Spatial Audio Reproduction: Towards Individualized Binaural Sound

Bill Gardner

Wave Arts, Inc.

Introduction

- Basic technologies for recording and reproducing audio are mature
- Compact disc (CD) reproduces audio signals at the limits of human perception
- Spatial audio reproduction remains a challenge

Overview

- Spatial audio reproduction
- Binaural audio via headphones
- Failed to live up to promise
- What can be done to fix
- Other spatial reproduction technologies

Binaural audio

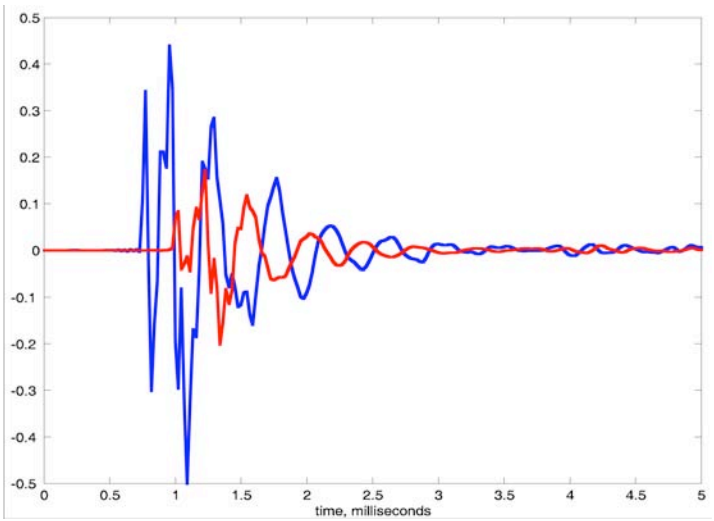
- Recording or reproducing signals at the ears
- Binaural audio recorded and reproduced at your own ears is stunningly realistic
- Signals at two eardrums completely define auditory experience
- Goal is headphone system which is perceptually indistinguishable from real listening

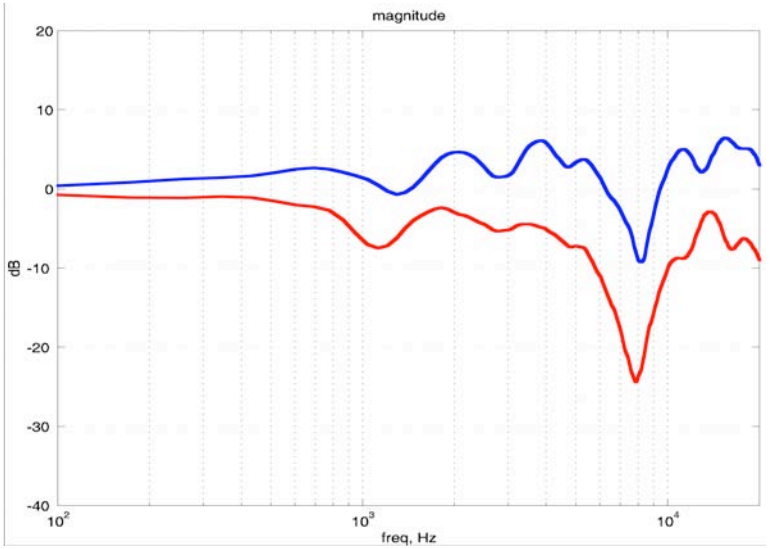
Spatial hearing

- Sound interacts with torso, head, external ears and arrives at the two ear canals
- Localization cues:
  - Time
  - Amplitude
  - Spectrum
- Dynamic cues

