THE AMBISONIC DECODER TOOLBOX: EXTENSIONS FOR PARTIAL COVERAGE LOUDSPEAKER ARRAYS

Aaron J. Heller, AI Center, SRI International, Menlo Park, CA US
Eric M. Benjamin, Surround Research, Pacifica, CA US

Linux Audio Conference, May 3, 2014

What is Ambisonics?

- Extensible, hierarchical system for representing sound fields
  - Says how something should sound, rather than specific speaker signals.
- Capture or creation
  - Microphone arrays
    - 2-D or 3-D
    - Natural B-format, Tetrahedral, Spherical arrays
  - Ambisonic Panners
- Reproduction
  - 2-D, “horizontal” or 3-D “with height” loudspeaker arrays
  - “Any” size or shape array of loudspeakers

What is an Ambisonic Decoder?

- In Ambisonics, the program format is independent of the reproduction layout.

- The decoder's task is to create the best *perceptual impression* possible that the sound field is being reproduced accurately, given the resources available
  - Bandwidth, number of speakers, configuration of speakers …

- We use the term "decoder" to mean the configuration for a decoding engine that does the actual signal processing

Goals for decoder design

- Mimic conditions of natural hearing
  - Constant amplitude gain for all source directions (*P*)
  - Constant energy gain for all source directions (*E*)
  - At low frequencies, correct reproduced wavefront direction and velocity (r_V)
  - At high frequencies, maximum concentration of energy in the source direction (r_E)
  - Matching high- and low-frequency perceived directions

- Getting r_E correct is the most difficult aspect
- Recent work shows that it is also the most important!

Designing Decoders

- Decoders for regular polygon and polyhedra loudspeaker arrays are easy to design
  - Build the speaker encoding matrix, K, by sampling the spherical harmonics at the speaker directions
  - Use pseudoinverse to find the basic decoding matrix M
  - rE guaranteed to point in same direction as rV
- However…
  - Room geometry or visual considerations often limit speaker placement
  - 3-D HOA requires placing more speakers above and *below* the listener

How you'd like to do it

AuraLab, San Francisco

A useful compromise

The Bubble, San Francisco

Tradeoffs

- Once we deviate from regular geometry
  - we must trade off localization accuracy for uniform loudness
  - Directions of rE and rV are not the same
- Localization degrades outside the area with a high density of loudspeakers
- Gerzon used nonlinear optimization for this
  - Many implementations: Wiggins, Moore & Wakefield, Tsang, BLaH
- Works well for small arrays (e.g., ITU 5.1)
- Convergence is slow for large HOA arrays (hrs)

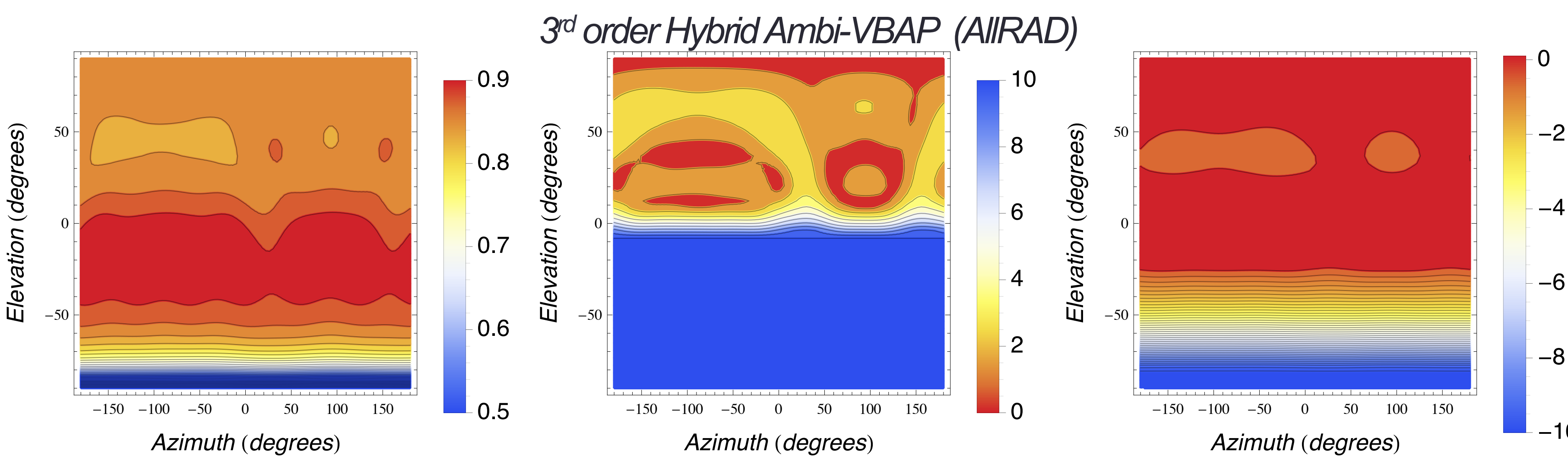
- IDHOA (Scaini and Arteaga) looks promising
  - Better objective function and zero out small coefficients

New Strategies in Toolbox

- Use an inversion technique suited to ill-conditioned matrices
  - Constant energy decoder
  - Truncated SVD
  - Energy limited
- Invert a well-behaved full-sphere virtual speaker array, map to a real array
  - Hybrid Ambisonic-VBAP
  - AllRAD (Zotter and Frank)
- Derive a new set of basis functions for which inversion is well behaved
  - Spherical Slepian Functions
  - EPAD (Zotter, Pomberger, Noisternig)

Are these decoders Ambisonic?

