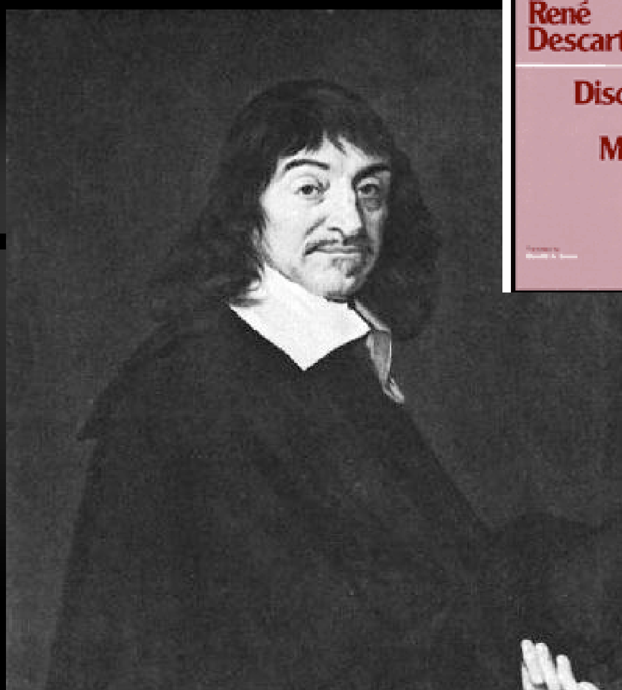
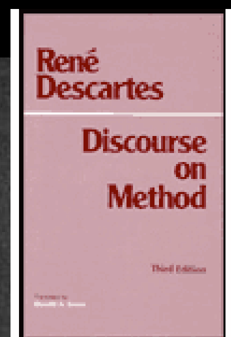
Virtual Acoustic Reflections

Gregory H. Wakefield
Interactive Systems Lab
CSE, PAT, OTO, EE:S

Cartesian Daemon

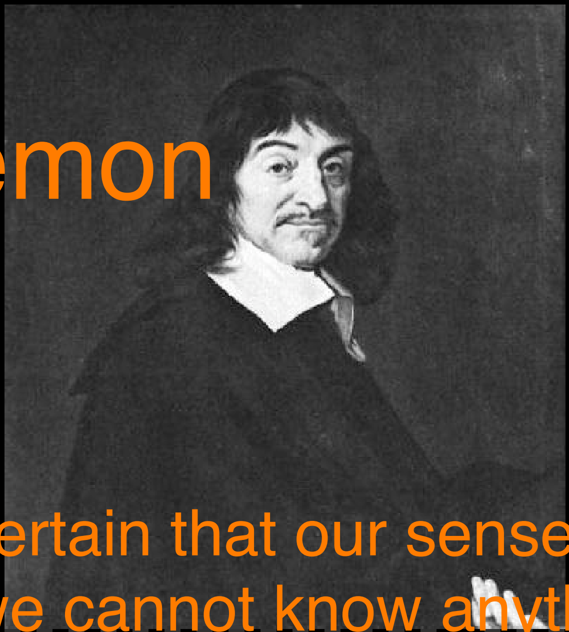




"I will suppose. . .that some evil demon of the utmost power and cunning has employed all his energies to deceive me." (17th century version)

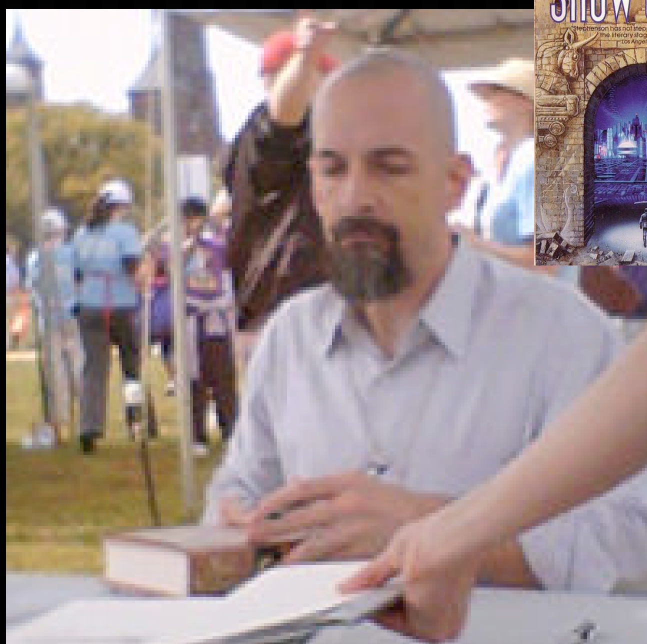
"I will suppose…that some mad scientist of the utmost power and cunning has placed my brain in a vat, and is cleverly stimulating my brain to deceive me otherwise." (20th century version)

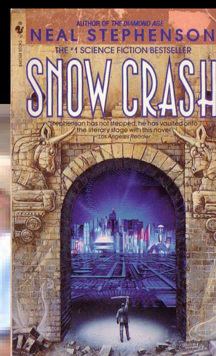
Cartesian Daemon



- If we cannot be certain that our senses do not deceive us, then we cannot know anything with certainty.
- We cannot know whether or not our senses deceive us.
- Therefore, we cannot know anything about the world.

The latest Cartesian Daemon



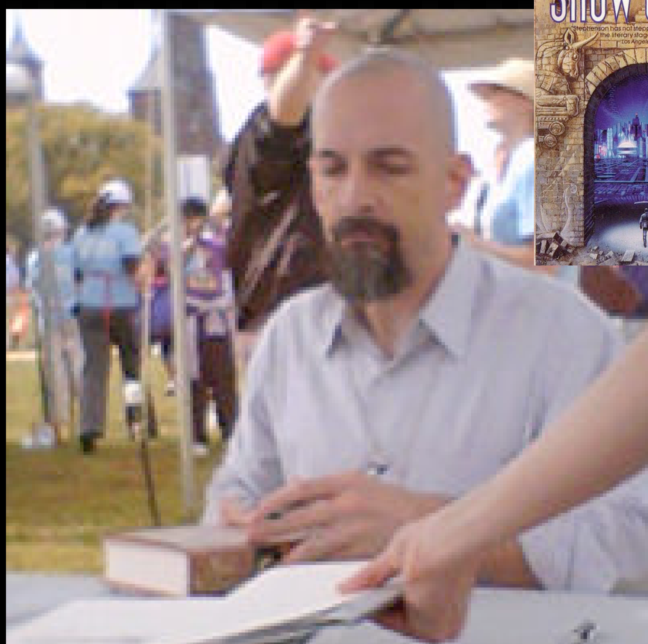


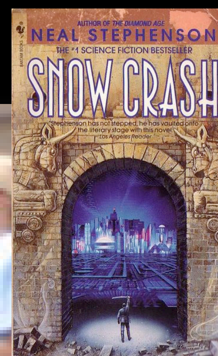
"I will suppose. . .that some evil demon of the utmost power and cunning has employed all his energies to deceive me."

"I will suppose…that some mad scientist of the utmost power and cunning has placed my brain in a vat, and is cleverly stimulating my brain to deceive me otherwise."

"I will suppose…that Bill Gates has placed me in a *metaverse* and has employed all his minions to deceive me."

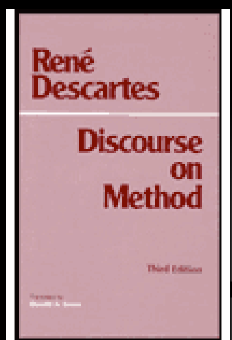
Neal Stephenson (1959 -)



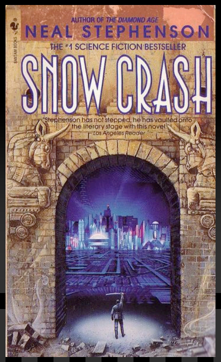


- Post-cyberpunk writer (see William Gibson)
- *Snow Crash (1992)*
    - Metaverse | Reality
    - Avatar embodiment

Well-engineered Games



. . .



. . .

- Epistemology
  - How does the world inform us of the world?
  - How do we “know what we know”?
  - Setting design criteria for VR, extended-reality, augmented reality, and interactive systems

Well-engineered Games

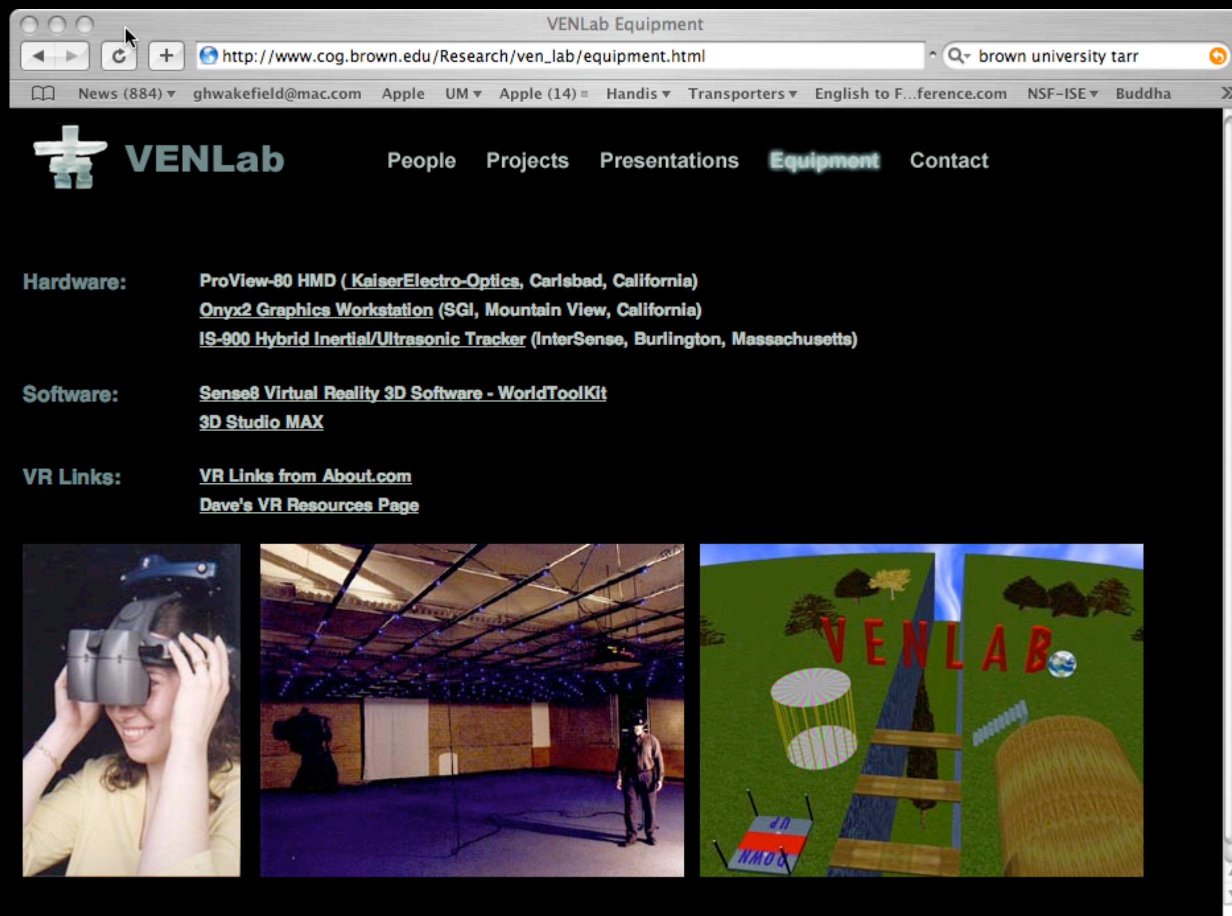
PERCEPTUAL PLUMBING

- Epistemology
  - How does the world inform us of the world?
  - How do we “know what we know”?
  - Setting design criteria for VR, extended-reality, augmented reality, and interactive systems

