An Improvement to Anthropometry-Based Head and Torso HRTF Models for Locations Near the Frontal Median Plane

Richard S. Juszkiewicz

April 6, 2007

OBJECTIVES

- Create personalized HRTFs from anthropometry

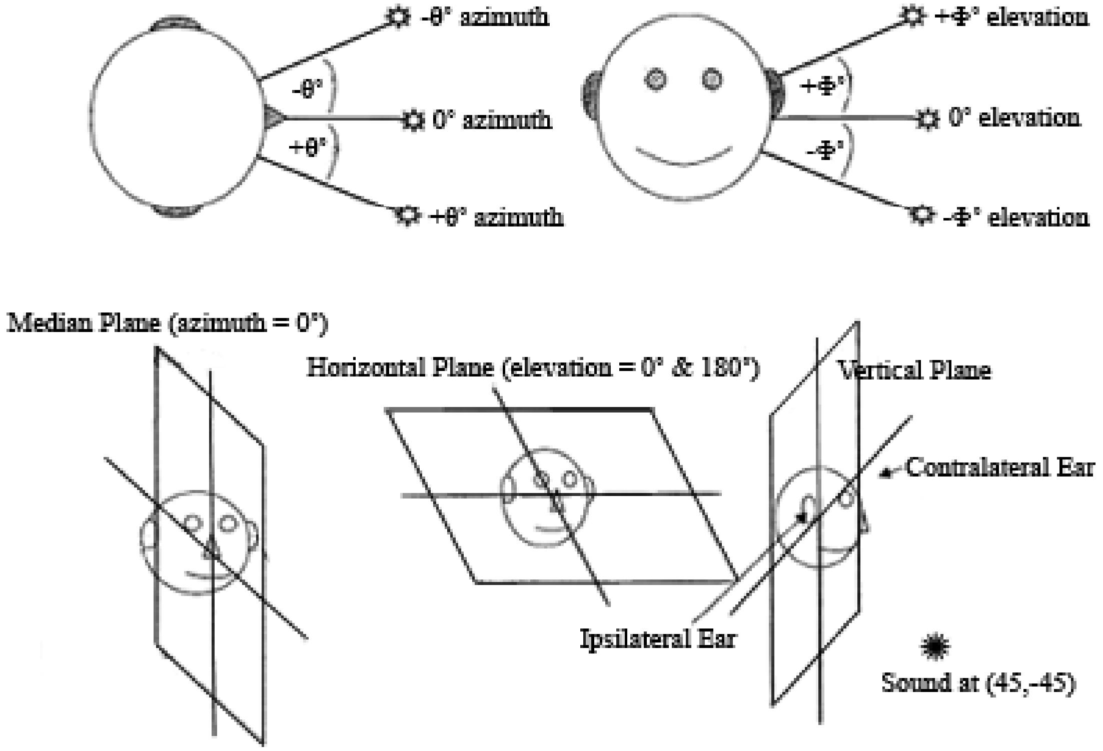
- Improve upon existing anthropometry-based synthesis methods in the process

- Keep algorithm computationally efficient

Content Overview

- Spatial Hearing
- HRTFs
- Head and Torso (HAT) model
- Pinna Model
- Objective Results
- Listening Test
- Subjective Results
- Future Work

COORDINATE SYSTEM & TERMINOLOGY



THE PHYSIOLOGY OF AUDITORY LOCALIZATION

- Process begins in the Lateral Superior Olive of the mid-brain

- A cross-correlation operation is performed on the signals that arrive at each ear.

- Temporal, loudness and spectral cues are among those interpreted by the cross-correlation process.

Interaural Time Difference (ITD)

- The relative difference in arrival times of an incident sound at each ear

- A sound will arrive earlier at the ear in which it is closest

- Not dependent upon range

- Dependent upon frequency

Interaural Intensity Difference (IID)

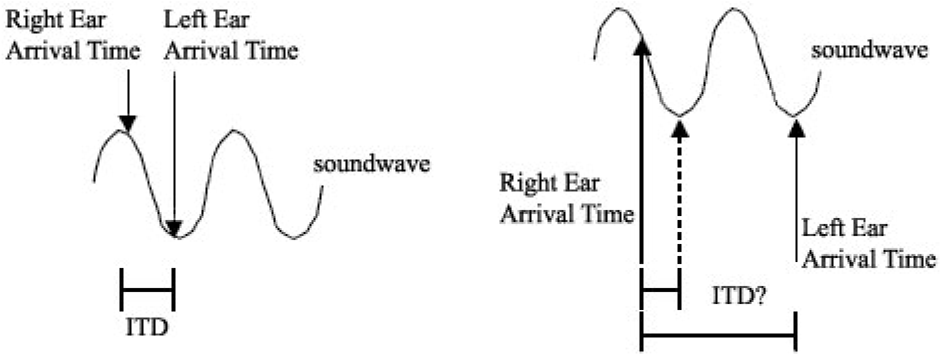
- The amplitude difference in the signals that arrive at each ear

- The human head acts as a filter that affects the magnitude response of the signals that arrive at each ear

- Frequency dependent

Duplex Theory

- ITD is only accurate for frequencies below approximately 1.5kHz where wavelengths are shorter than the diameter of the head



- IID is only accurate above 3kHz because head does not shadow low frequencies

- In the late 1800s Lord Rayleigh hypothesized that the IID and the ITD are used together by the brain for horizontal localization.

Spectral Cues

- In a plane of constant azimuth the ITDs and IIDs are all nearly equal for different elevations

- Spectral cues are used by the brain to determine elevation

- Spectral cues are caused by the pinnae and the torso

LOCALIZATION ACCURACY

- The just noticeable differences in each plane of hearing depend on multiple factors: closeness to median plane, familiarity of source and frequency content of the source.

  - Difference limen in horizontal direction: 1° near the frontal median plane and 3° near the interaural poles and twice as much in the rear hemisphere.

  - Difference limen in the vertical direction: most accurate in the median plane; 17° for continuous speech by an unfamiliar voice, 9° for continuous speech by a familiar person and 4° for white noise.

Head-Related Transfer Functions (HRTFs)

- A set of filters that capture the effects that the body has on an incident sound coming from a particular point in space

- Used in headphone based media players, virtual reality systems, military training, airplane cockpits, sonification of multi-variable data sets, etc.

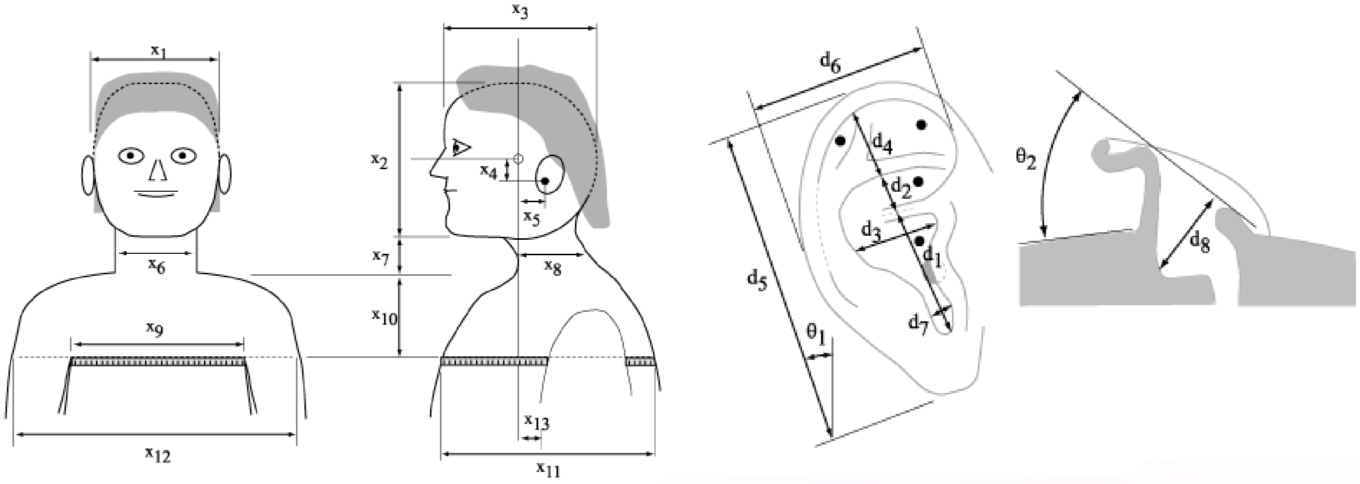
- They are very listener dependent particularly if accurate elevation localization is desired.

CIPIC HRTF DATABASE

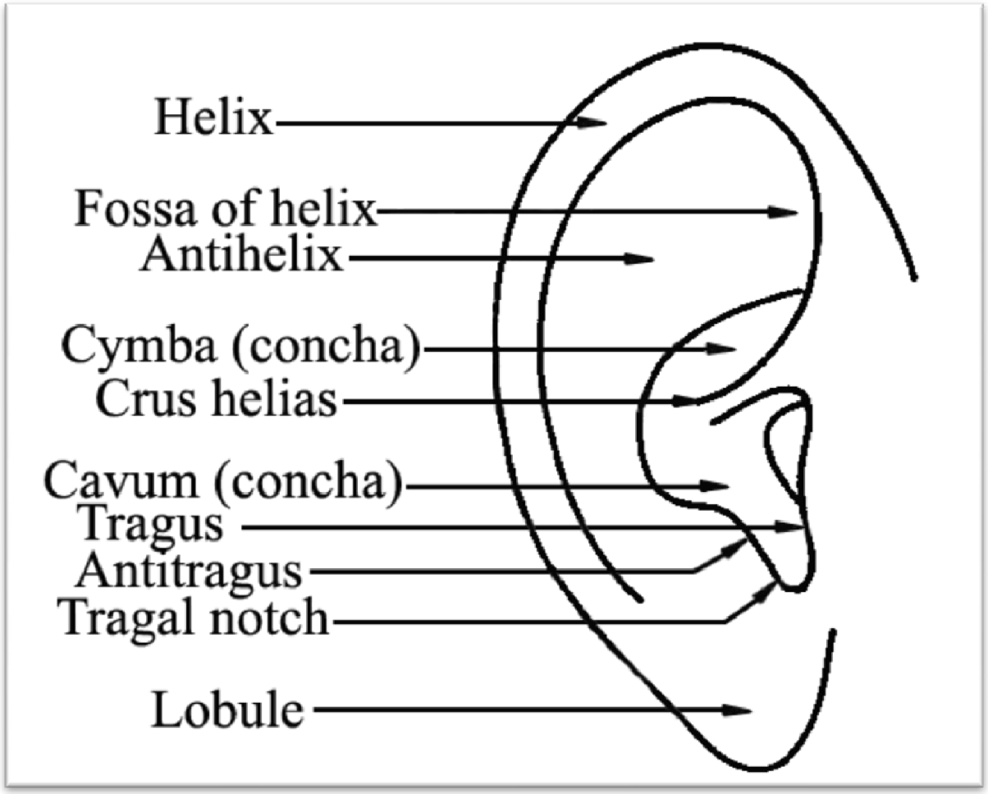
- Recorded at UC Davis

- Contains HRTFs for 45 subjects at 1250 spatial locations

- Also provides anthropometry for a number of subjects:

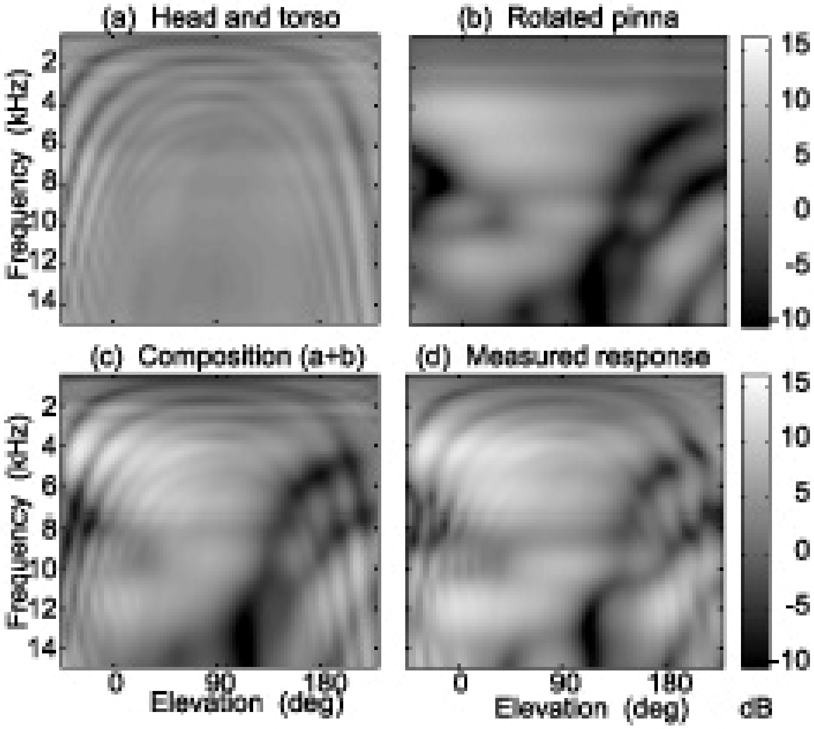


ANATOMY OF THE PINNA

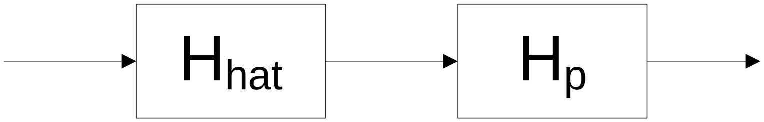


Return

Structural Decomposition



Azimuth constant at 20˚



Perceptually Significant Spectral Features

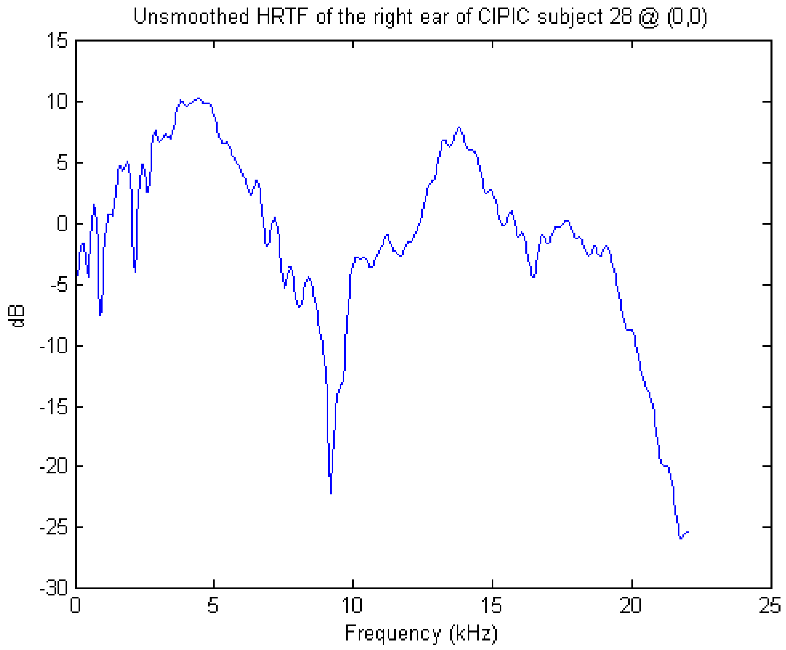
- Macroscopic peaks and notches are necessary at high frequencies (above 4kHz) to maintain accurate elevation localization

