Adaptive downmixing of Ambisonic signals using orthogonal basis transformations

Alex Tung

Schedule

1. Introduction
2. Background
   - Wave equation
   - Fourier Theory, Ambisonic solution
   - Encode-decode framework
   - Ambisonic transformations
3. Experiment
   - Proposed Solution, Setup
   - Progress

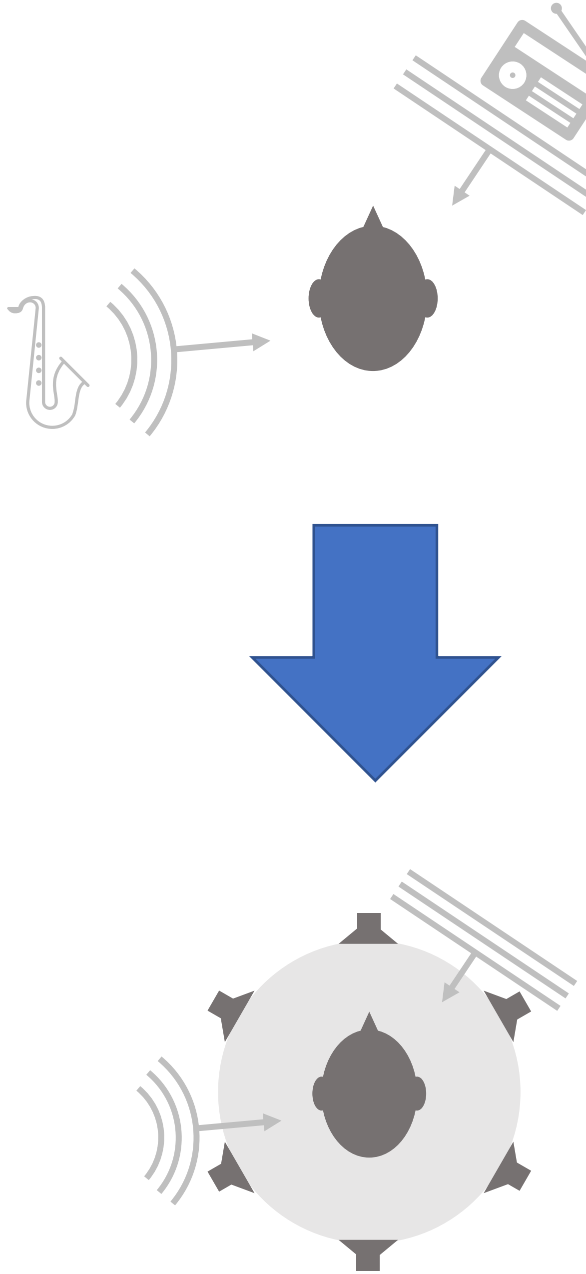
Introduction

Ambisonics is a popular multichannel encoding format for storing a three-dimensional acoustic field

- Geometrically robust for playback over different systems
- Can be encoded, decoded, and manipulated with matrix multiplications without the need to account for algorithm lag (e.g. impulse response tails, truncation)
- No theoretical limit for spatial resolution (order)

Problem: Increasing the resolution of an Ambisonic signal is costly from a systems design perspective: channel width increases with order $\propto (M+1)^2$, which presents a resolution-bandwidth tradeoff.

Is it possible to increase resolution of an Ambisonics playback chain while maintaining a minimal bandwidth?



