Analysis and Prediction of Protected-Ear Localization

Griffin D. Romigh, Ph.D.

Air Force Research Labs

2610 Seventh St., Area B, Bldg. 441

WPAFB, OH 45433

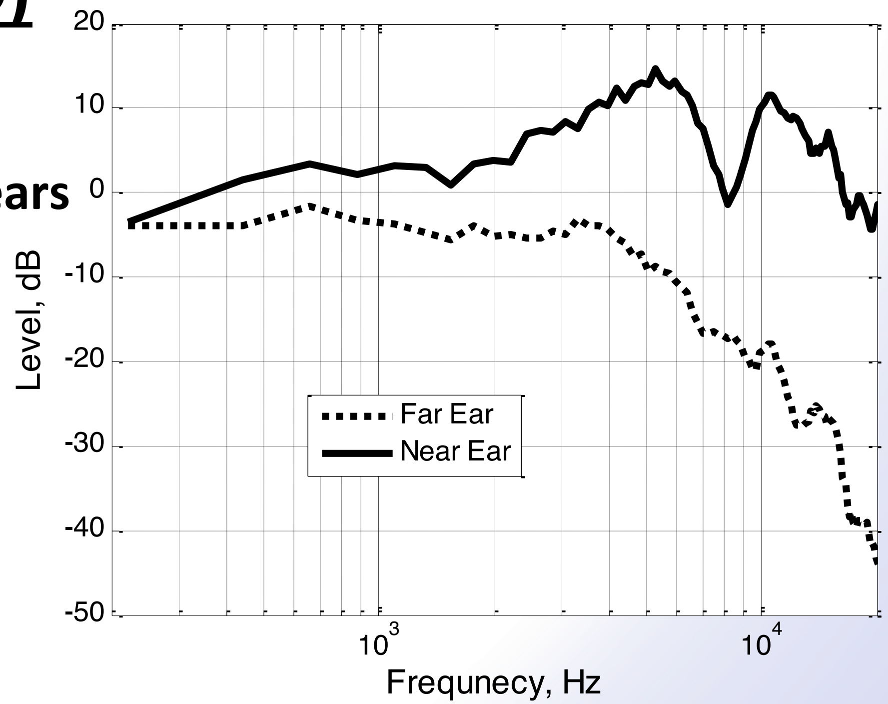
Overview

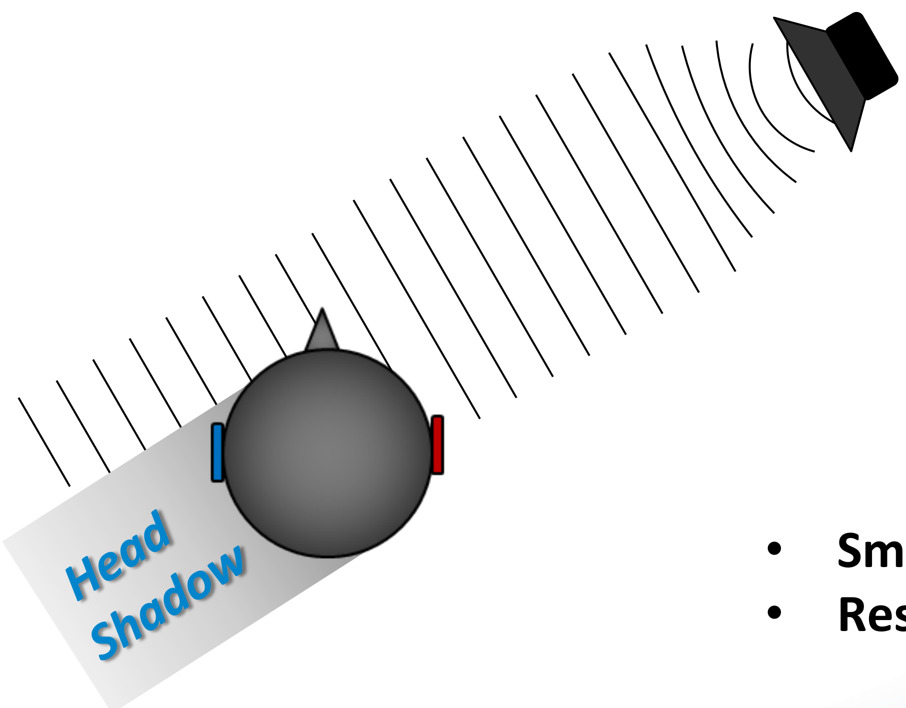
- Spatial Hearing and HRTFs
- A Different Approach
- An Efficient Representation
- Applying Bayesian estimation
- Modeling individual differences
- Summary of Contributions

Spatial Hearing

Interaural Level Difference (ILD)

- Sound energy is scattered by the head
- Less energy arrives at the far ear
- Results in a level difference at the two ears

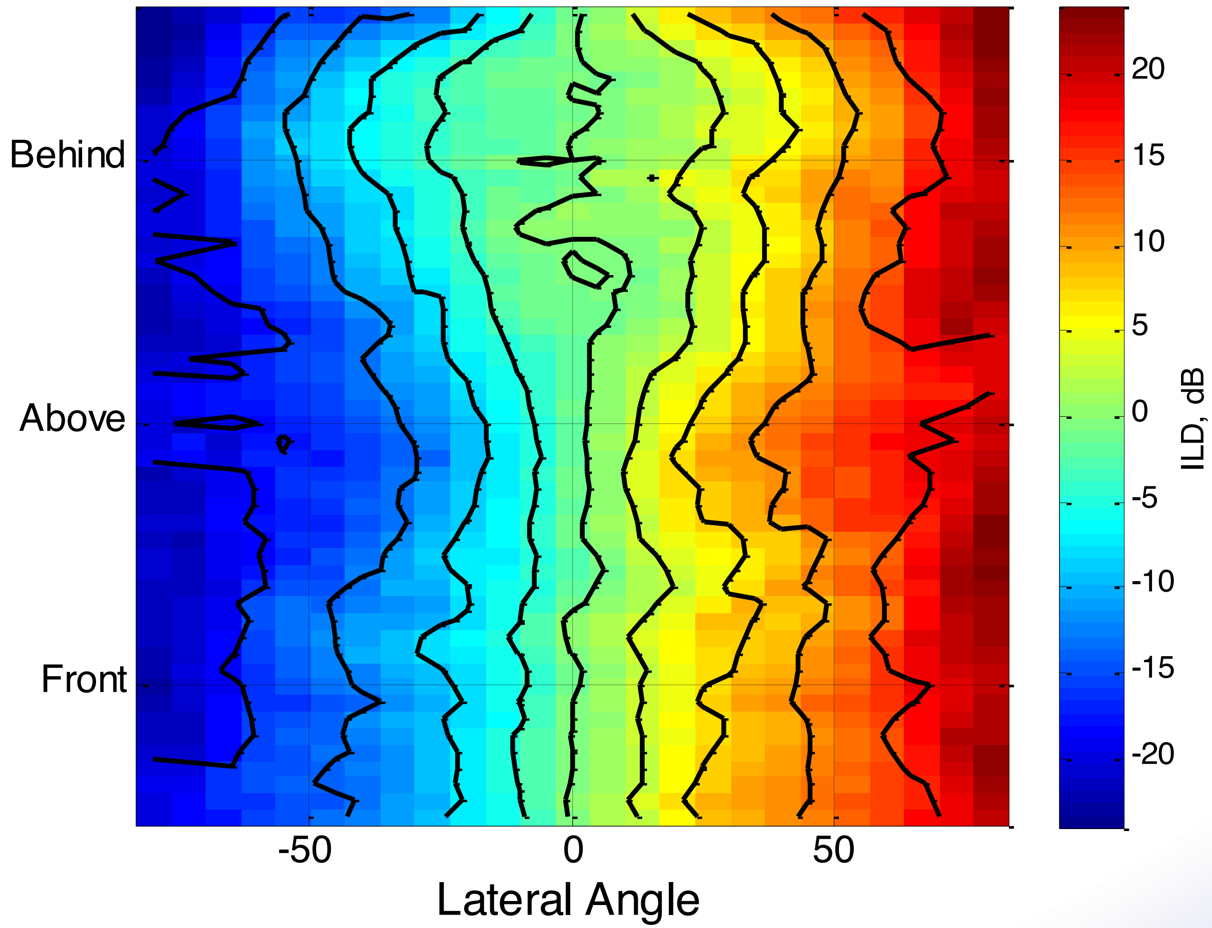


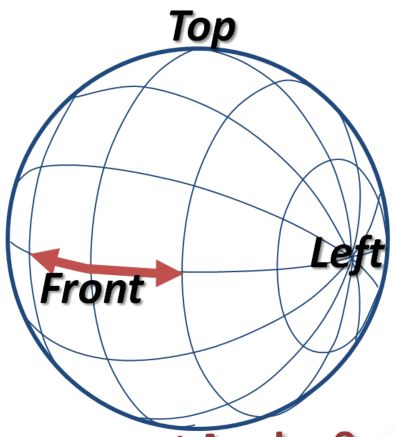


- Smaller wavelengths are attenuated more
- Results in larger ILDs at high frequencies

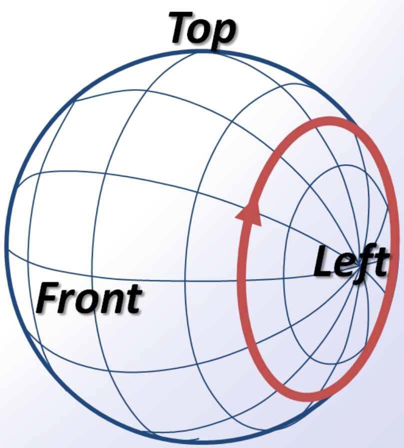
Spatial Hearing

Interaural Level Difference (ILD)





Lateral Angle, θ



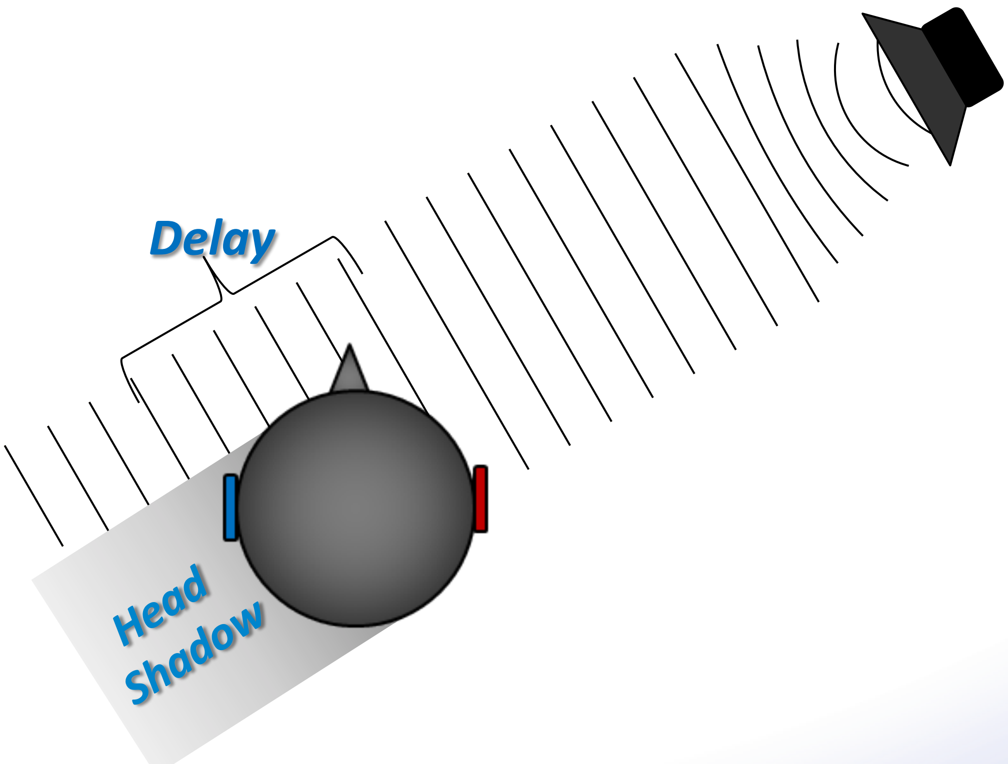
Intraconic Angle, ϕ

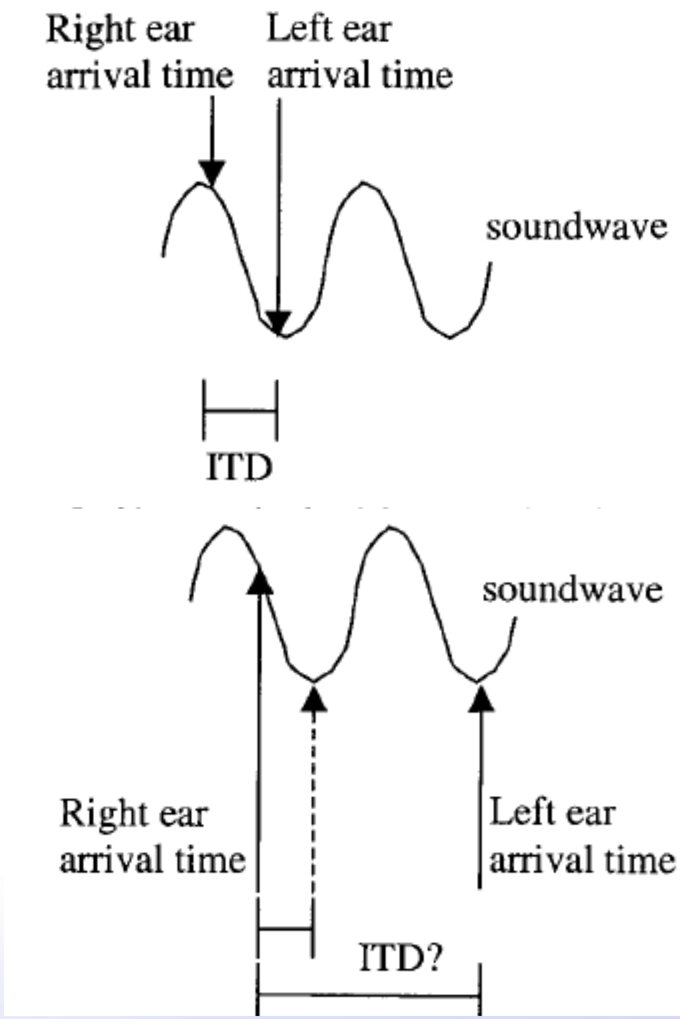
4

Spatial Hearing

Interaural Time Difference (ITD)