- Microscopic peaks are necessary at low frequencies to distinguish front from back and for subtle elevation localization effects

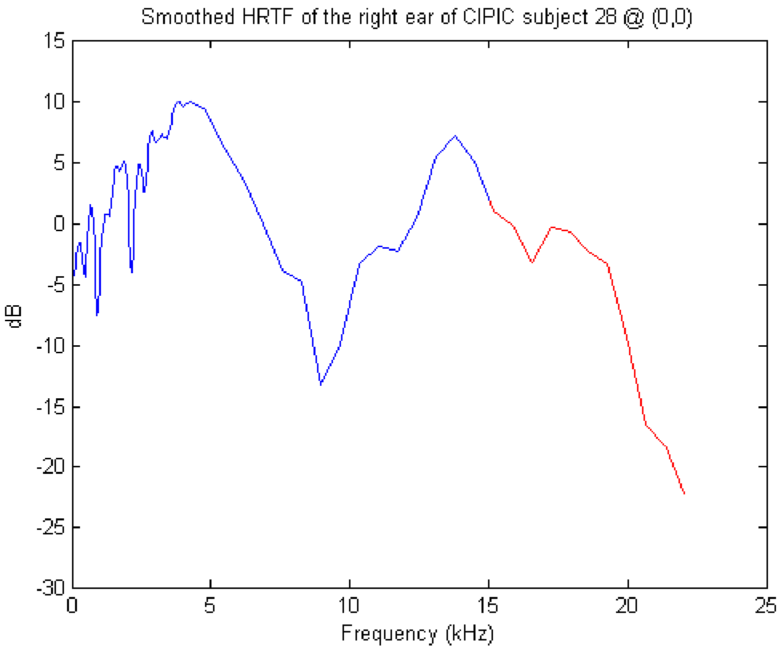
- The IID can be accurately approximated using a first order IIR filter without compromising horizontal localization

- Frequencies above 15kHz are perceptually irrelevant

SMOOTHING EXAMPLE



An unsmoothed HRTF



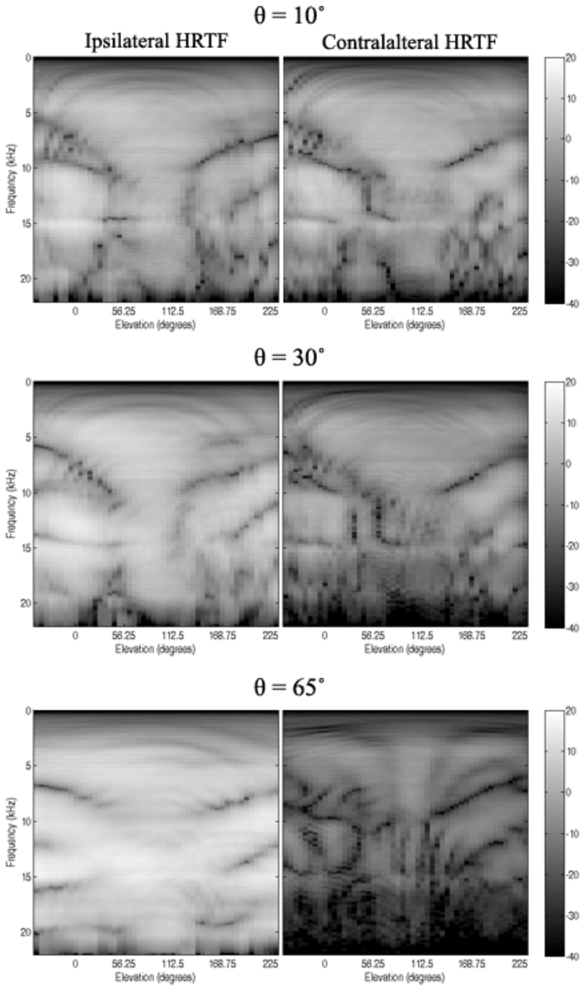
HRTF with minimal amount of necessary detail

Return

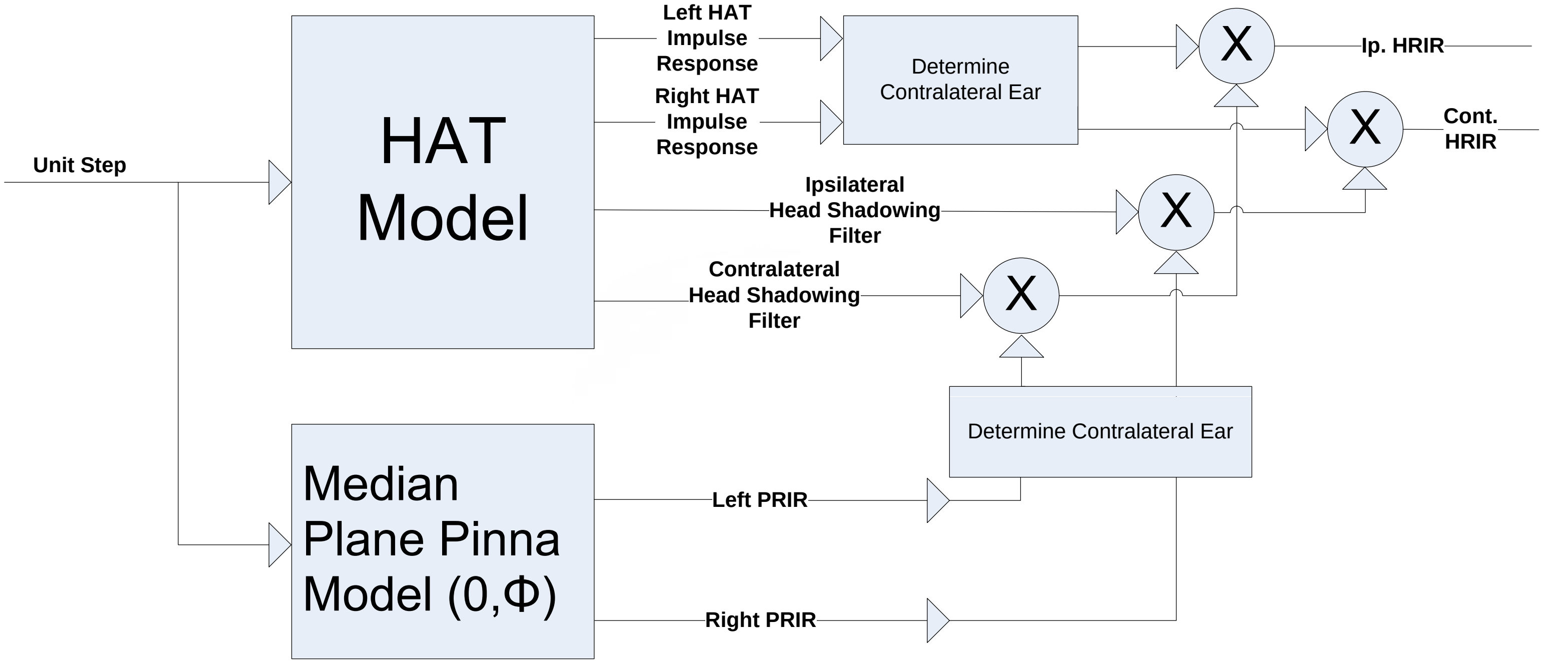
COMPLEXITY OF CONTRALATERAL HRTF

- An algorithm has been designed to approximate the contralateral HRTF from the known ipsilateral response and the response of a spherical head model

- A modified version of that algorithm is used in this implementation



PROPOSED SOLUTION


Unit Step
HAT Model
Median Plane Pinna Model (0,Φ)
Left HAT Impulse Response
Right HAT Impulse Response
Determine Contralateral Ear
X
Ip. HRIR
X
Cont. HRIR
Ipsilateral Head Shadowing Filter
X
Contralateral Head Shadowing Filter
X
Determine Contralateral Ear
Left PRIR
Right PRIR

Scope Limitations & Introduced Ideas

- Added non-symmetric ear offsets and a torso reflection coefficient that is dependent upon orientation and frequency to an existing HAT model.

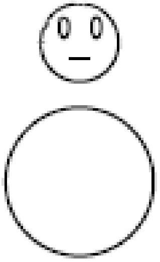
- Introduced pinna model possessing anthropometry-based elevation cues.

- Pinna model functionality limited to frontal hemisphere and azimuths within 30° of the median plane in either direction.

- The way in which all of the components of the model are interconnected is also novel.

Head and Torso (HAT) Model

- Approximates the anatomy of a human as a snowman with an invisible neck.



- Head is offset from the torso by a distance corresponding to the subject’s neck height

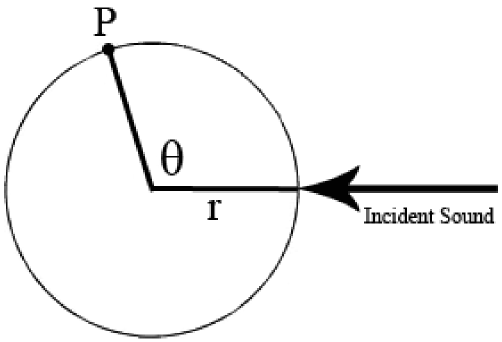
- Uses the acoustical response of a sphere and first order reflection principles to model the ITD, the IID, torso reflections and torso shadowing

Modeling the Acoustic Response of a Sphere

- A first order IIR filter designed by Brown and Duda accurately approximates the effects that the presence of a sphere has on an incident sound:

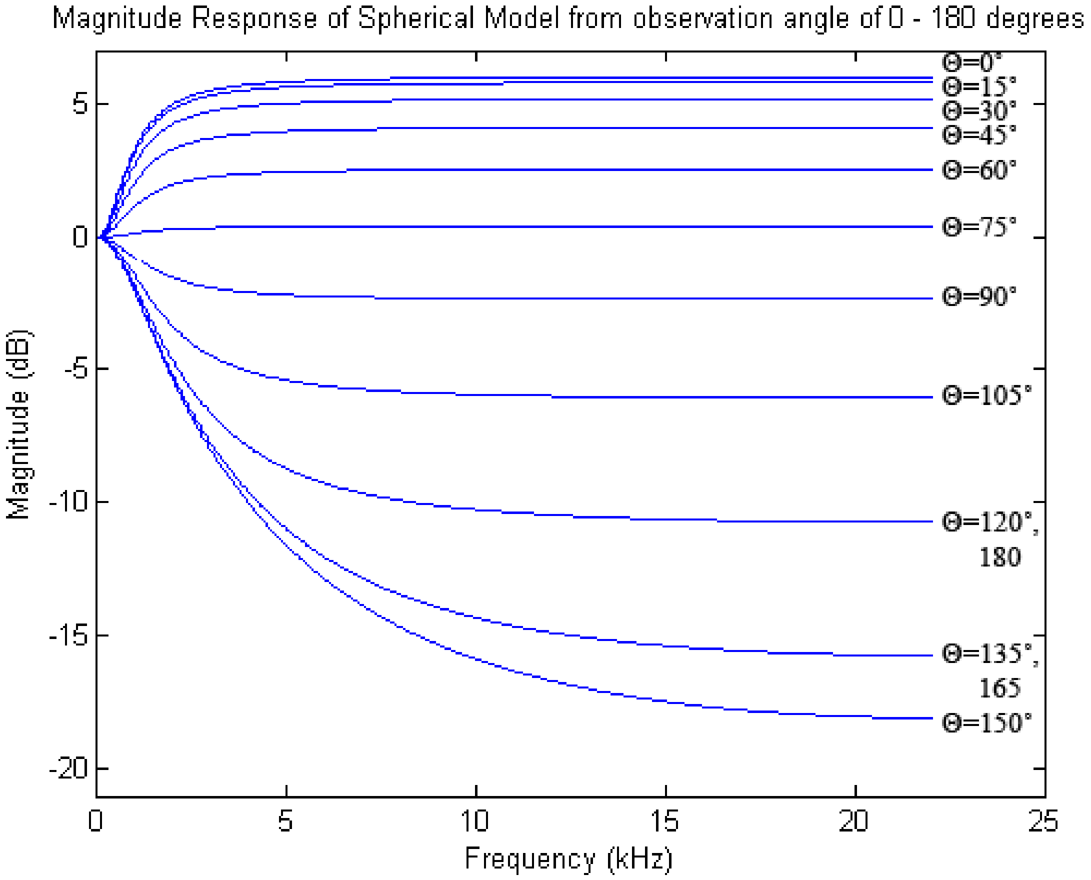
$$H(s,\theta,r)=\frac{\alpha\tau s+1}{\tau s+1}$$ where $$\alpha(\theta)=\left[1+\frac{\alpha_{\min}}{2}\right]+\left[1-\frac{\alpha_{\min}}{2}\right]\cos\left[\frac{\theta}{\theta_{\min}}\Pi\right]$$

and $$\tau=\frac{r}{2c}$$



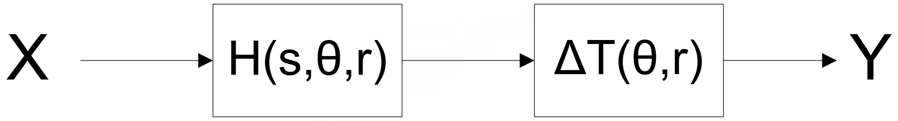
Constant values of $\theta_{\min}$ = 150˚ and $\alpha_{\min}$= .1 provide the most realistic sounding results

Frequency Response of Spherical Model



Modeling the Acoustic Time Delay of a Sphere

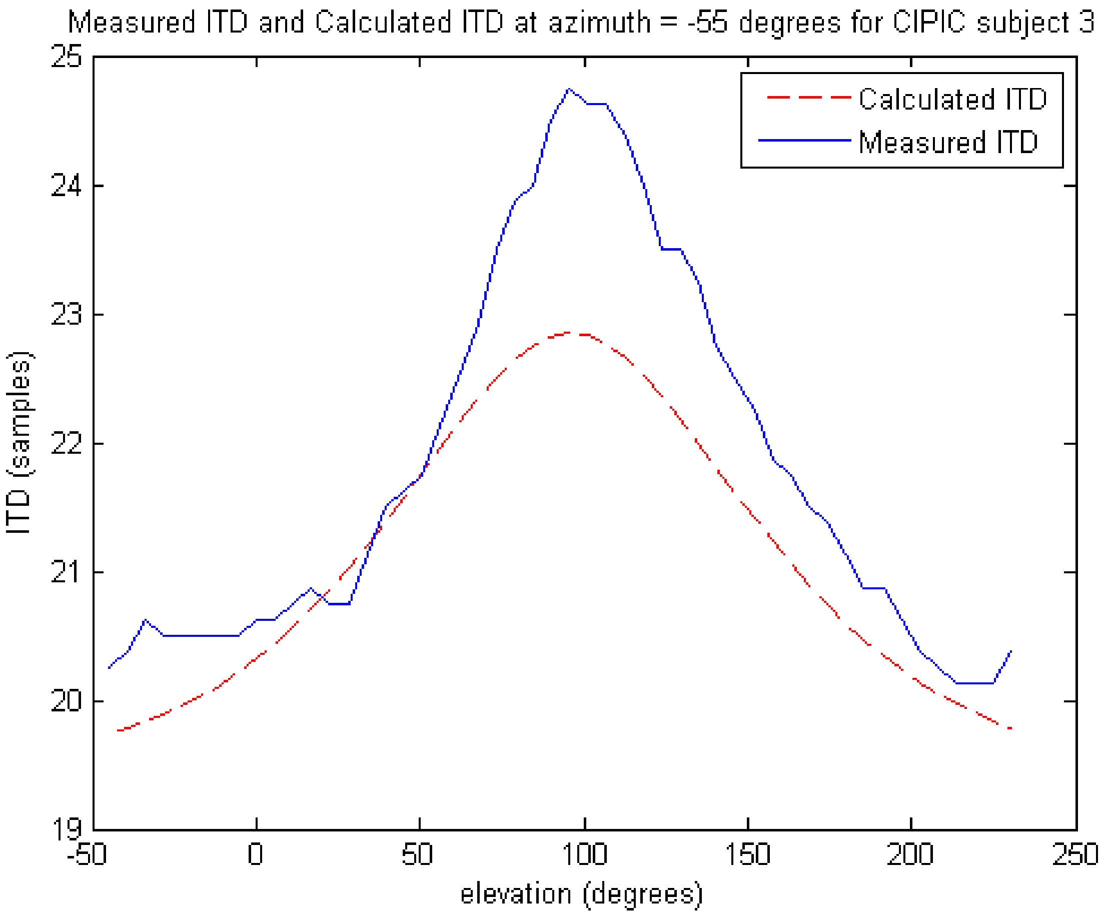
- The filter from the previous slide is cascaded in series with an FIR delay line to model the time delay that occurs due to the presence of the sphere



