Binaural

A Quick Introduction

Toplap Barcelona

Timothy Schmele

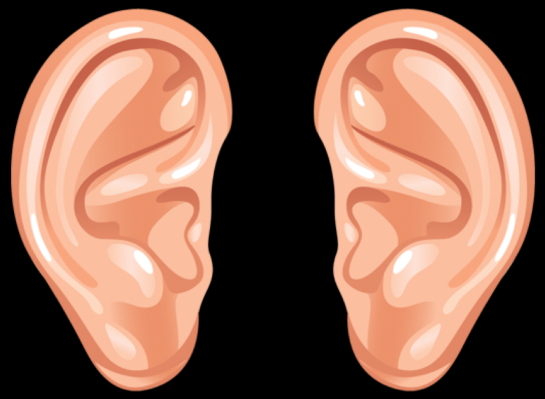
What we will cover

- Basics & historic overview
- Psychoacoustic background
- Head Related Transfer Function (HRTF)
- Capturing Binaural Filters
- Challenges with Binaural Audio Applications
- Ambisonics

Toplap Barcelona

Basics

- Binaural: listening with two (bi) ears (aural)
- 3D listening mainly possible because of outer ear shape
- Note: outer ear shape is unique like a fingerprint!



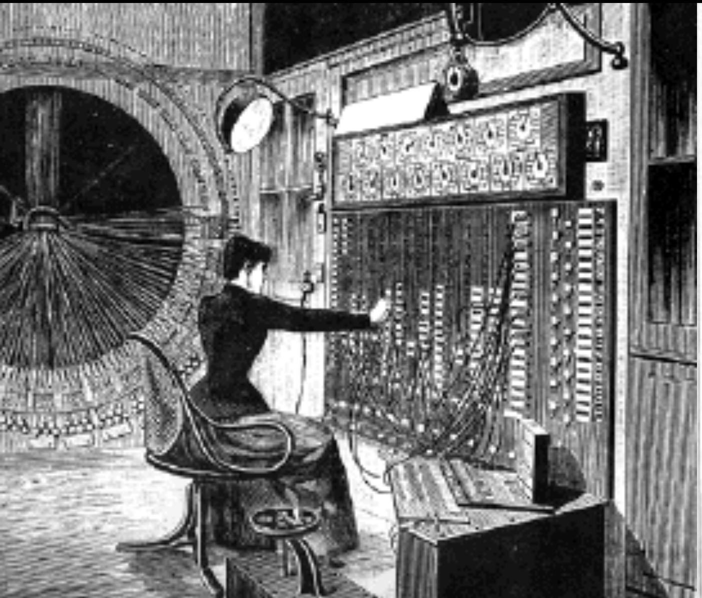
Toplap Barcelona

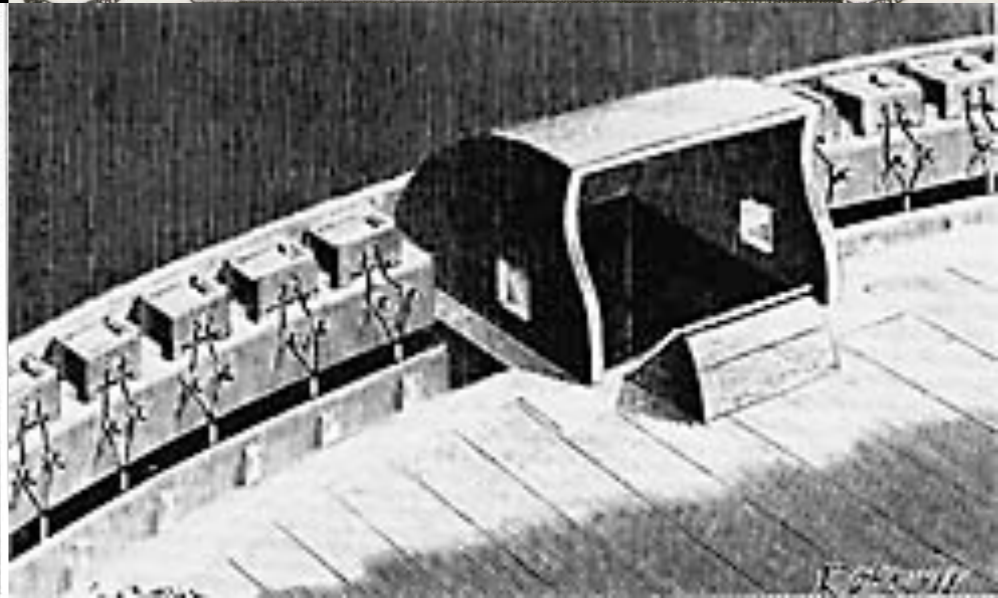
Basics

- *Binaural synthesis*: simulate effects that the outer ear imposes on the incoming sound
- Delivered over headphones
  - Tight control which ear receives which channel
  - Head Related Transfer Function (HRTF)
- ≠ Stereo
  - (gr.) solid, derived from stereoscopic
  - Synonymous with old-school two channel audio
- ≠ Transaural
  - Binaural transmitted over speakers
  - Requires cross-talk filters and is very fragile

History

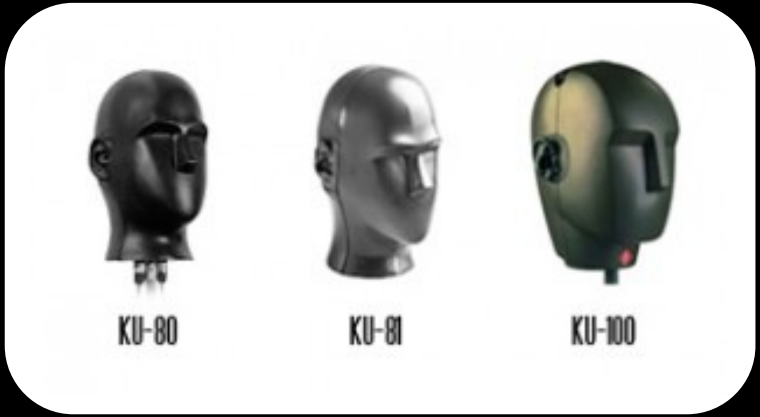
- *Théâtrophone*
  - Two channel live telephone transmission
  - Invented 1881 by Clément Ader in France
  - Pay-by-minute device, superseded by radio





