Sub-band Neural Interpolation of Binaural Room Impulse Responses for Personal Sound Zone Control

Yazhou Li, Lin Wang, Joshua Reiss,

Center for Digital Music, Queen Mary University of London

Queen Mary University of London

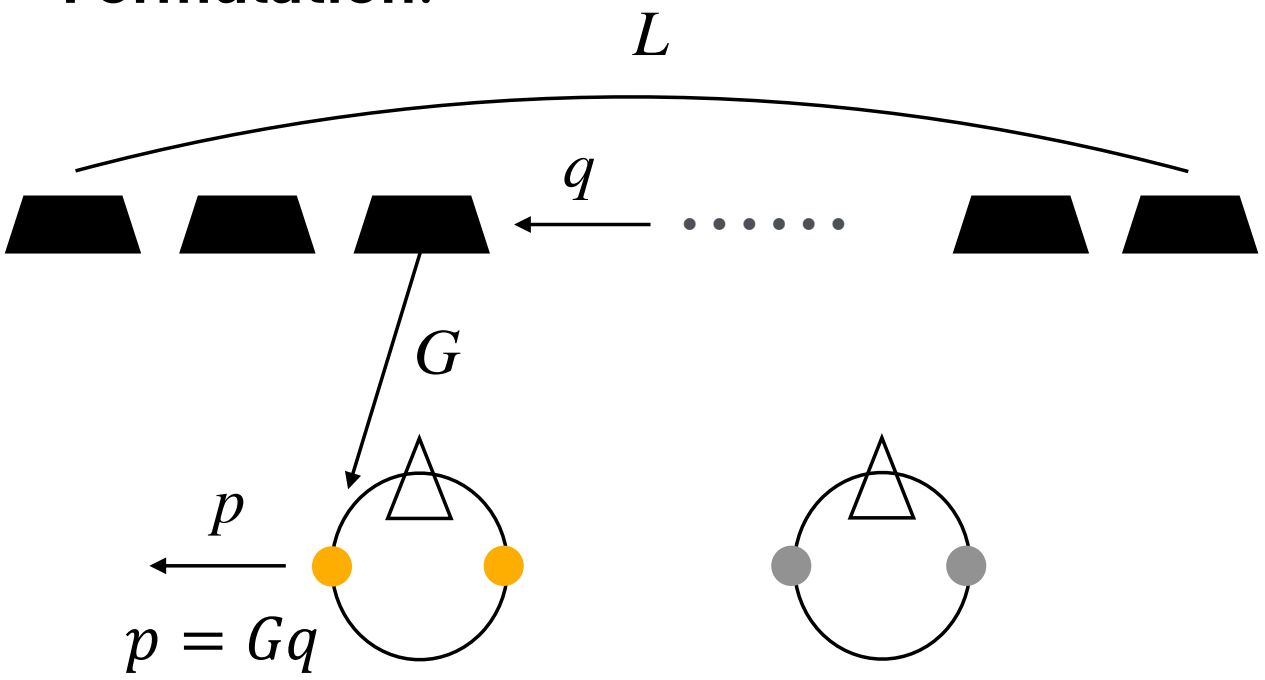
UK Research and Innovation

AI + MUSIC

centre for digital music

Background: Personal Sound Zone (PSZ)

Formulation:



Metric:

Acoustic Contrast $\frac{p_b}{p_d}$

Sound pressure level ratio between bright zone and dark zone

Optimization Method:

Pressure Matching (PM)

$$\min_q \|Gq - p_T\|^2 + \beta\|q\|^2$$

$$q = (G^{\mathrm{H}}G + \beta I)^{-1}G^H p_T$$

Introduction

- PSZ requires BRIRs for each listener position.
  - **Solution:** BRIR interpolation at unmeasured positions.
- BRIR interpolation methods
  - Traditional DSP methods:
    - Bilinear interpolation → smearing peaks problem
    - Dynamic time warping (DTW)[1]
  - Neural methods: implicit neural representation[2]
- **Problem:**
  - Interpolation methods are mostly used for perceptual applications.
- **Question:**
  - Can the interpolation methods work for sound field control?

