Spherical Sound Scene Analysis

Ramani Duraiswami

Perceptual Interfaces and Reality Laboratory

Institute for Advanced Computer Studies

University of Maryland, College Park

Joint work with Adam O'Donovan, Nail Gumerov, Dmitry Zotkin

Work supported by NSF, ONR/DARPA and UMIACS

http://www.umiacs.umd.edu/users/ramani ramani@umiacs.umd.edu

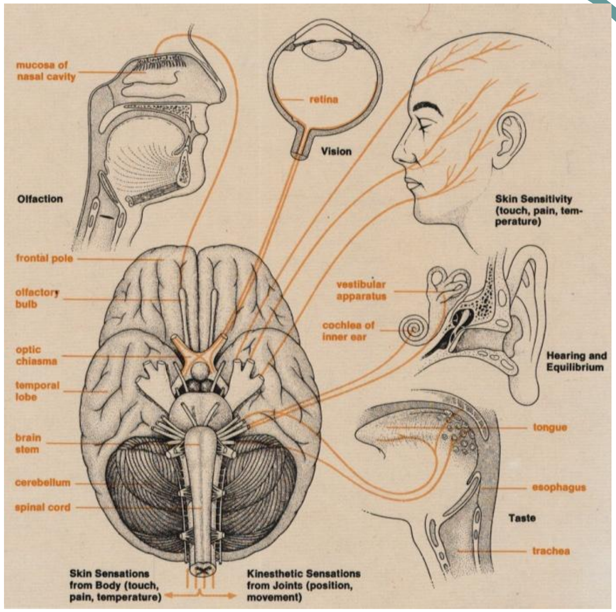
Outline

- Spatial Audio
- Modeling spatial sound
  - Spherical representations
  - Plane wave representations
- Spherical Microphone arrays
- Audio Camera
- Applications
- Recreating Auditory Reality
  - Head Related Transfer Functions
- Room and Concert Hall Acoustics
- Open Problems

Acoustical Scene Analysis

- The human perceptual system is a sophisticated sensing, measuring and computing system
- Measures audio along various dimensions useful for segregation
    - Spectral separation
    - Temporal modulations
    - Temporal separation
    - Temporal onsets/offsets

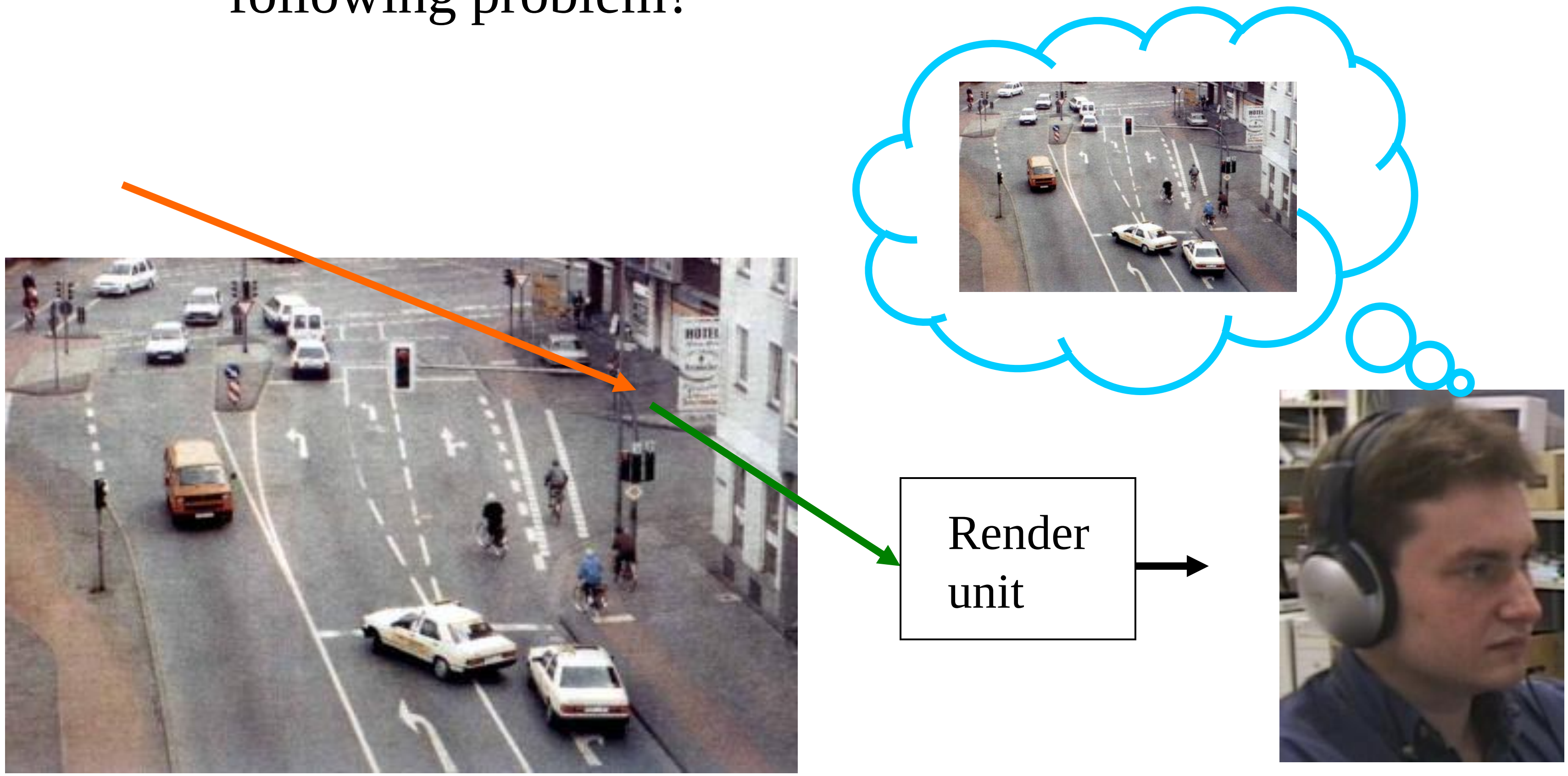
    * Spectral profile
    * Harmonicity
    * Spatial location
    * Ambience
- Designed by evolution to perform real-time measurements and take quick decisions
- Attention plays a significant role in deciding what is perceived
- Goal of today's talk --- last two items
- Create virtual reality – source at proper location



- Goal of virtual reality is to fool this system in to believing that it is perceiving an object that is not there

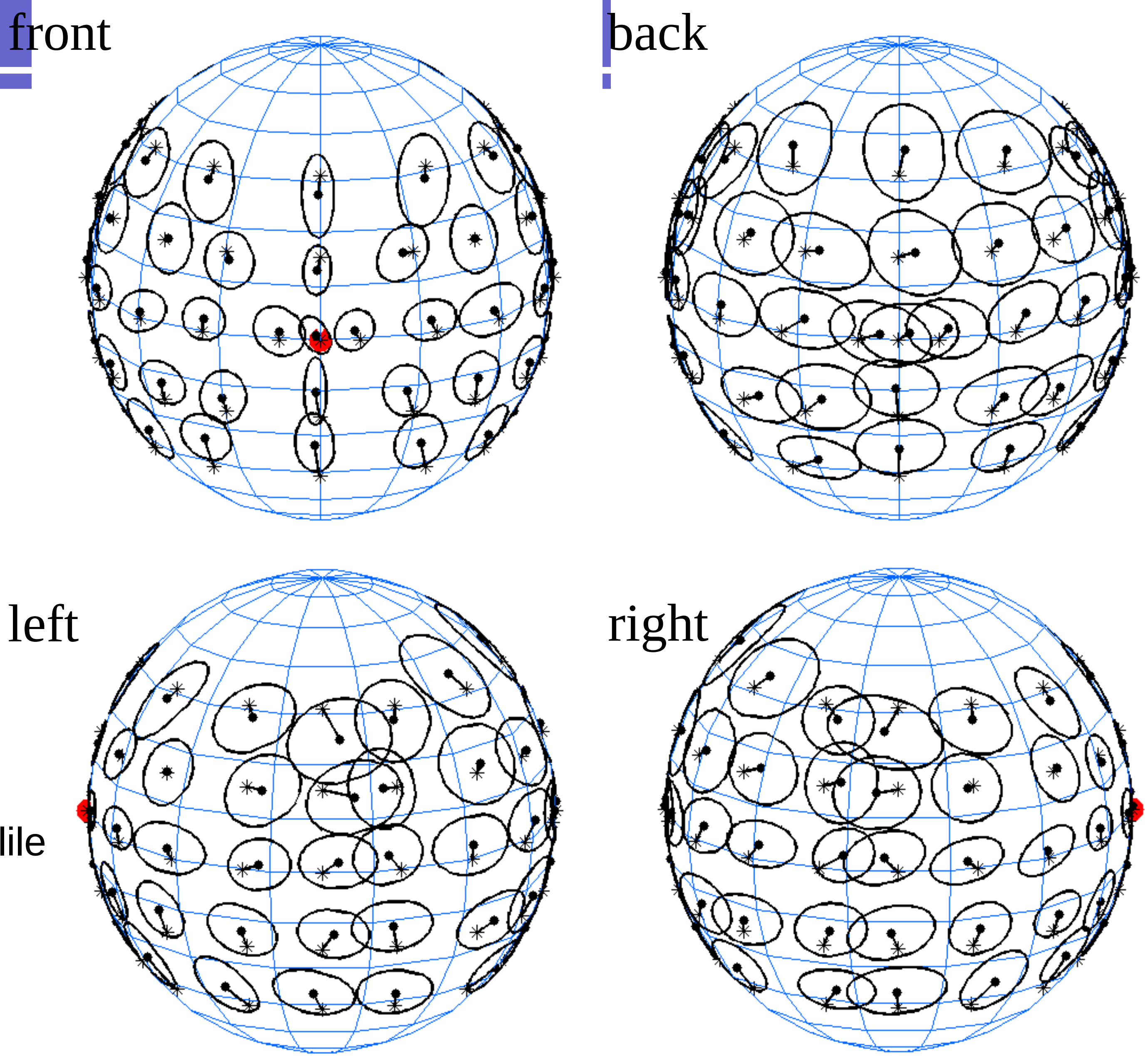
Problem we wish to solve

What theory can guarantee that we can solve the following problem?



Want to quantify error in measurement and error in reproduction using some theory. Want to do it without knowing the location of the sound sources. Allow interactivity and motion.

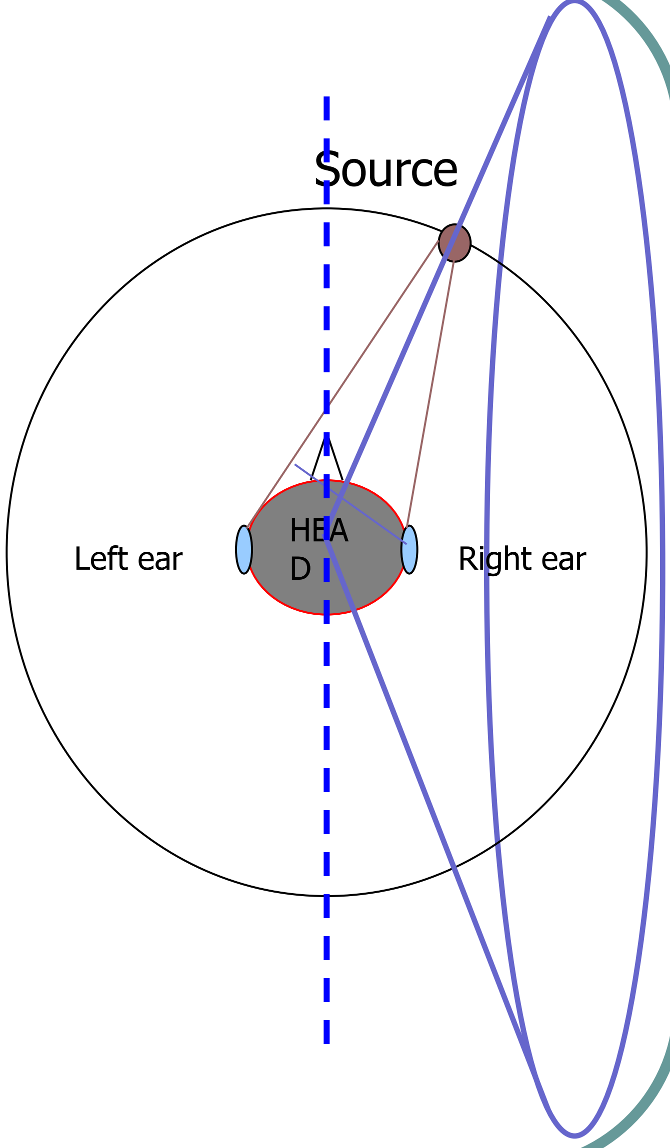
Human spatial localization ability



Best & Carlile
2003

How do we perceive sound location?

- Compare sound received at two ears
  - Interaural Level Differences (ILD)
  - Interaural Time Differences (ITD)
- Surfaces of constant Time Delay:
  $| x - x_L | - | x - x_R | = c\,\delta t$
  - hyperboloids of revolution
  - Delays same for points on cone-of-confusion
- Other mechanisms necessary to explain
  - Scattering of sound
    - Off our bodies
    - Off the environment
  - Purposive Motion

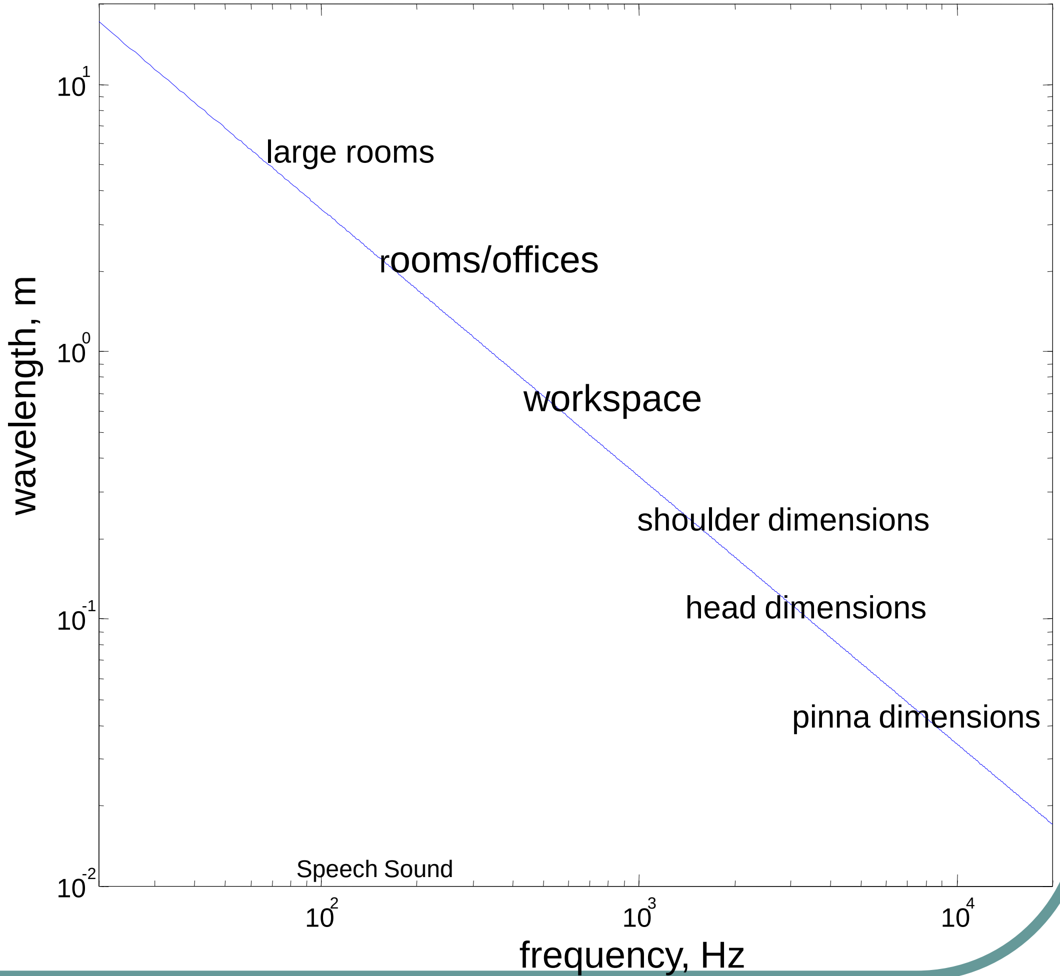


Audible Sound Scattering

- Sound wavelengths comparable to human dimensions and dimensions of spaces we live in.
- $f\lambda = c$
- When $\lambda >> a$
  wave is unaffected by object
  $\lambda \sim a$
  behavior of scattered wave is complex and diffraction effects are important.
  $\lambda << a$
  wave behaves like a ray