History

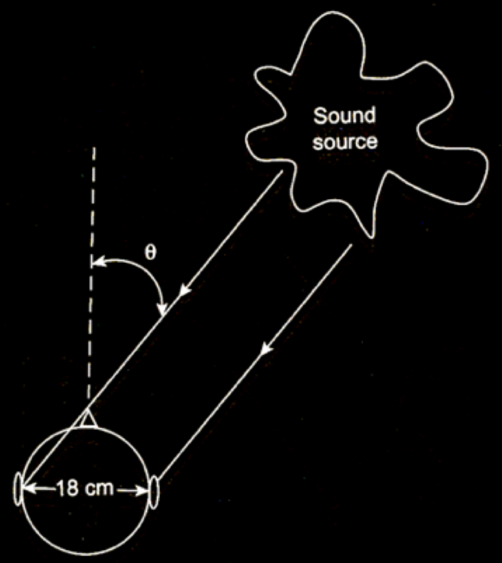
- First binaural hearing aid 1921 by Fletcher in the USA
- 1933 AT&T demonstrates binaural dummy head Oscar
- Neumann produces the KU since 1972
- First in ear microphones from Sennheiser 1974
- 2012: companies like 3Dio make affordable dummy ear microphones



Toplap
Barcelona

Psychoacoustics: Directional Hearing

- Difference in ear position produces differences in heard sound
- ITD: interaural time difference, difference in time of arrival between ears
- ILD: interaural level difference, shadowing effect of head
- ITD and ILD covered by *regular* two channel audio (stereo)
- “So, binaural …?”



Toplap Barcelona

Psychoacoustics: Directional Hearing

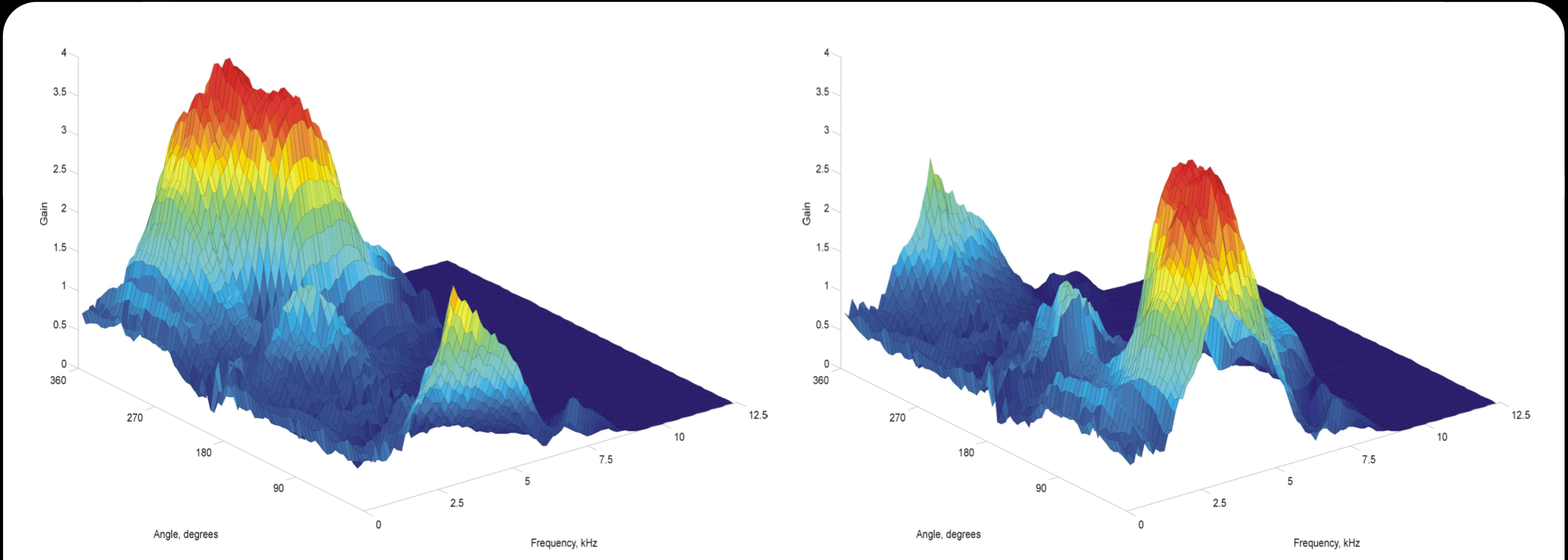
- Modern *binaural* addresses 3D listening, including the *vertical*!
- Sound reflects off of outer ear and torso differently depending on angle
- Physiology produces many reflections that enter at slightly different times
- Reflections sum together creating *comb filter* effects