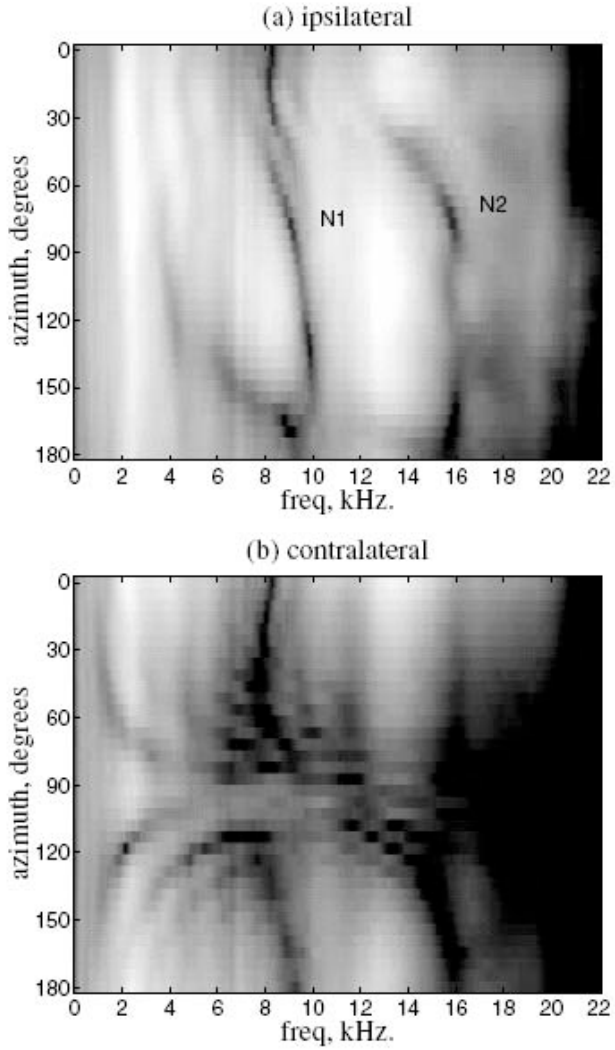
Head-Related Transfer Function (HRTF)