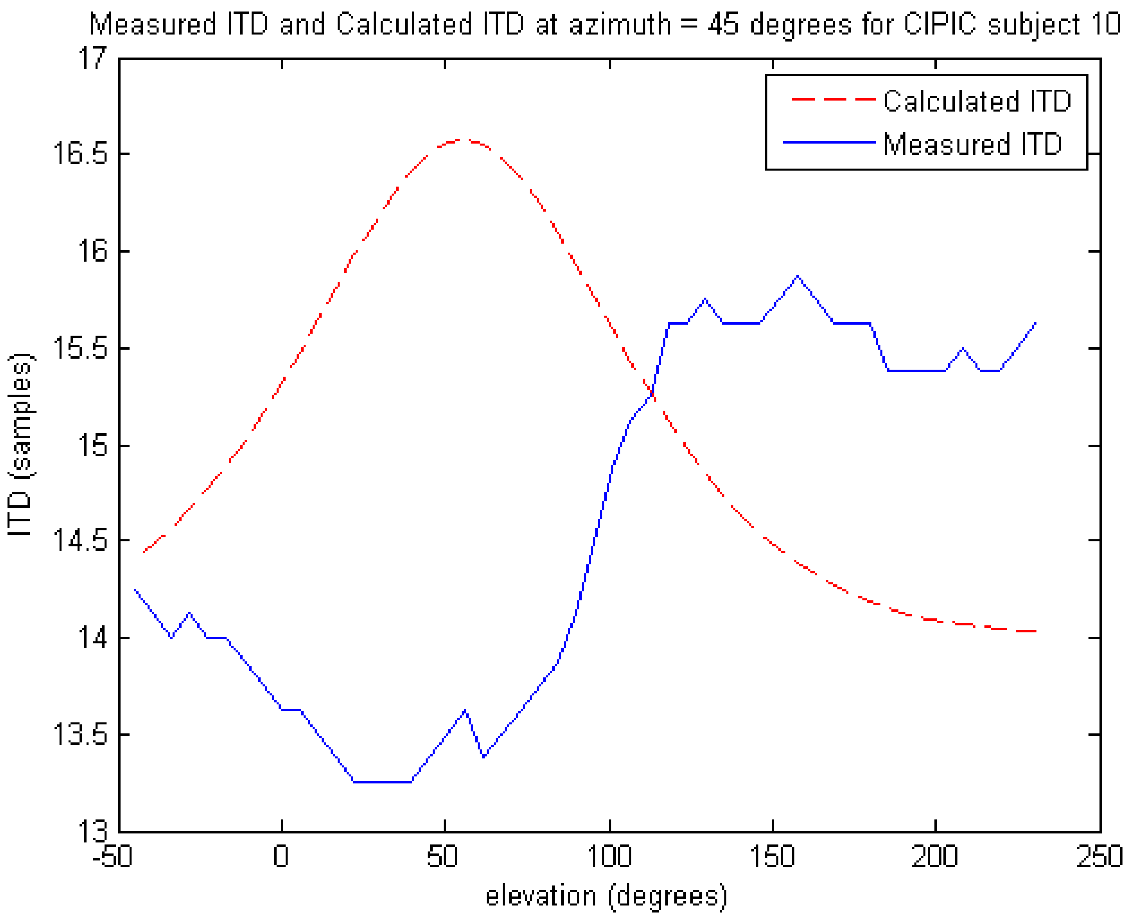
$$\Delta T = \begin{cases} -\dfrac{r}{c}\cos\theta & if\ 0 \le |\theta| < \dfrac{\Pi}{2} \\ \dfrac{r}{c}\left[|\theta| - \dfrac{\Pi}{2}\right] & if\ \dfrac{\Pi}{2} \le |\theta| < \Pi \end{cases}$$

This filter is used as the cornerstone to the HAT model. It models the ITD, the IID and torso shadowing effects.

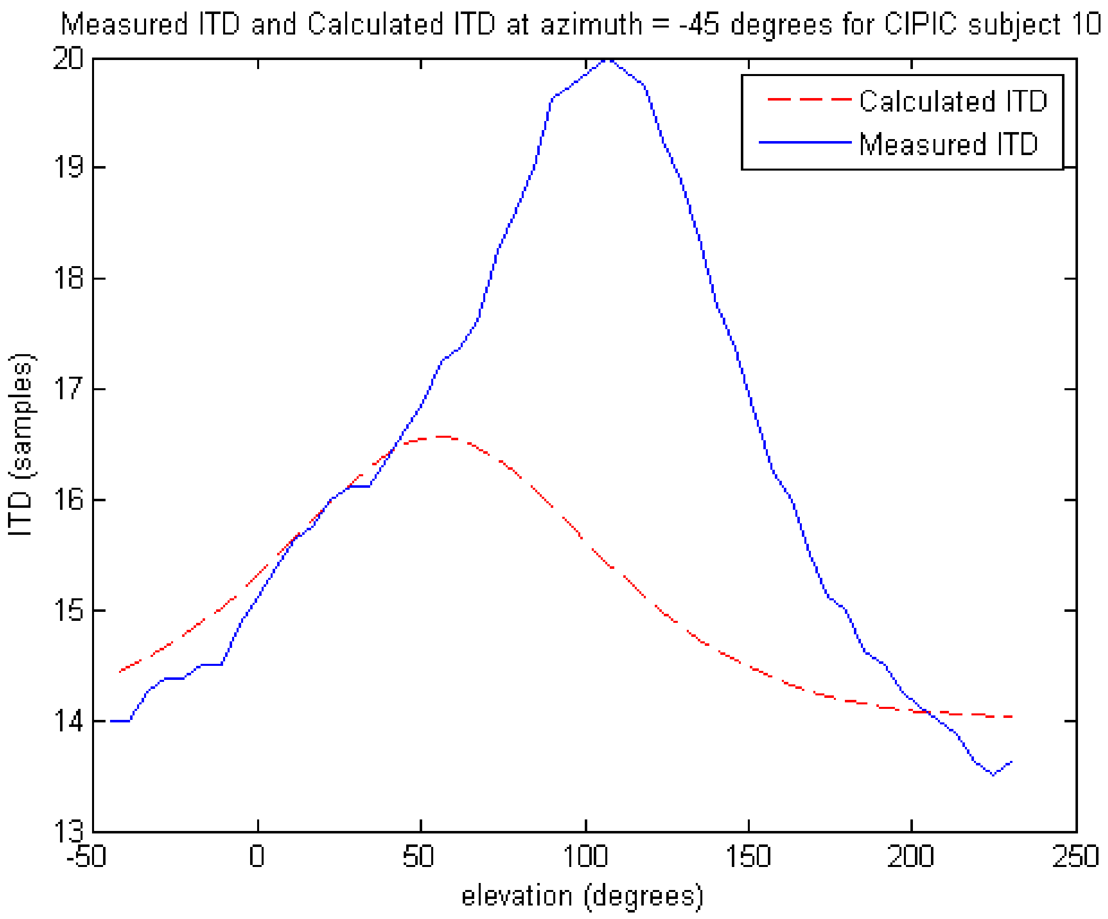
ITD Results


Measured ITD and Calculated ITD at azimuth = -55 degrees for CIPIC subject 3
ITD (samples)
25
24
23
22
21
20
19
-50
0
50
100
150
200
250
elevation (degrees)
Calculated ITD
Measured ITD

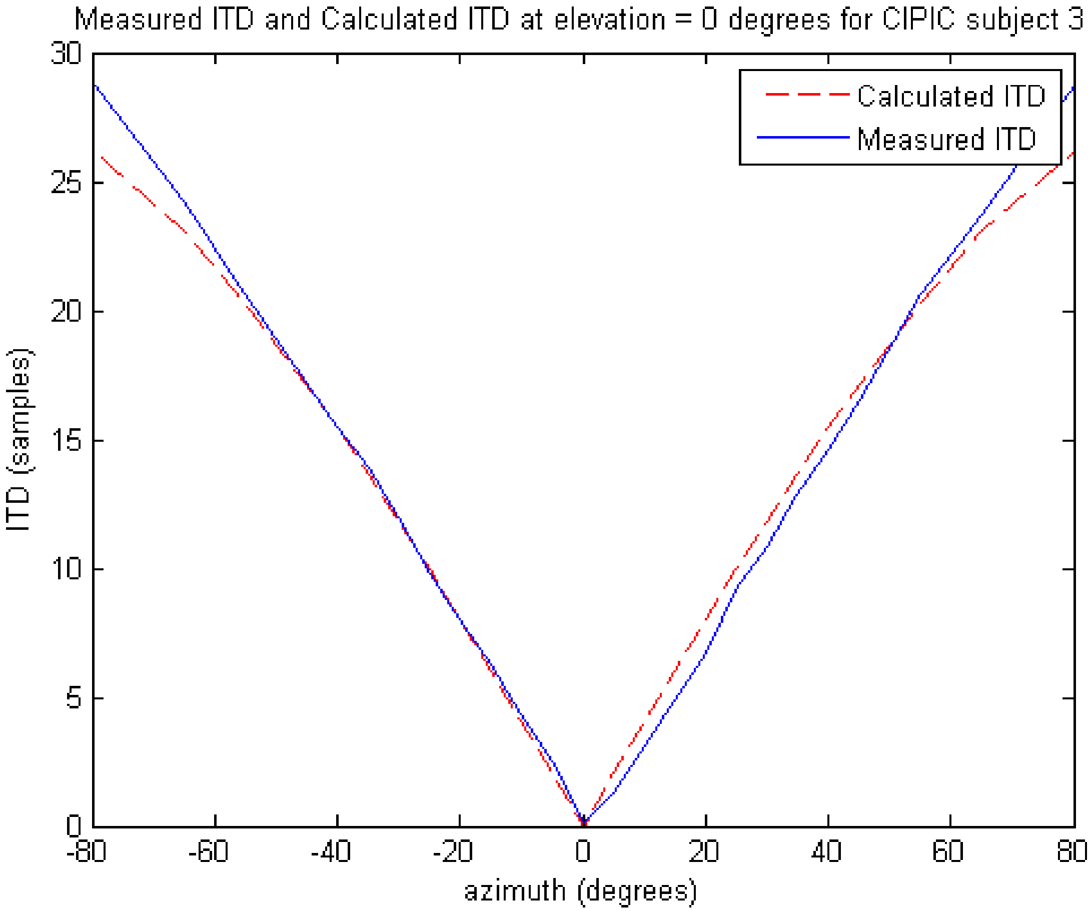
ITD Results


Measured ITD and Calculated ITD at azimuth = 45 degrees for CIPIC subject 10
ITD (samples)
17
16.5
16
15.5
15
14.5
14
13.5
13
-50
0
50
100
150
200
250
elevation (degrees)
Calculated ITD
Measured ITD

ITD Results


Measured ITD and Calculated ITD at azimuth = -45 degrees for CIPIC subject 10
Calculated ITD
Measured ITD
ITD (samples)
20
19
18
17
16
15
14
13
-50
0
50
100
150
200
250
elevation (degrees)

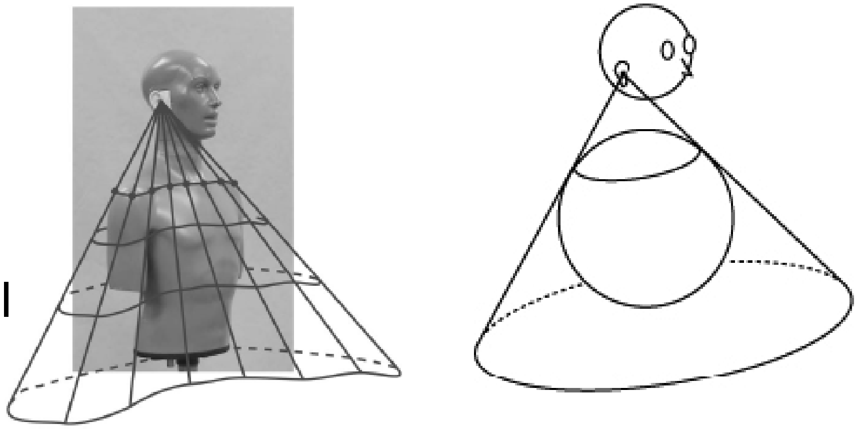
ITD RESULTS


Measured ITD and Calculated ITD at elevation = 0 degrees for CIPIC subject 3
Calculated ITD
Measured ITD
ITD (samples)
azimuth (degrees)
0
5
10
15
20
25
30
-80
-60
-40
-20
0
20
40
60
80

ROLE OF TORSO

- The torso that is also represented in the model as a sphere has two primary roles: it either shadows an incoming sound or reflects it.

An imaginary cone is created from all of the rays that extend from the ear and are tangent to the torso. If an incident originates inside of the torso shadow cone it is shadowed; all other sounds are reflected.

