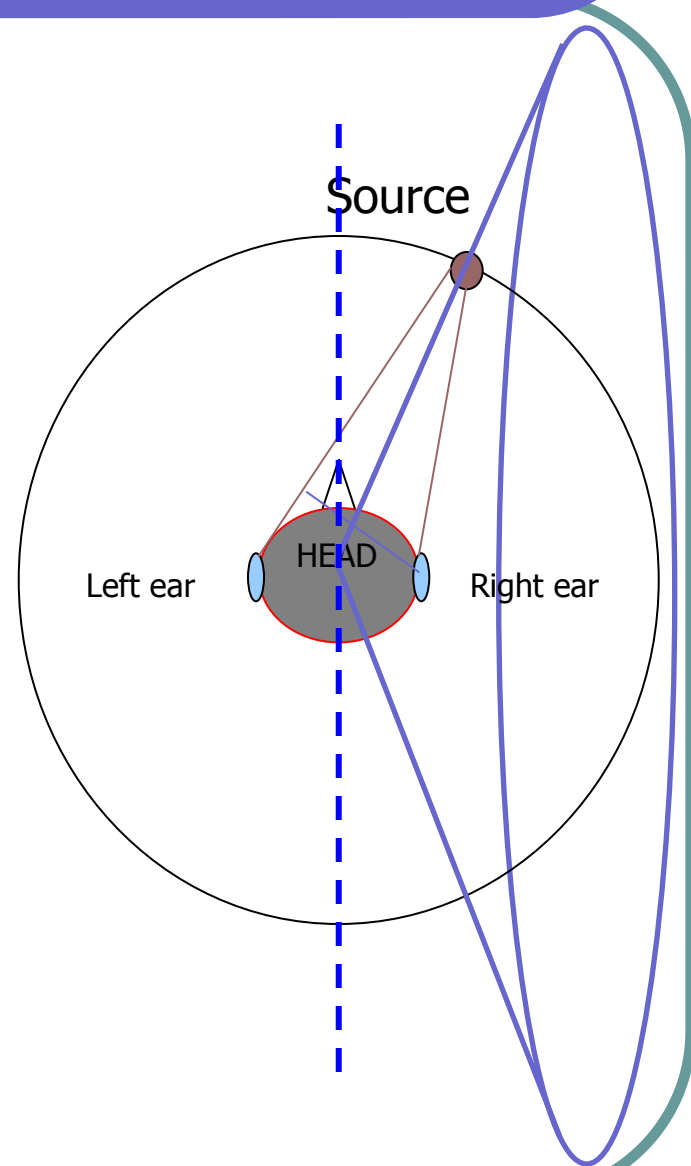


Introduction to HRTFs

How do we perceive sound location?

- Initial idea: Measure attributes of received sound at the two ears
- 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
- Level differences also vanishingly small
- Other mechanisms necessary to explain
 - Scattering of sound
 - Off our bodies
 - Off the environment
 - Purposive Motion



Sound and Human Spaces

- Sound wavelengths comparable to human dimensions and dimensions of spaces we live in.

- $f\lambda = c$

- When $\lambda \gg a$ wave is unaffected by object

$$\lambda \sim a$$

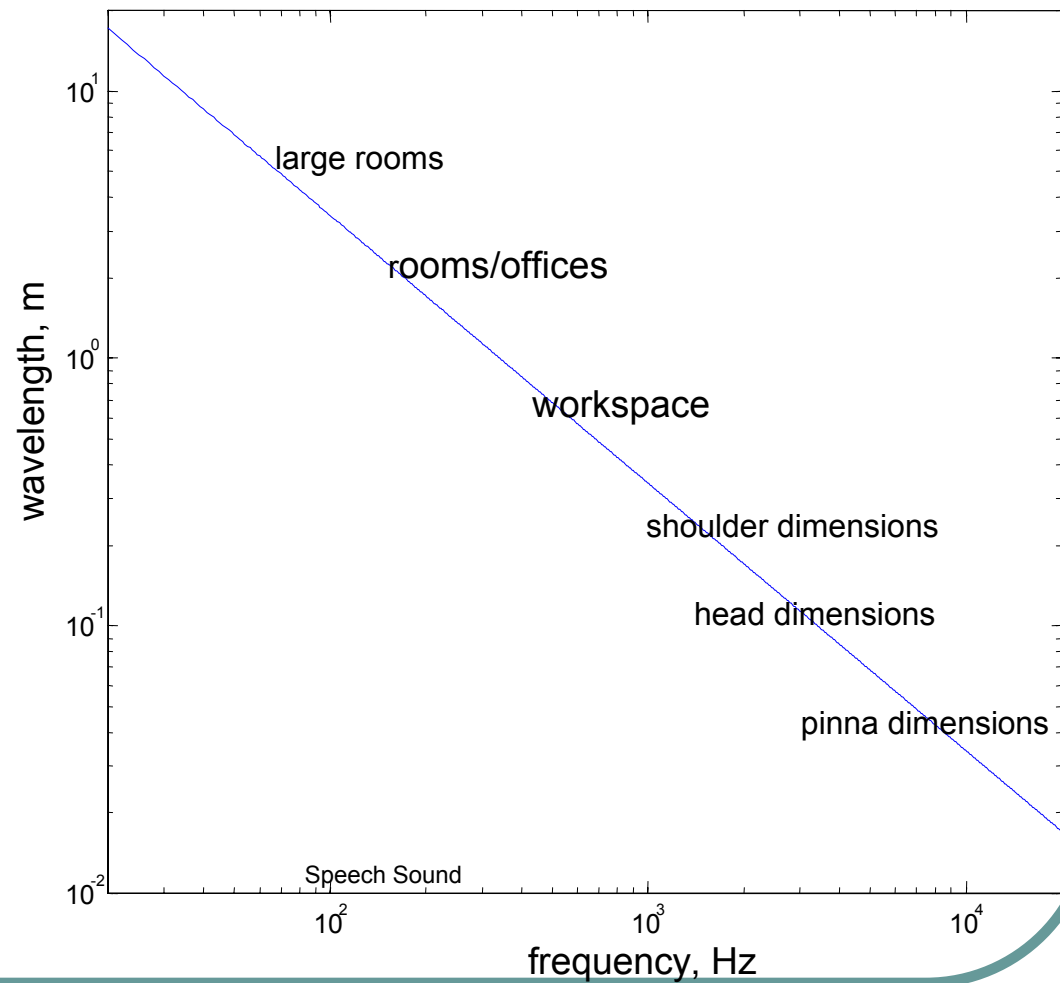
behavior of scattered wave is complex and diffraction effects are important.