Background

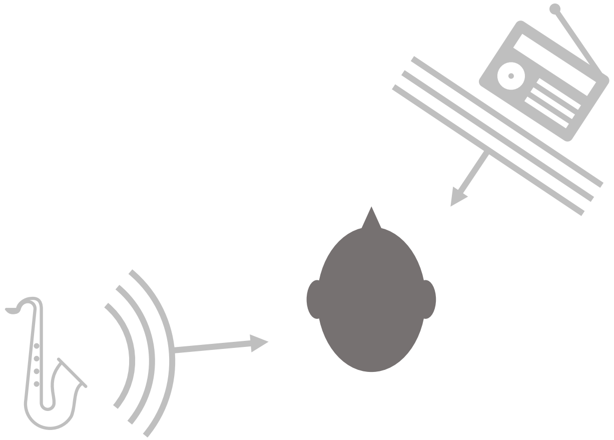
Wave Equation

- Linear acoustic propagation in three-dimensional space is governed by the three-dimensional wave equation

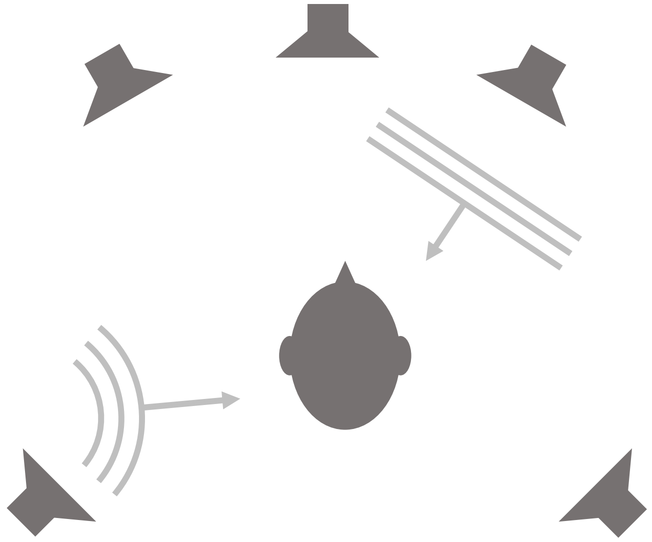
$$\frac{\delta^2 p}{\delta t^2} = c^2 \frac{\delta^2 p}{\delta \boldsymbol{x}^2}$$

$p =$ pressure
$t =$ time
$x =$ displacement
$c =$ prop. speed

- Spatial audio aims to capture and reproduce a given field of propagation with perceptual accuracy
  - **Capture:** microphone placement, array geometry, beamforming
  - **Replication/Playback:** speaker arrays, home theater standardization, binauralization



Original Soundfield



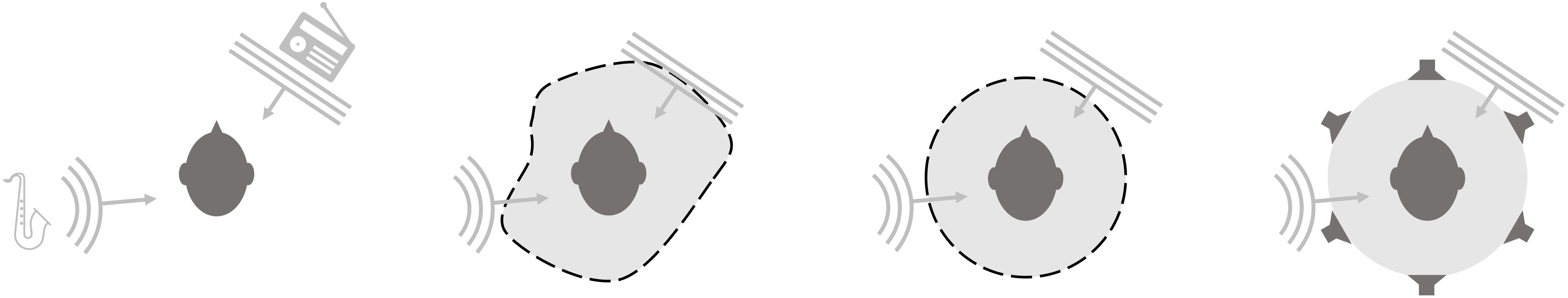
Reproduced Soundfield

Background

Wave Equation

- A sound field can be characterized by imposing boundary conditions on the wave equation
  - **Capture** → capture boundary conditions
  - **Playback** → reproduce interior conditions, change boundary conditions