- Describes transformation of sound from free-field to ear



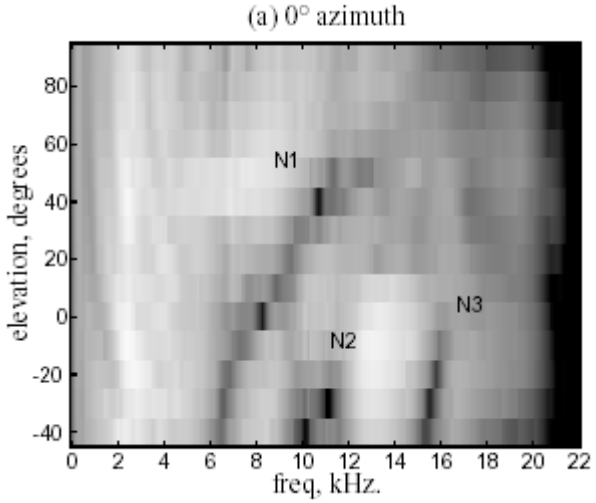


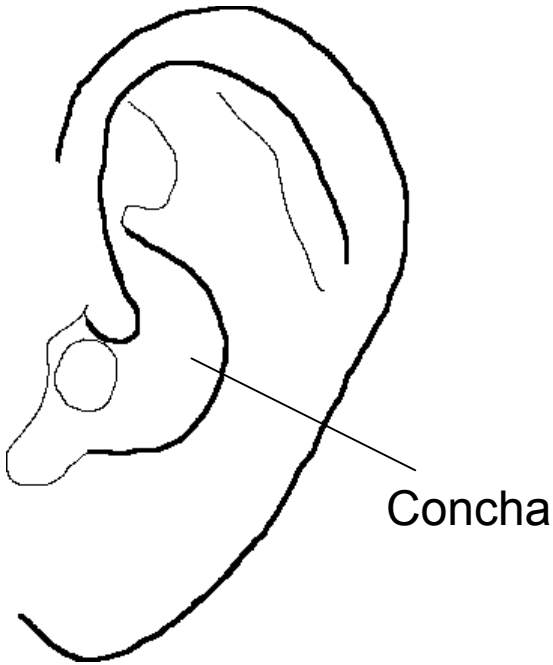
HRTF azimuth dependency



wave arts

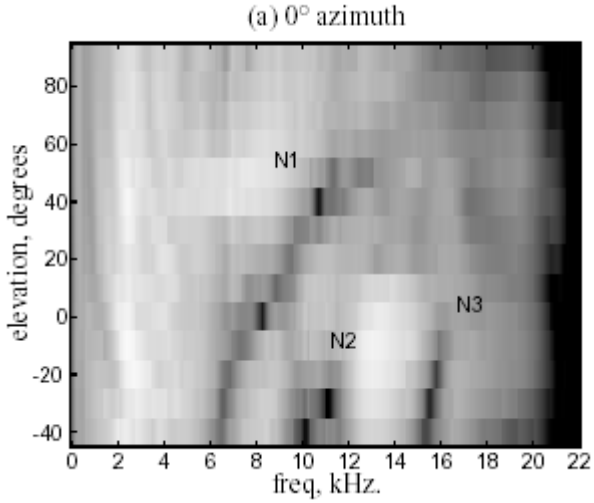
HRTF elevation dependency

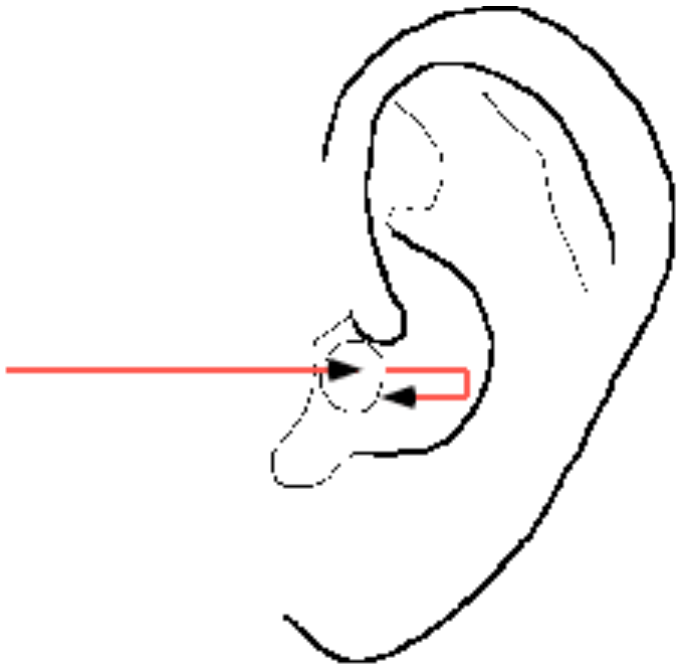




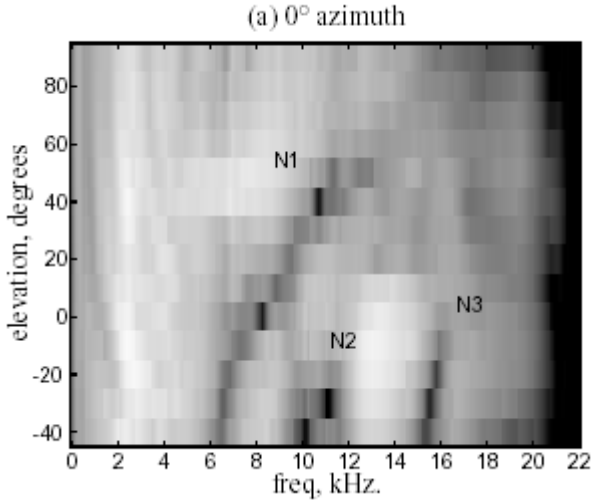
wave arts

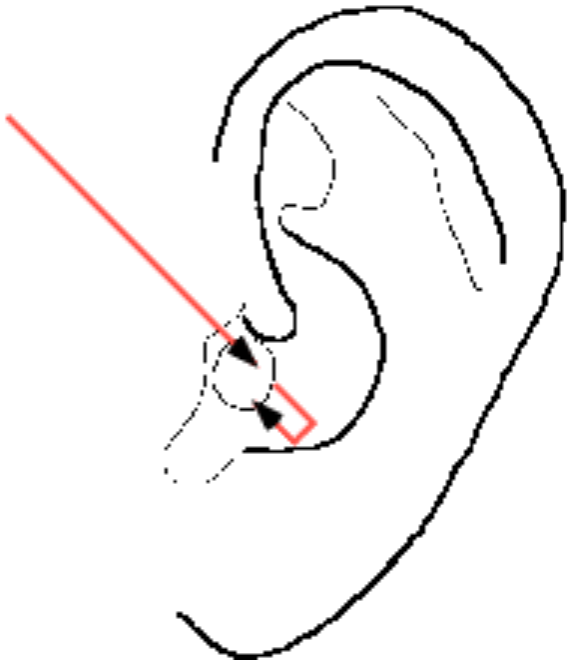