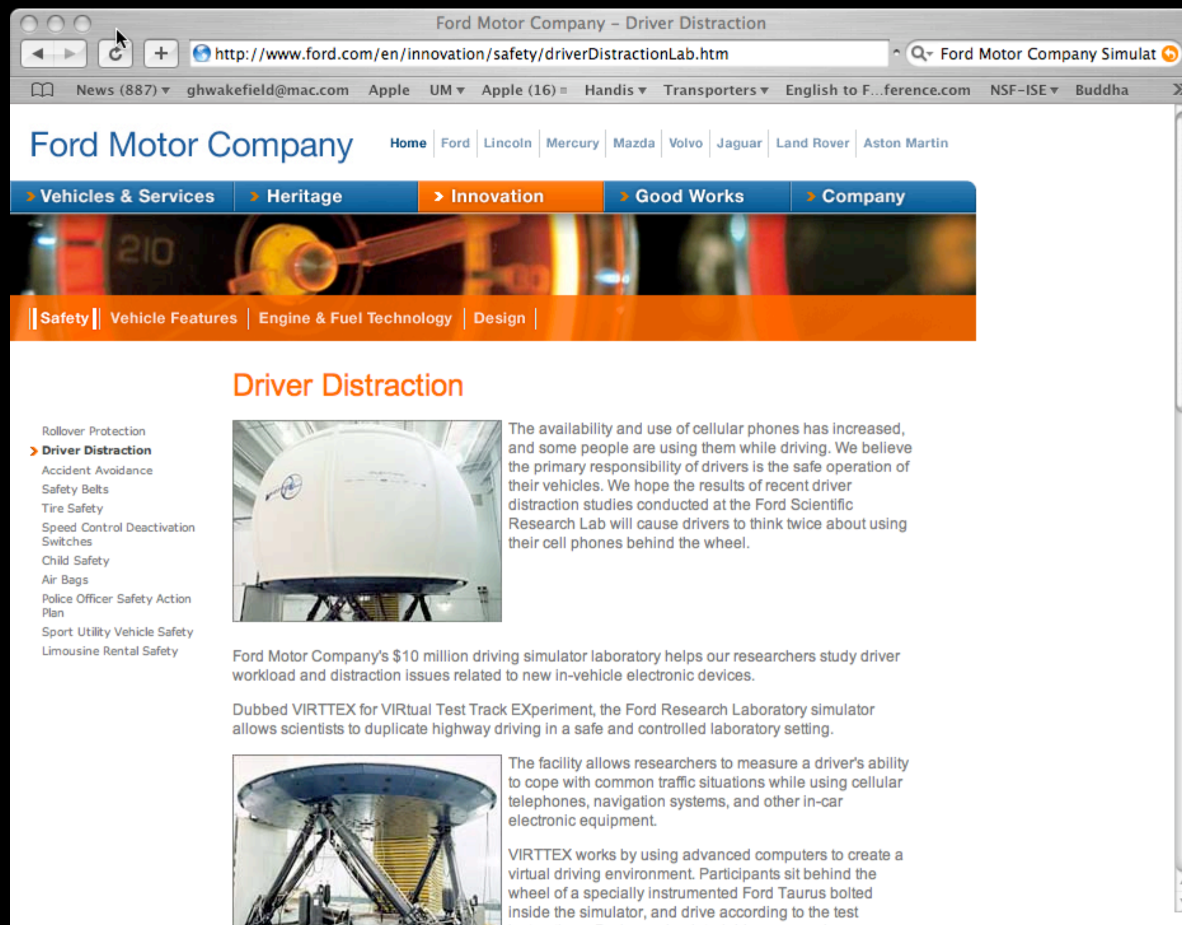
Rest of today…

- Examples of perceptual plumbers
  - Dominated by visual modality
- Auditory plumbing as a discipline
  - Modeling the listener in the environment
  - First-generation technologies
  - Current work
  - Caveats

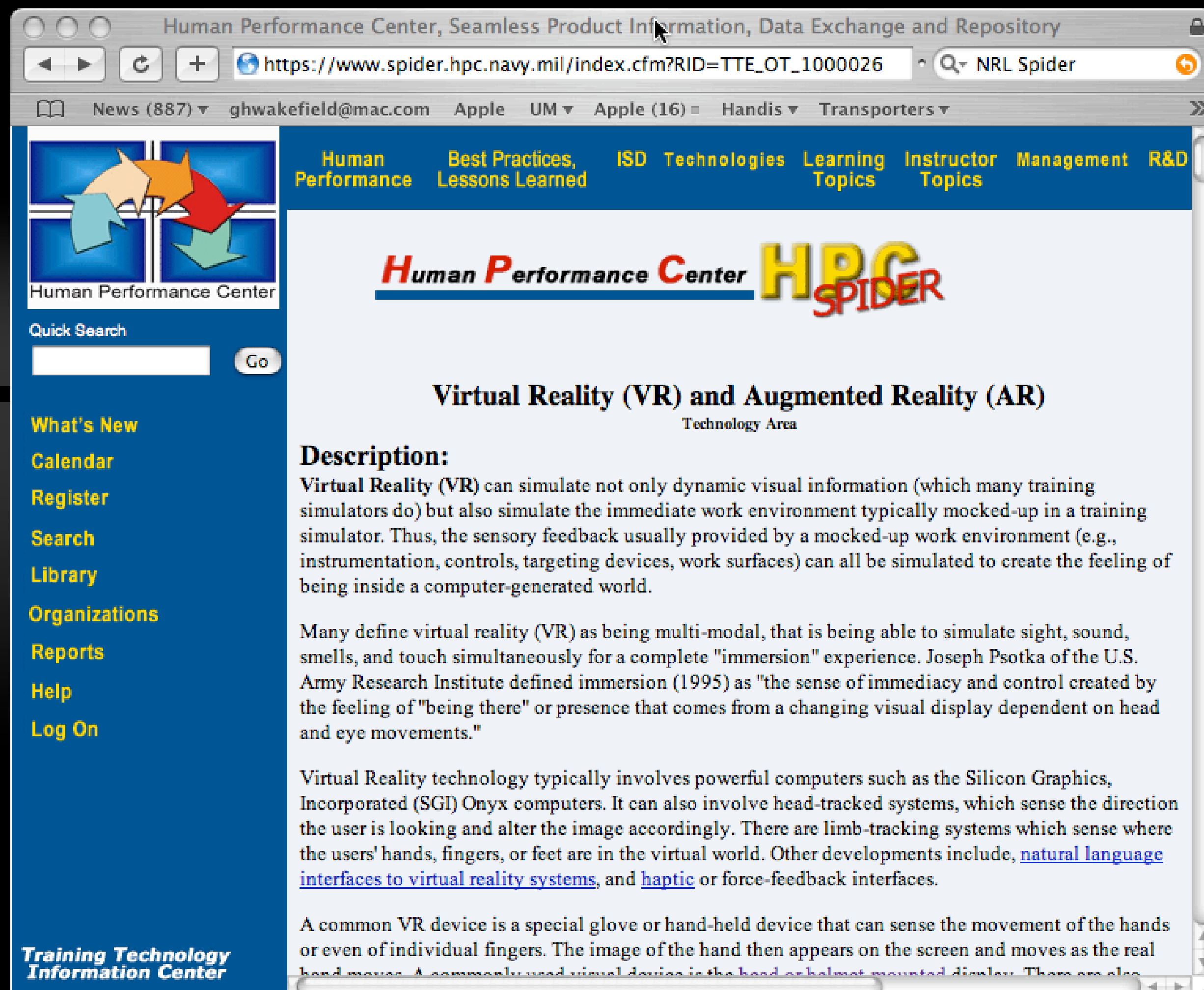
VENLab (Brown University)



Driver Distractor (Ford Motor Co.)

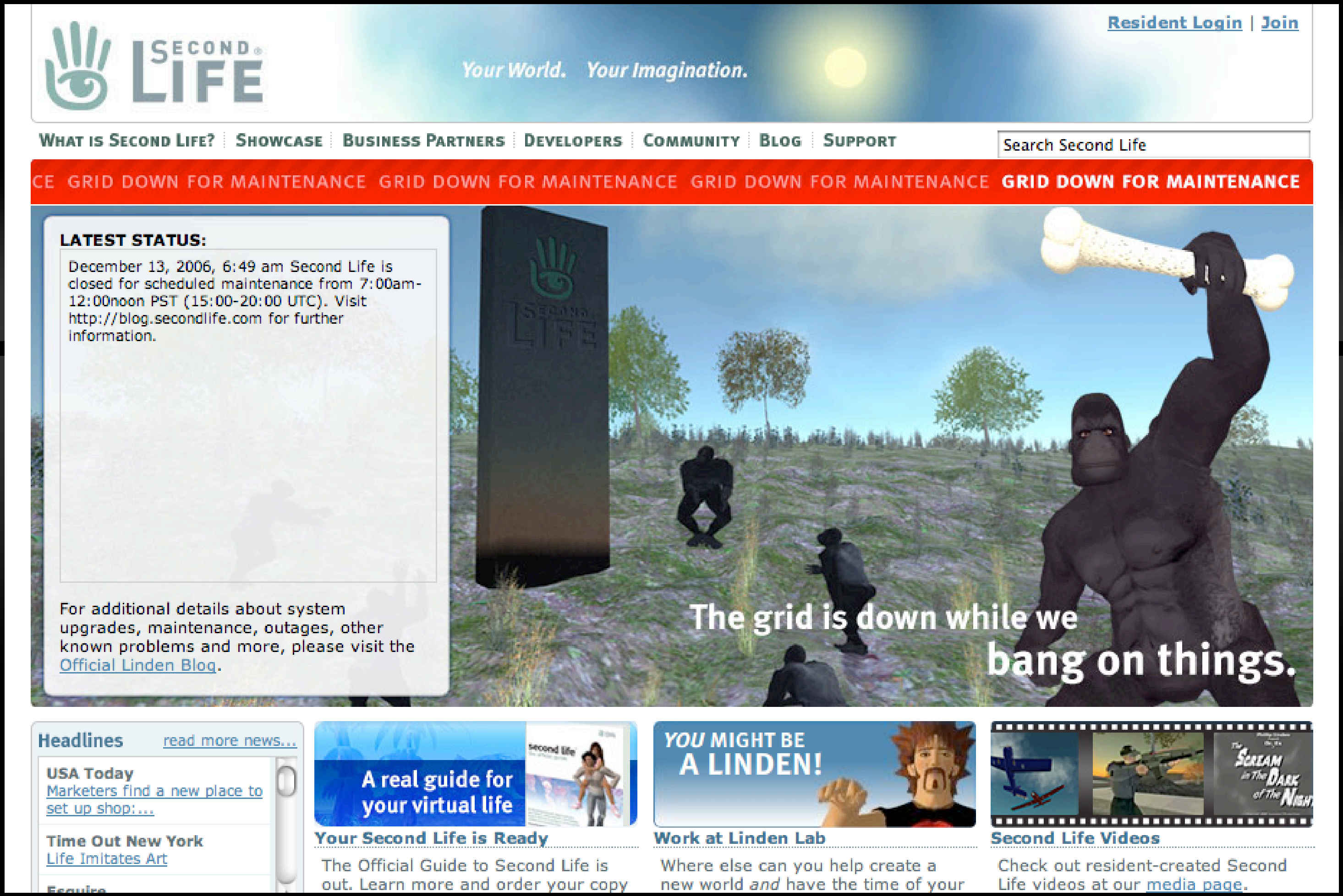


NRL/DoD



Human Performance Center, Seamless Product Information, Data Exchange and Repository
https://www.spider.hpc.navy.mil/index.cfm?RID=TTE_OT_1000026
NRL Spider
News (887) ghwakefield@mac.com Apple UM Apple (16) Handis Transporters
Human Performance Center
Human Performance
Best Practices, Lessons Learned
ISD
Technologies
Learning Topics
Instructor Topics
Management
R&D
Quick Search
Go
What's New
Calendar
Register
Search
Library
Organizations
Reports
Help
Log On
Training Technology Information Center
Human Performance Center
HPC SPIDER
Virtual Reality (VR) and Augmented Reality (AR)
Technology Area
Description:
Virtual Reality (VR) can simulate not only dynamic visual information (which many training simulators do) but also simulate the immediate work environment typically mocked-up in a training simulator. Thus, the sensory feedback usually provided by a mocked-up work environment (e.g., instrumentation, controls, targeting devices, work surfaces) can all be simulated to create the feeling of being inside a computer-generated world.
Many define virtual reality (VR) as being multi-modal, that is being able to simulate sight, sound, smells, and touch simultaneously for a complete "immersion" experience. Joseph Psotka of the U.S. Army Research Institute defined immersion (1995) as "the sense of immediacy and control created by the feeling of "being there" or presence that comes from a changing visual display dependent on head and eye movements."
Virtual Reality technology typically involves powerful computers such as the Silicon Graphics, Incorporated (SGI) Onyx computers. It can also involve head-tracked systems, which sense the direction the user is looking and alter the image accordingly. There are limb-tracking systems which sense where the users' hands, fingers, or feet are in the virtual world. Other developments include, natural language interfaces to virtual reality systems, and haptic or force-feedback interfaces.
A common VR device is a special glove or hand-held device that can sense the movement of the hands or even of individual fingers. The image of the hand then appears on the screen and moves as the real