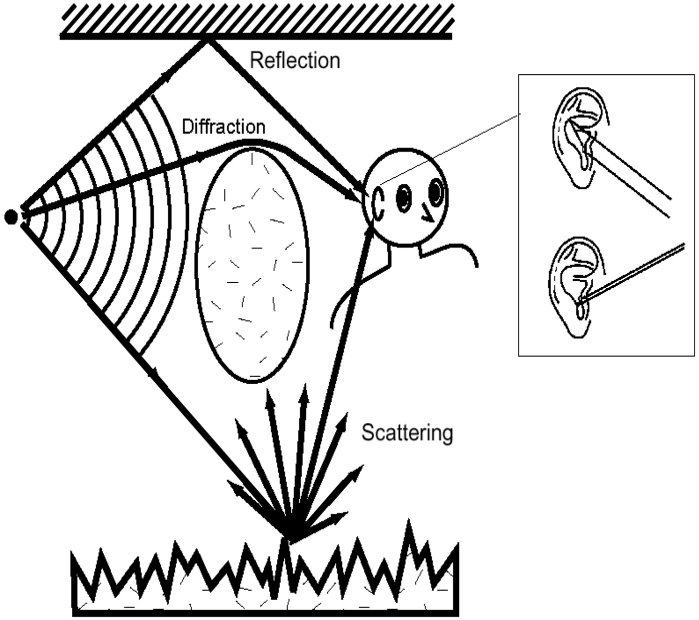
wavelengths are comparable to our rooms, bodies, and features

Not an accident but evolutionary selection!



distance cues

- Level variation - inverse square: -6dB per doubling of distance
- High frequency absorbance >4 kHz: - 1.6dB per doubling of distance
- Direct to reverberant E ratio: Direct E dependent on distance
- Near field binaural (1° ILD) variations with distance



Guiding principles

- **Axiom: To create the virtual scene, it is sufficient to recreate sound pressure levels at the eardrum**
  - Or a sufficiently fine approximation to it …
- Obtain sound field accurately
- Modify them using system dependent responses
- Linear systems can be characterized by impulse response (IR)
  - Knowing IR, can compute response to general source by convolution
- Response to impulsive source at a particular location
  - Scattering off person by Head Related Impulse Response (HRIR)
  - Room scattering by Room Impulse Response (RIR)
- Response differs according to source and receiver locations
  - Thus encodes source location
- HRTF and RTF are Fourier transforms of the Impulse response
  - Convolution is cheaper in the Fourier domain (becomes a multiplication)

Creating Auditory Reality

- Capture the Sound Source
- Rerender it by reintroducing cues that exist in the real world
- Scattering of sound off the human
  - Head Related Transfer Functions
- Scattering off the Environment
  - Room Models
- Head motion
  - Head/Body Tracking

Capturing sound: Mathematical formulation

- Analysis via wave-equation
  - Or its Fourier transform
  - (Human auditory system performs its own version of Fourier transform)
- Spherical coordinate system
  - Our head is relatively spherical
  - Our ability to characterize sources (linguistically and phenome-nologically) is direction based
  - Implies use of a spherical analysis

- Wave equation

$$\frac{1}{c^2}\frac{\partial^2 p'(\mathbf{r},t)}{\partial t^2} = \nabla^2 p'(\mathbf{r},t)\,,$$

- Subject to initial and boundary conditions
- Take Fourier Transform

$$\psi(x,y,z,w) = \int_{-\infty}^{\infty} p'(x,y,z,t) e^{-i\omega t} dt$$

- Helmholtz equation

$$\nabla^2\psi(\mathbf{r}) + k^2\psi(\mathbf{r}) = 0, \quad k = \frac{\omega}{c},$$

- Boundary value problem per frequency

Representation via spherical wavefunctions

- sound at a point
  - Satisfies the wave equation
  - Fourier transform satisfies Helmholtz equation
  - So we can represent the sound at a point in terms of the local point-eigenfunctions of the Helmholtz equation
  - Expand solutions in series, but truncate at *p* terms causing an error ε_p
  - Error depends on frequency
    - For a given sound of wavenumber *k* this gives us minimum order for sensible representation

$$\psi_{in}(k;\mathbf{r}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_n^m R_n^m(k;\mathbf{r}),$$

$$R_n^m(k;\mathbf{r}) = j_n(kr)Y_n^m(\theta,\varphi),$$

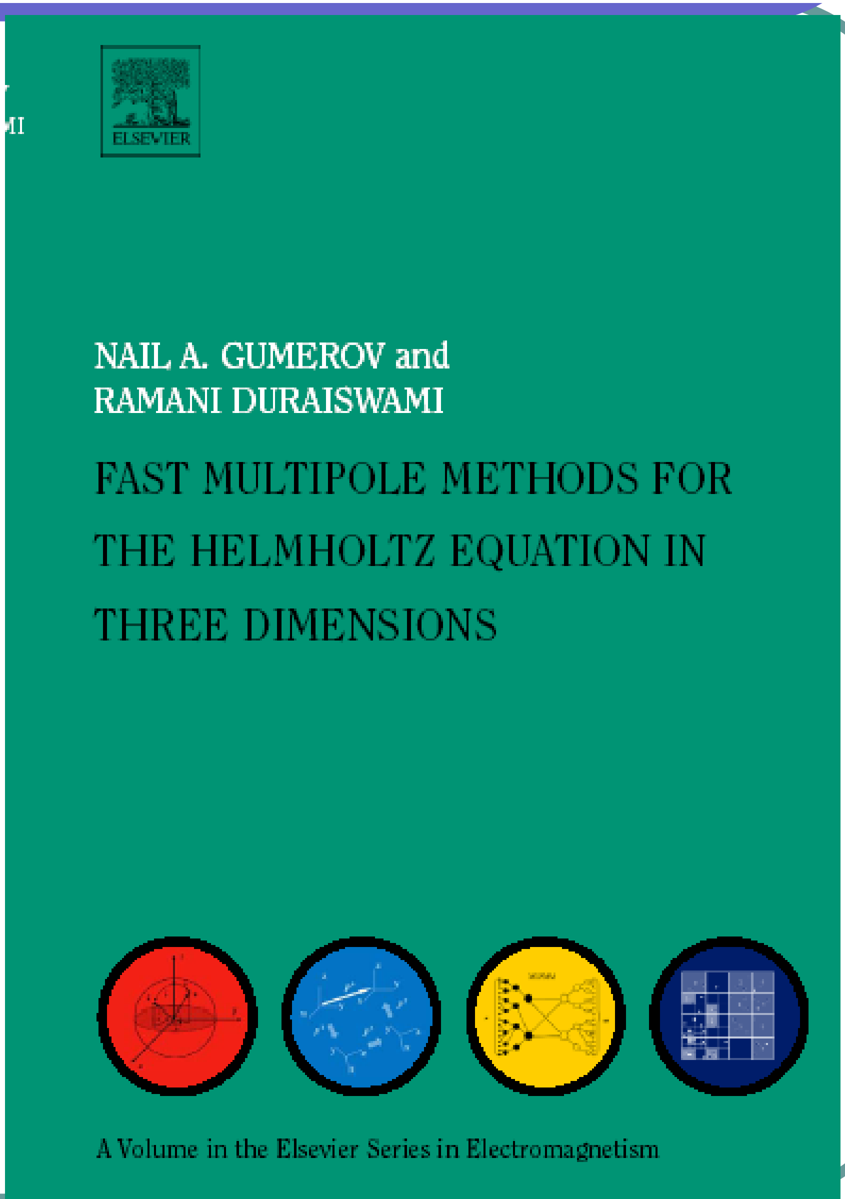
$$|\epsilon_p(\mathbf{s},\mathbf{r})| \lesssim \exp\left\{-\frac{1}{3}\left[2\frac{p-kR}{(kR)^{1/3}}\right]^{3/2}\right\} = \delta_p, \quad kR \gg 1.$$

Can also write this for radiating functions

$$S_n^m(k;\mathbf{r}) = h_n(kr)Y_n^m(\theta,\varphi)$$

Shameless plug

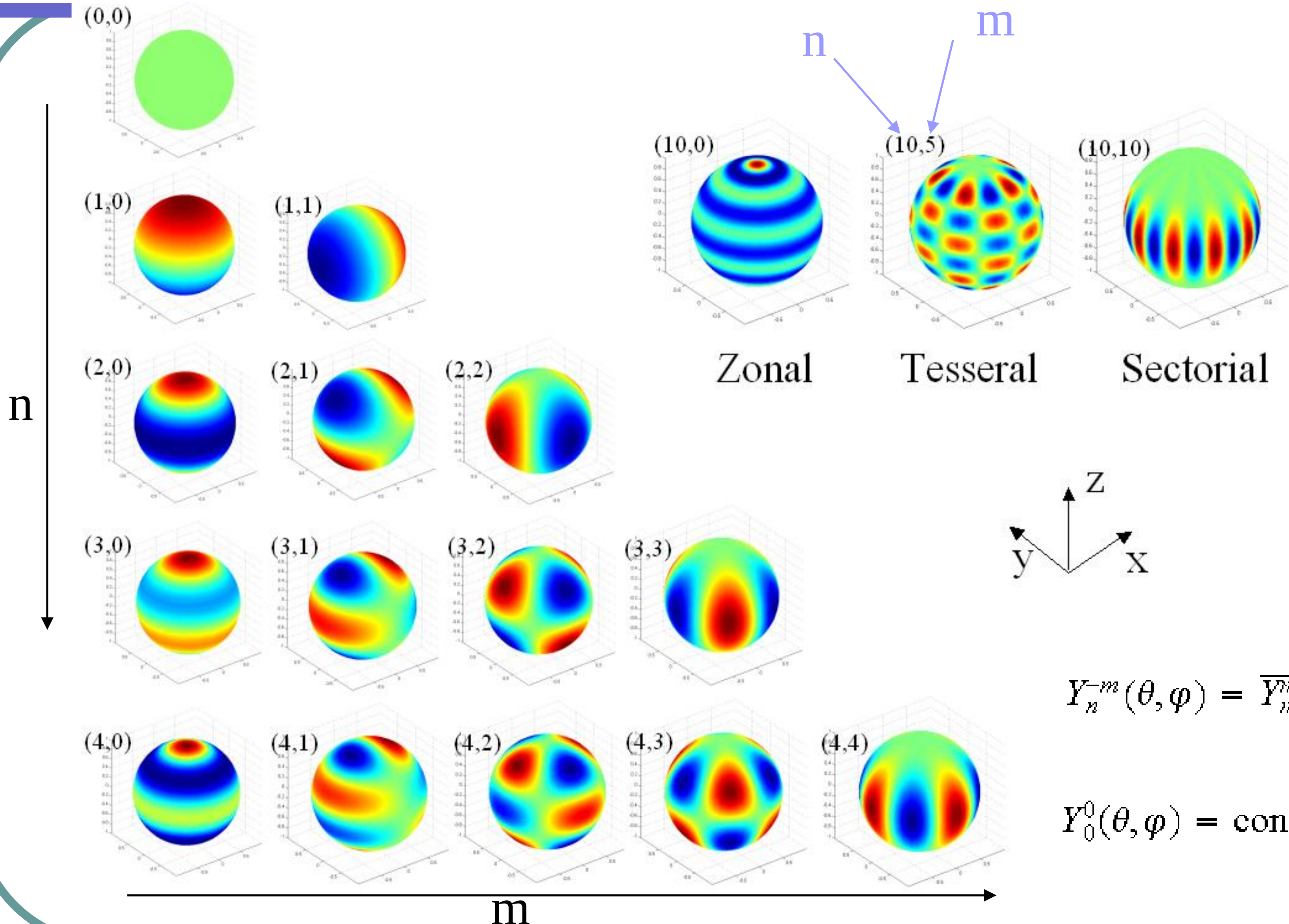
- Analysis of solutions of the Helmholtz equation in our book
  - Elsevier, 2005
- What do these basis functions look like?



Spherical Harmonics

$$Y_n^m(\theta,\varphi) = (-1)^m \sqrt{\frac{2n+1}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}}\, P_n^{|m|}(\cos\theta) e^{im\varphi},$$