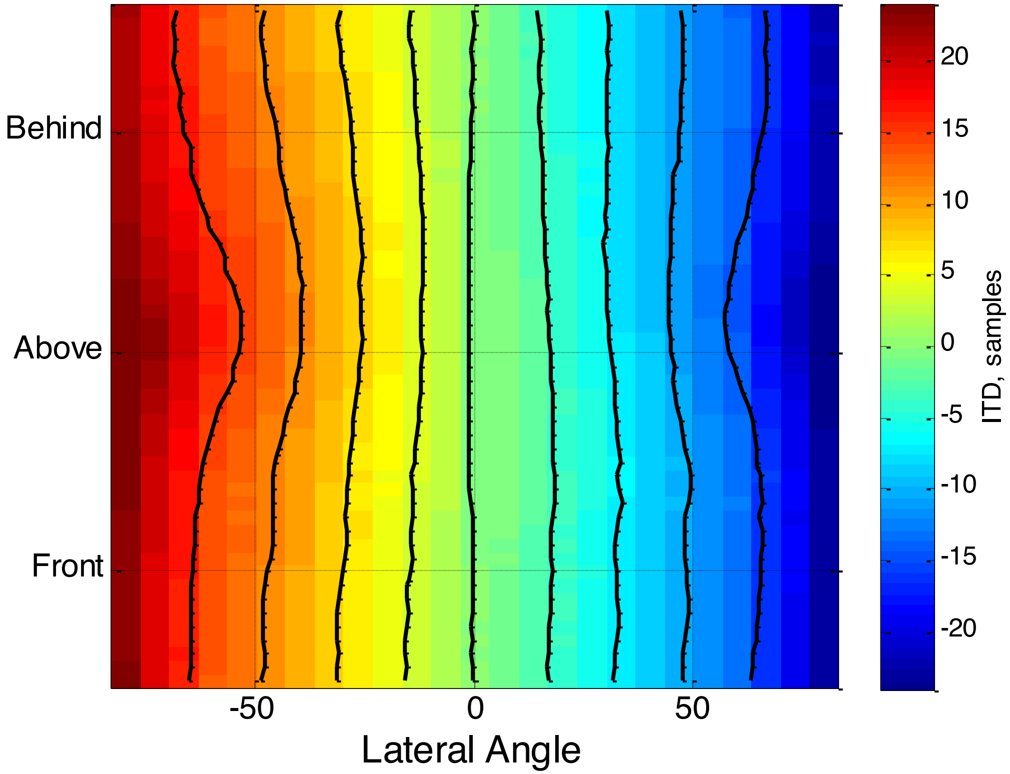
- Sound arrives at near ear before far ear
- Results in a arrival and phase difference
- Becomes ambiguous at high frequencies

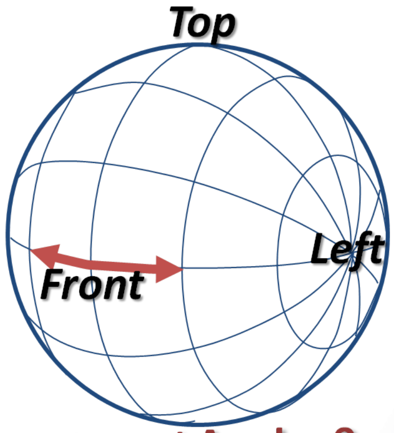




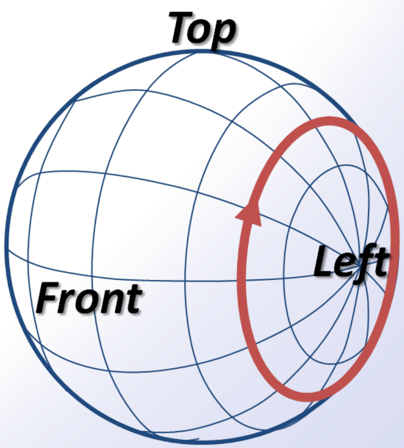
Spatial Hearing

Interaural Time Difference (ITD)





Lateral Angle, θ

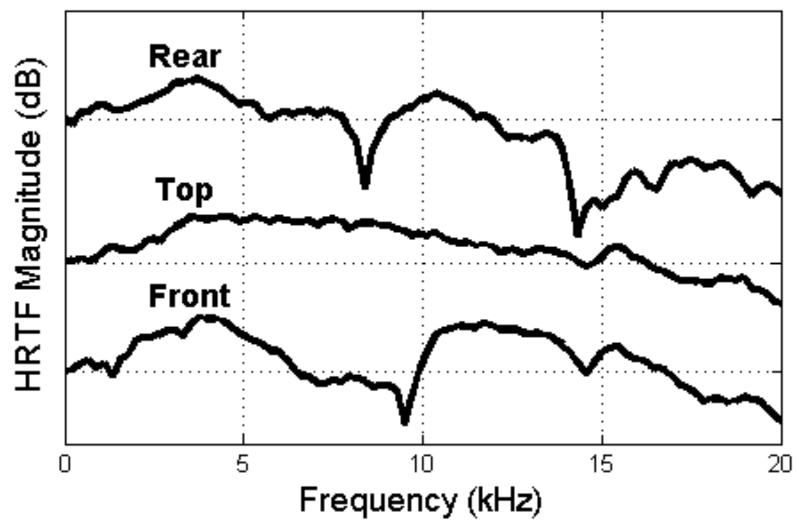


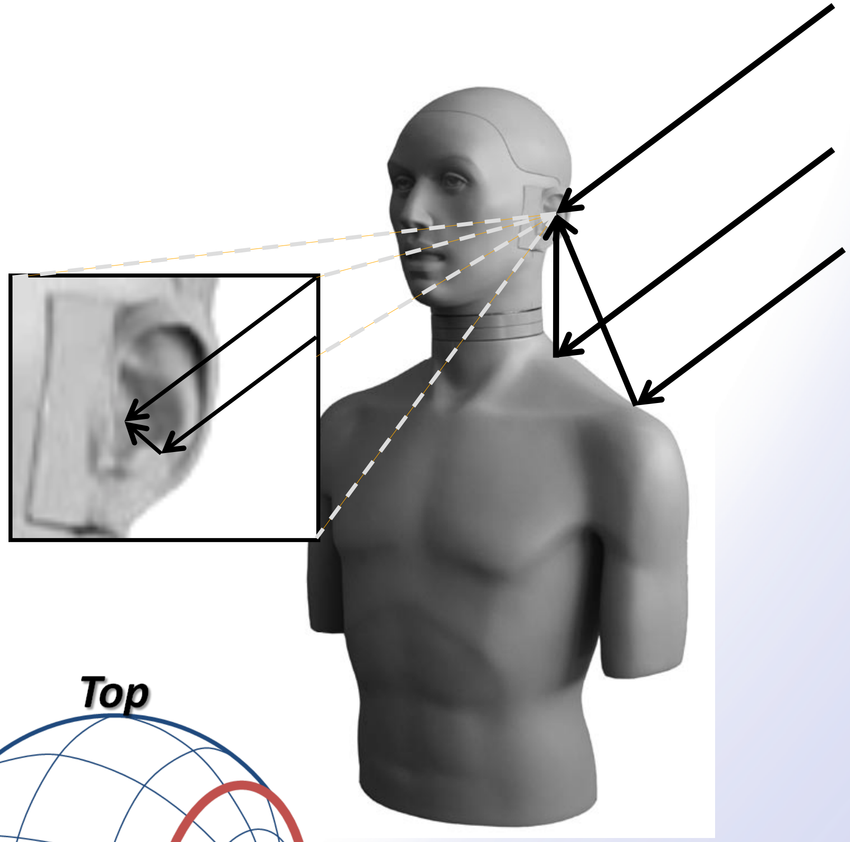
Intraconic Angle, ϕ

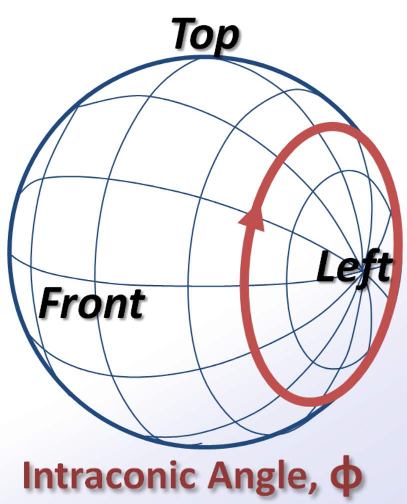
Spatial Hearing

Spectral Cues

- High frequency cues due to pinna
- Lower frequency cue due to shoulders
- Perceptually weighted to favor closer ear