Concha reflection at 0 deg elevation





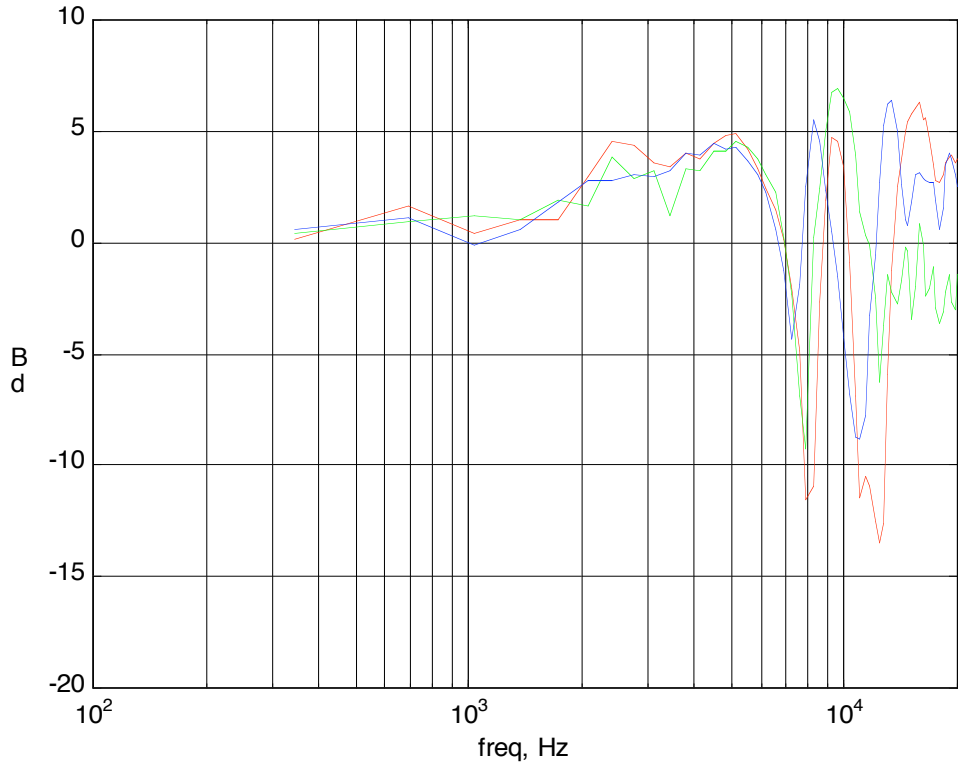
Concha reflection at 45 deg elevation





HRTFs from three humans

- External ears vary between humans
- High frequency cues are idiosyncratic

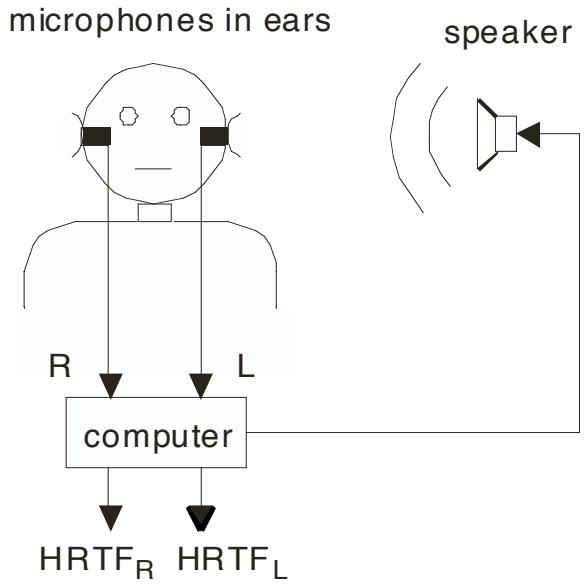


HRTFs provided by IRCAM

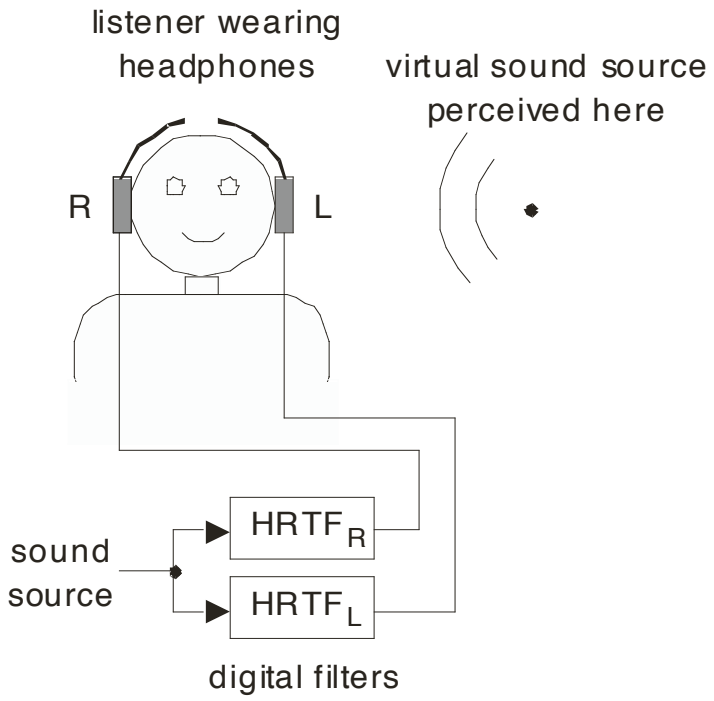
wave arts

Binaural synthesis

- Measurement



- Synthesis



Perceptual validation

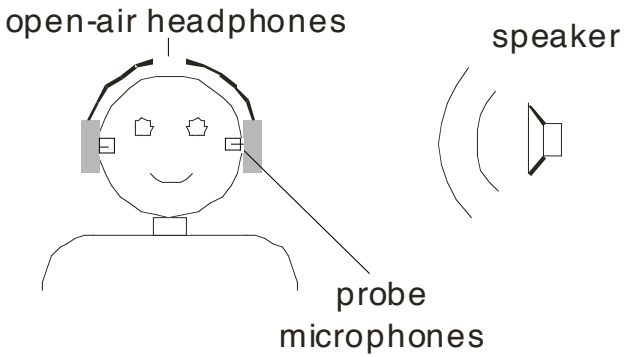
- Localization experiments
- Using individualized HRTFs
  - Localization performance unchanged from real listening (Wightman and Kistler, 1989)
- Using non-individualized HRTFs
  - Elevation errors
  - Front/back reversals (Wenzel et al., 1993)
  - Lack of externalization (sounds are perceived near head or inside head)