$$n = 0, 1, 2, \ldots; \qquad m = -n, \ldots, n.$$

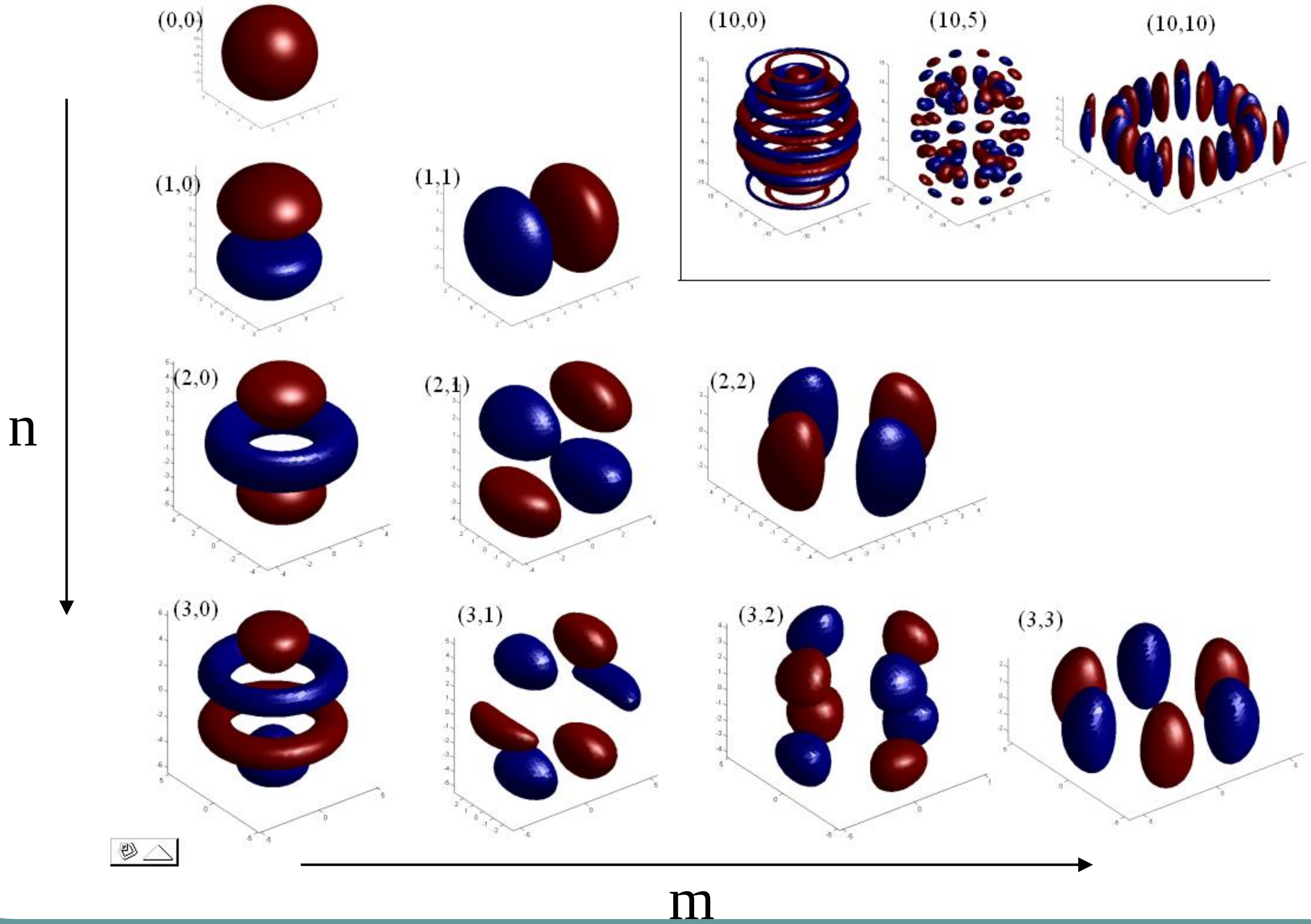


$$Y_n^{-m}(\theta,\varphi) = \overline{Y_n^m(\theta,\varphi)}.$$

$$Y_0^0(\theta,\varphi) = \text{const} = \sqrt{\frac{1}{4\pi}}.$$

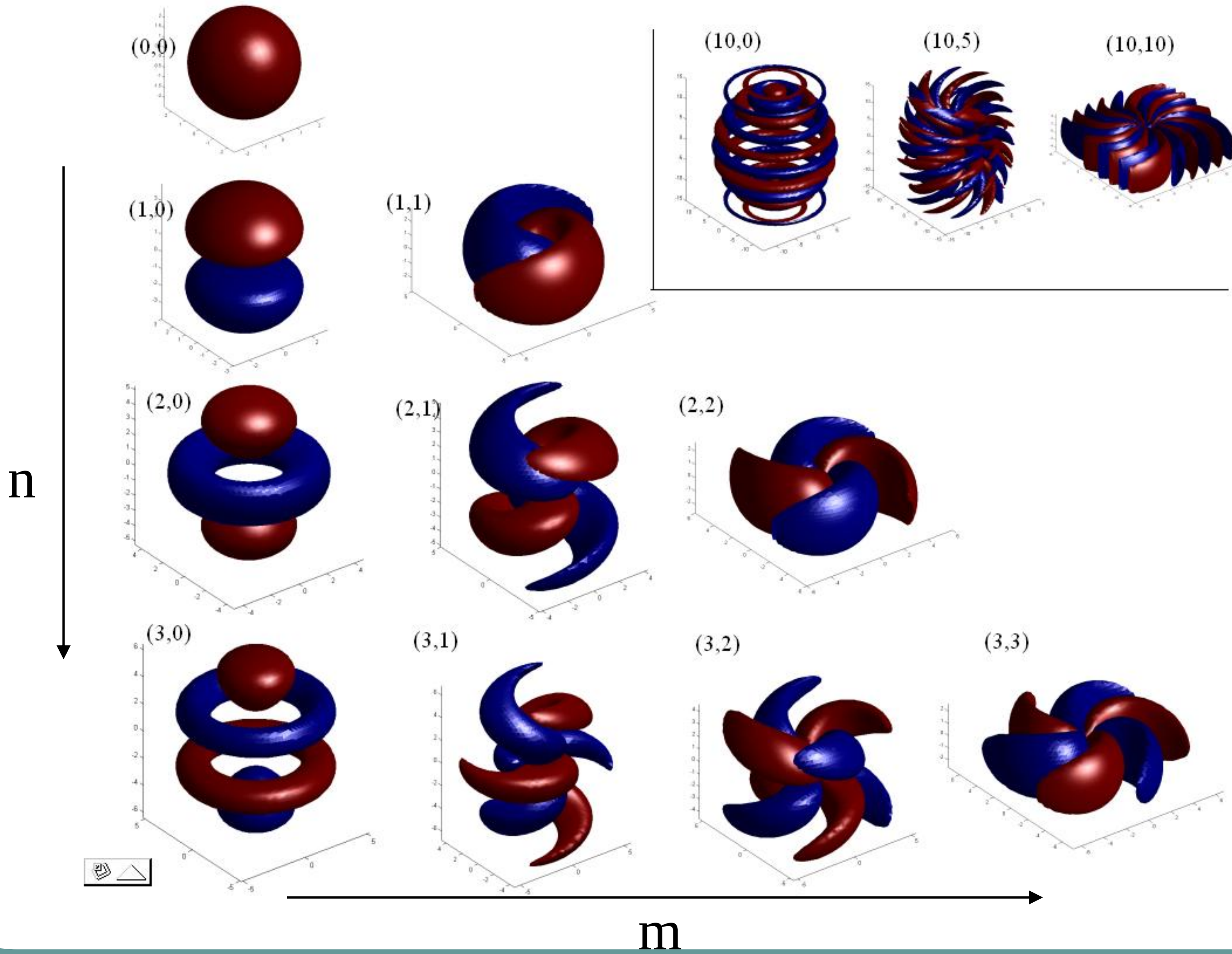
Isosurfaces For Regular Basis Functions

$$\mathrm{Re}\{R_n^m(\mathbf{r})\} = const$$



Isosurfaces For Singular Basis Functions

$$\mathrm{Re}\{S_n^m(\mathbf{r})\} = const$$



Real sound fields are quite different

- Created by relatively compact sources
- Sources are at a distance to the receiver
- Receiver is also relatively compact
- Source (of any order) far away appears as a plane-wave
- Plane-waves can also be used to form a basis!

Yet another representation (Plane Waves)

- any soundfield in regular region can be expressed as an integral form of plane waves.
  - Integral over a unit sphere at the point
  - Decomposes any sound field in to a set of planewaves of various strengths

$$\psi_{in}(\mathbf{r}) = \frac{1}{4\pi}\int_{S_u} e^{ik\mathbf{s}\cdot\mathbf{r}}\mu_{in}(\mathbf{s})\,dS(\mathbf{s}),$$

Plane waves

Coeffs

  - Can be connected to other representation by expanding plane-wave in terms of spherical functions

$$e^{ik\mathbf{s}\cdot\mathbf{r}} = 4\pi\sum_{n=0}^{\infty}\sum_{m=-n}^{n} i^n Y_n^{-m}(\mathbf{s})\,R_n^m(\mathbf{r}), \quad R_n^m(\mathbf{r}) = \frac{i^{-n}}{4\pi}\int_{S_u} e^{ik\mathbf{s}\cdot\mathbf{r}} Y_n^m(\mathbf{s})dS(\mathbf{s}),$$

  - In practice these integrals are evaluated via quadrature

$$\int_{S_u} F(\mathbf{s})\,dS = \sum_{j=0}^{L_Q-1} F(\mathbf{s}_j)\,w_j, \; F(\mathbf{s}) = \sum_{n=0}^{p-1}\sum_{m=-n}^{n} C_n^m Y_n^m(\mathbf{s}),$$

  - Approximation error in this case is related to error in the quadrature
  - Quadrature error formula relates L_Q to p

Sensor to capture sound in these representations

- Need some sensor to get the coefficients
  - Spherical microphone array
  - Similar to ambisonics: however the expansion depends on the frequency, and we know the error bound
  - if we want the sound to be valid in a domain the size of the head we can evaluate the needed order for a given error

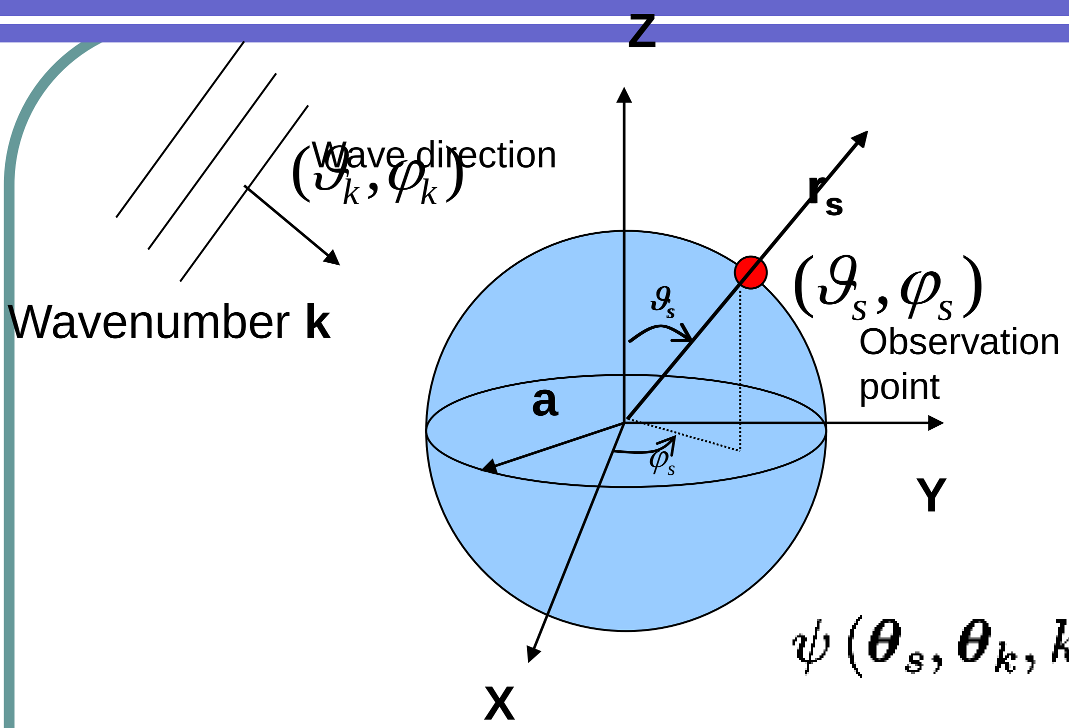
$$p \approx kR + A(kR)^{1/3}$$

  - To capture sound to order *p* we need a certain microphone design

Spherical Arrays

Plane wave

$$(\nabla^2 + k^2)\psi(\mathbf{r}) = 0$$



wave scattering from a rigid (sound hard) surface

find solution to Helmholtz equation which satisfies:

- the rigid surface, $\partial \psi / \partial n = 0$

radiation condition on ψ_{scat}.

`

$$\psi(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka)$$

$$= \left[\psi_{in}(\mathbf{r}_s, \mathbf{k}) + \psi_{scat}(\mathbf{r}_s, \mathbf{k})\right]\Big|_{r_s=a}$$

$$= 4\pi \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\boldsymbol{\theta}_k) Y_n^{m*}(\boldsymbol{\theta}_s),$$

$$b_n(ka) = j_n(ka) - \frac{j_n'(ka)}{h_n'(ka)} h_n(ka),$$

Meyer & Elko, 2002

Meyer and Elko's observation

Let the weight at each point be:

$$W_{n'}^{m'}(\boldsymbol{\theta}_s, ka) = \frac{Y_{n'}^{m'}(\boldsymbol{\theta}_s)}{4\pi i^{n'} b_{n'}(ka)}$$

Using orthonormality of spherical harmonics:

$$\int_{\Omega_s} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) d\Omega_s = \delta_{nn'}\delta_{mm'},$$

The output of the beamformer is:

The directional response of the plane wave

$$\int_{\Omega_s} \psi\left(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka\right) W_{n'}^{m'}(\boldsymbol{\theta}_s, ka) d\Omega_s = Y_{n'}^{m'}(\boldsymbol{\theta}_k).$$

Recall spherical harmonics are a basis for directions at a point

Combine spherical harmonics to get arbitrary beams

For example, the ideal beampattern looking at $\boldsymbol{\theta}_0$

$$\delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$$

can be expanded into:

$$2\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta})$$

In practice, with discrete spatial sampling, this is a finite number *N*.

So the weight for each microphone at $\boldsymbol{\theta}_s$ is:

$$w(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) = \sum_{n=0}^{\infty} \frac{1}{2i^n b_n(ka)} \sum_{m=-n}^{n} Y_n^{m*}(\boldsymbol{\theta}_0) Y_n^m(\boldsymbol{\theta}_s).$$

Then, the spatial response for the plane wave from $\boldsymbol{\theta}_k$ is:

$$\int_{\Omega_s} \psi\left(\boldsymbol{\theta}_s, \boldsymbol{\theta}_k, ka\right) w(\boldsymbol{\theta}_0, \boldsymbol{\theta}_s, ka) d\Omega_s = \delta(\boldsymbol{\theta}_k - \boldsymbol{\theta}_0)$$

Quadrature and Spherical Beamforming

- A quadrature formula provides layout and weights to obtain the integral.
- In practical spherical beamformer with finite number of microphones, this is a quadrature problem w.r.t. orthonormalities of spherical harmonics.

$$\int_{\Omega_s} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) d\Omega_s = \delta_{nn'}\delta_{mm'},$$

quadrature

$$\frac{4\pi}{S}\sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s) Y_{n'}^{m'}(\boldsymbol{\theta}_s) C_{n'}^{m'}(\boldsymbol{\theta}_s) = \delta_{nn'}\delta_{mm'},$$

Quadrature weights

$$(n = 0, \ldots, N_{eff}; m = -n, \ldots, n;$$

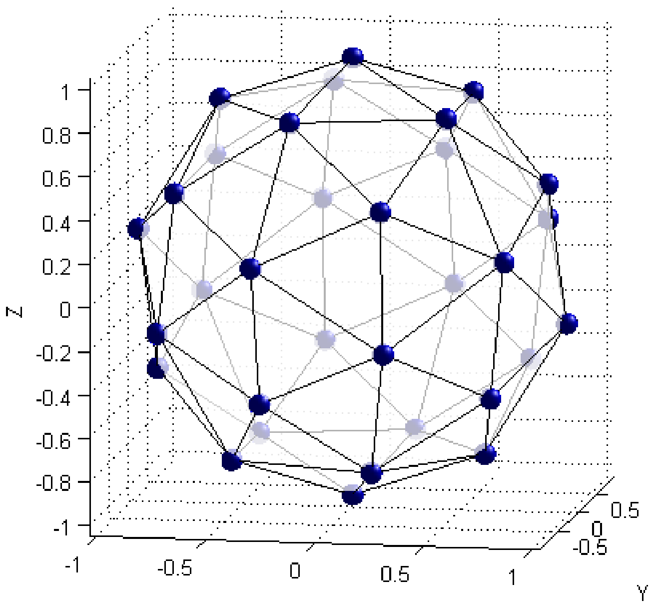
$$n' = 0, \ldots, N; m' = -n', \ldots, n'),$$

Bandlimit

Meyer and Elko: Uniform Layout Quadrature

- truncated icosahedron to layout 32 microphones.
- Unfortunately, It can be proven that only five regular polyhedrons exist: cube, dodecahedron, icosahedron, octahedron, and tetrahedron [Steinhaus99]
- Layouts are fixed and unavailable for arbitrary number of nodes.

The 32 nodes from face centers of a truncated icosahedron



Quadrature is the key

- Quadrature formula provides microphone locations on the sphere and weights for these

$$\frac{4\pi}{S}\sum_{s=1}^{S} Y_n^{m*}(\boldsymbol{\theta}_s)Y_{n'}^{m'}(\boldsymbol{\theta}_s) = \delta_{nn'}\delta_{mm'}$$

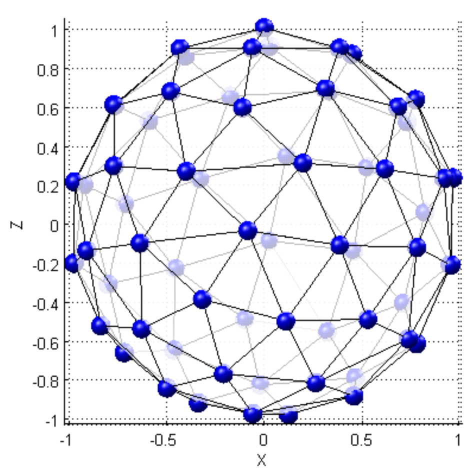
The number of microphones

The microphone angular positions

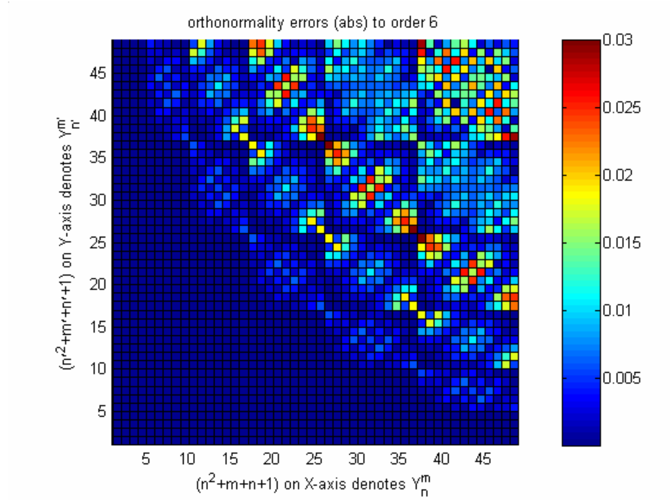
- Any formula of order *p* over the sphere should have more than $S = (p+1)^2$ nodes [Hardin&Sloane96, Taylor95].
- For bandwidth *p*, to achieve the exact quadrature using equiangular layout, we need $4(p+1)^2$ nodes [Healy96].
- For a Gaussian layout, we need $S = 2(p+1)^2$ [Rafaely05].
- Spherical t-design: use special layout for equal quadrature weights [Hardin&Sloane96]
  - used by Meyer & Elko, 2002

Microphone arrays via robust Fliege quadrature

- We use the Fliege nodes and an optimization based approach to obtain a robust set of quadrature points and weights, (Li & D, 2005)
- Idea: repel electrons on a surface of a sphere to find uniform sampling
- Sample sound field at these points
- Can use this idea to build "approximate" quadrature formulas which sample sound field much better -→
- Practically p^2 nodes give $O(p)$ analysis
- Shown to also degrade gracefully with frequency (Zotkin et al., 2010)



(a)



(b)

Capturing the sound field via spherical arrays

- From the recorded sound we can deduce the coefficients of the incident soundfield ψ_{in} (in the absence of the array)
- In Zhiyun Li's thesis (2005) and several papers the theory of spherical arrays was extended to
  - Allow arbitrary placement of microphones on sphere surface
  - Achieve highest order possible for a given number of microphones by developing robust quadrature over the sphere
  - Develop weights that are robust to noise, placement errors of microphones, and to individual microphone failure
  - Performing beamforming with them
  - Building and testing of spherical and hemispherical arrays
  - Developed devices work according to the theory!