Toplap Barcelona

Toplap Barcelona

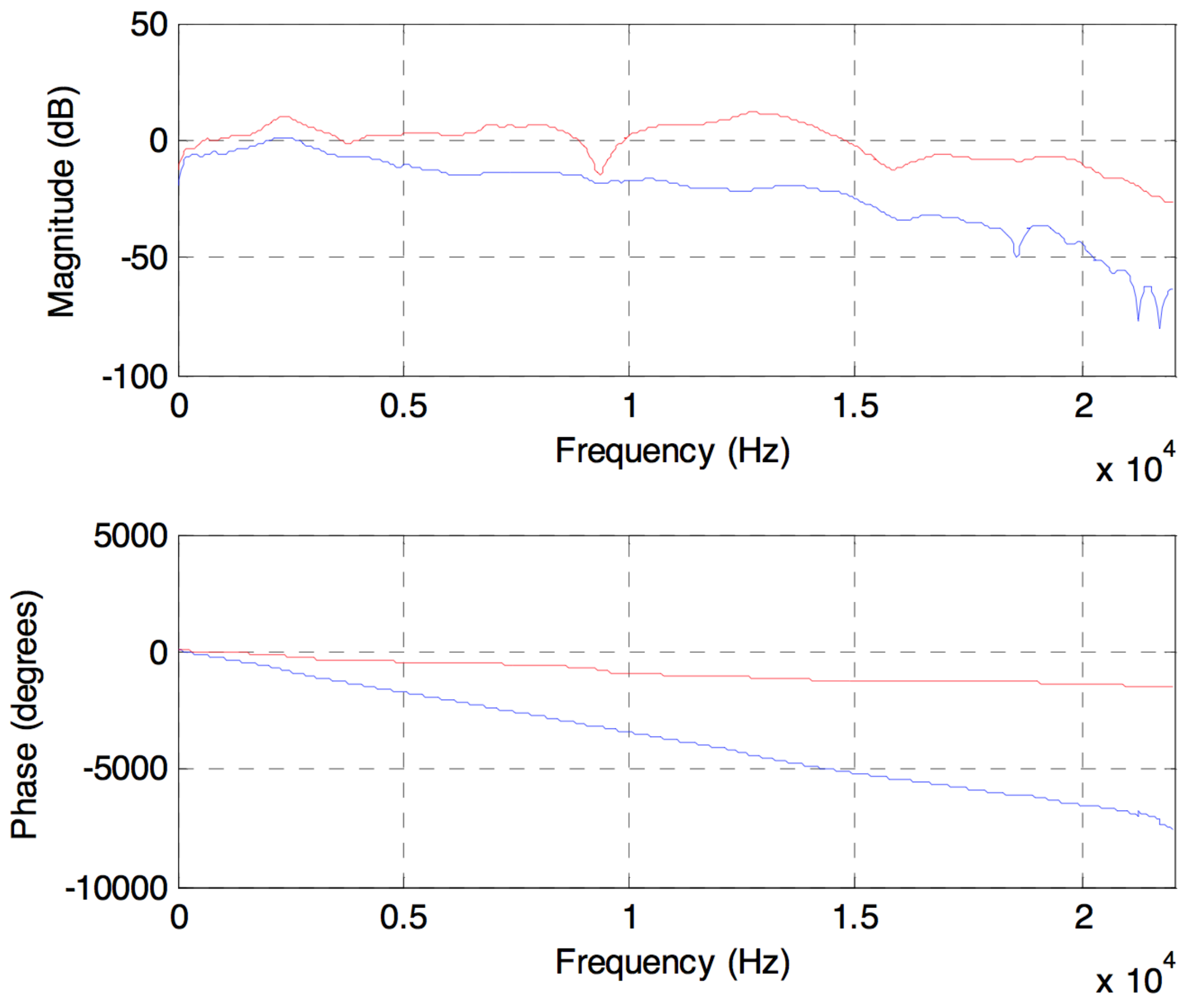
Head Related Transfer Function (HRTF)

- Describes amplitude of frequency, as a function of incident angle
- Summarizes all those tiny reflections into a single 3D filter
- Used to simulate the effect that our physiology has on incoming sound



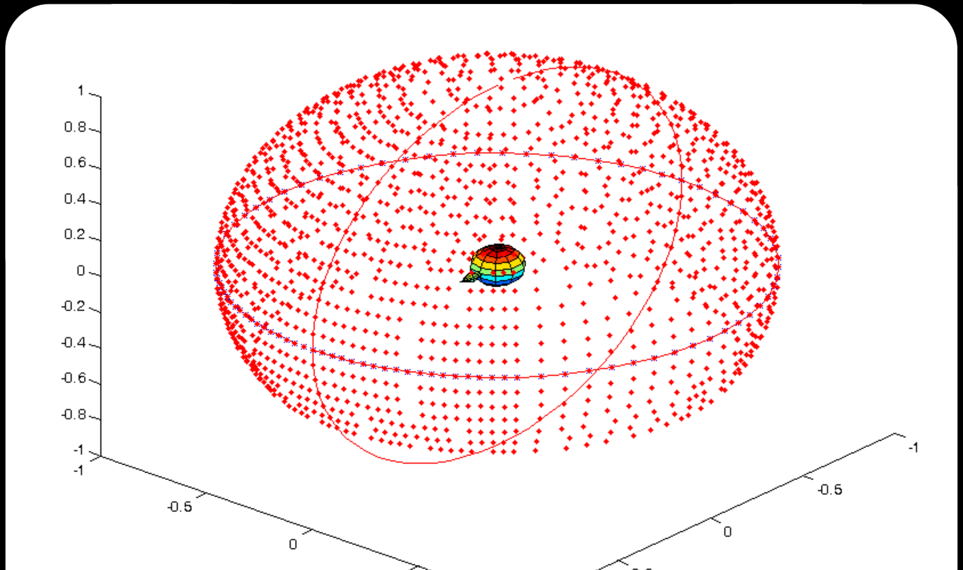
Head Related Transfer Function (HRTF)

- For example, we choose a single angle and receive the magnitude and phase response of the filter
- Here, example at 90°:



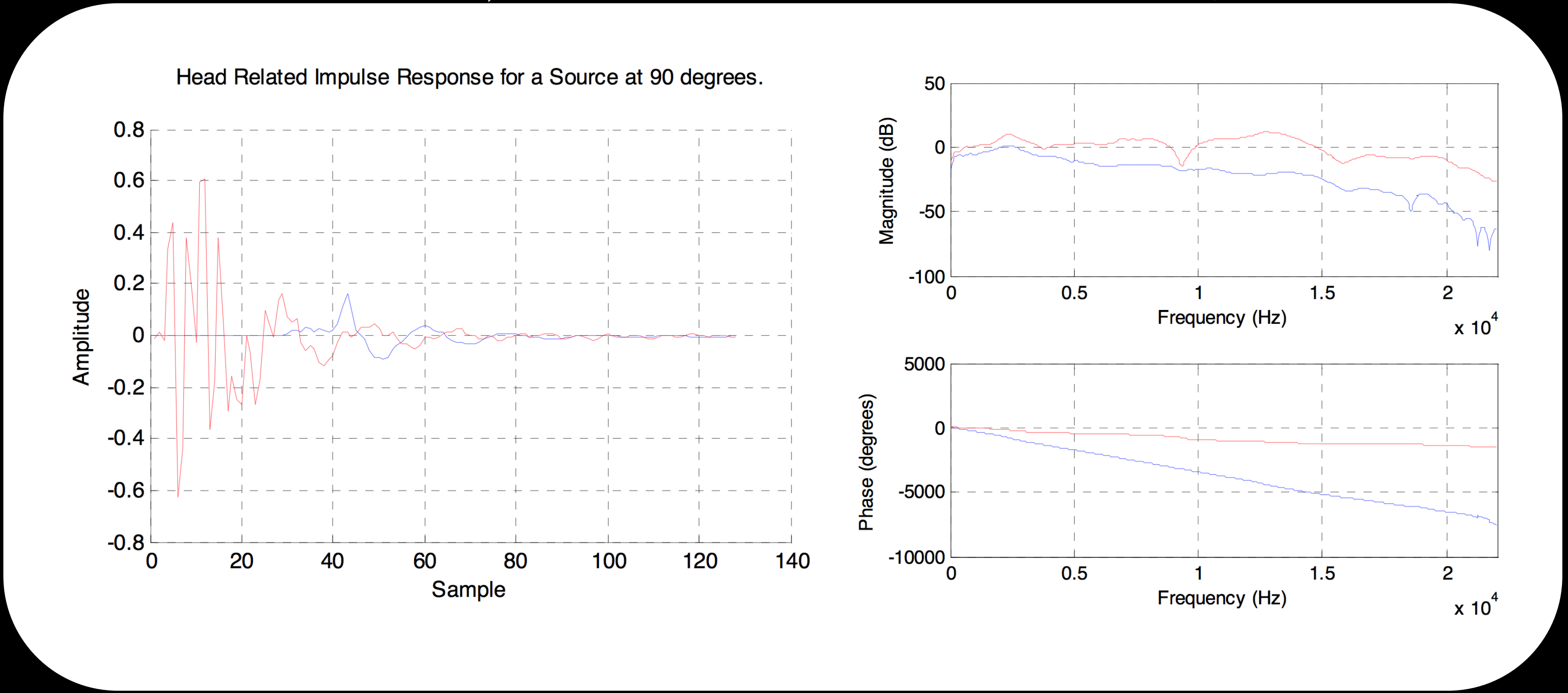
HRTF Measurement

- HRTF is built using many Head Related Impulse Responses (HRIR)
  - Impulse: click, balloon pop… also, sine sweep!
  - Response: how a system like a room or the human physiology modifies the click
  - Head related: in relation to the head orientation
- The HRTF is the fourier transform (frequency analysis) of the HRIR



