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

PRINCETON UNIVERSITY

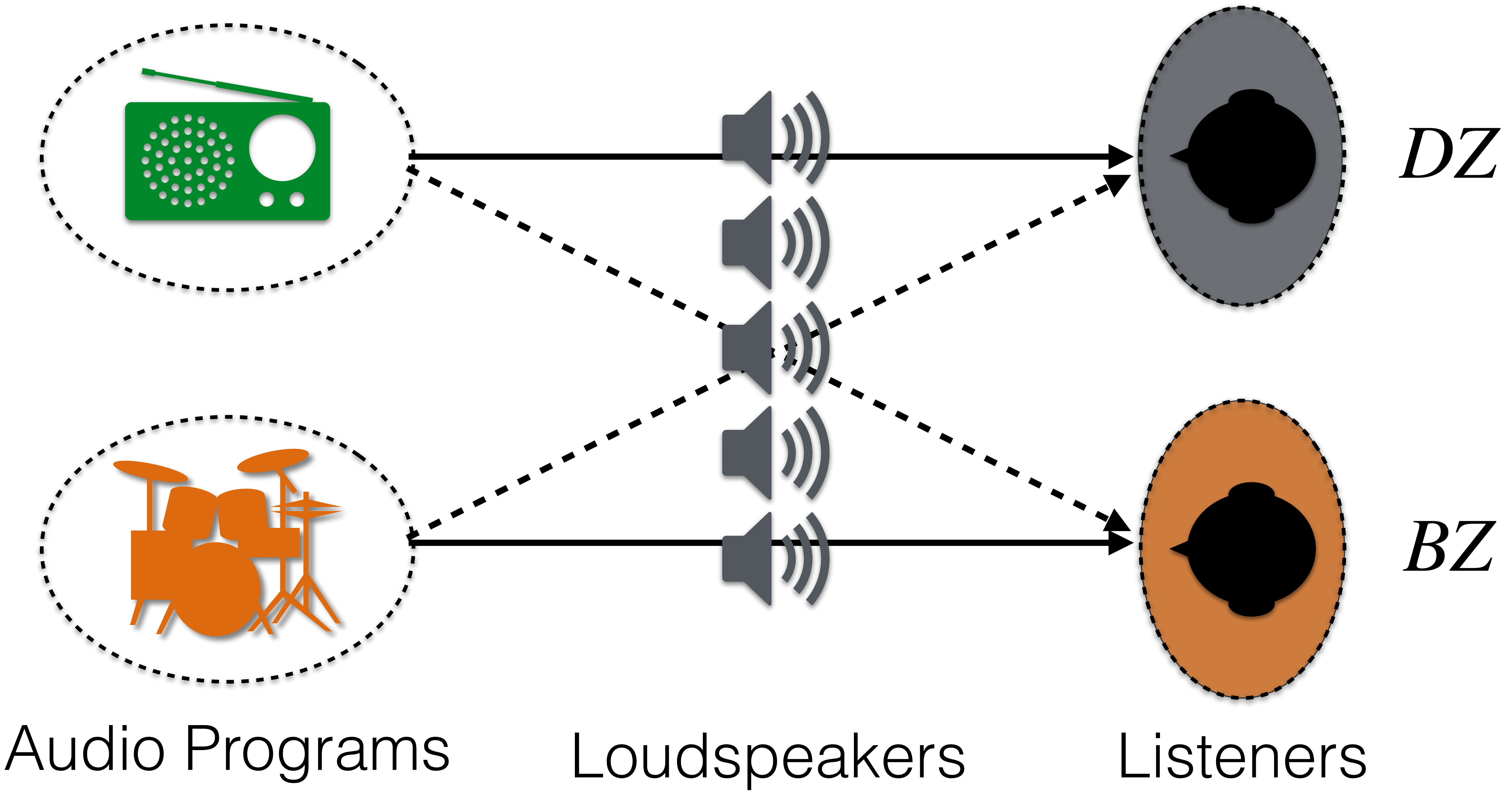
*E-mail: yqiao@princeton.edu

3D3A

Key questions

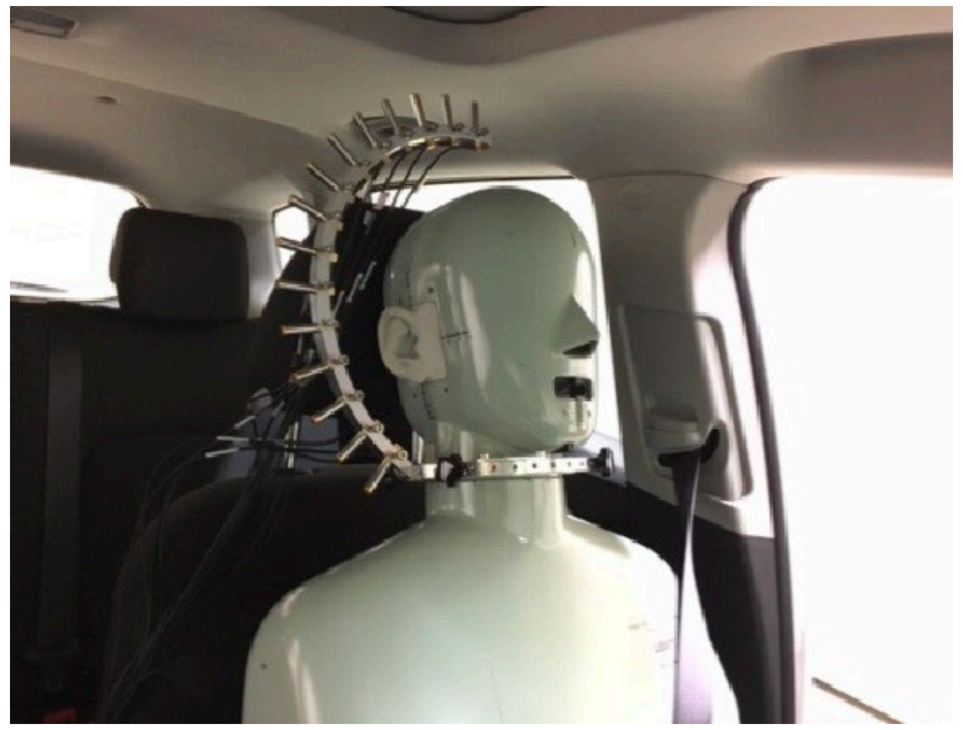
1. How to render spatial audio in personal sound zones?
2. How to optimize both aspects in a single problem?

Concept of PSZ[1]



