

# Optimization of Personal Sound Zones with Spatial Audio

**Yue Qiao\*** & Edgar Choueiri

3D Audio and Applied Acoustics (3D3A) Lab

Princeton University

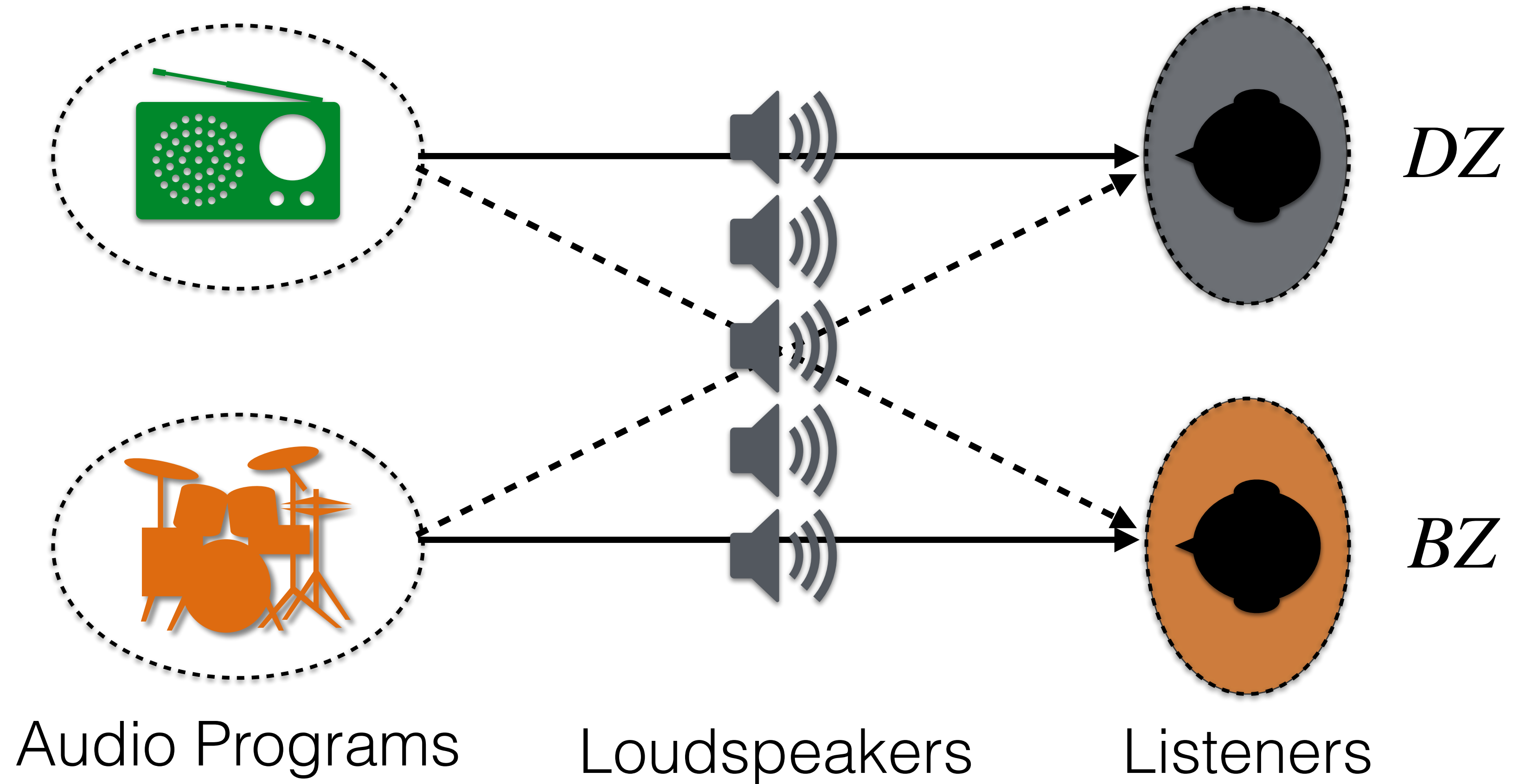
Presented at the AES AVAR Conference

Aug 16, 2022

# Key questions

1. How to render spatial audio in personal sound zones?
2. How to optimize both aspects in a single problem?

# Concept of PSZ<sup>[1]</sup>



# Example use cases



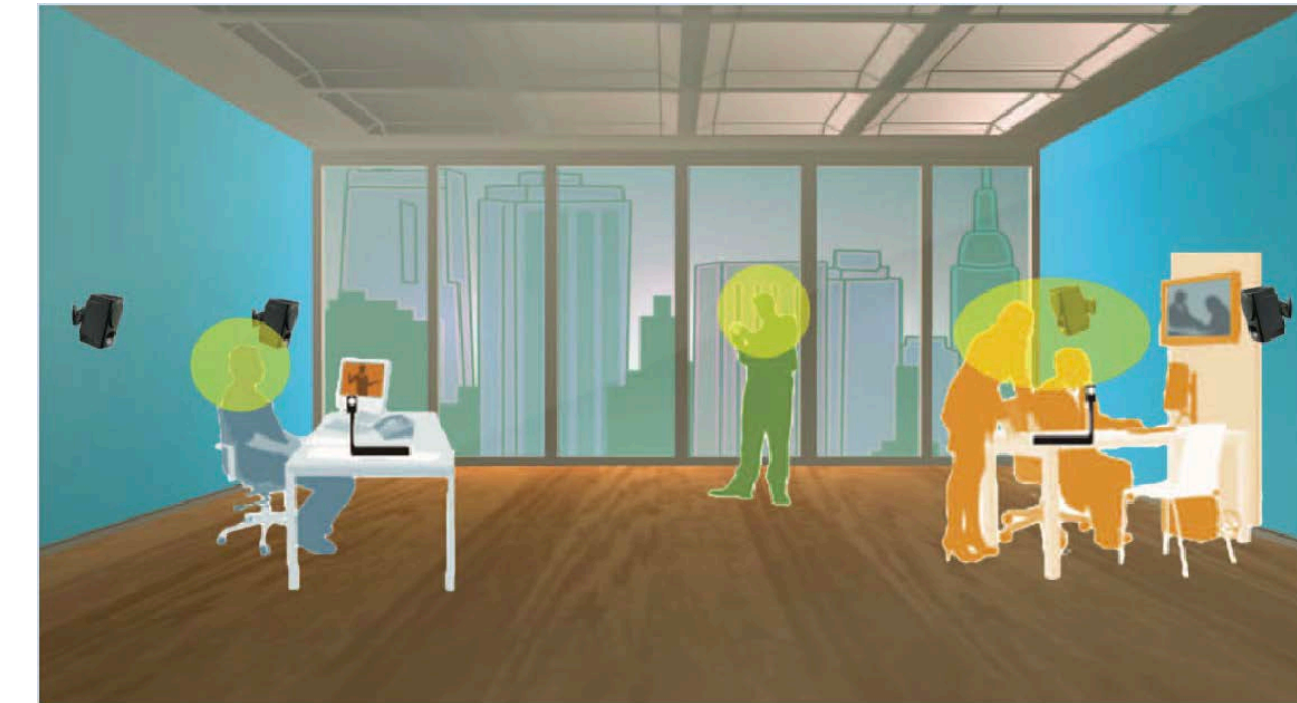
Cho and Chang, ICA, 2019



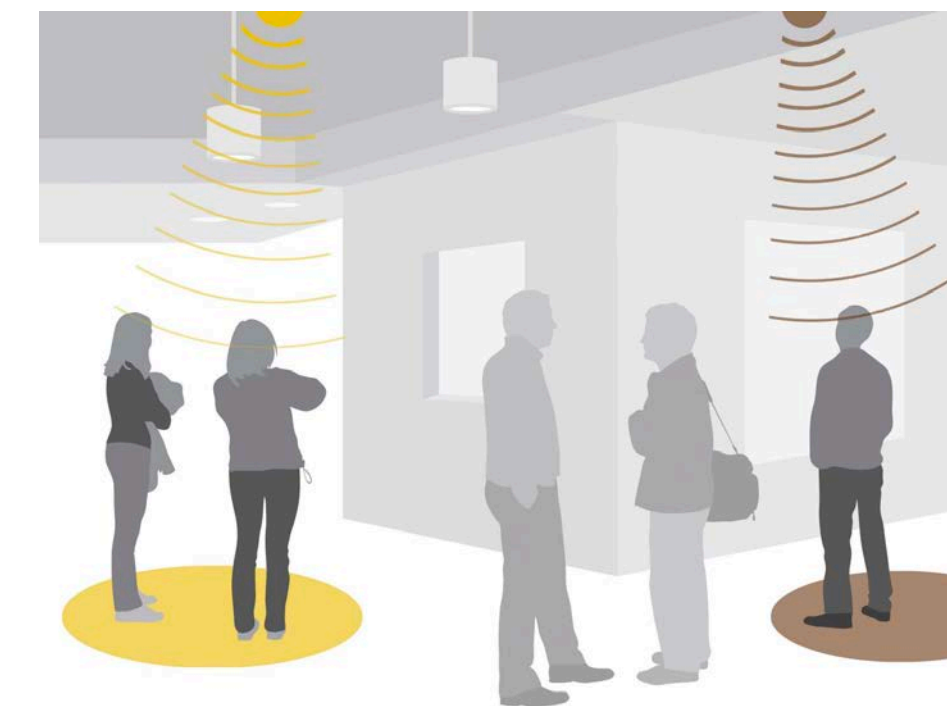
Ebri et al., AES Conv., 2020



Vindrola et al., JASA, 2021



Betlehem et al., IEEE Sig. Proc., 2015



Fraunhofer IDMT

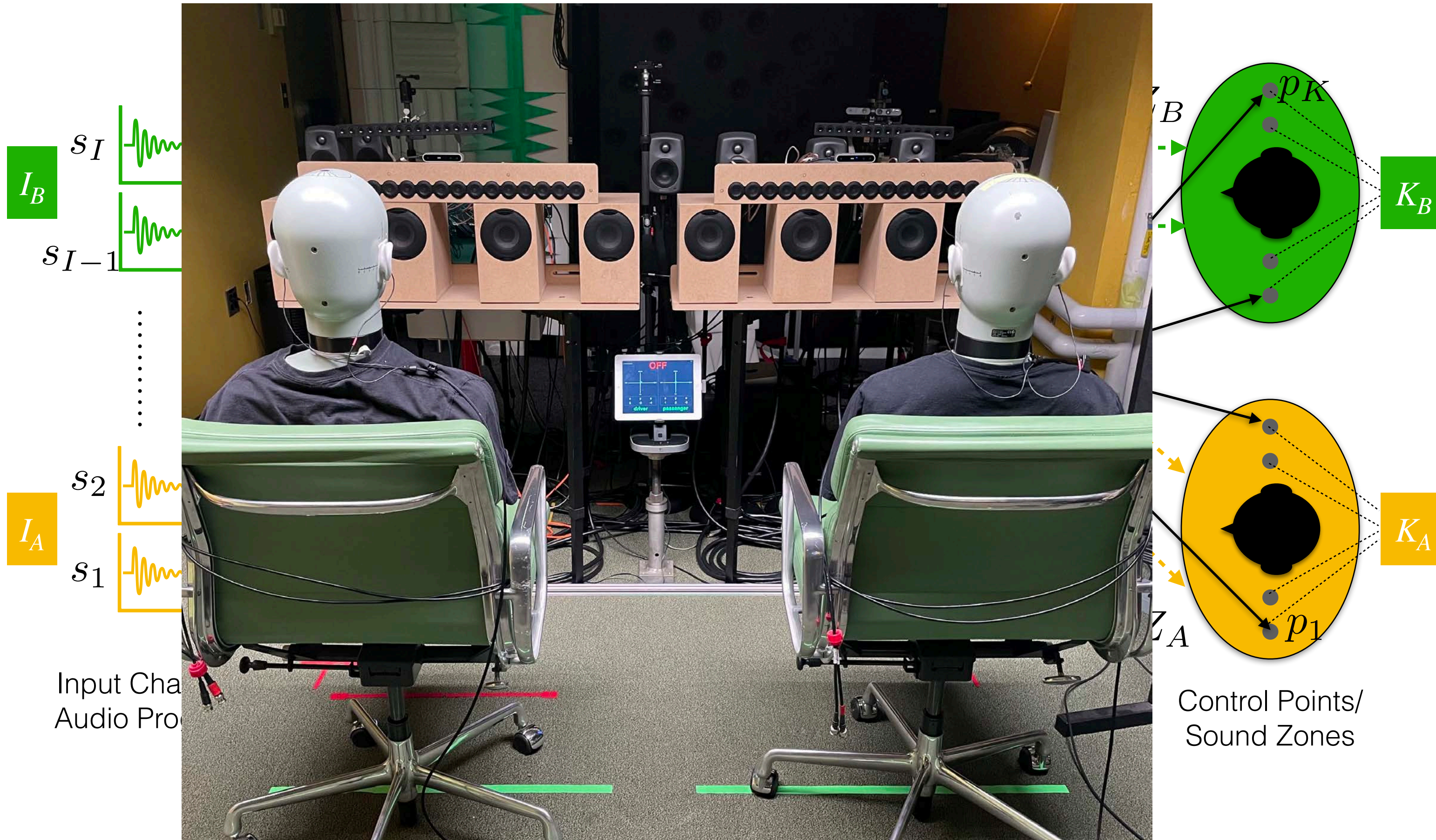
**Automotive cabins**

**Shared open space**

# PSZ with spatial audio

- Natural extension of crosstalk cancellation
  - Single program/listener -> multiple programs/listeners