Second Life



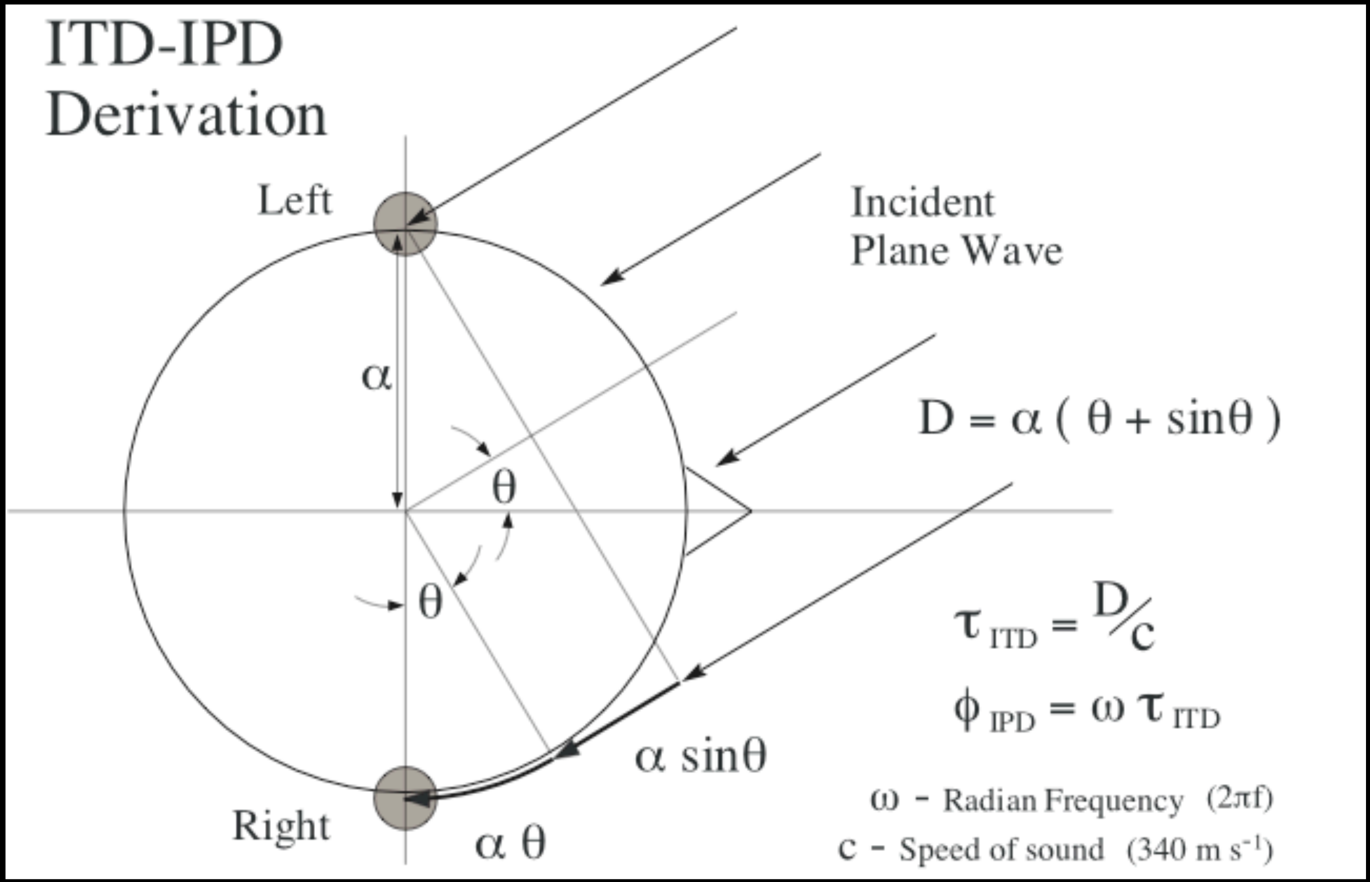
Resident Login | Join
SECOND LIFE
Your World. Your Imagination.
WHAT IS SECOND LIFE? | SHOWCASE | BUSINESS PARTNERS | DEVELOPERS | COMMUNITY | BLOG | SUPPORT
Search Second Life
CE GRID DOWN FOR MAINTENANCE GRID DOWN FOR MAINTENANCE GRID DOWN FOR MAINTENANCE GRID DOWN FOR MAINTENANCE
LATEST STATUS:
December 13, 2006, 6:49 am Second Life is closed for scheduled maintenance from 7:00am-12:00noon PST (15:00-20:00 UTC). Visit http://blog.secondlife.com for further information.
For additional details about system upgrades, maintenance, outages, other known problems and more, please visit the Official Linden Blog.
The grid is down while we bang on things.
Headlines read more news...
USA Today
Marketers find a new place to set up shop:...
Time Out New York
Life Imitates Art
A real guide for your virtual life
Your Second Life is Ready
The Official Guide to Second Life is out. Learn more and order your copy
YOU MIGHT BE A LINDEN!
Work at Linden Lab
Where else can you help create a new world and have the time of your
Second Life Videos
Check out resident-created Second Life videos at our media page.

Commercialization

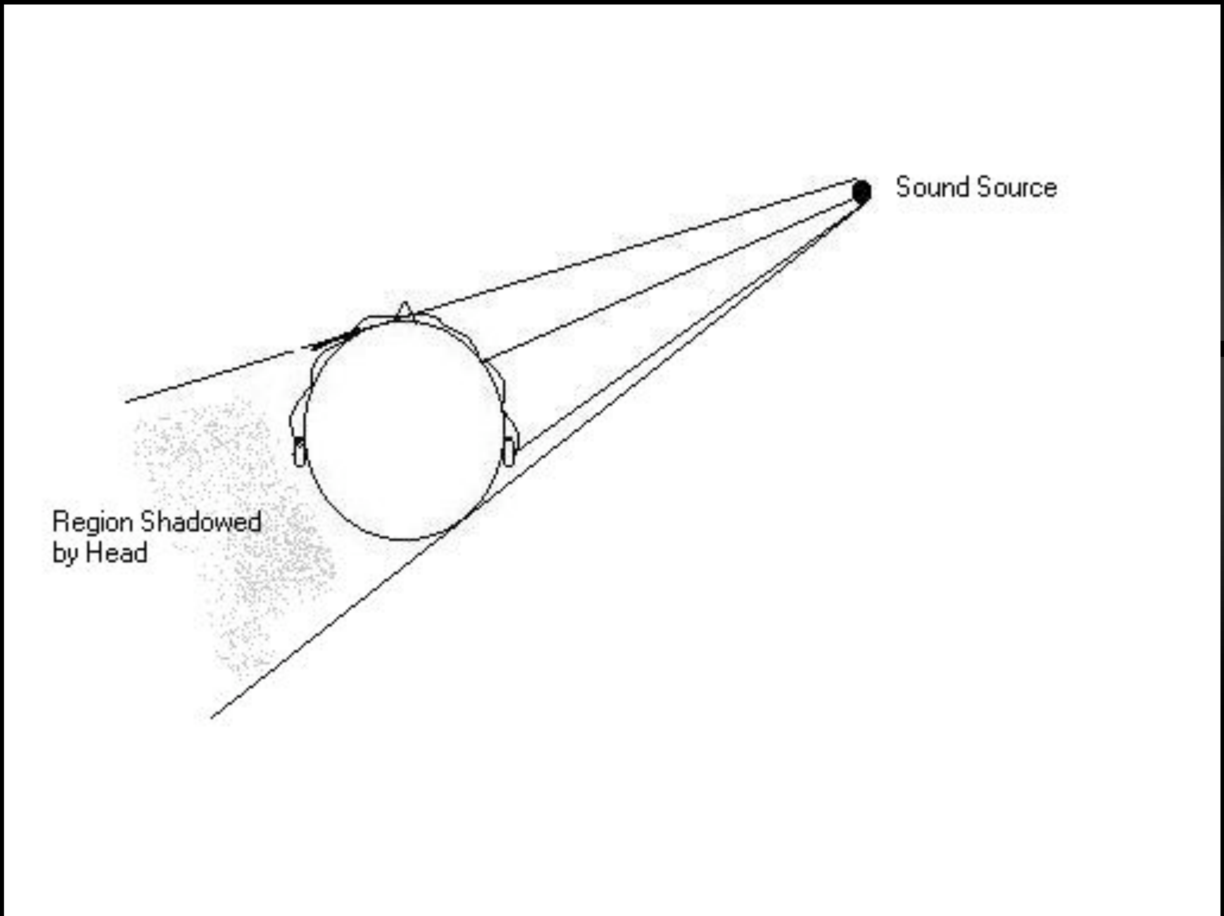
What about the world of sound?

- Physics of Propagation
  - Line-of-sight vs. bending around corners
  - Sources + Transmission Medium (reflections)
- Perception of Propagation
  - Vision: Object as reflectance, source ignored
  - Auditory: Object as source, reflectance ignored
- Perception of Object
  - Vision: “Time independence + Spatial dependence”
  - Audition: “Spatial independence + Time dependence”

Modeling the Listener in the Environment



Modeling the Listener in the Environment



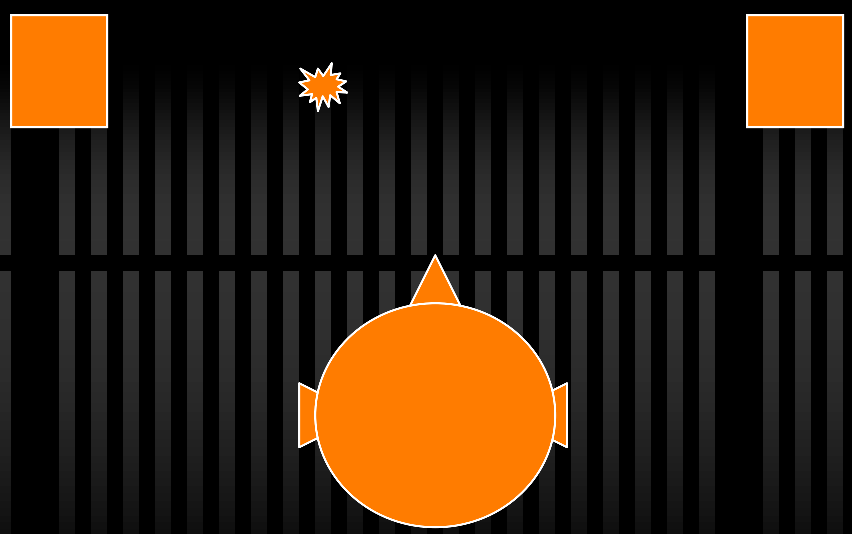
Modeling the Listener in the Environment

- Physics: Lord Rayleigh (1842-1919)
  - Difference in arrival times (Interaural Time Differences - ITD’s)
    - Transient wave
    - Continuous wave - spatial aliasing
  - Level difference (Interaural Level Differences-ILD’s)
    - Wavelength dependence
  - Duplex theory of auditory localization
    - Below 1-2 kHz, use ITD
    - Above 2-3 kHz, use ILD