- Choosing appropriate boundary conditions and constraints can yield an elegant solution
  - Listener is placed inside a closed boundary, sources originate from outside → Kirchhoff-Helmholtz integral
  - Circular/spherical boundary → consistent coordinate system centered around the listener → spatial Fourier theory
  - Discrete number of sources → discretization of integral → Fourier Series



Original Soundfield *Closed Boundary* *Radial Boundary* *Discretization*

Background

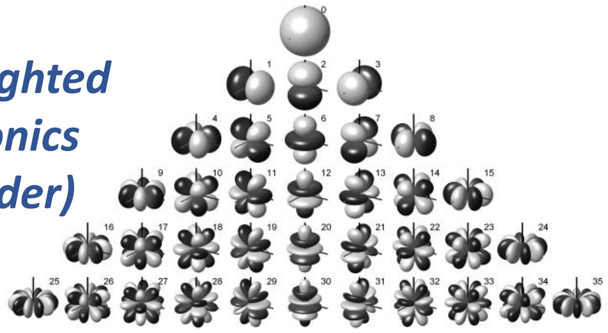
Fourier Theory and the Ambisonic Solution

- A plane wave incident at angle (θ, ϕ) recorded at the boundary of a sphere can be expressed as a linear combination of spatial basis functions

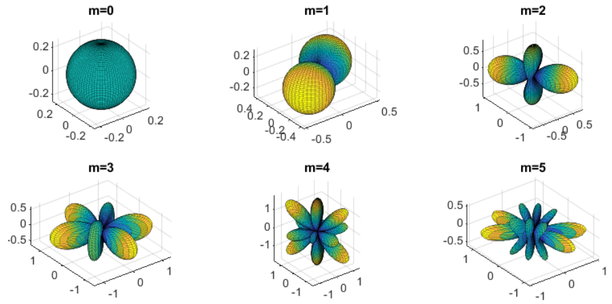
$$p(\vec{r}, t) = \sum_{m=0}^{M} \left[j_m(kr) \sum_{\substack{n \in [0,m] \\ \sigma = \pm 1}} B_{mn}^{\sigma}(t)\, Y_{mn}^{\sigma}(\theta, \phi) \right]$$

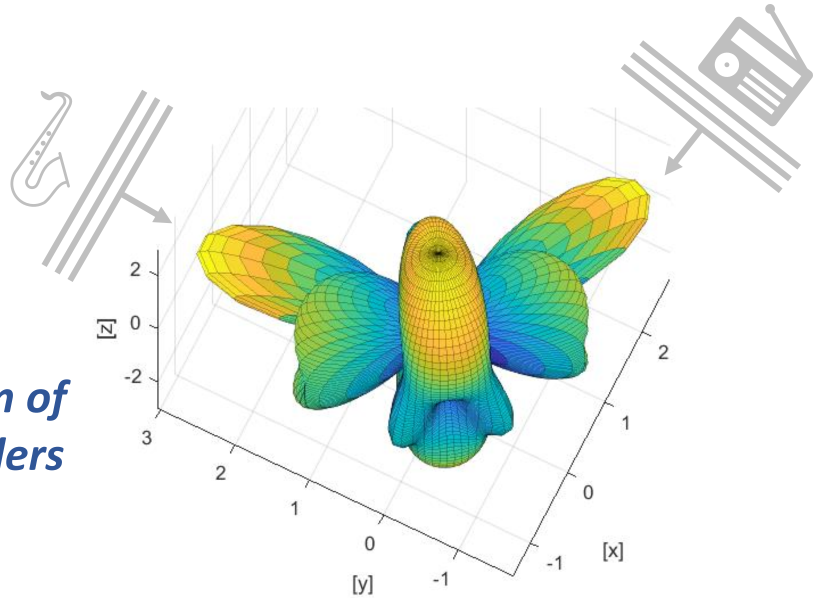
  - Spherical harmonics for each order m derived from Schmidt-normalized Legendre Polynomials (Y_{mn}) and Bessel functions of the first kind (j_m)
  - Can be extended to a combination of many incoming waves
  - Spatial resolution is related to the highest spatial frequency used (Nyquist-Shannon), infinite when $M \to \infty$
- **Ambisonics** is the representation framework used to encode these soundfields