[1] Druyvesteyn and Garas, JAES, 1997

Example use cases

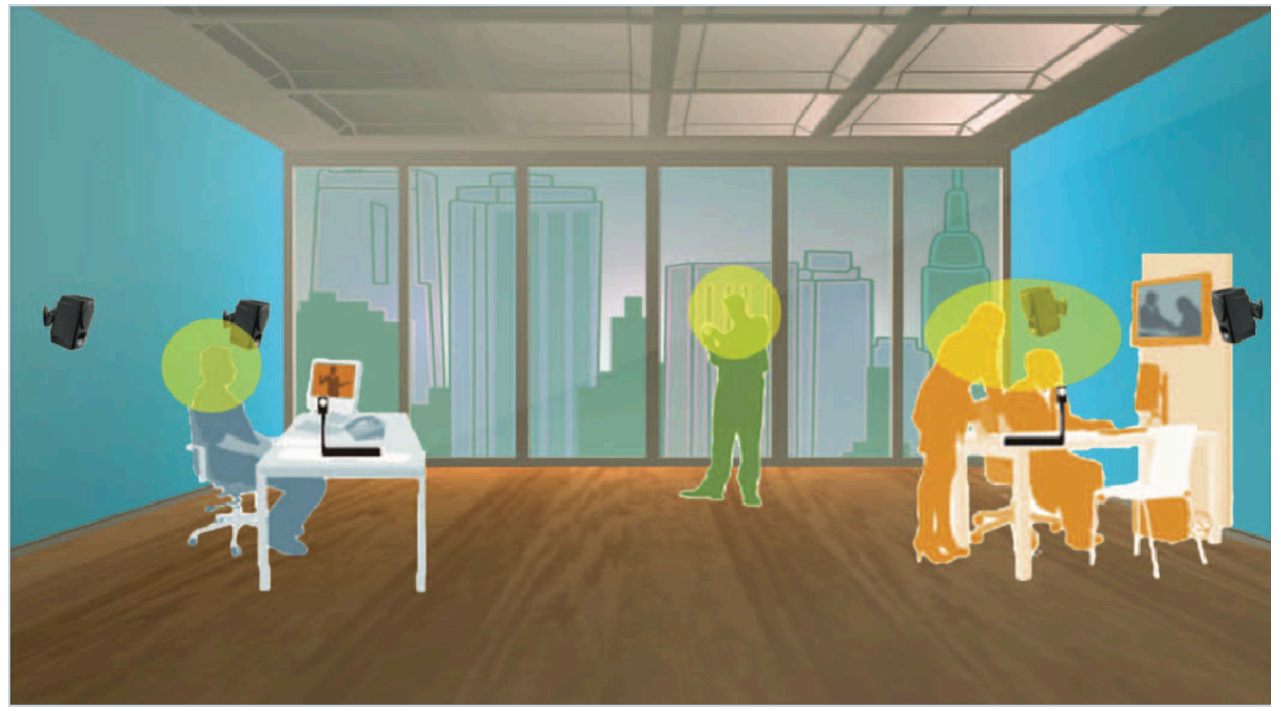


Cho and Chang, ICA, 2019

Vindrola et al., JASA, 2021

Ebri et al., AES Conv., 2020

Automotive cabins



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

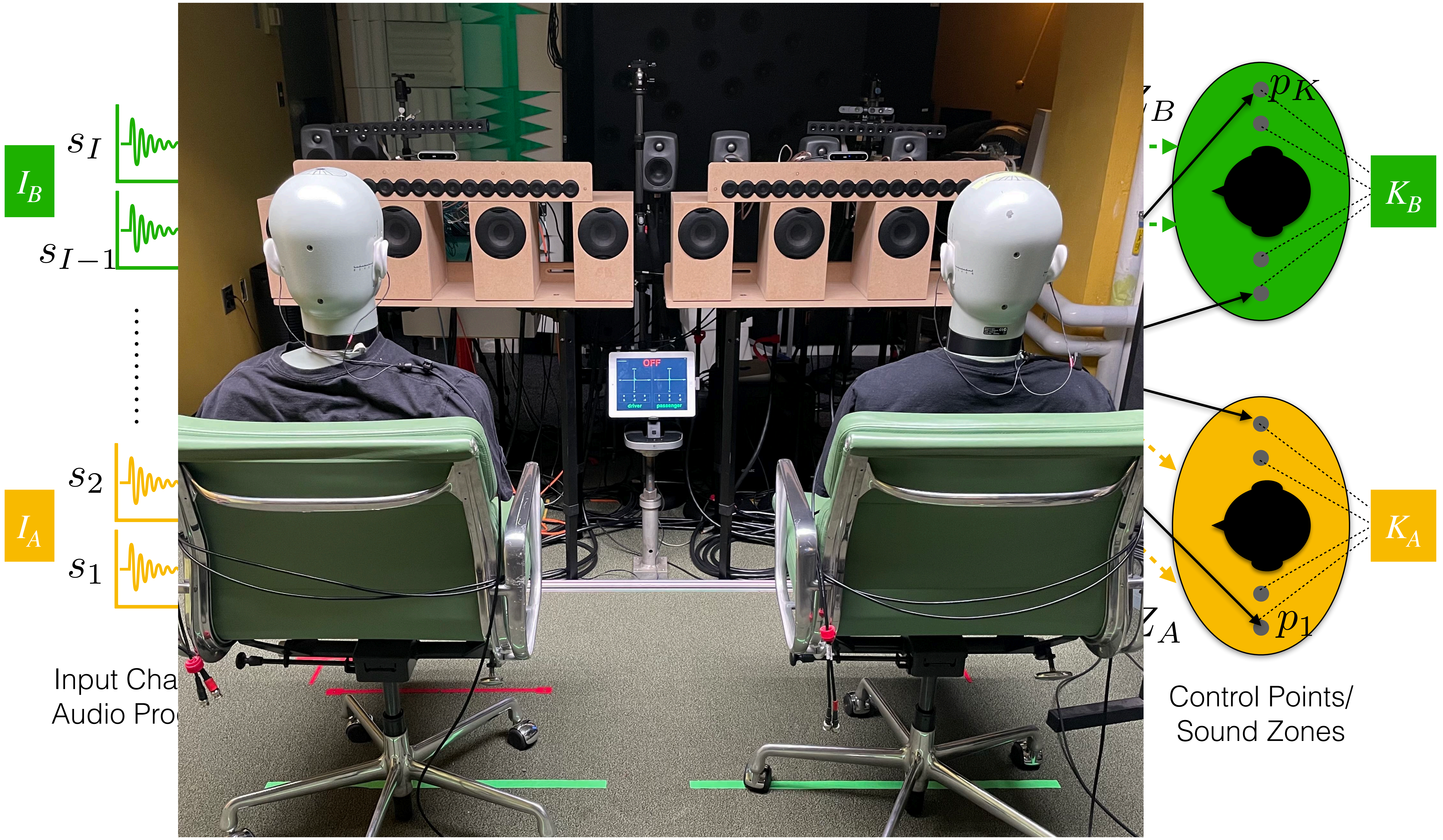
Fraunhofer IDMT

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)[2]

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

Acoustic Contrast Control (ACC)[3]

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

No control over phase

Not suitable for binaural audio

[2] Poletti, AES Conv. 125, 2008 [3] Choi and Kim, JASA, 2002

Rendering spatial audio for two listeners with PM

Specifying target pressure

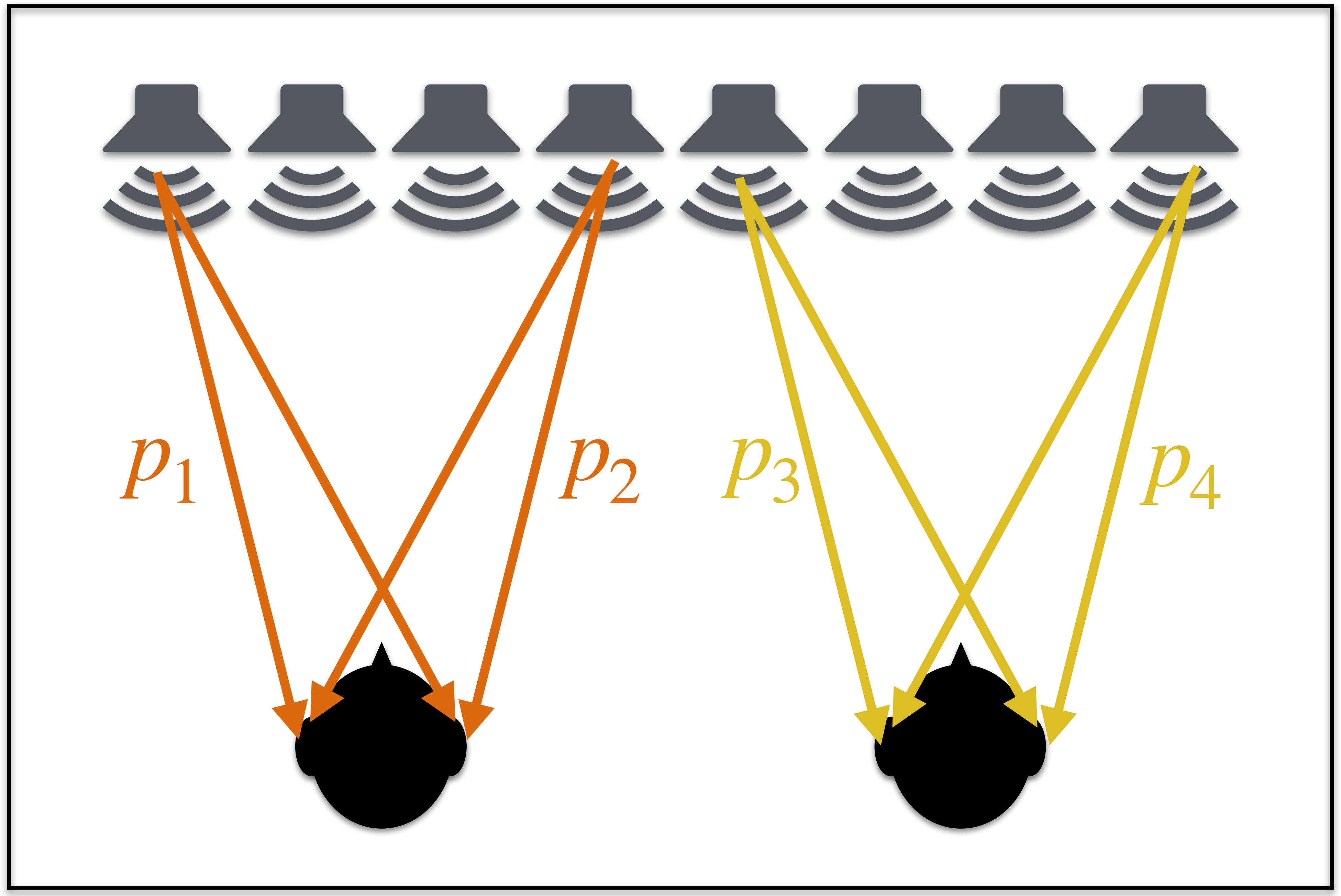
Pressure Matching (PM)[2]