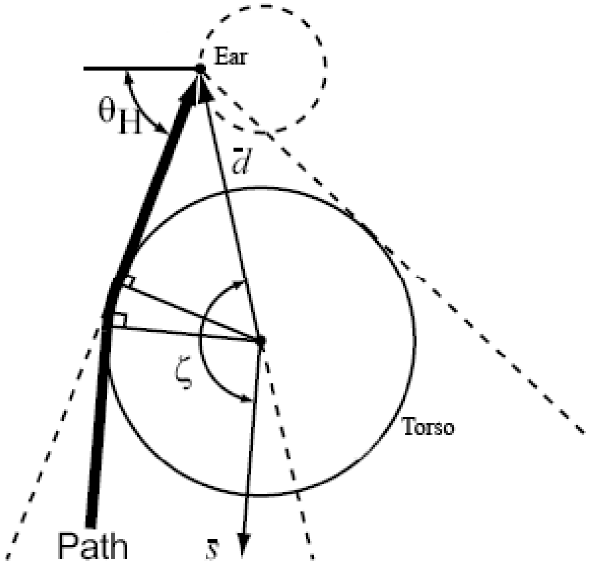


Torso shadow cones

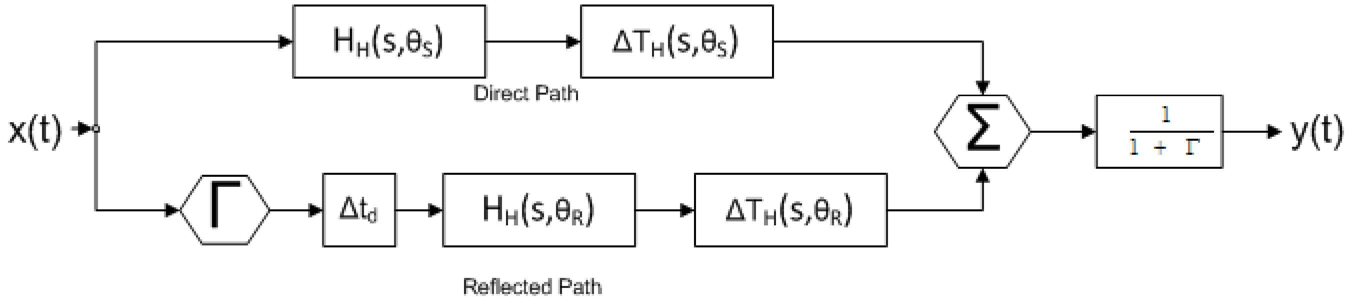
Torso Shadowing

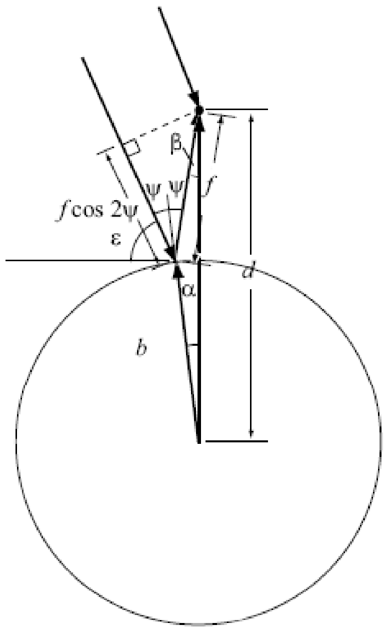
- When a sound originates in the torso shadow cone it is shadowed by the torso in the same way that the head shadows a sound.

- It arrives at the ear at an angle of incidence different from what it would be if it was unimpeded, since the path has been altered.

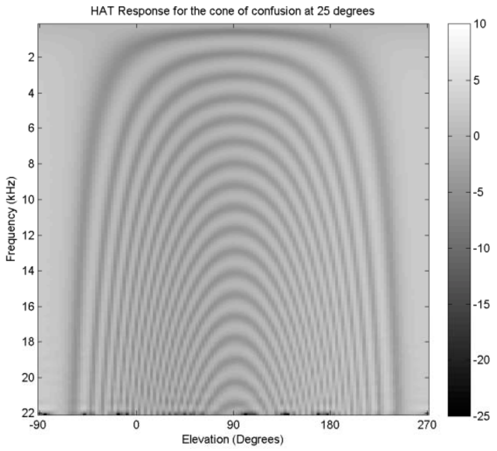


Torso Reflection

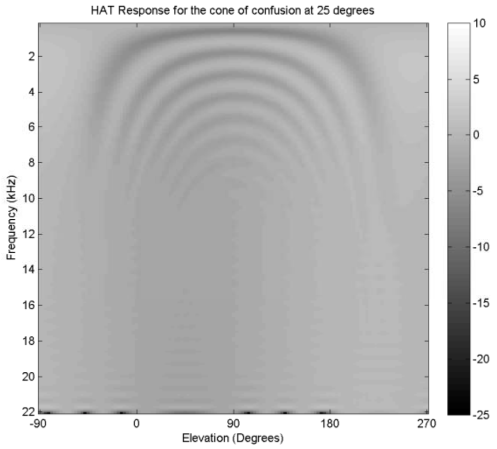




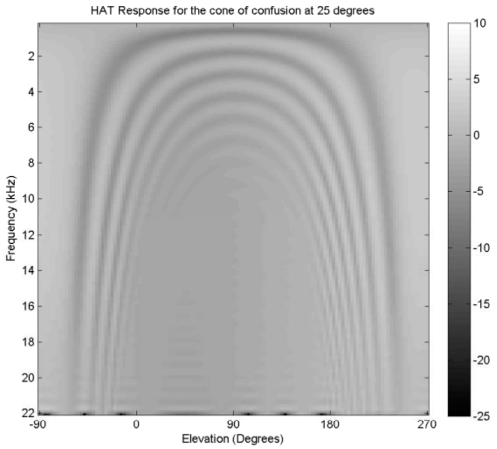
TORSO REFLECTION COEFFICIENT



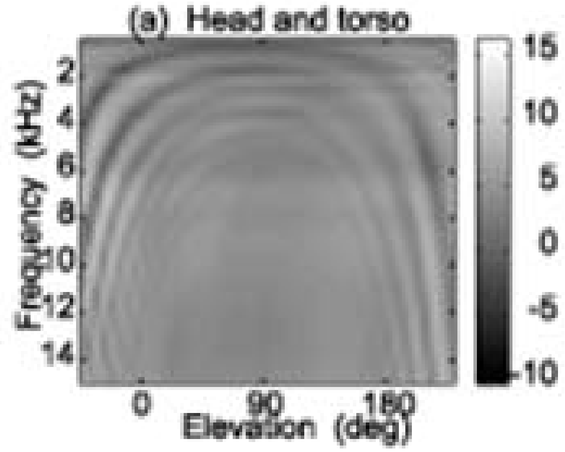
Constant reflection coefficient



Frequency dep. ref. coefficient



Freq. and orientation dep. ref. coefficient



Measured Response

Pinna Model

- Possible to model the most perceptually relevant characteristics of the pinna by cascading a few digital filters.

- Bandpass filters are used to account for the resonances of the pinna

- An FIR comb filter is used to model the notches that occurs due to the reflections off of the concha or helix.

- Pinna only affects frequencies above 3.5kHz

- Re-examining the smoothing slide reveals that only a few filters are needed to accurately model the pinna.

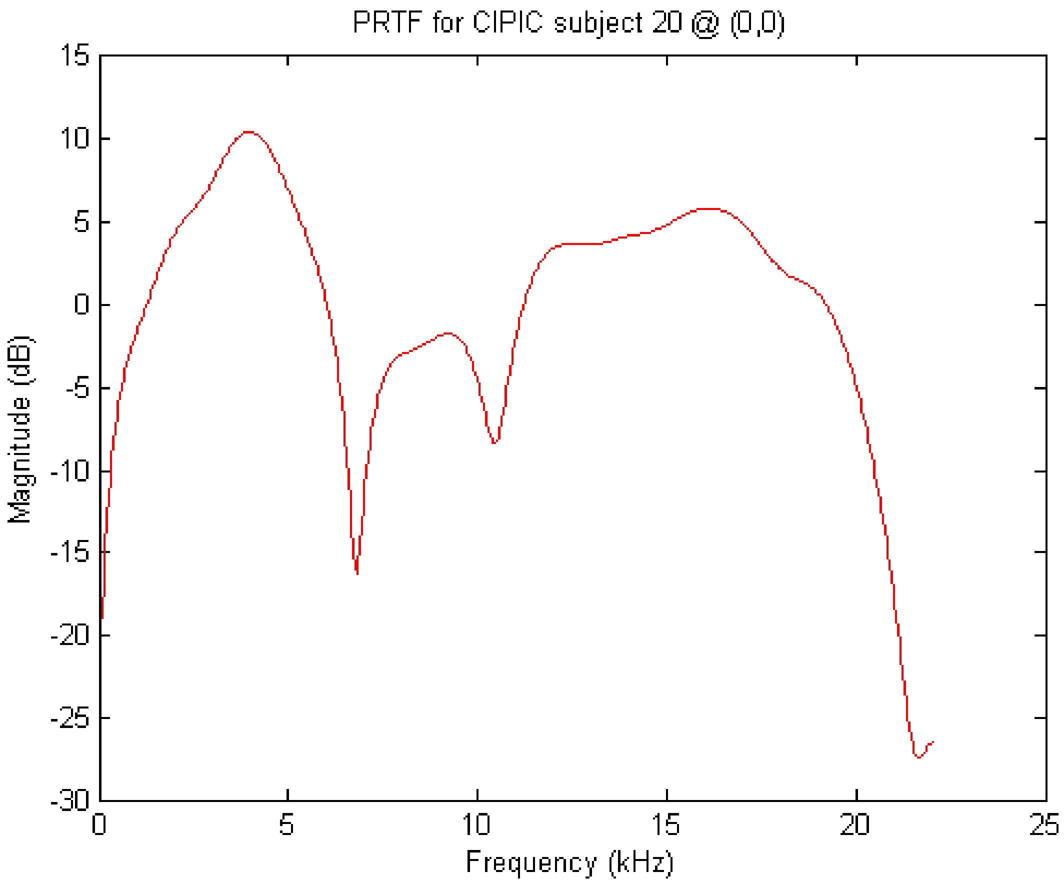
Pinna-Related Transfer Function (PRTF)

- Obtained by windowing HRIR (Head-Related Impulse Response)
  - .9ms half Hanning window

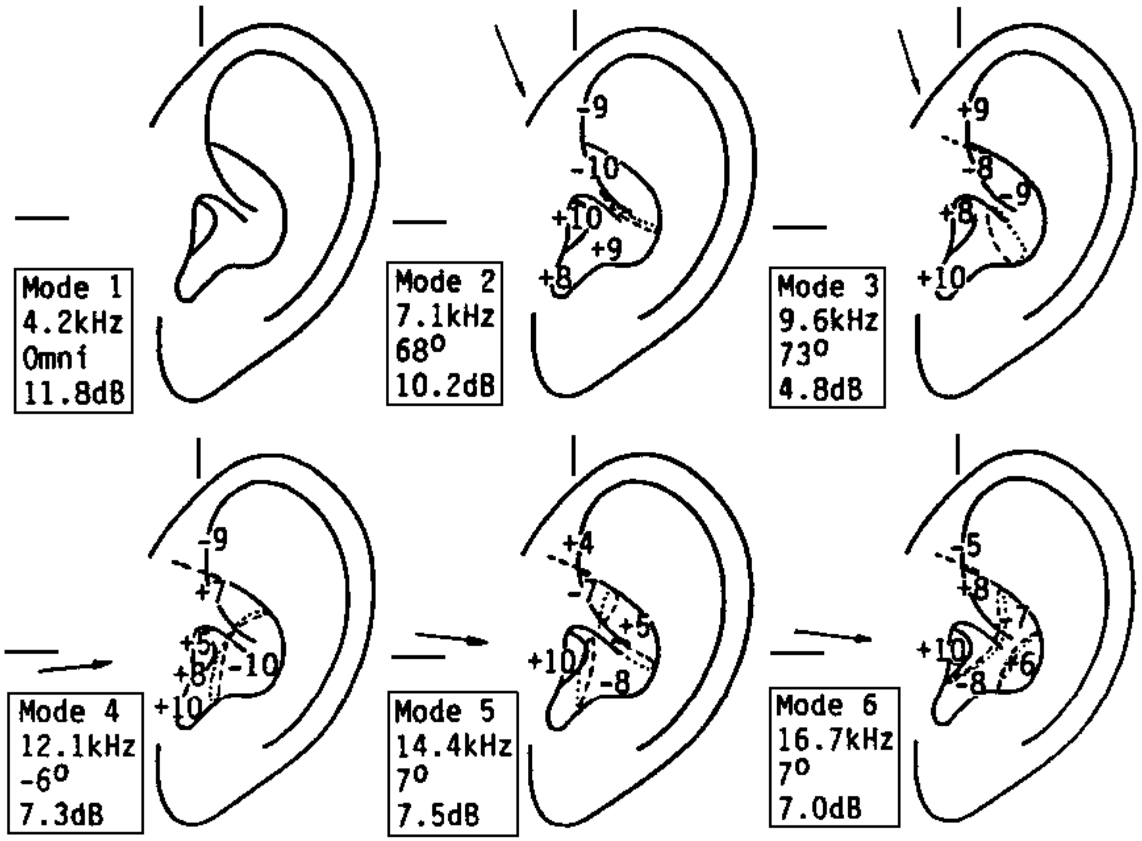
- Produces a smoothed version of the pinna response

- Used for analysis and testing purposes

PRTF Example at (0,0)



Resonances of the Pinna



Return

Modeling Resonances at (0,0)

$$H(z)=\frac{\left(\frac{G_0+G\beta}{1+\beta}\right)-\left(\frac{2\left(G_0\cos(\omega_0)\right)}{1+\beta}\right)z^{-1}+\left(\frac{G_0-G\beta}{1+\beta}\right)z^{-2}}{1-2\left(\frac{\cos(\omega_0)}{1+\beta}\right)z^{-1}+\left(\frac{1-\beta}{1+\beta}\right)z^{-2}}$$

$$\beta=\left(\sqrt{\frac{G_B{}^2-G_0{}^2}{G^2-G_B{}^2}}\right)\tan\left(\frac{\Delta\omega}{2}\right)$$

ω_0 = center frequency in rad/sec
G = gain
G_0 = DC offset.
$\Delta\omega$ = bandwidth of the filter in rad/sec (Q factor)
G_B = 3dB less than the gain of the filter (G)

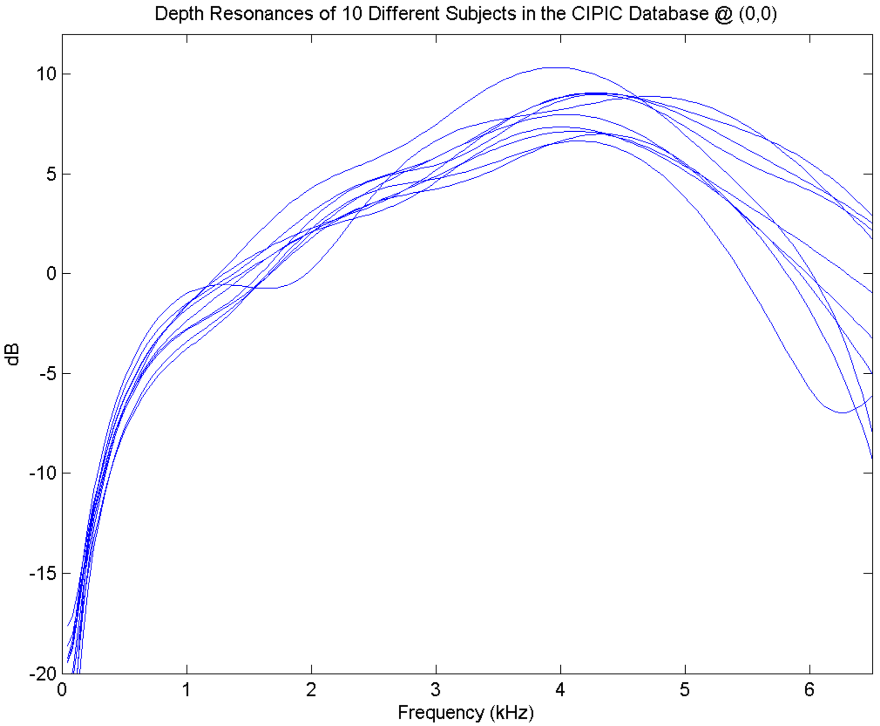
Omni-Directional Depth Resonance

- Center frequency of resonance is calculated by modeling the shape of the concha as a cylinder and calculating its resonance frequency using the depth and width of the concha as parameters