Improving externalization

- Dynamic head tracking
- Artificial reverberation

Real/virtual discrimination tests

- Apparatus allows direct comparison of real and virtual stimuli (Hartmann and Wittenberg, 1996)
- Virtual stimuli can be modified to determine limits of perception
- Individualized HRTFs needed for externalization

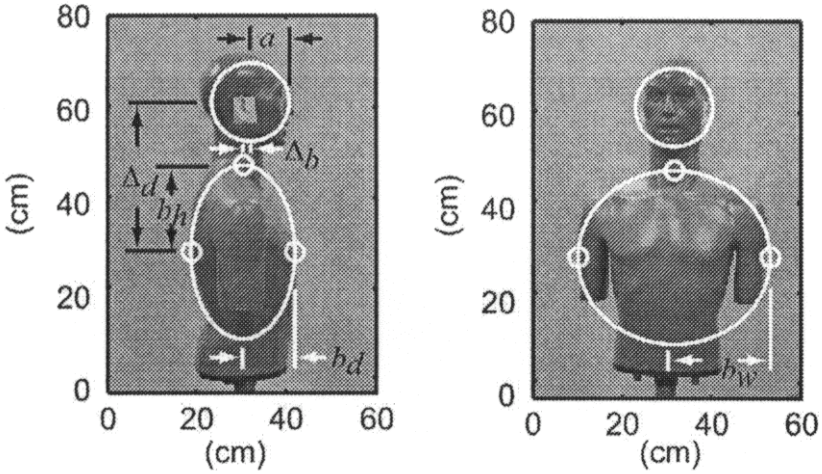


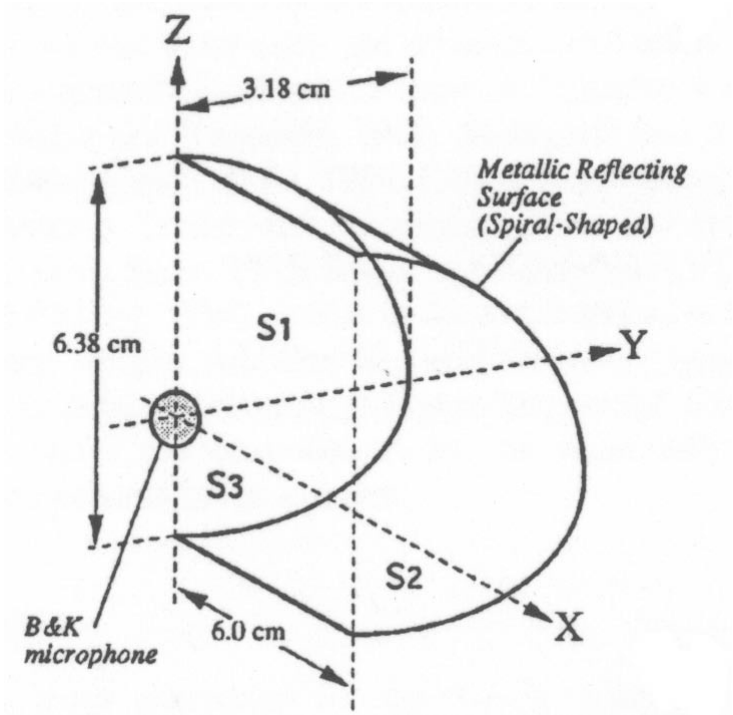
How to obtain individualized HRTFs?