First-Generation Auditory Plumbing

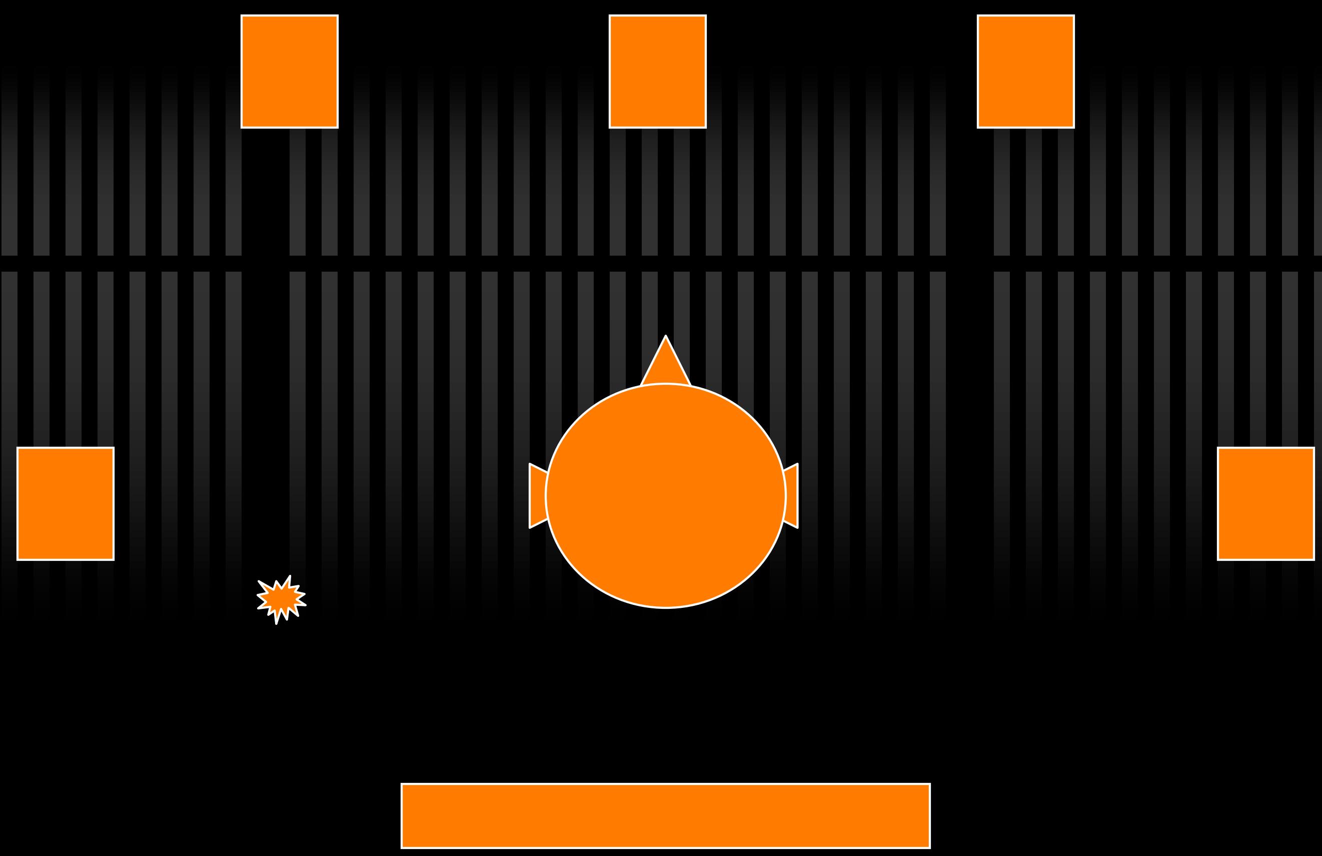
- Stereophony
  - stereos = "solid" and phōnē = "sound"
  - Adler (1881)
  - Fletcher (Bell Labs, 1930's)
  - Term coined: Western Electric (1937)
  - Publicly demonstrated: Carnegie Hall (1940)
- Recording Techniques
  - X-Y (Intensity stereophony)
  - A-B (Time-of-arrival stereophony)

Playback: Stereo Panning



Physics: Trading intensity and timing differences…

Surround Sound (5.1)



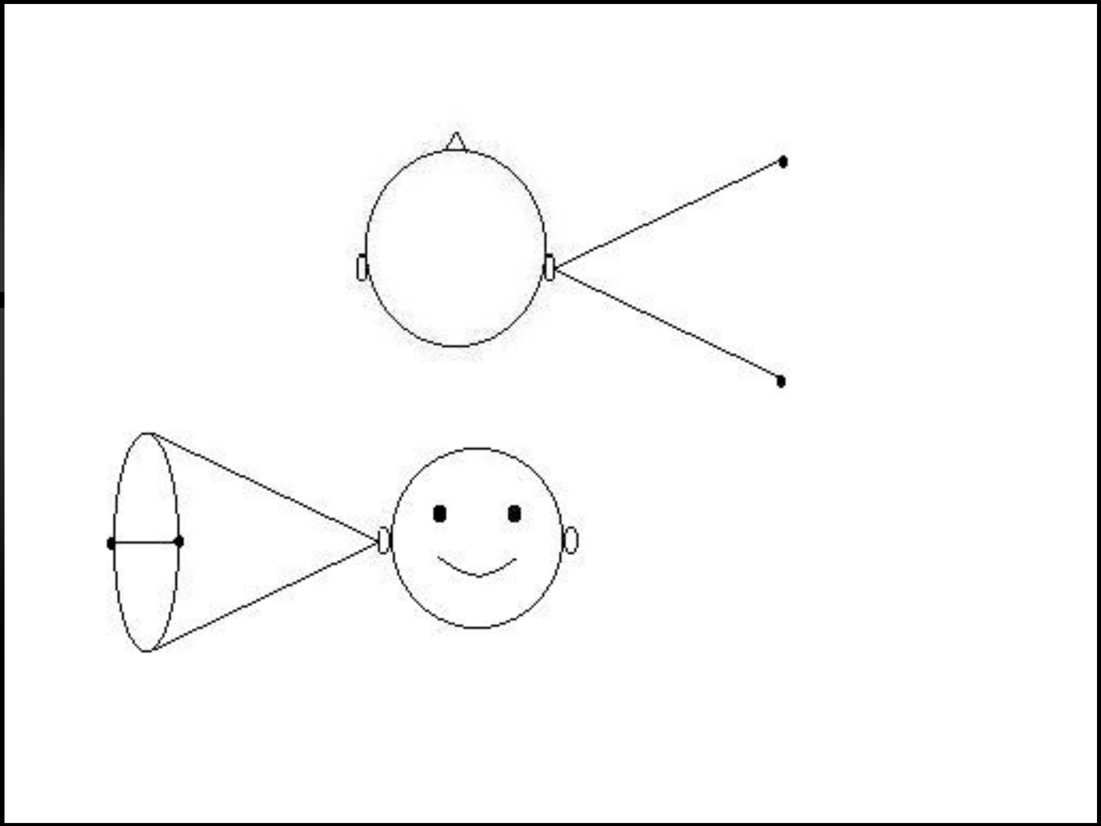
What about 3D Sound?

- Azimuth
  - ITD, ILD cues
  - Basis for stereophony and 5.1 sound
- Elevation
  - ???
- Range
  - ???

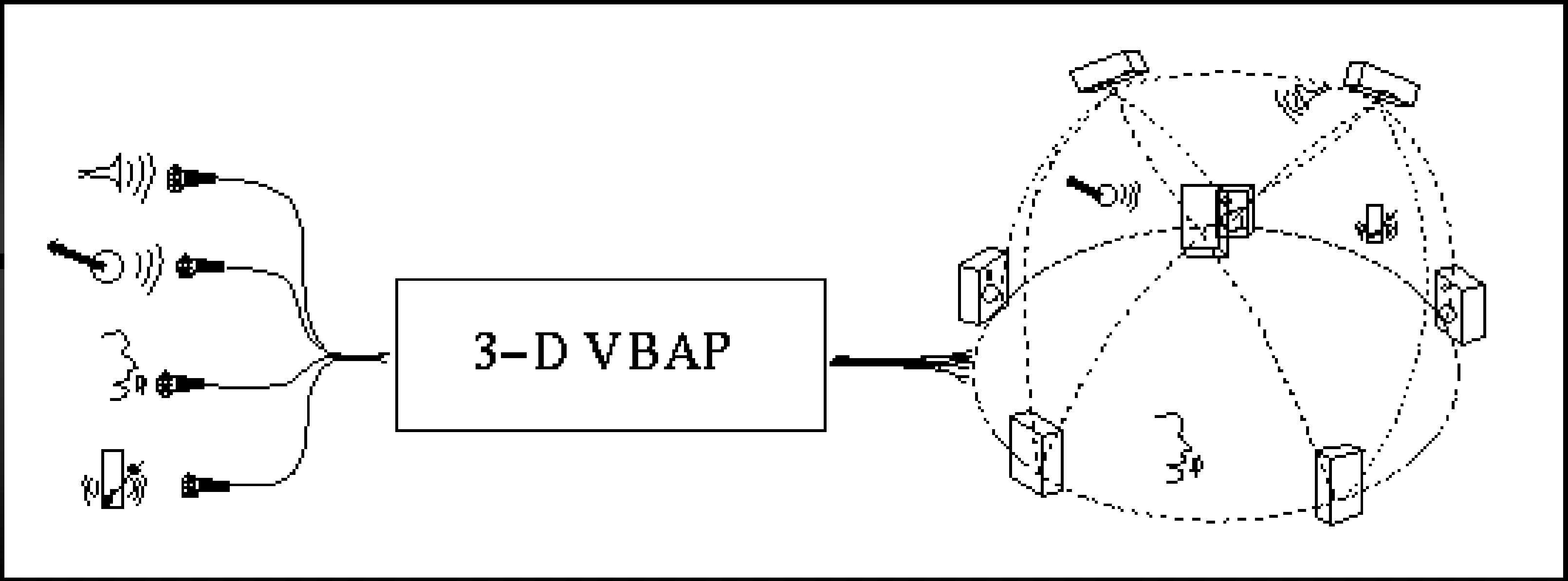
Range: Bigger Fish to Fry

- Helmholtz: Unconscious Inference
- Range perception: very poor
  - Sound is close, near, far away
  - Simulation:
    spectral coloration + reverberation

Elevation: Cone of Confusion



VBAP (Vector Based Amplitude Panning)



Solution: “M.1” in azimuth and elevation.
Ville Pulkki, D. Sc., HUT

3-D Commercialization

- ✓ Spatial diffusers
- ✓ THX
- ✓ VBAP

Costly commercial solutions…

- ✓ VR goggles vs.
- ✓ Large, acoustically balanced room +
  Optimized, balanced loudspeakers +
  Optimized, balanced amplifiers + …
- ✓ And did we bother to talk about the
  stereo hot-seat?