$$\lambda \ll a$$

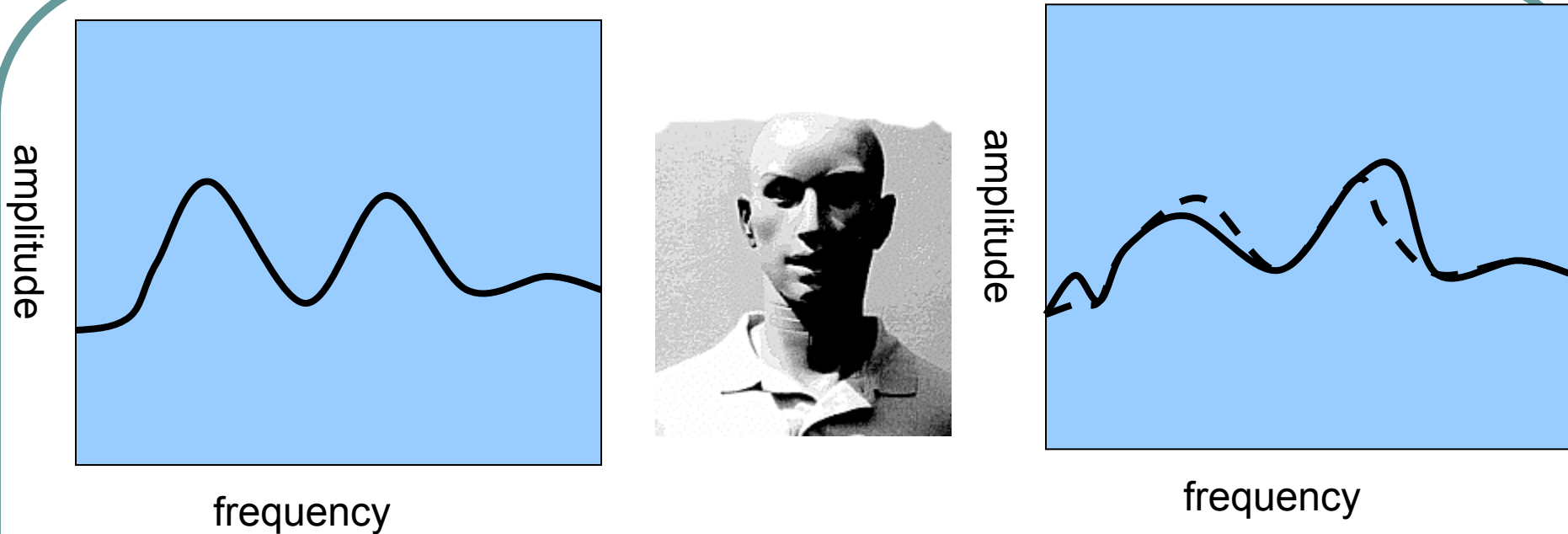
wave behaves like a ray

wavelengths are comparable to our rooms, bodies, and features

Not an accident but evolutionary selection!



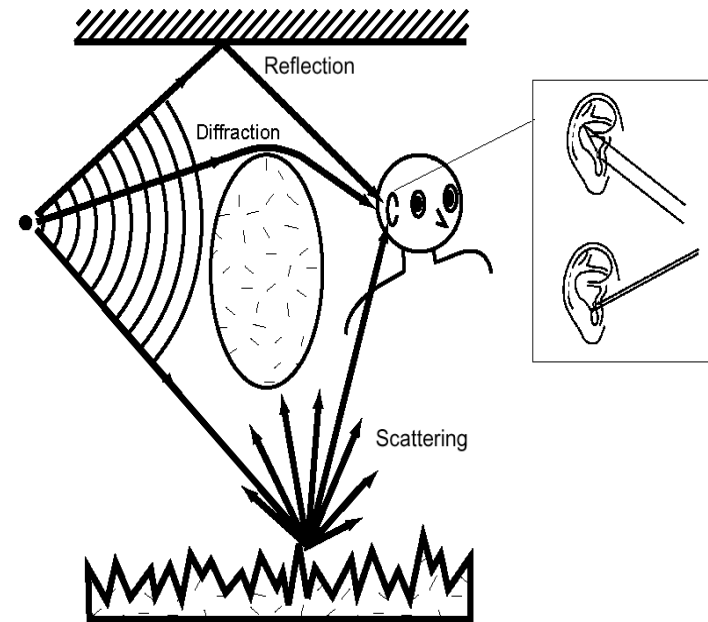
Scattering Cues



- Rather like moving a CD in light.
- As light direction or CD position changes ... colors change
- Neural system decodes these changes

Modeling Sound Scattering

- Interactions change received sound waves
- Scattering of body and ears
 - Bodies ~ 50 cm
 - Heads ~ 25 cm
 - Ears ~ 4 cm
 - Not much multiple scattering
- Scattering off surroundings
 - Rooms $\sim 2\text{m} - 10\text{m}$
 - More multiple scattering
 - Larger sizes \Rightarrow lower frequencies



Because of this separation of scales we can model these effects independently

Scattering characterization:

- 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)