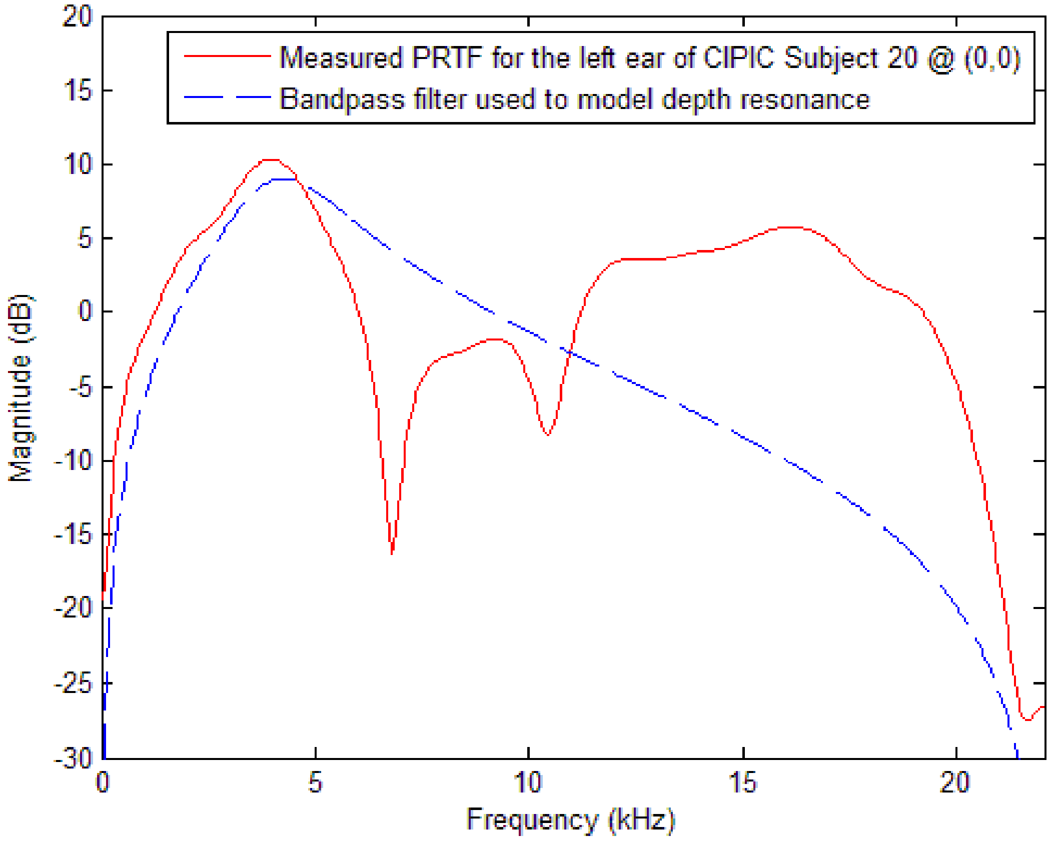
$$f_{depth} = \frac{c}{4(d + .411w)}$$

- Zeros at Nyquist and DC values

Gain = 9dB
Bandwidth = 2.5kHz



Depth Resonance Plot

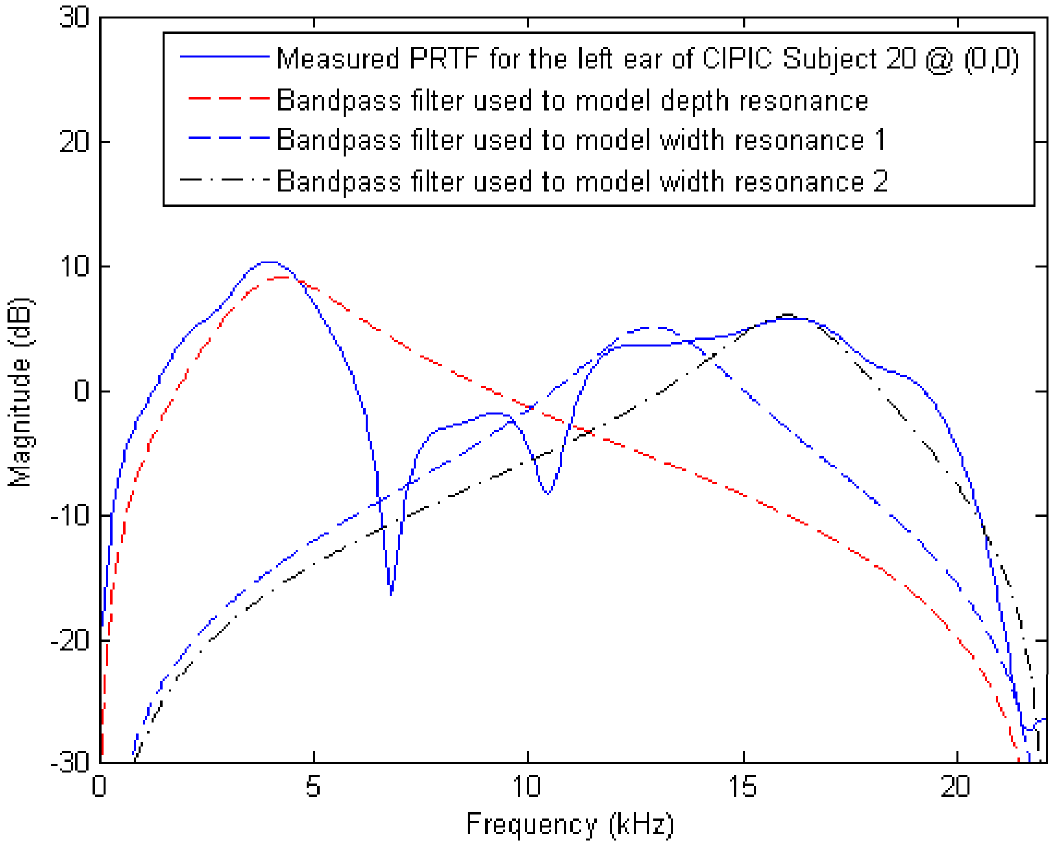


Width Resonances at (0,0)

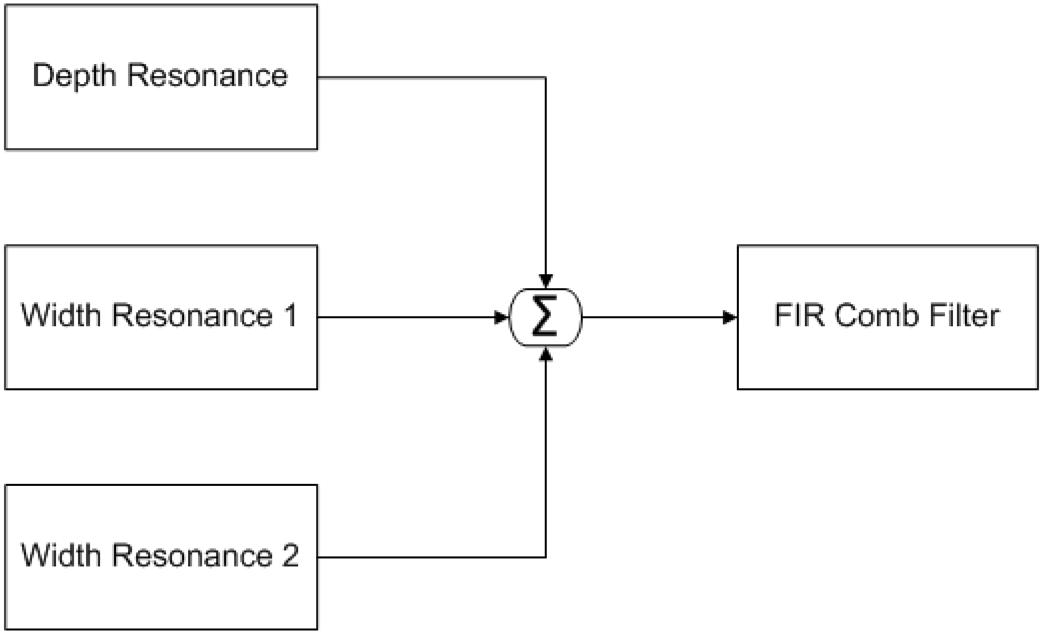
Width Resonance 1 (Mode 4)	Width Resonance 2 (Mode 6)
Fc = (see below)	Fc = 16kHz
G = 5dB	G = 6dB
BW = 3kHz	BW = 3kHz

$$f_c = \left\{ \begin{array}{ll} \dfrac{3}{t_d} & \text{, concha} \\ \dfrac{4}{t_d} & \text{, helix} \end{array} \right\}, \quad t_d = \frac{4 d_r}{c}$$

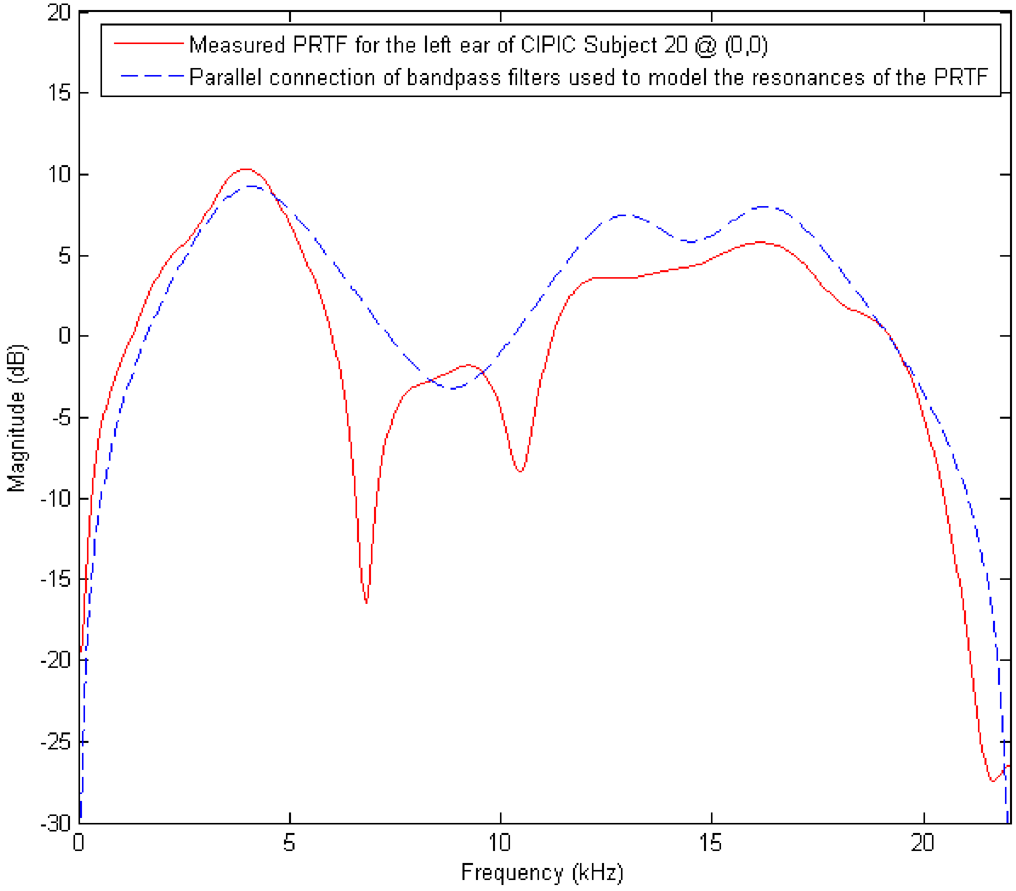
WIDTH RESONANCES PLOT AT (0,0)



Cascading of Modeled Resonances

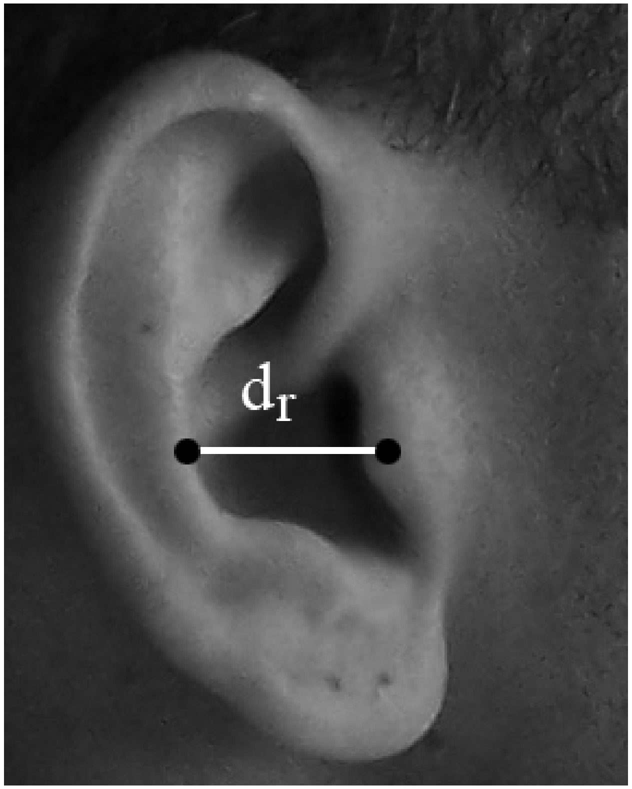


Cascading of Modeled Resonances Plot



Modeling Notches at (0,0)

- Notches are created from reflections off of the concha or the helix:



$$t_d = \frac{4d_r}{c}$$

MODELING NOTCHES AT (0,0)