Unweighted Harmonics (5^{th} Order)



Weighted Harmonics, Summed per Order





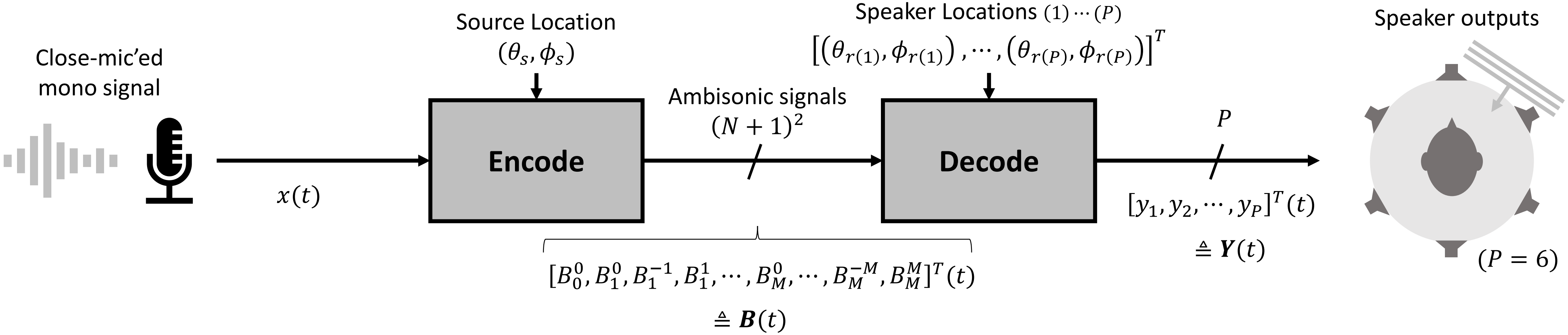
Sum of Orders

Upper image source: File:Spherical Harmonics deg5.png. (2021, May 23). Wikimedia Commons, the free media repository

Background

Encode-Decode Framework

Encoding and decoding can be implemented as time-invariant matrix multiplications



$$\boldsymbol{E} \triangleq \begin{bmatrix} E_0^0(\theta_s, \phi_s) \\ E_1^0(\theta_s, \phi_s) \\ \vdots \\ E_M^M(\theta_s, \phi_s) \end{bmatrix}$$

Encoding Matrix

$$\boldsymbol{D} \triangleq \begin{bmatrix} D_0^0(\theta_{r(1)}, \phi_{r(1)}) & D_0^0(\theta_{r(2)}, \phi_{r(2)}) & \cdots & D_0^0(\theta_{r(P)}, \phi_{r(P)}) \\ D_1^0(\theta_{r(1)}, \phi_{r(1)}) & D_1^0(\theta_{r(2)}, \phi_{r(2)}) & \cdots & D_1^0(\theta_{r(P)}, \phi_{r(P)}) \\ \vdots & \vdots & \ddots & \vdots \\ D_M^M(\theta_{r(1)}, \phi_{r(1)}) & D_M^M(\theta_{r(2)}, \phi_{r(2)}) & \cdots & D_M^M(\theta_{r(P)}, \phi_{r(P)}) \end{bmatrix}$$