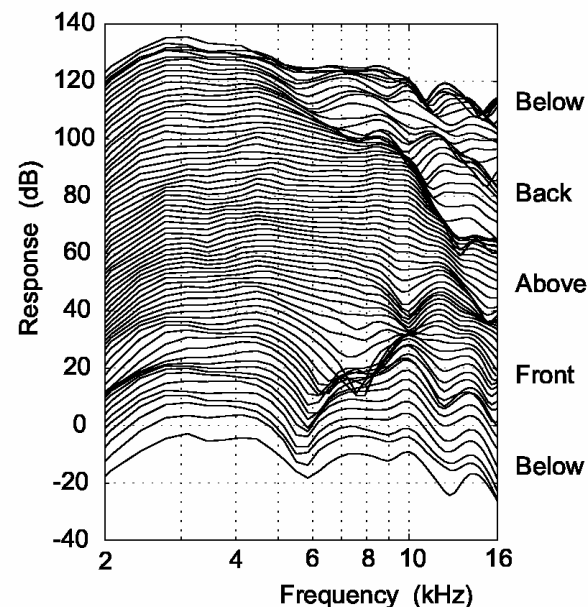
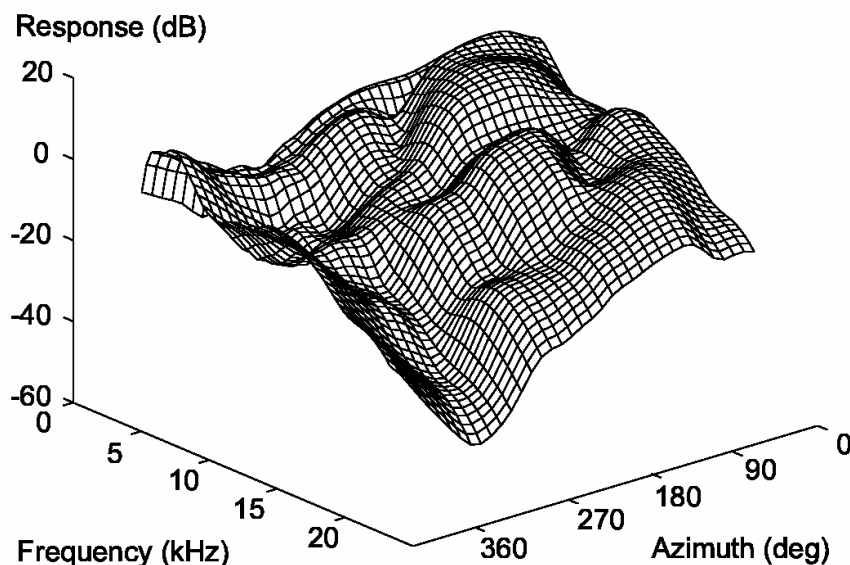
Creating Auditory Reality

- Reintroduce 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
- Three Legged Stool

Head Related Transfer Function

Head Related Transfer Function

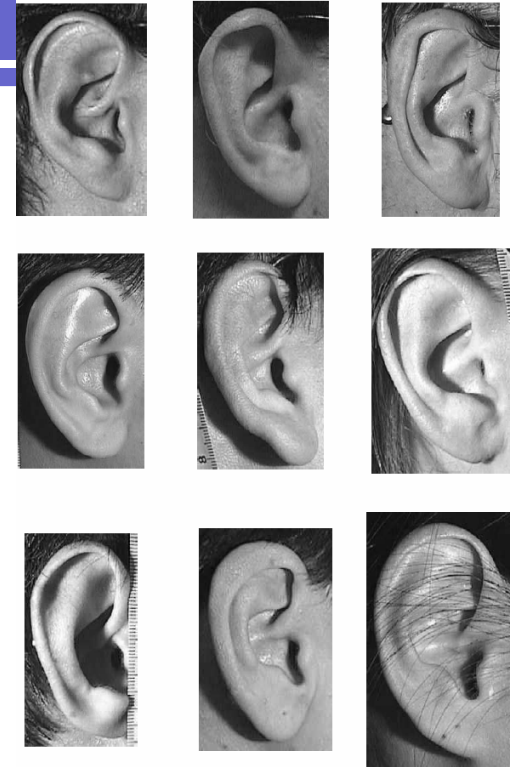
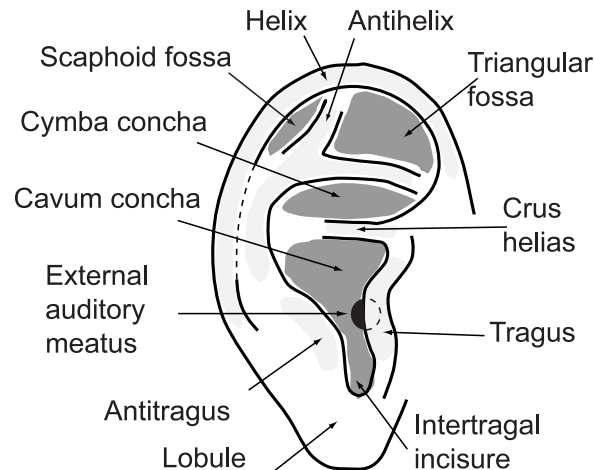
- Scattering causes selective amplification or attenuation at certain frequencies, depending on source location
 - Ears act as directional acoustic probes
 - Effects can be of the order of tens of dB
- Encoded in a Head Related Transfer Function (HRTF)
 - Ratio of the Fourier transform of the sound pressure 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 sized heads and torsos
- More importantly: ear shapes are very individual as well
- If ears are different \Rightarrow properties of scattered waves from them will be different.
- HRTFs will have to be individual.

