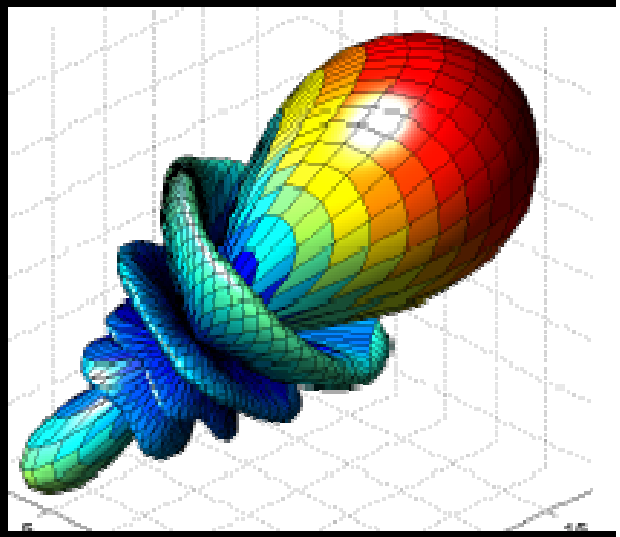
Supplemental Material for Paper 2523

Microphone Arrays as Generalized Cameras for Integrated Audio Visual Processing

Adam O'Donovan, Ramani Duraiswami and Jan Neumann

Spherical Microphone Arrays as Central Projection Sound Field Cameras

- Sound is collected at the microphones on a spherical array
- If we weigh the collected data appropriately and sum it, it is possible to pick out the sound from a particular direction, and suppress those from others.
- A direction on the sphere (a ray from the center along a particular direction) is characterized by a spherical delta function in two angles
- Using the spherical harmonic expansion for the delta function, truncating it to the order achievable by the array, we can collect sounds in ray-like beampatterns in any direction [see Refs. 20, 15]



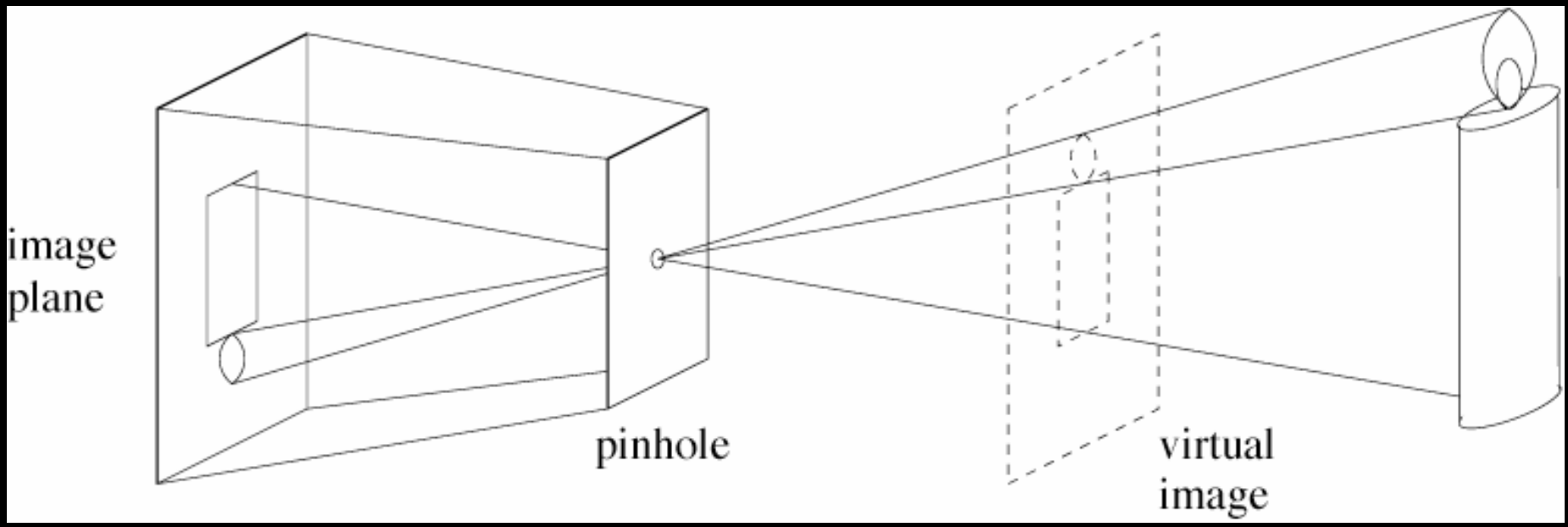
- The figure on the right shows an achieved spherical ray corresponding to a spherical harmonic expansion of order 6

Experimental spherical microphone array

- A 60 element spherical microphone array on a sphere of radius 0.09 m
- Beamweights computed according to References [20, 15]
- Note sound captured once can be processed either at the same time or later with different weights
- The beampattern is like the "point-spread function"

