Investigation of source localization and beamforming with a spherical microphone Array

Steve Philbert, University of Rochester, Department of Electrical and Computer Engineering

ABSTRACT

This project examines detection and localization of sound sources from a rigid spherical microphone array. The microphone array handles audio in two forms: raw, directly from individual elements; and ambisonics, a representation of audio as spherical harmonics. The detection and localization algorithms detect the number of active sources and the location of the active sources. Raw audio utilizes time difference of arrival (TDOA), while ambisonic audio relies on a grid of ambisonic decoders. The algorithms are run through simulations first, then optimized as plugins for use on a real-time digital audio workstation (DAW). A brain application runs parallel to the digital audio workstation which receives information on the number and location of active sources and translates this into the beam-forming plugins.

OVERVIEW

- The Eigenmike is a rigid spherical microphone with 32 elements. It has software to convert raw sound into an ambisonic field, then convert the ambisonic field into single channel beamformers. The beamformers require directional input (azimuth, and elevation), and a beam pattern (i.e. cardioid, hyper-cardioid, super-cardioid).
- The goal of the project is to implement an audio informed source localization algorithm that can detect audio sources, identify their location, and form a beam(s) in the direction of the audio. This operation should happen in real time.

Figure 1. Eigenmike

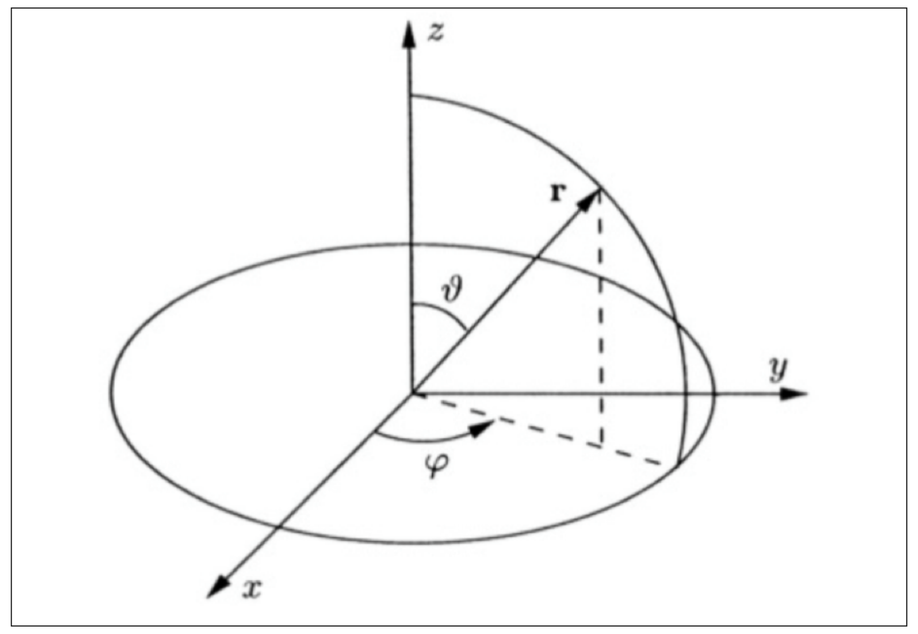

Figure 2. Spherical coordinates system used with the Eigenmike related to the cartesian coordinate system.

- Sources would ideally be musicians or talkers placed around the microphone array; on a stage or a conference room table
- Sources should not overlap one another, and should be approximately equal radius from the microphone
- Without visual clues; distance vs. intensity vs. orientation of source are difficult to differentiate
- This system does not train on source or position data; it should detect sources as they are heard

Detectors

Detectors

Time Difference of Arrival (TDOA)

1. Collect Audio at each element
   - Cross-correlate audio of each element with audio from all elements in the array
   - Find the index of the peak of the cross correlation for every element
   - Calculate the lag difference
2. Estimate number of sources
   - separation of sources and position; difficult task
3. Estimate locations of each source
   - if one maximum, assume source is in the direction of that element
   - if two maximums, assume source is in the direction of the vector normal to the bisection of those two elements
   - if three or more maximums, assume source is in the direction of the vector normal to a plane containing three elements

Ambisonics

1. Encode audio from elements array into ambisonic audio
2. Create a surface grid containing of many decoders
3. Decode ambisonics audio with decoders
4. Organize data by azimuth, elevation, magnitude (RMS)
   - Set threshold and find peaks above threshold
   - Number of peaks equals number of active sources
   - interpolate between peaks to find the exact locations