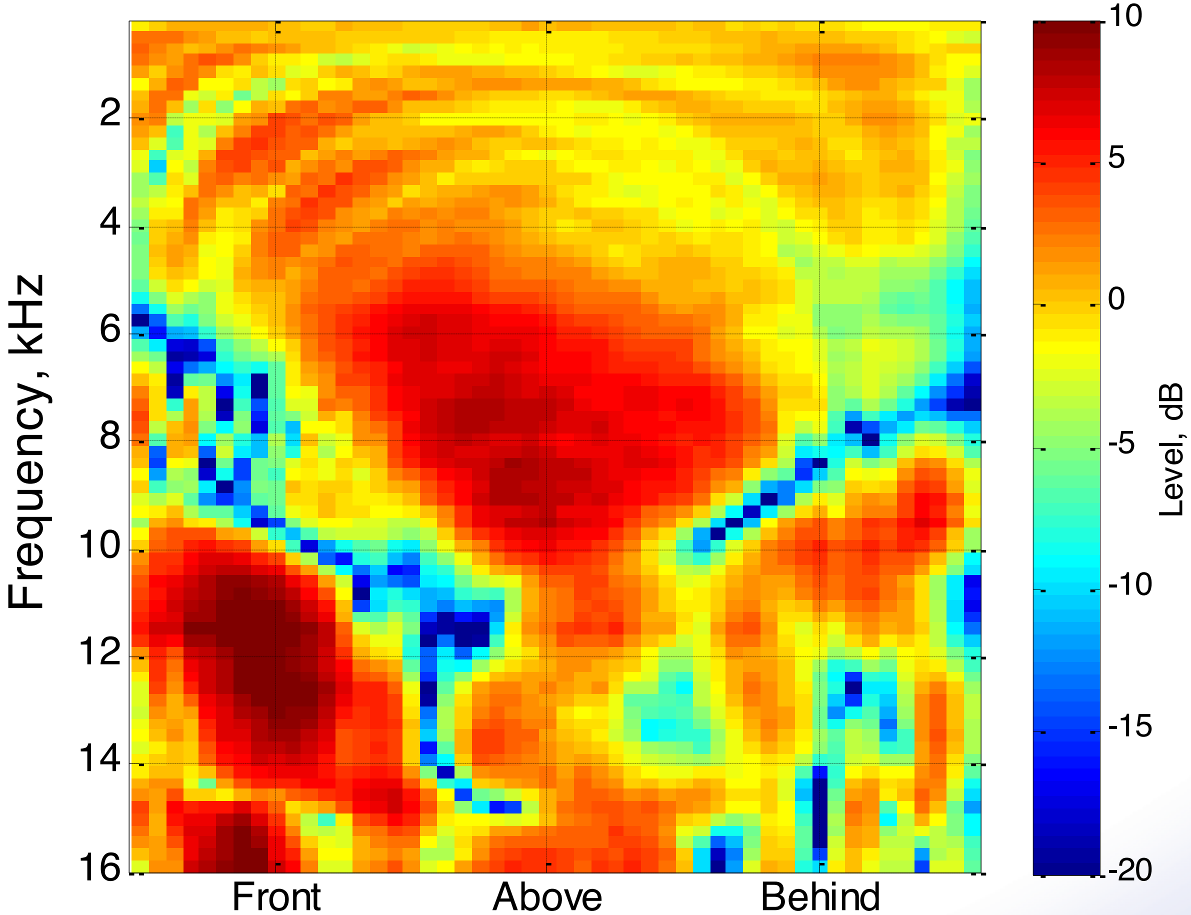


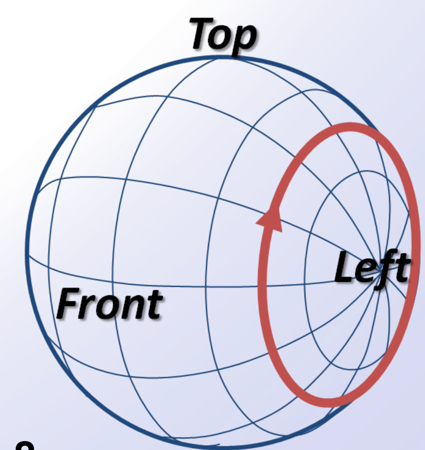


7

Spatial Hearing

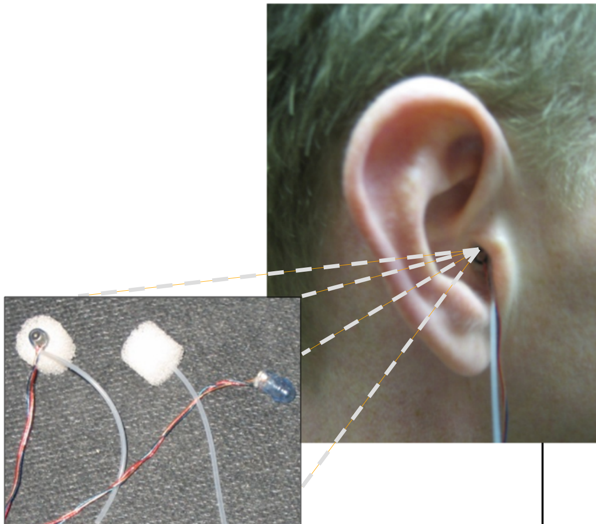
Spectral Cues

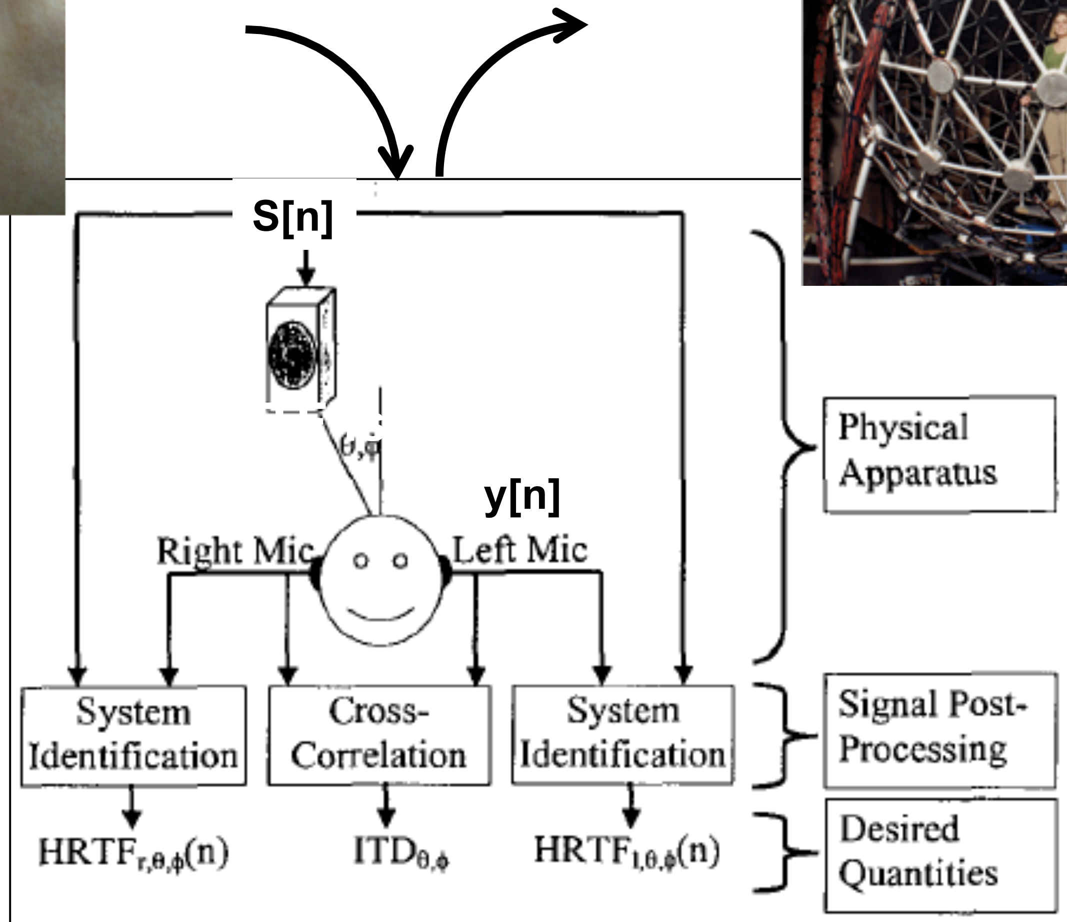




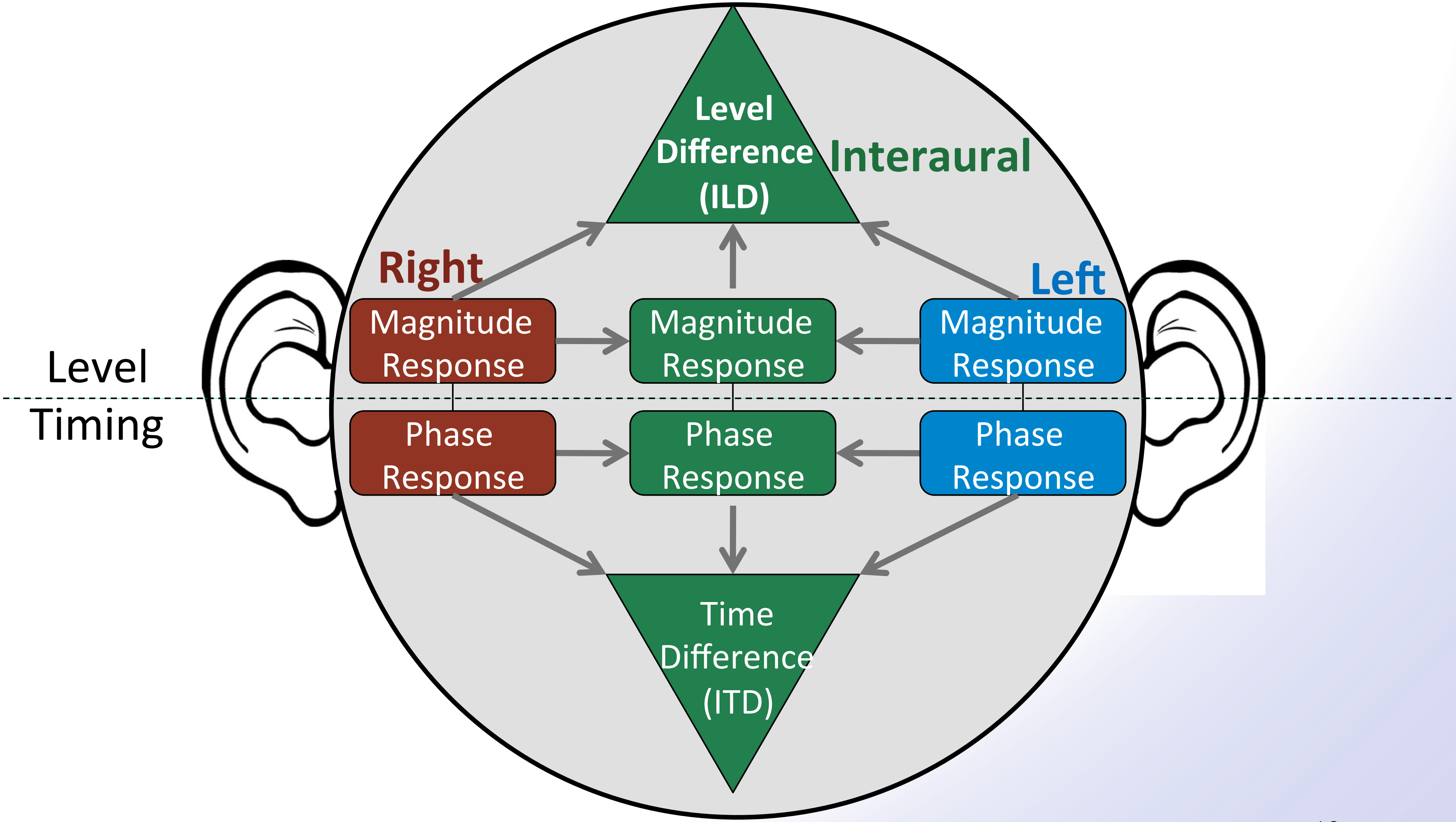
8 Intraconic Angle, ϕ

Head-Related Transfer Functions:



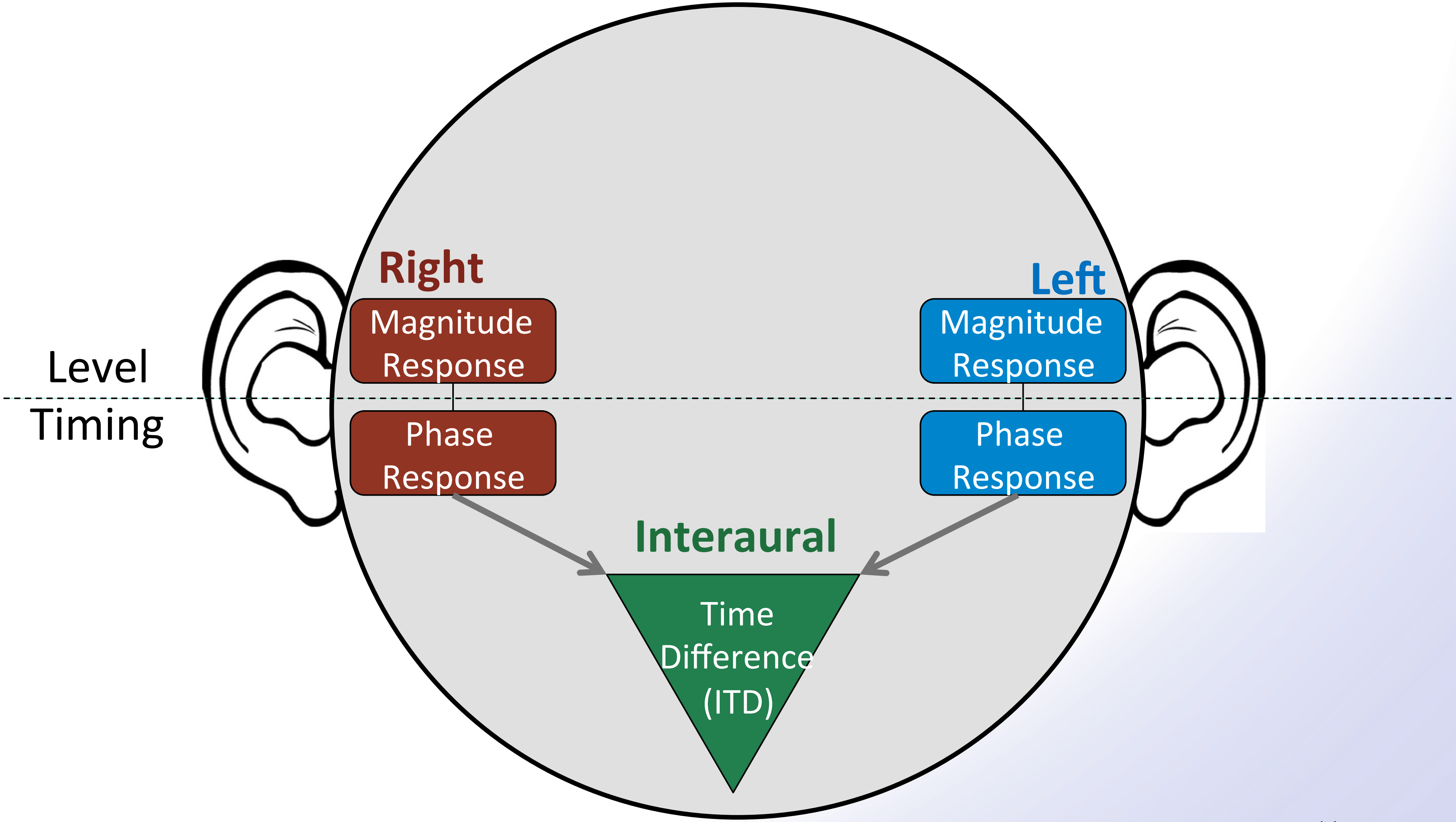


HRTF ← → Spatial Hearing Cues



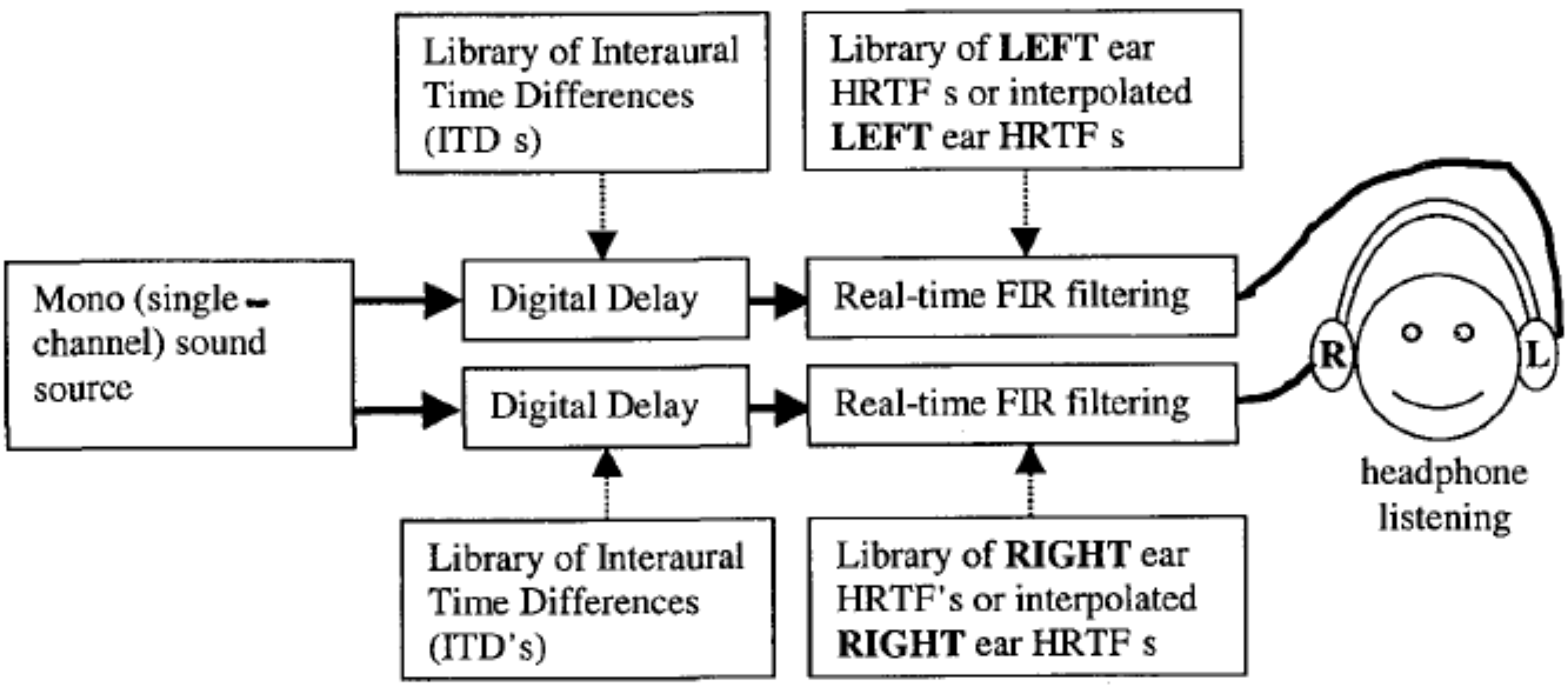
10

HRTF ← → Spatial Hearing Cues



11

Binaural Synthesis



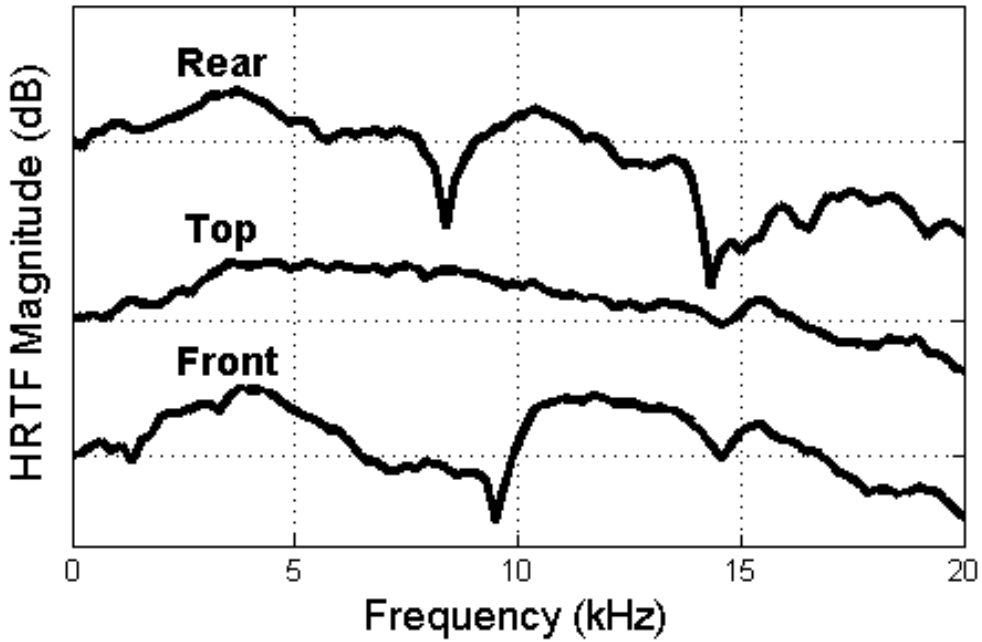
Spatial Auditory Displays:

<u>Spatial Auditory Displays</u>

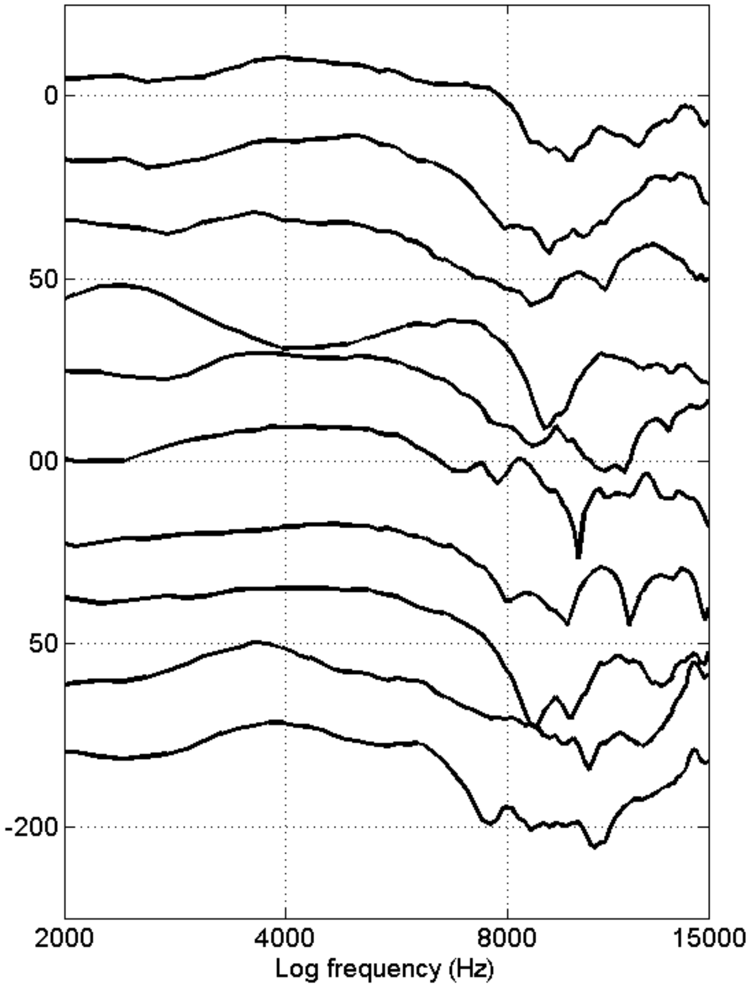
- Guidance systems
- Hearing Restoration
- Virtual Reality
- Augmented Reality