[1] Bruschi V., “An innovative method for binaural room impulse responses interpolation”, AES, 2020
[2] Qiao Y., “Neural modeling and interpolation of binaural room impulse responseswith head tracking”, AES, 2023

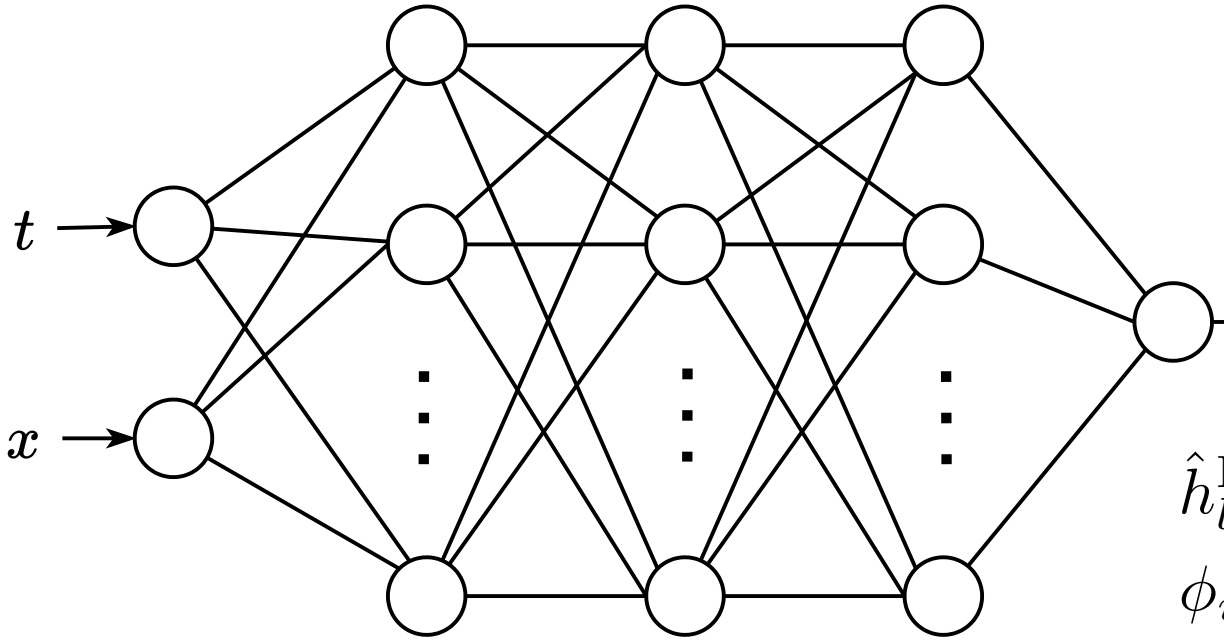
Problem Formulation

- **Input**: measured BRIRs at sparse listener positions
- **Output**: estimated BRIRs at unmeasured listener positions
- **Downstream task**: use estimated BRIRs for PSZ
- **Evaluation**:
  - interpolation accuracy: time-domain NMSE
  - PSZ performance: acoustic contrast

Method SIREN

Implicit Neural Representation (INR):
represents a continuous signal as a neural network that maps input coordinates to signal values.

Sinusoidal representation network (SIREN):
a specific INR architecture using sinusoidal activation functions, which is suitable for representing signals with fine spatial or temporal details.



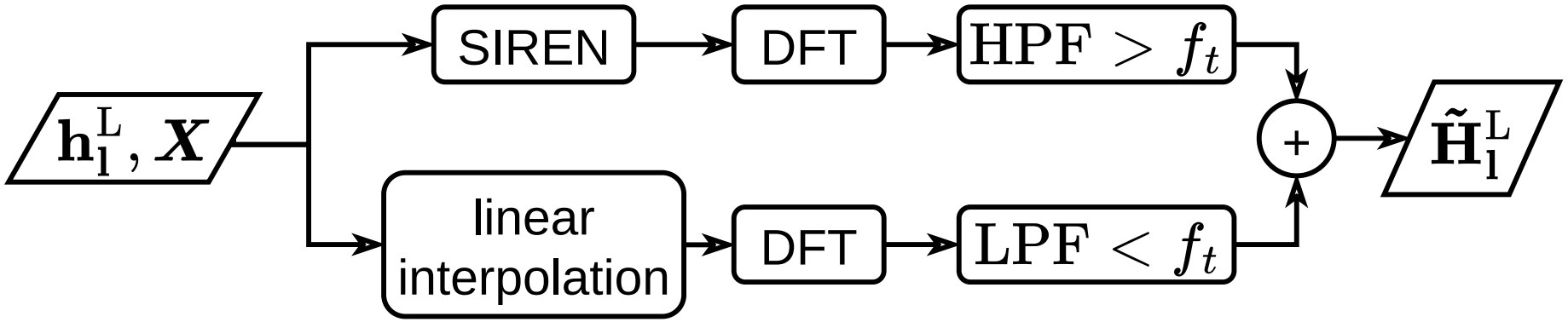
$$\hat{h}_l^{\mathrm{L}}(t, x) = \boldsymbol{W}_n(\phi_{n-1} \circ \phi_{n-2} \circ \ldots \circ \phi_0)(t, x) + \boldsymbol{b}_n,$$
$$\phi_i(\boldsymbol{x}_i) = \sin(\omega(\boldsymbol{W}_i \boldsymbol{x}_i + \boldsymbol{b}_i)),$$