Decoding Matrix

Background

Ambisonic Transformations

Adjust Gain: Scale all coefficients by constant value

Rotation: Apply linear map from $RIJK$ quaternions to spherical domain

Downmix: Discard higher-order coefficients

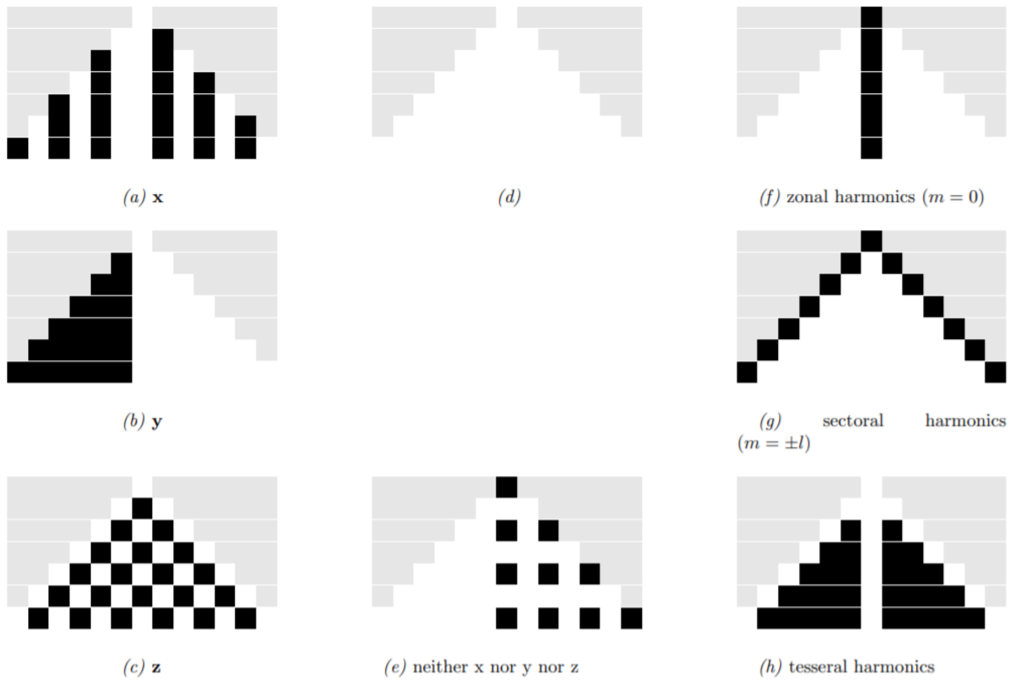
Spatial Mask: Selectively attenuate or discard specific coefficients based on spherical harmonic sector

$$\begin{bmatrix} \boldsymbol{\alpha} & 0 & 0 & 0 \\ 0 & \boldsymbol{\alpha} & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \boldsymbol{\alpha} \end{bmatrix}$$

Gain Factor $\boldsymbol{\alpha}$

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \mathbf{1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{1} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \mathbf{1} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \mathbf{0} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \mathbf{0} \end{bmatrix}$$