HRTFs: *Idiosyncrasy*

By location

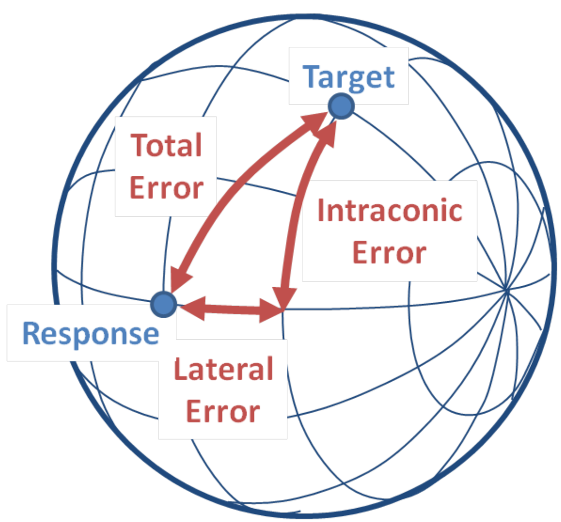


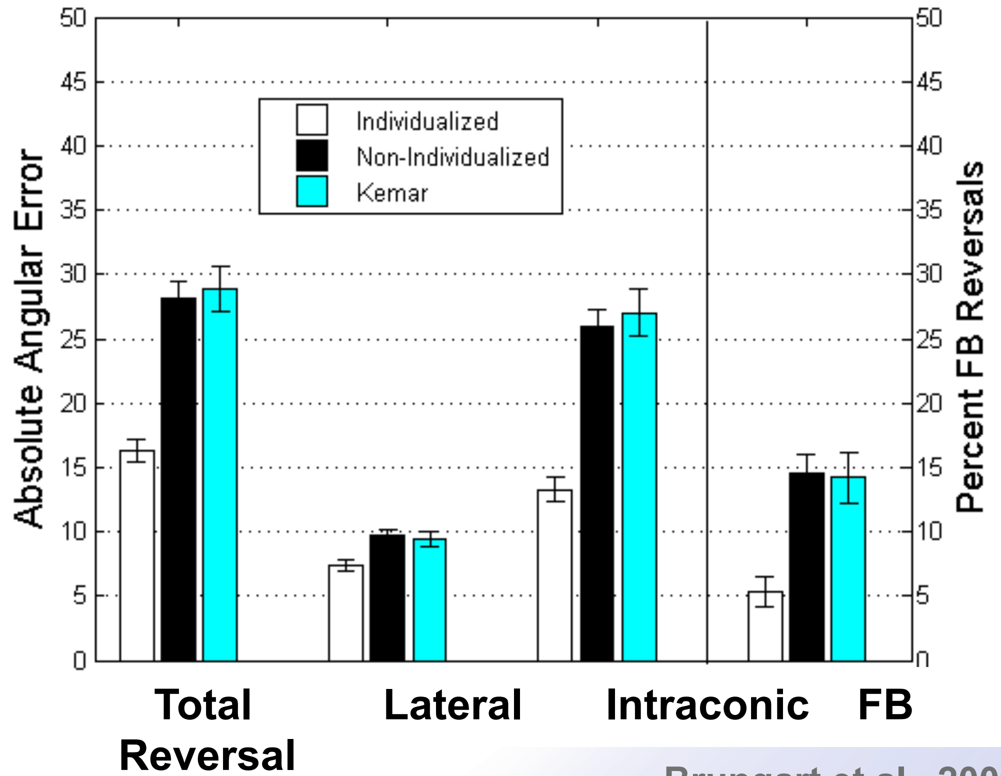
By individual



HRTFs: *Idiosyncrasy*

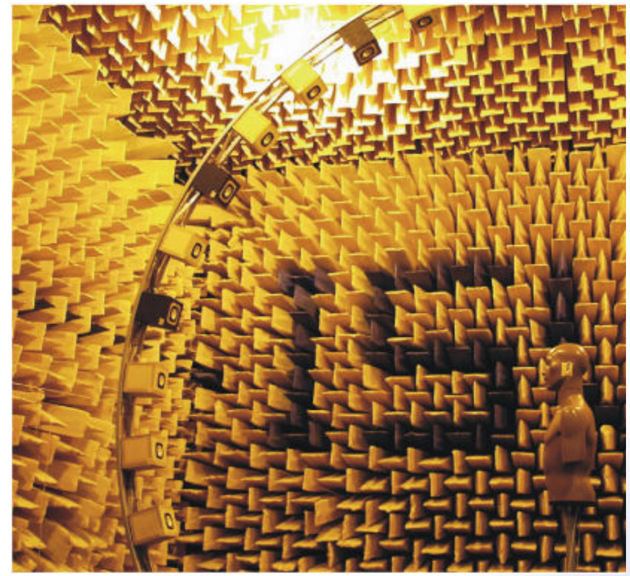
- **SADs need Individual HRTFs**
- **Otherwise:**
  1. **No sense of elevation**
  2. **Frequent FB Reversals**
  3. **Localized "In the Head"**





- Brungart et al., 2009

HRTFs: *Spatial Measurement*

	Fixed Spherical Array	**Rotating Arc Array**
		
Pros:	**Fast (5 – 10 min)**	**Cheaper, Temporary**
Cons:	**Expensive, Permanent**	**Slow (1 – 2 hours)**

The Problem

How can we get an HRTF for every spatial angle with as few physical measurements as possible?

Previous Methods:

Parallel

Measurement
-Reciprocity -Spectral asynchrony

- Same Equipment
- Less Time
- Perceptually Equivalent Performance

Baseline HRTF

- 277 locations
- 256 taps

Non-Acoustic Interpolation

Naive	Statistical
-Linear kNN -Spherical Basis	- Pattern Matching - Neural Net

- Less Equipment
- Less Time
- Perceptually Equivalent Performance

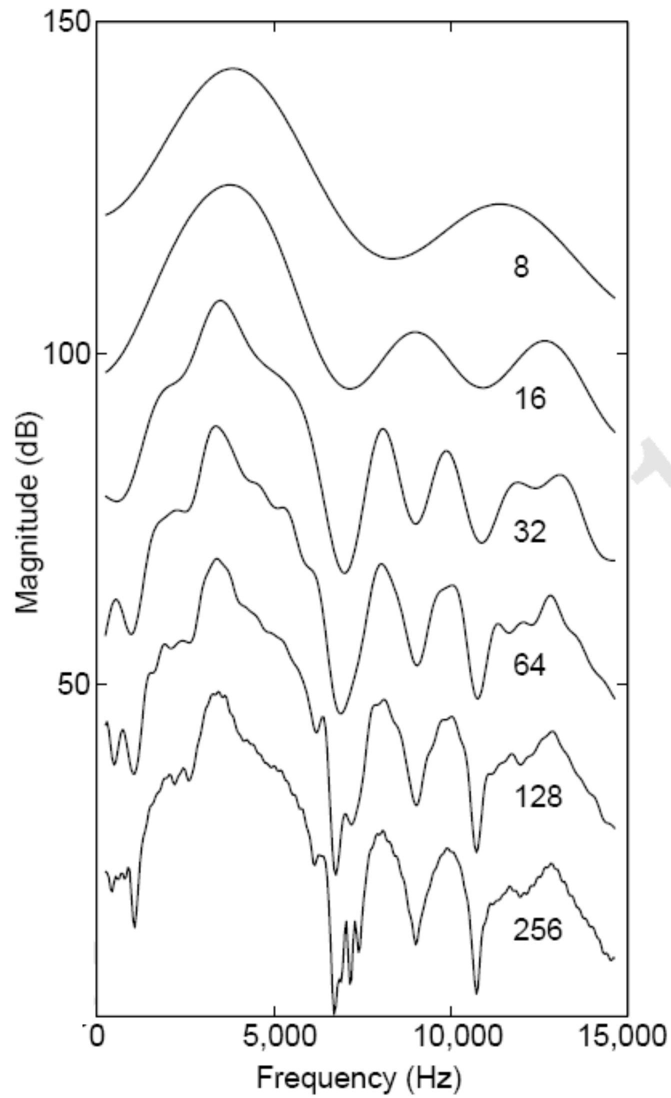
Subjective Selection	Structural Models	Generalization
-Most Externalized -Vertical Lift	-“Snowman” - Anthropometric	-Averaging -Super Subject

- Least Equipment
- Less Time
- Poor Performance

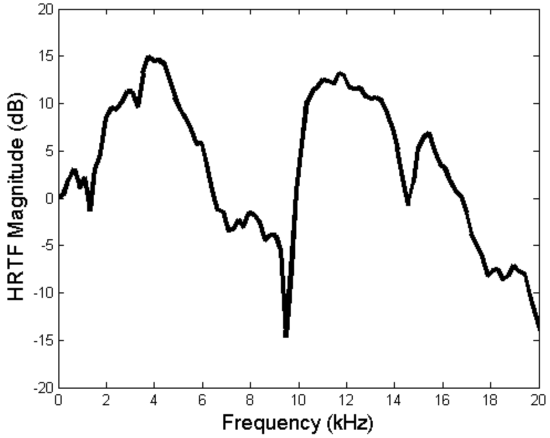
Irrelevant Spectral Details

- **Auditory system has limited spectral resolution**
- **This results in fine spectral details being averaged out**
- **Most impactful at high frequencies**

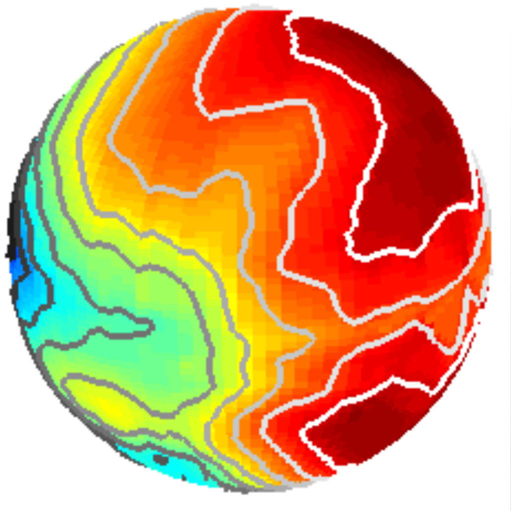
Maybe we can get away with smoothing the spatial detail



Spatial Representation:



$$|H_{\theta,\phi}(\omega)| \Leftrightarrow |H_{\omega}(\theta,\phi)|$$

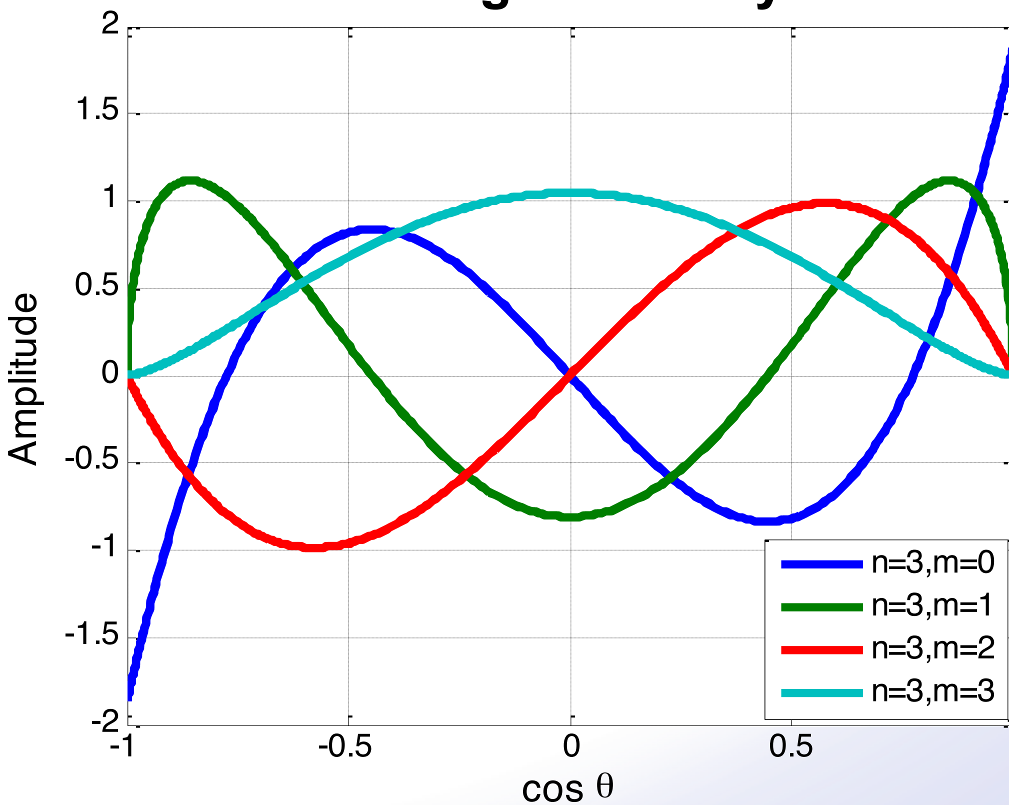


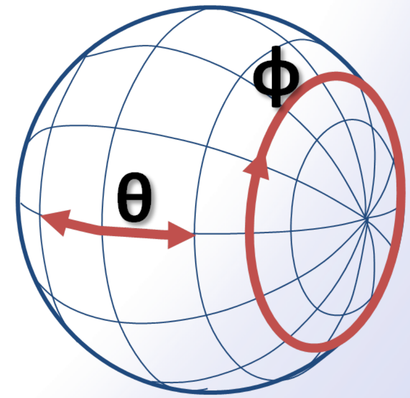
20

Spherical Harmonics

$$Y_{nm}(\phi, \theta) = \begin{cases} N_n^m P_n^m(\cos\theta)\cos(m\phi) & \text{if } m \geq 0 \\ N_n^m \underbrace{P_n^{|m|}(\cos\theta)}\sin(m\phi) & \text{if } m < 0 \end{cases}$$