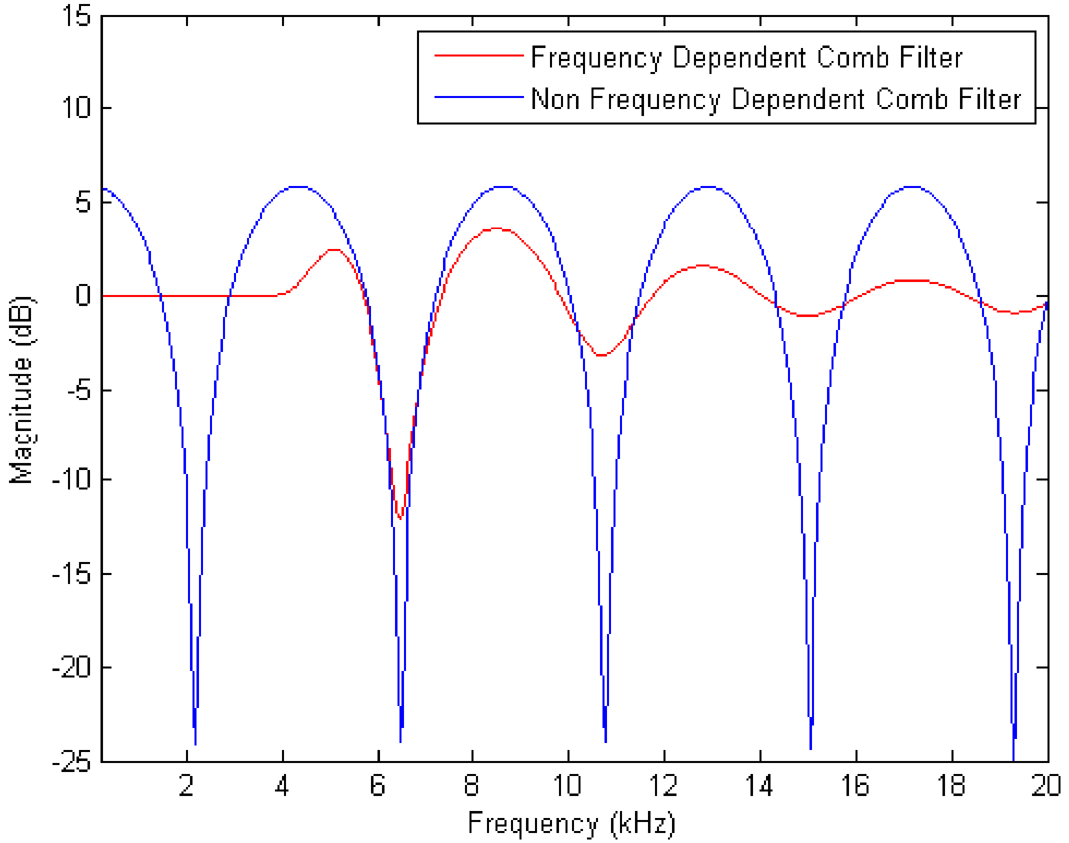
- Delay and add impulse response of cascaded resonances according to:

$$y(t) = x(t) + \Gamma x(t - t_d)$$

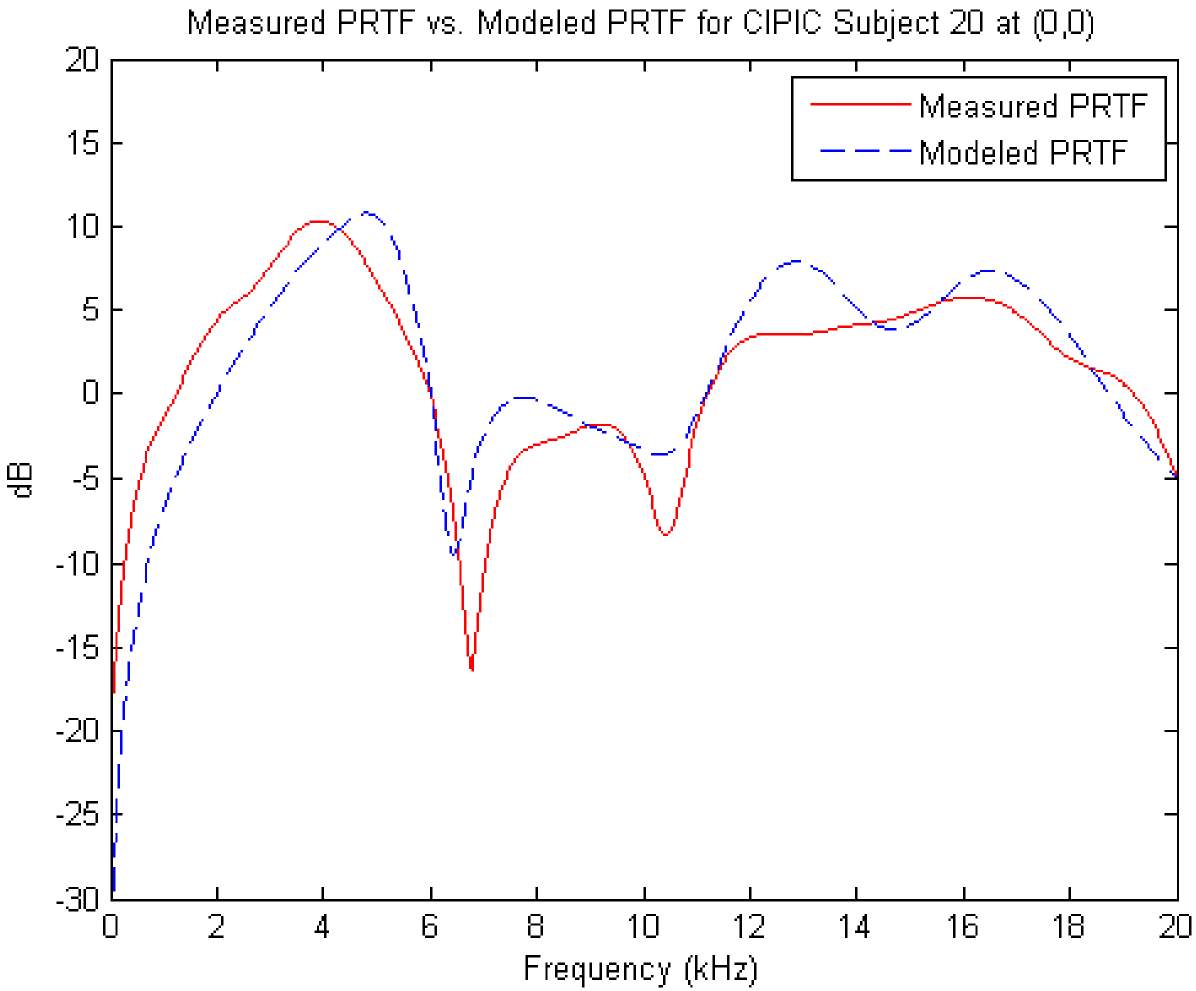
- Reflection coefficient is of band-pass nature and orientation dependent

- Time delays are very small, up and down-sampling by 10x is required to get necessary precision from the delay process

Effect of Reflection Coefficient



Complete PRTF Model at (0,0)



Elevation Dependencies

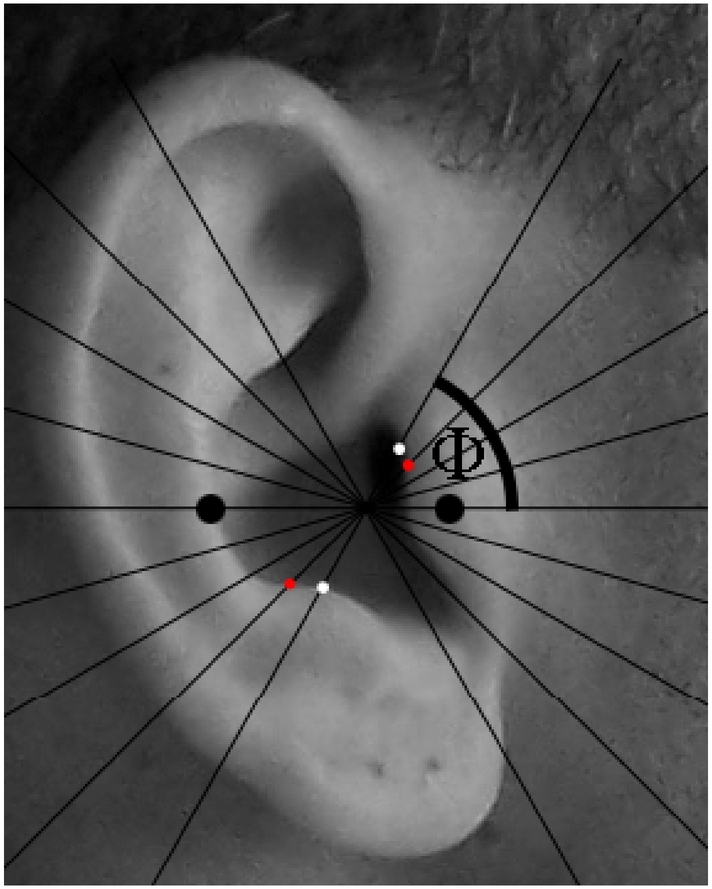
- Reflection distance changes with elevation

- Gain of width resonances change with elevation

- Bandwidth of depth resonance changes with elevation—due to the presence of height resonances (pinna modes 2 & 3)

- Response of reflection coefficient changes with reflection distance

ELEVATION DEPENDENCE (REFLECTION)



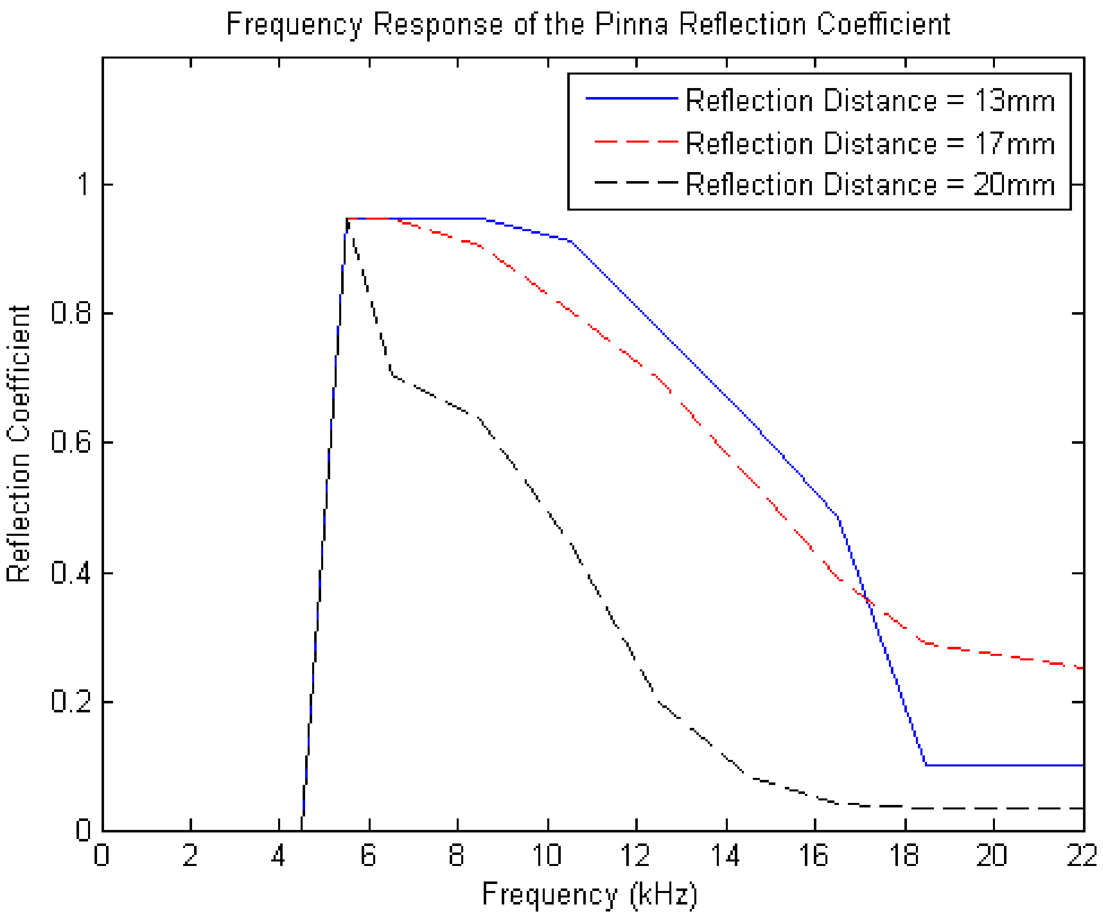
Gaining Precision by Up and Down Sampling Impulse Response

- Especially necessary for high elevations where reflection distances are very short

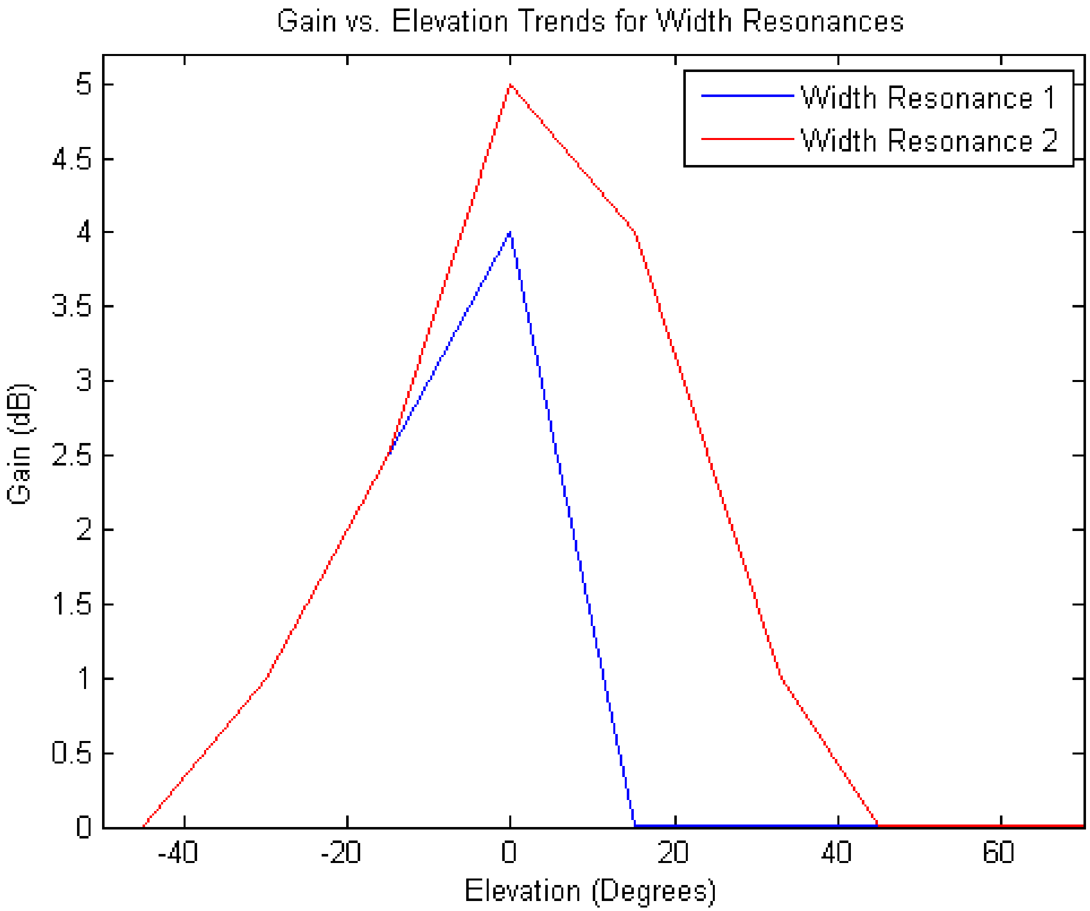
- Example:

Elevation Angle (Φ)	45°	60°
Reflection Distance	14mm	13.4mm
Corresponding Analog Time Delay	163.2653μs	156.2682μs
Analog Notch Location	**9.187kHz**	**9.5989kHz**
Corresponding Digital Time Delay (nearest sample precision @ 44.1kHz)	7 samples	7 samples
Corresponding Digital Notch Location	**9.450kHz**	**9.450kHz**
Corresponding Digital Time Delay (tenth of a sample precision @ 44.1kHz)	7.2 samples	6.9 samples
Digital Notch Location after up and down sampling	**9.1875kHz**	**9.58696kHz**