HOA to FOA Downmix



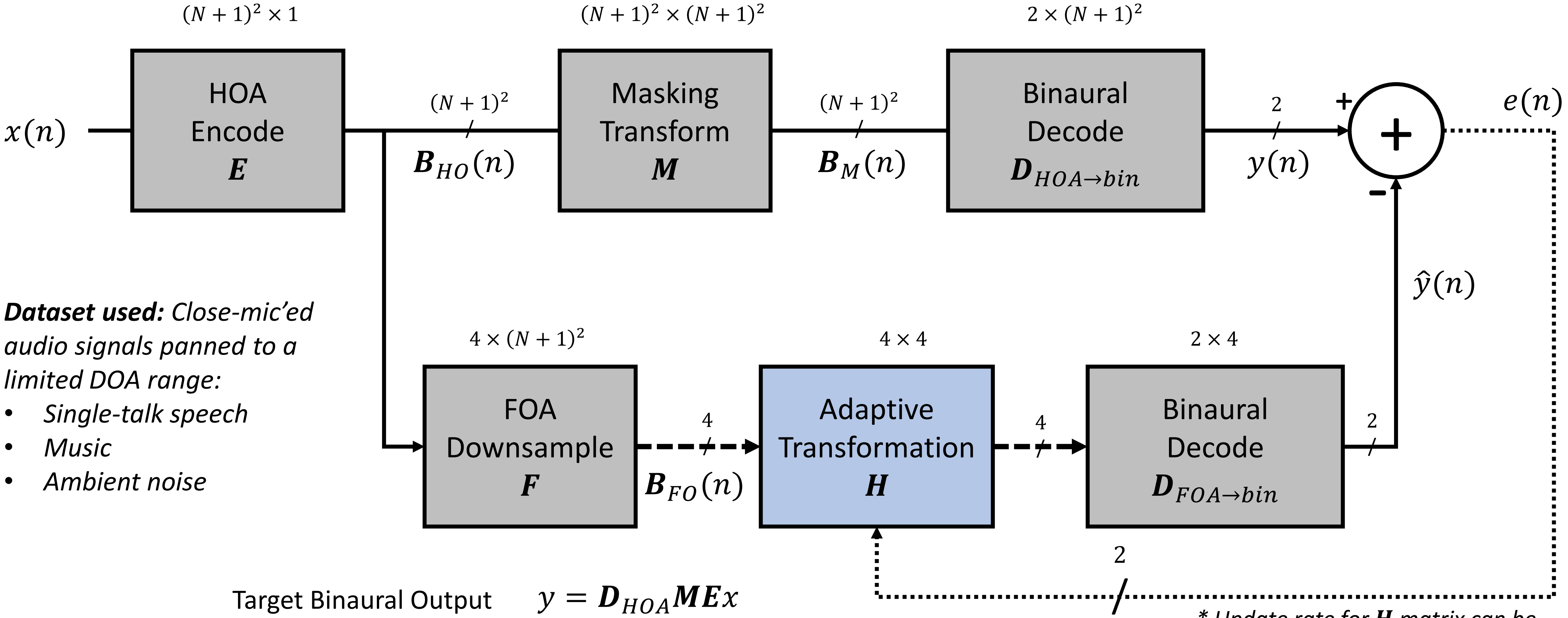
Spatial Masks

Lower image source: M. Chapman, Symmetries of Spherical Harmonics: Applications to Ambisonics

Experiment

Proposed Solution

$$\boldsymbol{H}^{(n+1)} = \boldsymbol{H}^{(n)} + \mu\big(x(n)\boldsymbol{F}\big)\big(e^*(n)\boldsymbol{D}_{FOA}^{\dagger}\big)$$



Dataset used: *Close-mic'ed audio signals panned to a limited DOA range:*

- *Single-talk speech*
- *Music*
- *Ambient noise*

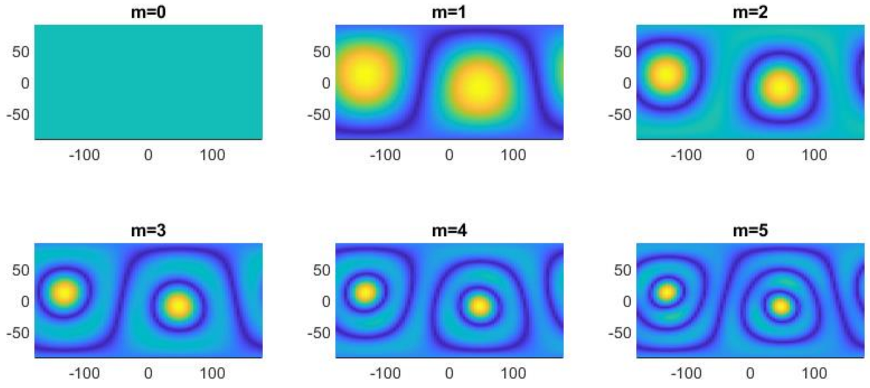
Target Binaural Output	$y = \boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x$
Predicted Binaural Output	$\hat{y} = \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x$
Binaural Error	$e = y - \hat{y}$