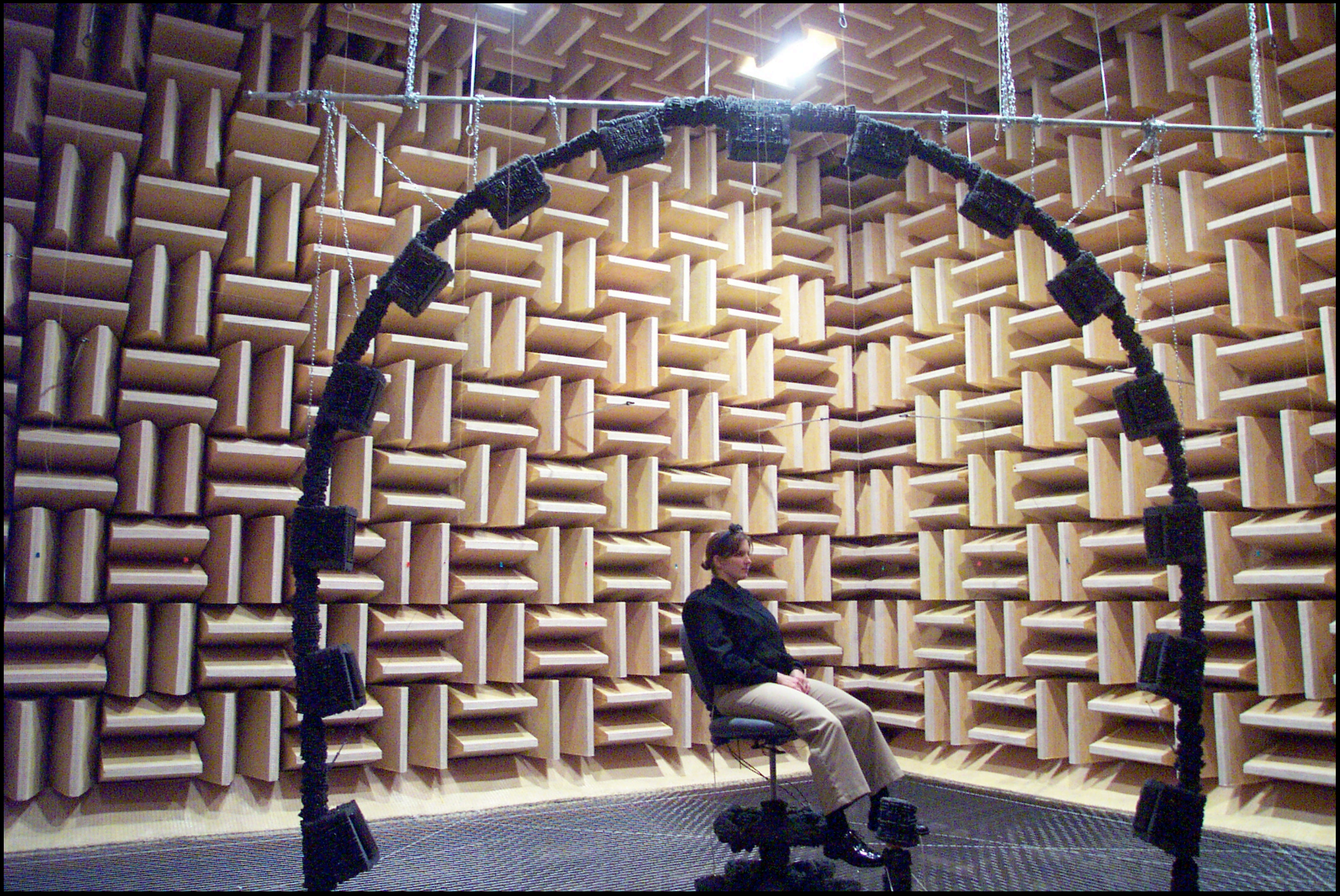
What about headphones?

1. Characterize the spatial radiativity pattern of the listener's head
2. Synthesize sources at desired locations
3. Playback over headphones so that the acoustic signals at the listener's left and right ears are identical to those they would have received in the original listening environment

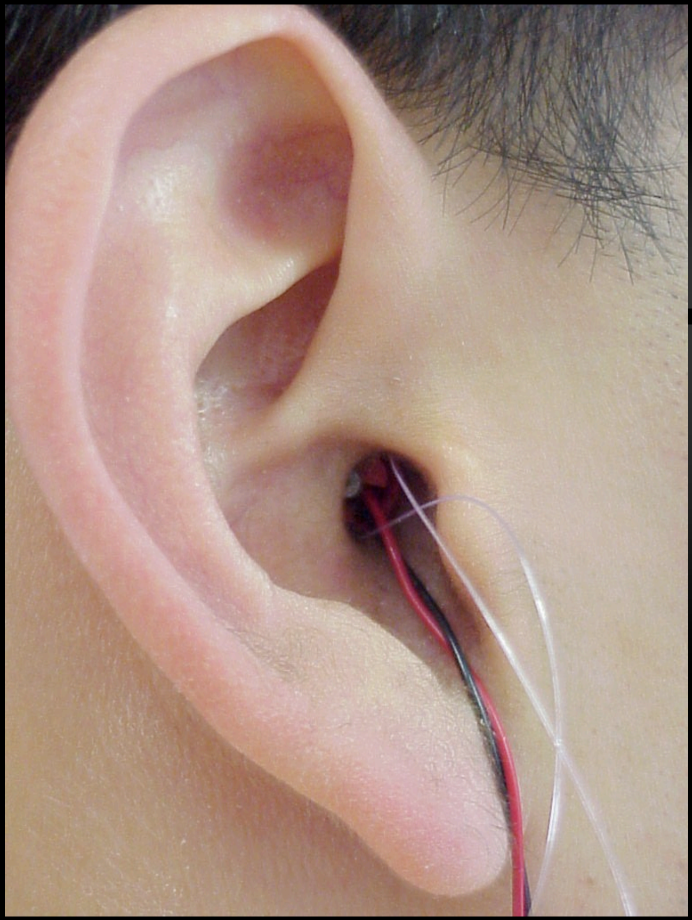
Measurement of Head-Related Transfer Functions

- Research history - short version of the list
  - NASA Ames - Wenzel/ Wightman and Kistler
  - Hartmann
  - Karjalainen (Finland, HUT)
  - Houtgast (Netherlands)
  - Wright-Patterson (Gilkey)
  - MIT/BU (Braida, Durlach, Colburn, Kulkarni, Shinn-Cunningham)
  - Duda and Algazi
  - Middlebrooks (UFl/UM)
  - Wakefield, Adams, Cheng, Runkle, Blommer, Santoro, Buell (UM, NSMRL)
- Equipment and Measurement Issues

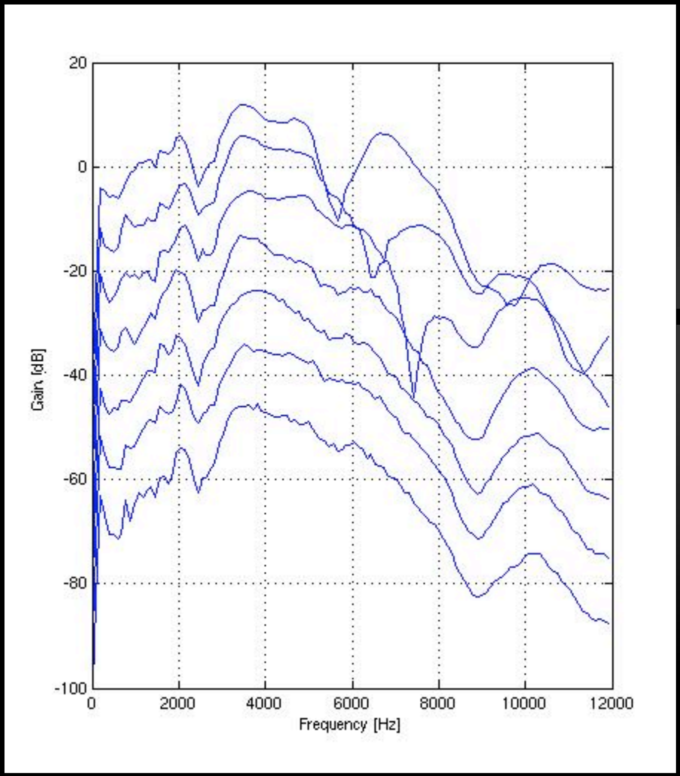
Anechoic Chamber/Loudspeaker System

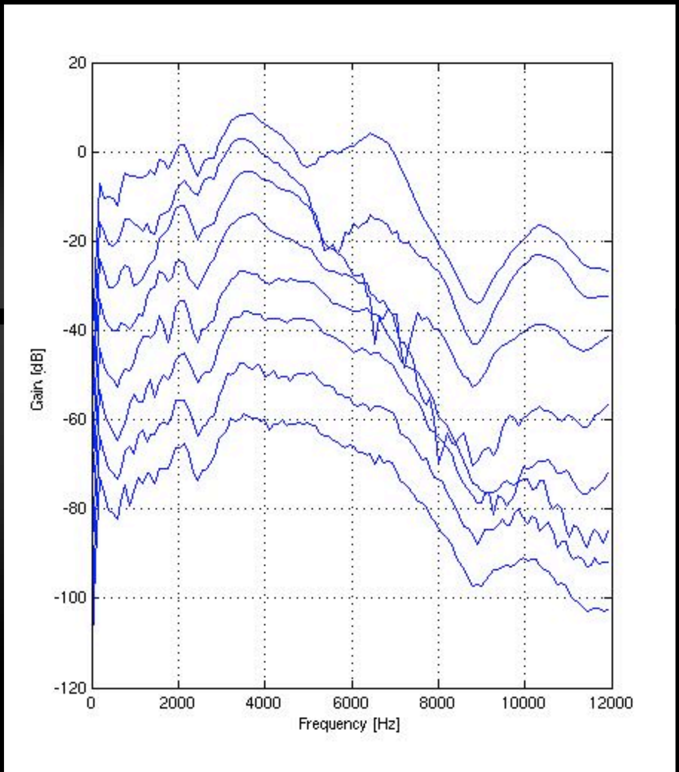


Probe-tube microphones



Results: Right Ear





90 deg (right)

0 deg (midline)

Typical data sets and costs

- 270-500 spatial locations
- Impulse responses: 150-300 samples
- 44100 samples/sec
- Rendering costs: 25 Mflops/source
- Low-order modeling: 2.5 Mflops/source
- PCA/Eigenfilter: 150 Mflops + 0.5 Mflop/source
- MIMO (Adams & Wakefield): 25-150 Mflops + fidelity criterion

Commercial Vendors

- AuSim (US)
  - Convolvotron w/specially-designed HRTFs
- Lake Technology (Australia)
  - Equalizer technologies
- Gaming “Near-field” systems

Want one?

- ✓ Azimuth performance with generic HRTFs: adequate
- ✓ Elevation performance with generic HRTFs: generally poor, particularly front-back confusions

Will there be a better one?

- Individualized HRTFs
  - Performance is never as good as free-field localization in anechoic chamber
  - Measurement error?
  - Rendering issues?
- Roughly 50% of population exhibit front-back confusions *even in free-field listening*
- Why? Role of head motion

Summary: Source Positioning

- Rendering azimuth and elevation of (point) source in anechoic environment
  - Azimuth (left-right)
    - Amplitude panning or timing disparity
    - Robust to differences among listeners
    - Robust to head motion
  - Elevation (including front-back)
    - Spectral coloration
    - Sensitive to differences among listeners
    - Sensitive to head motion

Additional rendering issues

- ✓ Rendering reflections
  - ✓ Computational explosion
    - ✓ Room-acoustic models
    - ✓ Auditory sensitivity spans 100 dB dynamic range: many many many reflections later
    - ✓ Speed of sound (1 ms per foot): challenging filter design problem
  - ✓ Perceptual constraints: 4 regimes (rough!)
    - ✓ Binaural summation (< 1 ms)
    - ✓ Precedence (1-30 ms)
    - ✓ Early late-arrivals (10-100 ms)
    - ✓ Late arrivals (30 ms - 10's of seconds)

