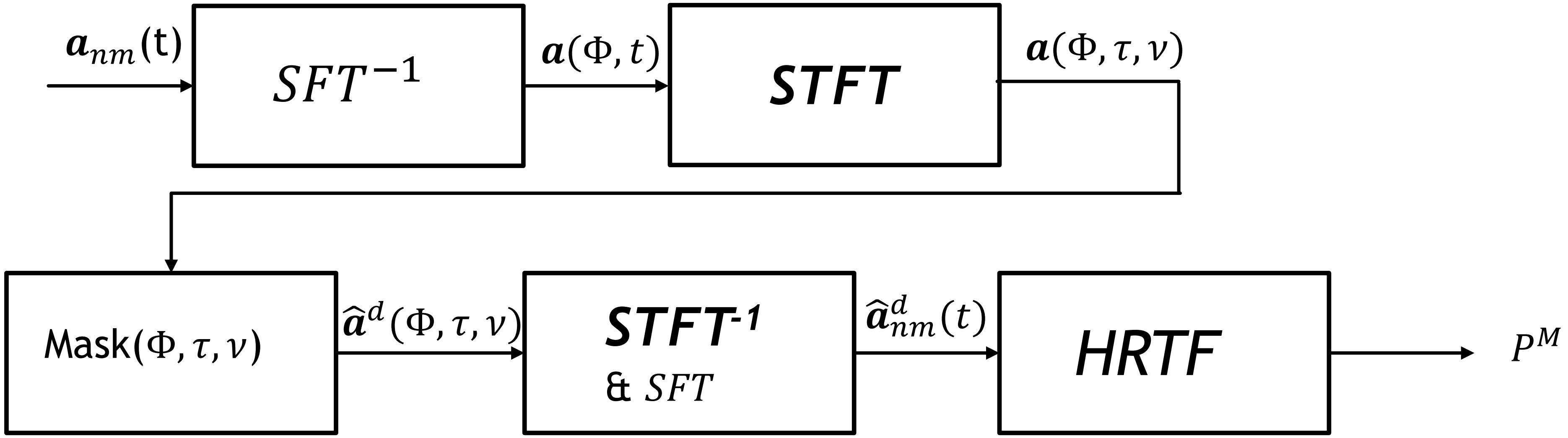
Wiener mask in the SH domain - TFSH



Head Related Transfer Function

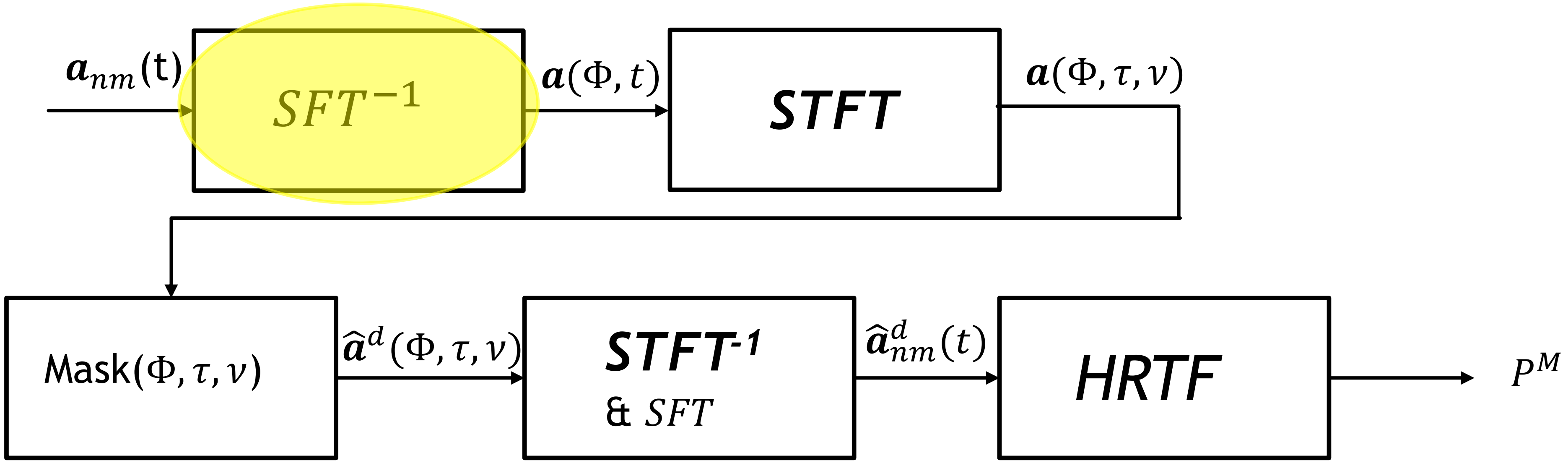
$$P^M = P^{Md} + P^{Mu}$$

Wiener mask in the spatial domain - TFS



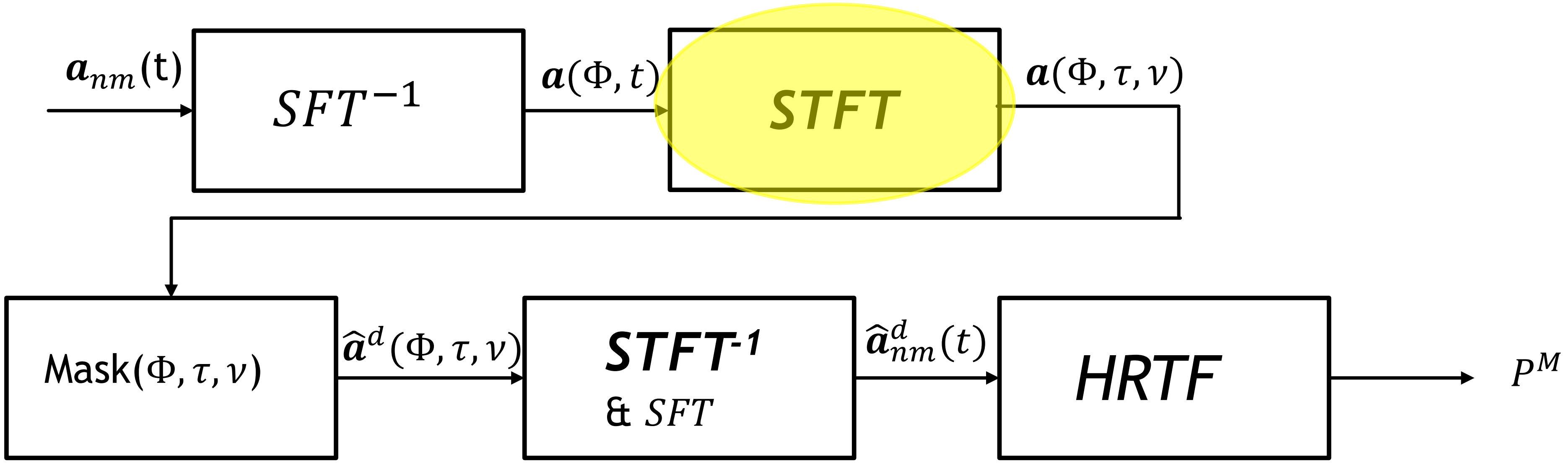
Herzog, A., & Habets, E. A. (2019, May). Direction Preserving Wiener Matrix Filtering for Ambisonic Input-output Systems. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 446-450). IEEE.