Expressions for incoming plane-wave strength

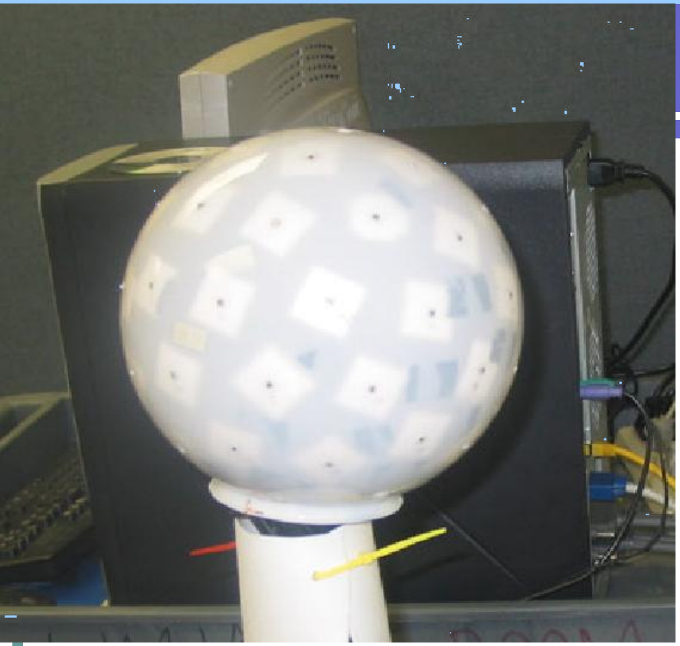
- solve for plane wave coefficients from particular directions $\mathbf{s}_l$ given measurements at microphones at locations $\mathbf{s}_j$
  - So this allows us to decompose any sound field in terms of a set of truncated plane waves

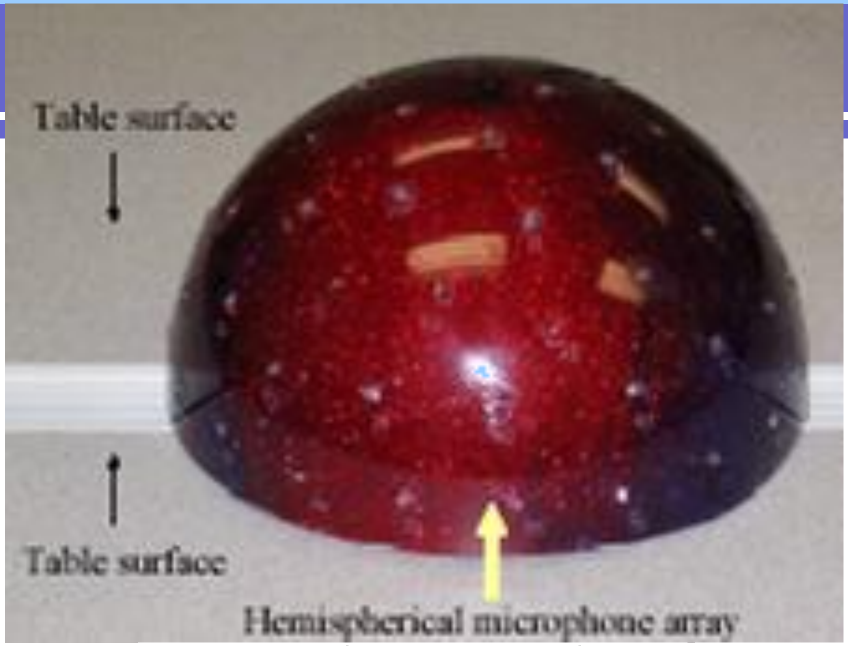
$$\mu_{in}\left(\mathbf{s}_l\right) = \sum_{j=0}^{L_M-1} w_j M\left(\mathbf{s}_l;\mathbf{s}_j\right)\psi_S\left(\mathbf{s}_j\right),$$

$$M\left(\mathbf{s}_l;\mathbf{s}_j\right) = -\frac{i\left(ka\right)^2}{4\pi}\sum_{n=0}^{p-1}\left(2n+1\right)i^{-n}B_n(ka)P_n\left(\mathbf{s}_l\cdot\mathbf{s}_j\right)$$

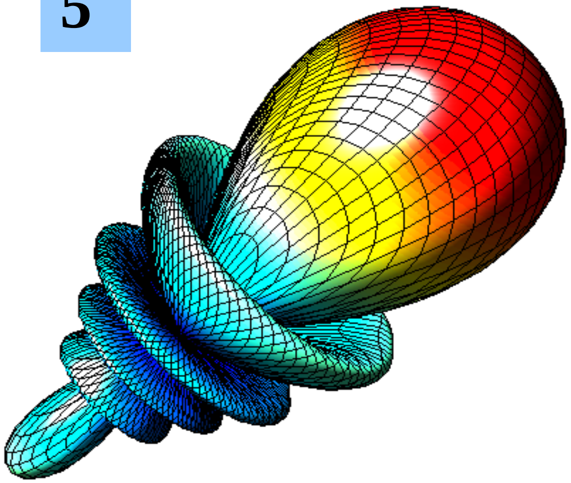
$$B_n(ka) = h_n'\left(ka\right) + \left(\sigma/k\right)h_n\left(ka\right),$$

Our Spherical Arrays: Experimental Results

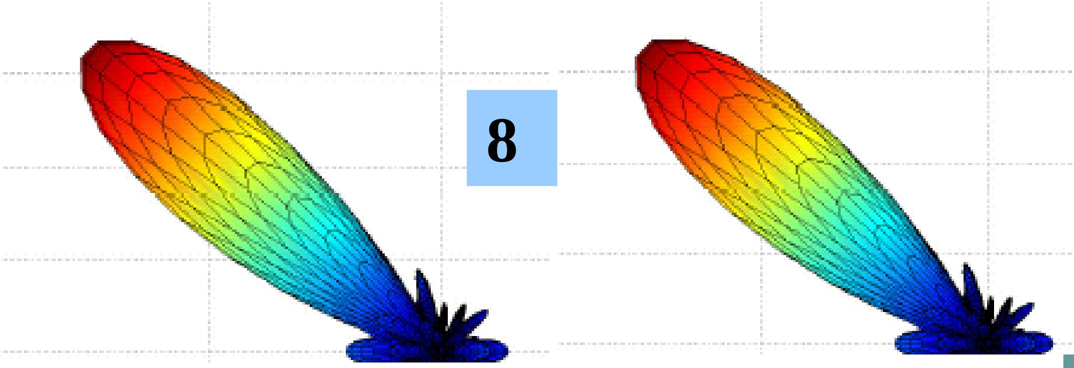




5



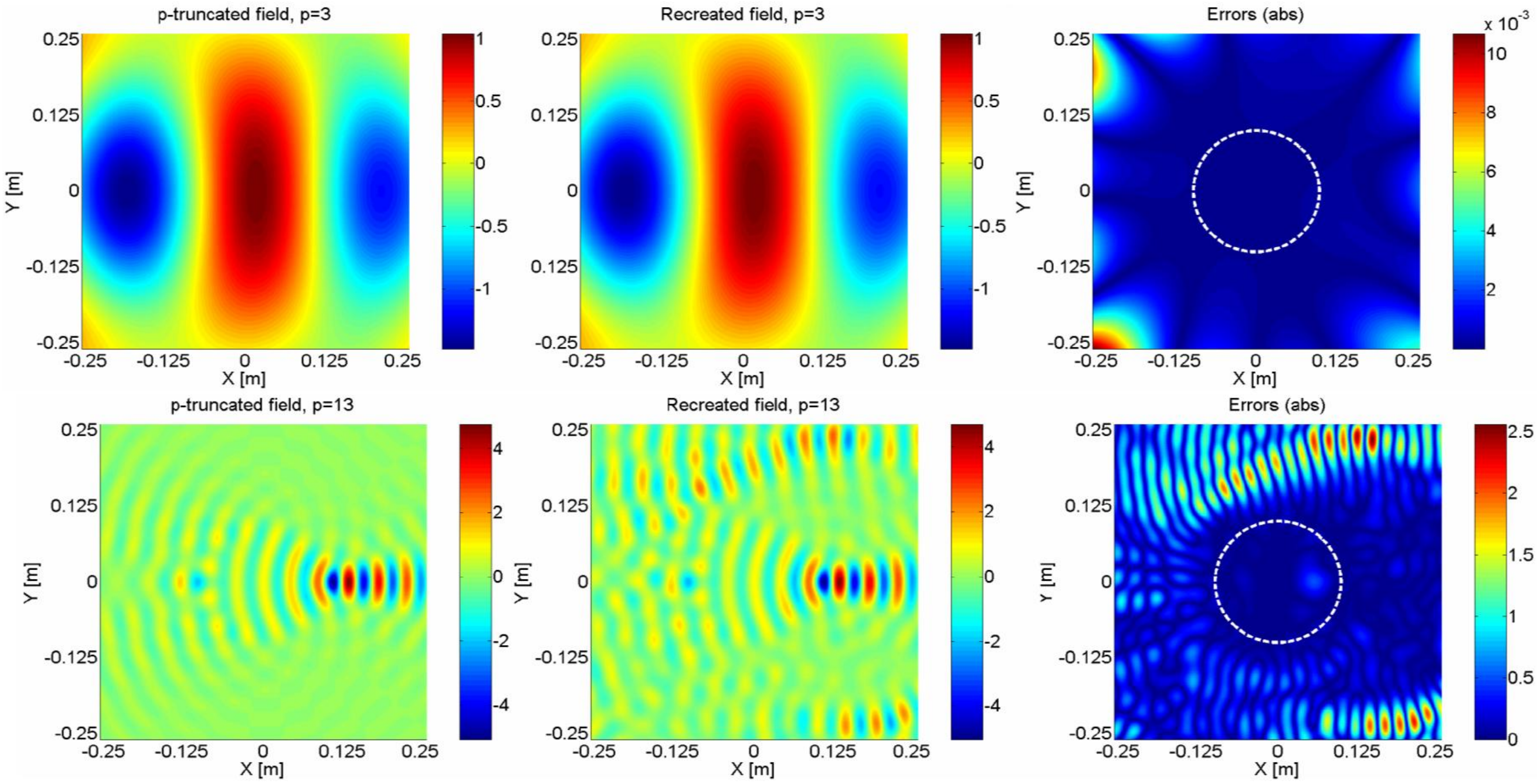
8



Can synthesize high order digital beams that can pick sounds from arbitrary directions!

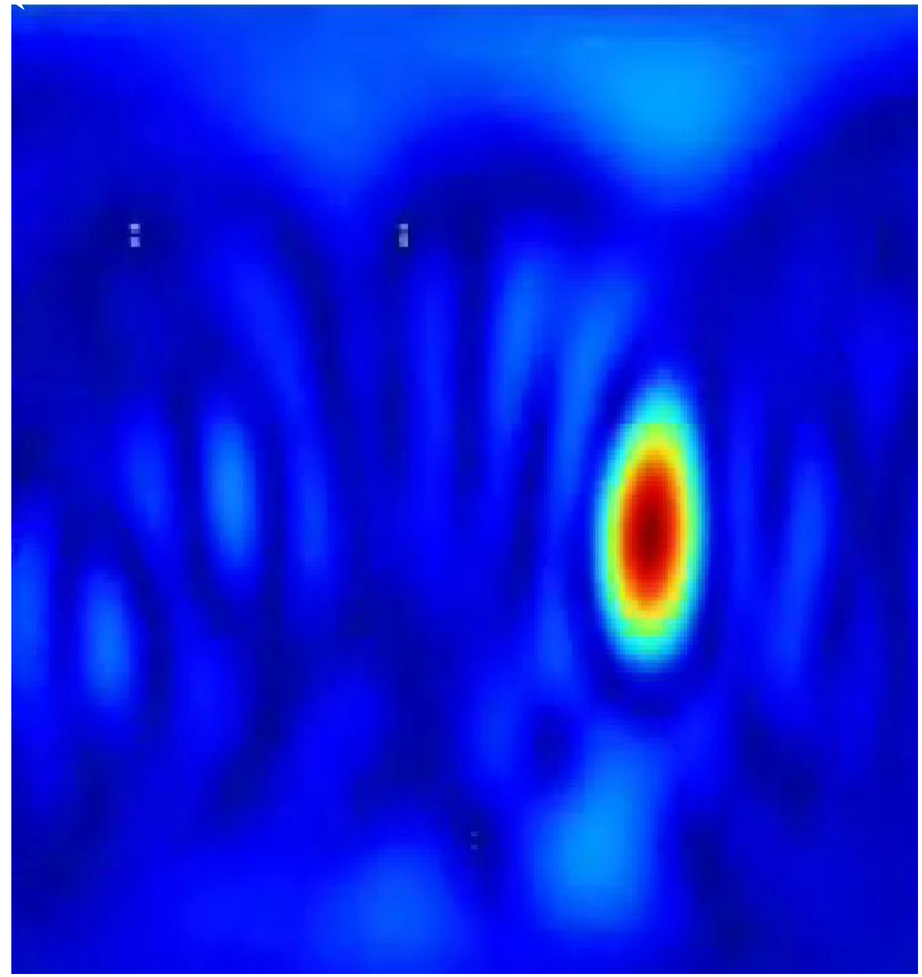
Synthetic results

- Reproduction of plane-waves truncated at various orders



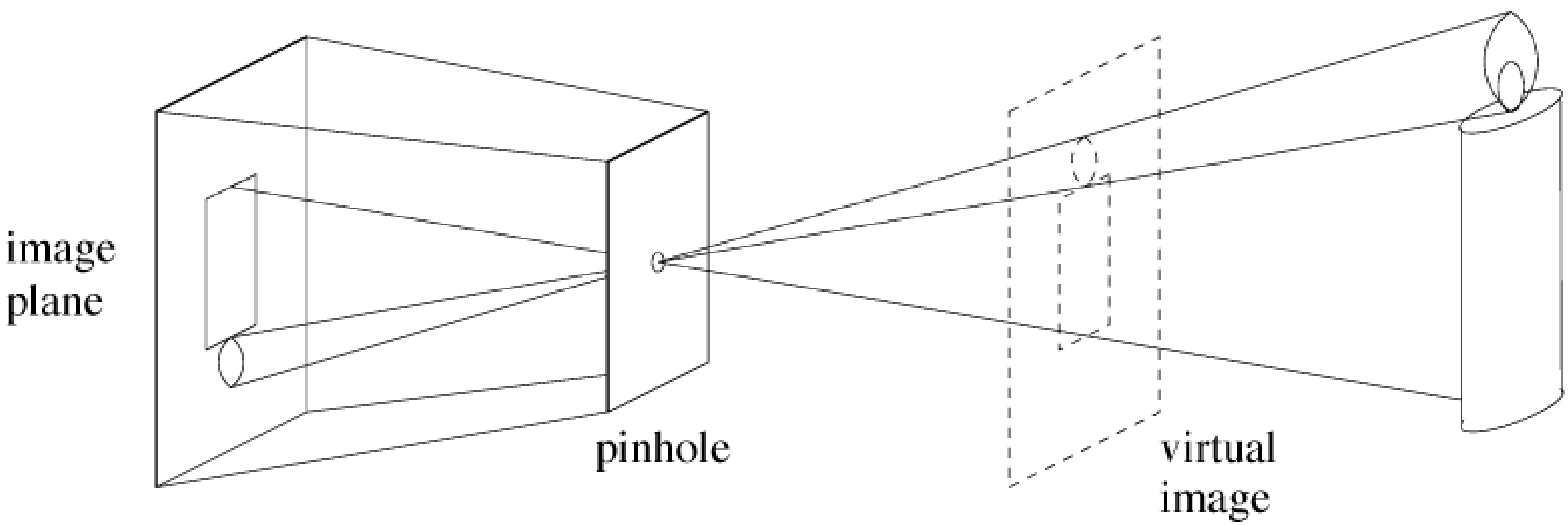
Audio Camera: Represent acoustic energy arriving from various directions as an image

- Each pixel intensity corresponds to acoustical energy in a given frequency band from direction (θ,φ)
- Map this to "Audio pixel" and compose audio image.
- Obtained via high-order beamforming using a spherical microphone array
  - Beamformer per pixel
- In this way we transform the spherical array into a camera for audio images