Spectral Bias Problem

- Neural networks learn low frequencies first[1]
- SIREN:
  - high-frequency components dominate during long training, low frequencies become noisy
  - choosing parameter ω not trivial
- Current solutions:
  - different initialization schemes[2]

[1] Rahaman N., “On the Spectral Bias of Neural Networks”, ICLR, 2019
[2] Novello T., “Tuning the Frequencies: Robust Training for Sinusoidal Neural Networks”, CVPR, 2025

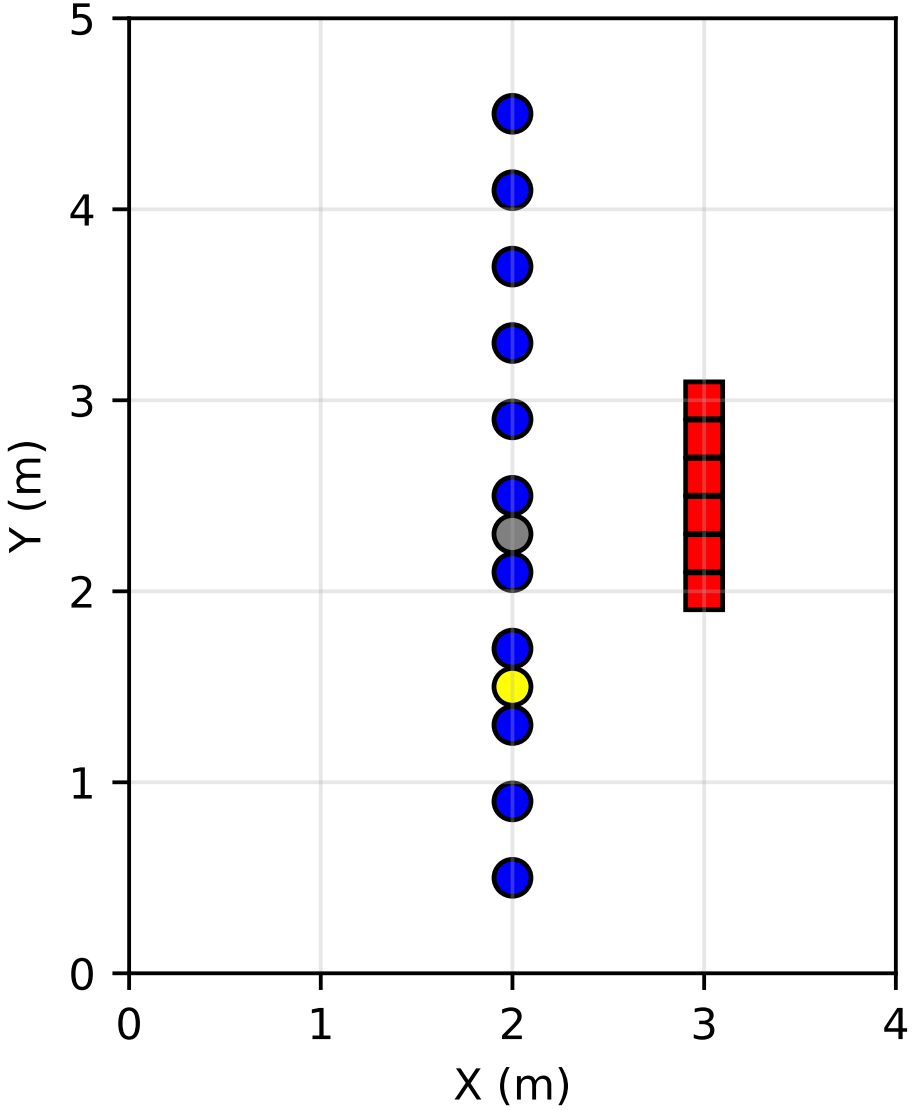
Method hybrid interpolation



Choice of turning frequency:

- fixed ft
- adaptive ft (not available in real applications)

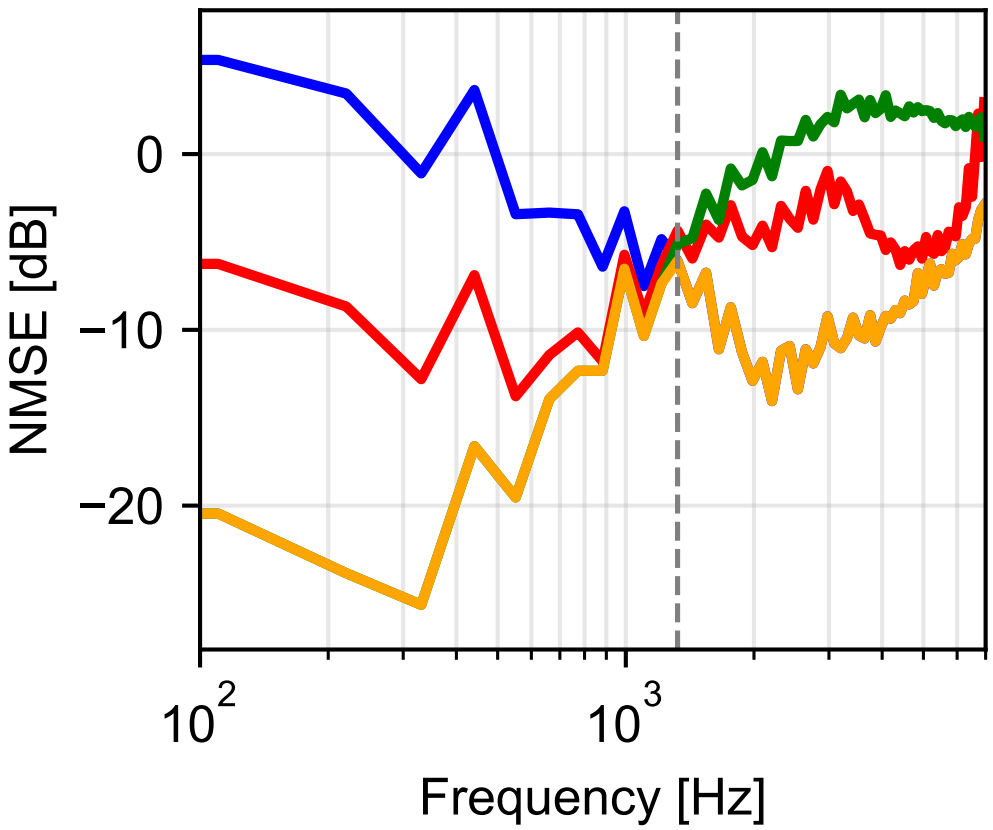
Simulation

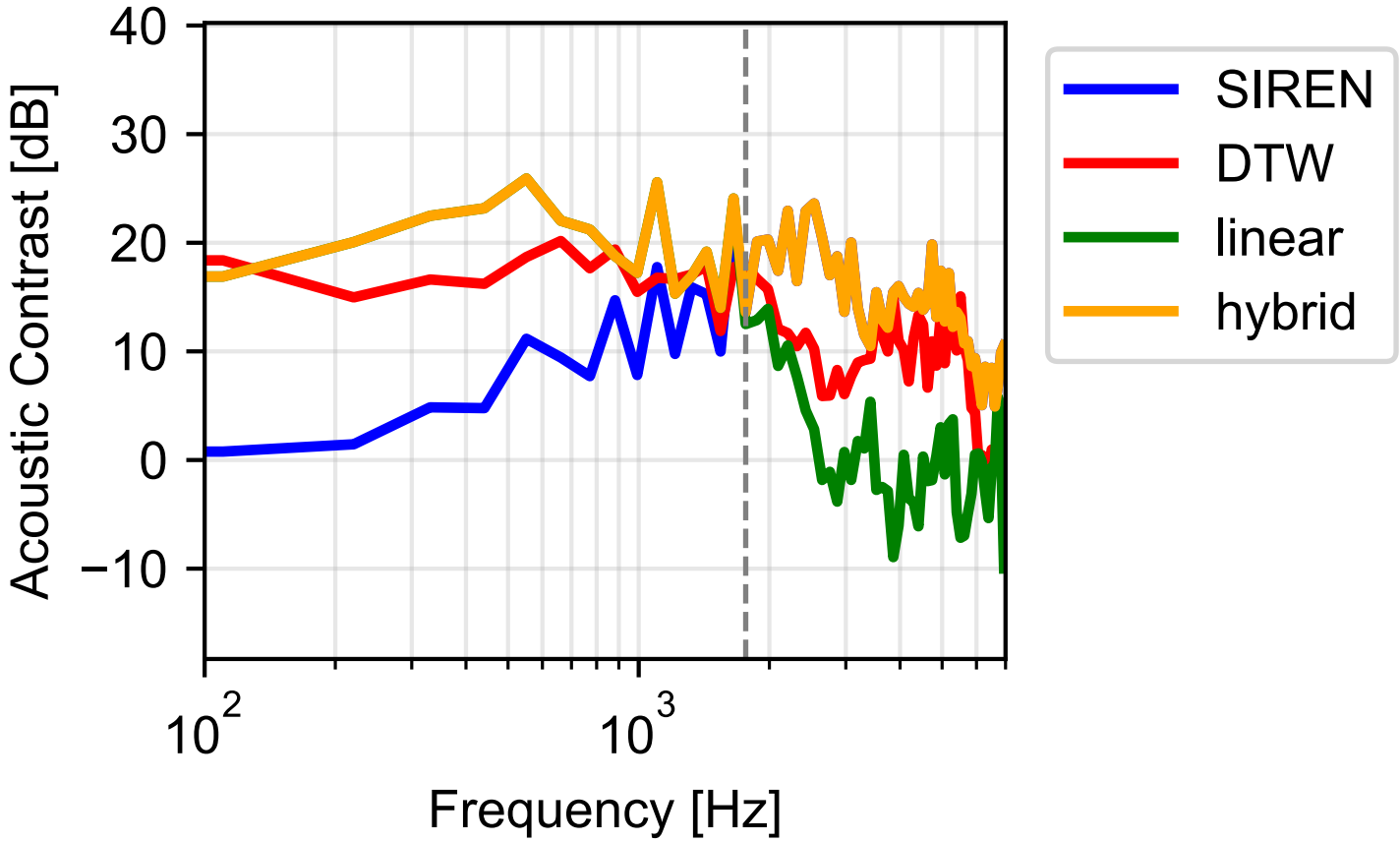


- 6 loudspeakers, Z measurement positions
  Z = 11, 21, …, 111 (Z=11 in the figure)
- BRIRs simulated using RAZR[1] toolbox
  + CIPIC KEMAR HRIRs
  RT60 = 0.1s, sr=44100Hz, 1000samples (22.7ms)
- One network for each loudspeaker and each ear
  (12 networks in total)
- Baseline: DTW

[1] Torben, W. “A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation” *JAES,* 2014