- Ambisonic theory specifies performance goals, not how to design a decoder
- We use the same criteria for these decoders
- But…
  - Apply them only to source directions in the covered part of the sphere
  - Require them be "well behaved" in other directions



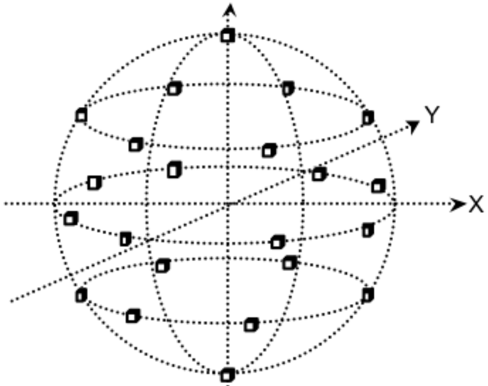
(a) r_E vs. Test Direction

(b) r_E Direction Error (degrees)

(c) Energy Gain (dB)

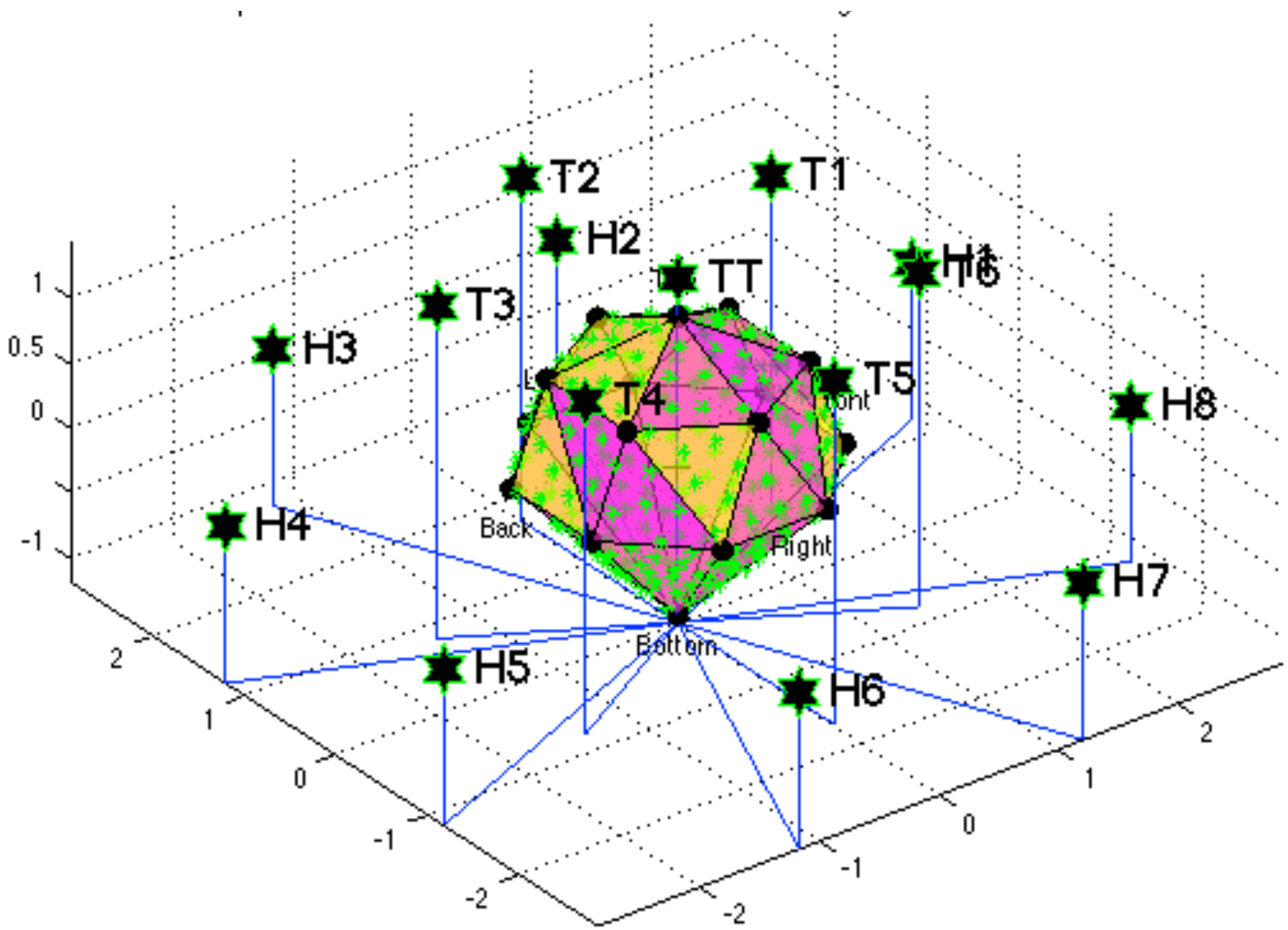
CCRMA Listening Room

- 22 identical loudspeakers in five rings
- Horizontal ring of 8 loudspeakers
- 2 rings of 6 loudspeakers, one 50° below horizontal and one 40° above
- 1 loudspeaker at each pole
- Array is almost regular
- Upper 15 used for hemispherical dome
- Full-sphere decoder described in our LAC2012 paper



Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

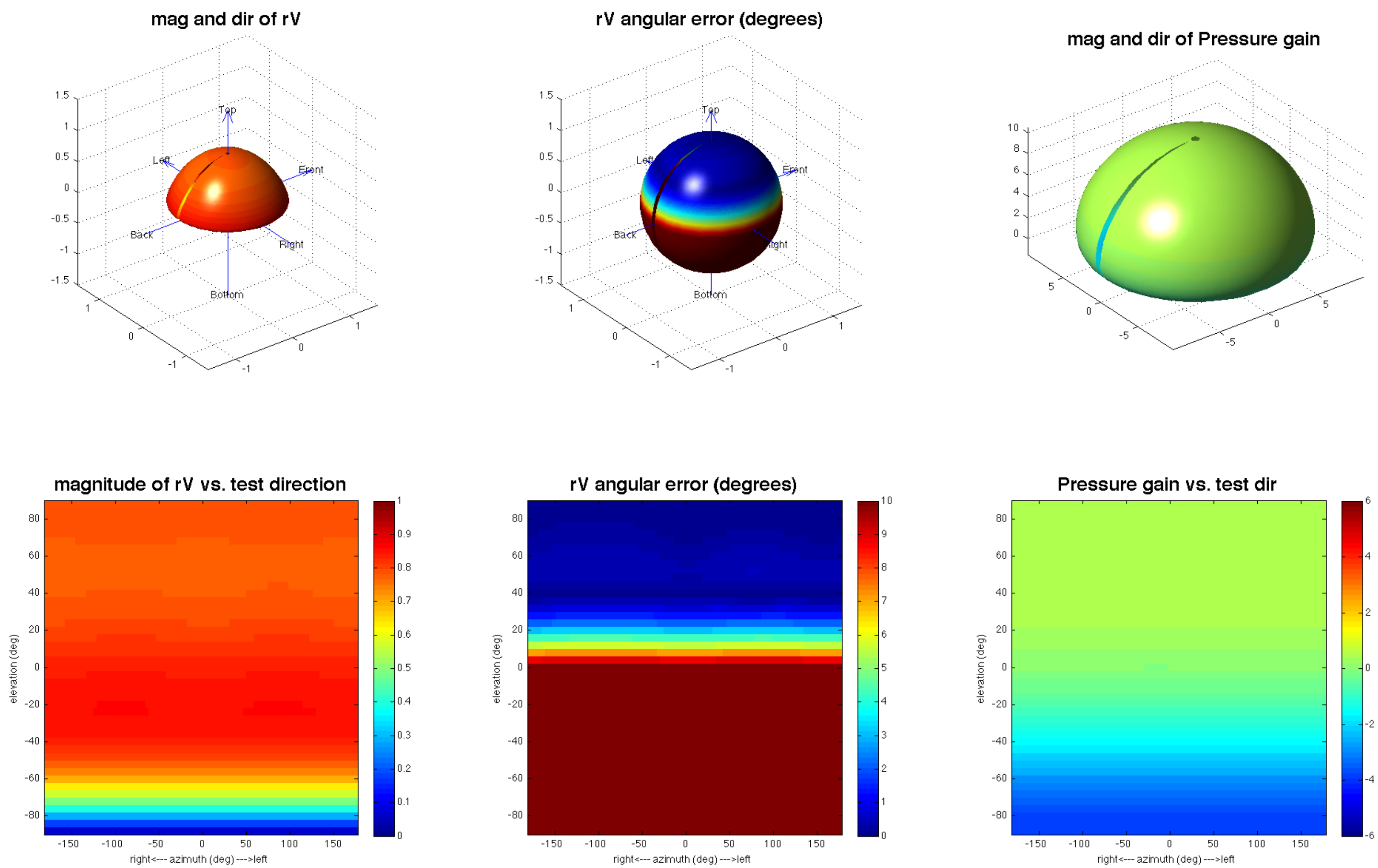
AllRAD – Hybrid Ambi-VBAP



- 240 point spherical design for virtual speaker array
- Dome of upper 15 loudspeakers of CCRMA Listening Room, 8-6-1
- Imaginary speaker at bottom
- Design procedure detailed in paper