  - Inter-aural Cancellation -> Inter-zone cancellation
- New medium for VR/AR applications
  - Headphone-free immersive experience
  - Head-externalized binaural reproduction
  - Transparent communication/interaction
  - Independent experience in shared space

# Designing a PSZ system



Pressure Matching (PM)<sup>[2]</sup>

$$\mathbf{g}^* = \arg \min_{\mathbf{g}} \|\mathbf{p}_T - \mathbf{H} \cdot \mathbf{g}\|^2$$

Acoustic Contrast Control (ACC)<sup>[3]</sup>

$$\mathbf{g}^* = \arg \max_{\mathbf{g}} \frac{\|\mathbf{H}_B \cdot \mathbf{g}\|^2}{\|\mathbf{H}_D \cdot \mathbf{g}\|^2}$$

No control over phase  
Not suitable for binaural audio

Rendering spatial audio for two listeners with PM

# Specifying target pressure

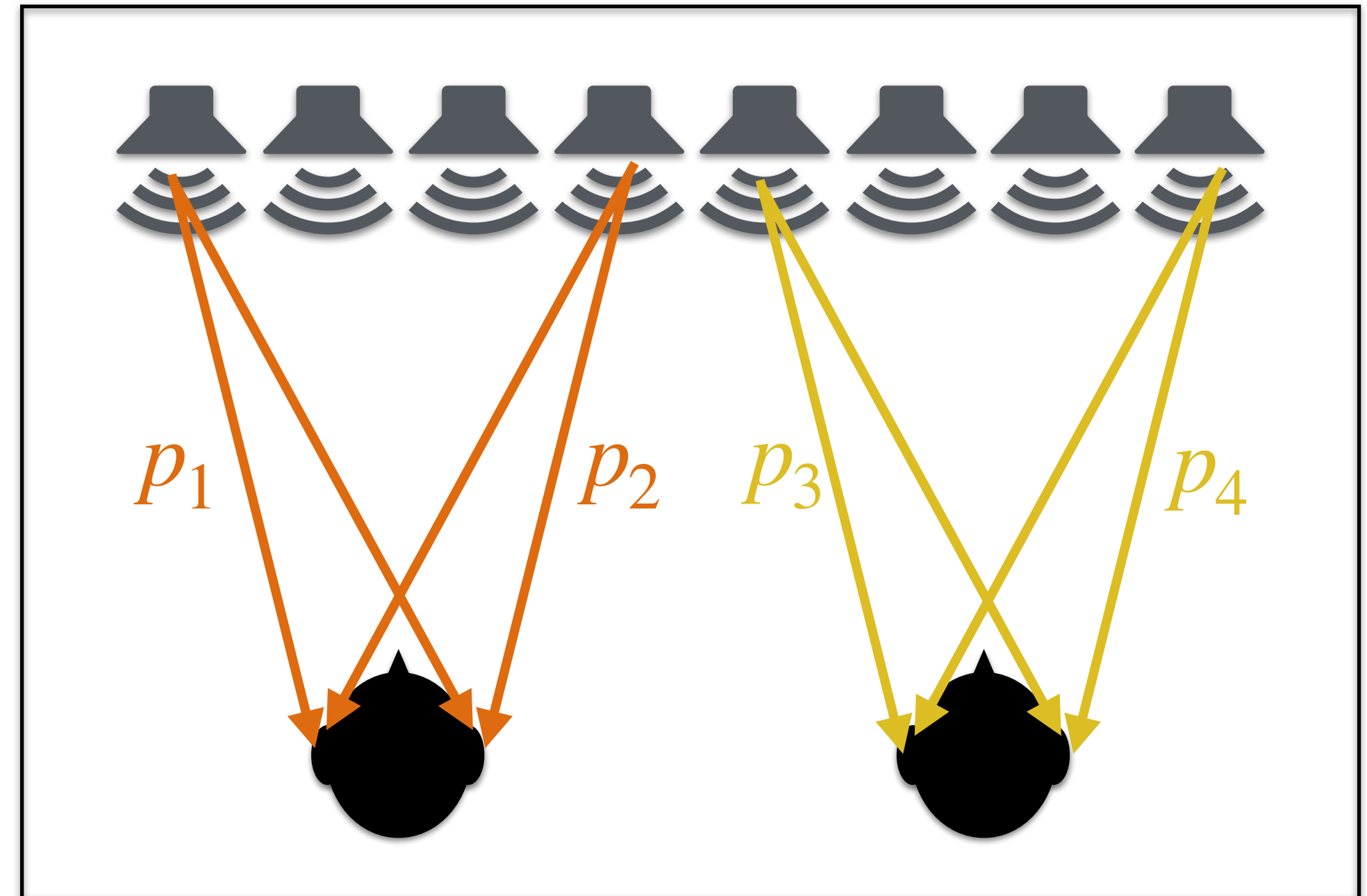
Pressure Matching (PM)<sup>[2]</sup>

$$\mathbf{g}^* = \arg \min_{\mathbf{g}} \|\mathbf{p}_T - \mathbf{H} \cdot \mathbf{g}\|^2$$