- Acoustic measurement impractical
- Would like fast, practical method
  - Measurements of listener
  - Based on image of head and ears
  - Calibration procedure

Simplified head models

- Model torso and head using spheroids (Algazi and Duda, 2002)
- Model ear using simple shapes (Lopez-Poveda and Meddis, 1996)
- Models inaccurate at high frequencies



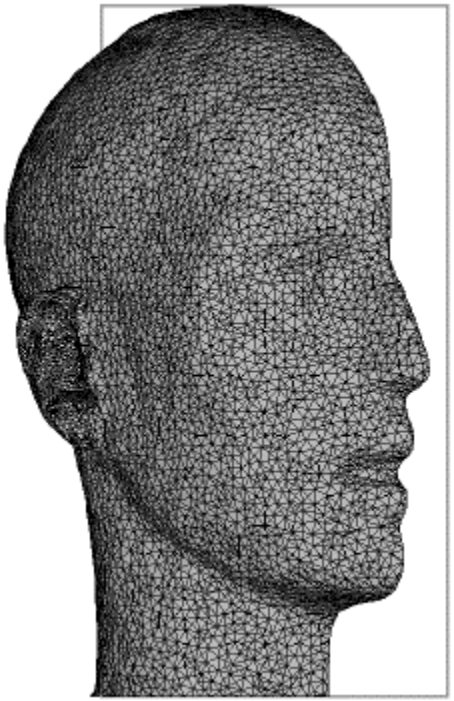


Statistical analysis

- Analyze sets of HRTFs to tease out low dimensional models
- PCA analysis of HRTFs (Kistler and Wightman, 1992)
- 5 principal components sufficient to reproduce individualized HRTFs
- Parameters need to be determined for each location and listener

Computational modeling

- Boundary element method (Kahana, 1998)
- 15,000 element model matches acoustical measurements accurately to 15 kHz
- Scanning head is difficult

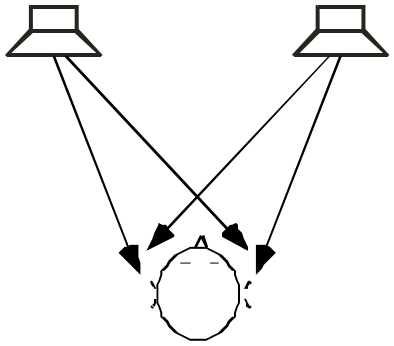


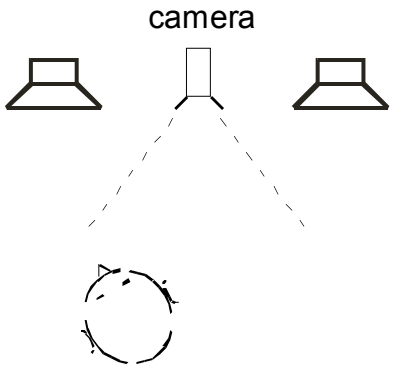
Deformable head model?

- Start with refined head mesh model
- Obtain subject's measurements using computer vision techniques
- Morph mesh model to fit measurements
- Calculate HRTFs using computational acoustics

Crosstalk-cancelled audio

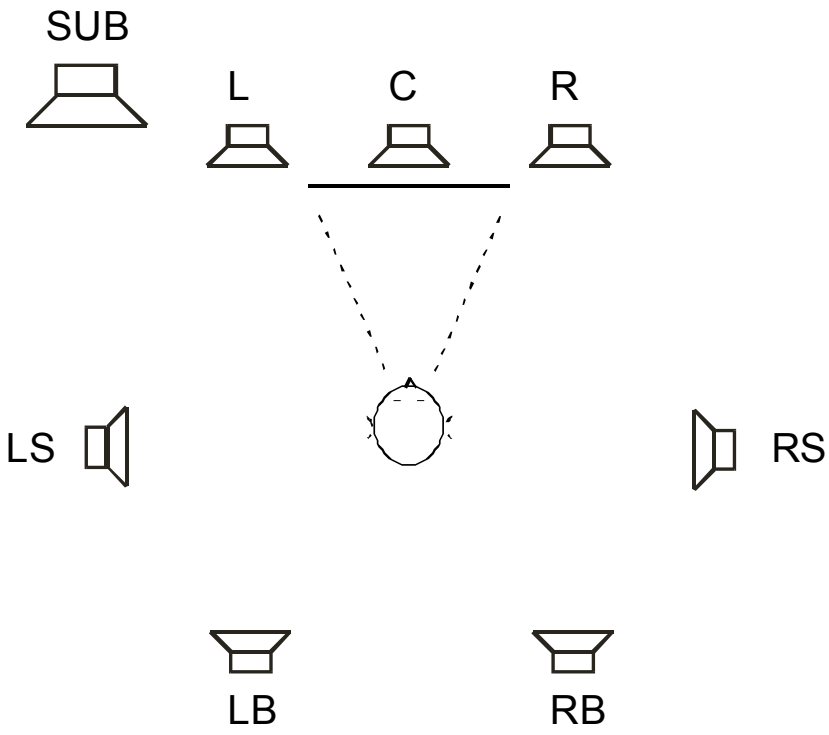
- Deliver binaural audio to listener via stereo loudspeakers
- Requires pre-processing to cancel crosstalk
- Works at low frequencies
- Listener must be fixed or tracked