Wiener mask in the spatial domain - TFS



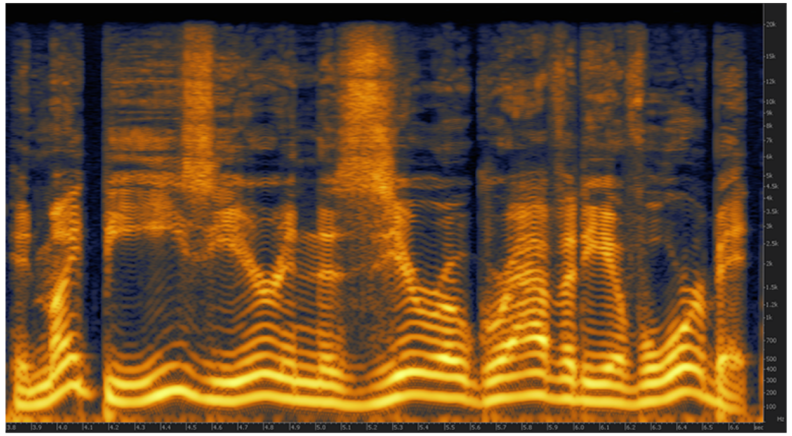
Inverse Spherical Fourier Transform

Wiener mask in the spatial domain - TFS

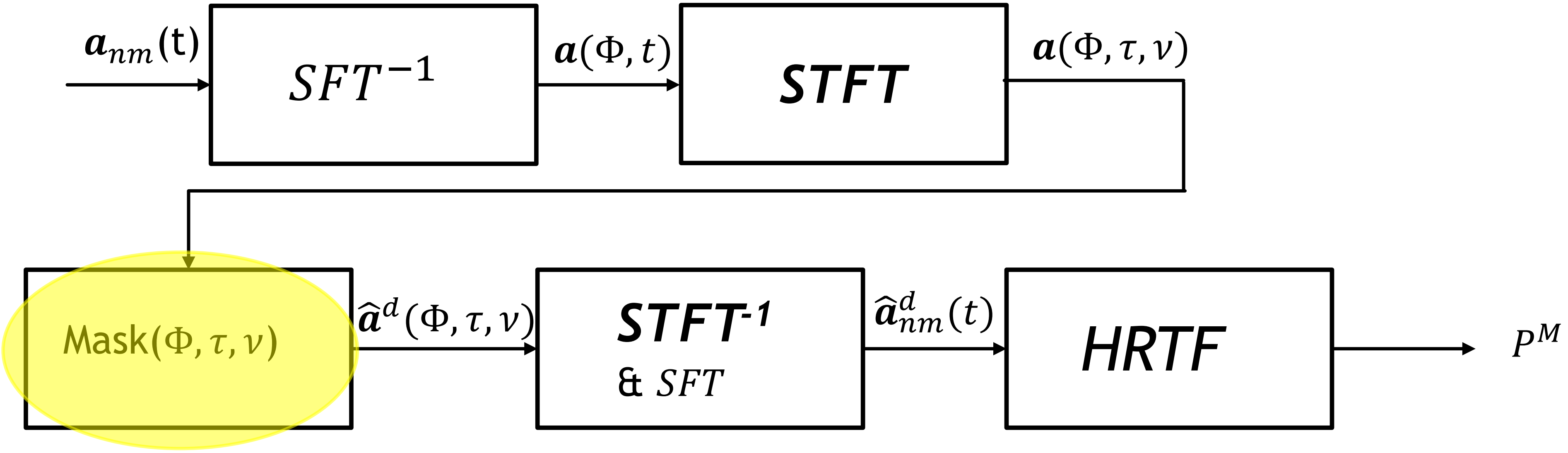


Short Time Fourier
Transform

$$\boldsymbol{a}(\Phi, \tau, \nu) = \boldsymbol{a}_d(\Phi, \tau, \nu) + \boldsymbol{a}_u(\Phi, \tau, \nu)$$



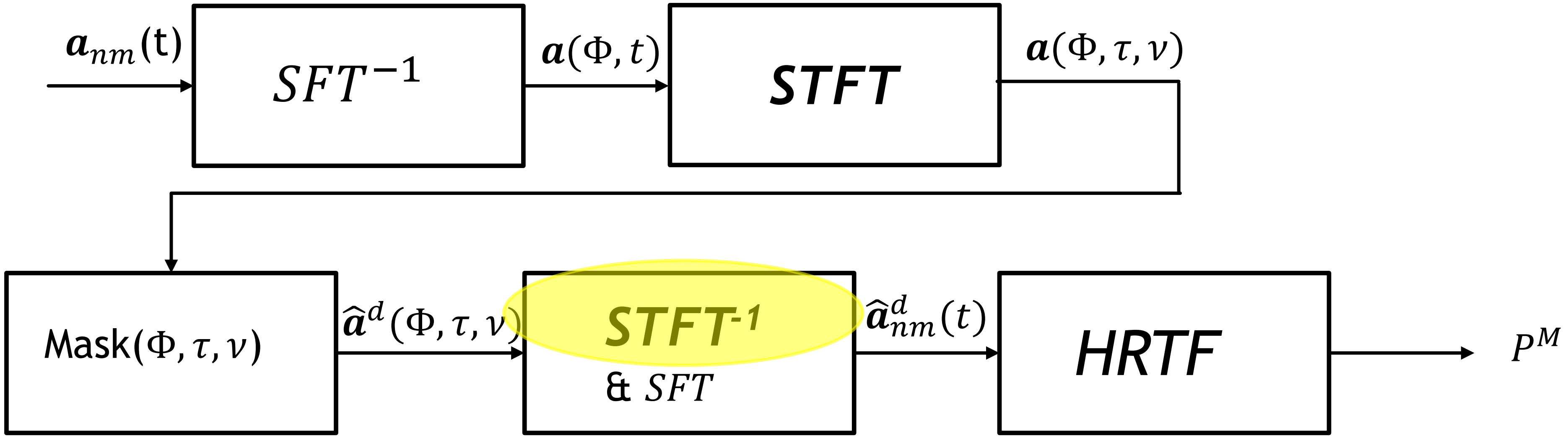
Wiener mask in the spatial domain - TFS



Wiener masking

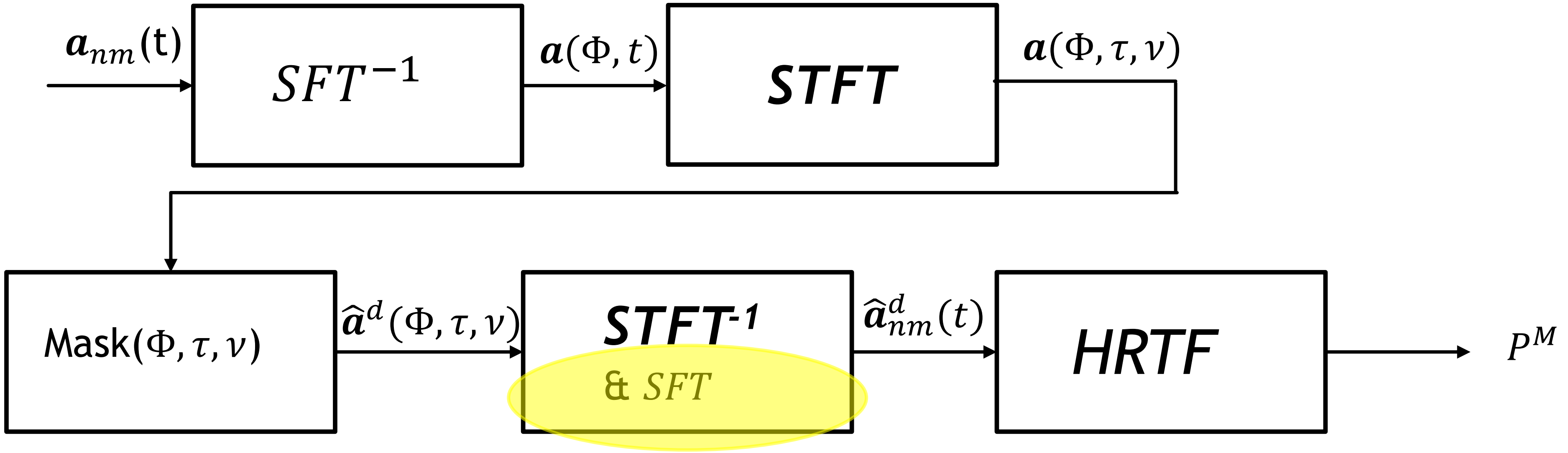
$$Mask(\Phi, \tau, \nu) = \frac{SNR(\Phi, \tau, \nu)}{SNR(\Phi, \tau, \nu) + 1}$$