Toplap Barcelona

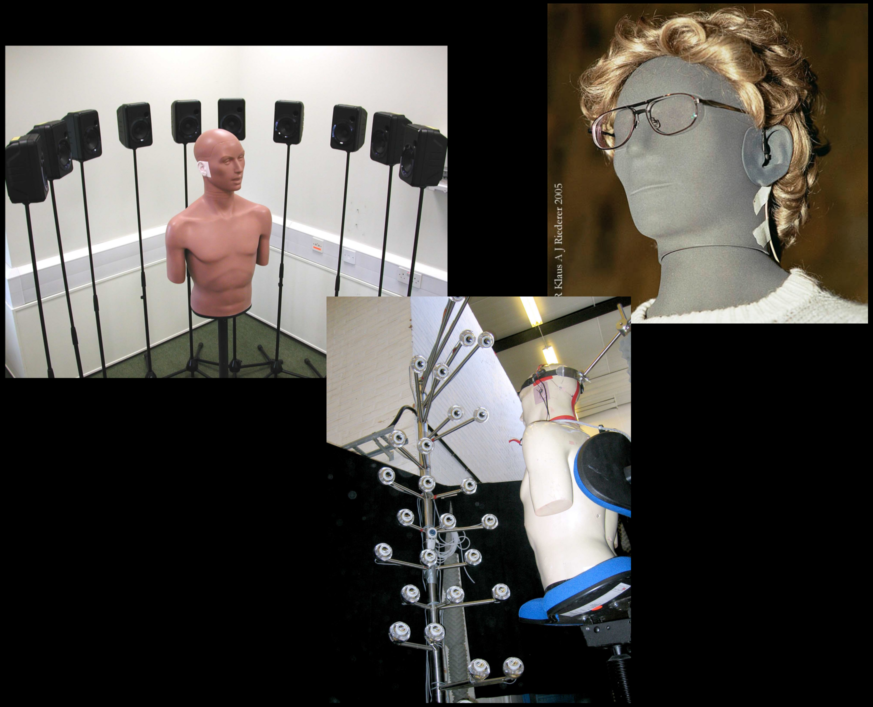
HRTF Measurement

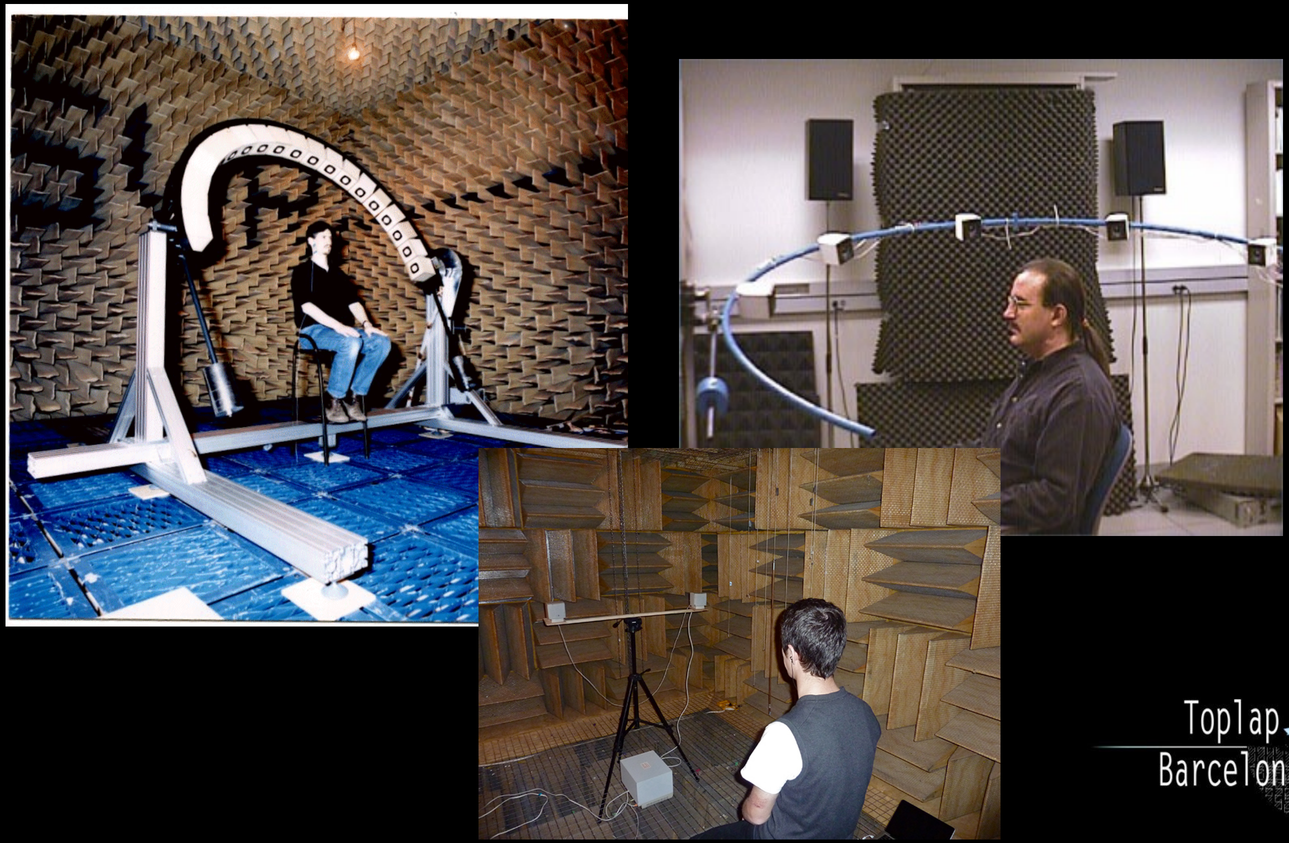


Toplap Barcelona

HRTF Measurement

- HRIRs are measured on human beings (CIPIC, IRCAM, …) or with artificial heads (Kemar, KU-100, …)





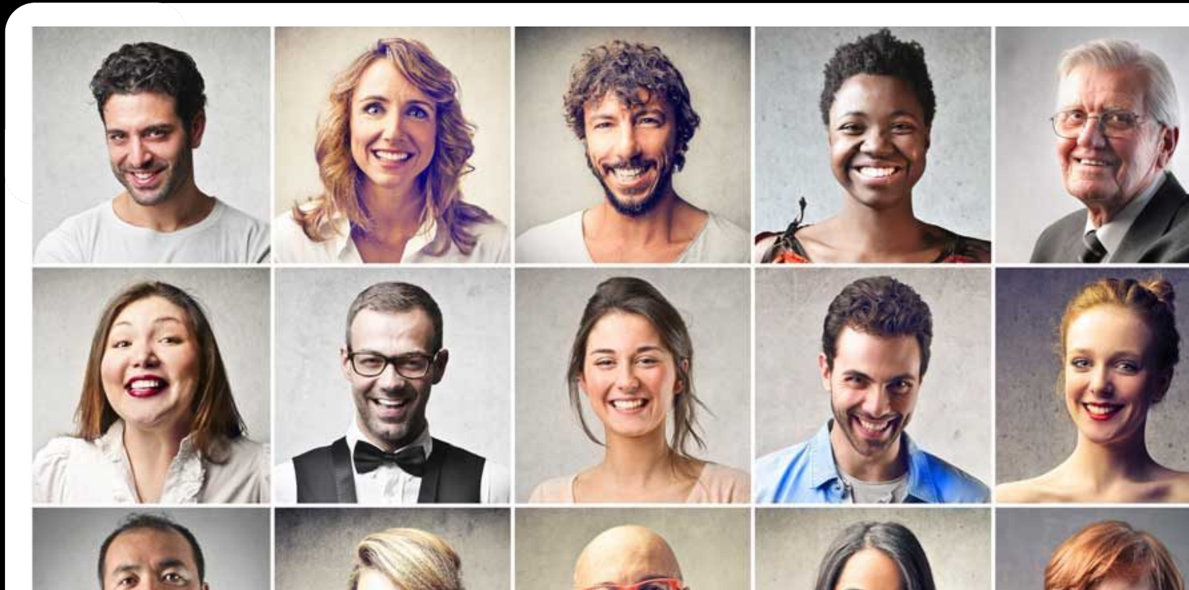
Toplap Barcelona

Binaural Reproduction

- If dummy head or in-ear recording: playback recording!
- When synthesizing:
  - Define sound source position
  - Get filter for given position
    - a) recover HRIR from HRTF for given position
    - b) choose closest HRIR from database
  - Apply filter to sound
    - If HRIR, treat as FIR and use convolution!