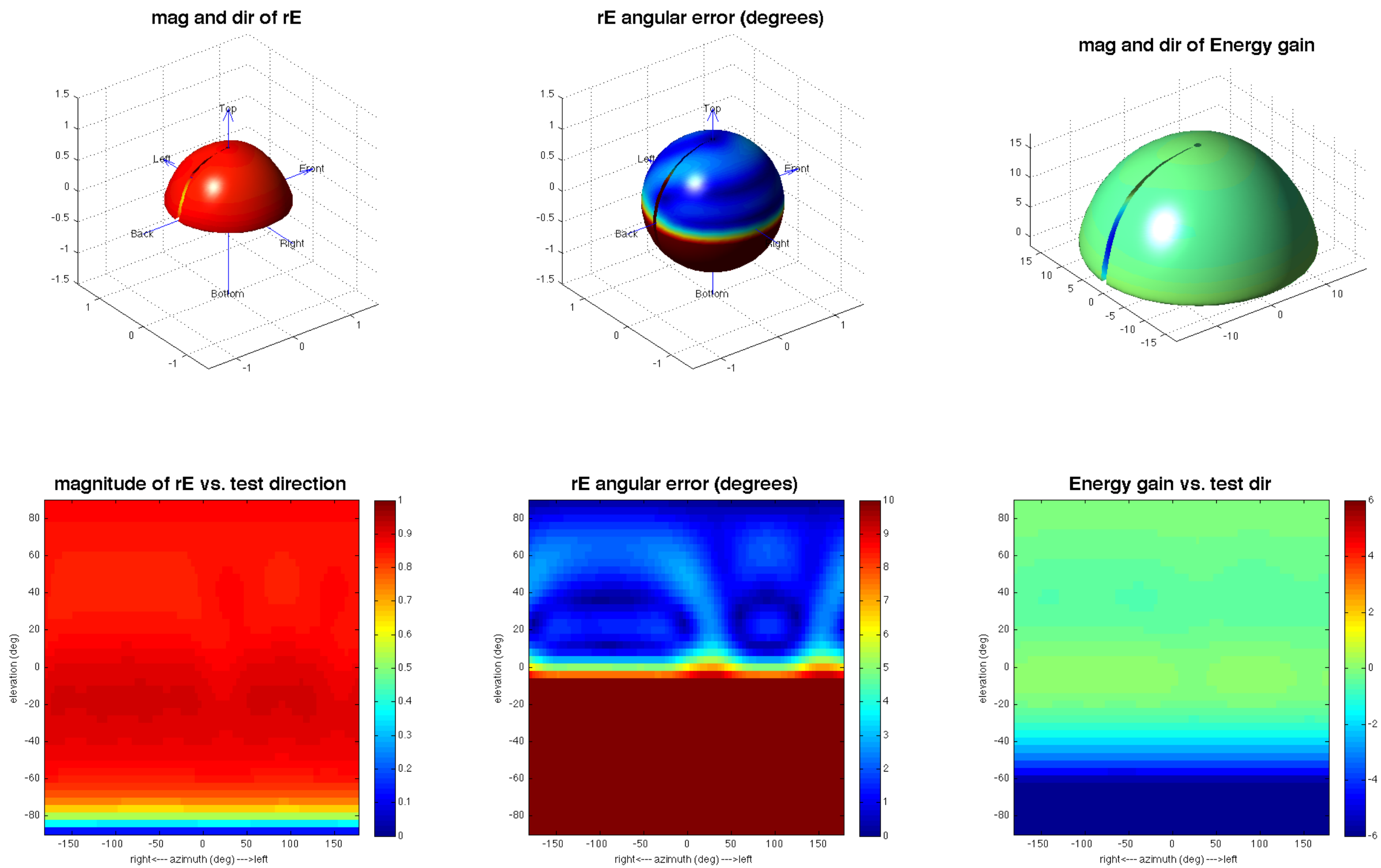
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

AllRAD performance r_V



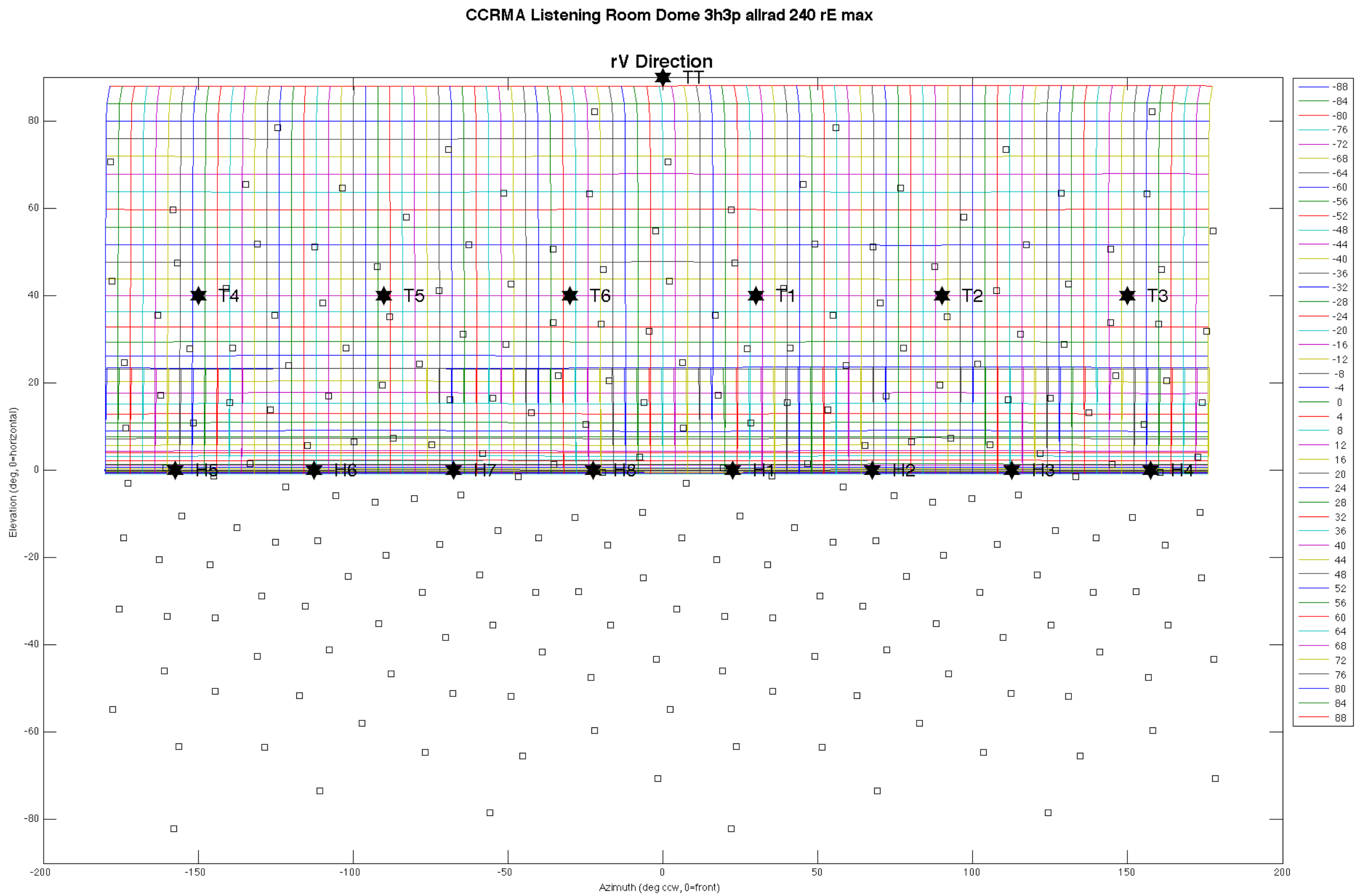
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