** Update rate for $\boldsymbol{H}$ matrix can be performed sample-wise or frame-wise*

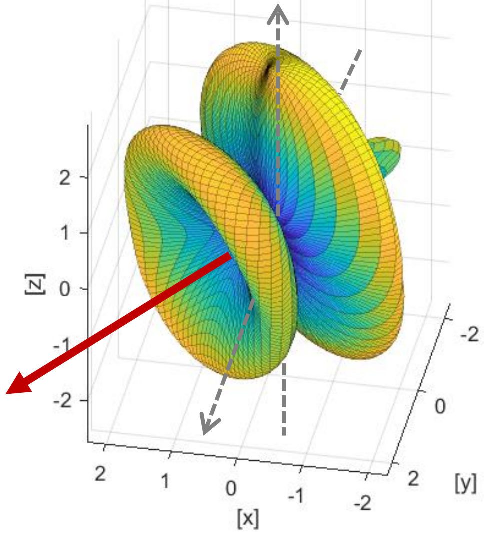
Experiment

Progress

Example: Three plane waves with DOA *$\boldsymbol{\theta} \in [\mathbf{30}, \mathbf{55}]^\circ, \boldsymbol{\delta} \in [-\mathbf{5}, \mathbf{0}]^\circ$ (weighted sums per order)*



(sum of orders)



Target Output $y = \boldsymbol{D}_{HOA}\boldsymbol{MEx}$

Predicted Output $\hat{y} = \boldsymbol{D}_{FOA}\boldsymbol{HFEx}$

Binaural Error $e = y - \hat{y}$

$$
\begin{aligned}
&= \boldsymbol{D}_{HOA}\boldsymbol{ME}x - \boldsymbol{D}_{FOA}\boldsymbol{HFE}x \\
&= (\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x \\
&= ((\boldsymbol{D}_H + \boldsymbol{D}_F)\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x
\end{aligned}
$$

$$
\begin{aligned}
&= (\boldsymbol{D}_{FOA}\boldsymbol{M} + \boldsymbol{D}_H\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x \\
&= (\boldsymbol{D}_H\boldsymbol{M} - \boldsymbol{D}_{FOA}(\boldsymbol{M} - \boldsymbol{HF}))\boldsymbol{E}x \\
&= (\boldsymbol{D}_H\boldsymbol{M} - \boldsymbol{D}_{FOA}(\boldsymbol{M} - \boldsymbol{HF}))\boldsymbol{E}x \qquad \to \min_H(\boldsymbol{M} - \boldsymbol{HF})
\end{aligned}
$$

Completed:

- Create framework for encoding directional signals into HOA
- Find dataset for training stimulus
- Plot encoded outputs to visualize soundfield quality

In Progress:

- Explore convergence with time-varying spatial distributions
- Explore higher-dimensional norm in FOA domain
- Try a higher-dimensional error using $\boldsymbol{D}_{FOA}^{-1}$ or Tikhonov regularized $\boldsymbol{D}_{FOA}^{\dagger}$