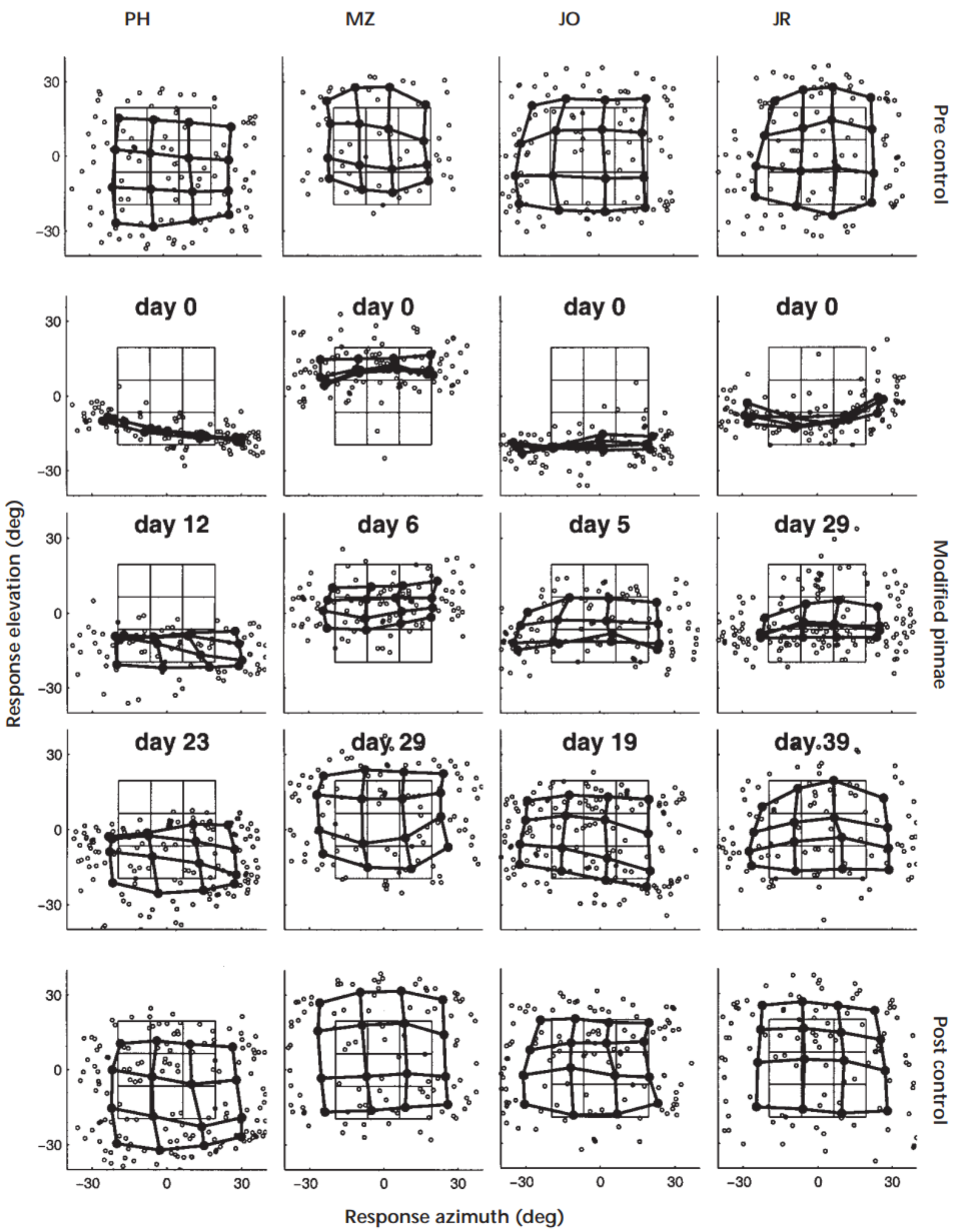
Every HRTF Is Unique!

- HRTF is personal, like a finger print
  - Every human’s physiology is unique!
- Brain learns personal incoming space-sound-characteristics since birth
- Other senses (visual, motoric, …) help cross-validate information
- Properties of artificial heads are scientific average, but sometimes still work badly for some listeners
- Databases with human measurements often supply many variants to cover as broad of a spectrum of society as possible



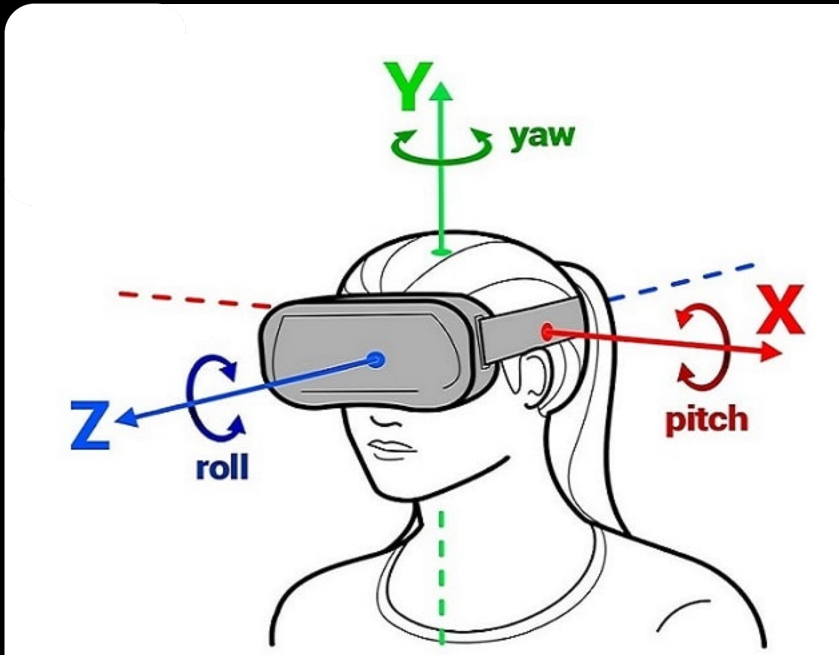
Every HRTF Is Unique!