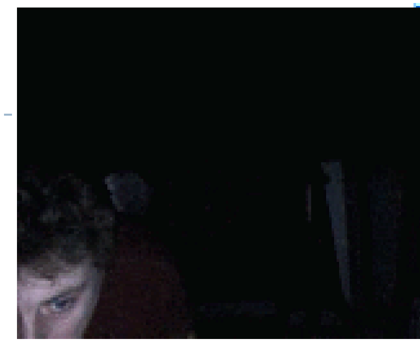


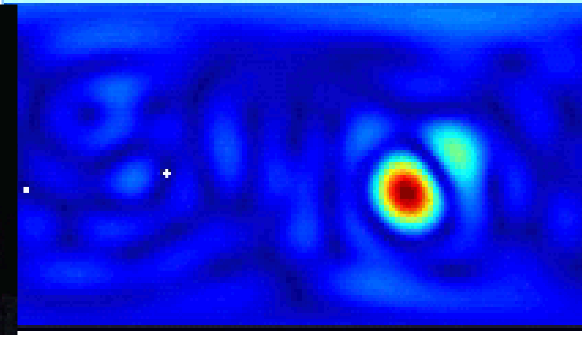
Approach: Combine microphone arrays and cameras

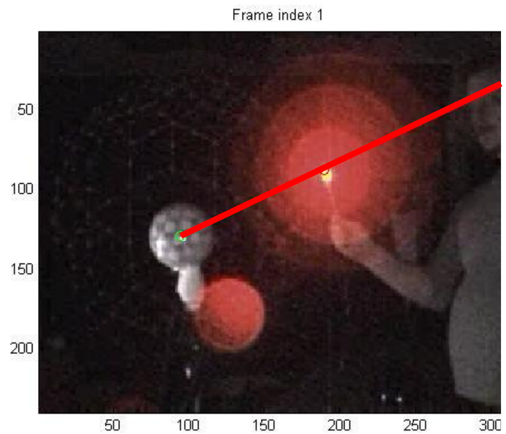
- Most Previous work
  - Audio and video processing is performed separately
  - Integration happens after estimations are performed
- Our work
  - Treat audio as a geometry sensor, and thus a "camera"
  - A joint analysis framework

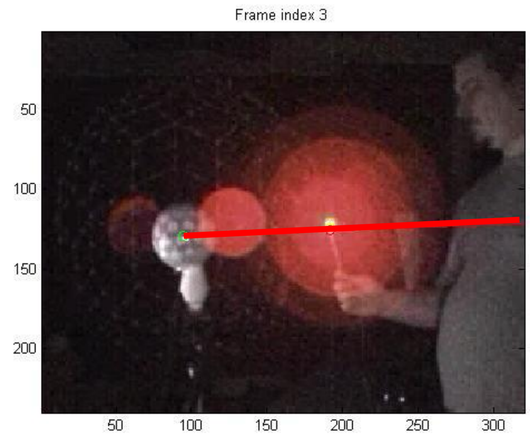


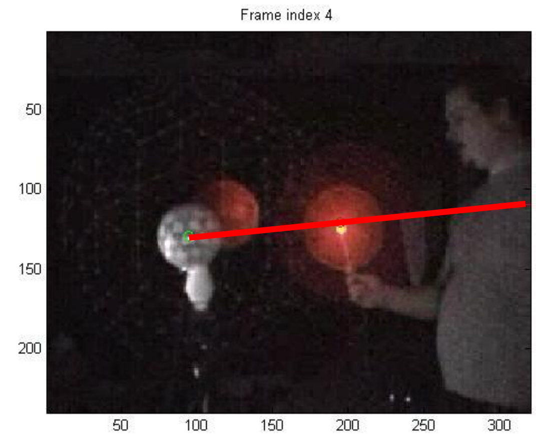
Calibration

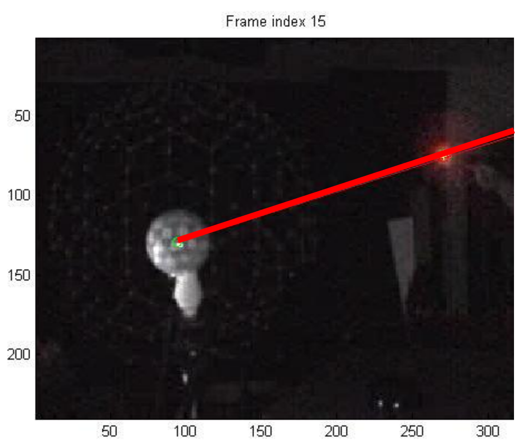


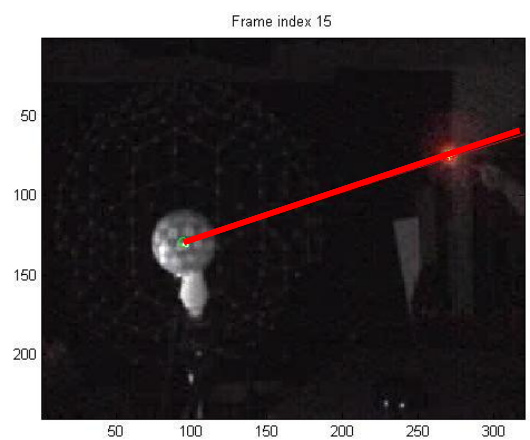






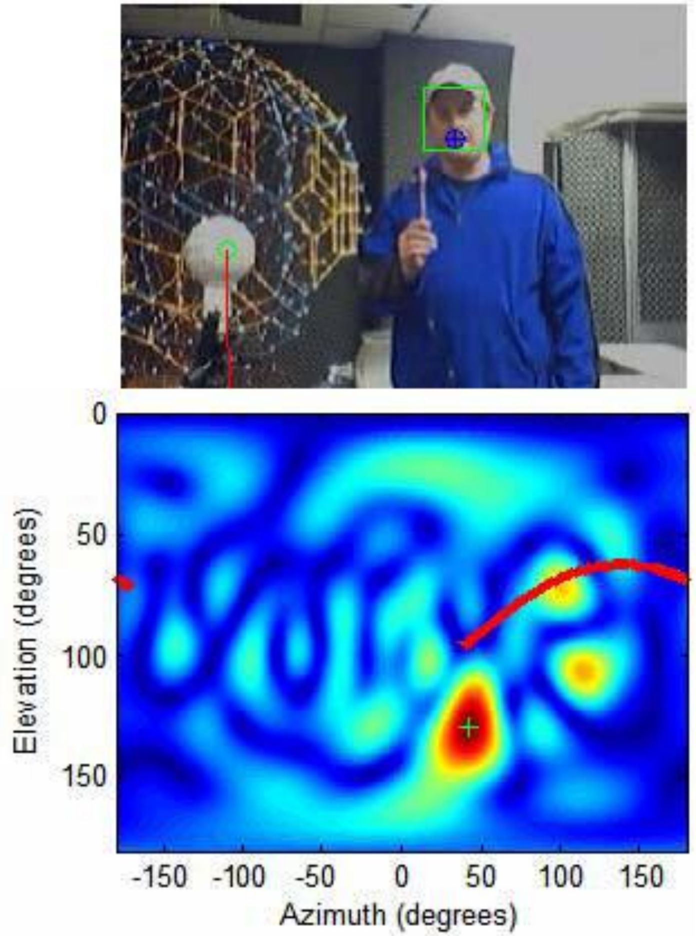






Vision Guided Beamforming

- Epipolar Constraint solves restricts search area
- Even in reverberant environments with complex distracters we can identify the beamforming direction.



Spherical beamforming for images

- Pros
  - Beamforming is digital … weights are known explicitly for each direction
  - Beamforming is independent
  - Can be done in time domain or frequency domain
- Cons
  - Each pixel requires the computation of a complex sum
  - Weights require special function evaluations
- Approach to speedup
  - Need to use speedup afforded by parallelism of computations
  - Use some math to reduce cost

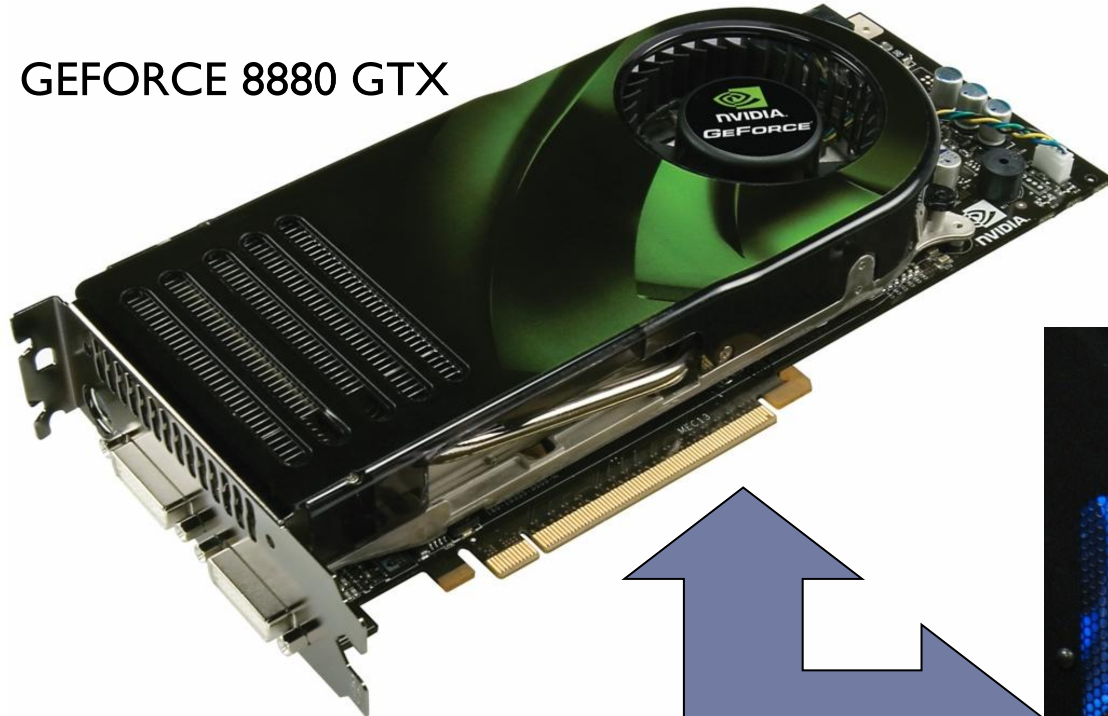
Making beamforming fast

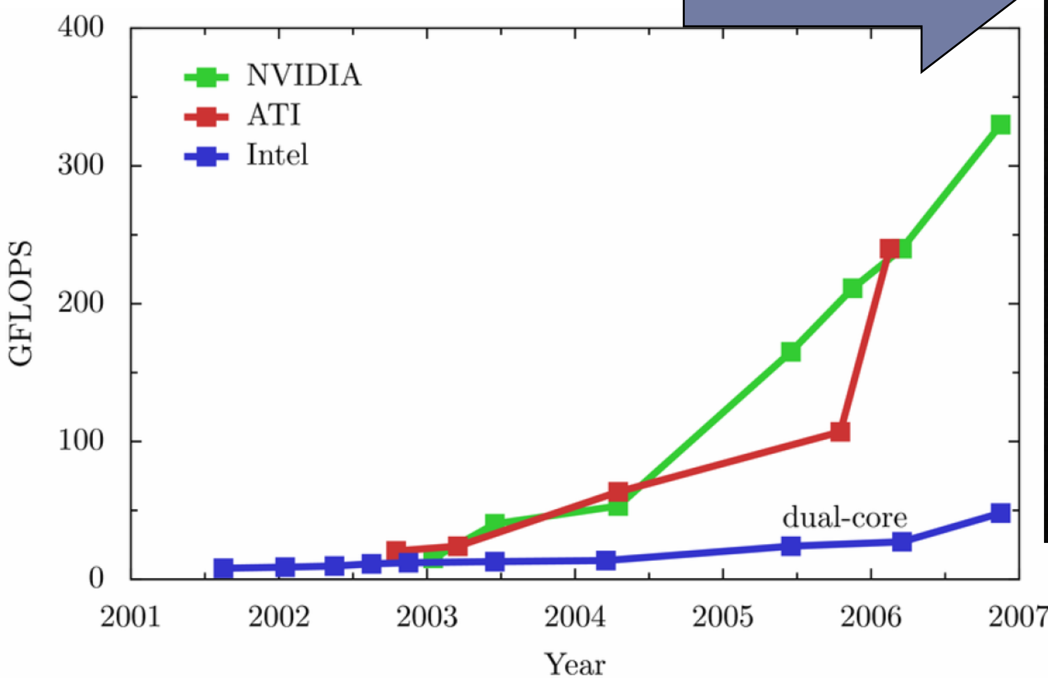
- Use spherical harmonics addition theorem

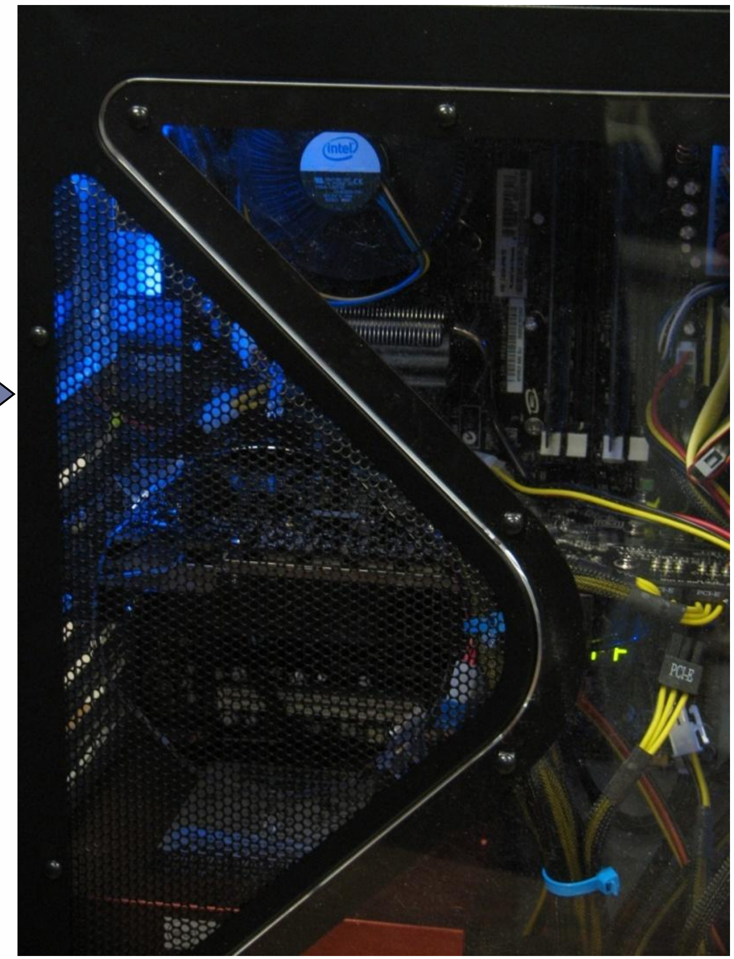
$$P_n(\cos\gamma) = \frac{4\pi}{2n+1}\sum_{m=-n}^{n} Y_n^{-m}(\mathbf{\Theta})\, Y_n^{m}(\mathbf{\Theta}_s)$$

  - Reduces *M* multiply and adds of spherical harmonics to one simple cosine evaluation
- Use Wronskian to simplify special function in b_n
- Use parallel processing
  - Each beamformer output is independent of the others
  - Trivially parallel
- Algorithm:
- For each pixel location
  - use table of known angle cosines for the given pixel, and given distribution of microphones,
  - perform weighted sum