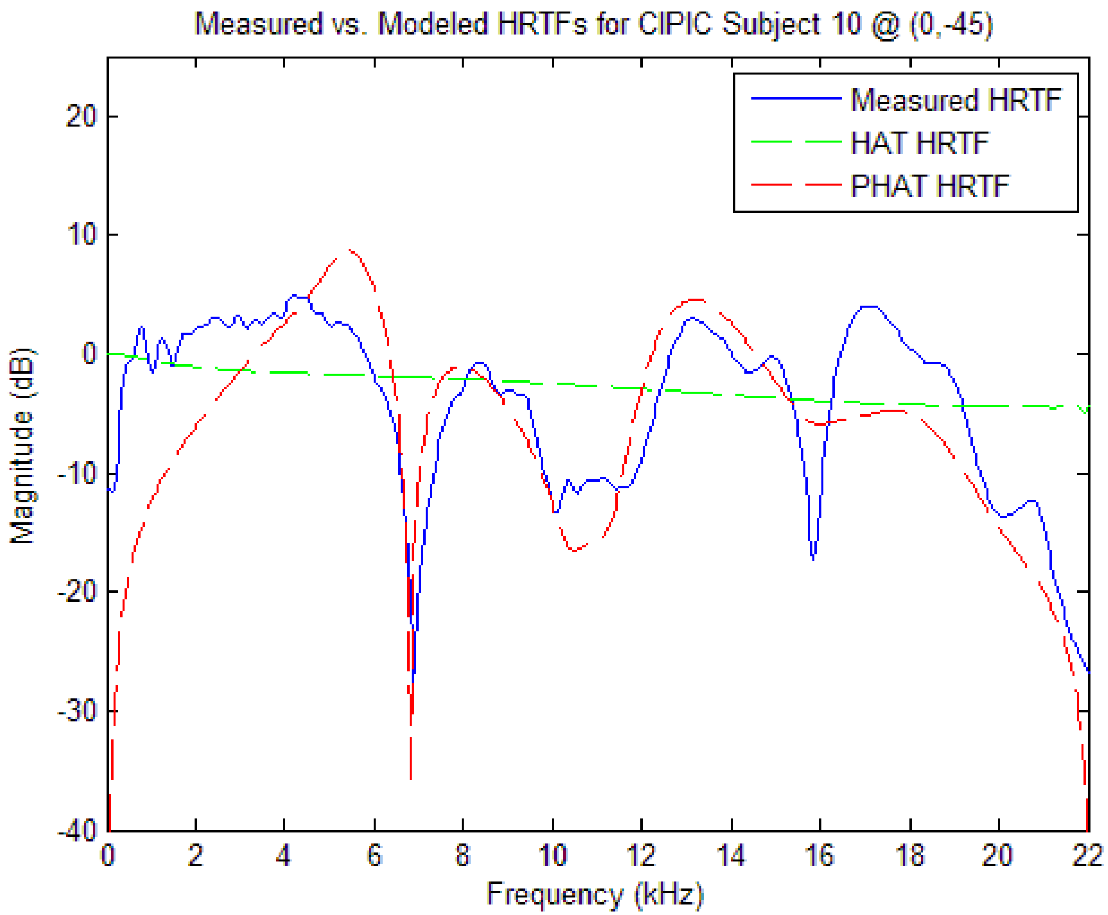
ELEVATION DEPENDENCE (REFLECTION COEFFICIENT)



Elevation Dependence (Resonances)

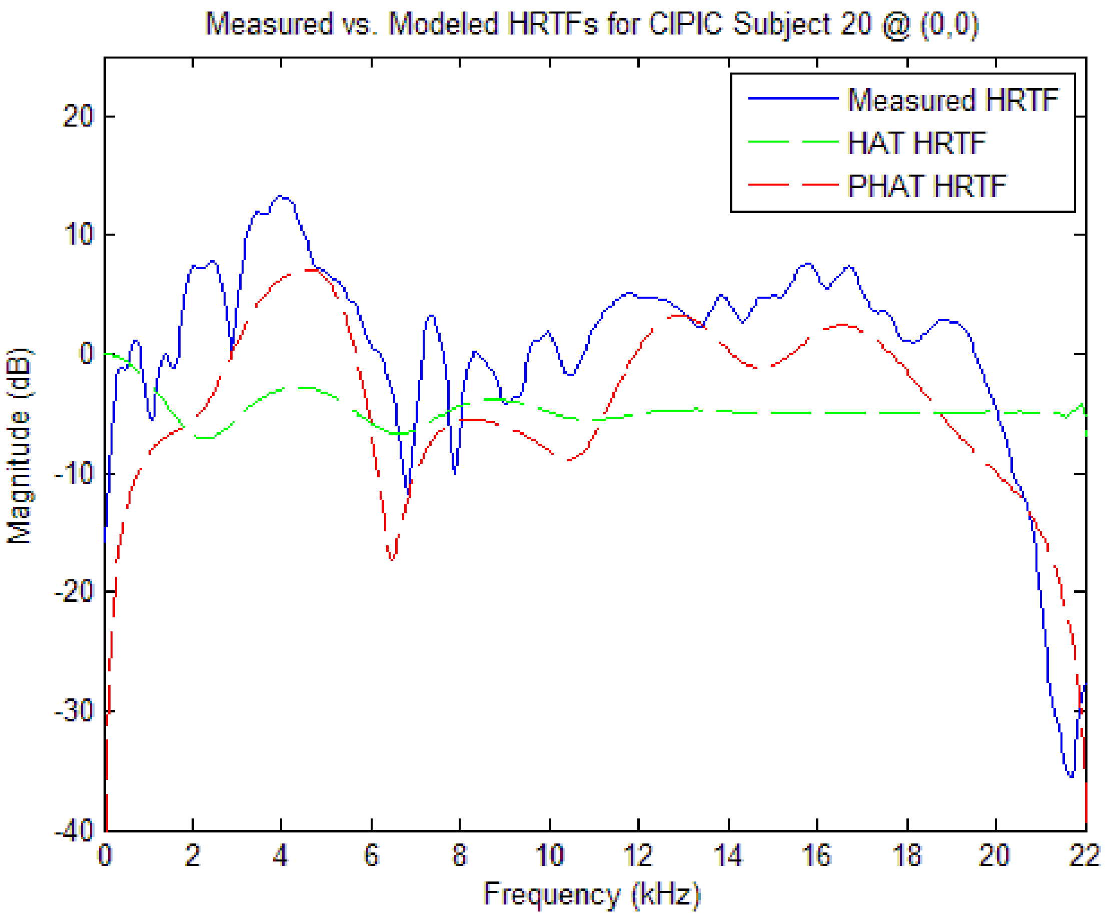


OBJECTIVE RESULTS AT (0,-45)

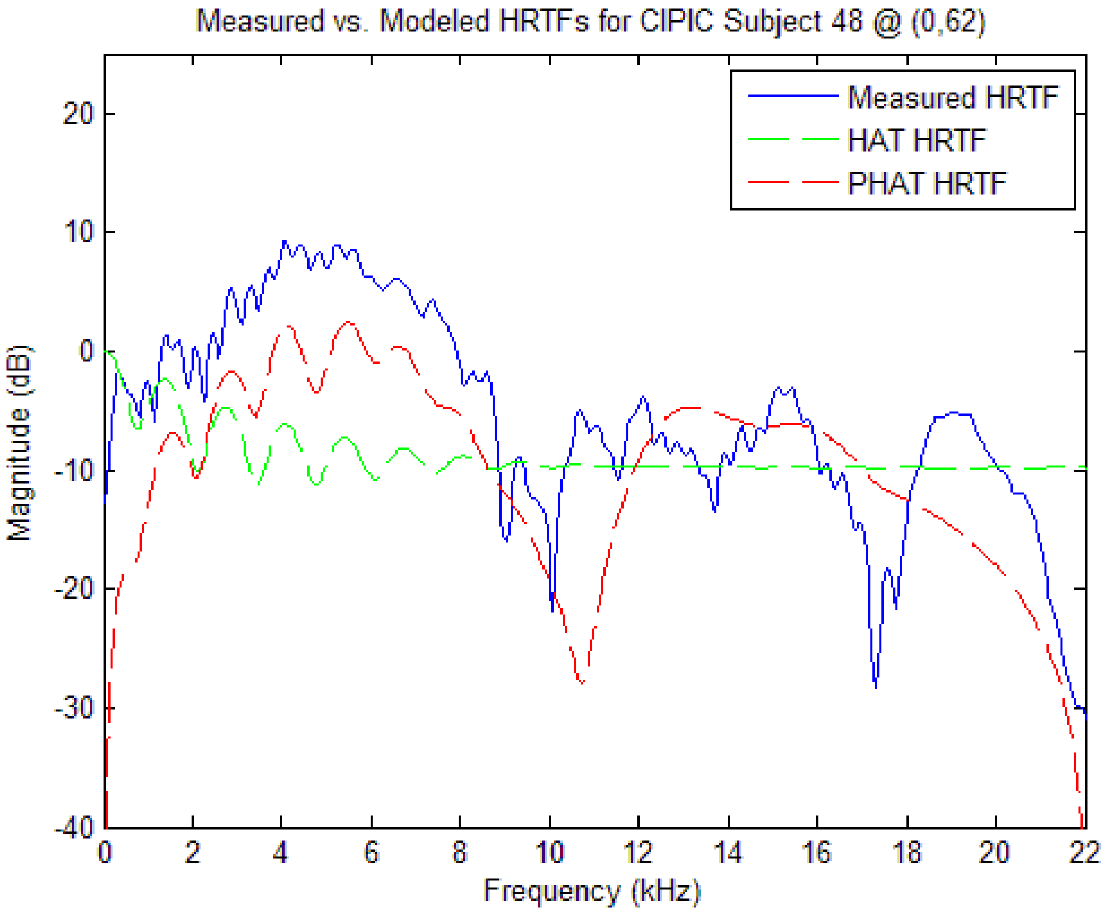


Return

OBJECTIVE RESULTS AT (0,0)



OBJECTIVE RESULTS AT (0,62)



LISTENING TEST PROCEDURE

- Take seven digital photographs of the subject’s anthropometry

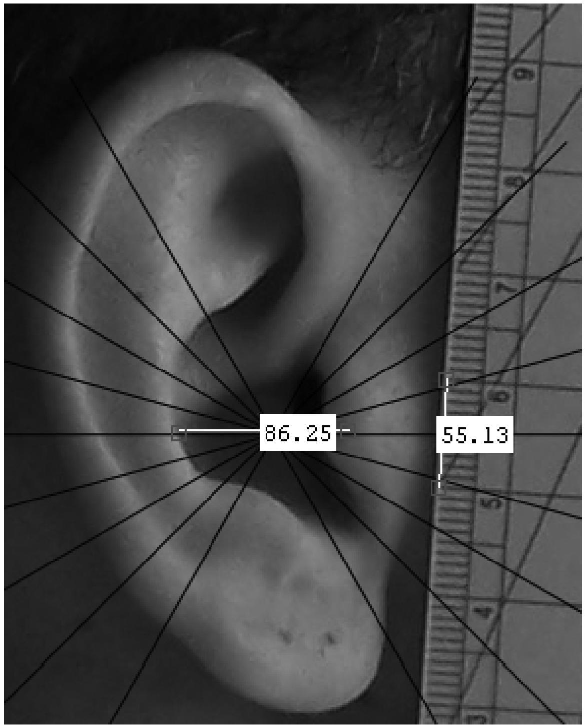
- Extract necessary measurements from the images

- Input the acquired anthropometry as parameters to the HAT and PHAT algorithms at select spatial locations

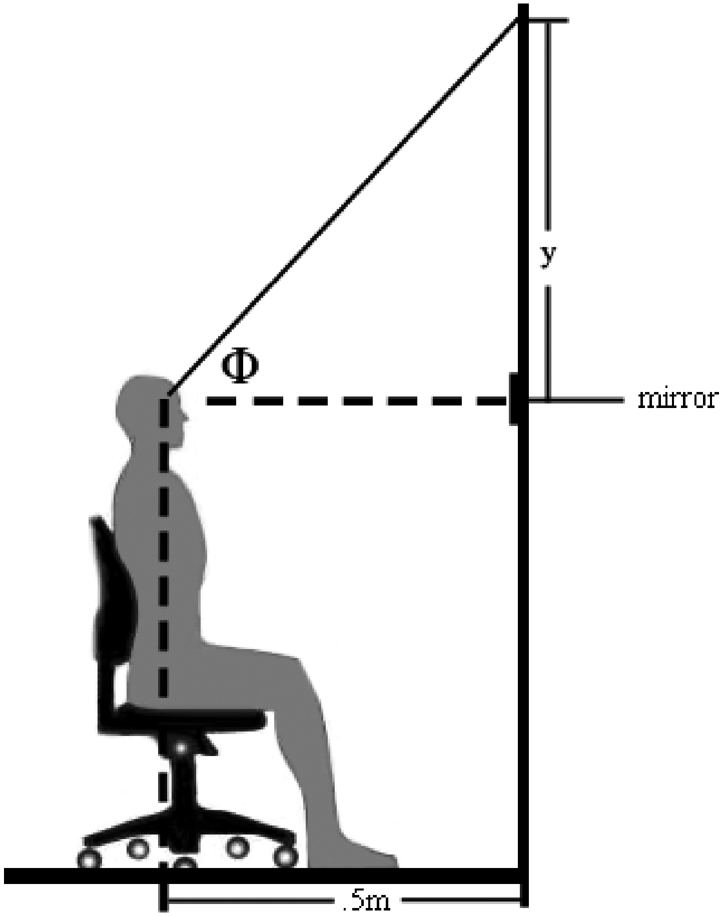
- Generate customized HRTFs using each algorithm

- Filter test sound

- Playback filtered sounds for listeners and have them localize the stimulus



Listening Test Setup



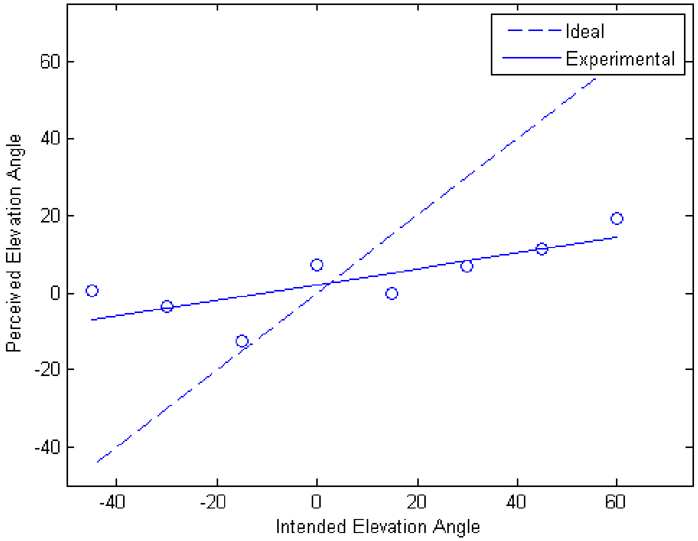
SUBJECTIVE RESULTS

- 15 subjects tested representing a wide variety of pinna shapes and body dimensions

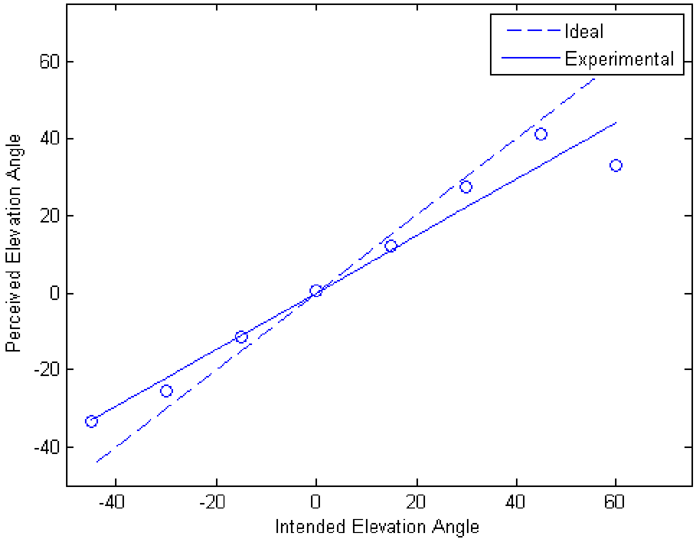
- Results produce three distinct groups:
  - **Group I** – good localization performance at all locations **(47% of subjects)**
  - **Group II** – good localization performance at an azimuth of 20°, poor localization performance at an azimuth of 0° **(20% of subjects)**
  - **Group III** – poor localization performance at all locations **(33% of subjects)**