AllRAD performance r_E



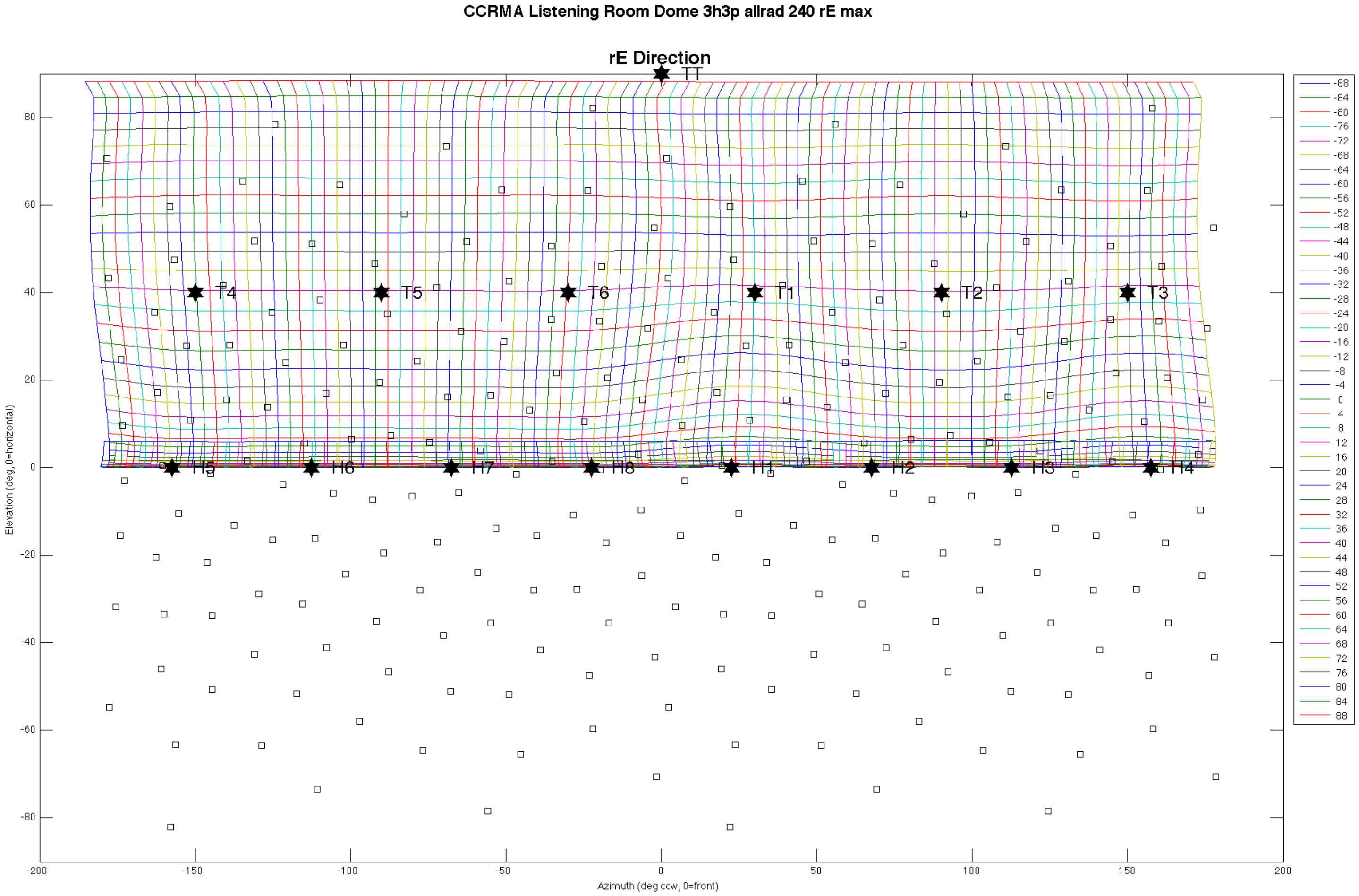
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