Wiener mask in the spatial domain - TFS

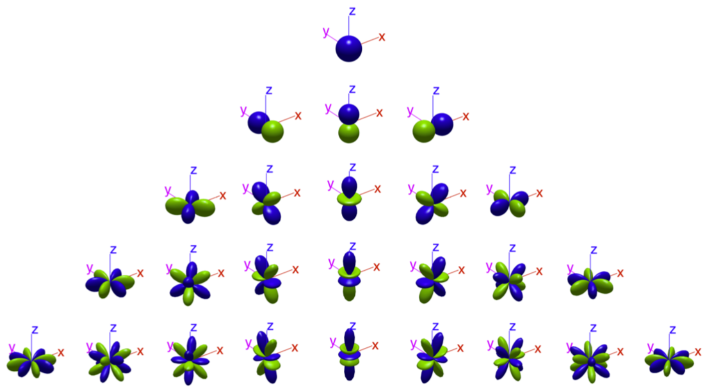


Inverse Short Time Fourier Transform

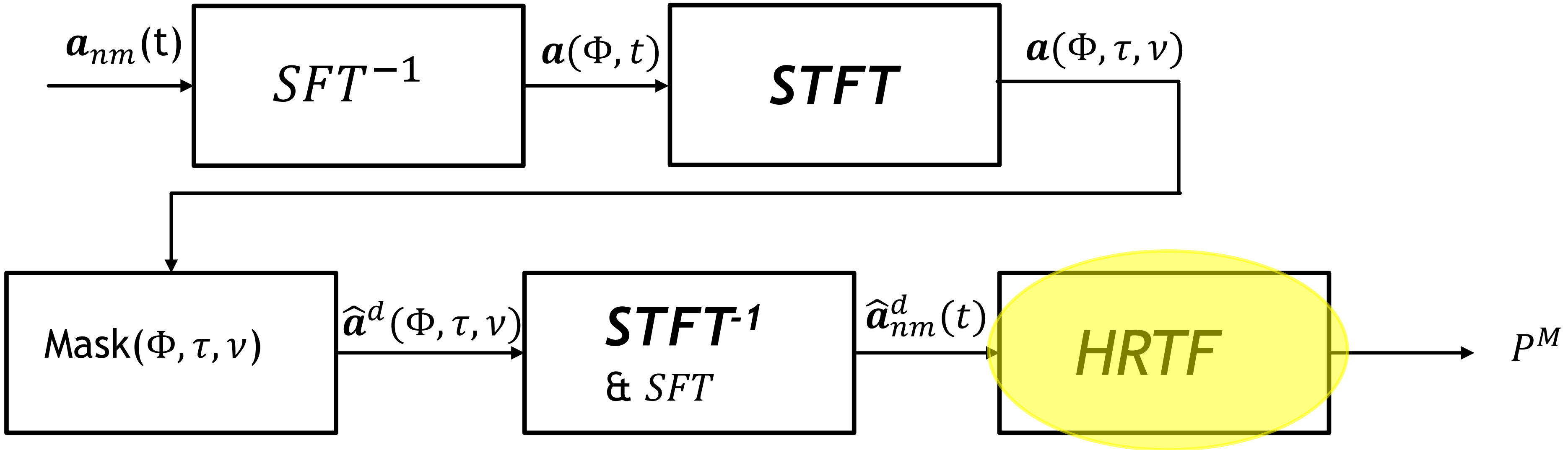
Wiener mask in the spatial domain - TFS



Spherical Fourier Transform



Wiener mask in the spatial domain - TFS



Head Related Transfer Function

$$P^{M} = P^{Md} + P^{Mu}$$

Objective performance measures

Objective performance measures

- Overall quality:

Objective performance measures

- Overall quality:
  - Signal to noise ratio gain (G_{SNR}):

$$G_{SNR} = \frac{SNR_{out}}{SNR_{in}}, \qquad SNR_{out} = \frac{||P^{Md}||^2}{||P^{Mu}||^2}, \qquad SNR_{in} = \frac{||P^{d}||^2}{||P^{u}||^2}$$

- P^{d}- unprocessed desired binaural signals
- P^{u}- unprocessed desired binaural signals
- P^{Md}- masked desired binaural signals
- P^{Mu}- masked undesired binaural signals

Objective performance measures

- Overall quality:
  - Signal to noise ratio gain (G_{SNR}):

$$G_{SNR} = \frac{SNR_{out}}{SNR_{in}}, \qquad SNR_{out} = \frac{||P^{Md}||^2}{||P^{Mu}||^2}, \qquad SNR_{in} = \frac{||P^{d}||^2}{||P^{u}||^2}$$

  - signal to distortion ratio (SDR):

$$SDR = \frac{||P^{d}||^2}{||P^{d} - P^{Md}||^2}$$

Objective performance measures

- Overall quality:
  - Signal to noise ratio gain (G_{SNR}):

$$G_{SNR} = \frac{SNR_{out}}{SNR_{in}}, \qquad SNR_{out} = \frac{||P^{Md}||^2}{||P^{Mu}||^2}, \qquad SNR_{in} = \frac{||P^{d}||^2}{||P^{u}||^2}$$

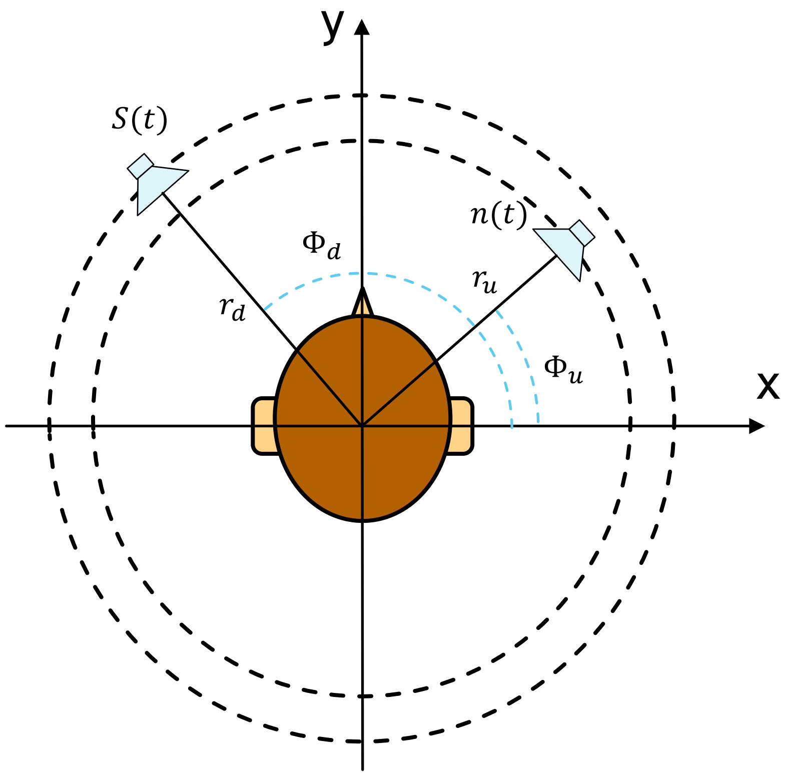
  - signal to distortion ratio (SDR):