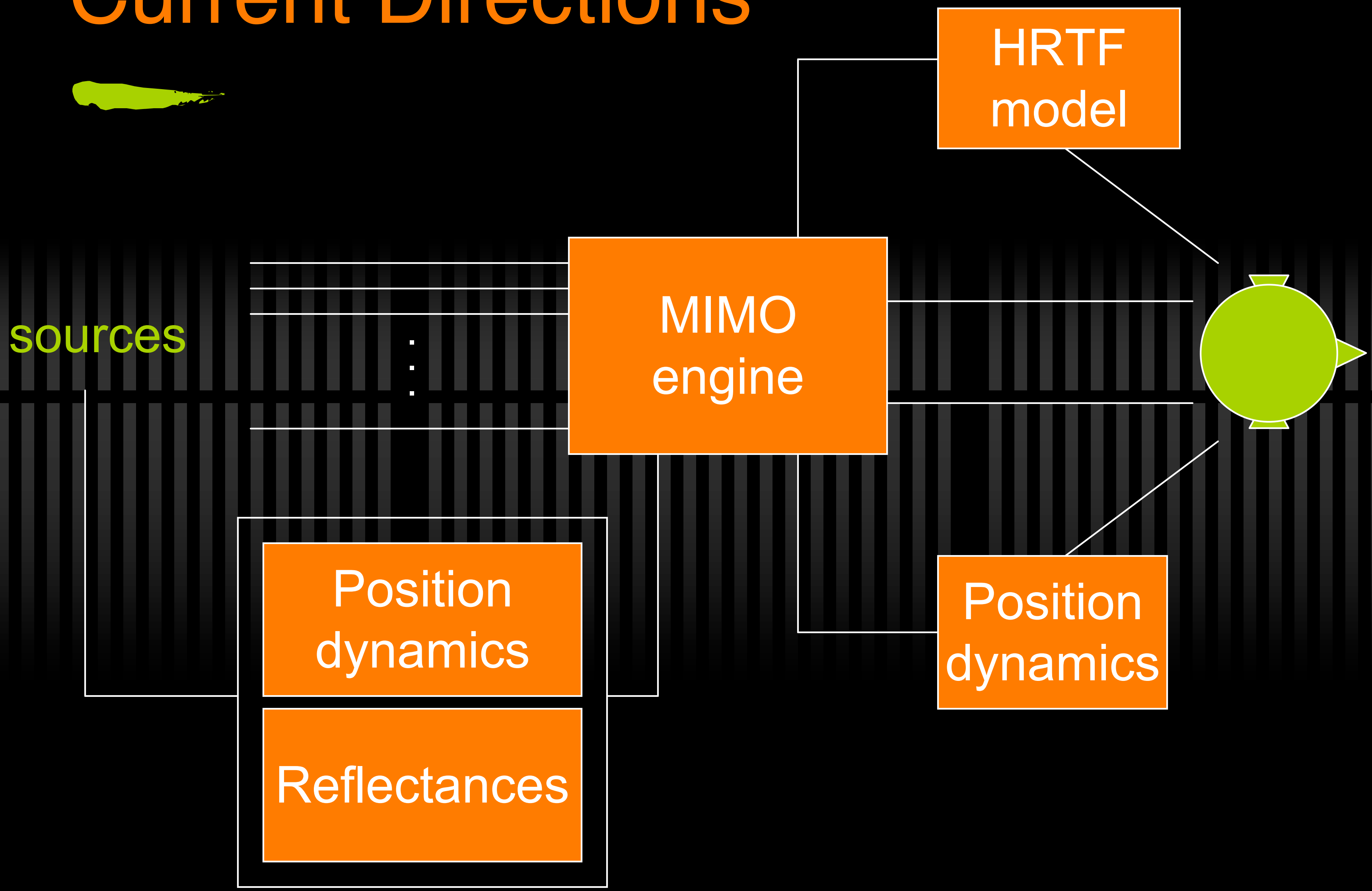
Additional rendering issues

- Rendering motion
  - Head motion
    - Interpolation among HRTFs
  - Source motion
    - Doppler effects
    - Rotational vs. lineal

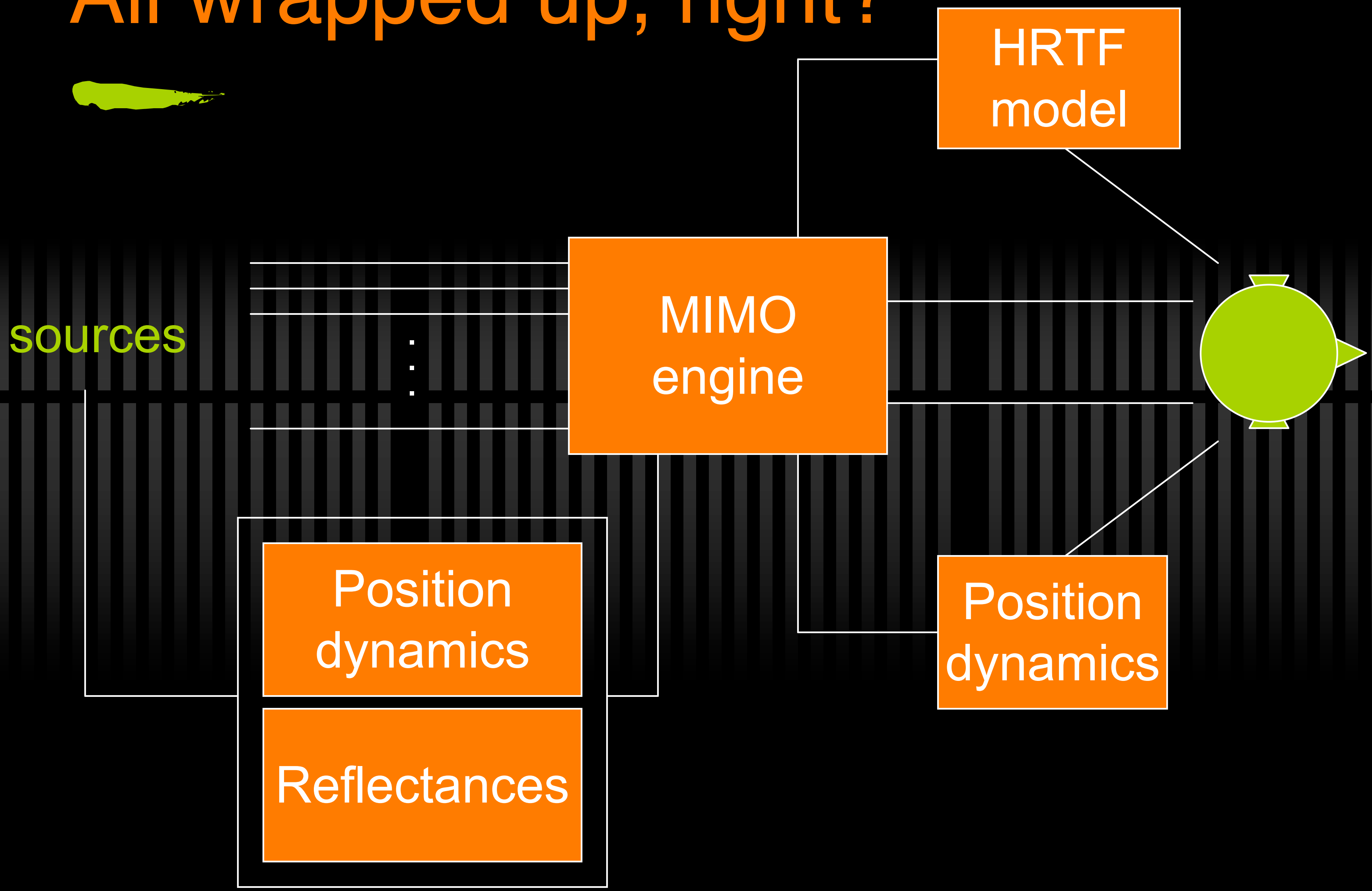
Additional rendering issues

- Rendering spatial extent
  - Types of spatially extended sources
    - Flock of geese
    - Flow-noise in aircraft or automotive interiors
    - Non-uniform spatial radiativity patterns
  - ???

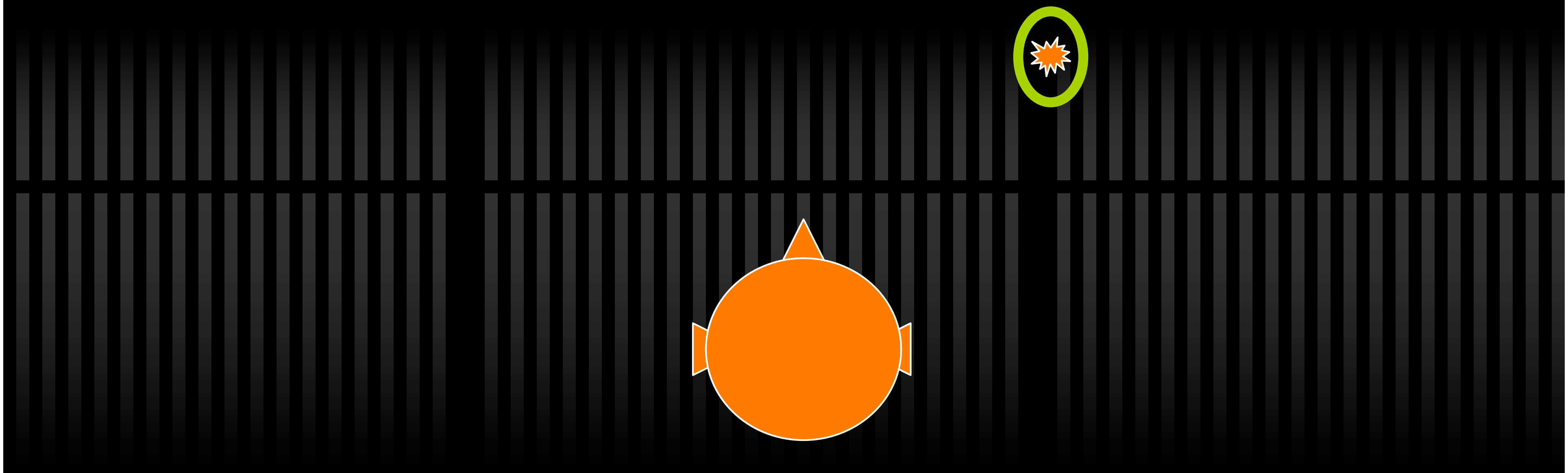
Current Directions



All wrapped up, right?

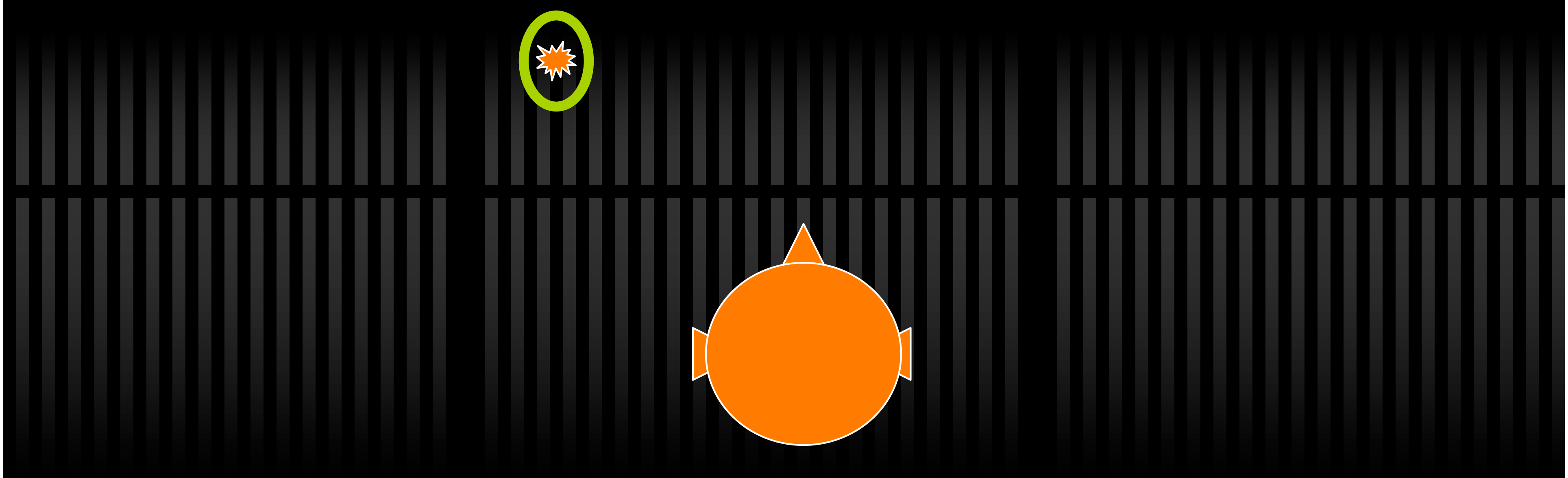


Caveat:
The Single-Source Mindset



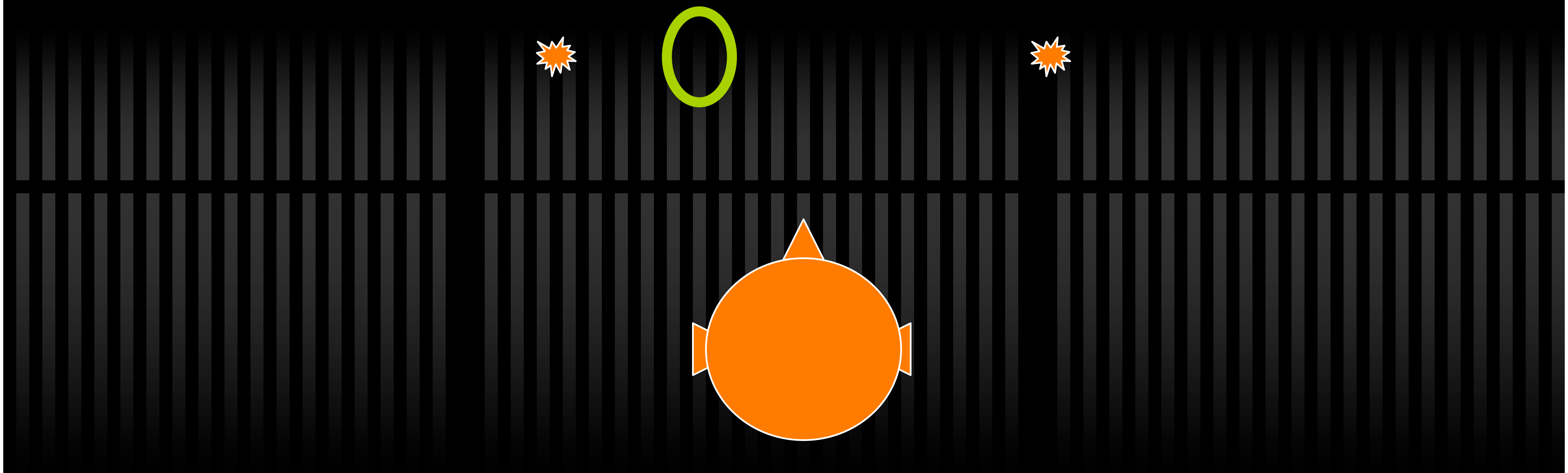
Where’s the source?