AllRAD r_V direction grid



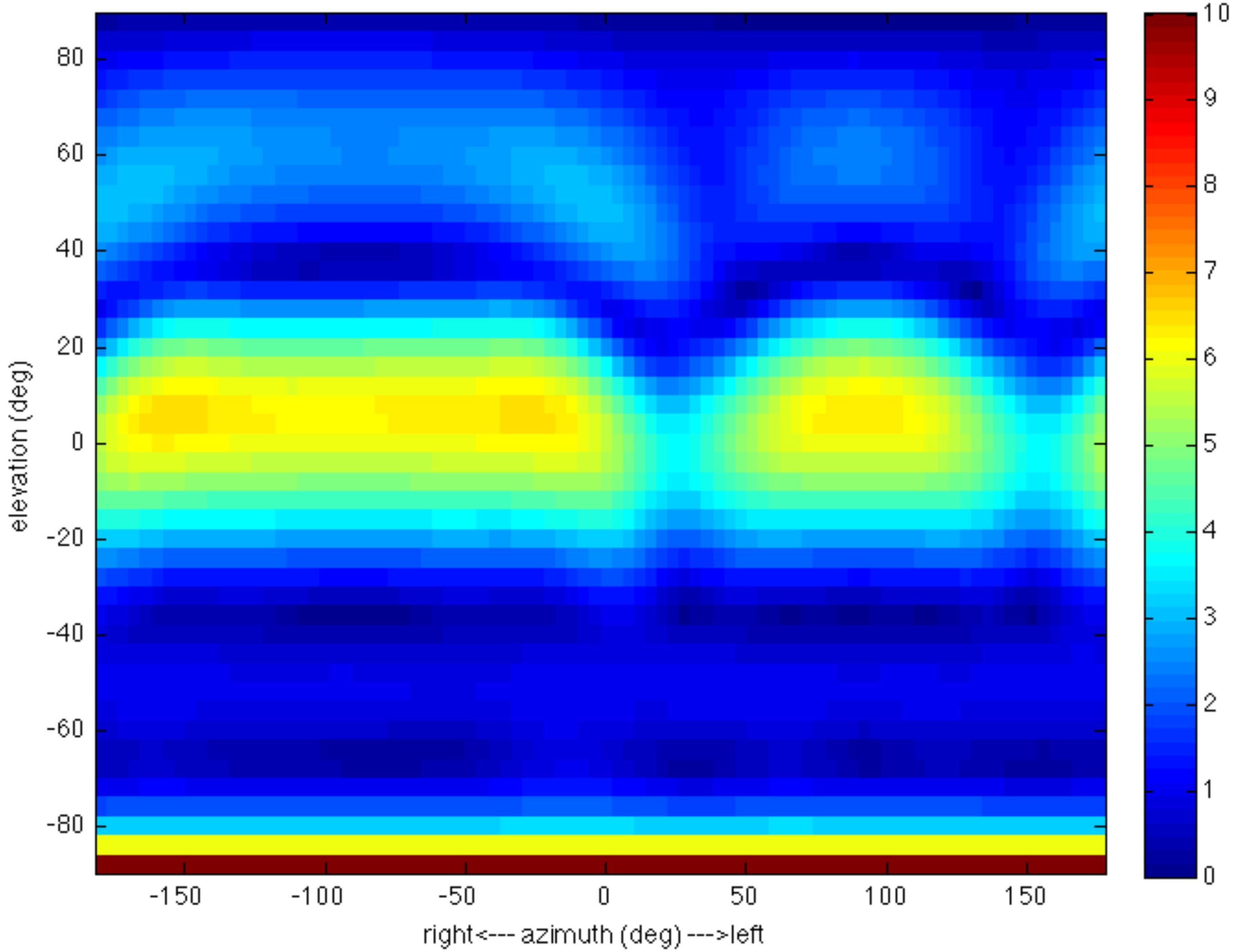
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

AllRAD r_E direction grid



Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

CCRMA Listening Room Dome 3h3p allrad 240 rE max
rV rE Direction Difference



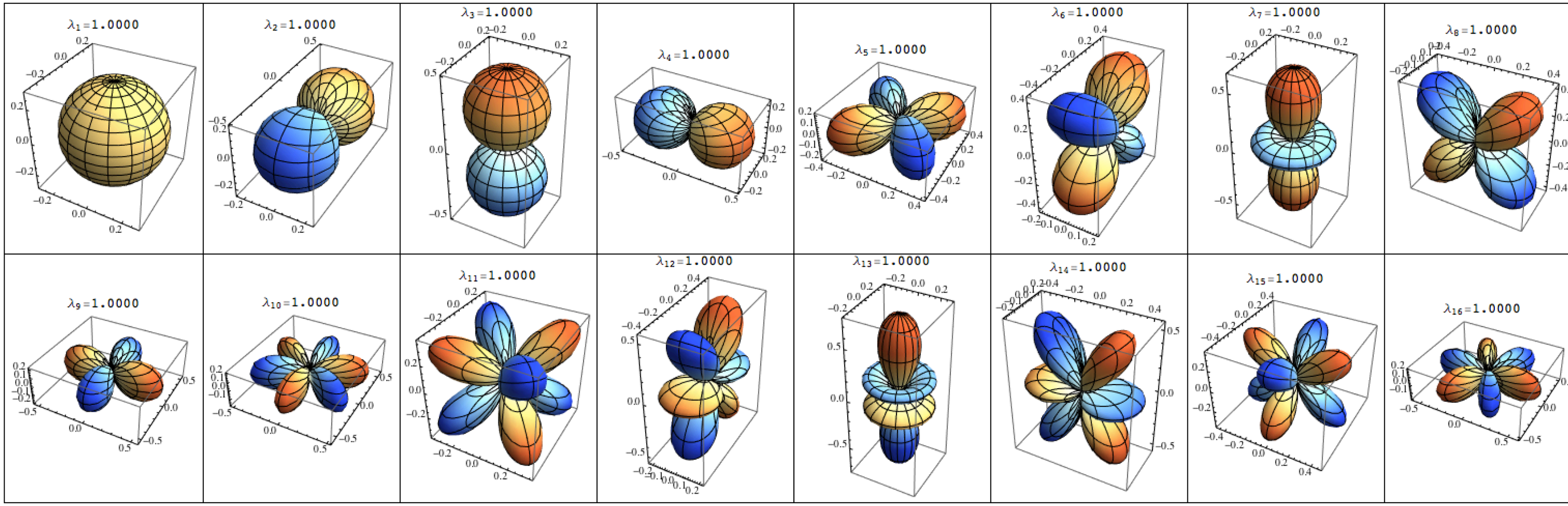
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