Several spatially distributed open cavities and protuberances



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 x 25 measurements
- Takes 40 minutes to several hours
- Subject given feedback to keep pose relatively steady
- Hoop is usually $>1\text{m}$ away (no range data)



HRTFs can be calculated

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

$$p'(x, y, z, t) = \int_{-\infty}^{+\infty} P(x, y, z; \omega) e^{-i\omega t} d\omega$$

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 boundary conditions:

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

Sommerfeld radiation condition
(for infinite domains):

$$\lim_{r \rightarrow \infty} r \left(\frac{\partial P}{\partial r} - ikP \right) = 0$$

HRTFs can be computed

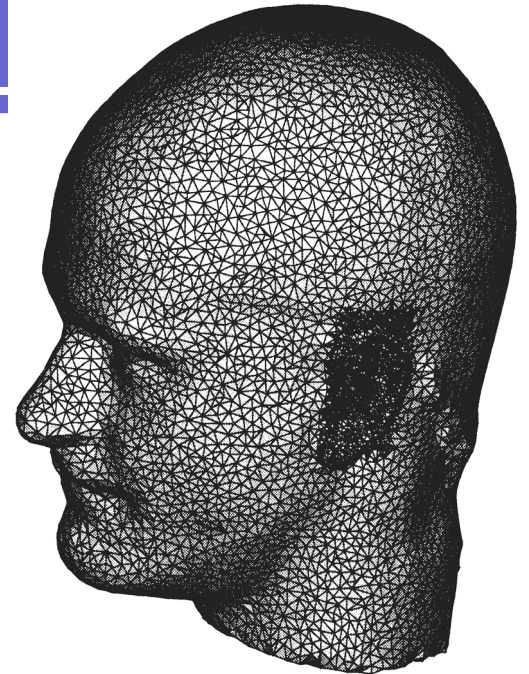
- Boundary Element Method
- Obtain a mesh
- Using Green's function G

$$G(\mathbf{x}, \mathbf{y}) = \frac{e^{ik|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}$$

- Convert equation and b.c.s to an integral equation

$$C(x)p(x) = \int_{\Gamma_y} \left[G(x, y; k) \frac{\partial p(y)}{\partial n_y} - \frac{\partial G(x, y; k)}{\partial n_y} p(y) \right] d\Gamma_y$$

- Need accurate surface meshes of individuals



Issues with Measured HRTFs

- HRTF measurements take a while to do
 - Tedious both for the subject and experimenter
- Angular resolution necessary is not clear
- Measurement angular resolutions vary considerably
- Range HRTF data is essentially unavailable
 - May be important for simulating nearby sources
- Despite this sampling, we still need interpolation
- Because of the expense of measuring HRTFs and their relative scarcity
 - “... if we only had individualized HRTFs our system would be perfect and we could do wonders ...”

Algorithms for VAS Synthesis

- We developed a set of algorithm and a system for creation of the virtual auditory space (VAS)
- Render sounds so that they appear to be external and come from some point in space
- Goals of the system:
 - Deal with “technical” issues of sound rendering
 - Audio rendering pipeline (similar to graphics notion)
 - Provide a baseline system into which computed HRTFs can be plugged in for psychophysical testing
 - Test some simple HRTF customization methods

Synthesis of VAS

- Three sets of cues to be reproduced:
 - Static
 - Dynamic
 - Environmental
- Static: HRTF set
- Dynamic: Head tracking (Polhemus)
- Environmental: Room model

Static Localization Cues

- Set of HRTFs stored as Impulse Responses
 - Convolve the sound with appropriate IRs
- HRTF interpolation
 - Interpolate amplitudes and add back the ITD

Dynamic Localization Cues

- Playback is done through headphones
 - Head movements (rotations + translations) must be compensated for
 - Otherwise, the source rotates with you and is perceived as being inside the head
 - Polhemus sensor is mounted on headphones
- Simple geometric computations stabilize the virtual audio scene w.r.t. moving listener

Environmental Cues

- Naive rendering with HRTF creates an audio scene that is “flat” and does not have depth
 - Sound is perceived at the correct DOA but excessively close to the head
- The reason is that we do not live in anechoic rooms
- Room reverberation is present and is important
 - Provides externalization
 - Provides depth perception

Environmental Modeling

- Can be modeled (image model) or measured in “good” environment to create pleasant perception
- Image model creates virtual sources that are reflections of the true source in room walls
- These virtual sources also have their own directions and must be rendered as such
 - There are now many sources to render and long overall impulse response (seconds)

Rendering Pipeline

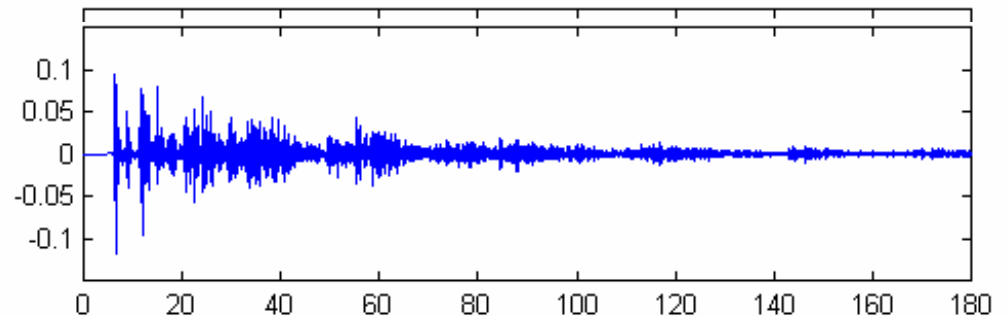
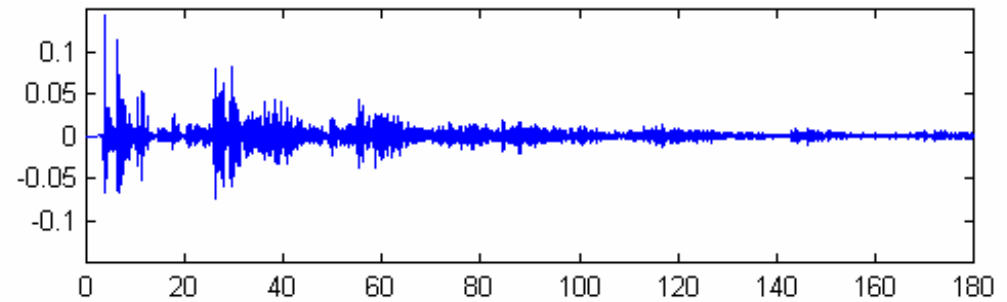
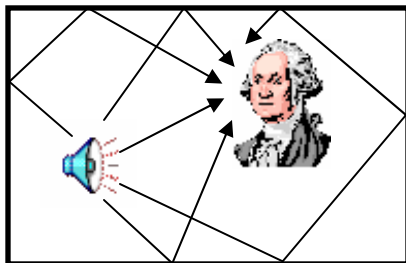
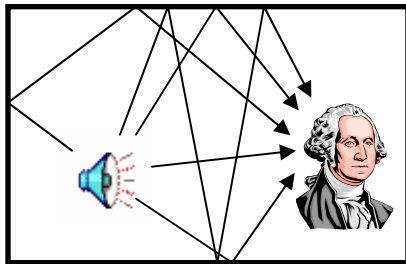
- Convolution in time domain is very slow [$O(N*N)$]
- Convolution in frequency domain is fast [$O(N*\log N)$] but introduces some latency
- Superimpose image sources IRs in one rendering filter and do frequency-domain convolution with it
- In our pipelined processing, the unavoidable latency is used to update the rendering filter
 - Recompute image sources in response to motion
 - The rest of filter is fixed for a given room geometry
 - Adjust on the fly to the computational power available