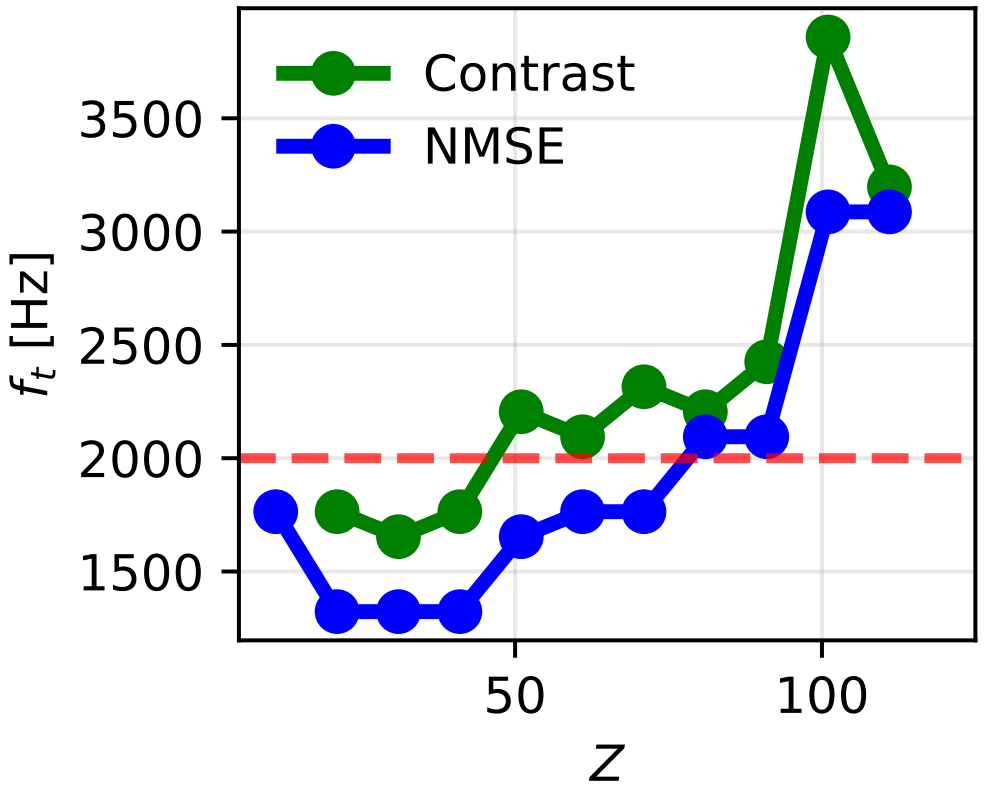
Results

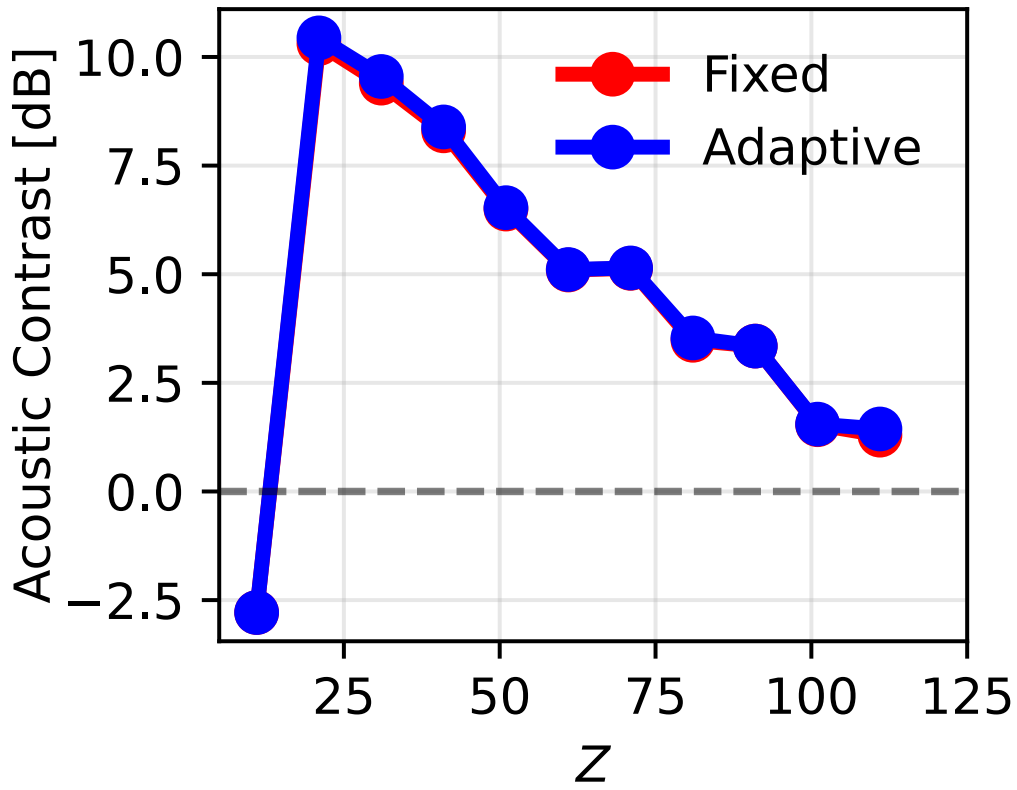




Method	NMSE (dB) ↓	AC (dB) ↑
Linear	−1.96	4.96
DTW	-4.79	10.41
SIREN	−7.12	12.85
Hybrid	**−9.89**	**15.40**

Results: Turning Frequency Analysis





Adaptive ft gives slight gains, fixed 2000 Hz is generally sufficient.

Conclusion

- SIREN interpolated BRIRs suffer from the spectral bias problem
- The proposed hybrid interpolation method improves PSZ performance.
- A fixed turning frequency of 2000 Hz is sufficient in most settings.
- Future work:
  - alternative neural architectures for full-band interpolation
  - physical experiment validation