Spherical Slepian Functions

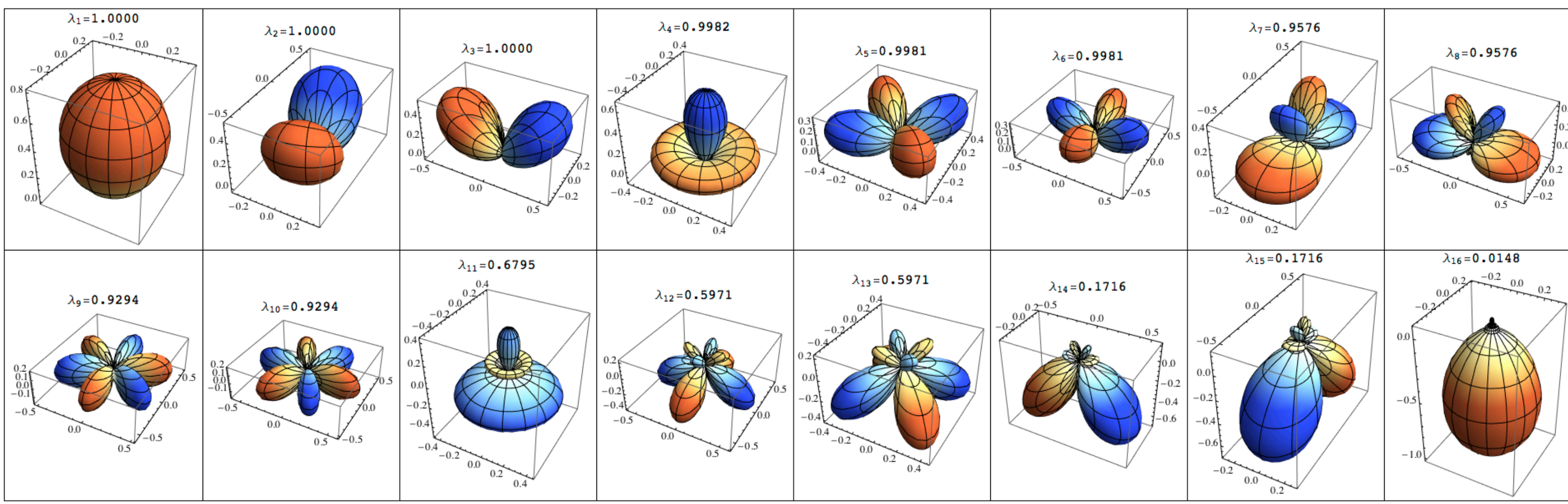
- Linear combinations of spherical harmonics
- Produce a new set of basis functions that are zero outside the region of interest on the sphere
- Remain orthogonal within the region
- Used in satellite geodesy to model earth's gravitational and magnetic fields from incomplete data

- In Ambisonic decoding, we can specify a region of the sphere, a dome or a ring, and derive a well behaved set of basis functions for that region.
- Design procedure detailed in paper

3rd order spherical harmonics (blue = inverted polarity)