Breaking up the Filter

- Convolution is linear
- Early reflections are more important and time separated
 - Important for determining range
- Later reflections are a continuum
 - important for “spaciousness,” “envelopment,” “warmth,” etc.
- Create early reflections filter on the fly
 - reflections of up to 5th or 6th order (depending on CPU resources)
 - These are convolved with their HRTF
 - Stick appropriate HRIR at the arrival location
- Tail of room impulse response is approximated depending on room size
 - This part is pre-computed and mixed with source

Sequential creation of the room impulse response

- Start with the pre-computed tail of IR (reflections 4th order and up)
- Quickly compute the reflections of order 0-3 for current geometry
- (Reflection of order 0 is just the direct arrival)
- Stick them onto this generic tail
- parts that are perceptually important are updated in real time



Studying the HRTF

- Measure it for a manikin (a dummy with mold ears)
- Measure and compute it for spheres, snowman model
- Measure/study the HRIR/HRTF in an anechoic/infinite environment
- Study the combination: HRIR+Room Impulse Response
- Model its range dependence
- What features in the HRTF lead to localization?
- Can the HRTF be learned/related to anthropometry?

The CIPIC HRTF database

- A carefully collected database of HRTFs plus anthropometry

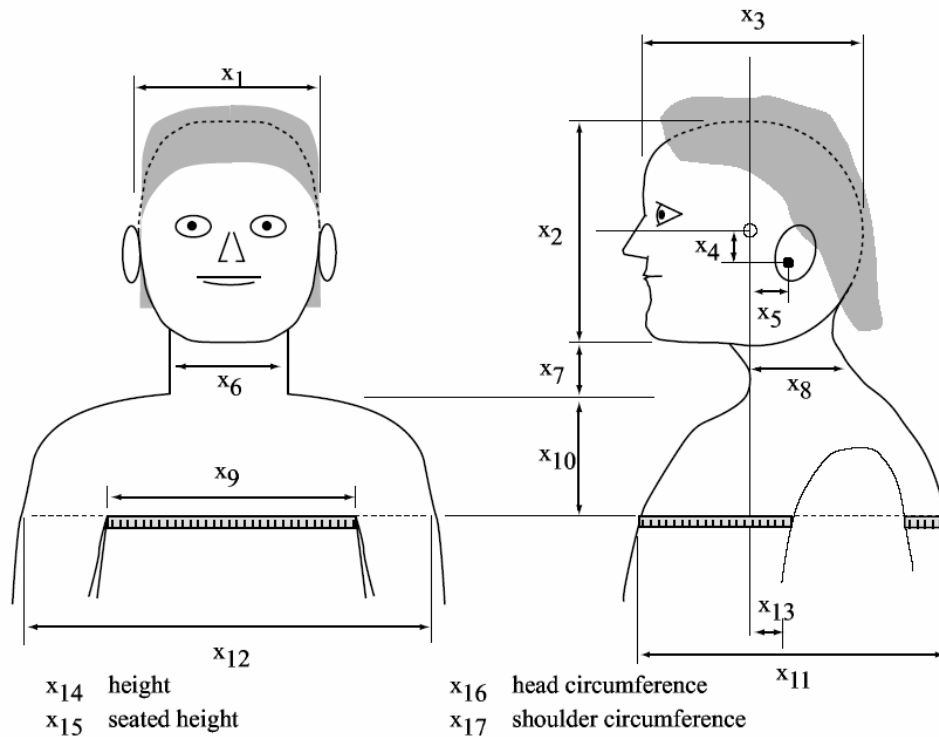
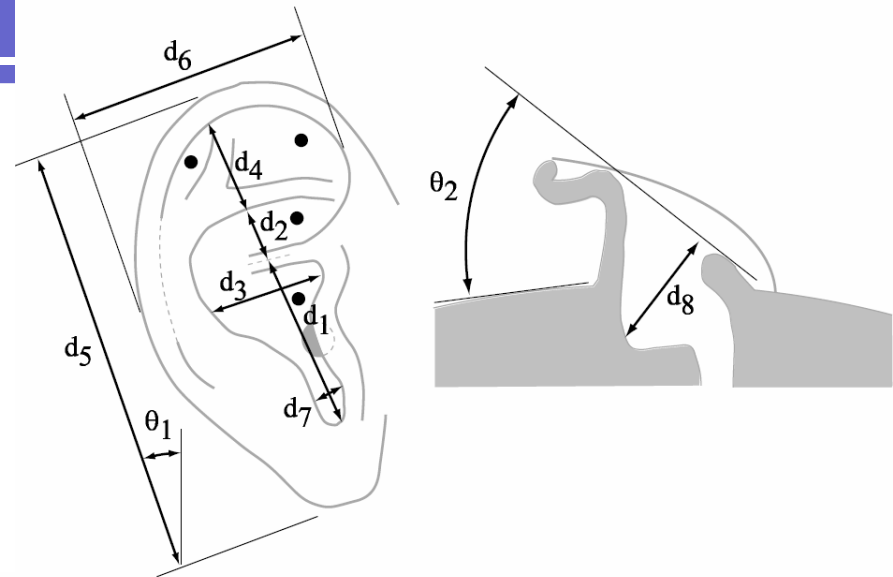