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

Mono programs as input

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

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



Specifying target pressure

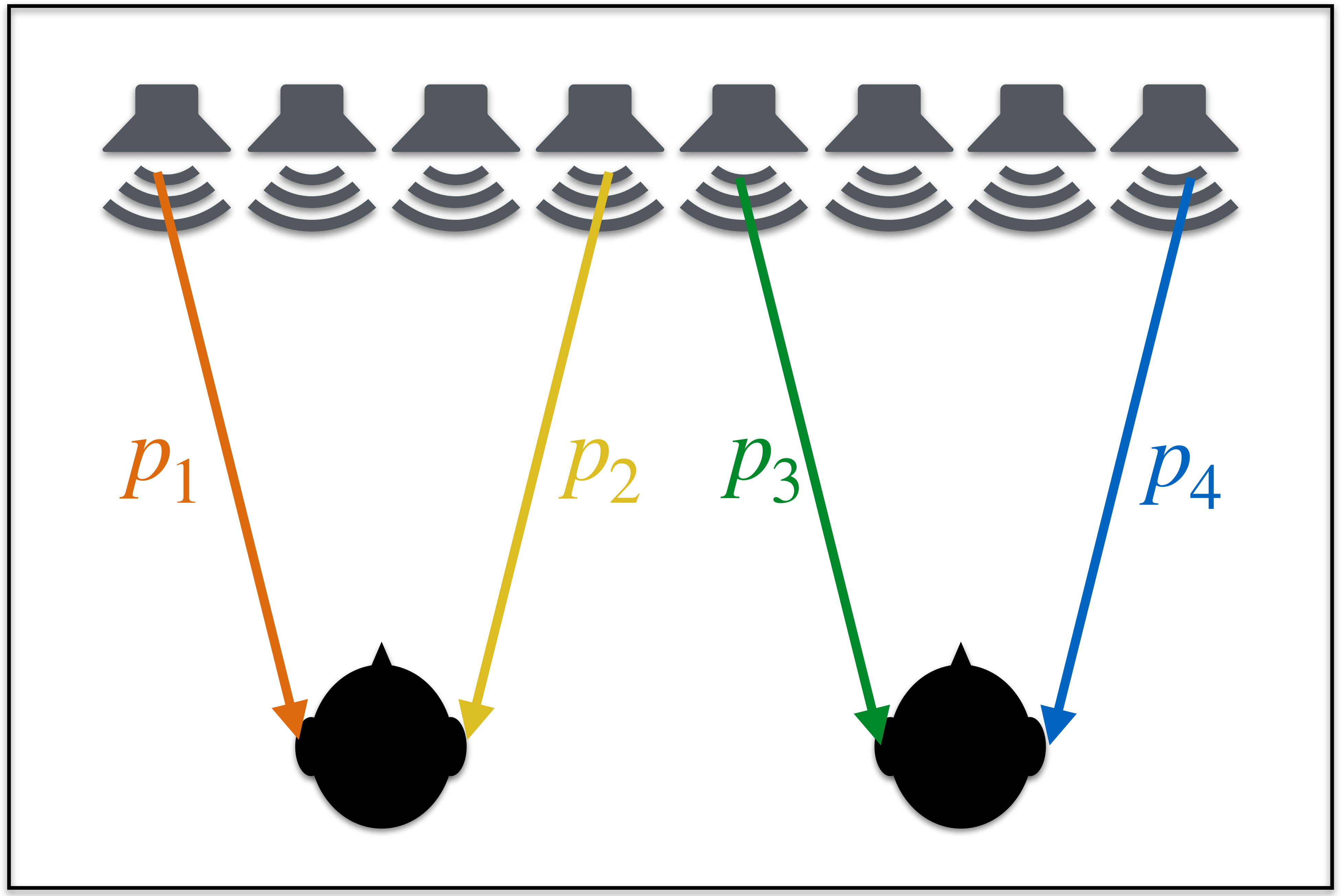
Pressure Matching (PM)[2]

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

Binaural programs as input

4 input channels ⟶ 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

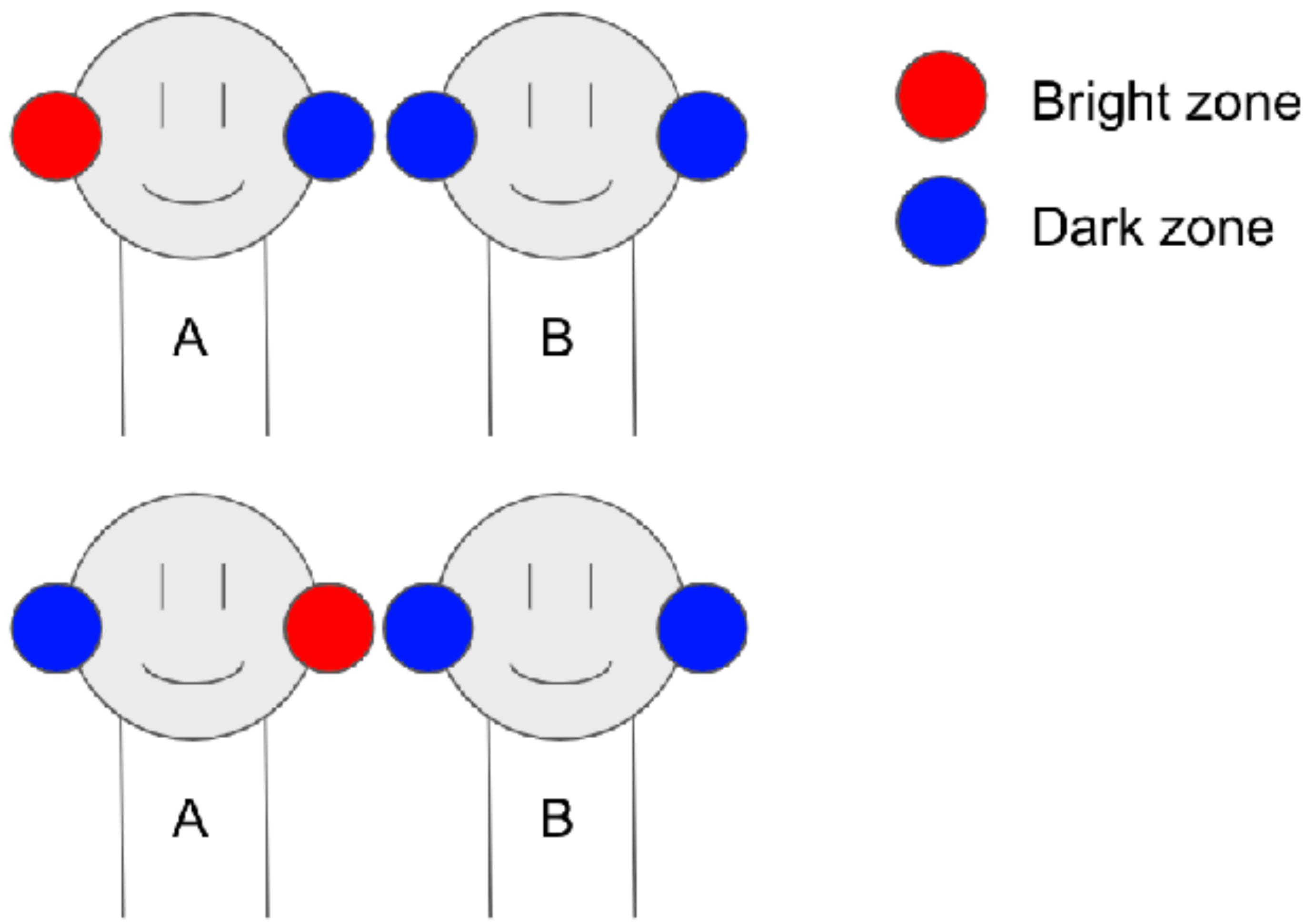
   - Acoustic Contrast (AC)

   - poor performance -> distraction by other audio

2. Isolation between ears

   - Crosstalk Cancellation (XTC)