Mono programs as input

2 input channels  $\longrightarrow$  2  $\mathbf{p}_T$  vectors

$$\mathbf{p}_{T,1} = \begin{bmatrix} p_1 \\ p_2 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{p}_{T,2} = \begin{bmatrix} 0 \\ 0 \\ p_3 \\ p_4 \end{bmatrix}$$





# Specifying target pressure

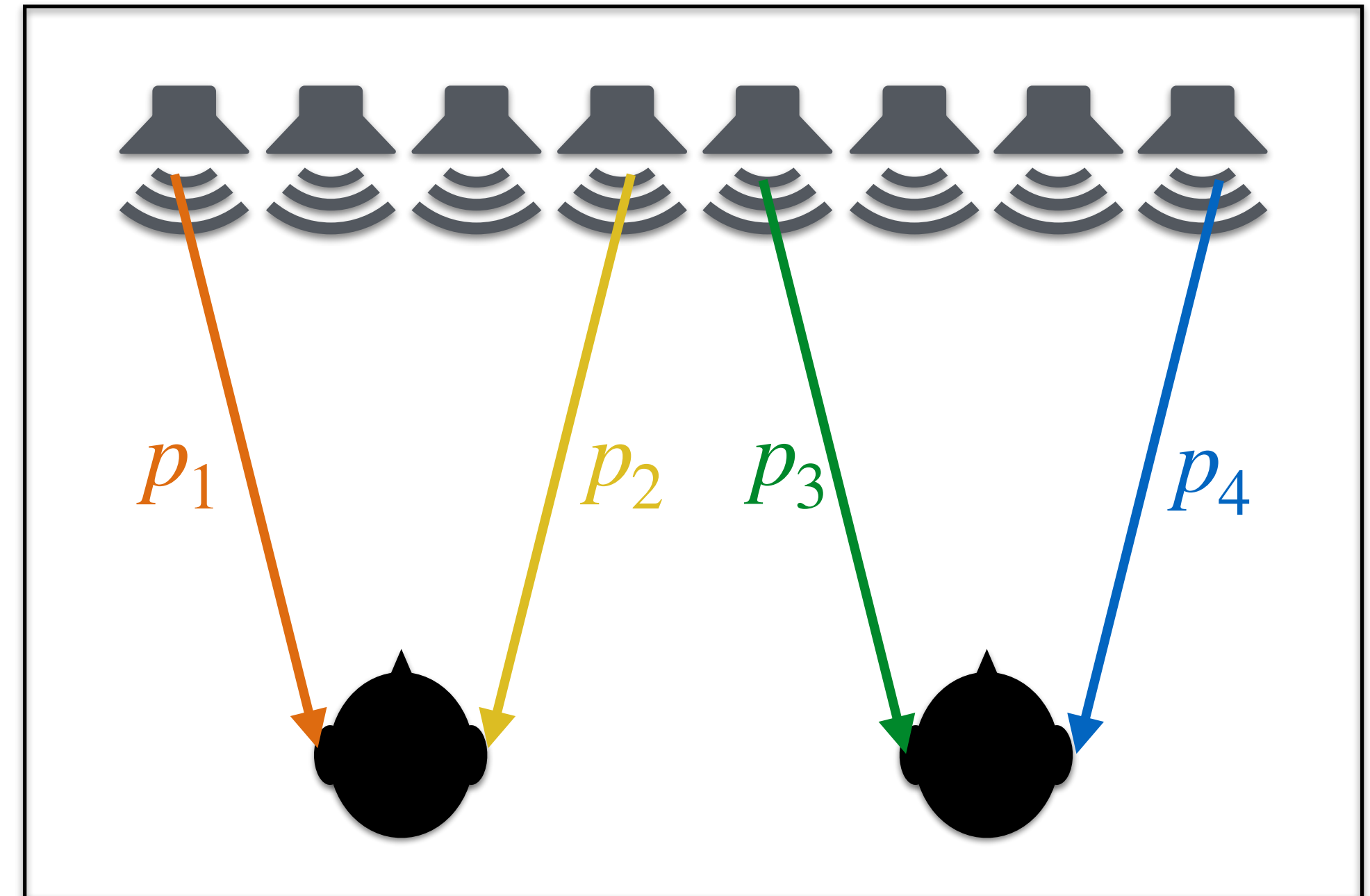
Pressure Matching (PM)<sup>[2]</sup>

$$\mathbf{g}^* = \arg \min_{\mathbf{g}} \|\mathbf{p}_T - \mathbf{H} \cdot \mathbf{g}\|^2$$

Binaural programs as input

4 input channels  $\longrightarrow$  4  $\mathbf{p}_T$  vectors

$$\mathbf{p}_{T,1} = \begin{bmatrix} p_1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{p}_{T,2} = \begin{bmatrix} 0 \\ p_2 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{p}_{T,3} = \begin{bmatrix} 0 \\ 0 \\ p_3 \\ 0 \end{bmatrix} \quad \mathbf{p}_{T,4} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ p_4 \end{bmatrix}$$



# Performance metrics

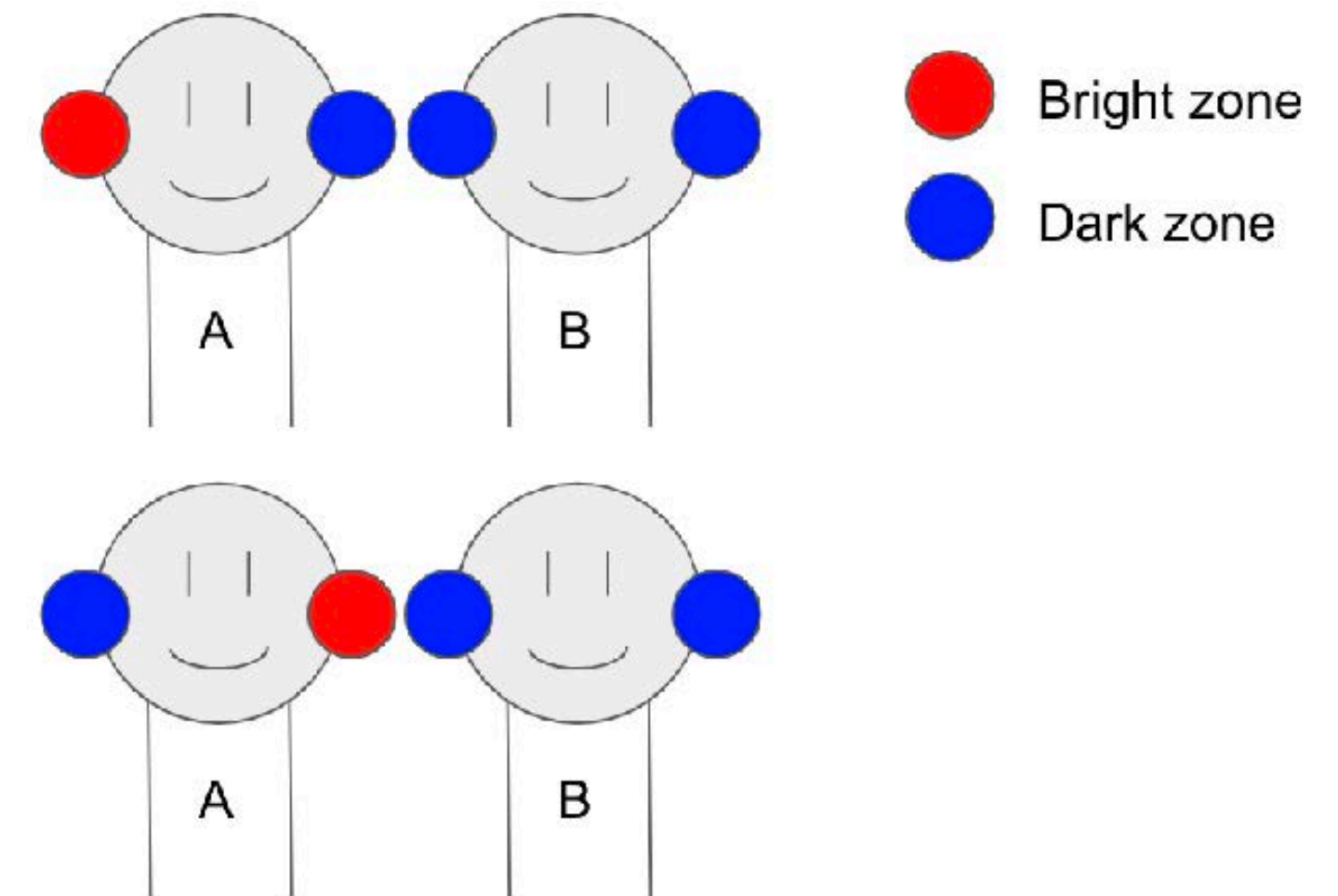
Two aspects of isolation performance:

## 1. Isolation between listeners

- Acoustic Contrast (AC)
- poor performance -> distraction by other audio

## 2. Isolation between ears

- Crosstalk Cancellation (XTC)
- poor performance -> lack of envelopment



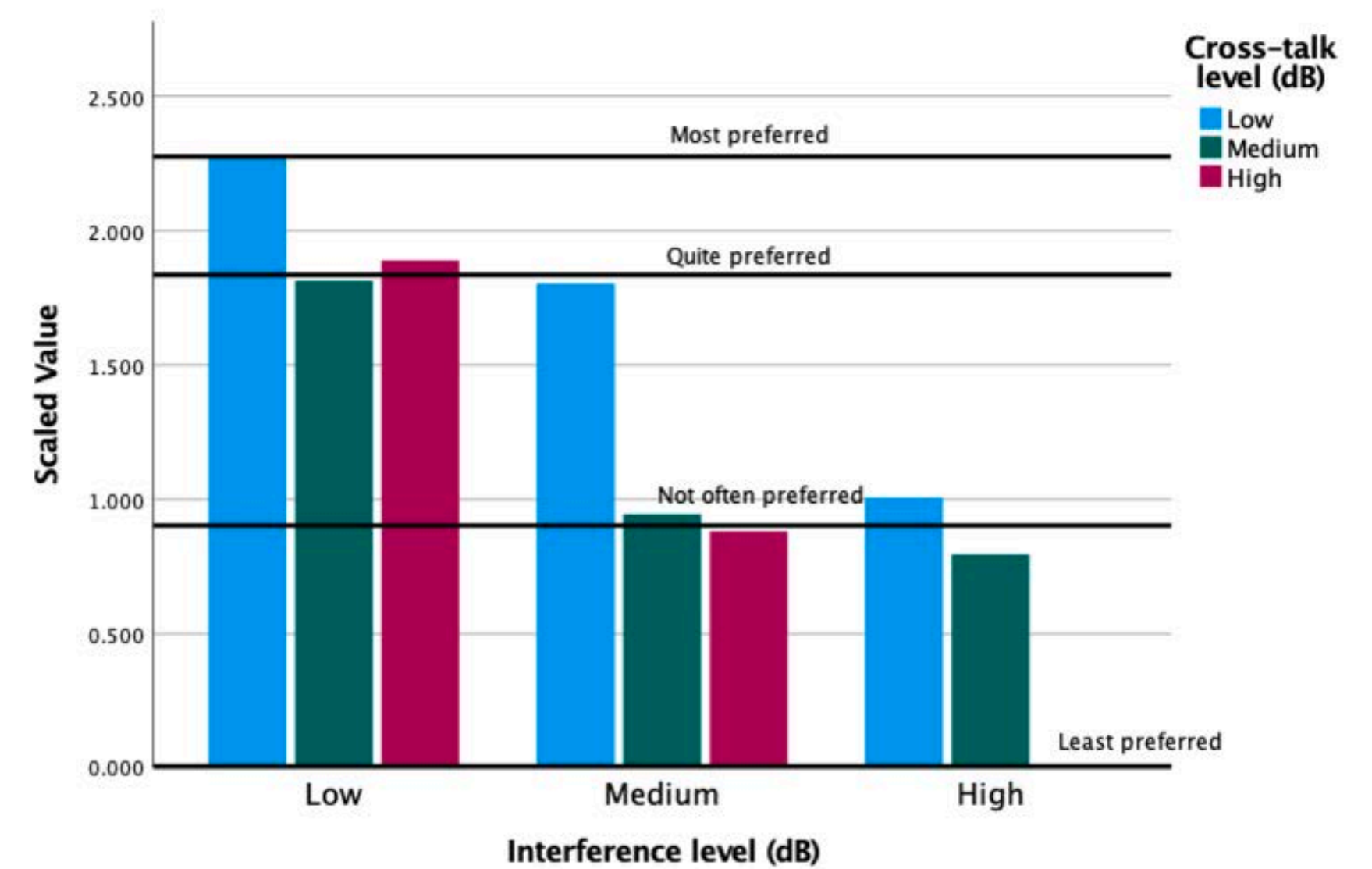
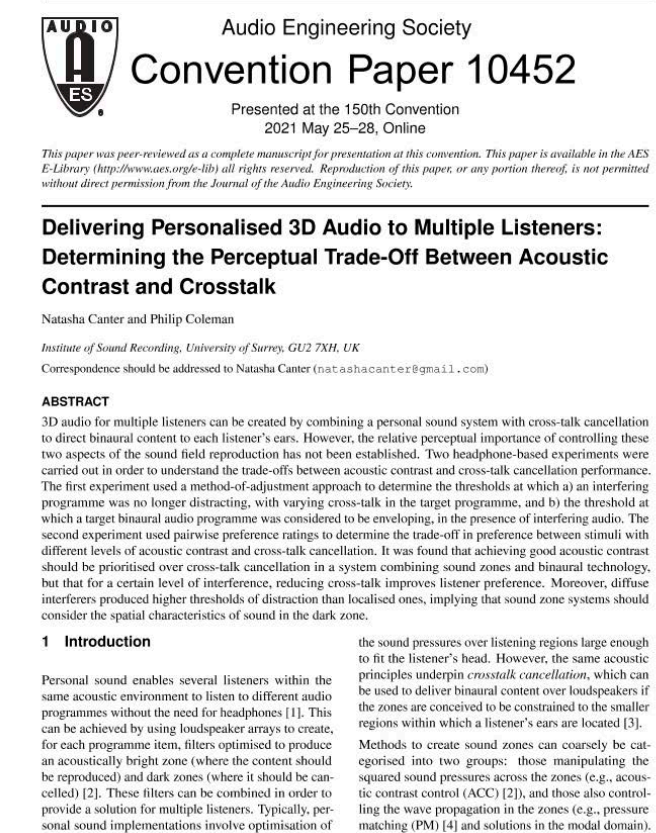
Canter and Coleman, AES Conv., 2021

Both aspects matter!

But which one is perceptually more important?

# Perceptual trade-offs between AC and XTC

- Headphone-based subjective experiments
- Manually adjusted interference & crosstalk levels
- Stimuli: pop (+film); classical (+pop); film (+pop)
- Main takeaways
  - Interference and crosstalk are perceptually uncorrelated
  - Program combination affects the interference threshold
  - AC should be prioritized over XTC when both are present

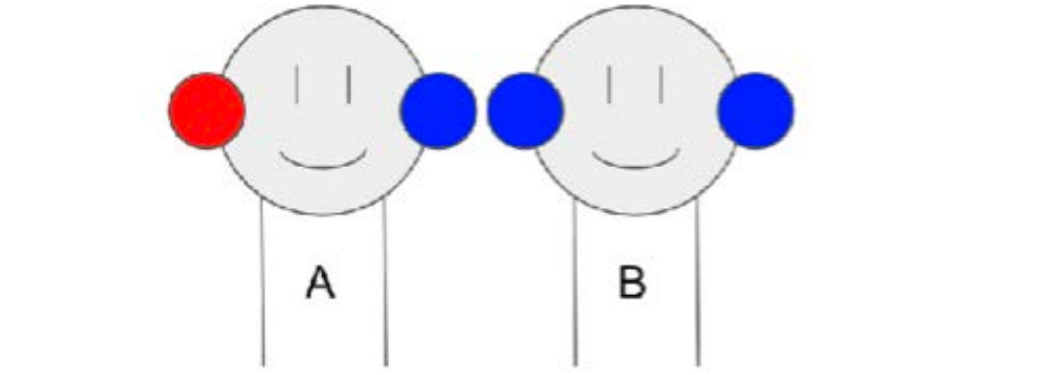


Canter and Coleman, AES Conv., 2021

Finding the optimal trade-off between AC and XTC

Revisit the PM cost function

$$J = \|\mathbf{H}\mathbf{g} - \mathbf{p}_T\|^2$$



$$\mathbf{p}_T = \begin{bmatrix} p_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$J = \underbrace{\|\mathbf{h}_{B2}^H \mathbf{g}\|^2}_{\text{XTC}} + \underbrace{\|\mathbf{H}_D \mathbf{g}\|^2}_{\text{AC}} + \underbrace{\|\mathbf{h}_{B1}^H \mathbf{g} - p_1\|^2}_{\text{Reproduction Error}}$$

Both aspects are treated with same priority

Add a weighting parameter  $\alpha$  to control the priority

$$J = \alpha \|\mathbf{h}_{B2}^H \mathbf{g}\|^2 + \|\mathbf{H}_D \mathbf{g}\|^2 + \|\mathbf{h}_{B1}^H \mathbf{g} - p_1\|^2$$