Figure 2: Head and torso measurements

Var	Measurement	μ	σ	%
x_1	head width	14.49	0.95	13
x_2	head height	21.46	1.24	12
x_3	head depth	19.96	1.29	13
x_4	pinna offset down	3.03	0.66	43
x_5	pinna offset back	0.46	0.59	254
x_6	neck width	11.68	1.11	19
x_7	neck height	6.26	1.69	54
x_8	neck depth	10.52	1.22	23
x_9	torso top width	31.50	3.19	20
x_{10}	torso top height	13.42	1.85	28
x_{11}	torso top depth	23.84	2.95	25
x_{12}	shoulder width	45.90	3.78	16
x_{13}	head offset forward	3.03	2.29	151
x_{14}	height	172.43	11.61	13
x_{15}	seated height	88.83	5.53	12
x_{16}	head circumference	57.33	2.47	9
x_{17}	shoulder circumference	109.43	10.30	19

Variability in body part measurements

Variability in pinna measurements



d_1	cavum concha height	1.91	0.18	19
d_2	cymba concha height	0.68	0.12	35
d_3	cavum concha width	1.58	0.28	35
d_4	fossa height	1.51	0.33	44
d_5	pinna height	6.41	0.51	16
d_6	pinna width	2.92	0.27	18
d_7	intertragal incisure width	0.53	0.14	51
d_8	cavum concha depth	1.02	0.16	32
θ_1	pinna rotation angle	24.01	6.59	55
θ_2	pinna flare angle	28.53	6.70	47

Correlation between dimensions

- Correlations are weak
- Indicate that one cannot predict another feature from measurements of any one
- Presumably the same applies to the HRTF
- The low frequency parts of the HRTF are reasonably well predicted

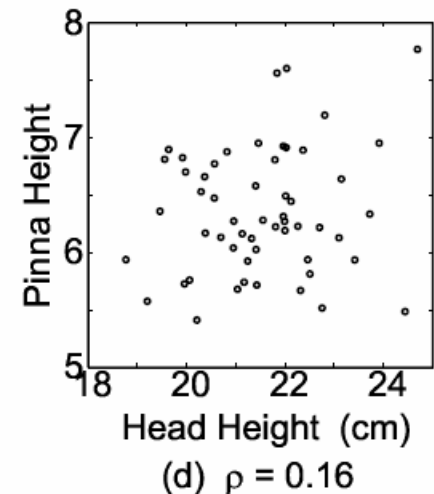
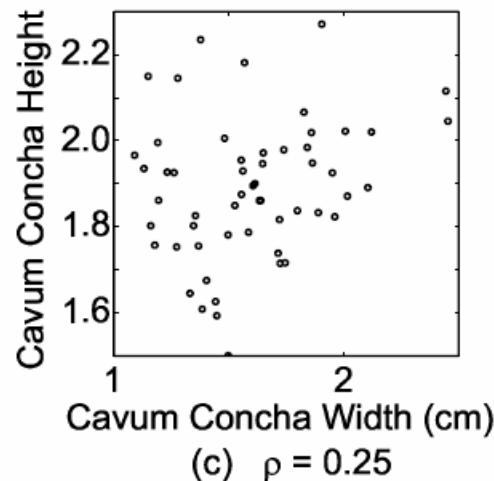
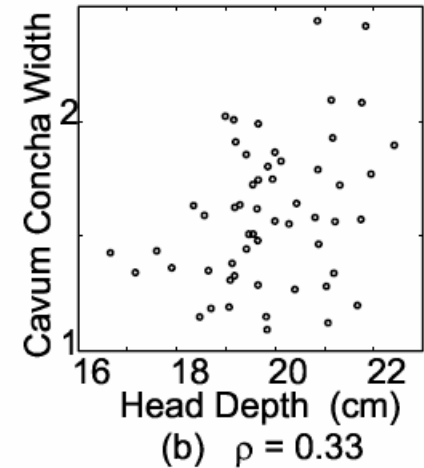
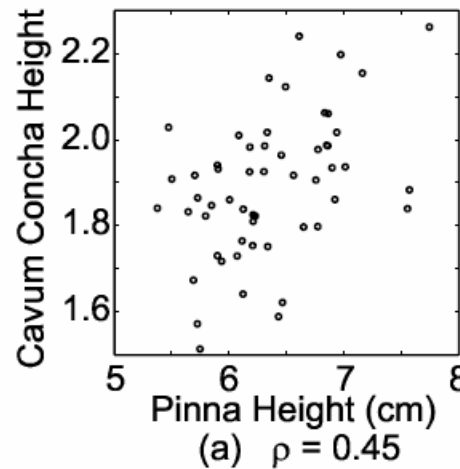