   - poor performance -> lack of envelopment



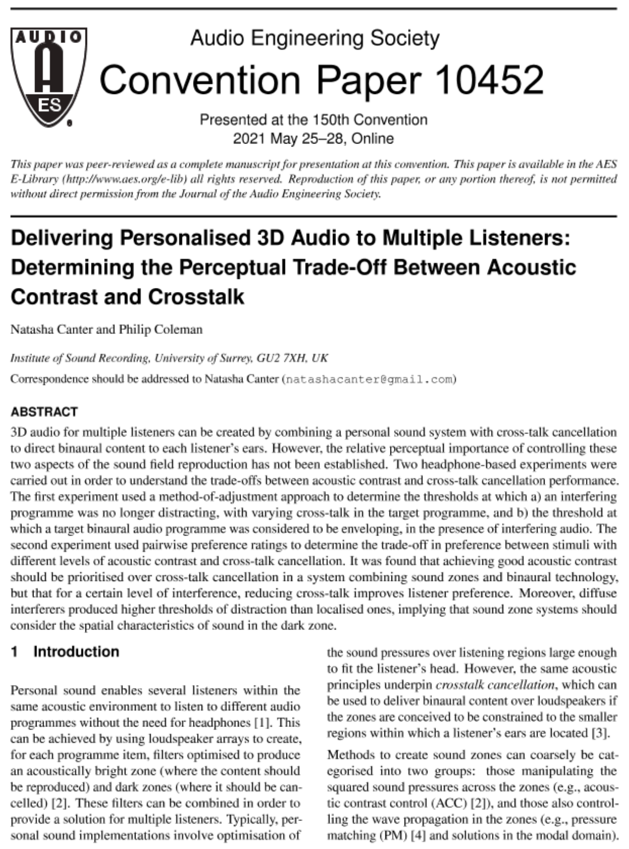
Canter and Coleman, AES Conv., 2021

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


Audio Engineering Society
Convention Paper 10452
Presented at the 150th Convention
2021 May 25–28, Online

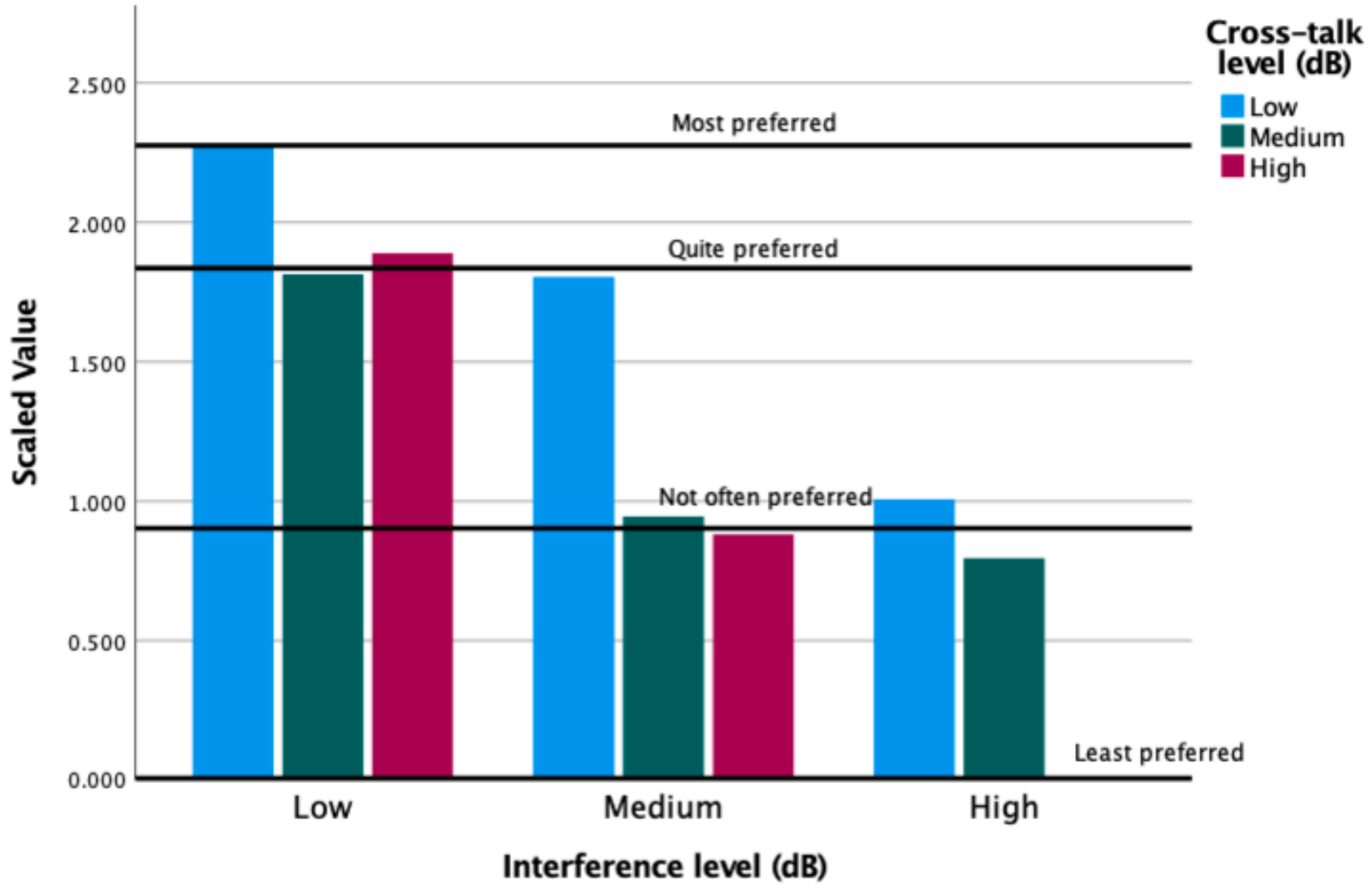
Delivering Personalised 3D Audio to Multiple Listeners: Determining the Perceptual Trade-Off Between Acoustic Contrast and Crosstalk

Natasha Canter and Philip Coleman

Institute of Sound Recording, University of Surrey, GU2 7XH, UK

ABSTRACT

1 Introduction



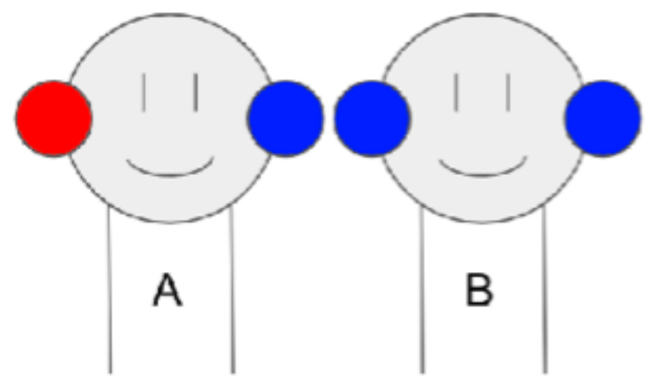
Canter and Coleman, AES Conv., 2021

Finding the optimal trade-off between AC and XTC

Revisit the PM cost function

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

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



XTC | AC | Reproduction Error

Both aspects are treated with same priority

Add a weighting parameter α to control the priority

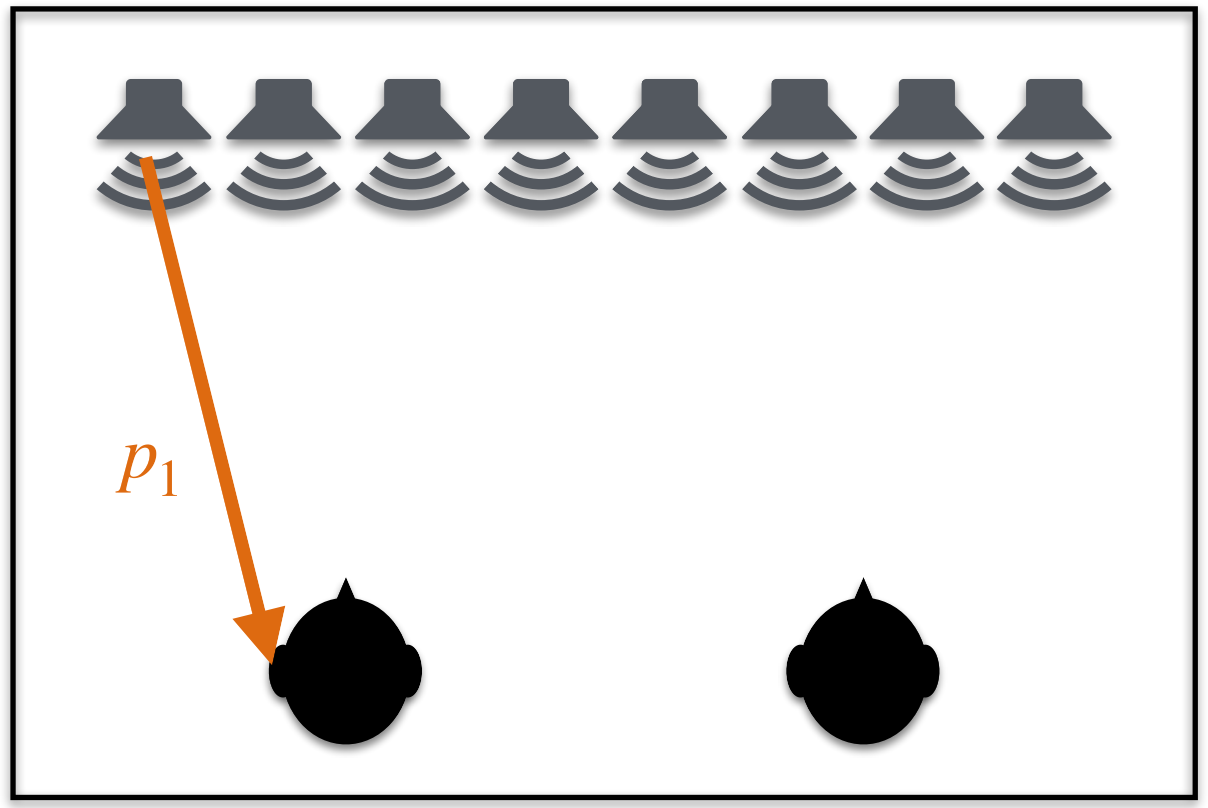
$$J = \alpha\|\mathbf{h}_{B2}^H\mathbf{g}\|^2 + \|\mathbf{H}_D\mathbf{g}\|^2 + \|\mathbf{h}_{B1}^H\mathbf{g} - p_1\|^2$$