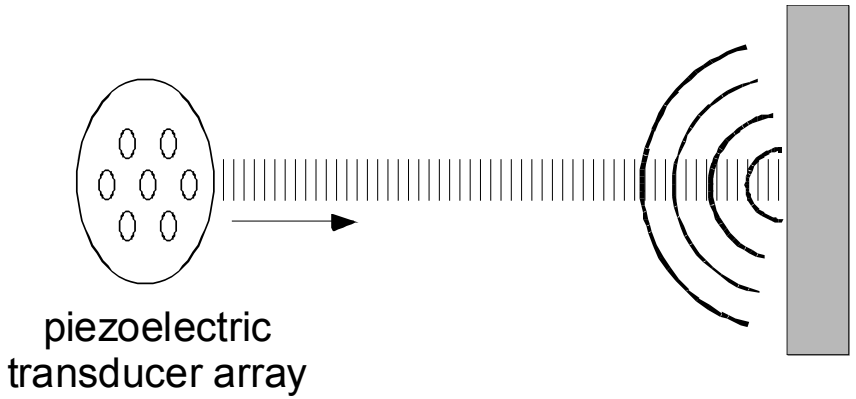
Multichannel audio

- 5.1 and 7.1 surround systems developed for movie theatres are now widely used in homes
- Emphasis on frontal localization
- Doesn't require individualization



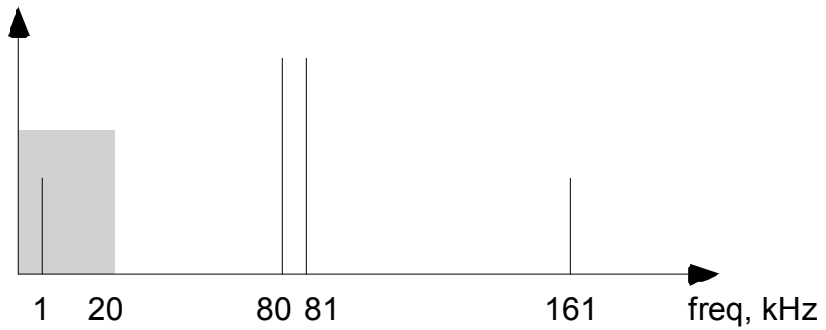
Ultrasonic audio

- Highly directional ultrasonic beam creates audible sound (Pompei, 1999)
- Based on non-linear properties of air
- Tones inter-modulate at high intensity



Ultrasonic audio

- 80 kHz and 81 kHz tones will produce audible 1 kHz difference tone
- Requires pre-processing to reduce distortion
- Acceptable fidelity, poor low frequency response



Summary

- Discussed spatial audio reproduction techniques
- Binaural audio has promise for audio reproduction indistinguishable from real listening
- Requires individualized playback
- Proposed method to individualize playback for a particular listener