$$SDR = \frac{||P^{d}||^2}{||P^{d} - P^{Md}||^2}$$

- Spatial cues of the residual noise:
  - Inter-aural level difference (ILD) - for $f > 1{,}500 H_z$
  - Inter-aural cross correlation time (IACC_t) - for $f < 1{,}500 H_z$

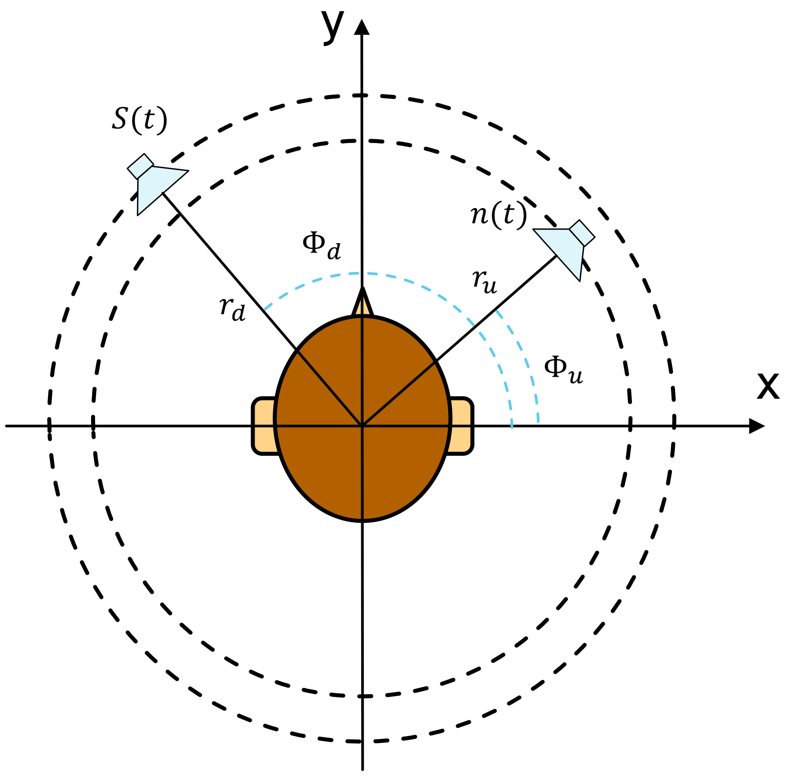
Monte Carlo simulation - setup

- This simulation is repeated for various:
  - Speakers
  - SNRin values
  - Noise types
  - Rooms
  - Source directions (Φ_d, Φ_u)
  - Source distances (r_d, r_u)
- Total of 1728 realizations for all combinations

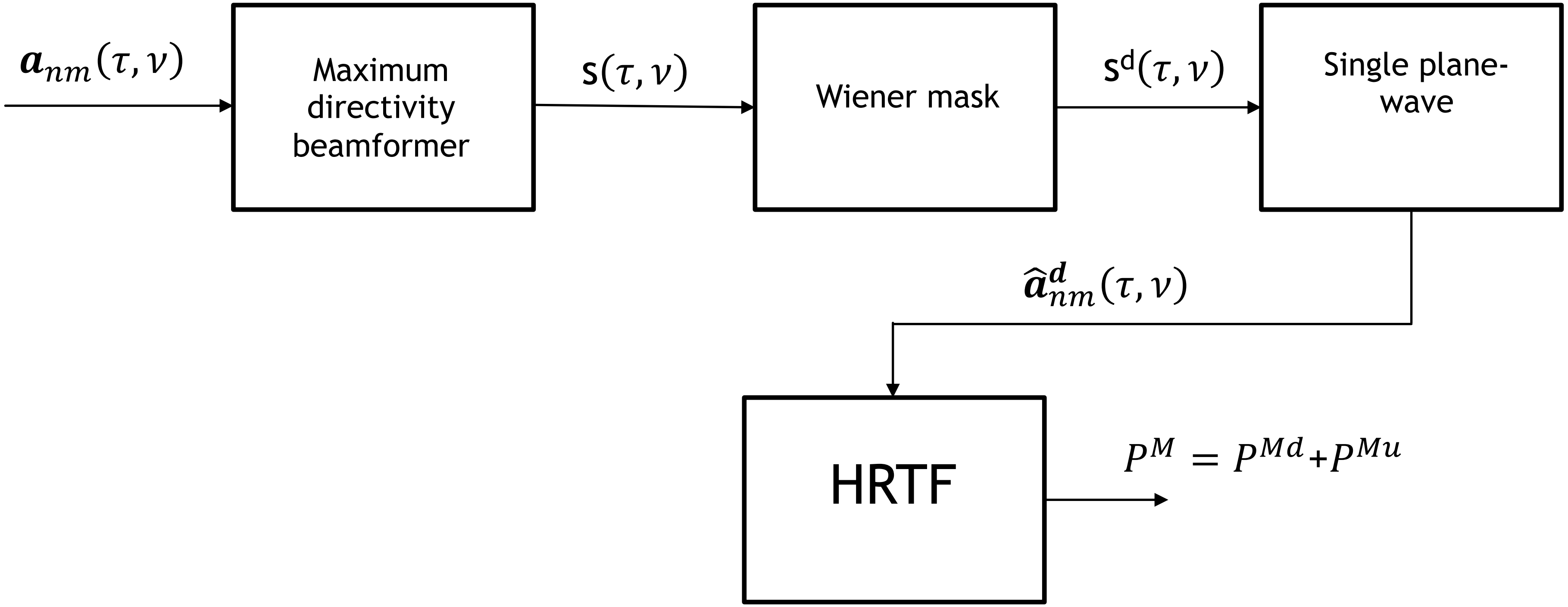


Monte Carlo simulation - setup

- This simulation is repeated for various:
  - Speakers
  - SNRin values
  - Noise types
  - Rooms
  - Source directions (Φ_d, Φ_u)
  - Source distances (r_d, r_u)
- Total of 1728 realizations for all combinations
- Three methods investigated:
  - TFS masking
  - TFSH masking
  - Beamforming + TF masking