α ↘ { AC ↗ Less distraction; XTC ↘ Less envelopment }

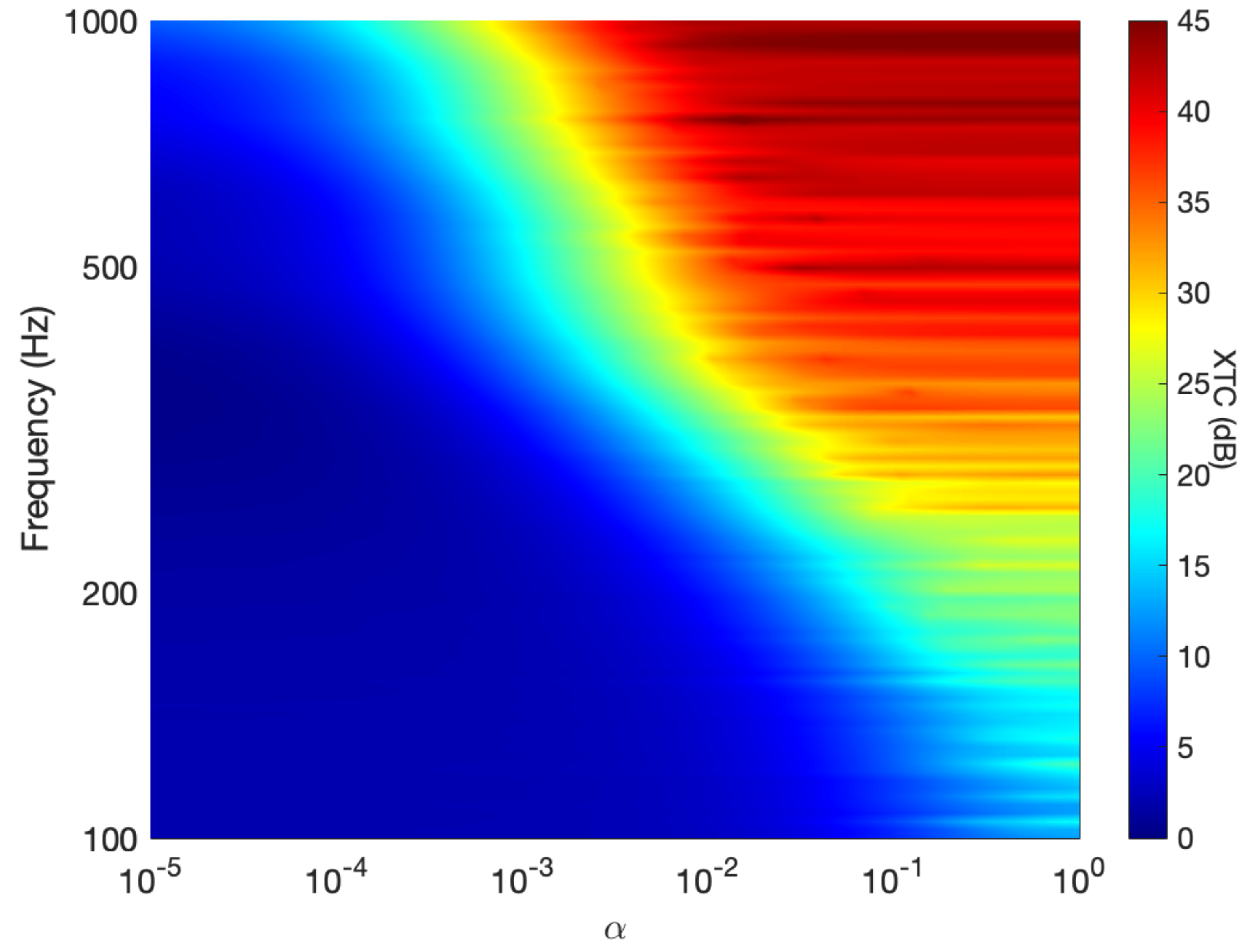
Simulated examples

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

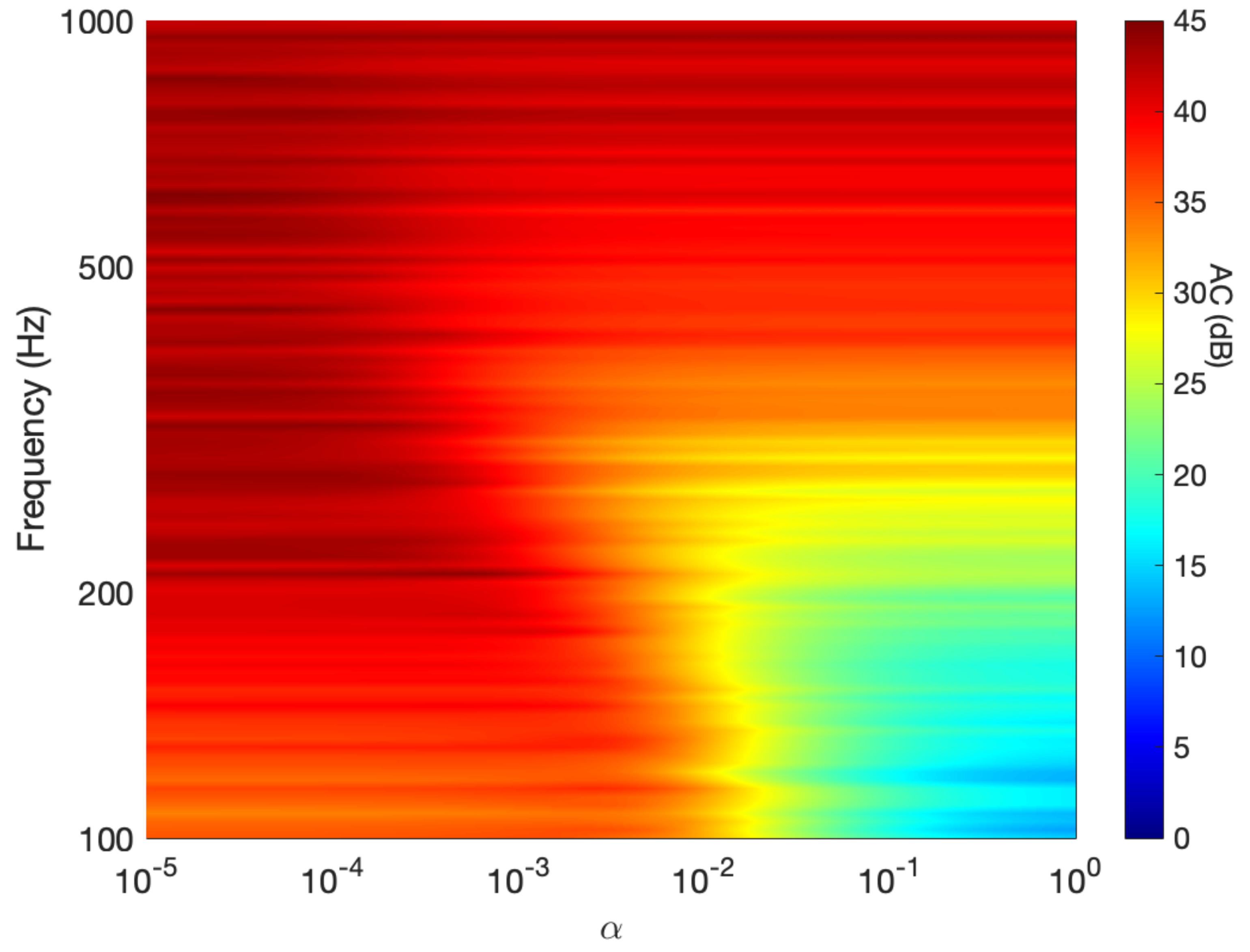
$$\mathbf{p}_T = \begin{bmatrix} p_1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$