$\alpha$ 

$\left\{ \begin{array}{l} \text{AC} \\ \text{XTC} \end{array} \right.$

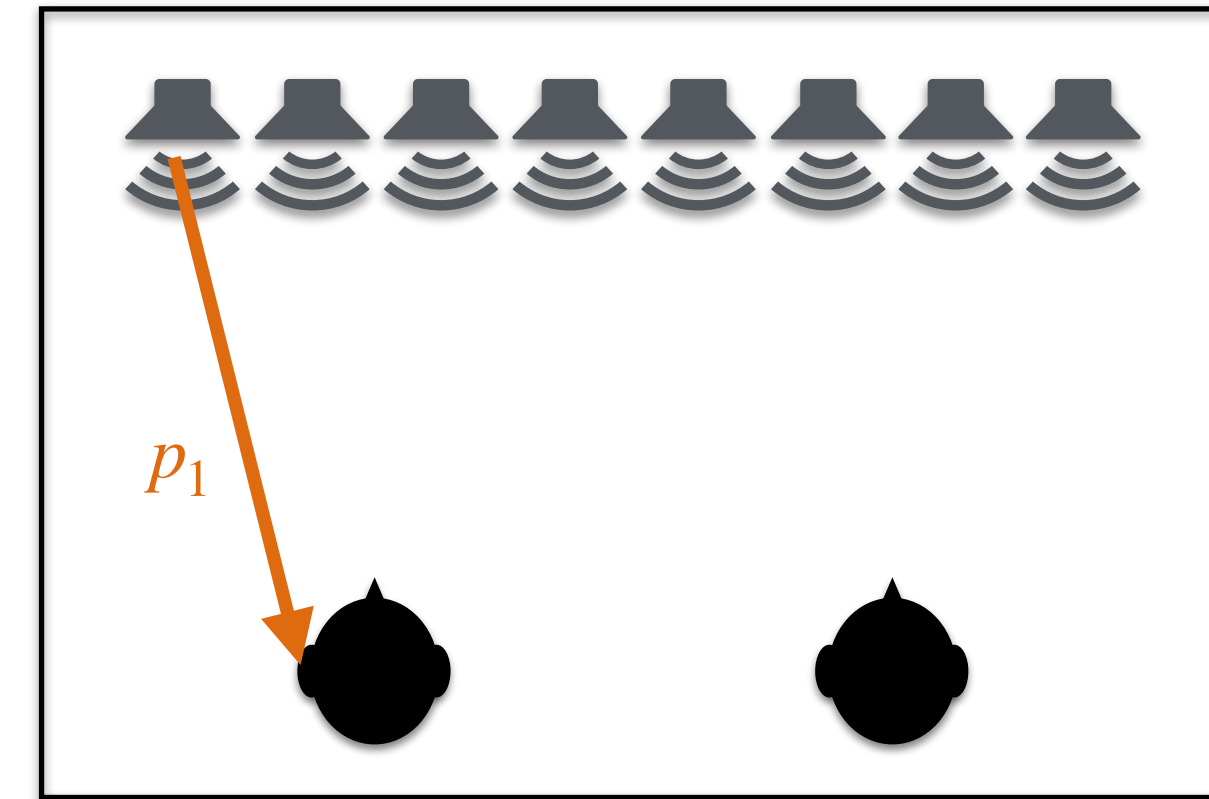
Less distraction

Less envelopment

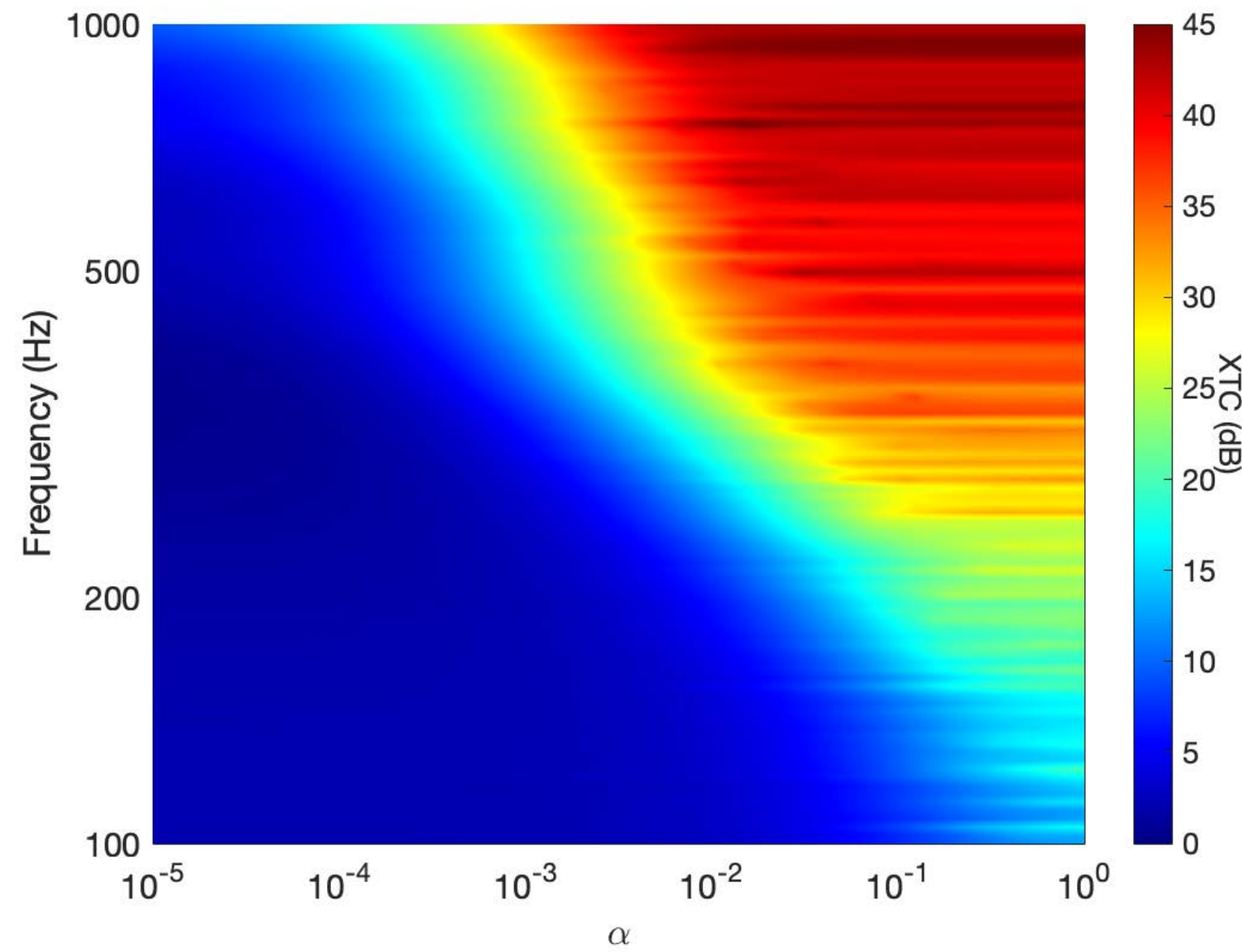
# Simulated examples

- Free-field condition with point sources
- randomly perturbed transfer functions
- Constant regularization @ 100-1000 Hz

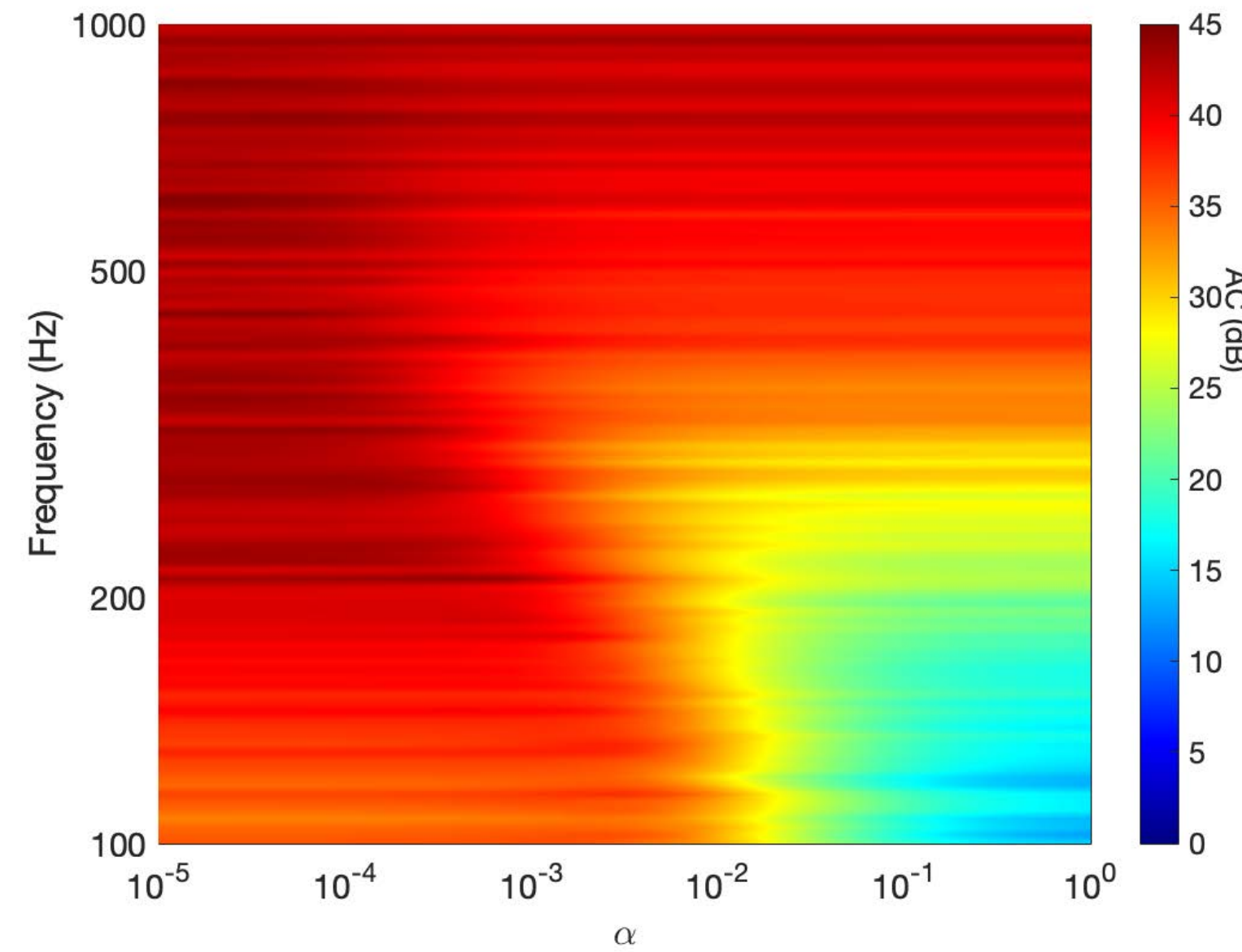
$$\mathbf{p}_T = \begin{bmatrix} p_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