Associated Legendre Polynomials

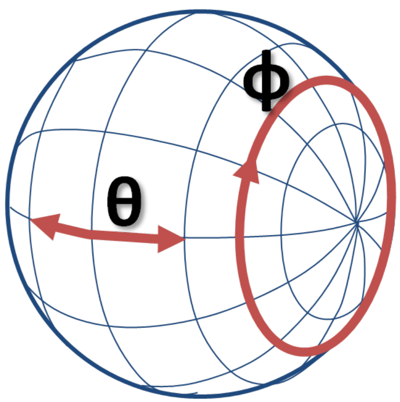




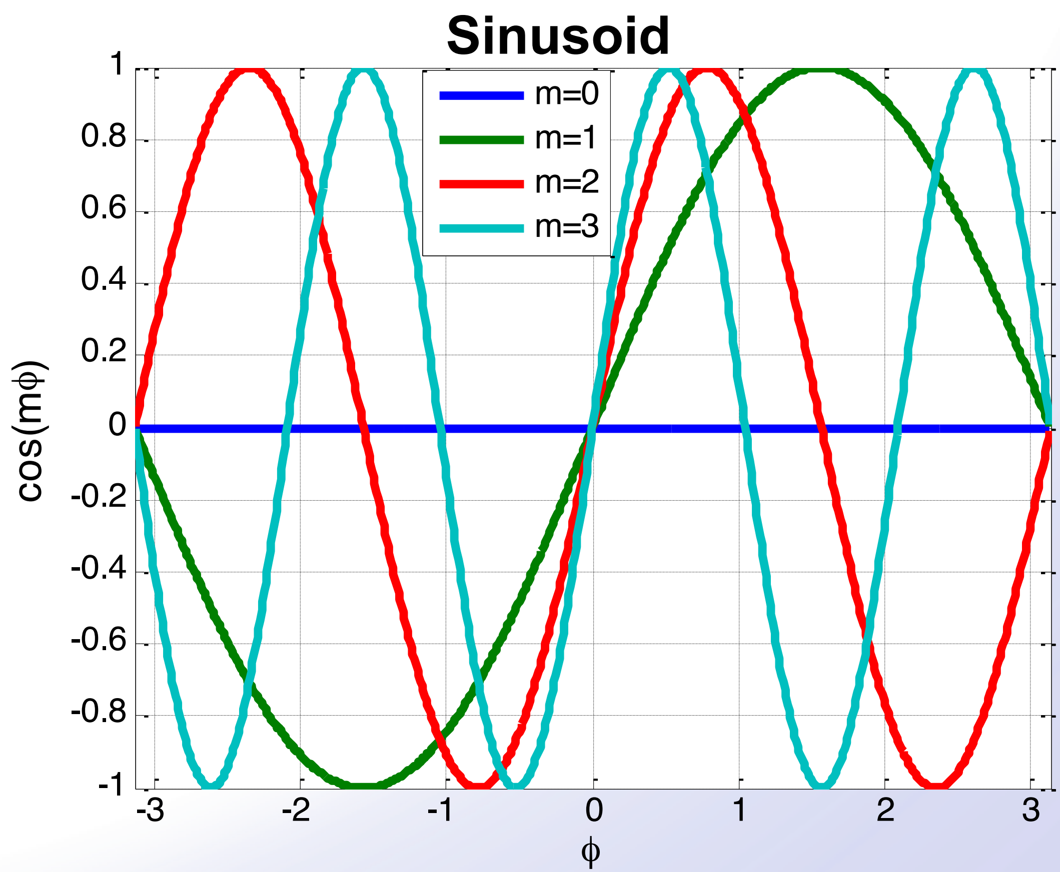
Orthogonal functions for lateral angle

Spherical Harmonics

$$Y_{nm}(\phi,\theta) = \begin{cases} N_n^m P_n^m(\cos\theta)\cos(m\phi) & \text{if } m \geq 0 \\ N_n^m P_n^{|m|}(\cos\theta)\underbrace{\sin(m\phi)} & \text{if } m < 0 \end{cases}$$



Orthogonal functions
in intraconic angle



Spherical Harmonics

Orthonormal basis over the continuous sphere

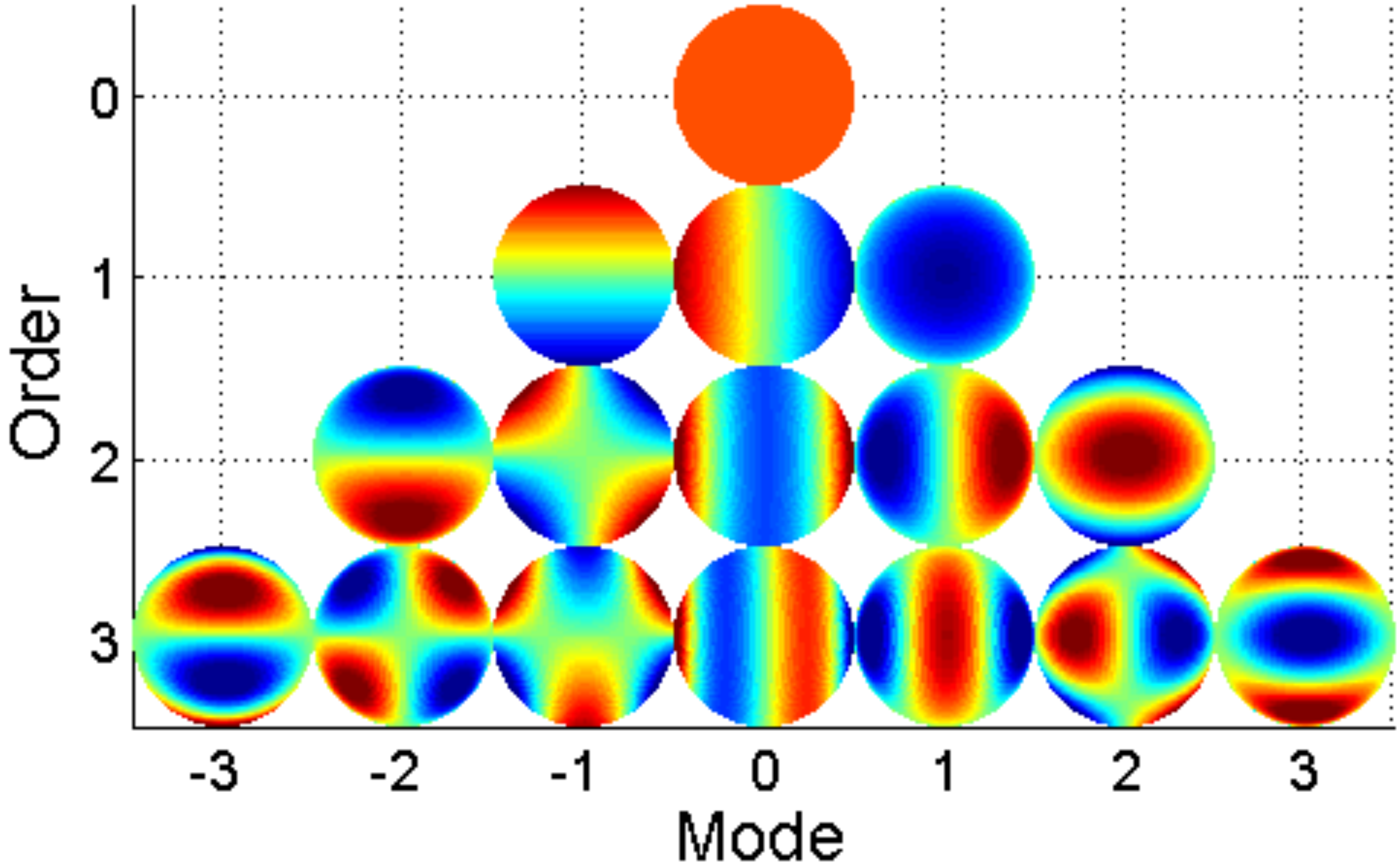
$$\int_{\theta=-\pi/2}^{\pi/2} \int_{\varphi=-\pi}^{\pi} Y_{nm}(\varphi,\theta)^{*} \, Y_{n'm'}(\varphi,\theta) = \delta_{nn'} \, \delta_{mm'}$$

Allow us to represent any square integrable spherical function with a set of SH coefficients

$$f(\phi,\theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_{nm}(\phi,\theta) C_{nm}$$

****** We can do Fourier analysis on a sphere *****

Spherical Harmonics



24

Practical SH Expansion

Re-cast problem into system of linear equations

$$\mathbf{f} = \mathbf{Y}\mathbf{c}$$

$$\text{where } \mathbf{f} = [f(\phi_0, \theta_0), f(\phi_1, \theta_1), \cdots, f(\phi_S, \theta_S)]^T$$

$$\mathbf{c} = [C_{00}, C_{1-1}, C_{10}, C_{11}, \cdots, C_{PP}]^T$$

$$\mathbf{Y} = \begin{bmatrix} Y_{00}(\phi_1, \theta_1) & Y_{-11}(\phi_1, \theta_1) & \cdots & Y_{PP}(\phi_1, \theta_1) \\ Y_{00}(\phi_2, \theta_2) & Y_{-11}(\phi_2, \theta_2) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ Y_{00}(\phi_S, \theta_S) & Y_{-11}(\phi_S, \theta_S) & \cdots & Y_{PP}(\phi_S, \theta_S) \end{bmatrix}$$

Truncation Order

# of samples

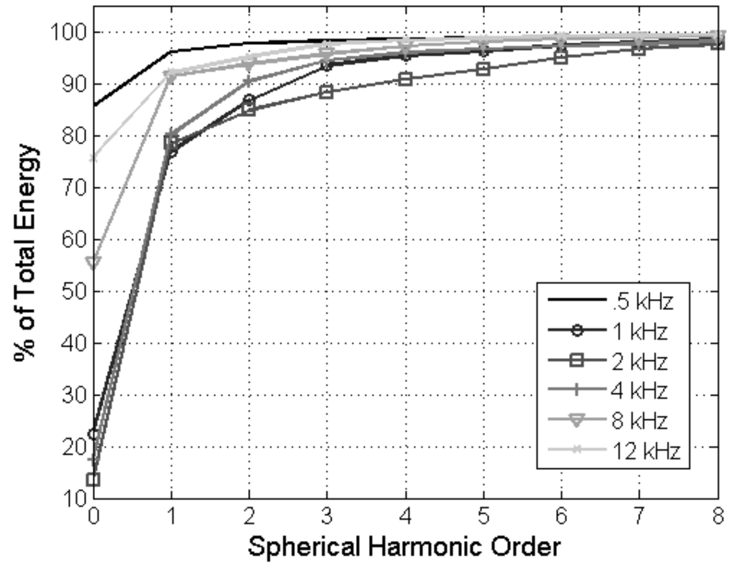
Simple least-squares solution

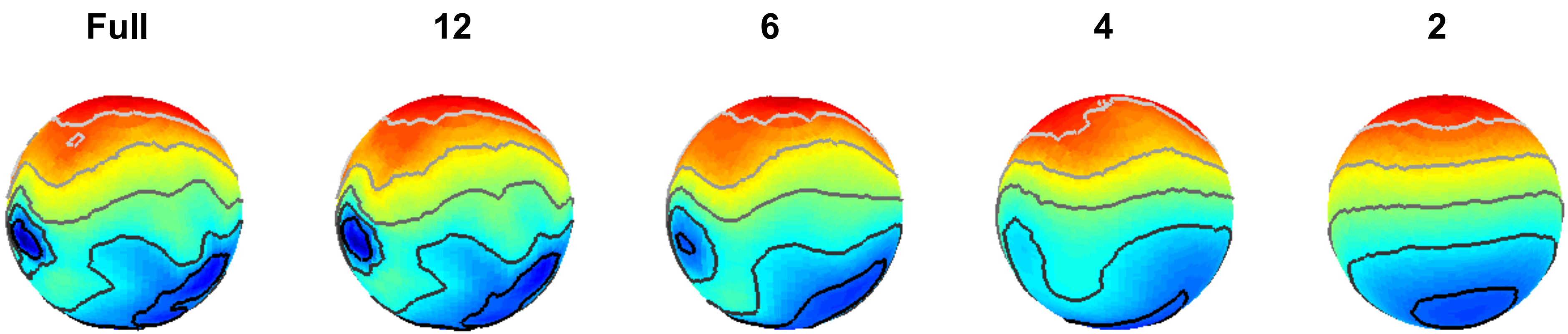
$$\hat{\mathbf{c}} = (\mathbf{Y}^{\mathbf{T}}\mathbf{Y})^{-1}\mathbf{Y}^{\mathbf{T}}\mathbf{f}$$

Spatial Smoothing:

Truncating the expansion provides spatial smoothing

$$f(\phi,\theta) = \sum_{n=0}^{P} \sum_{m=-n}^{n} Y_{nm}(\phi,\theta) C_{nm}$$

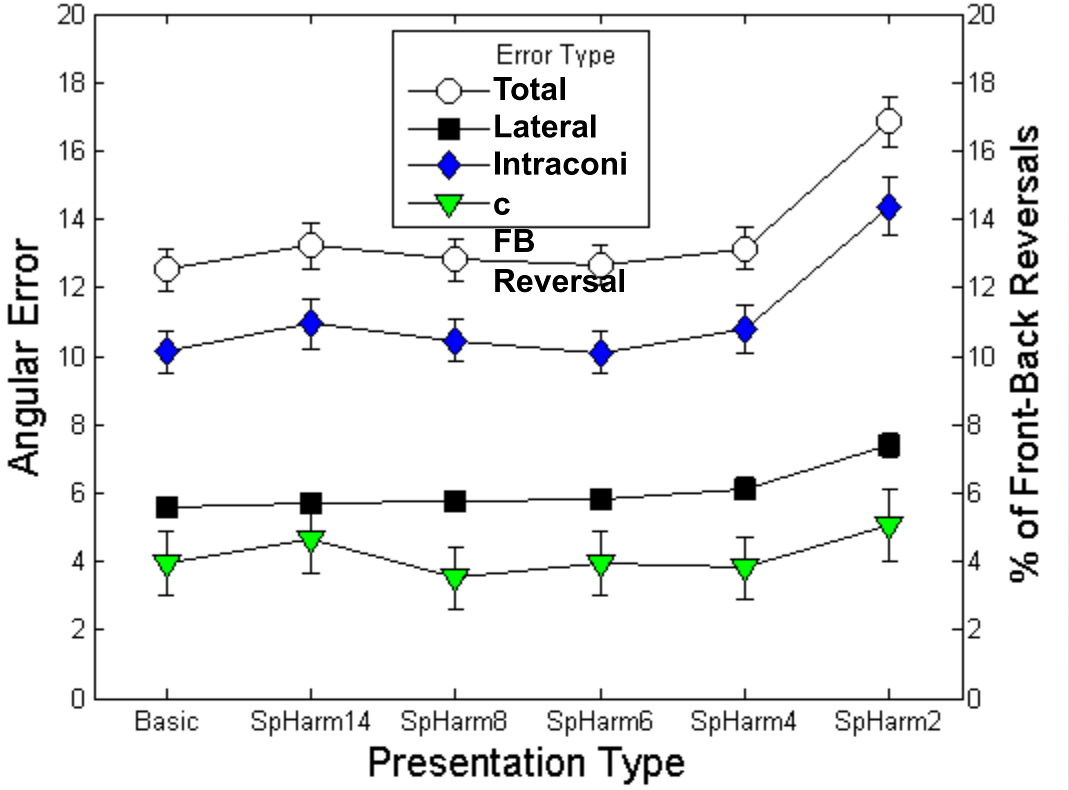


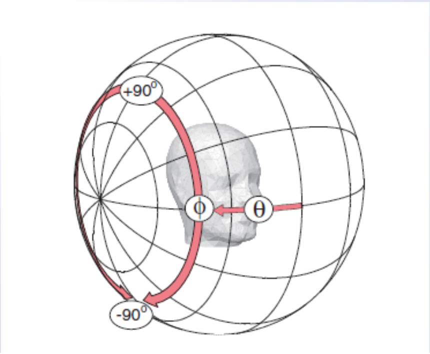


Perceptual Evaluation

<u>Localization task</u>

- 8 Subjects
- 250-ms noise bursts
- 245 locations





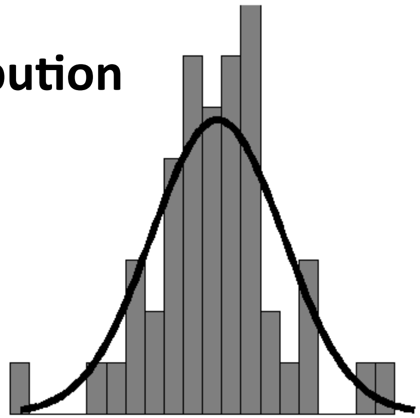
Recap…

- **New SH-based HRTF representation**
  - Spatially continuous
  - Reduces irrelevant spatial variation
  - Localization equivalent to full HRTF
  - Reduces # of parameters by 95% w.r.t. baseline HRTF

Can non-individualized HRTFs provide information about a new HRTF measurement?