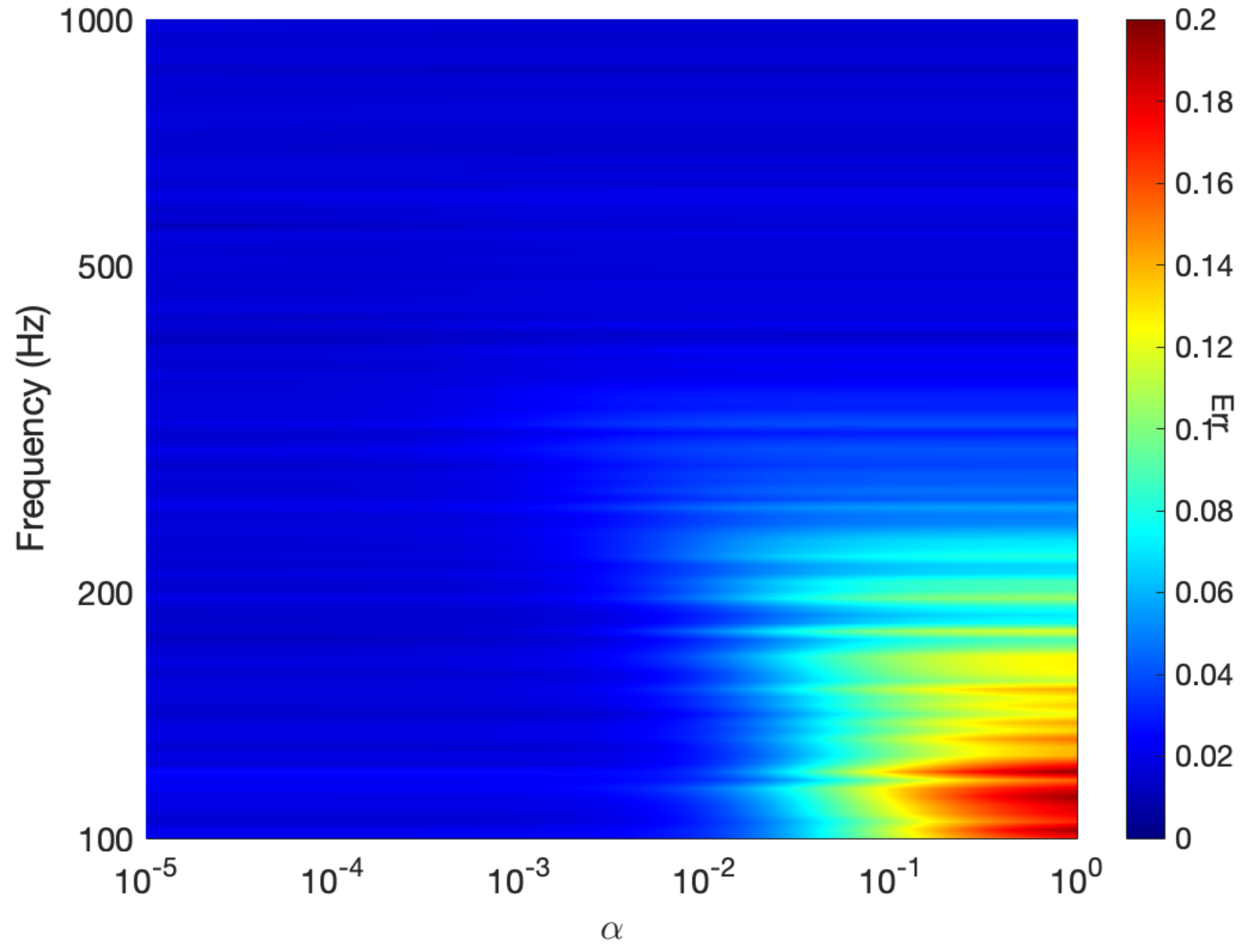
Crosstalk Cancellation (XTC)



Acoustic Contrast (AC)



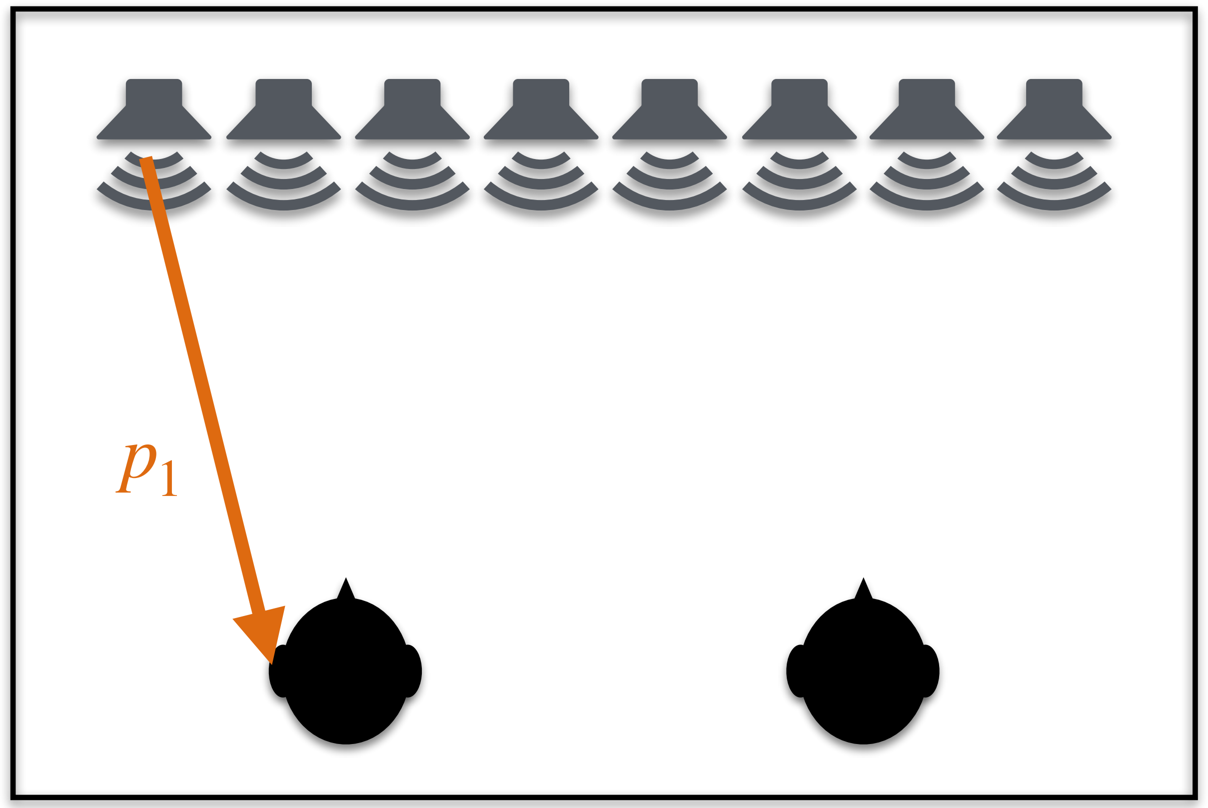
Normalized Reproduction Error (Err)