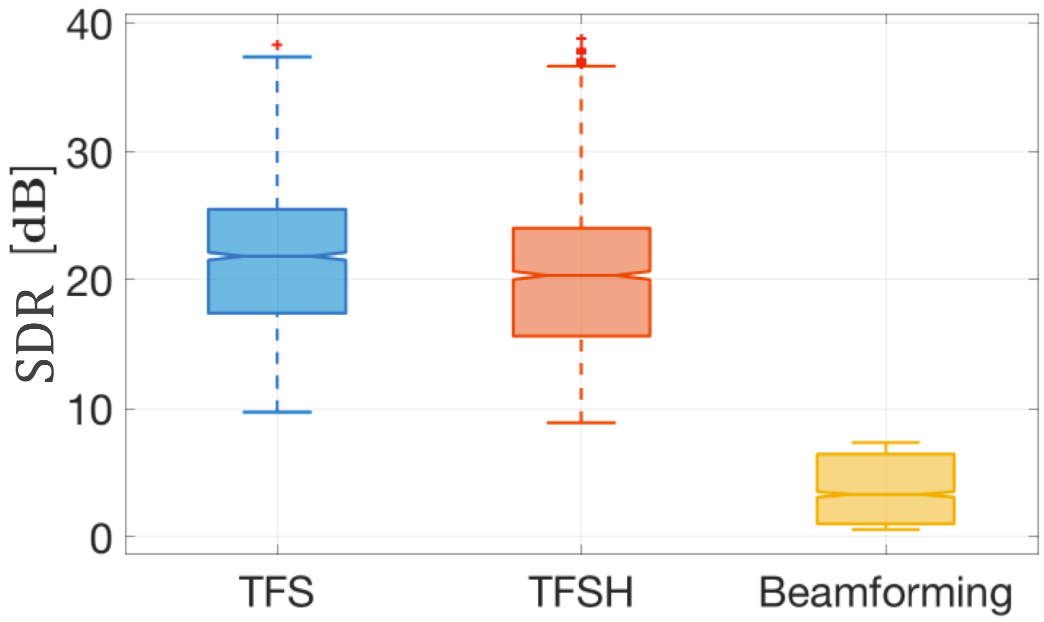


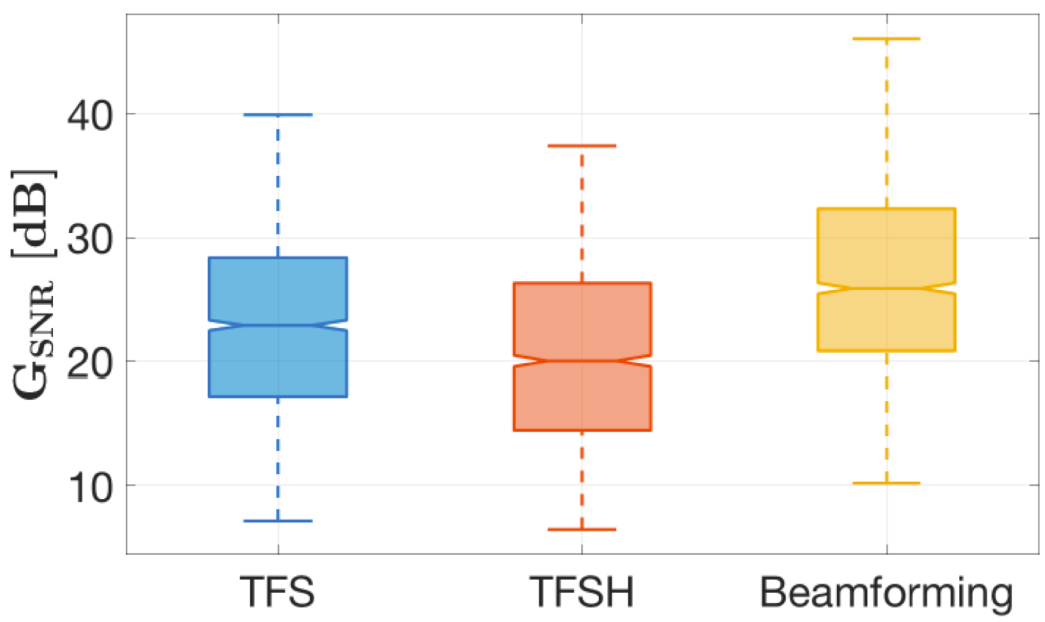
Beamforming – low-end reference



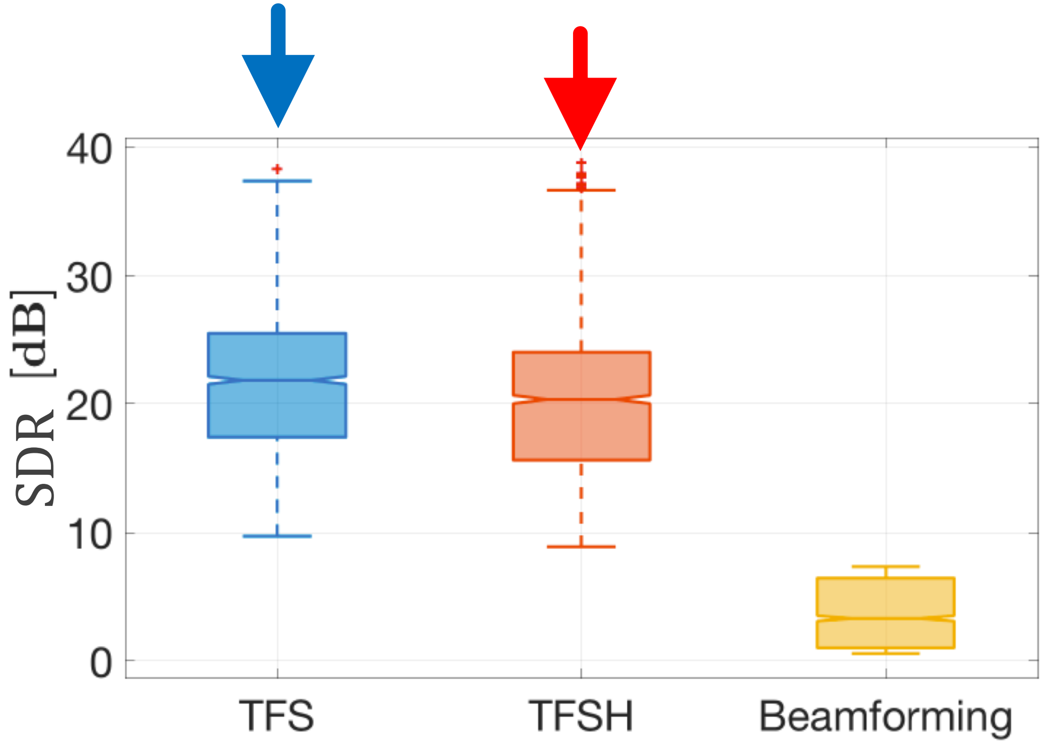
Moore, Alastair H., et al. "Binaural mask-informed speech enhancement for hearing aids with head tracking." *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2018.