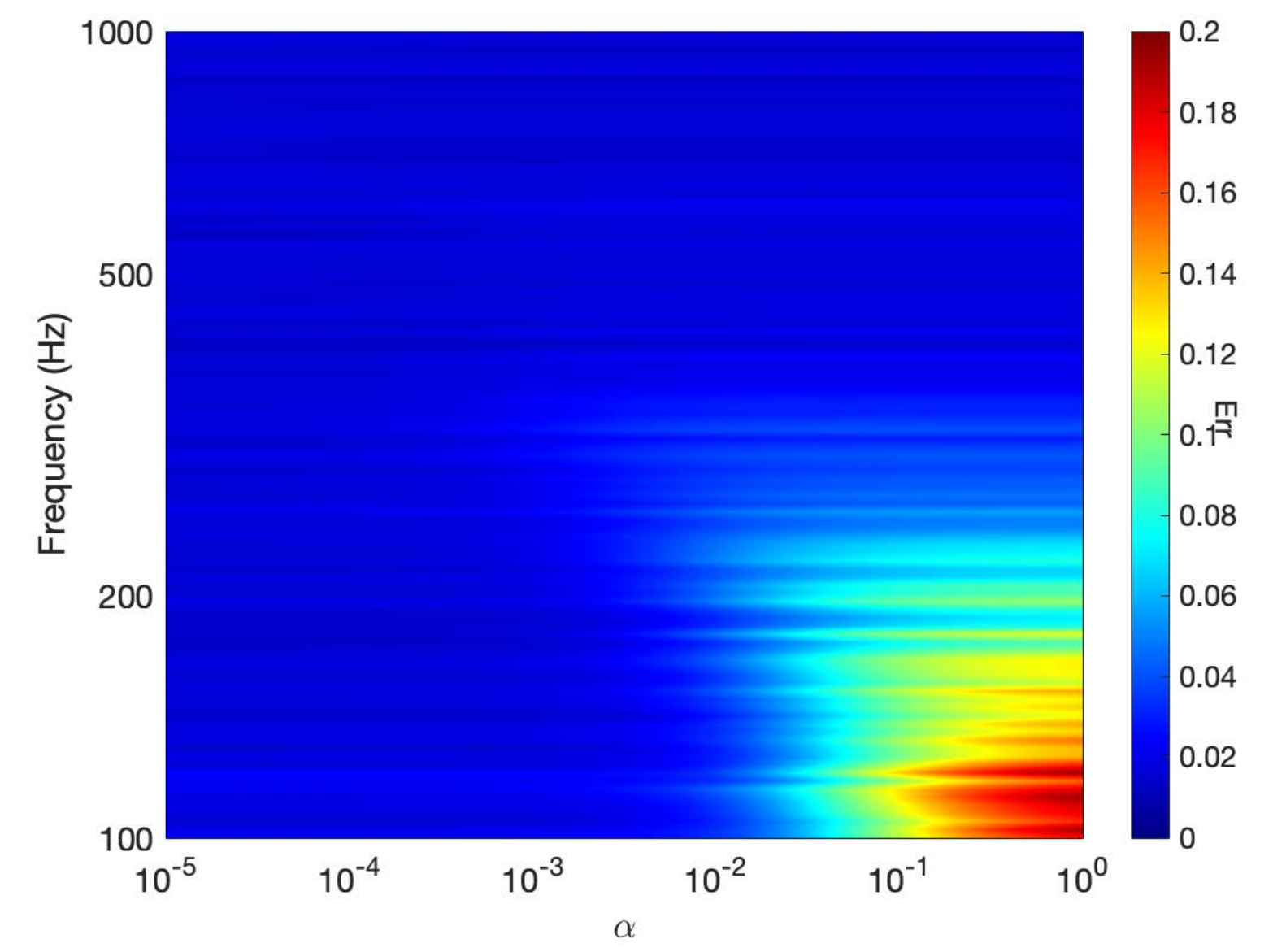
Crosstalk Cancellation (XTC)



Acoustic Contrast (AC)