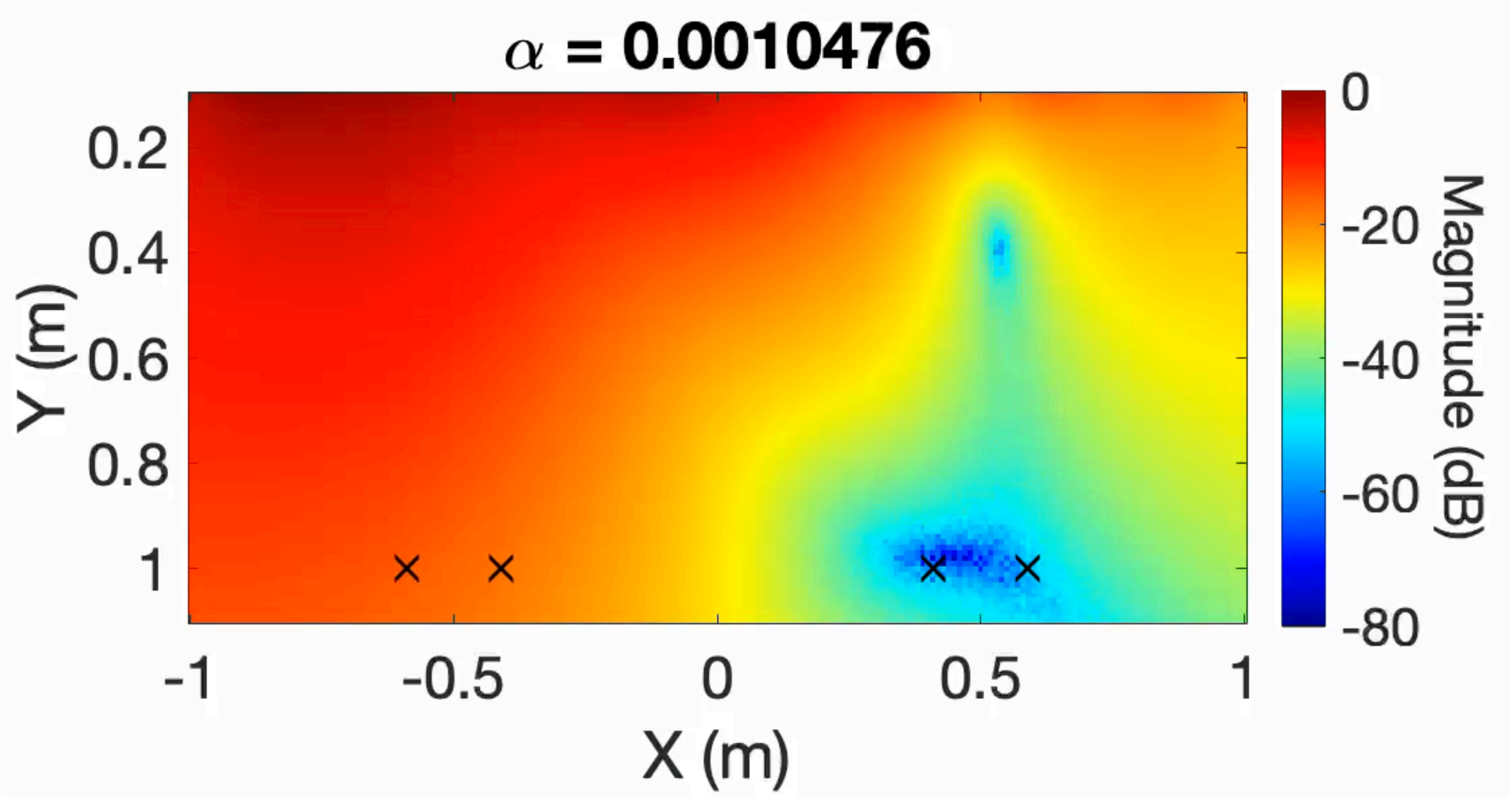


Simulated examples

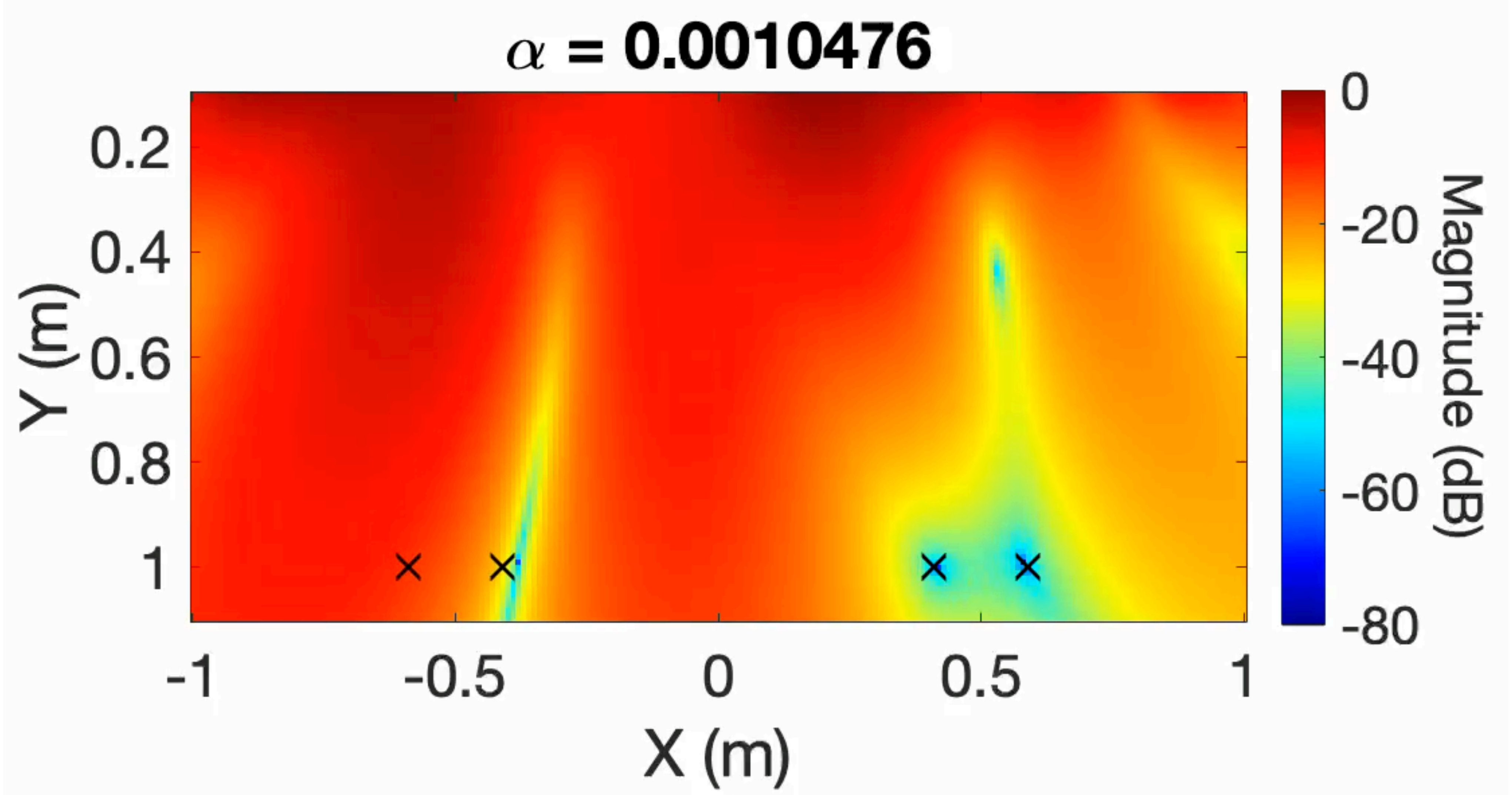
- Free-field condition with point sources
- randomly perturbed transfer functions
- Constant regularization @ 100-1000 Hz