SUBJECTIVE RESULTS (GROUP I, Θ=0°)

HAT results at an azimuth of 0 for Group I
Regression Slope = 0.20

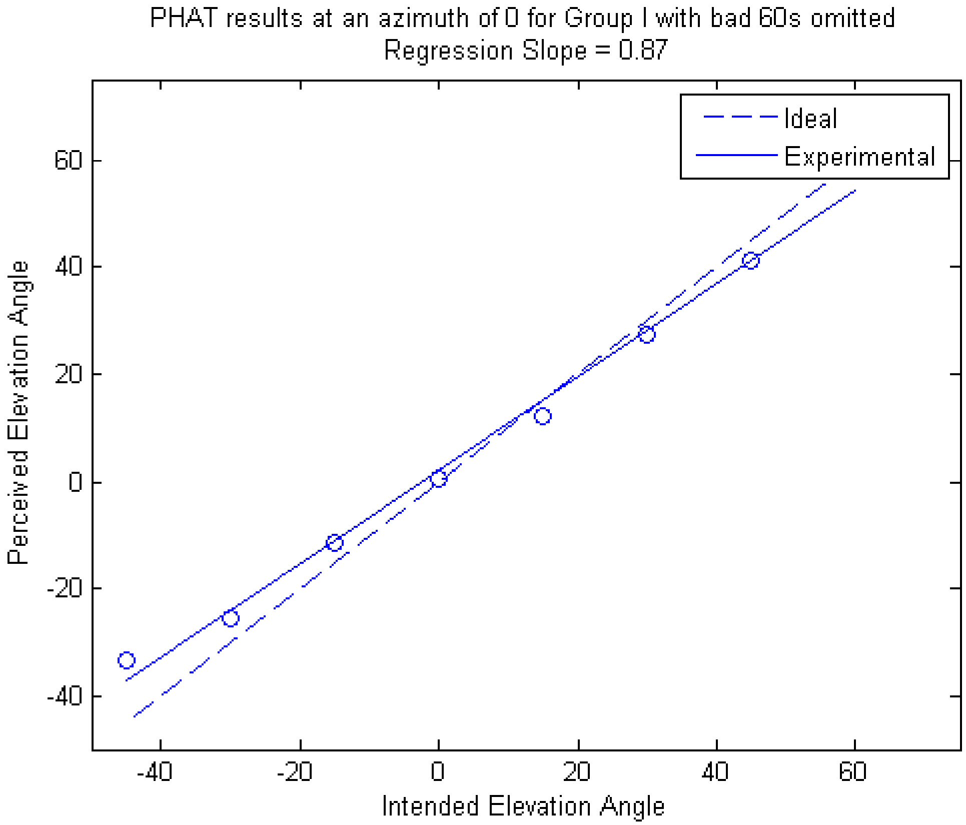


PHAT results at an azimuth of 0 for Group I
Regression Slope = 0.73

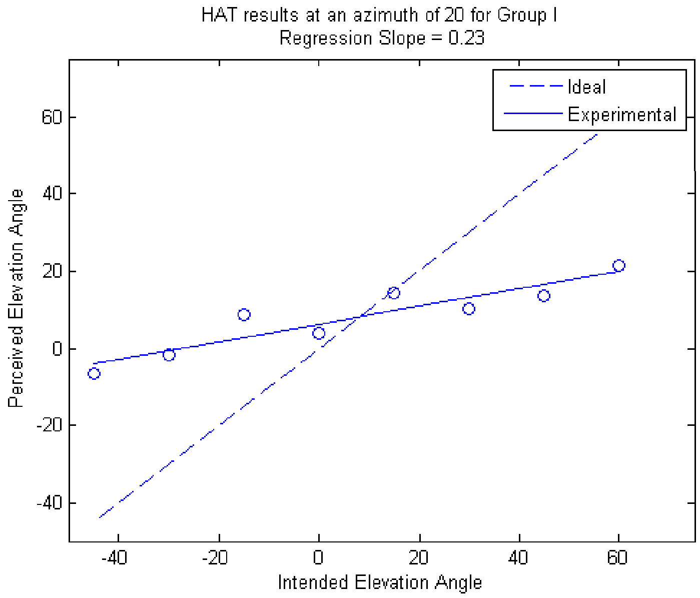


Link

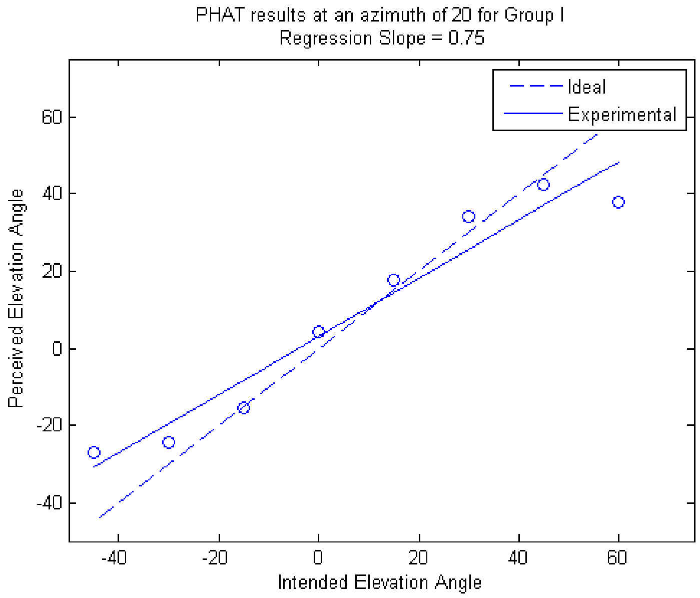
Subjective Results (Group I, Θ=0° with outliers removed)



Subjective Results (Group I, Θ=20°)

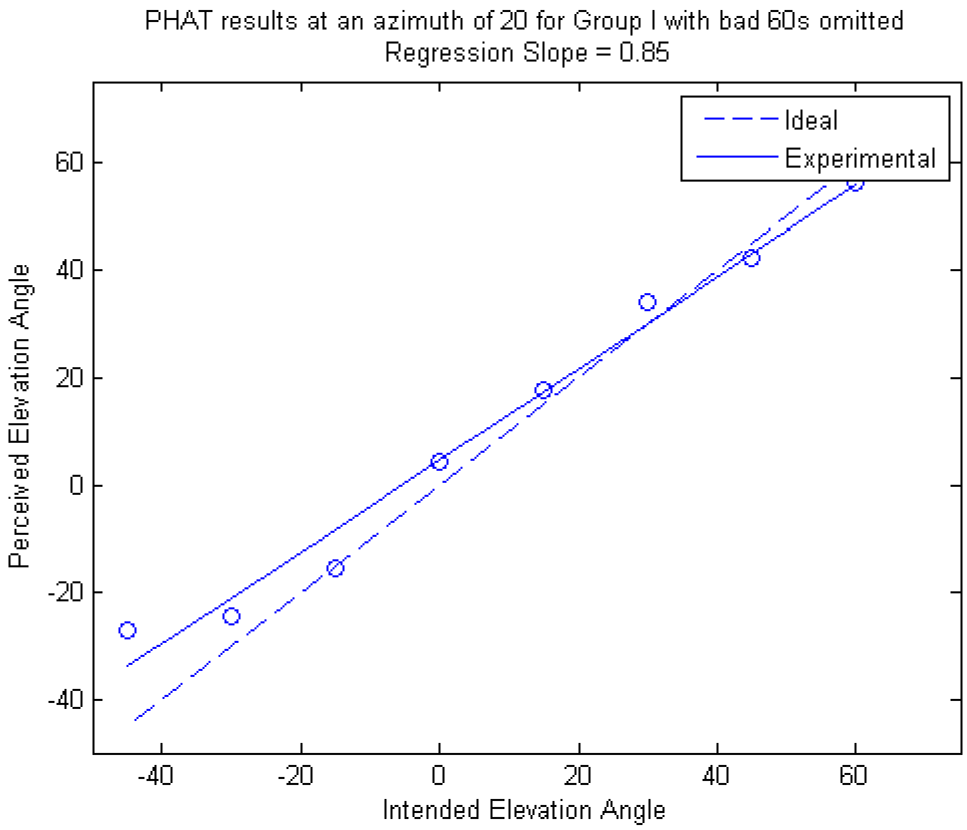


HAT results at an azimuth of 20 for Group I
Regression Slope = 0.23

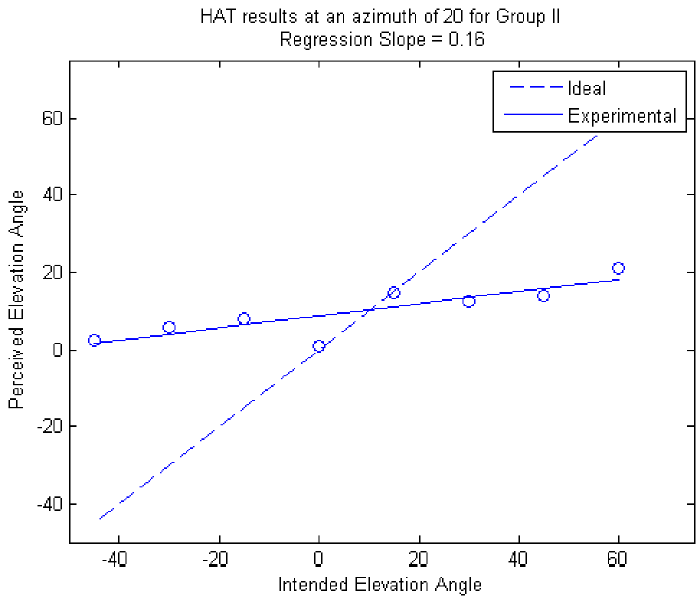


PHAT results at an azimuth of 20 for Group I
Regression Slope = 0.75

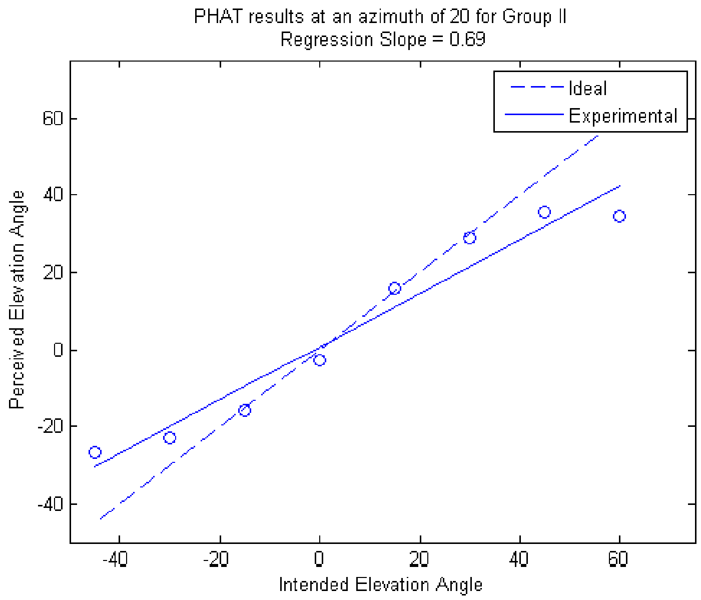
Subjective Results (Group I, Θ=20° with outliers removed)



Subjective Results (Groups I & II, Θ=20°)

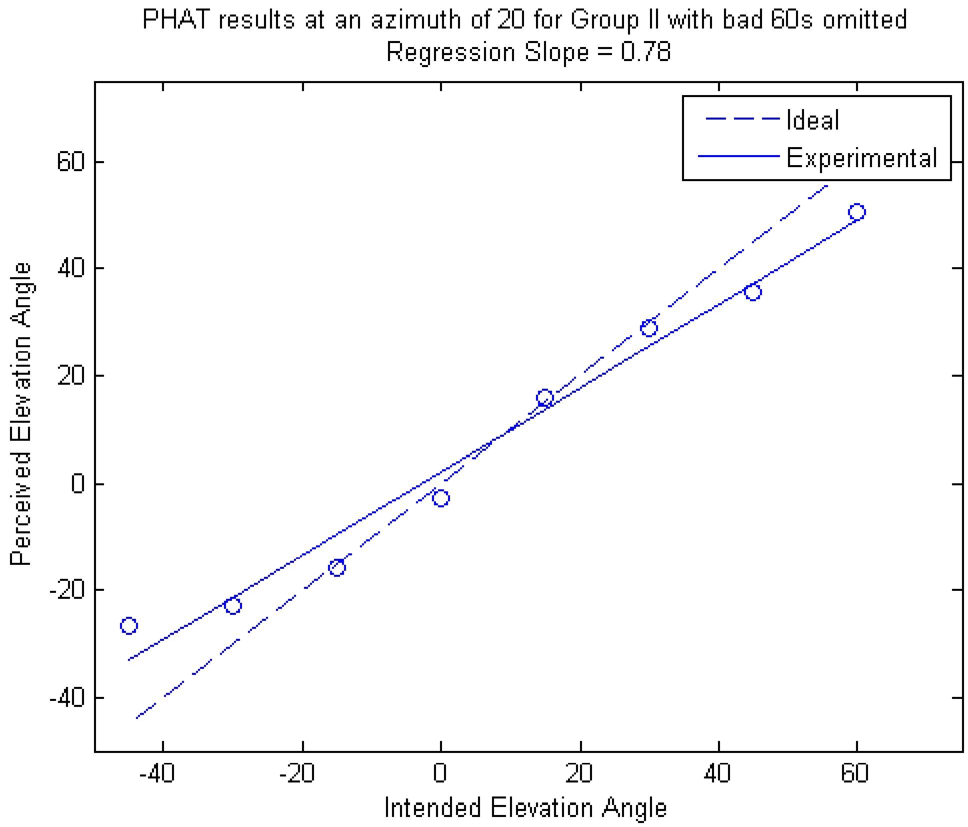


HAT results at an azimuth of 20 for Group II
Regression Slope = 0.16



PHAT results at an azimuth of 20 for Group II
Regression Slope = 0.69

SUBJECTIVE RESULTS (GROUPS I & II, Θ=20° WITH OUTLIERS REMOVED)



CONCLUSIONS

- Delay and add pinna model breaks down if reflection distance is less than 10.2mm

- Model is more effective for people with wide and non-flat pinnae

- PHAT model consistently out performs the HAT model despite its identified flaws

- Actual values of the resonances' gains do not matter as much as their gains relative to each other

- The pinna's primary reflector for some people is not always either the concha or the helix. The model will fail in such cases (about 20% of the subjects in the CIPIC database)

- Some subjects commented that they only heard timbral differences in the sounds and were unable to localize them

Possible Future Work

- Extend the model to function for all azimuths in the frontal hemisphere using a 3-D model of the ear

- Link gains of resonances to anthropometry

- Study the nature of the ear’s primary reflector more in depth to solve the 10.2mm break down problem

- Add effective externalization to the model

- Establish a more definitive method for superimposing the coordinate system onto an image of the ear

QUESTIONS