GEFORCE 8880 GTX

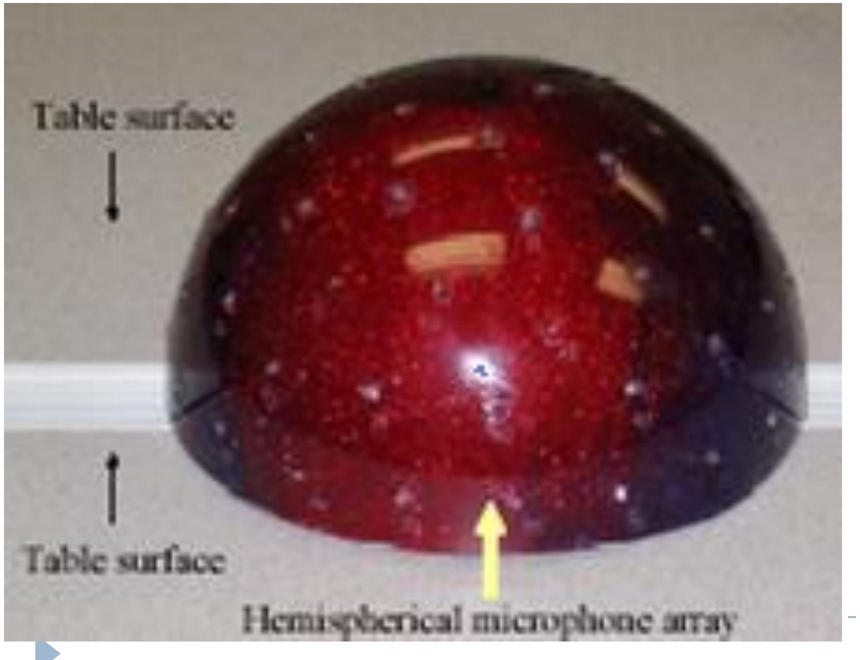




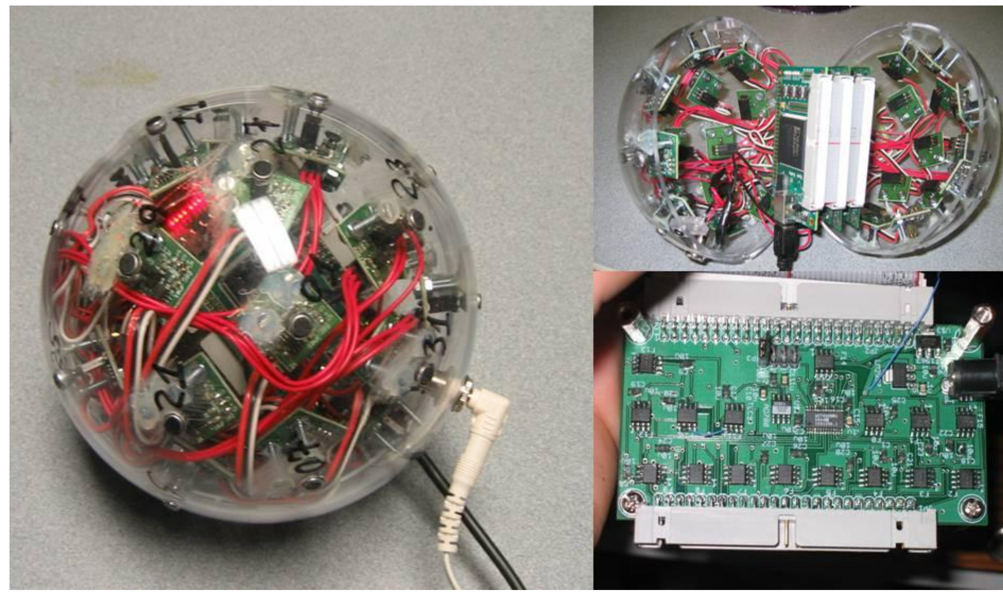


Evolution of the array architecture

- Lamp shade
- Bowling Ball
- PCI card based capture
  and PC -based processing



- 32-channel array
- 3 custom 12-bit ADCs boards
- Programmable anti-aliasing filter @ each channel
- 32 pre-amp mini-boards
- USB 2.0 interface via Xilinx FPGA
- Total speed up to 2.5 Msamples / second
- Digitally programmable

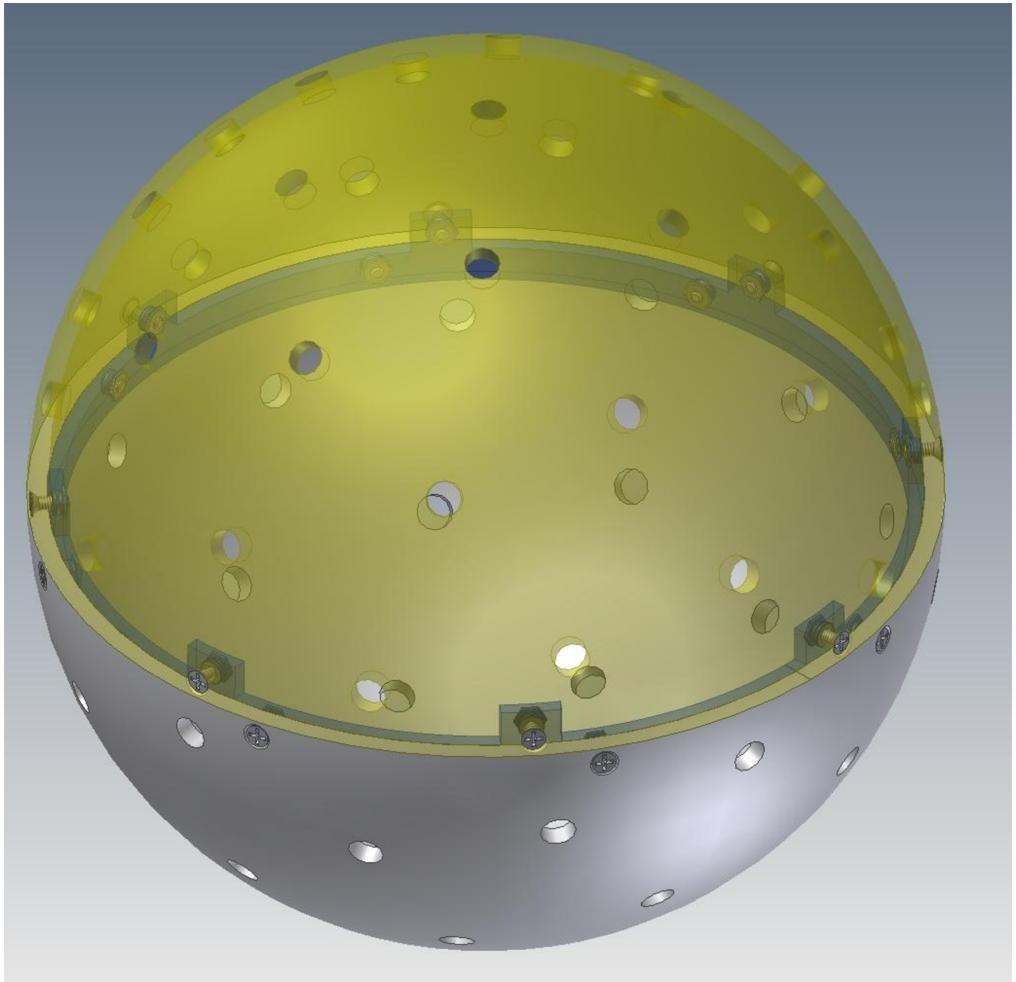


Newer arrays 2008-2009

- Integrated camera
- 64 microphones
- Power via USB or via separate power channel

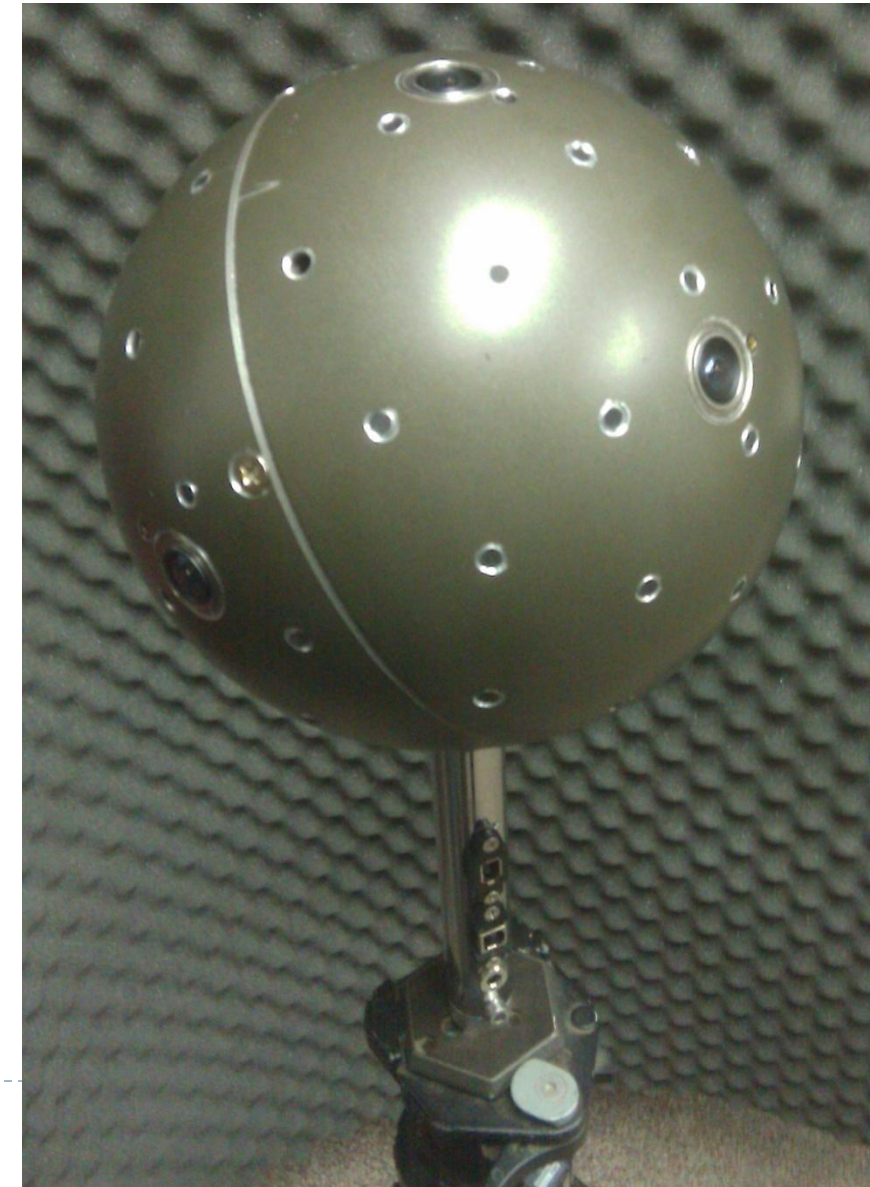
Newer arrays 2009-2010

- Integrated panoramic camera array
- 24 bit A/D
- Aluminum rugged construction
- Smaller electronics



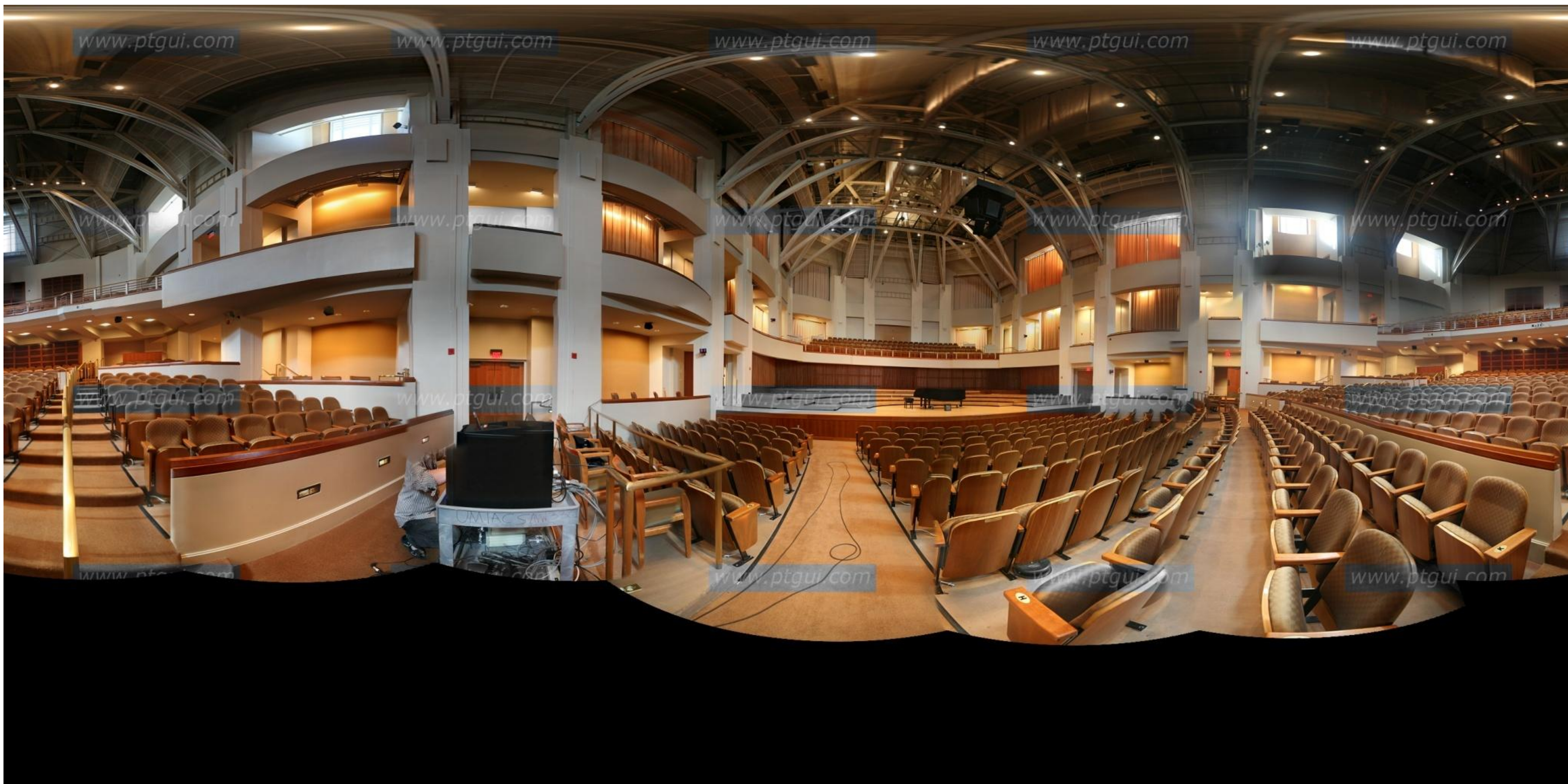
VisiSonics

- Company launched to develop audio visual spherical arrays and associated applications software
- Panoramic audio-visual real-time streams
- Contact adam.o@visisonics.com