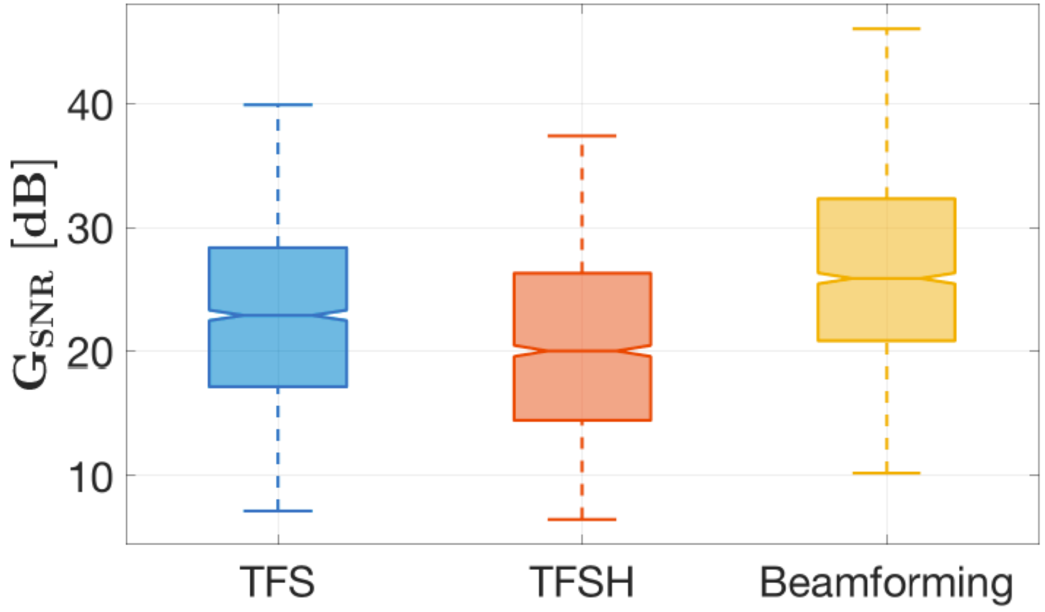
Distortion and SNR gain



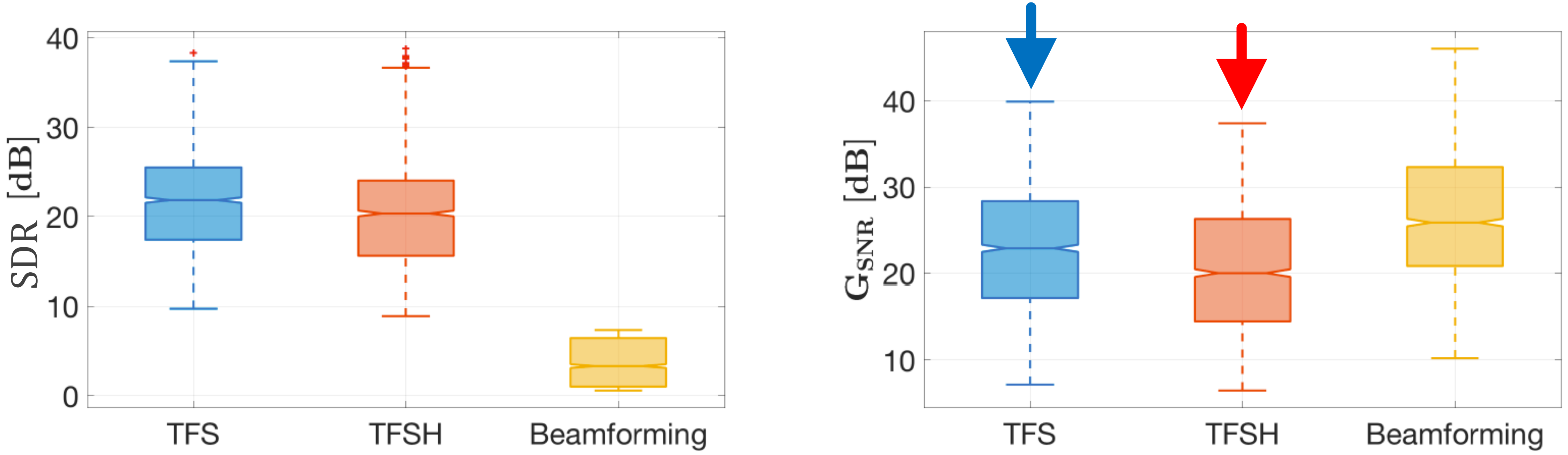


Distortion and SNR gain

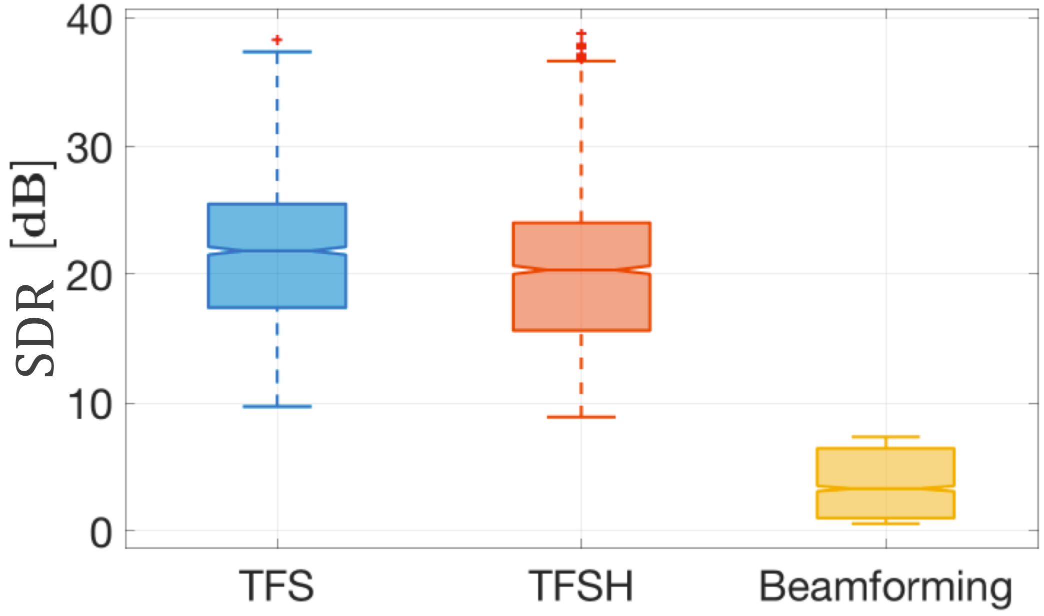


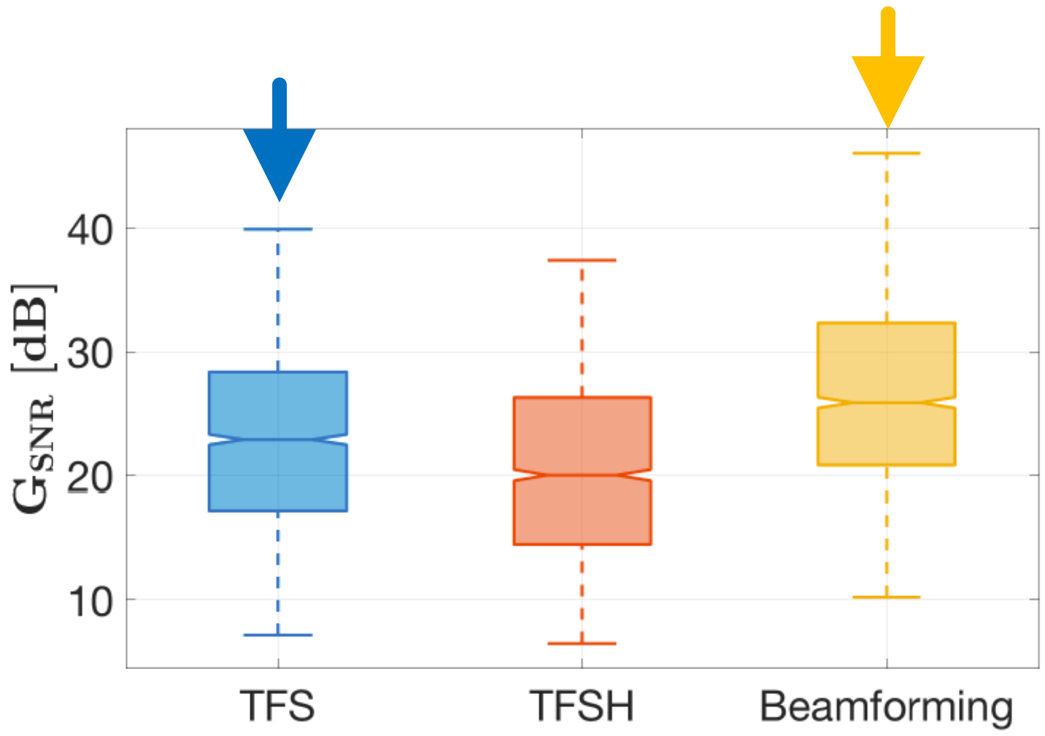


Distortion and SNR gain

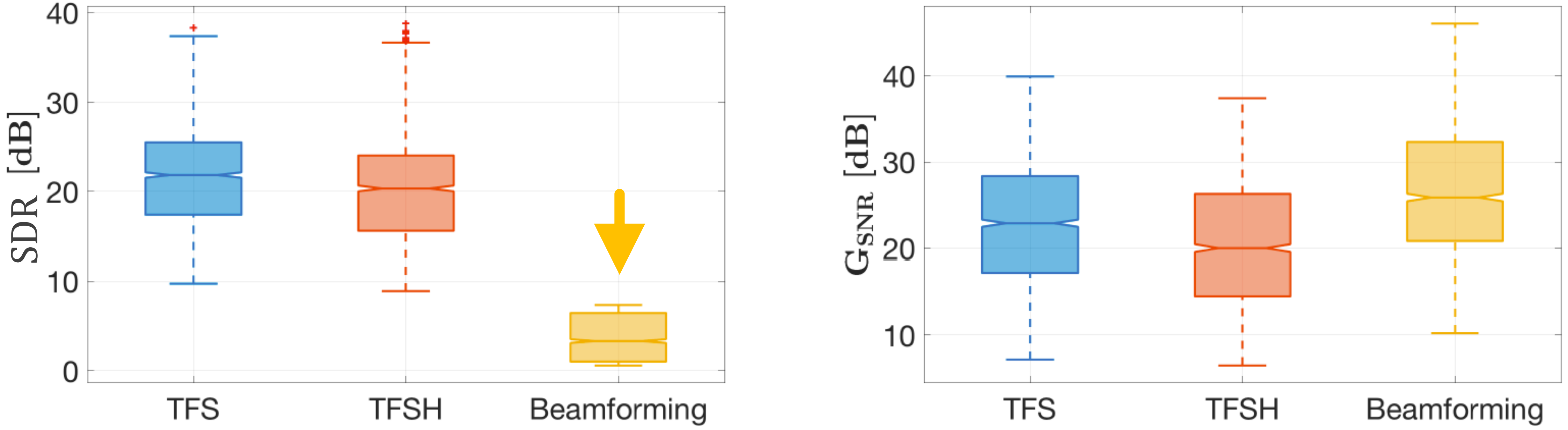


Distortion and SNR gain