Caveat:
The Single-Source Mindset



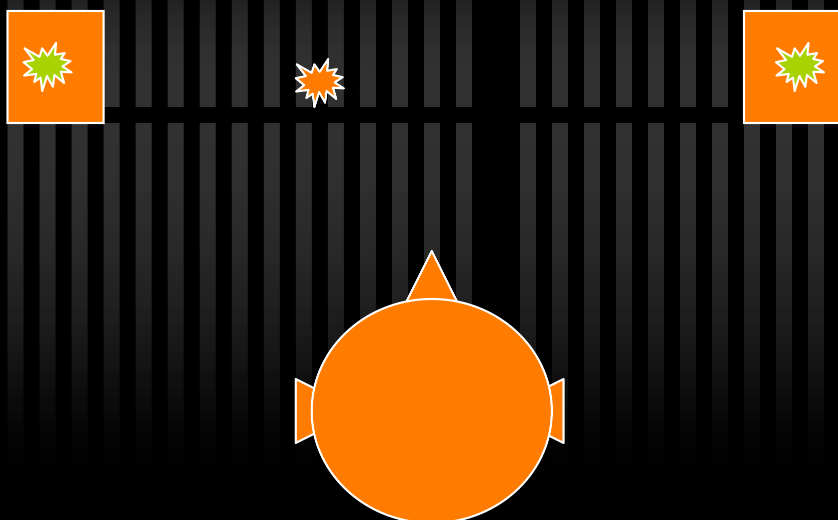
Where’s the source?

Caveat:
The Single-Source Mindset



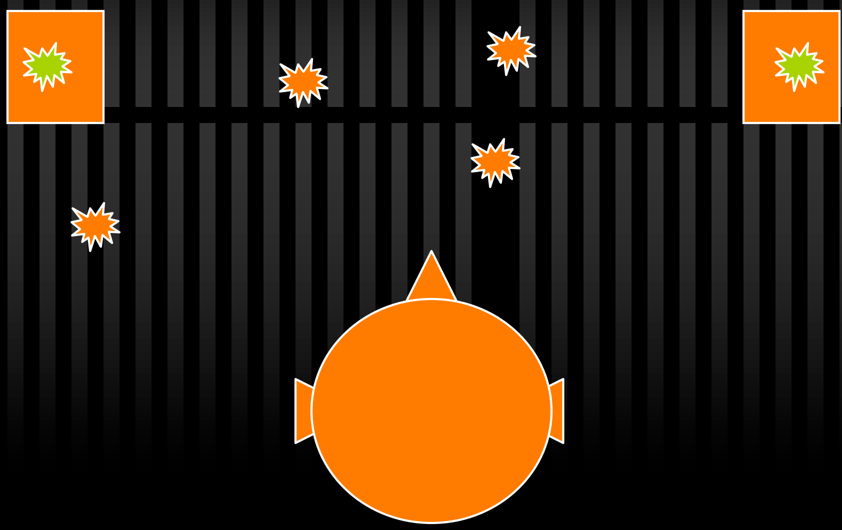
Where are the sources?

Binaural phenomena



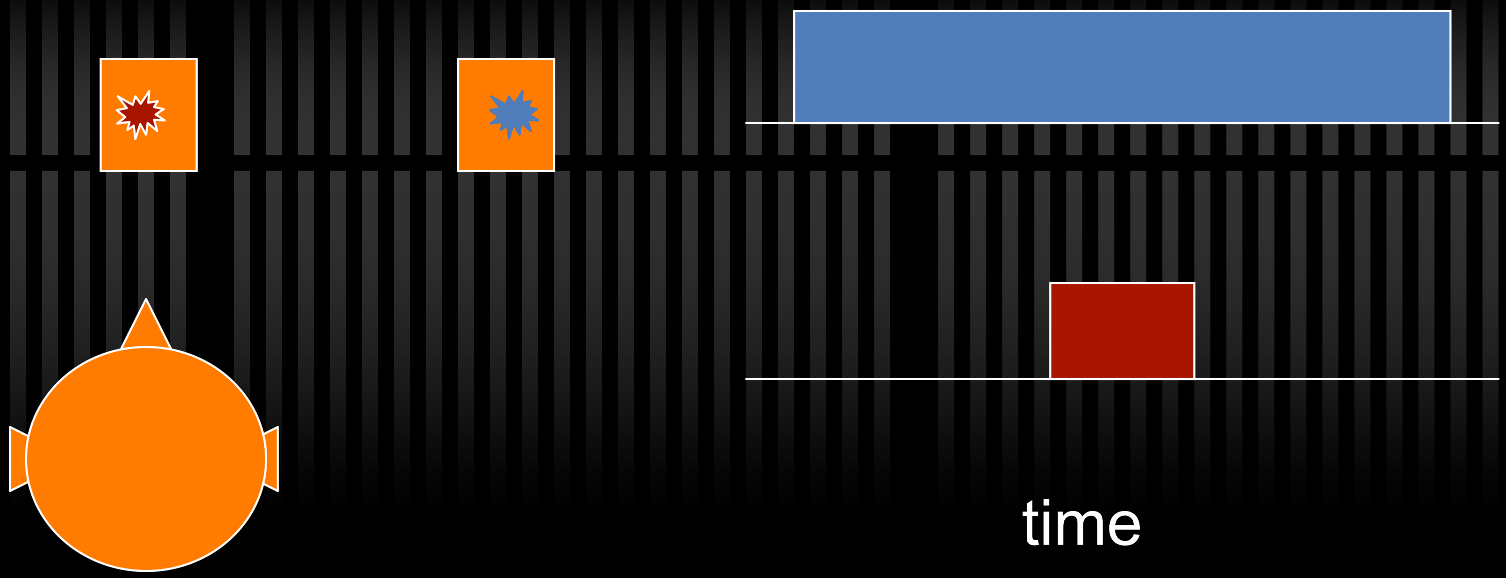
- Fusion
  - Multiple acoustic sources are fused into a single source
  - Source location dependent on relative timing of acoustic sources
  - Precedence effect
    - Reverb
    - Spectral effects
  - Franssen effect
  - Stereophony

Binaural phenomena



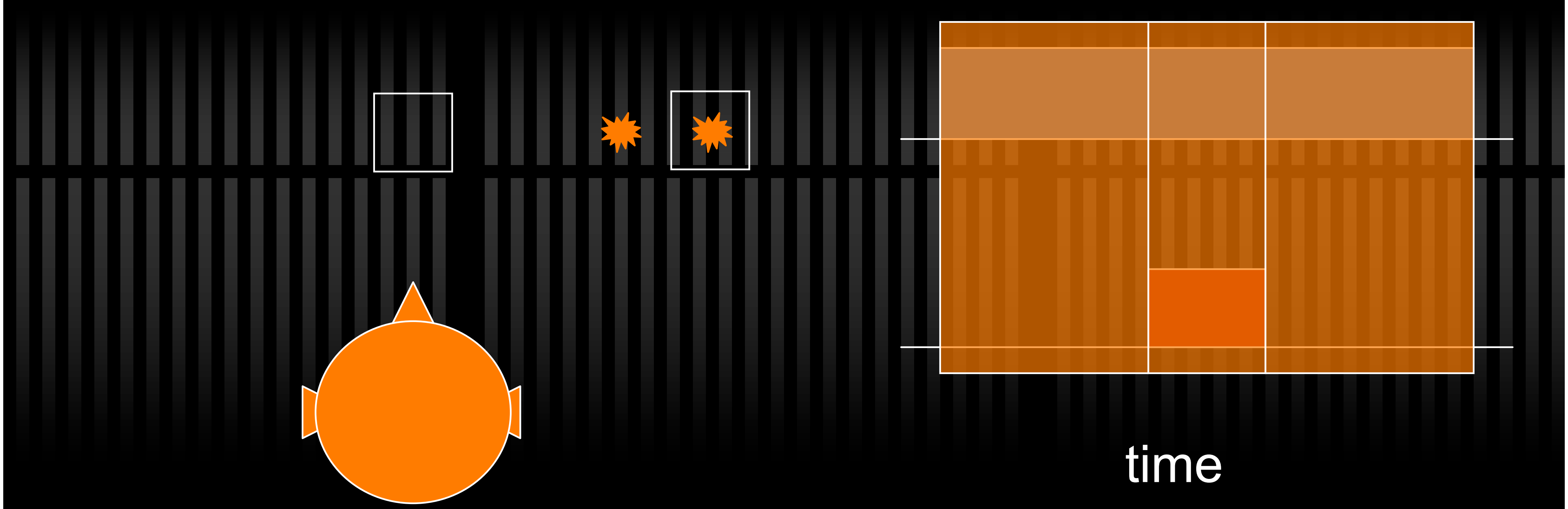
- Fission
  - Multiple acoustic sources may split into additional perceived sources
  - Source location dependent on relative timing of acoustic sources
  - Stereophony

Binaural phenomena



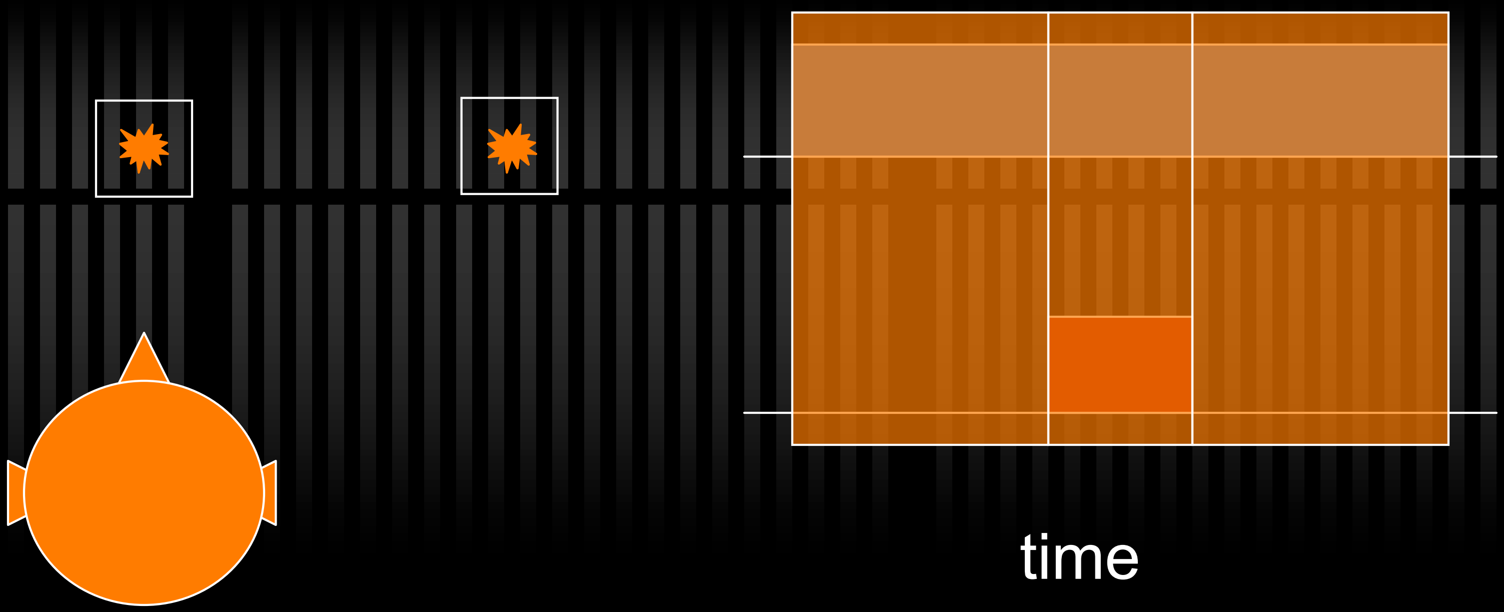
What should you hear?