Bayesian HRTF Model

Model all HRTFs as belonging to the same underlying distribution



$$\mathbf{f} = \mathbf{Y}\mathbf{c} + \mathbf{n}$$

$$\mathbf{c} : \mathcal{N}(\mathbf{m_c}, \mathbf{R_{cc}})$$

$$\mathbf{n} : \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

} **Independent**

Non-individual information is incorporated through hyper-parameters

$$\mathbf{R_{cc}} = \begin{bmatrix} \sigma_{00}^2 & 0 & \cdots & 0 \\ 0 & \sigma^2_{-11} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{PP} \end{bmatrix}$$

$$\mathbf{m_c} = \begin{bmatrix} E[C_{00}] \\ E[C_{-11}] \\ \vdots \\ E[C_{PP}] \end{bmatrix}$$

Bayesian Estimation

Estimation via MMSE Estimator

$$\hat{\mathbf{c}} = \mathbf{E}[\mathbf{c}|\mathbf{f}] = \mathbf{m_c} + \mathbf{R_{cc}}\mathbf{Y}^{\mathbf{T}}(\mathbf{Y}\mathbf{R_{cc}}\mathbf{Y}^{\mathbf{T}} + \sigma^{\mathbf{2}}\mathbf{I})^{-\mathbf{1}}(\mathbf{f} - \mathbf{Y}\mathbf{m_c})$$

Estimated SH coefficients for individual

Difference between individual HRTF and average HRTF at measurement locations

Estimator is based on how the HRTF is different from average…

Bayesian Estimation

Estimation via MMSE Estimator

$$\hat{\mathbf{c}} = \mathbf{E}[\mathbf{c}|\mathbf{f}] = \mathbf{m_c} + \mathbf{R_{cc}}\mathbf{Y}^{\mathbf{T}}(\mathbf{Y}\mathbf{R_{cc}}\mathbf{Y}^{\mathbf{T}} + \sigma^{\mathbf{2}}\mathbf{I})^{-\mathbf{1}}(\mathbf{f} - \mathbf{Y}\mathbf{m_c})$$

Average SH coefficients

Innovations from individualized measurements (bias)

Assuming hyper-parameters are already known…

Estimating Hyper-parameters

We have fixed unknown model parameters….

$$\mathbf{c} : \mathcal{N}(\mathbf{m_c}, \mathbf{R_{cc}})$$

Classical Estimation (MVUB)

$$\hat{\mathbf{m}}_c = \frac{1}{M} \sum_{i=1}^{M} \mathbf{c_i}$$

$$\hat{\sigma}_j^2 = \frac{1}{M-1} \sum_{i=1}^{M} (\mathbf{c}_i[j] - \hat{\mathbf{m}}_c[j])^2$$

$$\mathbf{R_{cc}} = \begin{bmatrix} \sigma_{00}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{-11}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{PP}^2 \end{bmatrix}$$

Assuming we have M individuals SH coefficients…

Estimating Hyper-parameters

We have fixed unknown model parameters….

$$\mathbf{c} : \mathcal{N}(\mathbf{m_c}, \mathbf{R_{cc}})$$

But we can't measure SH coefficients. We need a way to estimate both simultaneously.

<u>Classical Estimation (MVUB)</u>

$$\hat{\mathbf{m}}_c = \frac{1}{M}\sum_{i=1}^{M} \mathbf{c_i}$$

$$\hat{\sigma}_j^2 = \frac{1}{M-1}\sum_{i=1}^{M} (\mathbf{c}_i[j] - \hat{\mathbf{m}}_c[j])^2$$

$$\mathbf{R_{cc}} = \begin{bmatrix} \sigma_{00}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{-11}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{PP}^2 \end{bmatrix}$$

Assuming we have M individuals' SH coefficients…

Expectation-Maximization

Compute parameters and hyper-parameters iteratively

1. **Initialize $\mathbf{R_{cc}}$ and $\mathbf{m_c}$ to arbitrary values**

2. **Calculate Bayesian estimates of SH coefficients**

$$\hat{\mathbf{c}} = \mathbf{m_c} + \mathbf{R_{cc}}\mathbf{Y^T}(\mathbf{Y}\mathbf{R_{cc}}\mathbf{Y^T} + \sigma^2\mathbf{I})^{-1}(\mathbf{f} - \mathbf{Y}\mathbf{m_c})$$

3. **Update estimates of $\mathbf{R_{cc}}$ and $\mathbf{m_c}$ using new coefficient values**

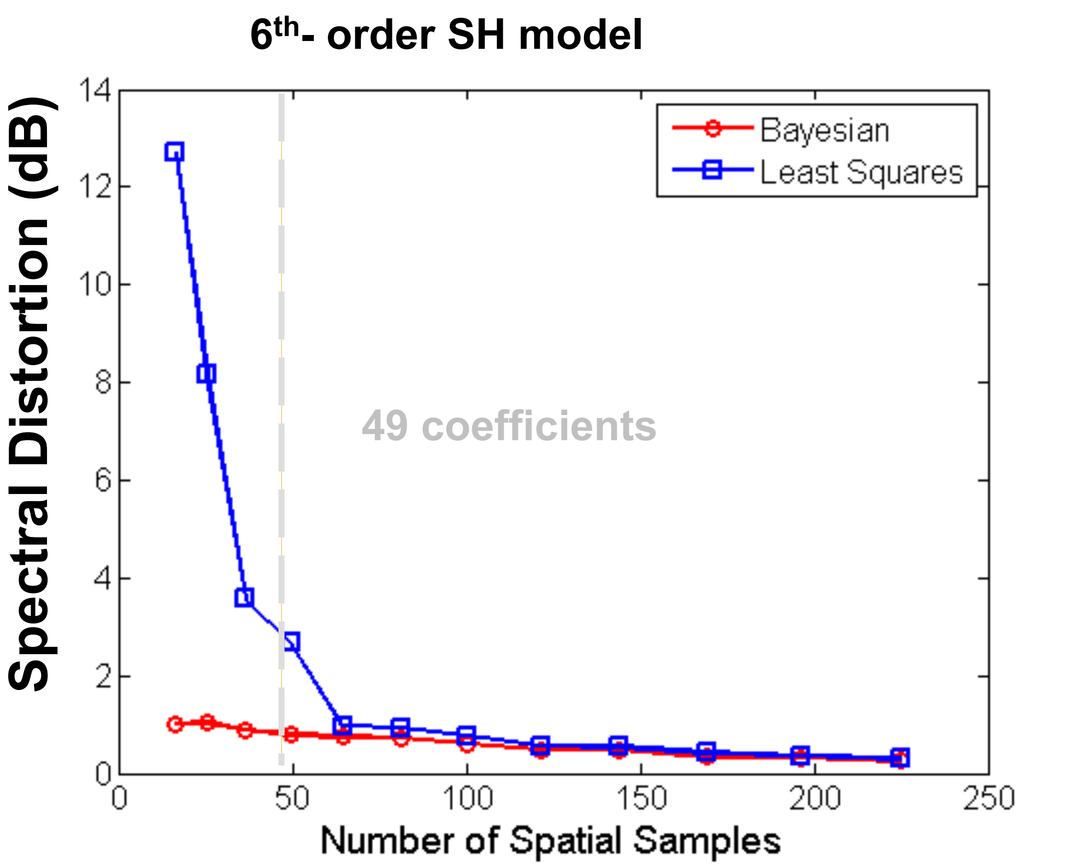
$$\hat{\mathbf{m}}_c = \frac{1}{M}\sum_{i=1}^{M}\mathbf{c_i} \qquad \hat{\sigma}_j^2 = \frac{1}{M-1}\sum_{i=1}^{M}(\mathbf{c}_i[j] - \hat{\mathbf{m}}_c[j])^2$$

4. **Repeat 2 and 3 until estimates converge**

Computational Performance

<u>Training the model</u>
- EM based
- 44 subjects
- 274 spatial samples

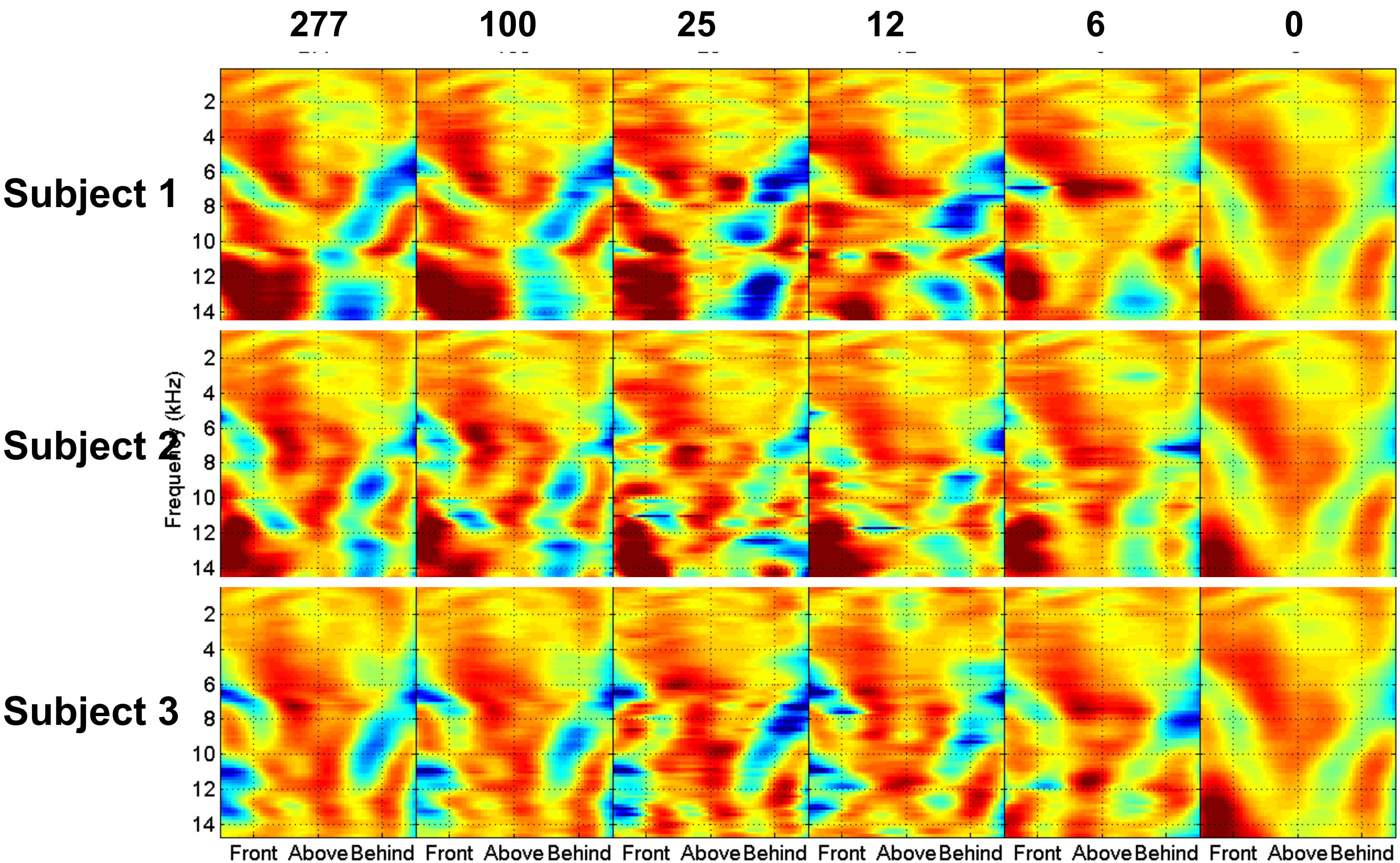
<u>Testing the model</u>
- Bayesian estimation
- 10 subjects
- varied # of samples



Better reconstruction performance with fewer spatial samples

35

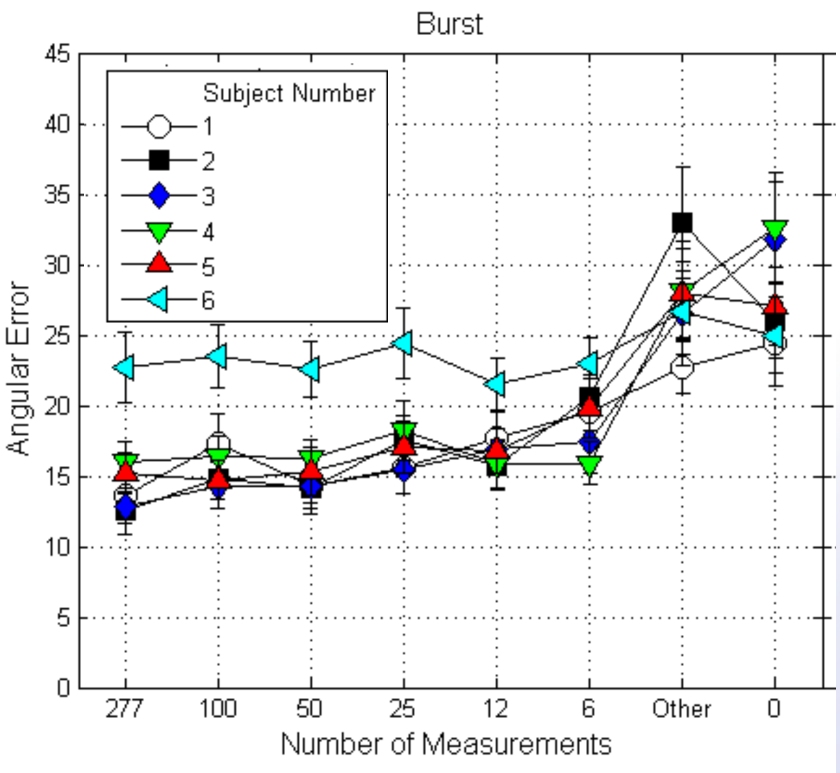
Computational Performance



36

Perceptual Evaluation

Localization Task
- **6 Subjects**
- **250-ms noise bursts**
- **245 locations**



Equivalent performance with as few as 12 measurements

37

Recap…

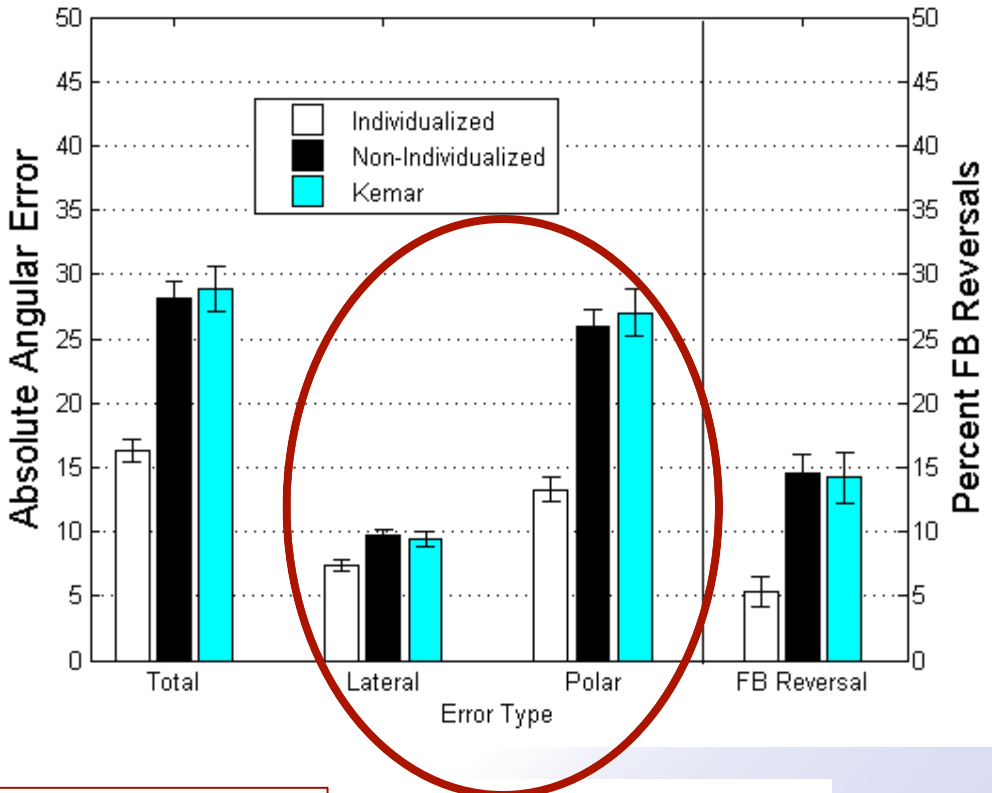
- **Bayesian HRTF model**
  - Models general HRTF distribution as MVN
  - Individualized HRTF represents a single sample

- **Bayesian HRTF Estimation**
  - Non-individualized HRTFs provide "template"
  - Individualized measurements personalize the template
  - Much fewer measurements are needed (~ 12 distributed)

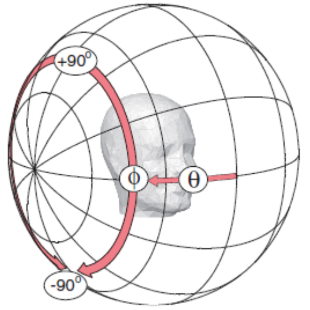
How do HRTFs differ amongst individuals?

