Motivation/Goals

- The first-stage goal of the HRTF project
- Ultimately, obtain the passive tracker for the head position and orientation
  - To be used in the first working demo
  - The first demo will use head and torso
  - (\( \Rightarrow \) LF only since HF depend mostly on pinna)
- Want to develop a passive system and verify it with ground truth

Existing trackers

- Active tracker
  - Commercially available systems
    - Polhemus
    - Ascension
    - Vicon
  - Cost a lot
- Want to develop our own ultrasonic-based tracker for head position and orientation

Active position tracker

- Place an ultrasonic emitter on the head
  - Note that person have headphones anyway
- Place the microphone array in the ceiling
- Determine the TDOAs by
  - Generalized cross-correlation
  - Cross-spectrum phase shift
  - Sterausis
- Determine the position of the emitter

Active orientation tracker

- Use three non-collinear emitters
  - Some patterns are better than the others
  - From coordinates of all emitters one can compute the orientation of the head
- Separate them in time, frequency or pattern
  - Have to have non-sine-wave anyway (there is no features in the sine wave)
  - Easiest (to receive) is to separate in time
Practical issues
• Sampling frequency
• Microphone arrangement
• Some pattern for the emitters (Golay codes)
• Some filter for the receivers
• No ultrasonic in room (motion sensor)
• Synchronize emitters with the receiving end
• Get precision sufficient for head orientation

Semi-active position tracker
• Markers (beacons) on the headphones
• Color patch, or some icon, or bright/IR LED
• Several video cameras in the room
  – So that every point is visible from at least 2 cameras
• Theoretically if 2 cameras see the beacon its 3D-position can be computed
• Not so easy in practice…

Practical issues
• Camera auto-calibration
• Tracking (Kalman filtering with motion models; papers on feature points tracking)
• Parallel processing
• Markers selection/ recognition
  – Need really big icons to recognize them…
• LED synchronization with the cameras

Passive tracker
• Head position: contour tracking from multiple cameras
• Head orientation: facial features, ears with headphones and hairline
• Need to have a face big in frame to robustly detect facial features
• Zoom is slow, but can point zoomed camera to the person’s face. Still not high-fidelity.

Close-range passive tracker
• A listener sitting in front of the workstation
  – Teleconferencing
  – Entertainment/games
• In this case, we can have constant close-up view of the person’s head
  – Tracking of facial features is easy
• For the small head motion this is sufficient
  – 30 fps, 33ms latency in the ideal case

Practical issues
• All as in semi-passive tracker, and…
• Face and facial feature detection
• Complex shape contour tracking
  – Physics-based deformable models
  – Active contours (snakes)
• I don’t have much expertise here (yet) beyond reading papers
• Is long-range facial feature detection do-able now?
• Leave the question for the later discussion
Sound synthesis

- Use the obtained head position and orientation to select or compute the HRTF
- Convolve the sound(s) with HRIR(s)
- Convolve with the room response function
- Render through DAC
- Dynamically change the head position and orientation in accordance with tracker

Previous research at UMd

- Have working audio-video tracking system
  - Two cameras and two microphone arrays
  - Tuned for the speech-band
  - Uses skin color face tracker or background subtraction
  - Condensation tracker for the head
- Will use principles and algorithms for the active and semi-active tracker
- W4 system, close-range head orientation tracker

Picture of the system

Tracker usage

- An informal validation of the system
- Demonstration for the LF sounds
- Demonstration for the full-band sound with non-individualized pinna HRTF or with two-reflection pinna model
- Effects of small head motion on the stability of the perceived virtual environment

Thank you