Distortion and SNR gain

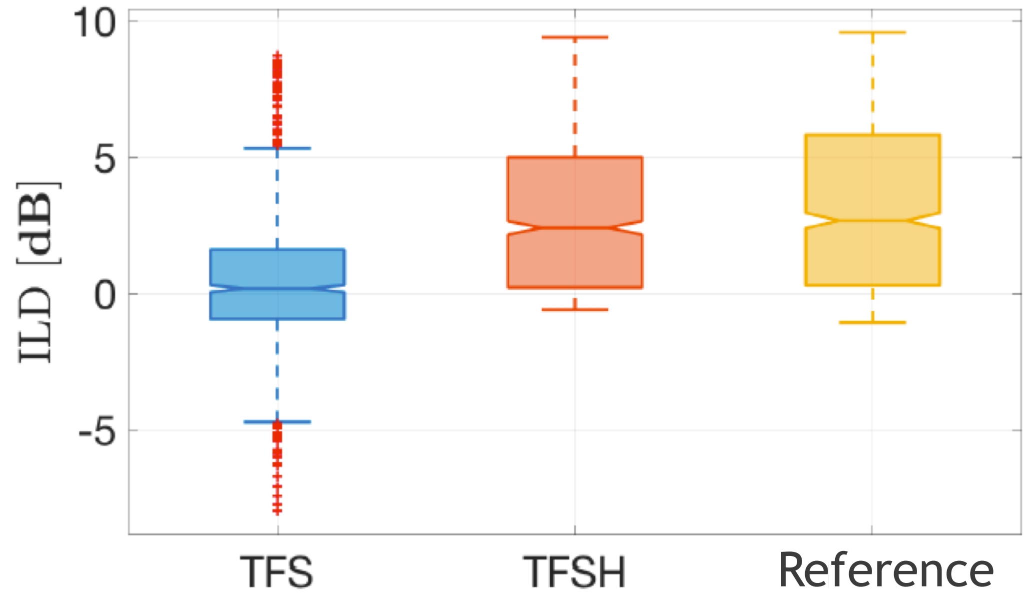


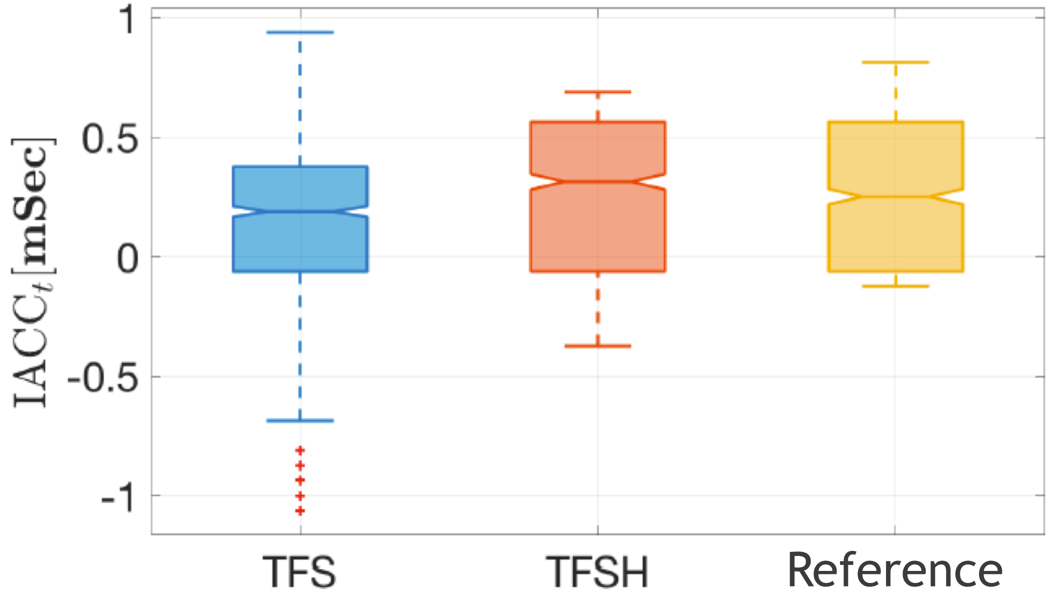
IACCt and ILD of the residual noise

- The same methods are applied only to the undesired sound field:
  - TFS
  - TFSH
  - Reference - unprocessed sound field generated by the noise source