Further Model Reduction

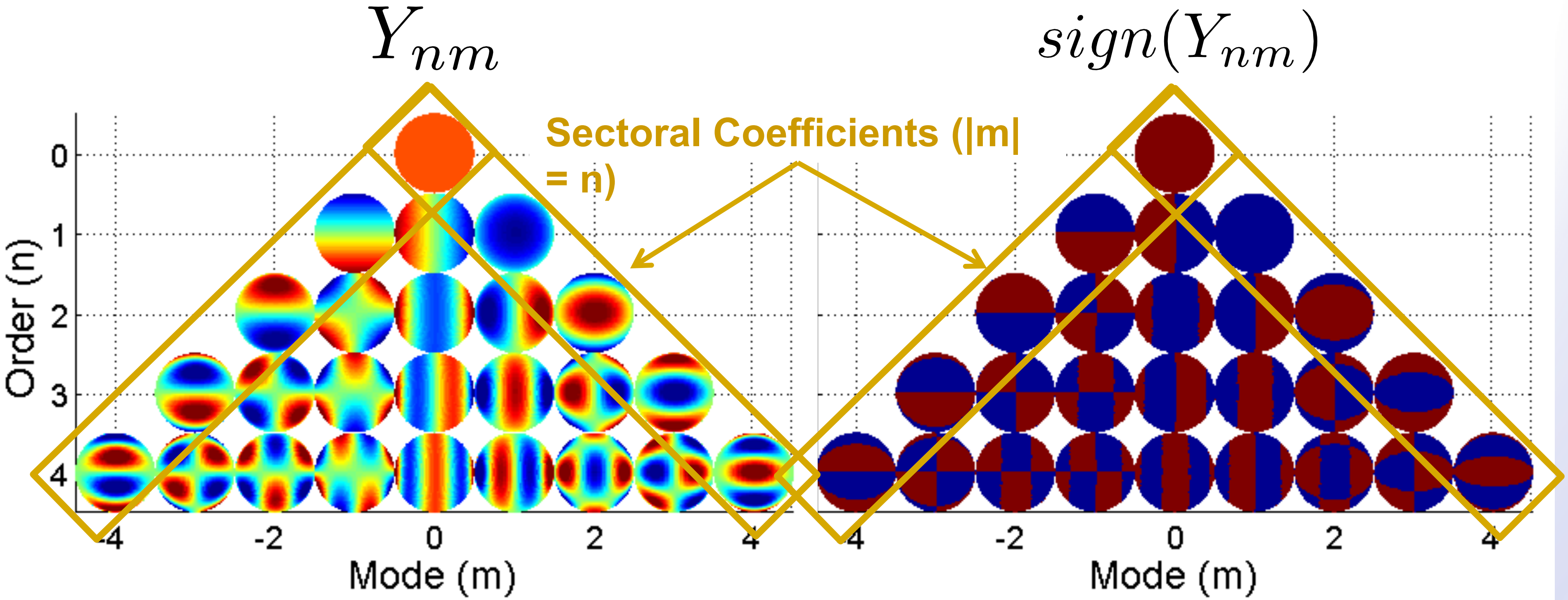
- **Non-individual localization is bad mostly in polar dimension**
- **Implies inter-subject differences in HRTFs account for polar cue difference**



If we can separate out polar cues we might only need to estimate those!



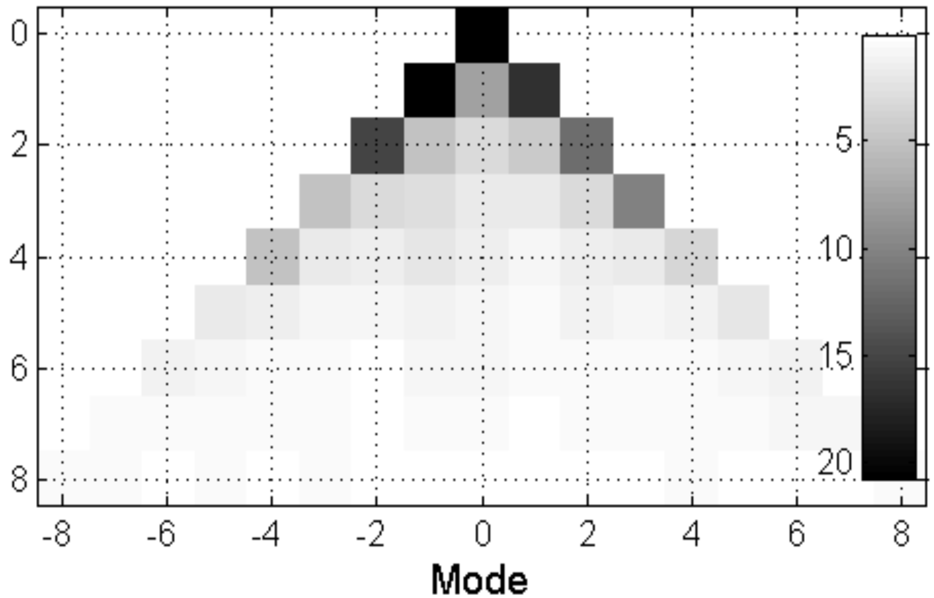
Further Model Reduction



Sectoral coefficients capture mostly intraconic variation

Further Model Reduction

Inter-subject Variance



Sectoral coefficients contain most of the inter-subject variance

These coefficients may be all that need to be individualized

41

Sectoral HRTF Model:

Separate individual (Sectoral) and non-individual (Lateral) features.

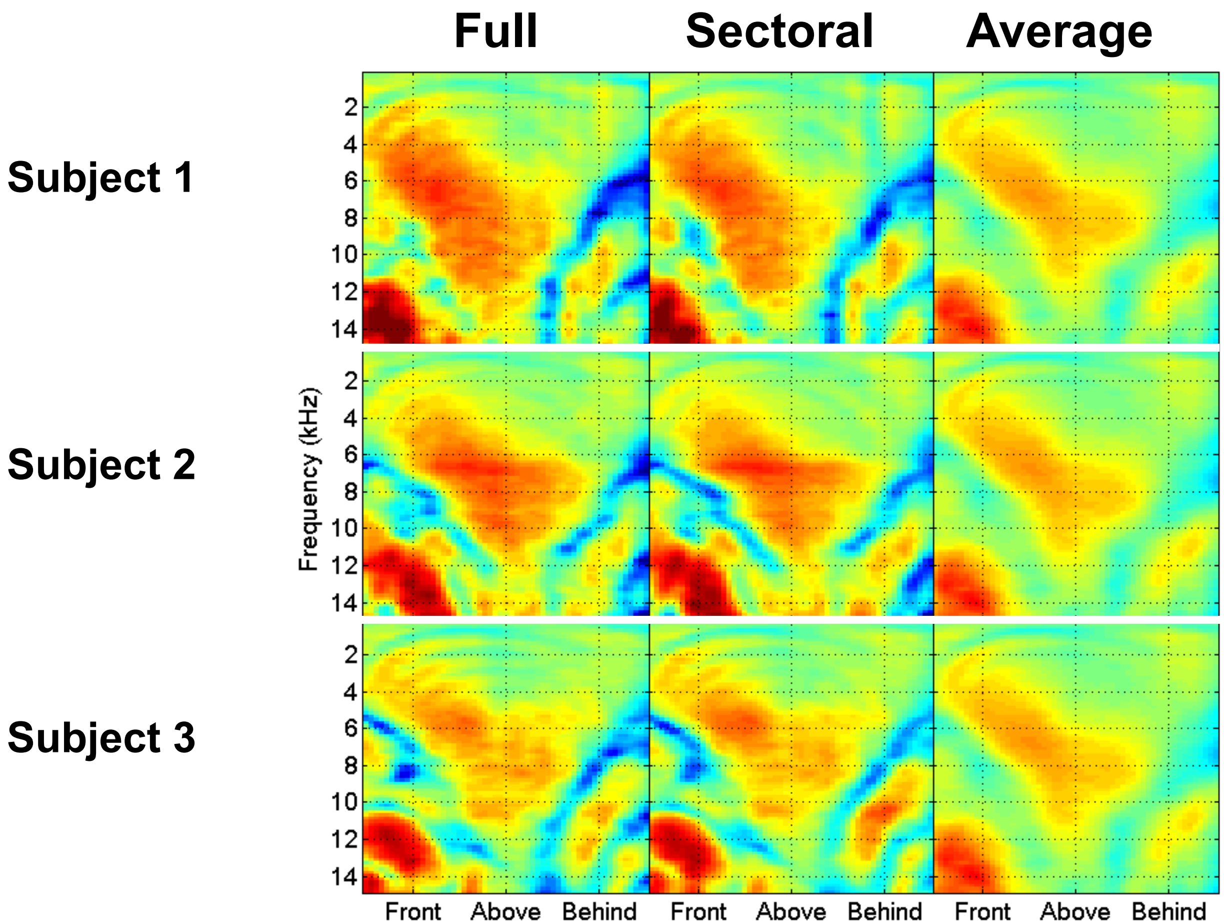
$$
\begin{aligned}
\mathbf{h} &= \mathbf{h}_{\mathrm{Sec}} + \mathbf{h}_{\mathrm{Lat}} \\
&= \mathbf{Y}_{\mathrm{Sec}}\mathbf{c}_{\mathrm{Sec}} + \mathbf{Y}_{\mathrm{Lat}}\bar{\mathbf{c}}_{\mathrm{Lat}}
\end{aligned}
$$

- $\mathbf{Y}_{\mathrm{Sec}}$: *Sectoral Basis Functions*
- $\mathbf{c}_{\mathrm{Sec}}$: *Sectoral Coefficients*
- $\mathbf{Y}_{\mathrm{Lat}}$: *Lateral Basis Functions*
- $\bar{\mathbf{c}}_{\mathrm{Lat}}$: *Average Lateral Coefficients*

Only sectoral coefficients need to be estimated. The rest can be average values.

42

Sectoral HRTF Model:



Sectoral model does capture the intraconic HRTF features

43

Estimating the Sectoral HRTF:

Estimate Sectoral HRTF with average lateral coefficients.

$$\begin{aligned}\hat{\mathbf{h}}_{\mathbf{Sec}} &\approx \mathbf{h} - \mathbf{Y}_{\mathbf{Lat}}\bar{\mathbf{c}}_{\mathbf{Lat}} \\ &\approx \mathbf{Y}_{\mathbf{Sec}}\mathbf{c}_{\mathbf{Sec}}\end{aligned}$$

Average Coefficients

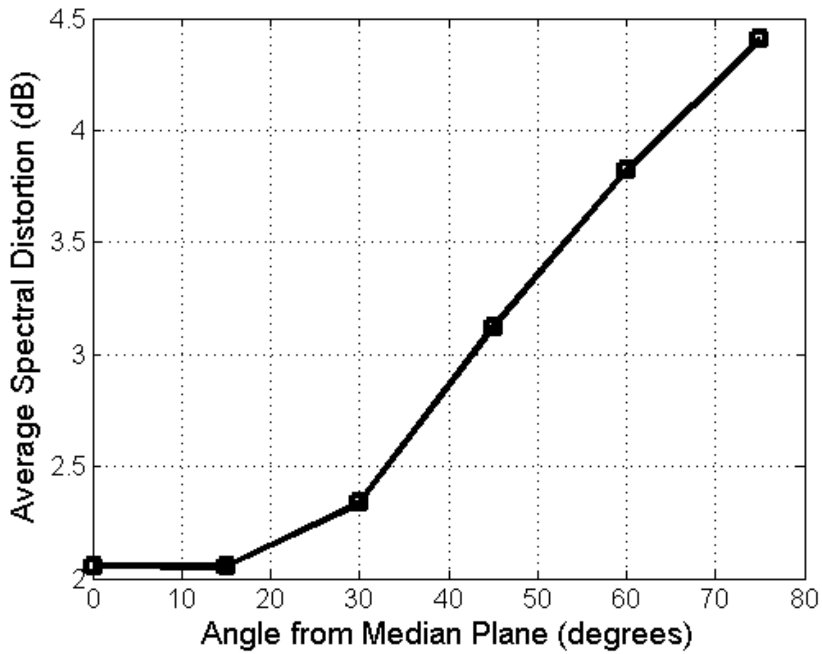
Now use Bayesian technique with Sectoral basis functions.

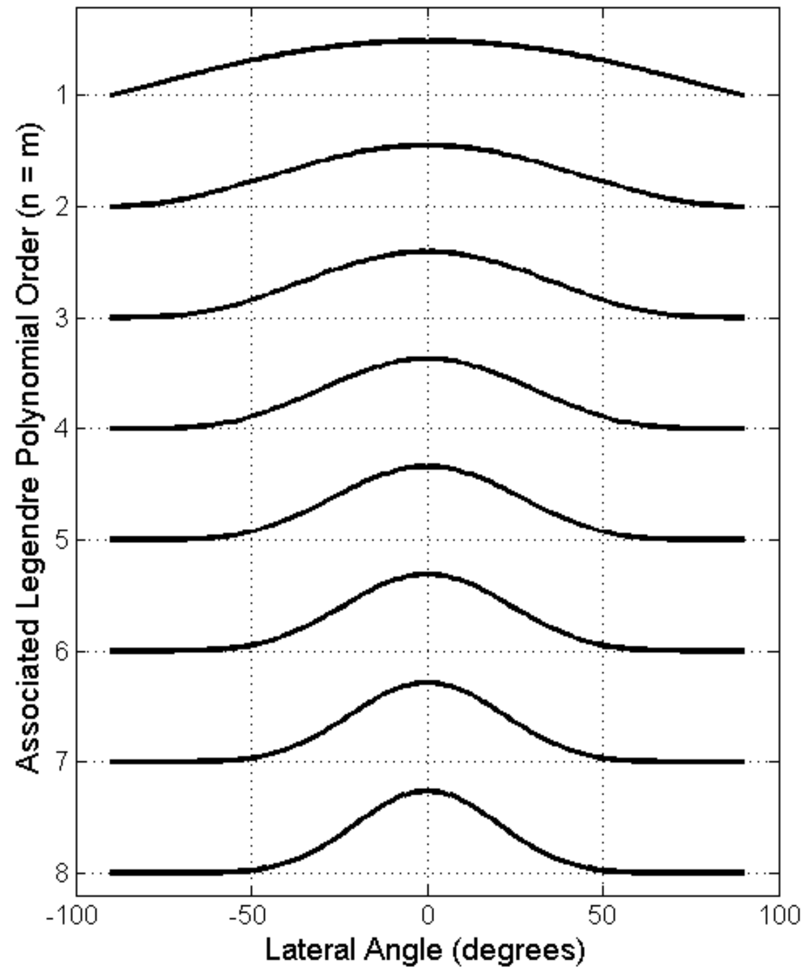
$$\begin{aligned}\hat{\mathbf{c}}_{\mathbf{Sec}} &= E[\mathbf{c}|\mathbf{h}_{\mathbf{Sec}}] \\ &= \bar{\mathbf{c}}_{\mathbf{Sec}} + \mathbf{R}_{\mathbf{Sec}}\mathbf{Y}_{\mathbf{Sec}}{}^{\mathbf{T}}(\mathbf{Y}_{\mathbf{Sec}}\mathbf{R}_{\mathbf{Sec}}\mathbf{Y}_{\mathbf{Sec}}^{\mathbf{T}} + \sigma^{\mathbf{2}}\mathbf{I})^{-\mathbf{1}}(\hat{\mathbf{h}}_{\mathbf{Sec}} - \mathbf{Y}_{\mathbf{Sec}}\bar{\mathbf{c}}_{\mathbf{Sec}})\end{aligned}$$

Estimated Sectoral HRTF

Why the median plane?