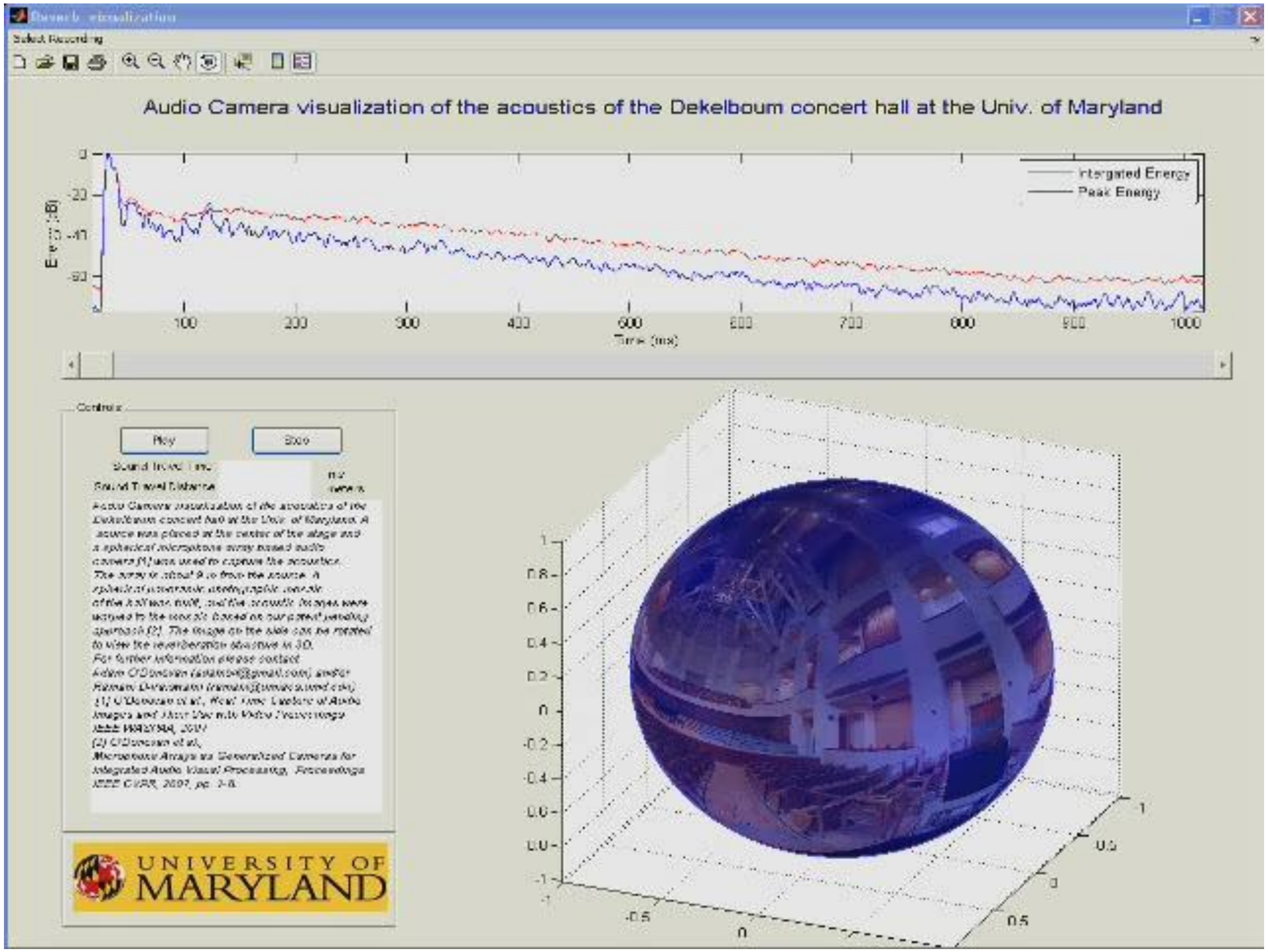


Dekelbaum theater at Clarice Smith Performing arts Center at UMD

- Mercator projection created from 24 snapshots

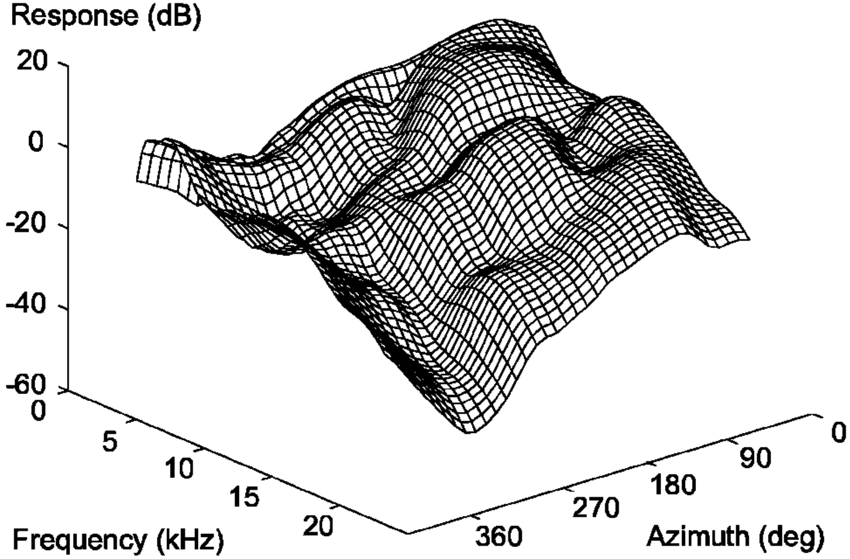


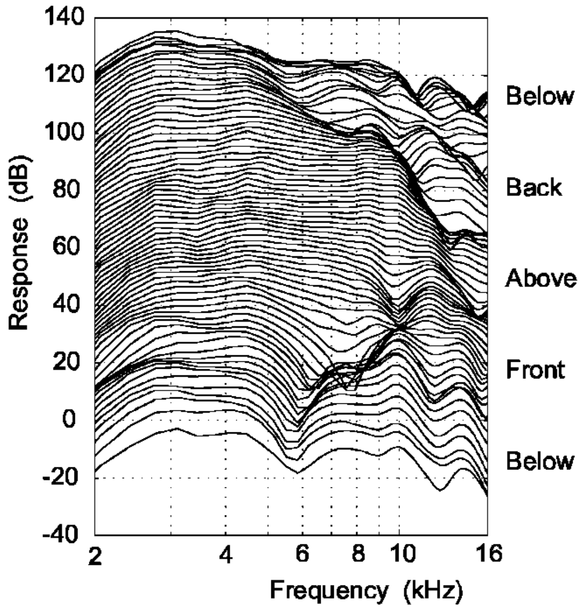
Studying Reverberation



Head Related Transfer Function

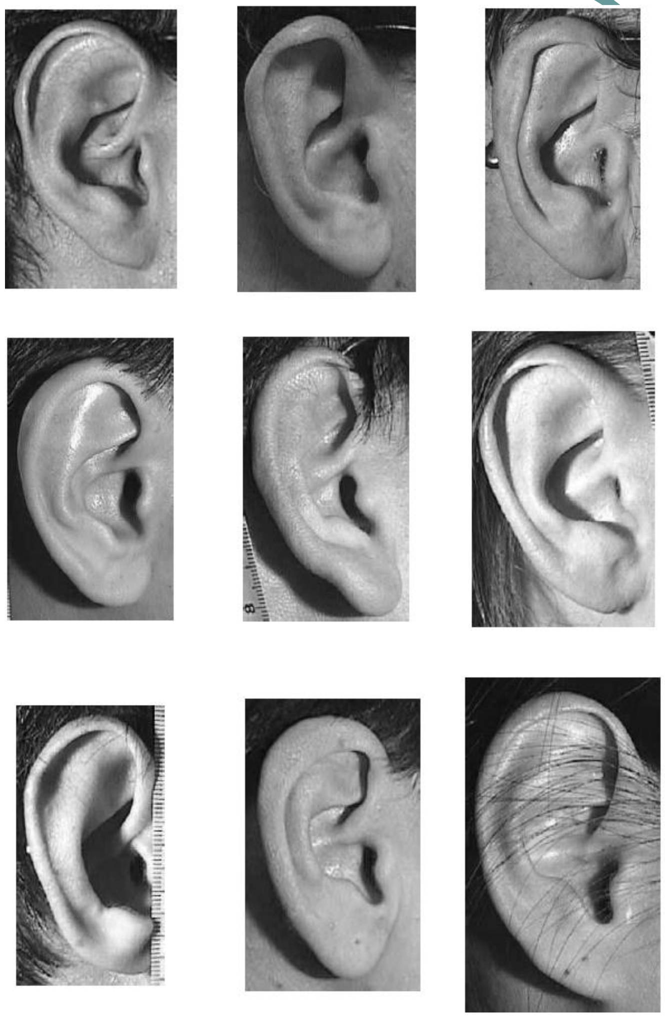
- Scattering causes selective amplification or attenuation at certain frequencies, depending on source location
  - Ears act as directional acoustic probes
  - Effects can be of the order of tens of dB
- Encoded in a Head Related Transfer Function (HRTF)
  - Ratio of the Fourier transform of the sound pressure level at the ear canal to that which would have been obtained at the head center without listener





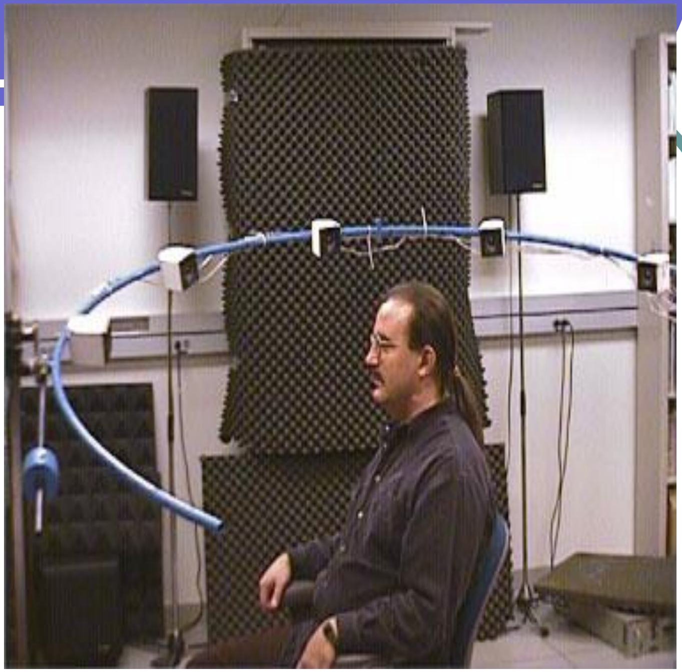
HRTFs are very individual

- Humans have different sizes and shapes
- Ear shapes are very individual as well
  - Before fingerprints, Alphonse Bertillon used a system of identification of criminals that included 11 measurements of the ear
- Even today ear shots are part of
  - Mugshots & INS photographs
- If ear shapes and body sizes are different
  - Properties of scattered wave are different
  - HRTFs will be very individual
- Need individual HRTFs for creating virtual audio



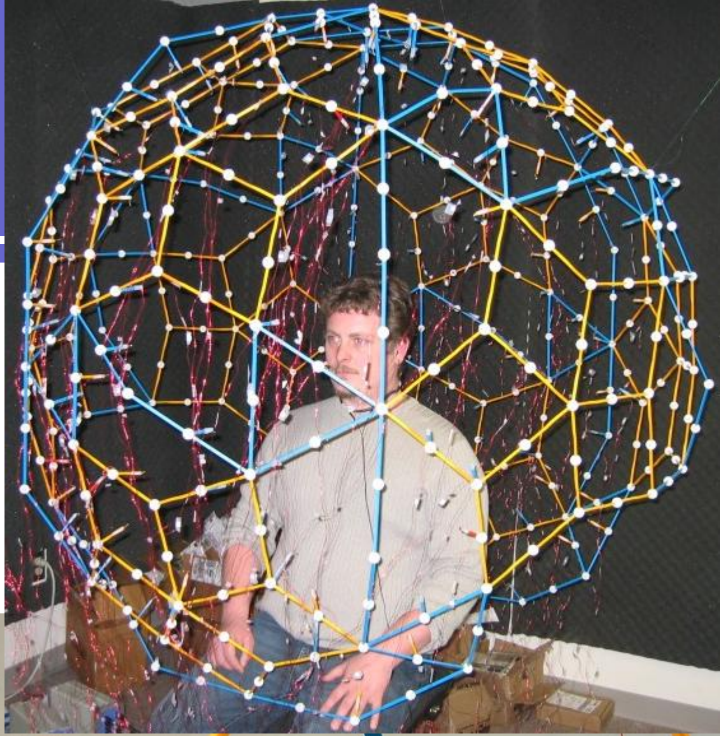
Typically measured

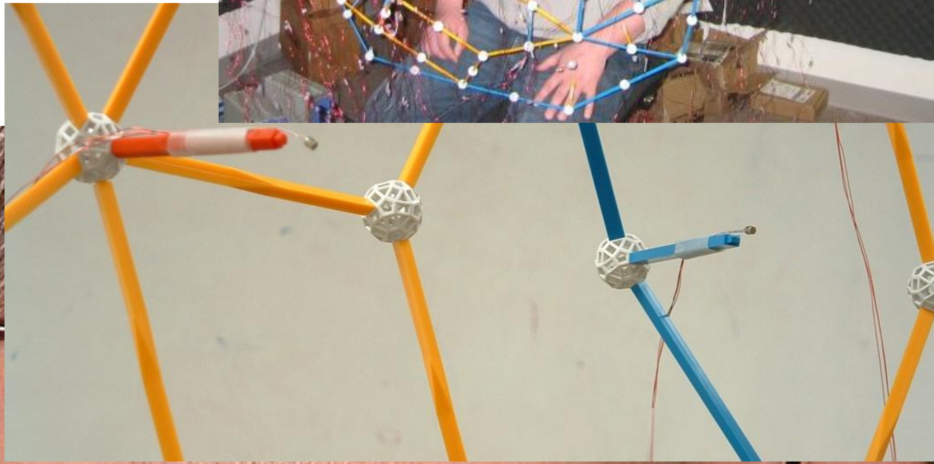
- Sound presented via moving speakers
- Speaker locations sampled
  - e.g., speakers slide along hoop for five different sets, and hoop moves along 25 elevations for 50 x25 measurements
- Takes 40 minutes to several hours
- Subject given feedback to keep pose relatively steady
- Hoop is usually >1m away (no range data)

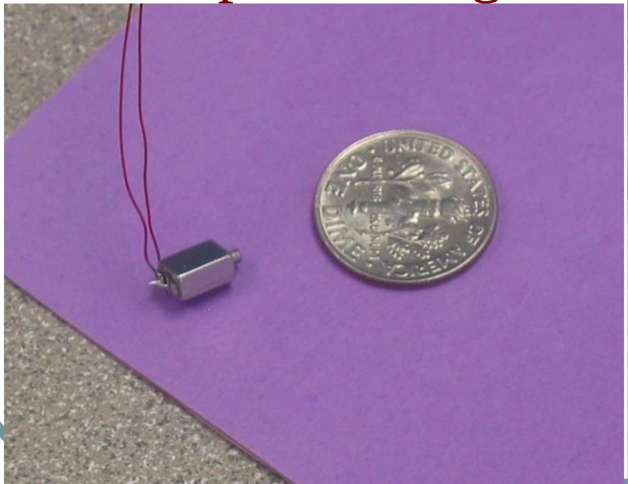


Approach

- Turned out headphone drivers
- Array of tiny microphones
- Send out a highpass signal and measure received signal
- Use analytical anthropometric representation for low frequencies
  and compose
- Extrapolate range

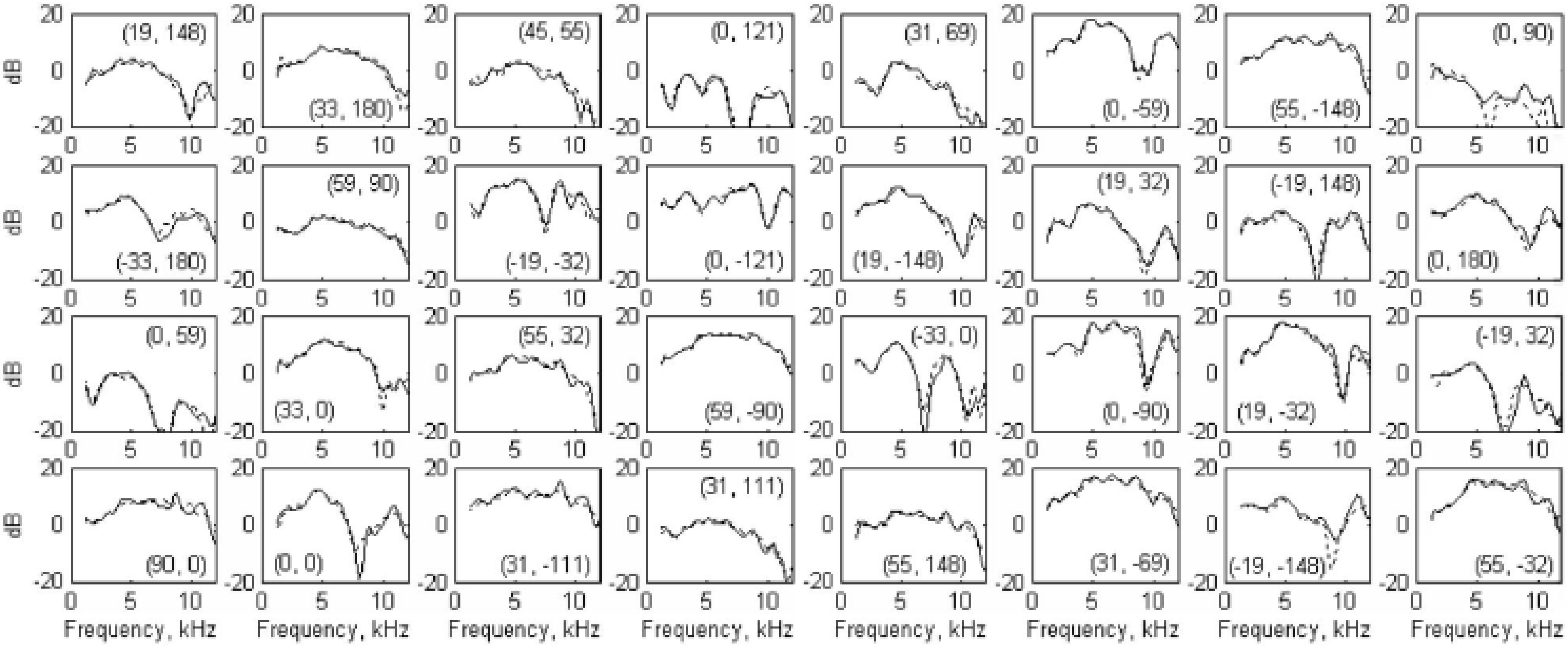






Comparisons

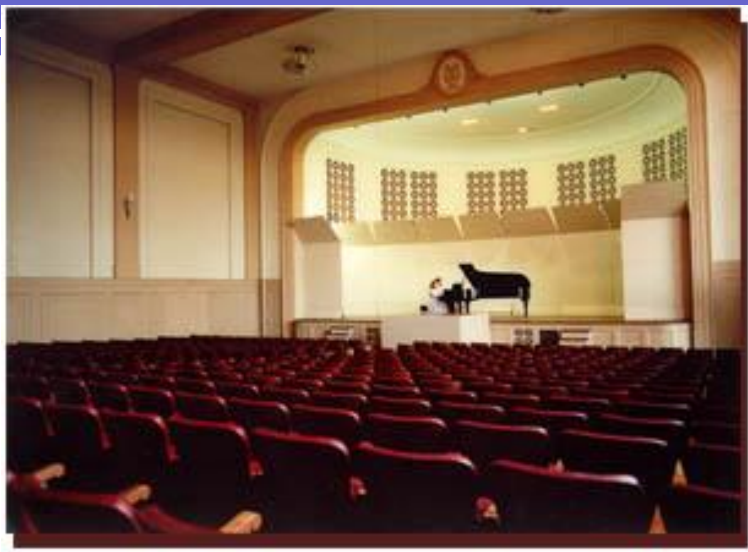
- Direct vs. Reciprocal (Zotkin et al. 2006, JASA)



Dmitry N. Zotkin, Ramani Duraiswami, Elena Grassi, and Nail A. Gumerov, "Fast head-related transfer function measurement via reciprocity," Journal of the Acoustical Society of America, pp. 2202-2214, volume 120, no. 4, October 2006

Decouple HRTFs and Recordings

RECORDING



PLAYBACK

- Place microphones at a remote location (e.g. concert hall)
- Replay spatialized audio at a remote location
- Must play it for many users
- Use HRTFs at the client side

HRTF based playback

- Scattering response of anatomy, measured at ear locations to plane waves from direction (θ, ϕ)

$$H^L(k;\theta,\varphi) = \frac{\psi^L(k;\theta,\varphi)}{\psi^0(k;r)}, \quad H^R(k;\theta,\varphi) = \frac{\psi^R(k;\theta,\varphi)}{\psi^0(k;r)},$$

- We have decomposed the sound field in to plane-waves. So all we need to do is take the product and sum

$$\psi^L = \sum_{l=1}^{L_Q} w_l H_l^L \mu_{in}(\mathbf{s}_l), \quad \psi^R = \sum_{l=1}^{L_Q} w_l H_l^R \mu_{in}(\mathbf{s}_l),$$

- **No need to localize sound sources first!**

HRTF-based spatial scene capture and rendering algorithm

Initial Input: array radius, microphones and HRTF locations, desired order p

Preliminary (offline) processing:

quadrature weights of order p for microphone and HRTF grid

using error bound determine appropriate p for each k.