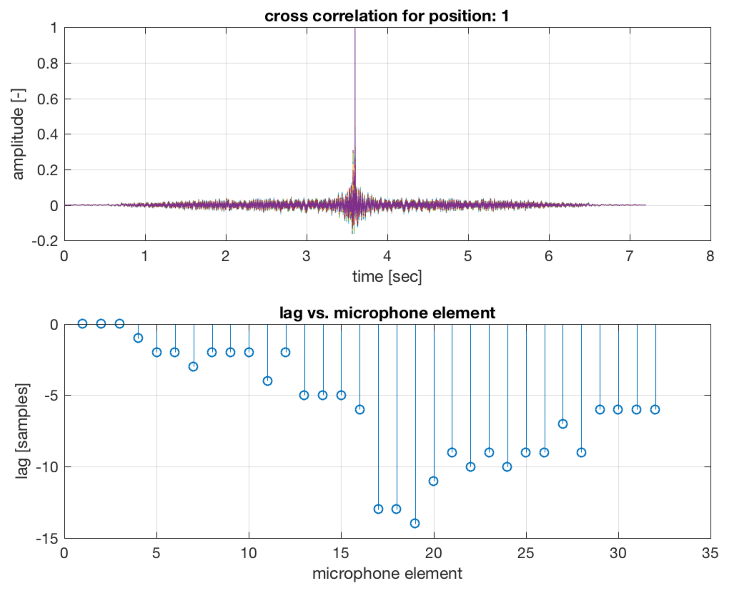

Figure 3. TDOA detector

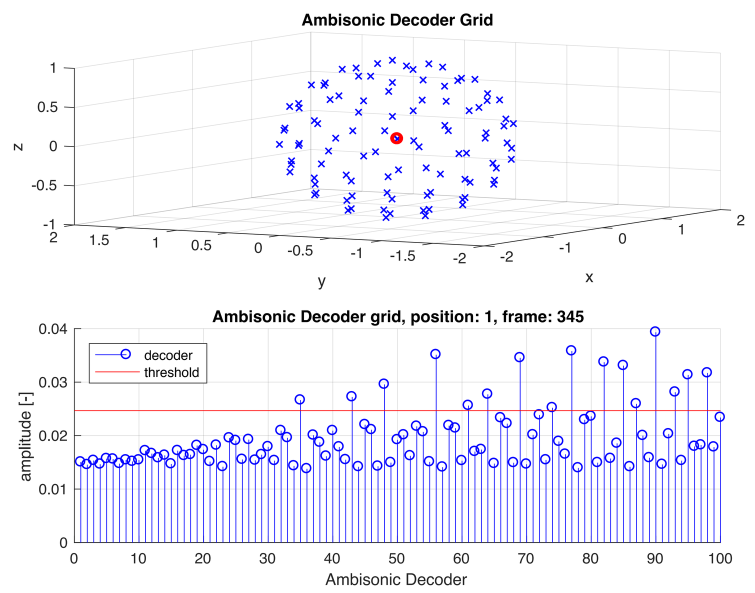

Figure 4. Ambisonic detector

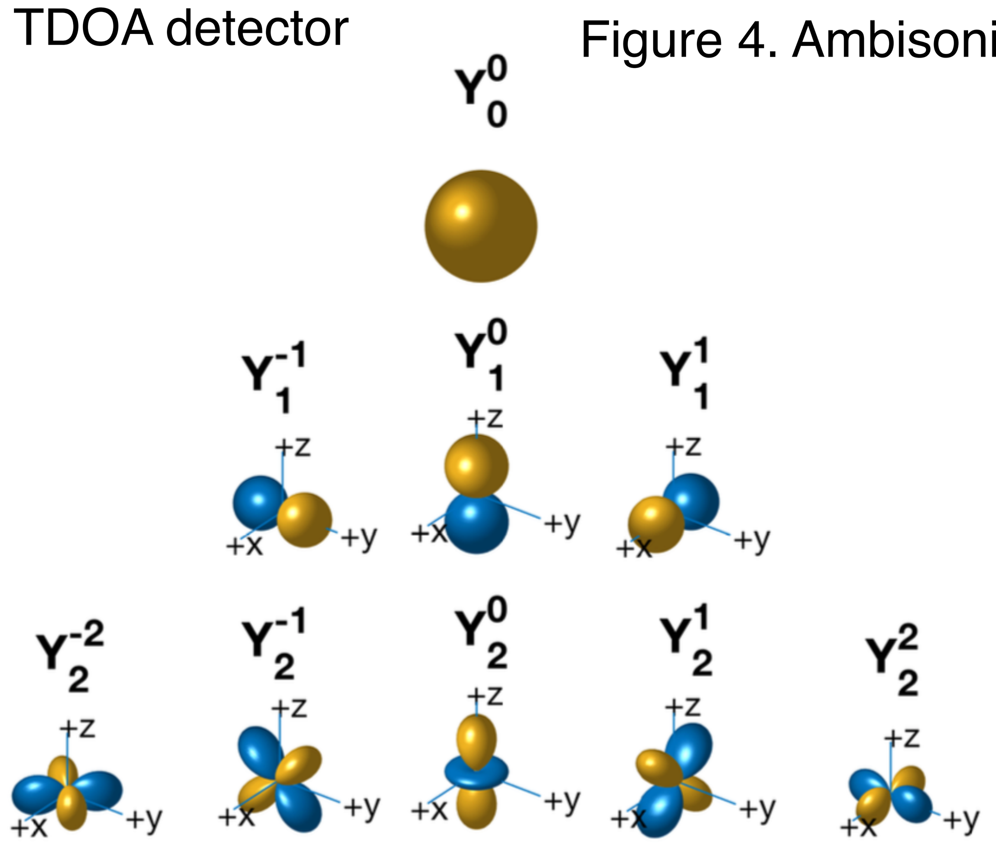

Figure 5. Spherical Harmonics shown up to second order

IMPLEMENTATION

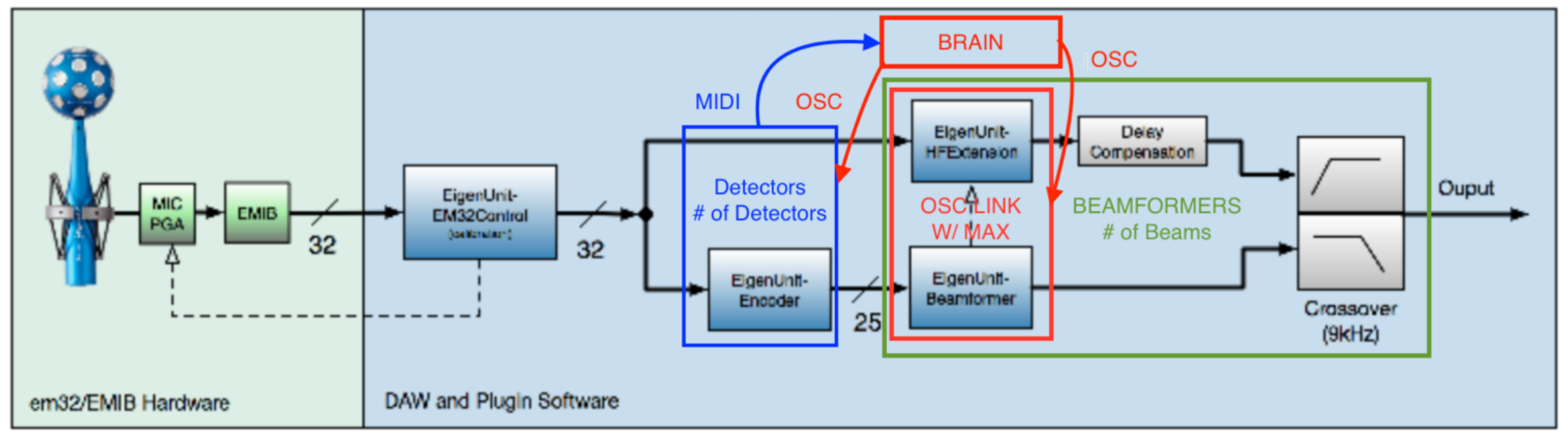

Figure 5. Implementation of detectors and beamformers

- EigenUnit EM32Control ~ sets gain and calibration of microphone
- EigenUnit Encoder ~ converts raw audio into ambisonics
- EigenUnit HFExtension ~ handles spatial aliasing
- EigenUnit Beamformer ~ forms beams from azimuth, elevation, and beam type.
- Brain ~ Controls the number, type, and direction of beams ~ manages all plugins ~ handles communication (OSC, MIDI) between plugins
- Detector32 ~ Time difference of Arrival implementation
- Detector25 ~ Ambisonic implementation
- Detector1 ~ Measures RMS for verification of beams; also measures intensity from ambisonic W channel.

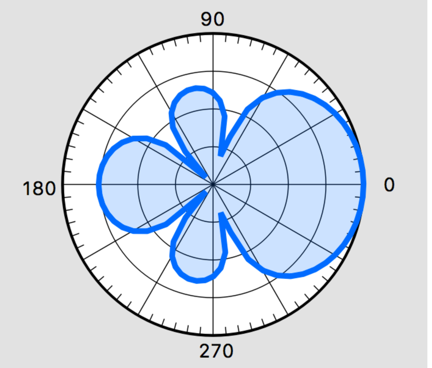

Figure 6. hyper-cardioid 2nd-order beam

CONCLUSION

- The time difference of arrival detector works well for single sources. It is difficult to estimate multiple active sources with spherical array, because sound sources and element positions are correlated. Cross correlation gives the time of arrival, not the distance of arrival or source from which it arrived.
- The Ambisonics detector separates the source and distance correlation by encoding location into spherical harmonics. The spherical geometry of the array makes this a good option. However a large number of decoders is needed to accurately locate a source.
- The implementation uses plugin provided with the Eigenmike; there are simpler methods to implement detector and beam forming circuits and projects. This was an overview of the possibilities of localization and beam forming.

Future work areas

- Thresholding is full range, this could be band limited
- Investigation into optimizing the layouts of surface grid decoders
- Use video informed source localization to track the locations of talkers and musicians; this would work well in a spherical layout.
- Other array geometries may provide better results for certain applications, as well as combining microphone arrays to obtain unique positions and cover different angles.