- Brain can learn to use "several" ears!
- Experiment on right modifies outer ears of participants → vertical localization collapses!
- Participants learn to "listen" with new, modified ears over a month
- When ears returned to original, localization is not impeded!



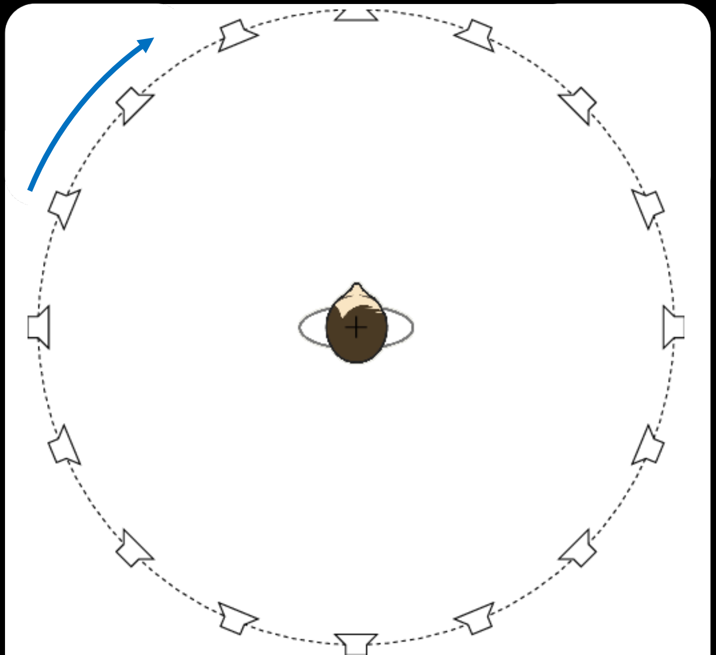
Externalization

- Binaural often sounds as if coming from inside the head
- Problem: sound through headphones and personal head movement is coupled!
  - In nature, a sound source will remain in it’s place
  - → moving the head will not make sound source move in same direction
- Solution: head tracker
  - Track rotation movement of head
  - Artificially move sound sources in opposite direction
- Also, reverb can help with externalization!



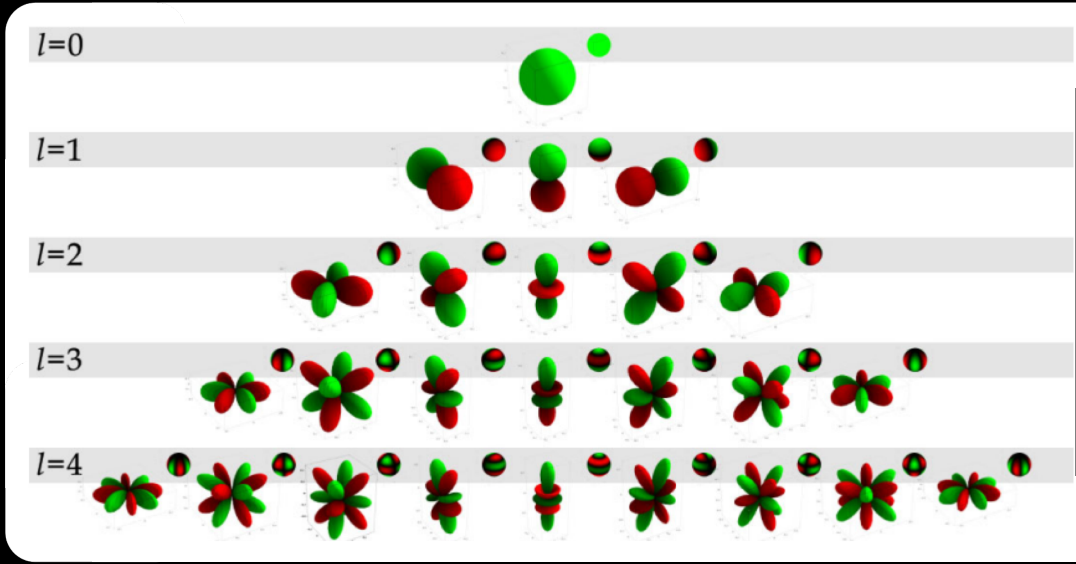
Interpolation

- HRIRs often measured in discrete points
  - A smooth movement cannot jump from measuring point to measuring point
  - Intermediate positions need to be estimated
- Interpolation in: Amplitude, Frequency, Impulse, …
- Most common: virtual speaker panning!
  - Treat HRIR measurement points as positions of virtual speakers
  - Use any 3D panning algorithm to move between measurement points
  - VBAP: Vector Base Amplitude Panning
  - Ambisonics →



Ambisonics

- Physical description of *directional* sound pressure (aka. velocity) in a point
- Using a decomposition into *spherical harmonics*
- Higher orders add *spherical harmonics*, adding information
- Basic 1st order has 4 channels:
  - 1 omnidirectional pressure channel
    - taken from 0th order
  - 3 directional velocity channels
  - 1st order Velocity components can be seen as perpendicular figure-8 microphones
- *Spherical harmonic* representation using 4 channels is also called *B-Format*



Ambisonic Microphones

- Can record 3D in Ambisonic *B-format* “directly”
  - Note: the raw capsule output is *not* the desired *B-Format*!