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





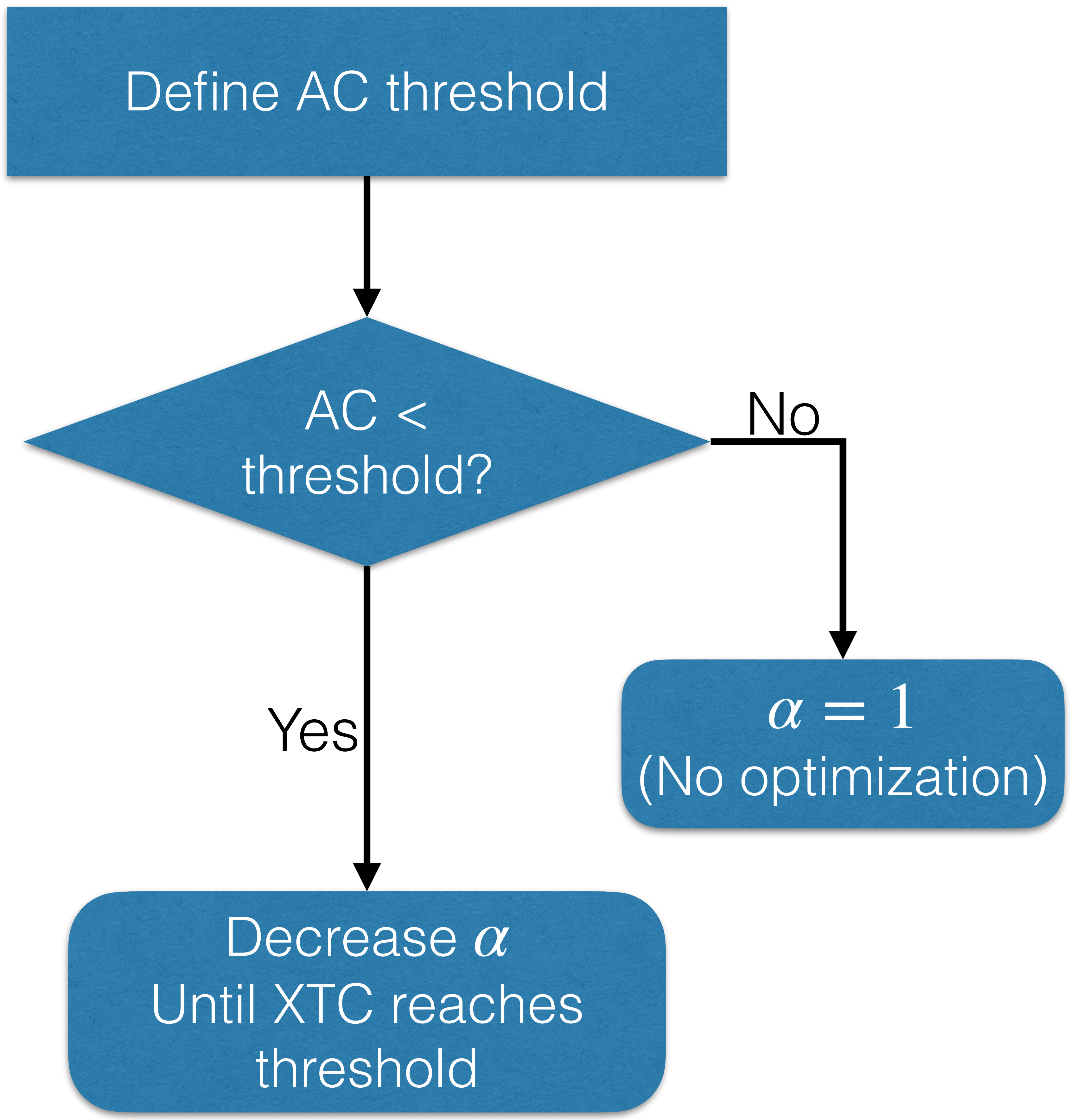
200 Hz



500 Hz

Optimizing the trade-off

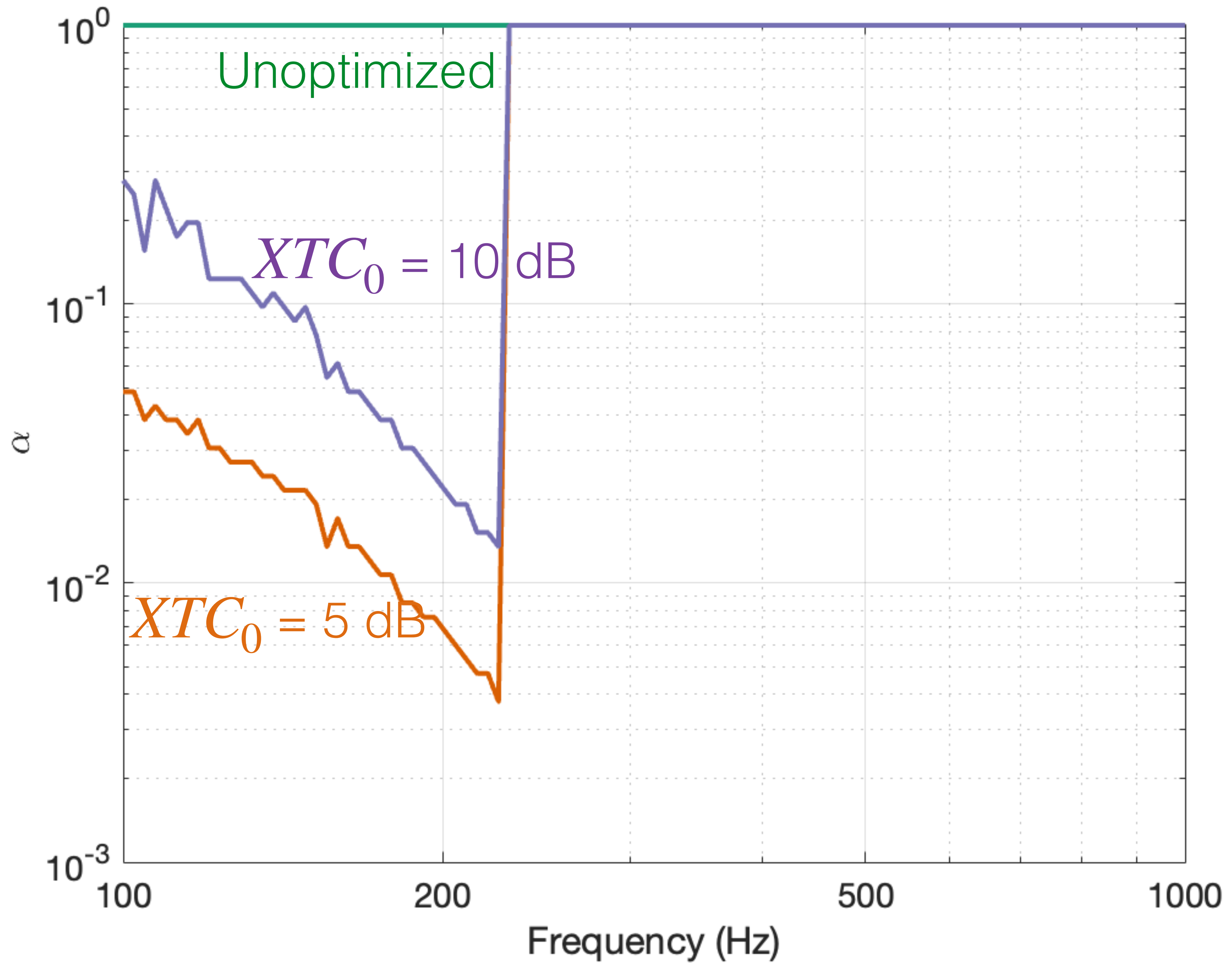
- Principle: trading off XTC for higher AC
- Observations
  - AC and Err are mostly affected by α at low frequencies
  - XTC is affected at almost all frequencies



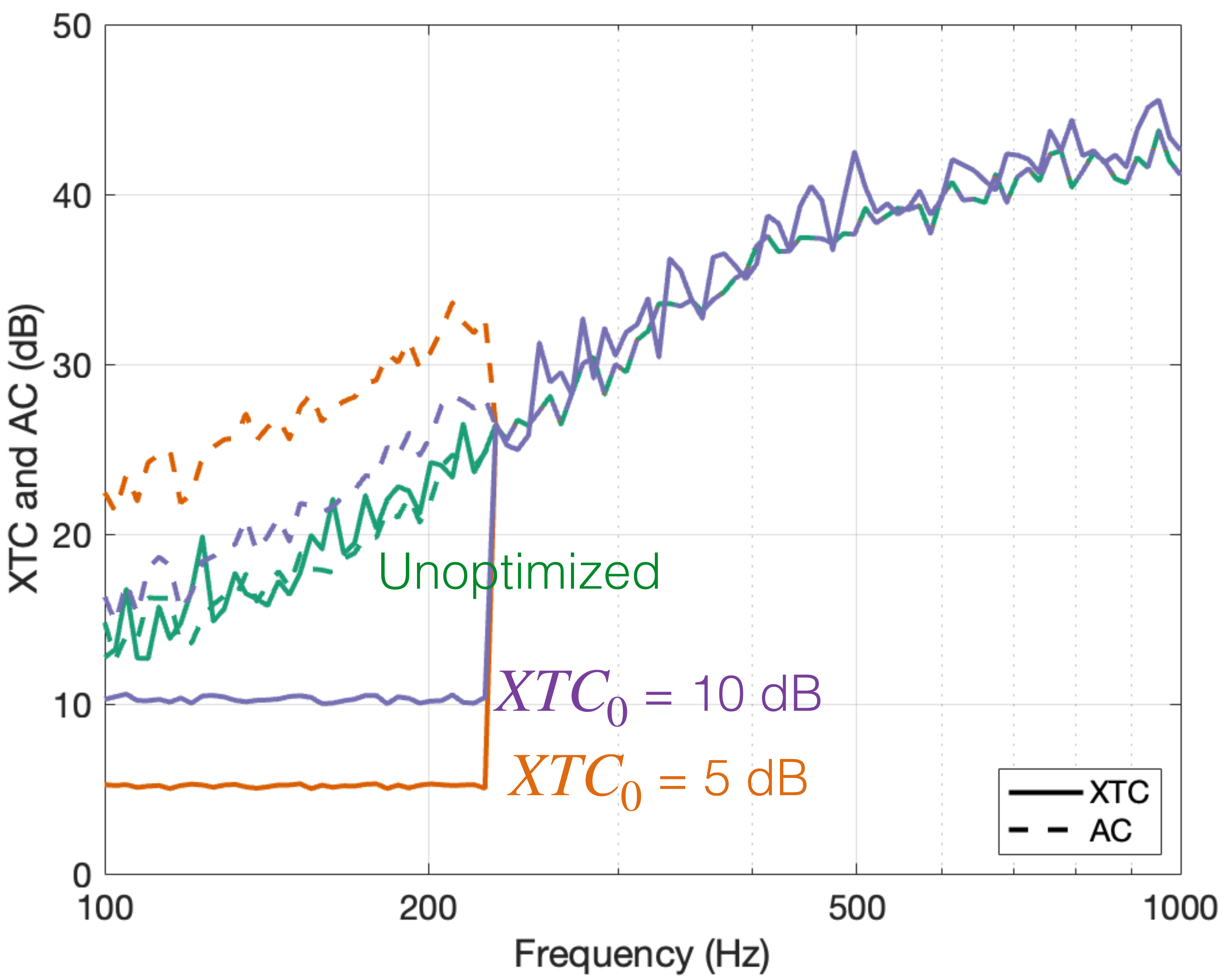
Optimizing the trade-off

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

Optimal α



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

*E-mail: yqiao@princeton.edu