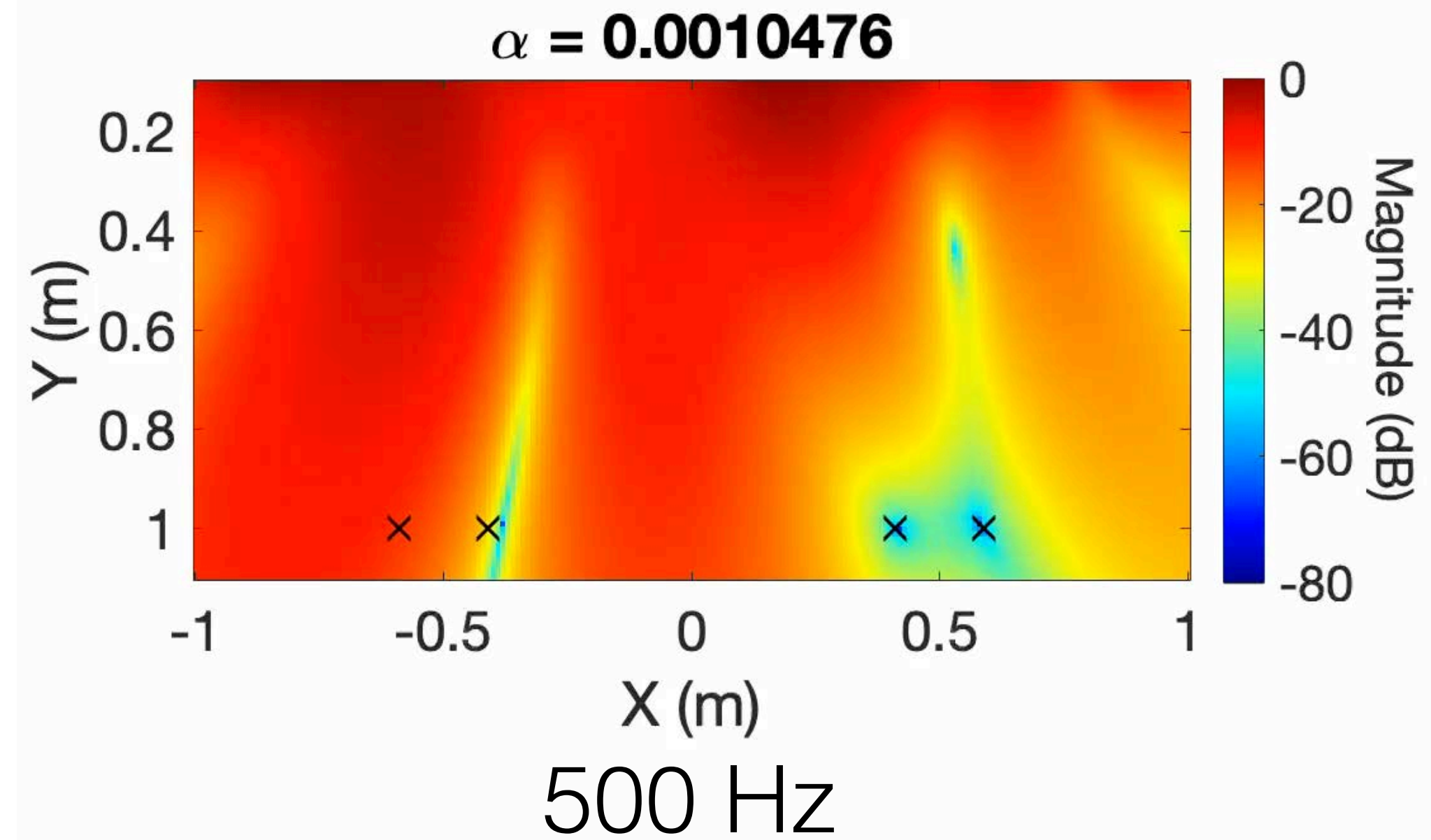
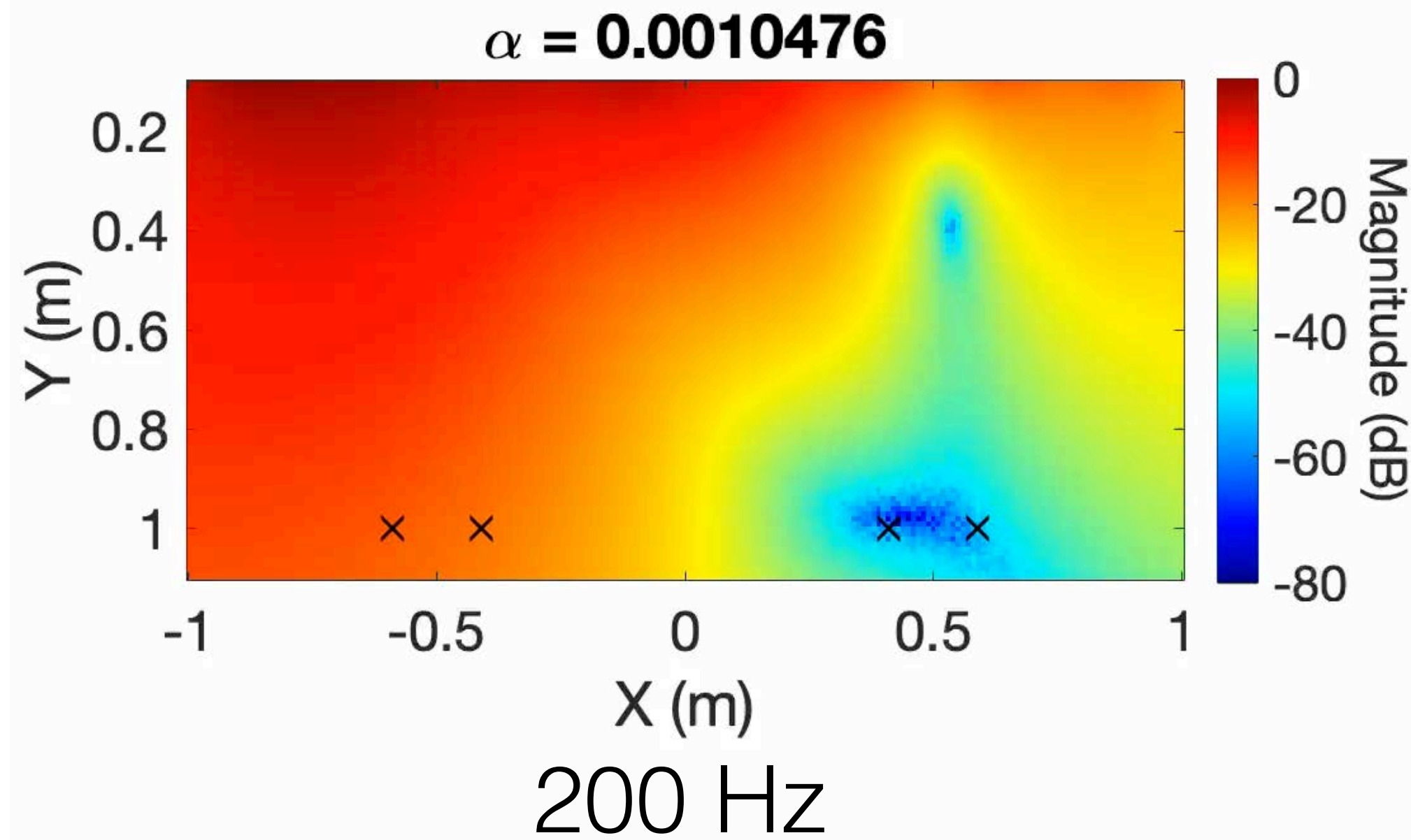
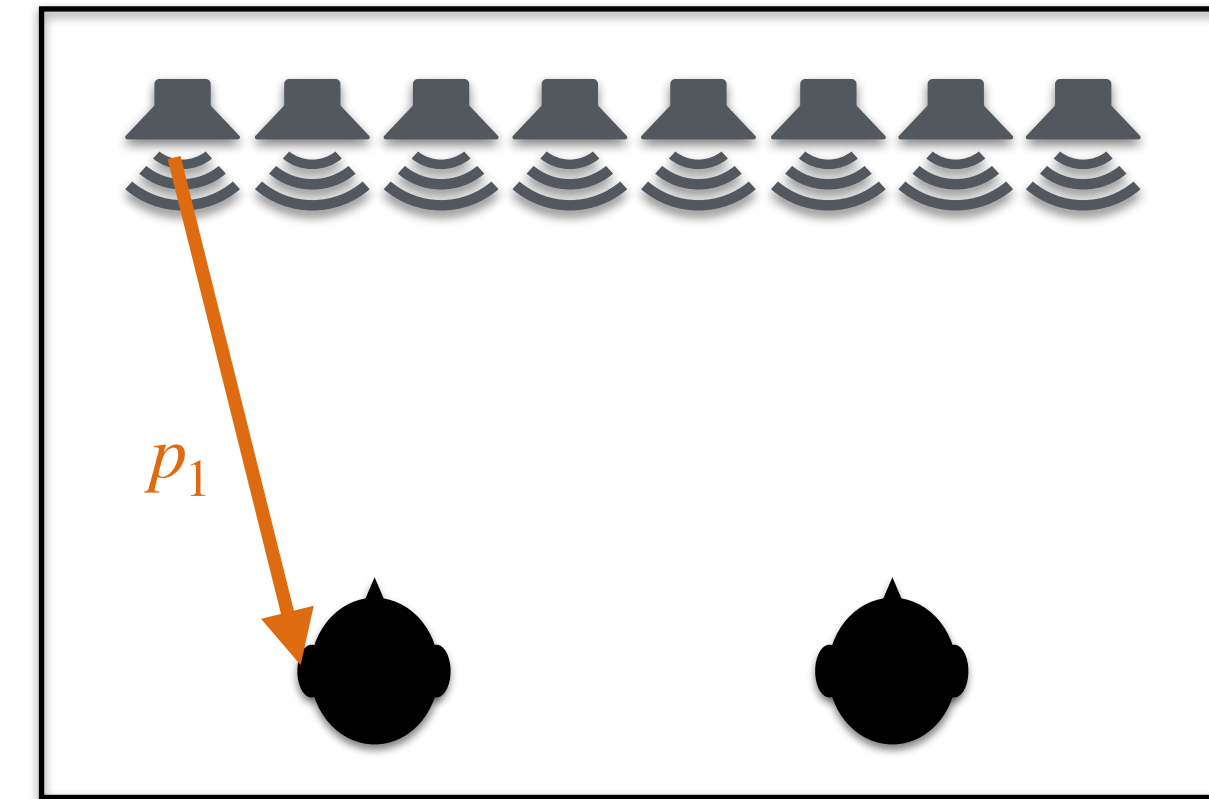
Normalized Reproduction Error (Err)



# Simulated examples

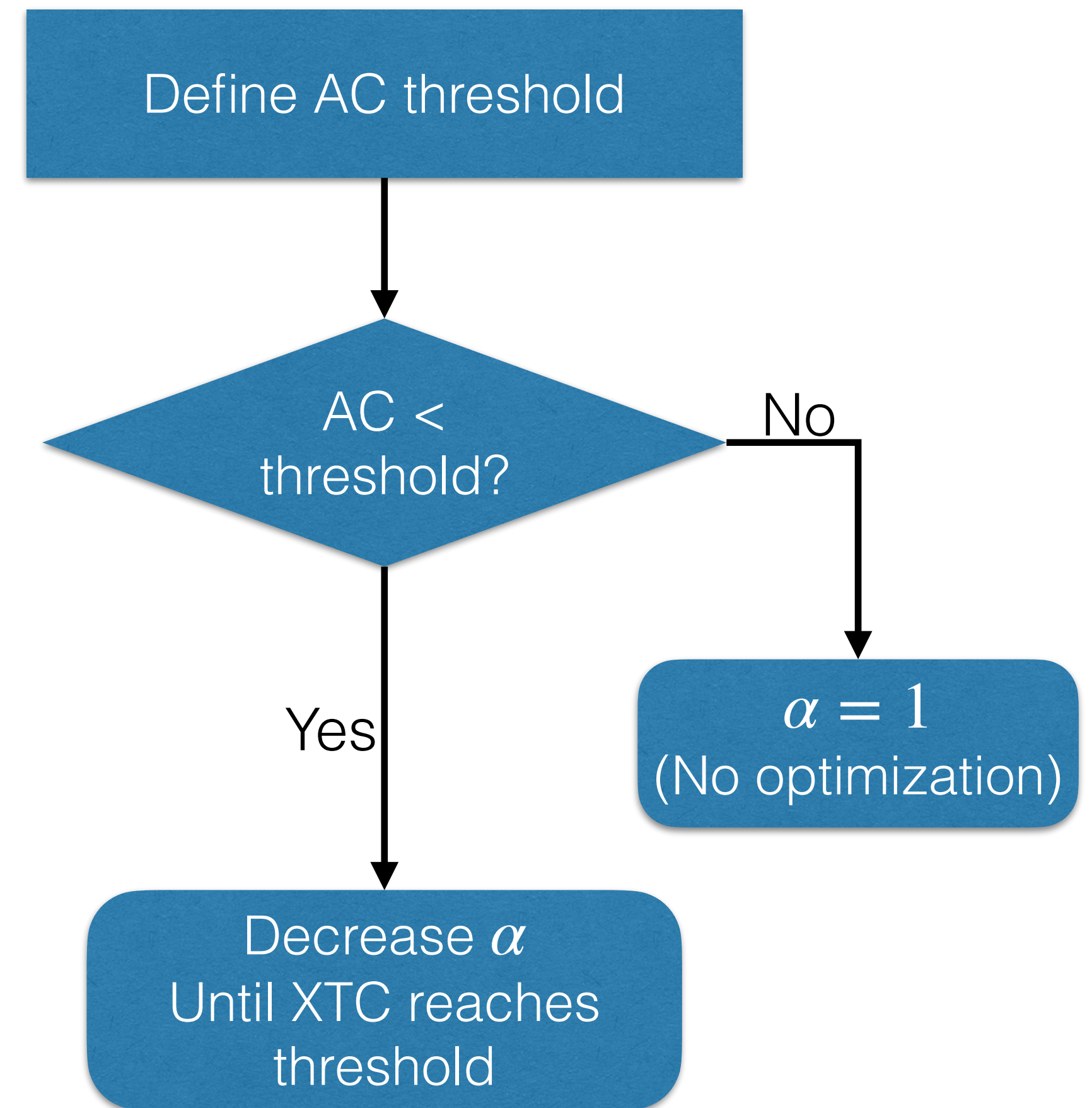
- Free-field condition with point sources
- randomly perturbed transfer functions
- Constant regularization @ 100-1000 Hz

$$\mathbf{p}_T = \begin{bmatrix} p_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