IACCt and ILD of the residual noise

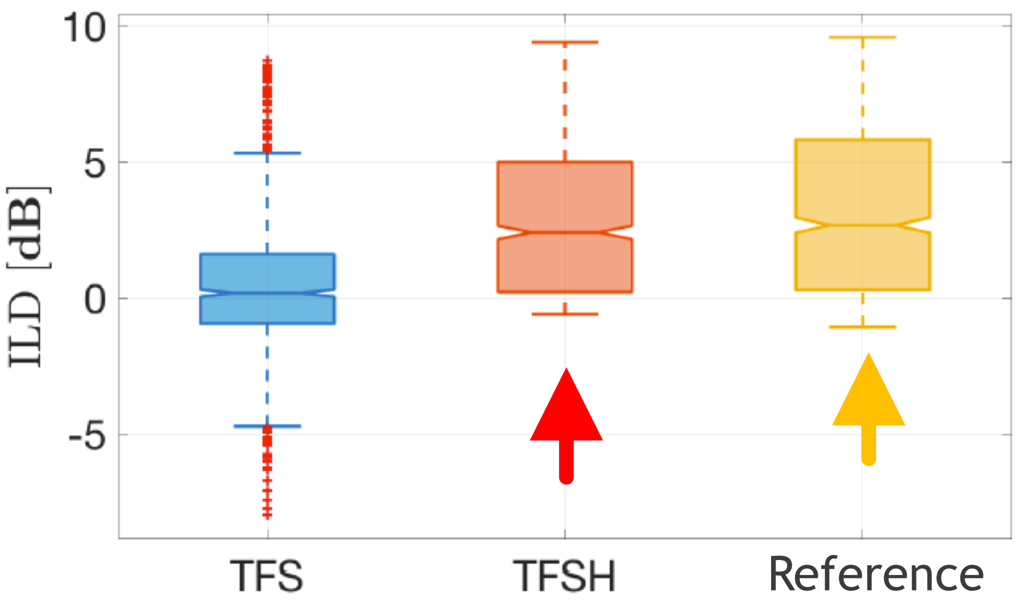
- The same methods are applied only to the undesired sound field:
  - TFS
  - TFSH
  - Reference - unprocessed sound field generated by the noise source

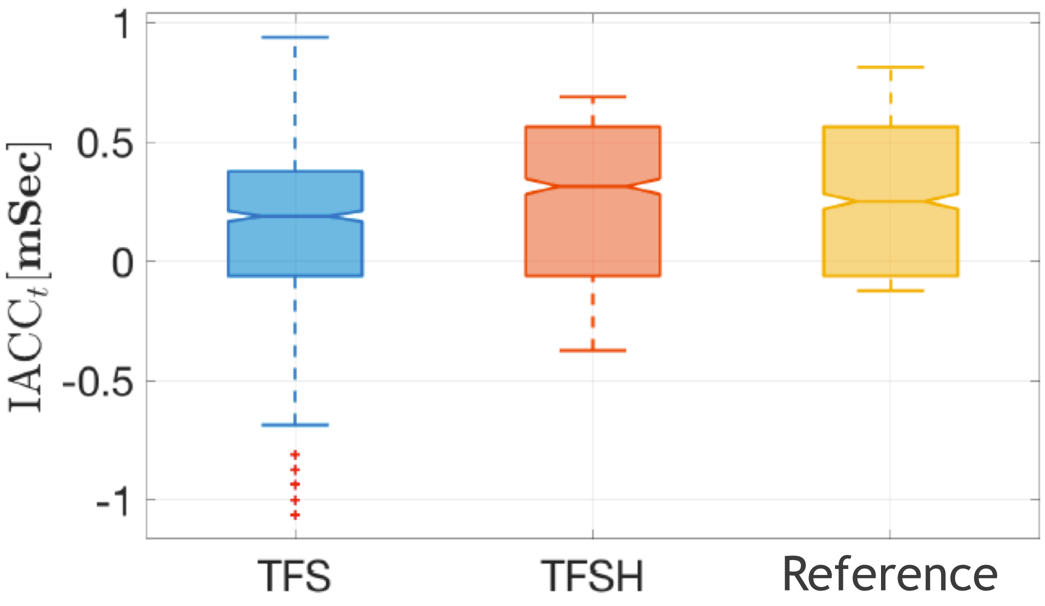




IACCt and ILD of the residual noise

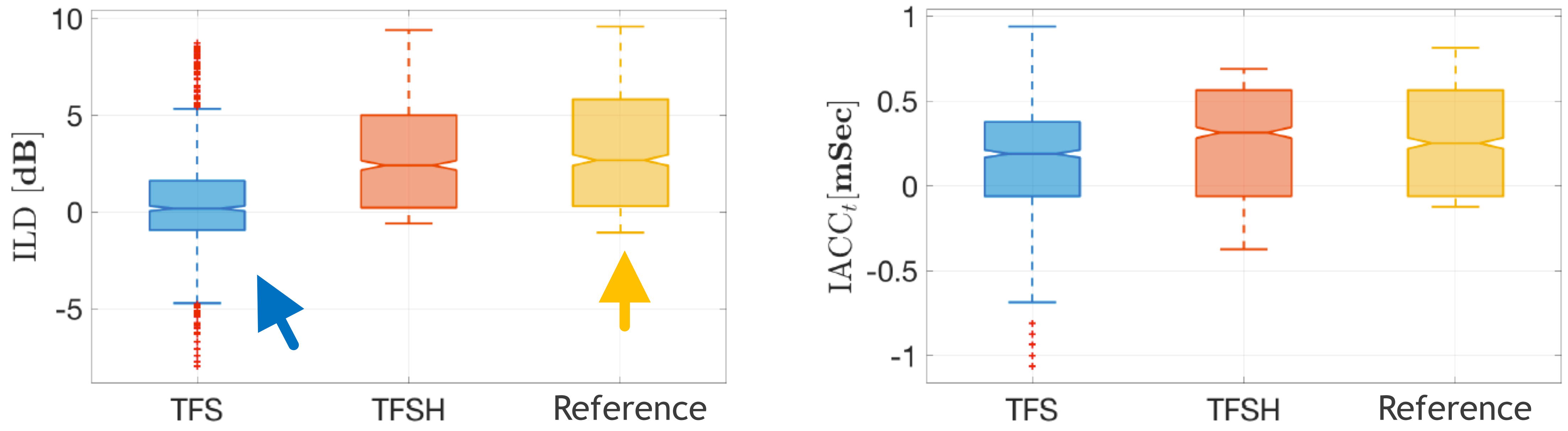
- The same methods are applied only to the undesired sound field:
  - TFS
  - TFSH
  - Reference - unprocessed sound field generated by the noise source





IACCt and ILD of the residual noise

- The same methods are applied only to the undesired sound field:
  - TFS
  - TFSH
  - Reference - unprocessed sound field generated by the noise source



Listening tests

- Two listening tests were conducted
  - Listening test 1 – overall quality
  - Listening test 2 - residual noise DOA
- The results of these tests are correlated to the objective analysis
- Details in the paper

Conclusions

- The TFS method:
  - Preserves the desired sound field better than the TFSH method
  - May change the DOA of the residual noise
- The TFSH method:
  - Preserves the DOA of the residual noise
  - The preservation of the desired sound field depends on acoustic parameters

Thank you!