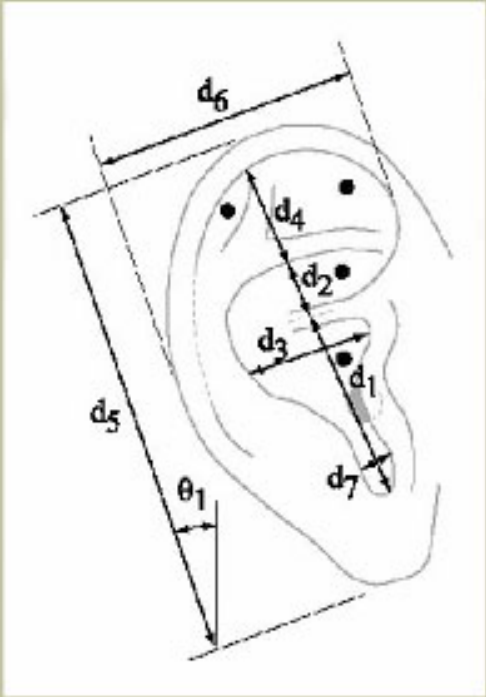
Figure 4: *Selected scatterplots*

HRTF Personalization using the Database

- HRTF is created by sound scattering
 - We can match features of the scatterer and hope for better-fitting HRTF
- CIPIC database, UC/Davis, 45 persons
 - Ear measurements are included
- Take ear picture, find ear parameters and locate the best-matching subject in the database
- Localization performance is improved by 20-30%
- Subjective experience is also better
- Number of front-back confusions is reduced

Database Personalization

Untitled



Subject name
dz
Head1
Head2
Right ear
Left ear
Done
Calibrate

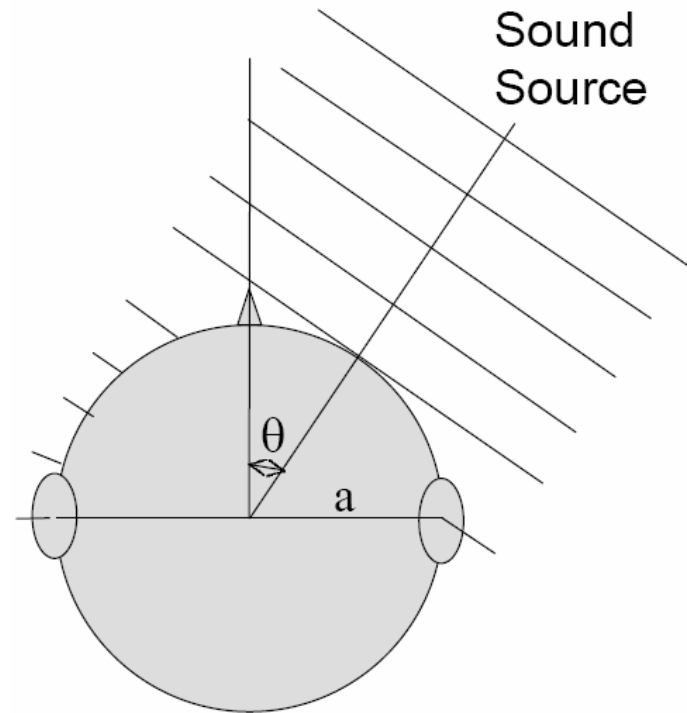
Use the reference image to make measurements. Remember to calibrate using the scale in the background before you start any measurements. Left click the mouse on the initial point and the final point. Do the measurements in order as shown in the reference figure. Note that the cavum concha depth (d_8) and the pinna rotation angle and the pinna flare angle are not measured. The closest id is chosen based on 7 measurements. After you have finished the measurements press Done to save the results in a file.

Analytical methods

- Sphere model
- Head and Torso Models

The “HRTF” of a sphere

- A sphere scatters sound
- Solution of the Helmholtz equation to a plane-wave from infinity (already used in designing the spherical array)
- Can also write the solution for a source at a particular point
- Careful study of these solutions for a sphere