Cost Function

$$
\begin{aligned}
J_s \triangleq e^H e &= \left((\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x\right)^{\mathrm{H}}(\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x \\
&= x^H \boldsymbol{E}^H (\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})^{\mathrm{H}}(\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x \\
&= x^H \boldsymbol{E}^H \left((\boldsymbol{D}_{HOA}\boldsymbol{M})^{\mathrm{H}} - (\boldsymbol{D}_{FOA}\boldsymbol{HF})^{\mathrm{H}}\right)(\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{HF})\boldsymbol{E}x \\
&= x^H \boldsymbol{E}^H \left((\boldsymbol{D}_{HOA}\boldsymbol{M})^{\mathrm{H}}(\boldsymbol{D}_{HOA}\boldsymbol{M}) - (\boldsymbol{D}_{HOA}\boldsymbol{M})^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HF}) - (\boldsymbol{D}_{FOA}\boldsymbol{HF})^{\mathrm{H}}(\boldsymbol{D}_{HOA}\boldsymbol{M}) + (\boldsymbol{D}_{FOA}\boldsymbol{HF})^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HF})\right)\boldsymbol{E}x \\
&= x^H \boldsymbol{E}^H \left(\cancel{(\boldsymbol{D}_{HOA}\boldsymbol{M})^{\mathrm{H}}(\boldsymbol{D}_{HOA}\boldsymbol{M})} - 2(\boldsymbol{D}_{HOA}\boldsymbol{M})^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HF}) + (\boldsymbol{D}_{FOA}\boldsymbol{HF})^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HF})\right)\boldsymbol{E}x \\
&= x^H \boldsymbol{E}^H \left((\boldsymbol{D}_{FOA}\boldsymbol{HF})^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HF}) - 2(\boldsymbol{D}_{HOA}\boldsymbol{M})^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HF})\right)\boldsymbol{E}x \\
&= (\boldsymbol{D}_{FOA}\boldsymbol{HFE}x)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x) - 2(\boldsymbol{D}_{HOA}\boldsymbol{ME}x)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x) \\
&= (\boldsymbol{D}_{FOA}\boldsymbol{HFE}x)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x) - 2\left(([\boldsymbol{D}_{HOA[0,1]}] + [\boldsymbol{D}_{HOA[2,N]}])\boldsymbol{ME}x\right)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x) \\
&= (\boldsymbol{D}_{FOA}\boldsymbol{HFE}x)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x) - 2\left([\boldsymbol{D}_{HOA[0,1]}]\boldsymbol{ME}x\right)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x) + 2\left([\boldsymbol{D}_{HOA[2,N]}]\boldsymbol{ME}x\right)^{\mathrm{H}}(\boldsymbol{D}_{FOA}\boldsymbol{HFE}x)
\end{aligned}
$$

Quadratic form

$$
\boldsymbol{D}_{HOA} = \Bigg(\underbrace{\begin{bmatrix} d_{L(0,0)} \\ d_{R(0,0)} \end{bmatrix}\begin{bmatrix} d_{L(1,-1)} & d_{L(1,0)} & d_{L(1,1)} \\ d_{R(1,-1)} & d_{R(1,0)} & d_{R(1,1)} \end{bmatrix}}_{[\boldsymbol{D}_{HOA[0,1]}]}\underbrace{\begin{bmatrix} d_{L(2,-2)} & \cdots & d_{L(2,2)} \\ d_{R(2,-2)} & \cdots & d_{R(2,2)} \end{bmatrix}\cdots\begin{bmatrix} d_{L(N,-N)} & \cdots & d_{L(N,N)} \\ d_{R(N,-N)} & \cdots & d_{R(N,N)} \end{bmatrix}}_{[\boldsymbol{D}_{HOA[2,N]}]}\Bigg) \triangleq [\boldsymbol{D}_{HOA[0,1]}] + [\boldsymbol{D}_{HOA[2,N]}]
$$

$$
[\boldsymbol{D}_{HOA[0,1]}] = \begin{bmatrix} \alpha\boldsymbol{D}_{FOA} & \begin{matrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \end{matrix} \end{bmatrix} \quad s.t.\ \ \alpha \in \mathbb{R}
$$

$$
\begin{aligned}
ee^{H} &= (\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F})\boldsymbol{E}x\big((\boldsymbol{D}_{HOA}\boldsymbol{M} - \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F})\boldsymbol{E}x\big)^{H} \\
&= (\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x - \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)\ \big((\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x - \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)\big)^{H} \\
&= (\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x - \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)\big((\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x)^{H} - (\boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)^{H}\big) \\
&= (\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x - \boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)\big((\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x)^{H} - (\boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)^{H}\big) \\
&= (\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x)(\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x)^{H} - (\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x)(\boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)^{H} - (\boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)(\boldsymbol{D}_{HOA}\boldsymbol{M}\boldsymbol{E}x)^{H} + (\boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)(\boldsymbol{D}_{FOA}\boldsymbol{H}\boldsymbol{F}\boldsymbol{E}x)^{H}
\end{aligned}
$$

$$
ee^{H} \in \mathcal{F}^{2\times 2}
$$