Fusion mechanism



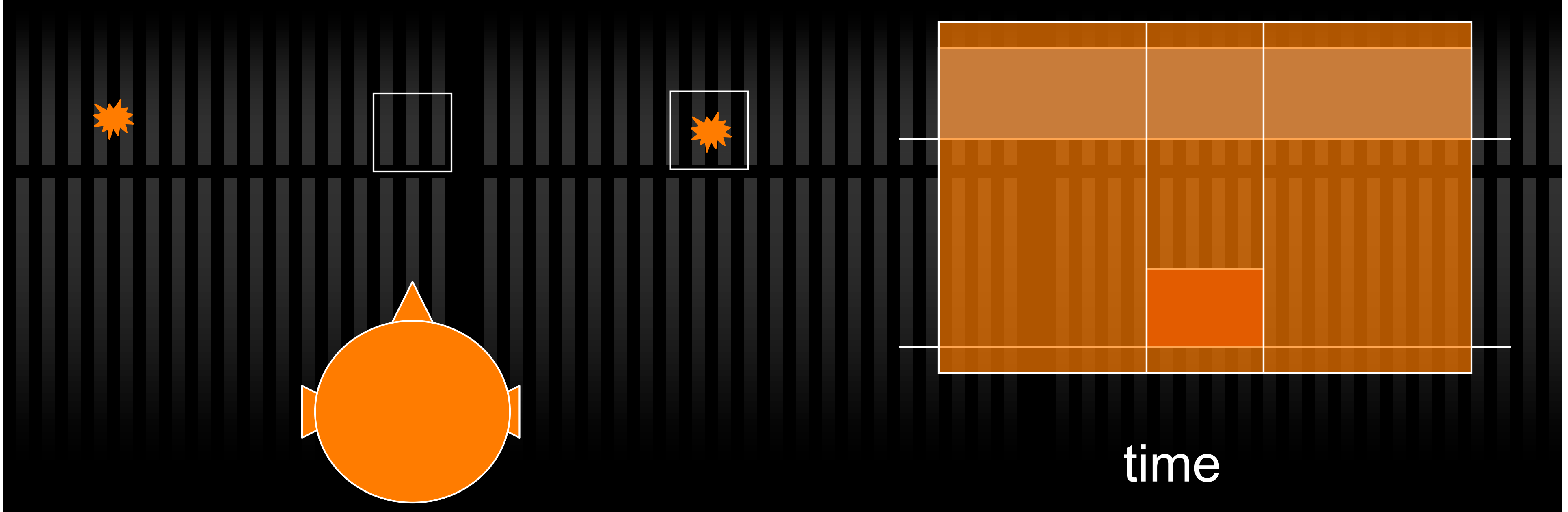
What should you hear?

Fusion mechanism



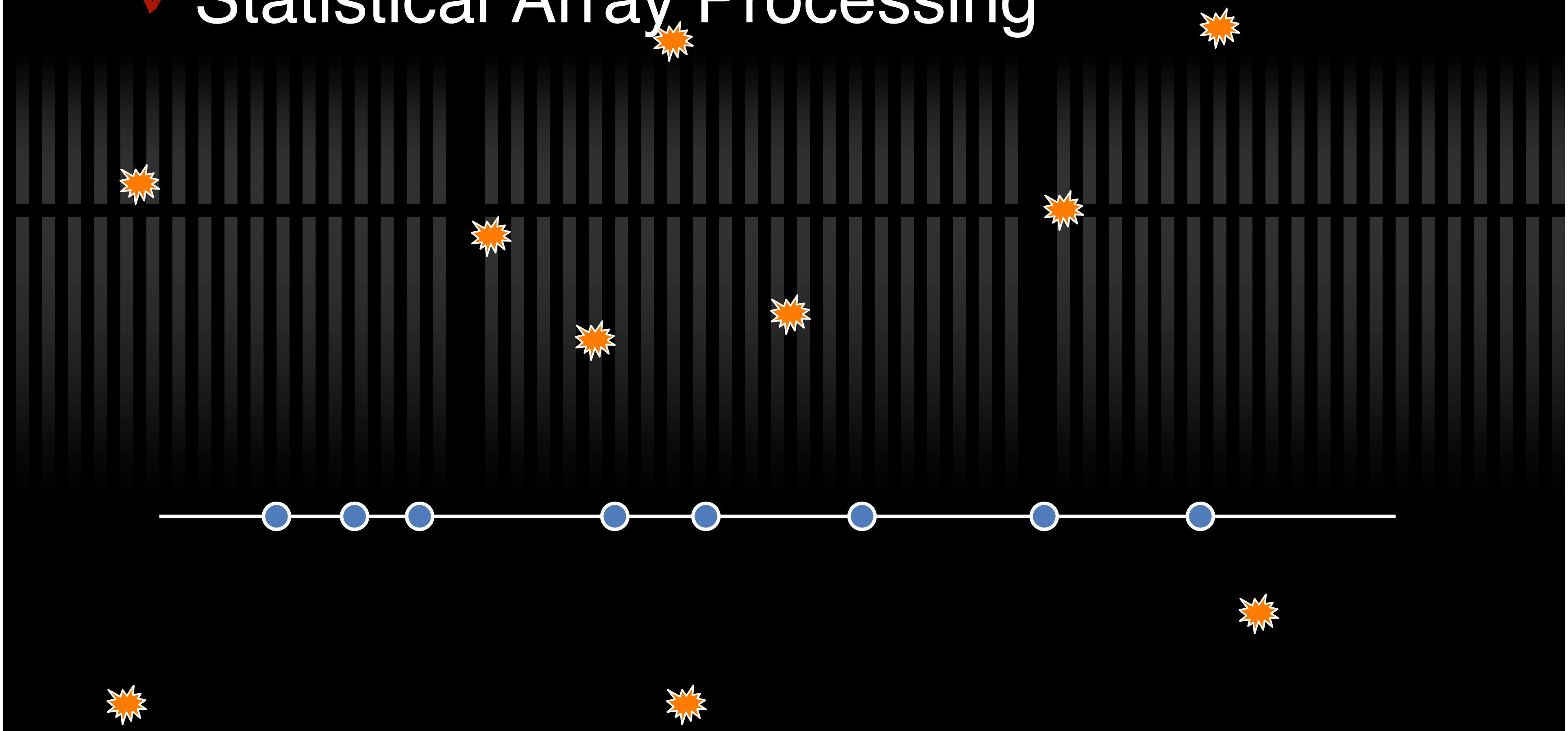
What should you hear?

What you do hear



SAPS and Perception

- ✓ Statistical Array Processing



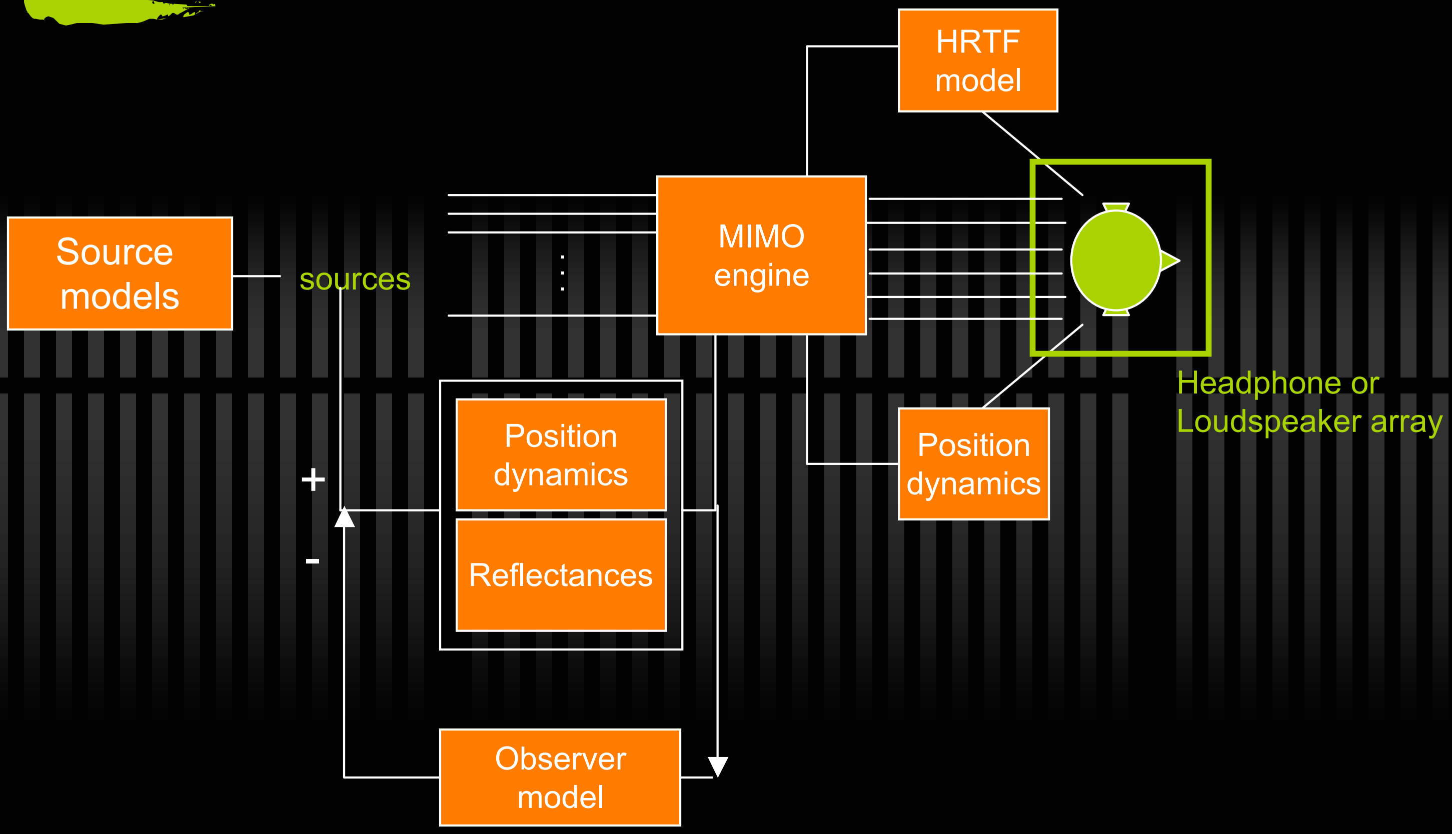
One-to-One vs. Many-to-One

- N sensors, M narrowband sources
  - $M \leq N$
    - One-to-one
    - Location is recoverable
  - $M > N$
    - Many-to-one
    - Additional constraints necessary to choose one solution

Mathematics of Spatial Hearing: Many-to-2

- Therefore, we should expect
  - Fusion
  - Fission
  - And everything in between

Wrap-Up



HRTF model
Source models
sources
MIMO engine
Headphone or Loudspeaker array
+
-
Position dynamics
Reflectances
Position dynamics
Observer model