Online processing of data frames:

For frame i

  Input data from each of the L_M microphones of length T

  Prepare data and convert to frequency domain

  for k ($k_{\min}$ to $k_{\max}$)

    select $p(k)$

    do fitting at the HRTF grid nodes

      build $\psi(k)$ at the sphere center

      Evaluate $\psi_{Left}(k)$ and $\psi_{Right}(k)$ at the HRTF grid nodes

  next k

  Perform an Inverse FFT to obtain sound in the time domain

  Perform any filtering modifications

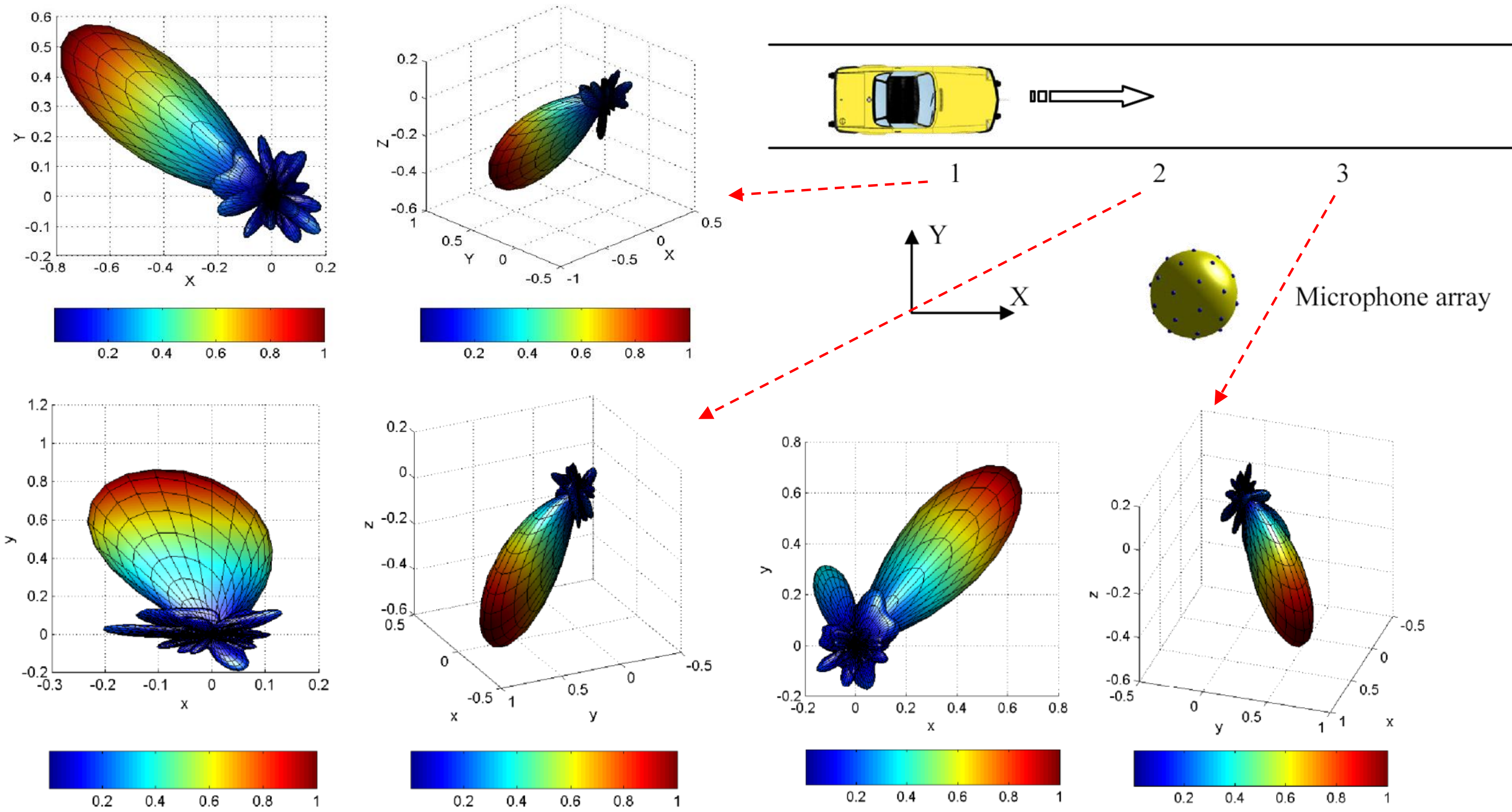
  Playback

next frame

Beamforming a Traffic Scene

Beamforming results

Experiment setup



HRTFs can be computed

Wave equation: $$\frac{\partial^2 p'}{\partial t^2} = c^2\left(\frac{\partial^2 p'}{\partial x^2}+\frac{\partial^2 p'}{\partial y^2}+\frac{\partial^2 p'}{\partial z^2}\right) = c^2\nabla^2 p'$$

Fourier Transform from Time to Frequency Domain $$P(x,y,z,w) = \int_{-\infty}^{\infty} p'(x,y,z,t)e^{-i\omega t}dt$$

Helmholtz equation: $$\nabla^2 P + k^2 P = 0$$

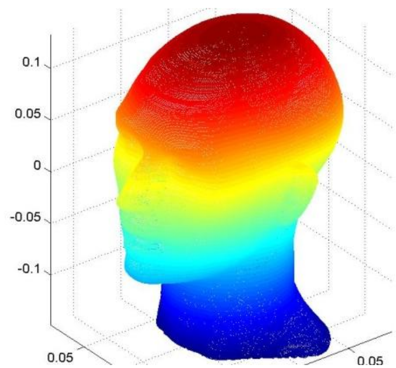
Boundary conditions:

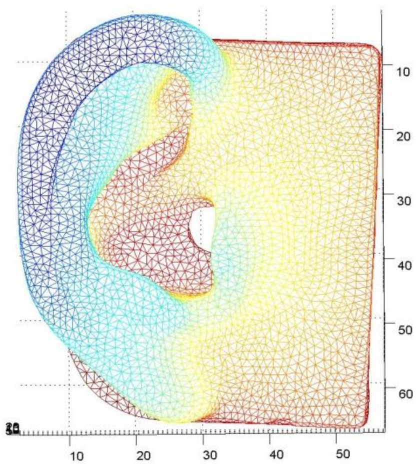
Sound-hard boundaries: $$\frac{\partial P}{\partial n} = 0$$

Sound-soft boundaries: $$P = 0$$

Impedance conditions: $$\frac{\partial P}{\partial n} + i\sigma P = g$$

Sommerfeld radiation condition $$\lim_{r\to\infty} r\left(\frac{\partial P}{\partial r} - ikP\right) = 0$$

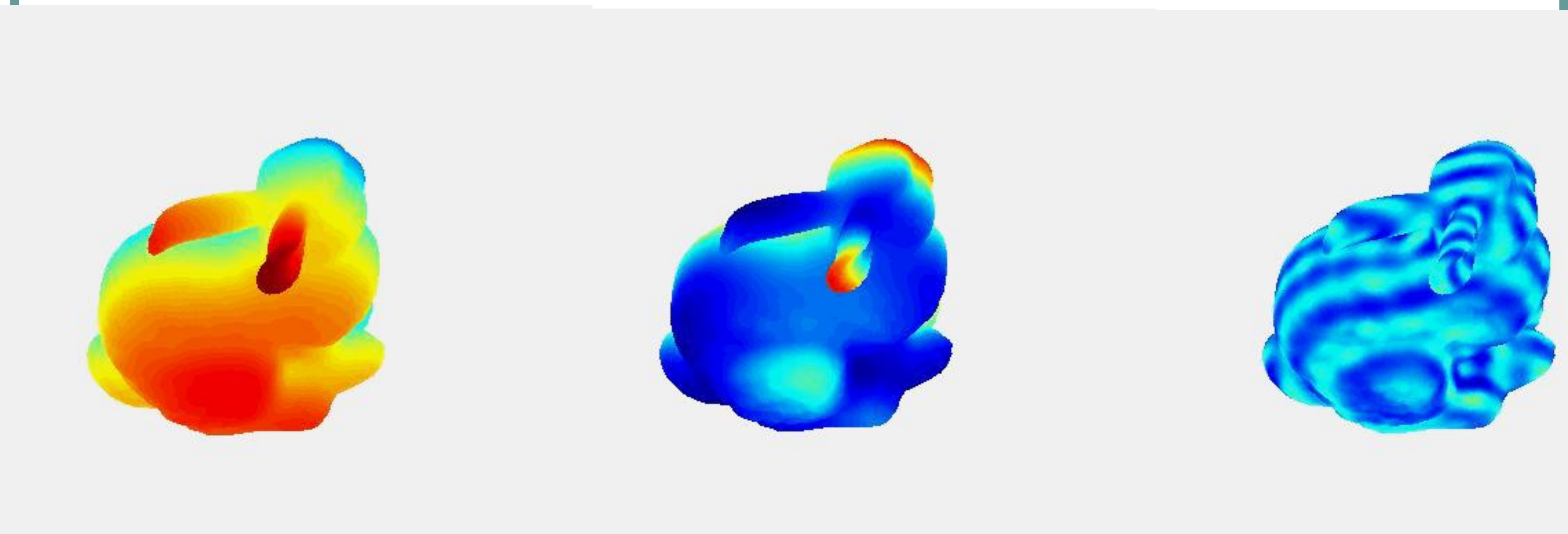




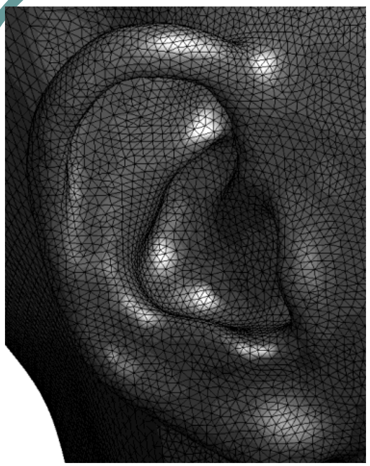
FMM Accelerated BEM

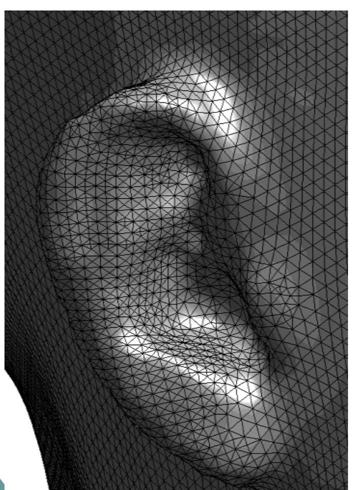
Nail A. Gumerov and Ramani Duraiswami. FMM accelerated BEM JASA 2009.

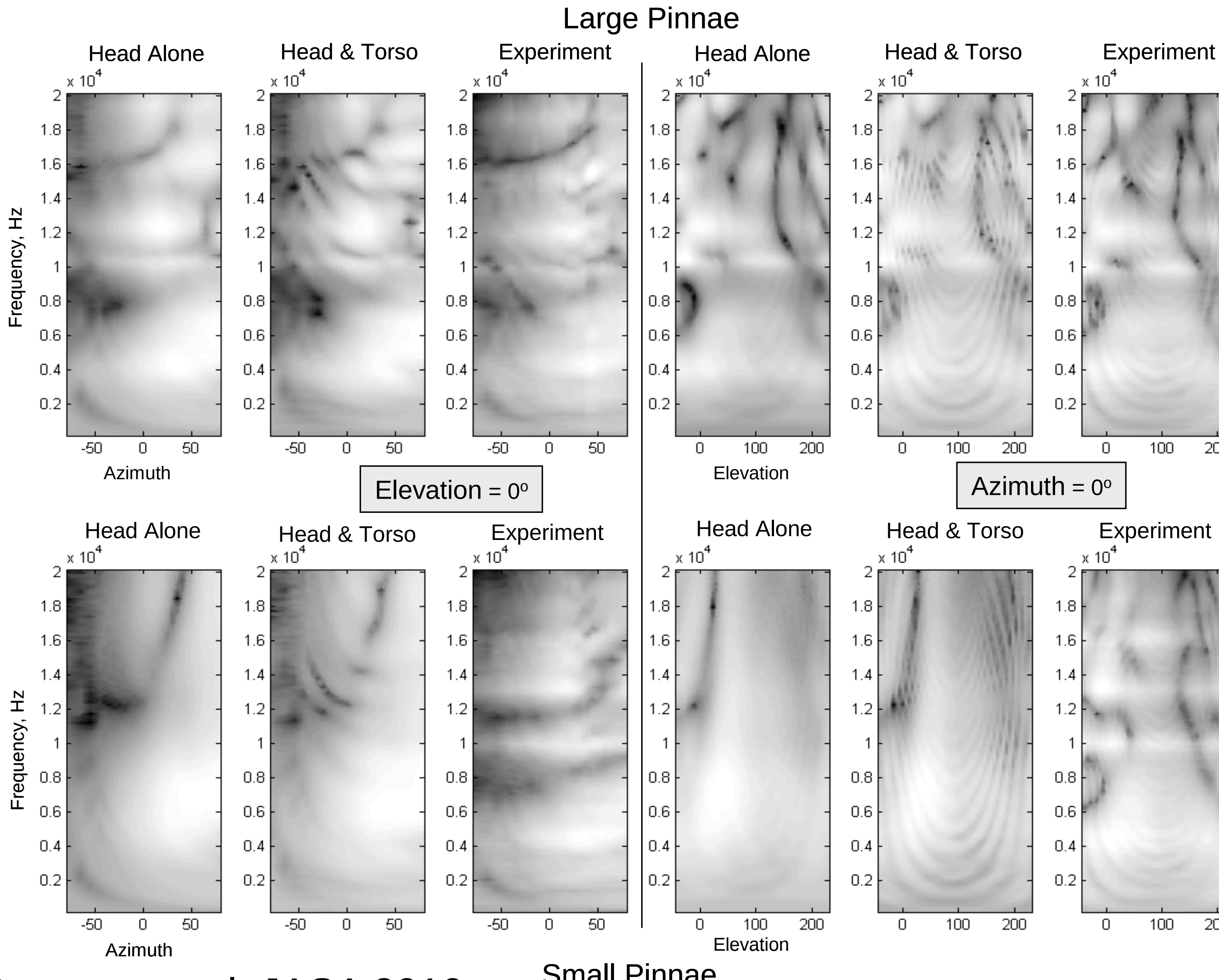
f=250 Hz f=2.5 kHz f=25kHz



Computing HRTFs: Effect of the mesh







Gumerov et al. JASA 2010

Other audio visual research

- Speaker identification using arrays
- Environment identification
- Lenses for audio cameras (telephoto!)
- Scene reproduction
- Using reverberation to improve beamforming
- …

Conclusions

- Higher order ambisonics (sound scene analysis) can be done rigorously
  - With error bounds
- Error bounds are too strong
  - Interesting things are done with lower order
  - We have lot of data (e.g. 64 mics times 16 bits times 44.1 kHz)
- Can we trade prior knowledge about the signal/environment to improve recognition?
  - Sparse representations/compressed sensing
- Knowing the scene visually helps in building prior knowledge