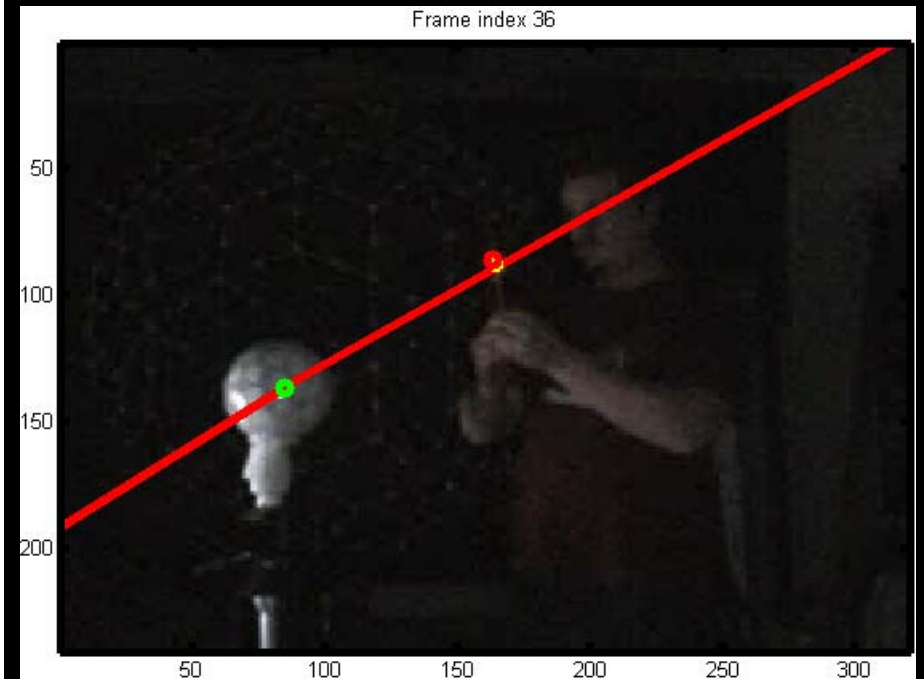
Central projection camera

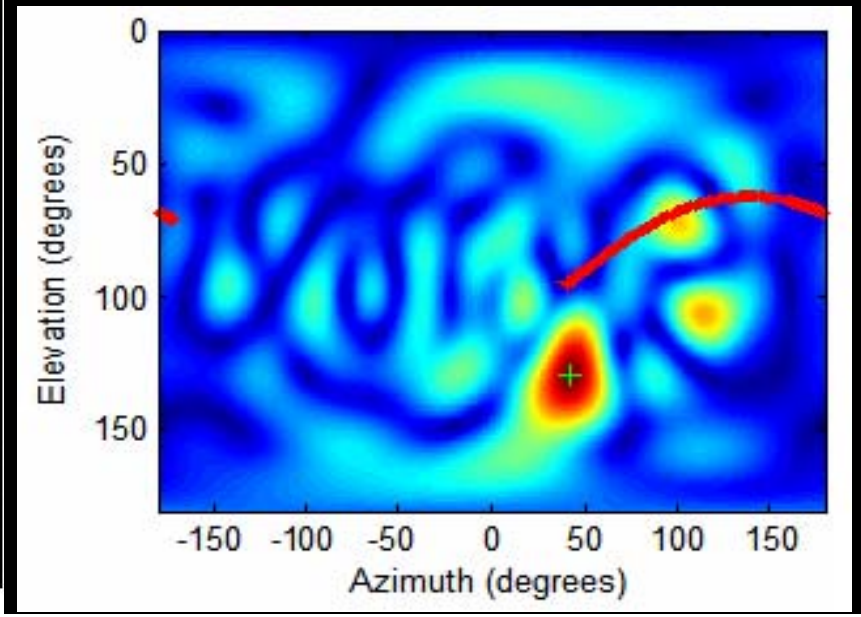
- In a central projection camera the straight line joining an object point and its image on the pixel plane passes through the camera center of projection



Camera-Spherical Array Epipolar Geometry

- As the array and camera are both central projection devices, corresponding points lie on an epipolar plane
- A calibration target containing a microspeaker and a light emitting diode on a wand (right) was constructed
- Epipolar lines from the array in the camera (left), and camera in the array (right) are shown as red lines below





Demonstration

- Using the calibrated camera/array, face detection, and a face template, localize speaker in video
- Epipolar line from the video in soundfield image provides a region for beamformer to search for
- Using this information the microphone array can lock on to a weak sound source, in the presence of a distracting source, and beamforming can suppress the distracter
- Movie shows a subject speaking while a loud music source plays. The video shows the epipolar line of the loud source from the array in the camera image
- The soundfield movie shows the intensity of the sound at a single frequency. Peaks in the magnitude are seen at the music source and speech source, and periodically in the walls from reverberation
- Speech is not understandable because of the loud music in the single microphone recording
- The beamformed speech, which uses the video information to locate the weaker sound source, is intelligible