- Higher order microphones also exist:

Ambisonic Synthesis

- You can also synthesize the pressure and velocity components of Ambisonics → i.e. synthesize *B-format*
- Used to pan sources in a 3D setup using Ambisonics
- Once in *B-Format* both synthesized and recorded sources can be freely combined and played back together from a single *B-format* mix-down

$$W(t) = \sum_i s_i(t)/\sqrt{2}$$

$$X(t) = \sum_i s_i(t) \cos\theta_i \cos\delta_i,$$

$$Y(t) = \sum_i s_i(t) \sin\theta_i \cos\delta_i,$$

$$Z(t) = \sum_i s_i(t) \sin\delta_i,$$

Toplap Barcelona

Ambisonic Transformation

- Ambisonic signals in B-Format can be spatially manipulated, like warp, push, zoom…
- Most common: rotation
  - Done using simple trigonometric rotation matrices
  - Here, an example for 1st order
  - Higher orders (>3rd) usually discretize the sound sphere and rotate each point individually
- Rotation matrices often applied in binaural for head rotation compensation (externalization) using a head tracker

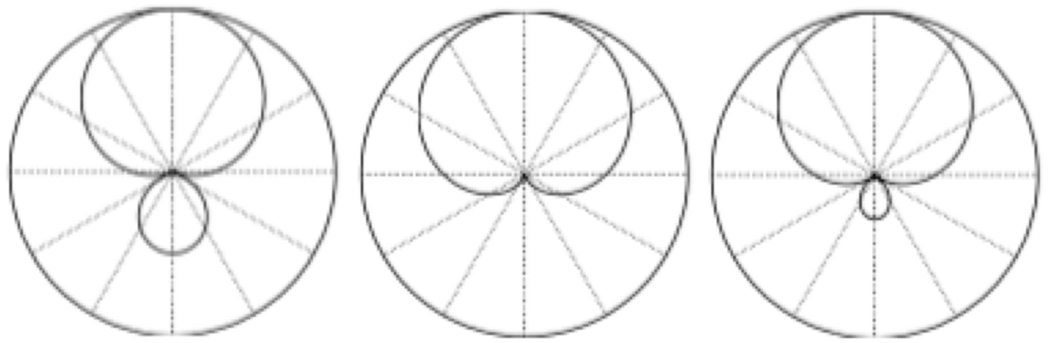
$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Ambisonic Decoding

- *Spherical Harmonics* decoded to speaker system using “virtual microphone” technique
- Microphone characteristic determines quality of decoder
  - in-phase: avoid out of phase signals
  - max-r_e: maximize psychoacoustic directionality vector r_e

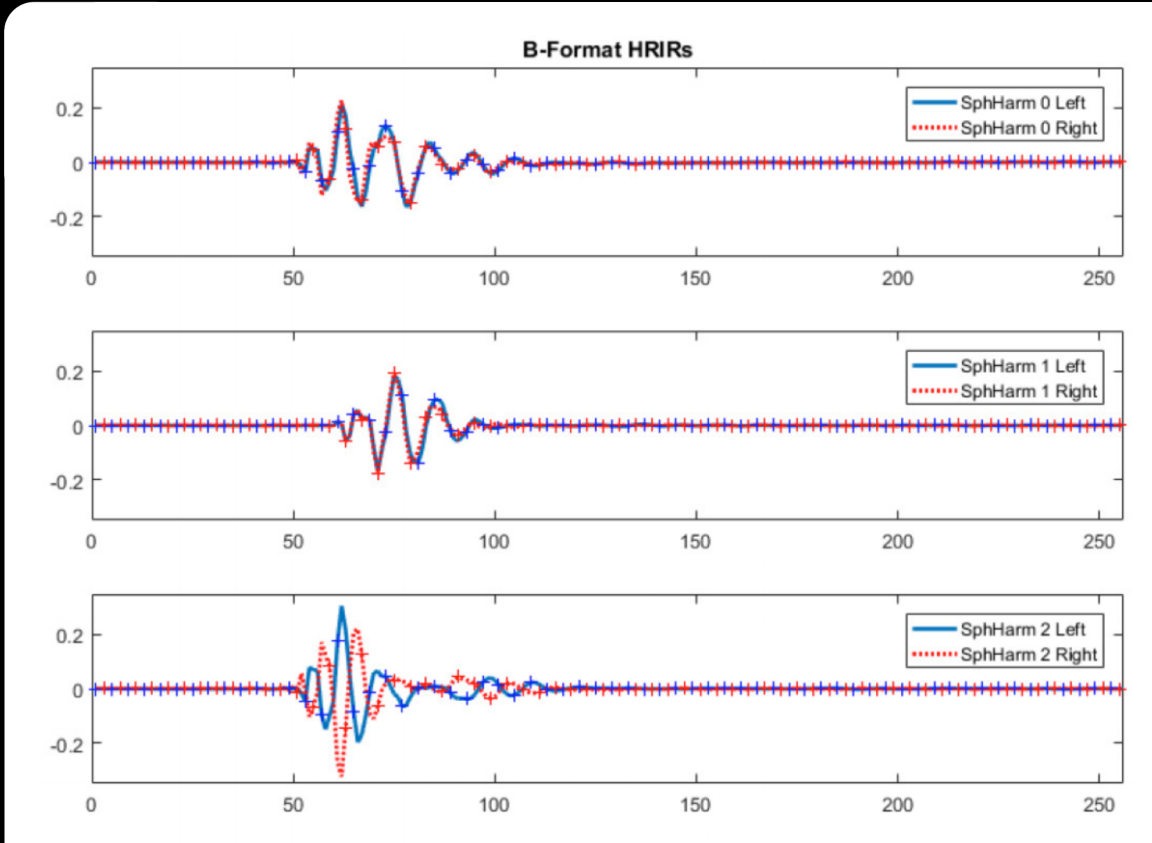


regular | in-phase | max-r_e

Toplap Barcelona

Binaural FIR Filters in Ambisonic Format

- Combine HRIRs into Ambisonic *B-Format* FIR filters
  - Reduce 100s of HRIR filters to 8 FIR filters in 1st order for *entire* 3D sphere!
  - Set of 4 FIR filters per ear
  - 1st filter is omnidirectional pressure
    - i.e. spherical harmonic 0
  - 3 following filters describe binaural effect in x, y and z dimension respectively
    - i.e. spherical harmonic 1, 2 and 3



Note: image is for a horizontal 2D setup, hence only 3 channels total!