- **Bad DC estimate off midline**
- **Sectoral harmonics contain no energy off the midline at high orders**



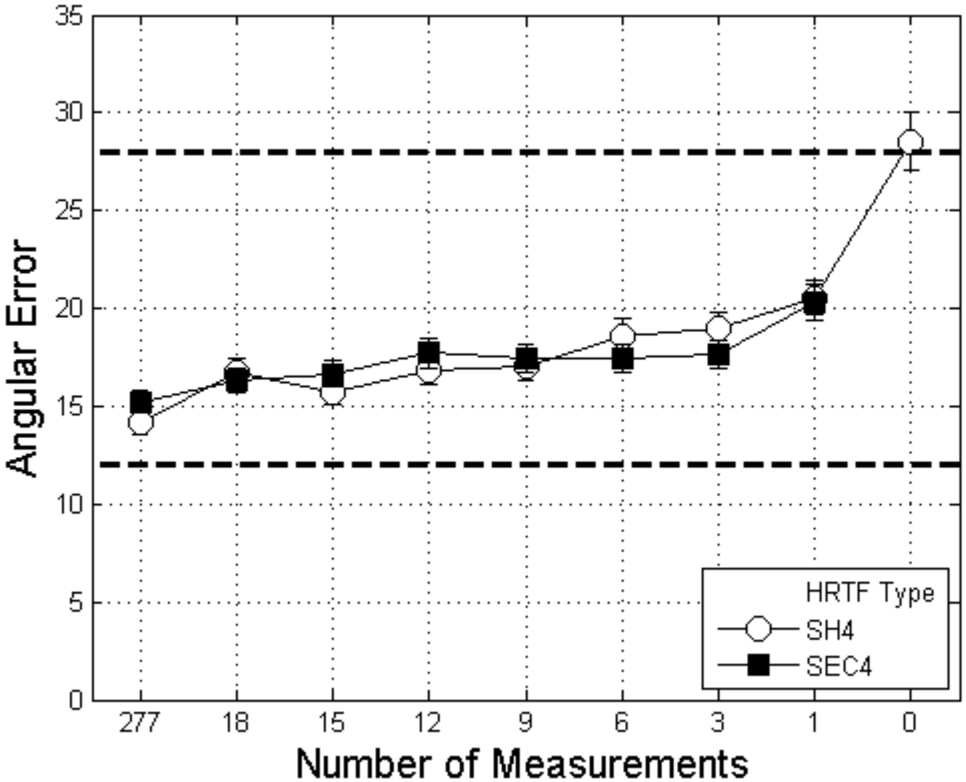


45

Perceptual Evaluation

Localization Task

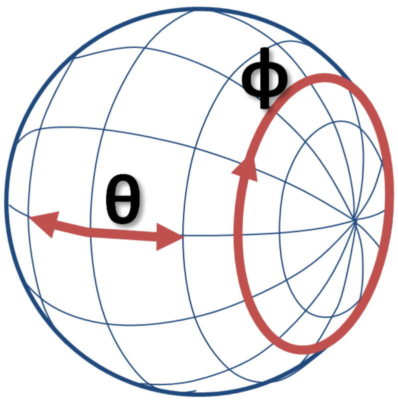
- **6 Subjects**
- **250-ms noise bursts**
- **245 locations**
- **HRTFs - Full 4th-Order (SH4) - 4th-Order Sectoral (SEC4)**

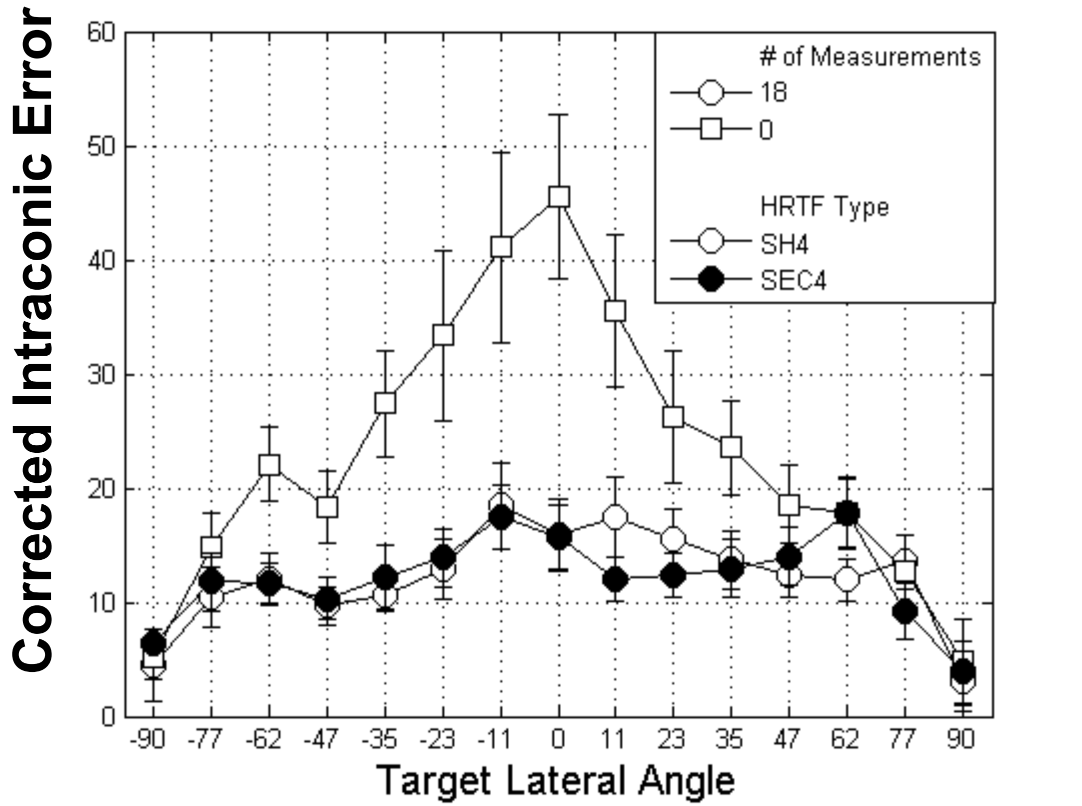


•Statistically similar performance with as few as 12 measurements

•No performance difference from Full SH model

Perceptual Evaluation





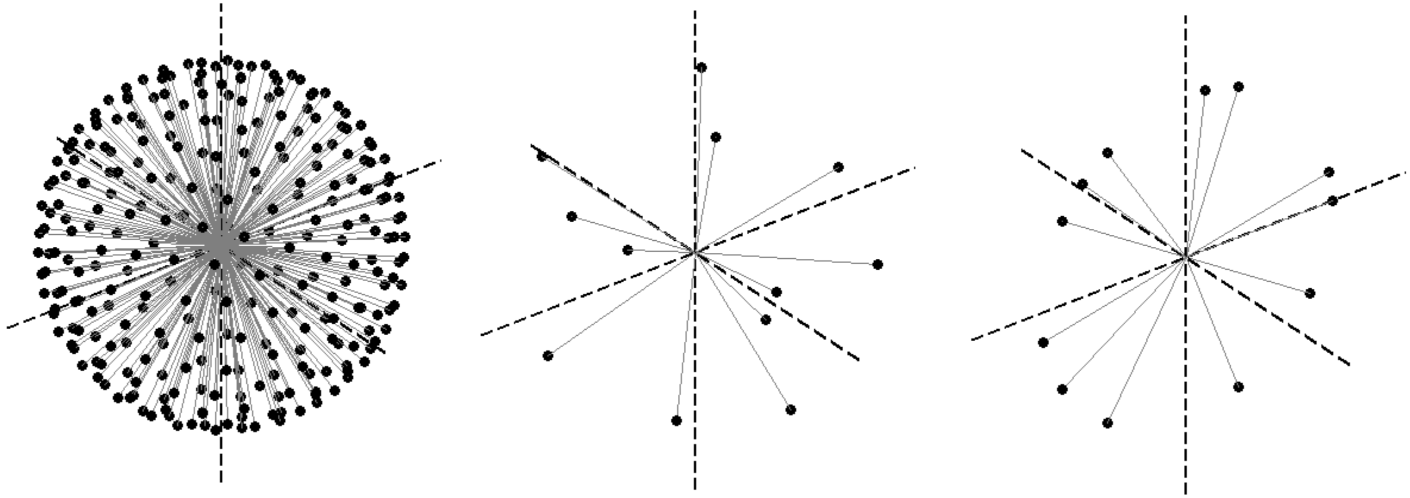
Maintains good performance off the midline

47

Recap…

- **Sectoral HRTF Model**
  - Sectoral coefficients contain large inter-subject variance
  - Only sectoral coefficients need to be individualized
  - The rest of the coefficients can be replaced with average
  - 98% fewer parameters w.r.t. baseline HRTF

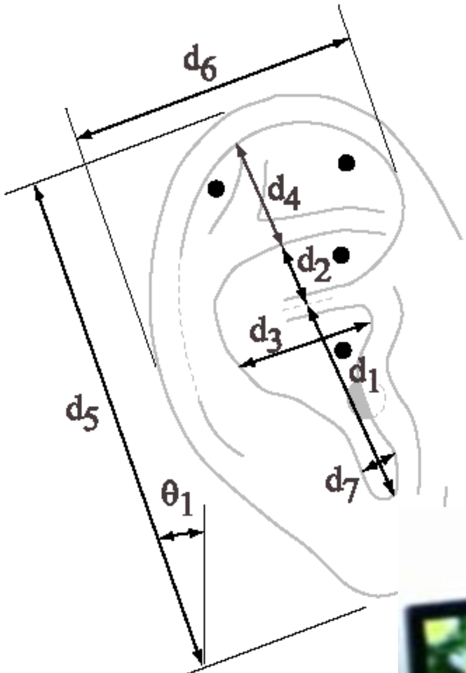
- **Median-Plane Estimation**
  - Sectoral harmonics vary mainly in intraconic dimension
  - Values can be estimated from median plane measurements

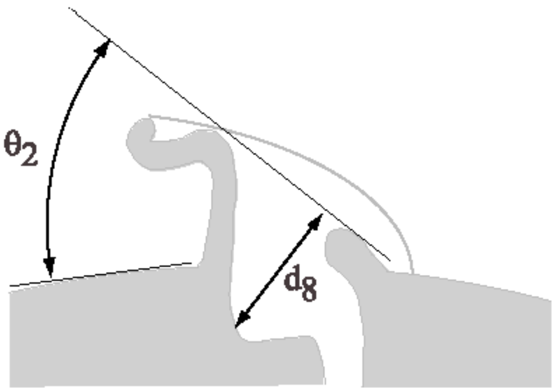


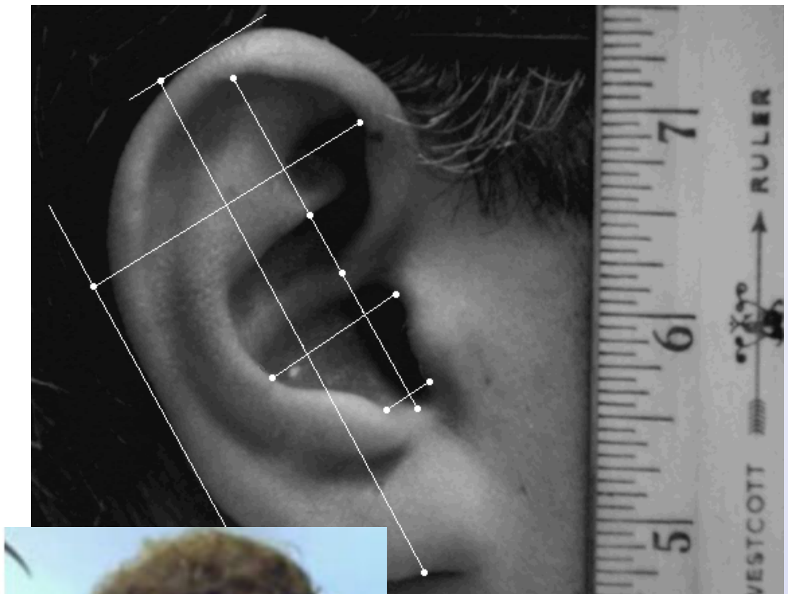
Thank You

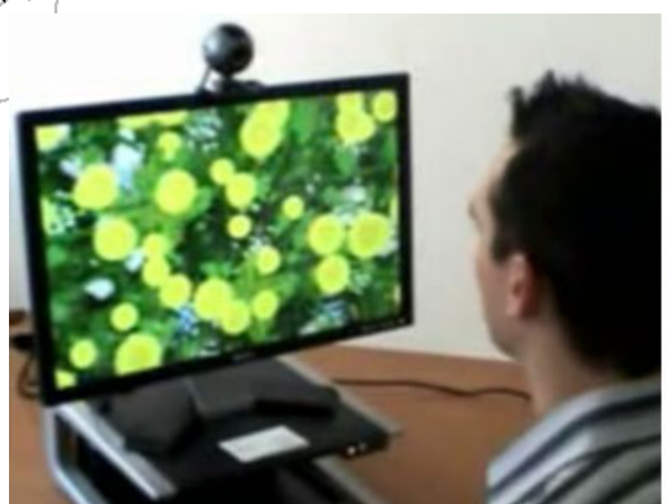
Project Ideas

Head-tracking and/or prediction of anthropometric parameters via webcam



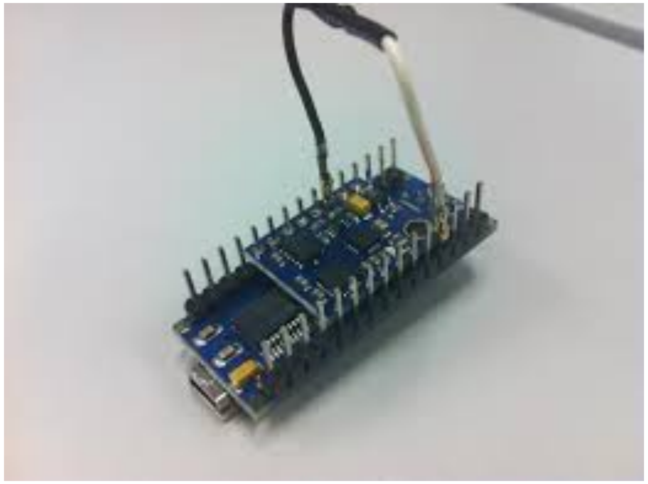






Project Ideas

HRTF measurement using a single speaker and a head tracker



Project Ideas

HRTF-based sound source localization/segregation from a binaural recording (many recordings available)