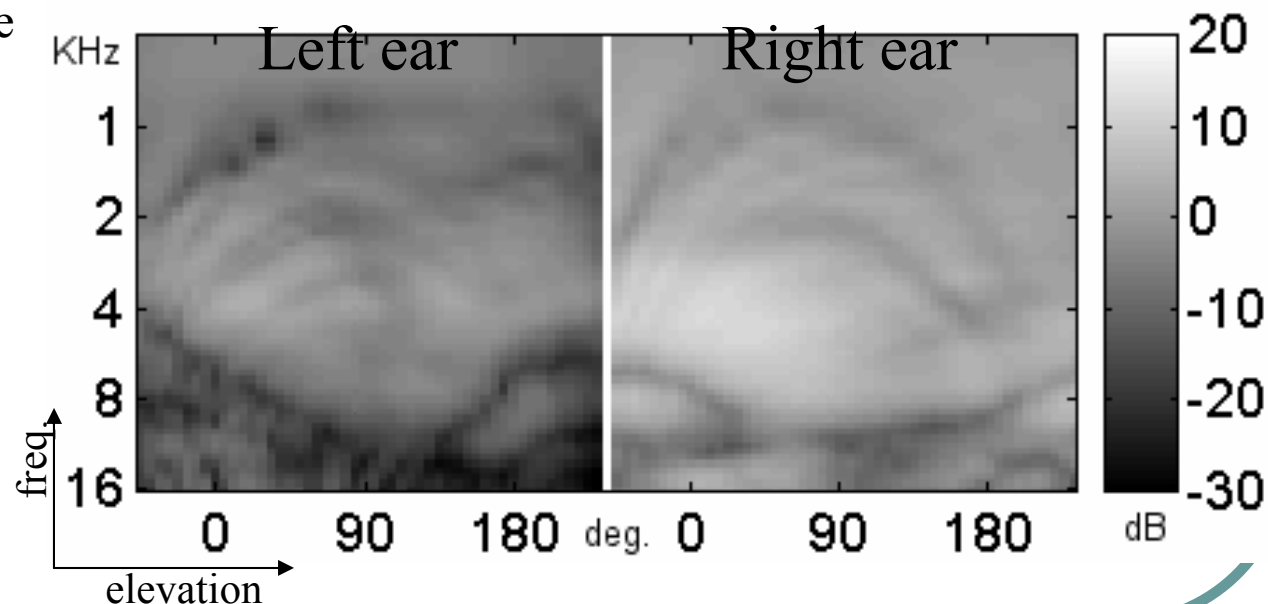


Sound Shadowing and Bright Spot



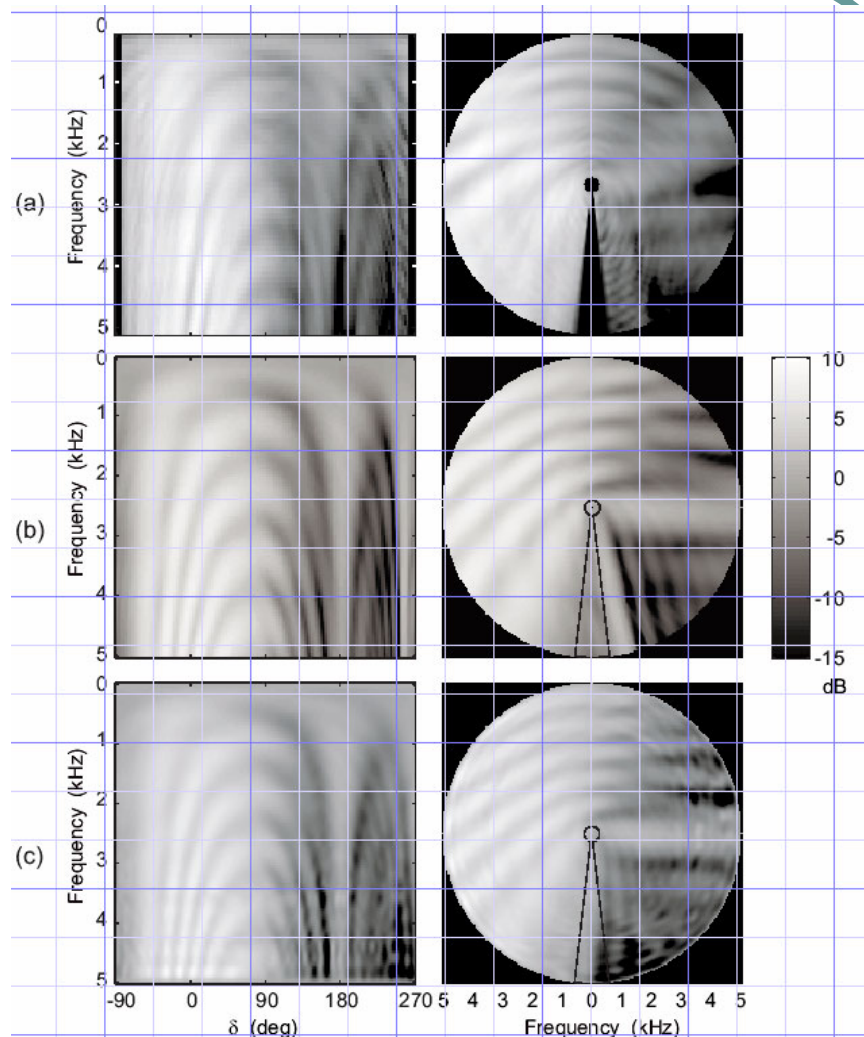
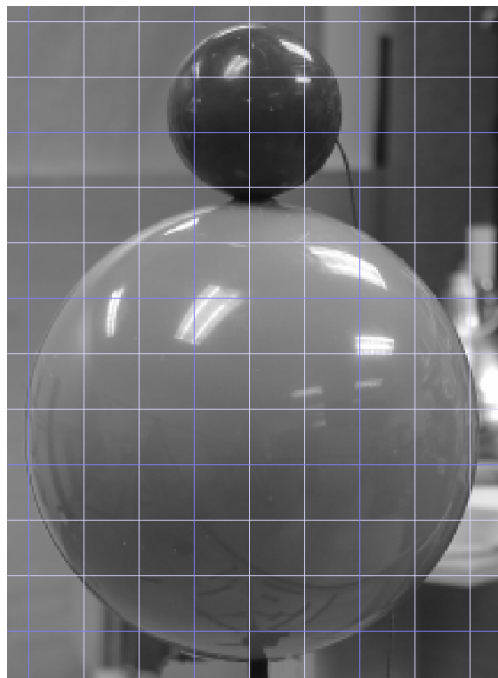
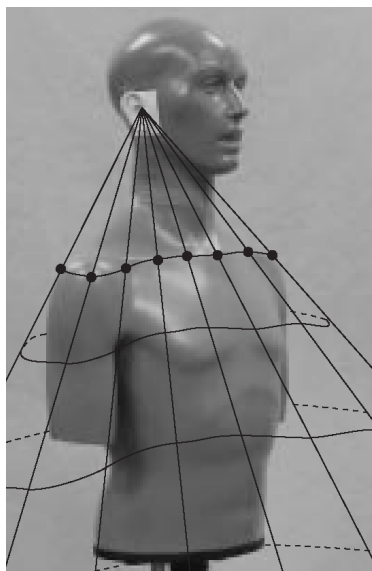
CIPIC HRTF Database

- Algazi et al (WASPAA 2001)
- 45 HRTFs + ear dimensions - available on web
- Demonstrate significant person-to-person variability
 - of both HRTFs and ear shapes
- Allow study of HRTFs
- Contralateral and ipsilateral ear HRTF slices for the 45 degree azimuth and varying elevation for a human
 - Torso reflections are the wide arches in low frequencies
 - Pinna notches are the dark streaks in high frequency range
- However, there are no models yet that let us go directly from geometry to response



First crack at personalization: Spherical HRTFs?

- We compared measured HRTFs for a mannequin, a bocce+bowling ball, and computed.
- Validate strategy



Algazi et al, "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am., 112, 2053-2064, 2002.

Readings for these lectures

- Papers: Spherical Model: Duda and Martens, 1998,
- Head and Torso Models: Algazi et al. 2002,
- Recreation of Spatial Audio: Zotkin et al. 2004,
- The CIPIC HRTF Database: Algazi et al. 2001
- See class web page for links