# Optimizing the trade-off

- Principle: trading off XTC for higher AC
- Observations
  - AC and Err are mostly affected by  $\alpha$  at low frequencies
  - XTC is affected at almost all frequencies



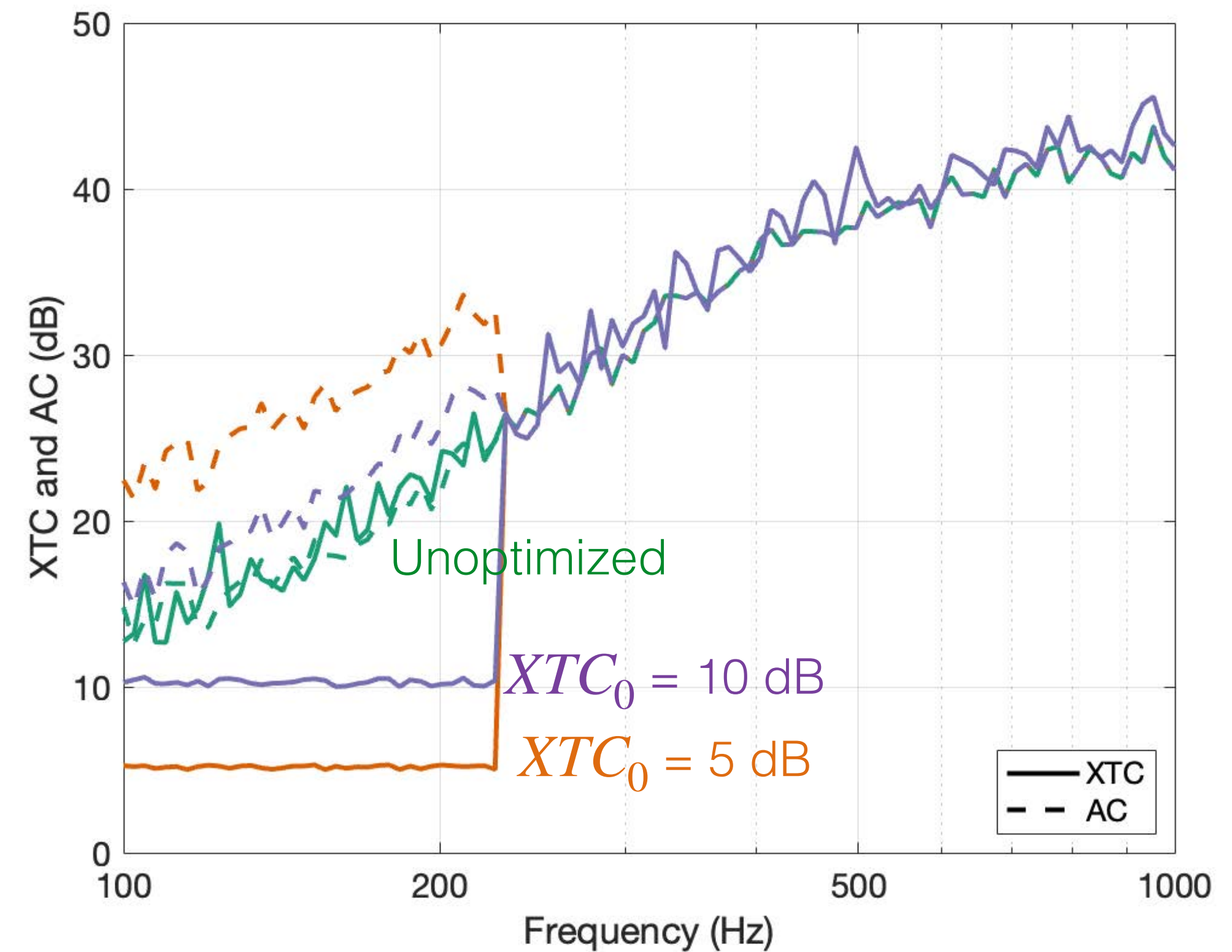
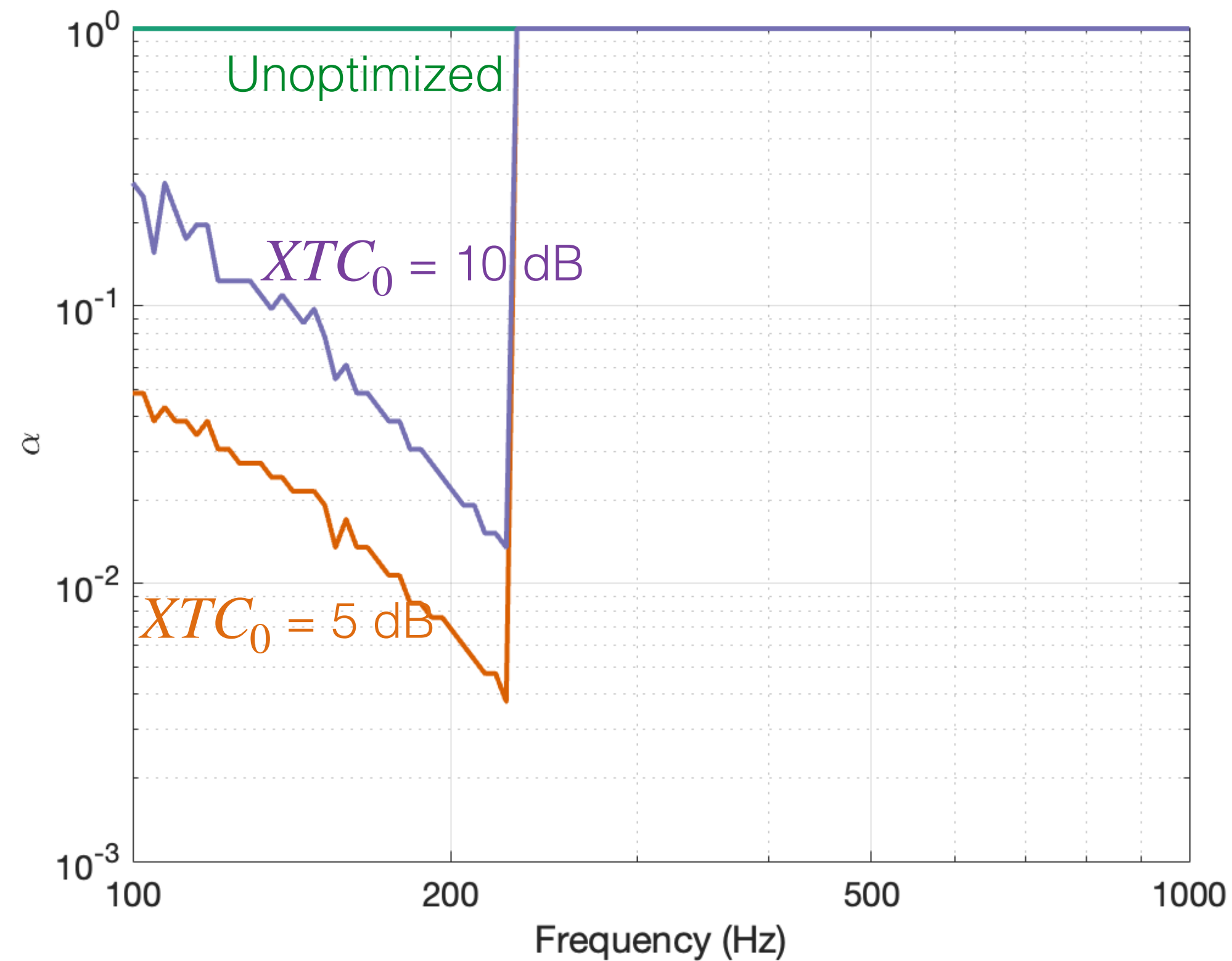


# Optimizing the trade-off

- For  $AC_0 \leq 25$  dB,  $XTC_0$  is set to 5 or 10 dB

Optimal  $\alpha$

Optimized XTC & AC



# Takeaways

- Less distraction is preferred over better spatialization when both are present
- Trade-off can be optimized by adjusting the weights in the PM cost function
- Trade-off mostly exists at low frequencies
  - High frequencies: independently addressed by beamforming

## Caveats

- Established subjective preferences were based on full-range stimuli
- Optimization parameters need to be tuned for each case
- Reproduction error is unconstrained during optimization (might lead to distortion issues)

## Future directions

- Incorporating other metrics
  - tonal coloration
  - dynamic range loss
- Objective & subjective evaluation with different threshold levels
- Adaptive solutions with head tracking

# Optimization of Personal Sound Zones with Spatial Audio

**Yue Qiao\*** & Edgar Choueiri

3D Audio and Applied Acoustics (3D3A) Lab

Princeton University

Presented at the AES AVAR Conference

Aug 16, 2022