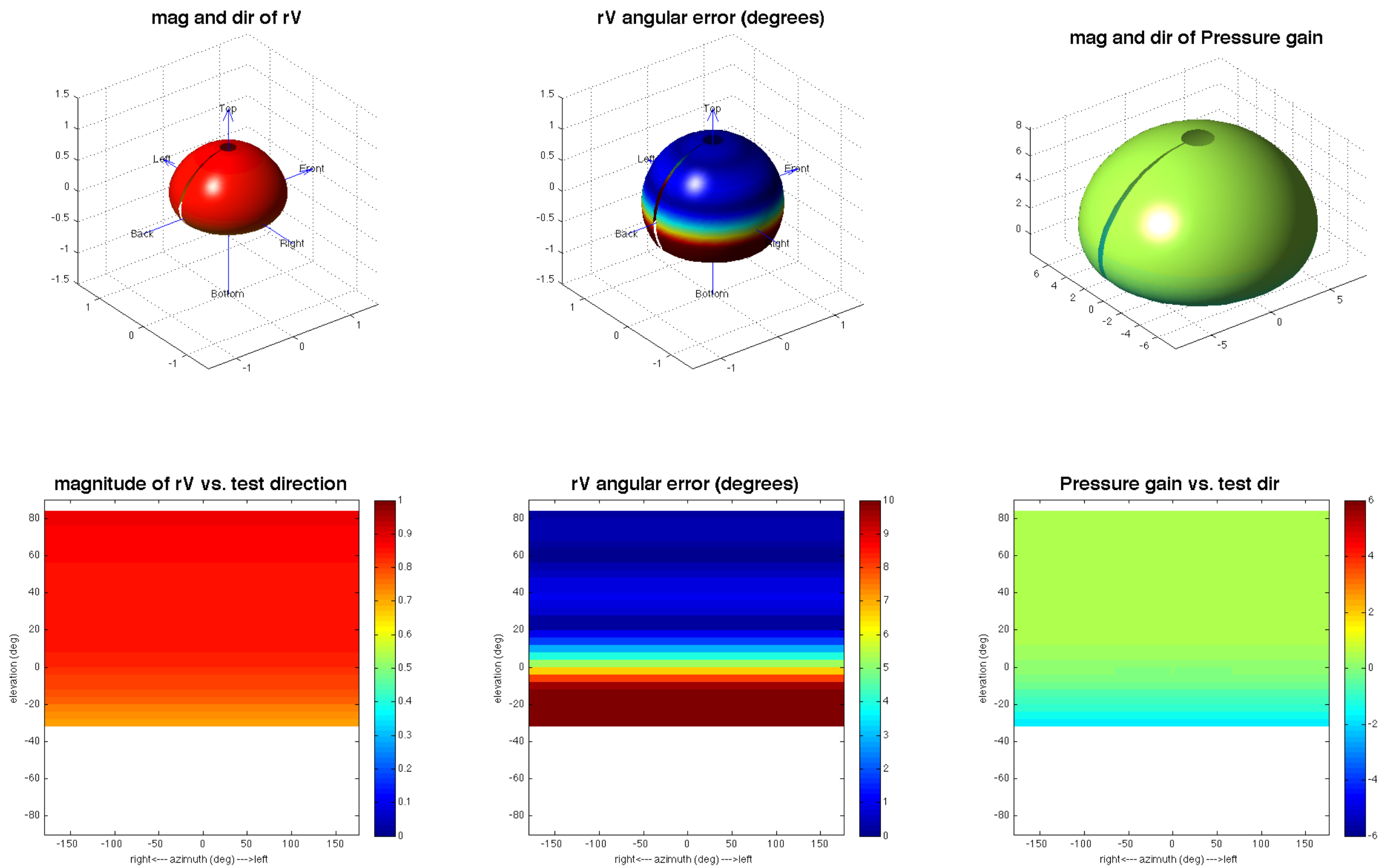


3rd order spherical Slepian functions for +90° to -30° dome (first 13 used for decoder)



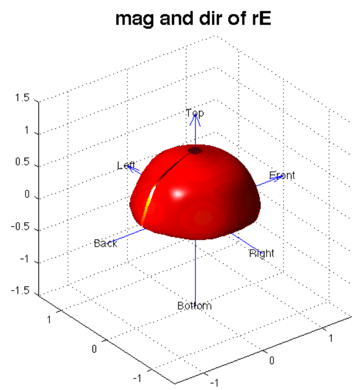
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

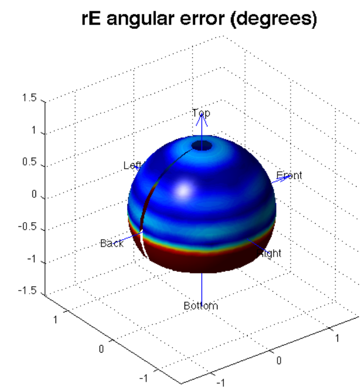
Spherical Slepian performance r_V

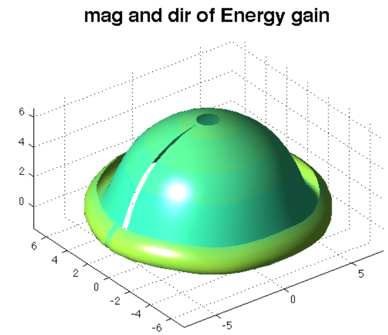


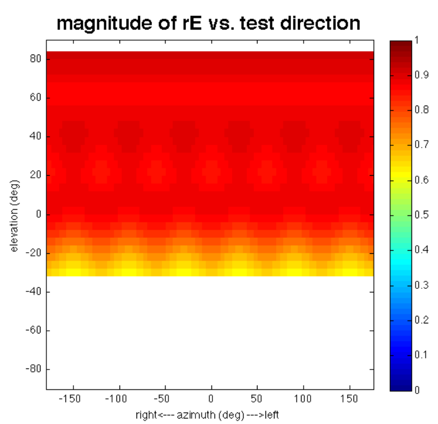
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

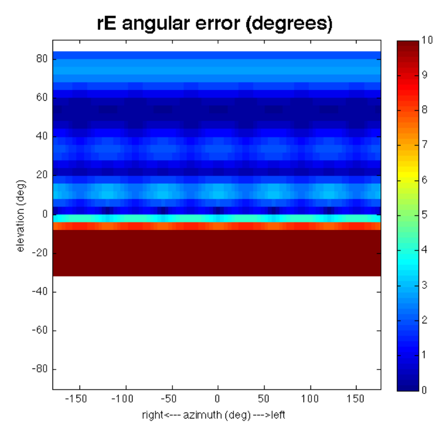
Spherical Slepian performance r_E

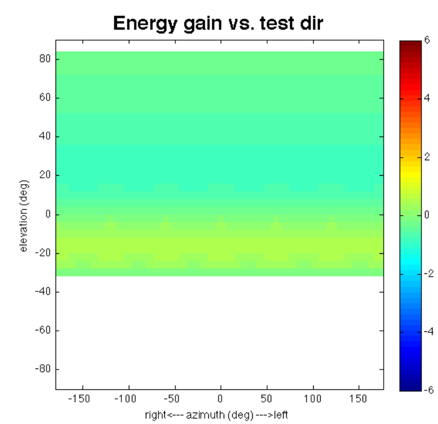






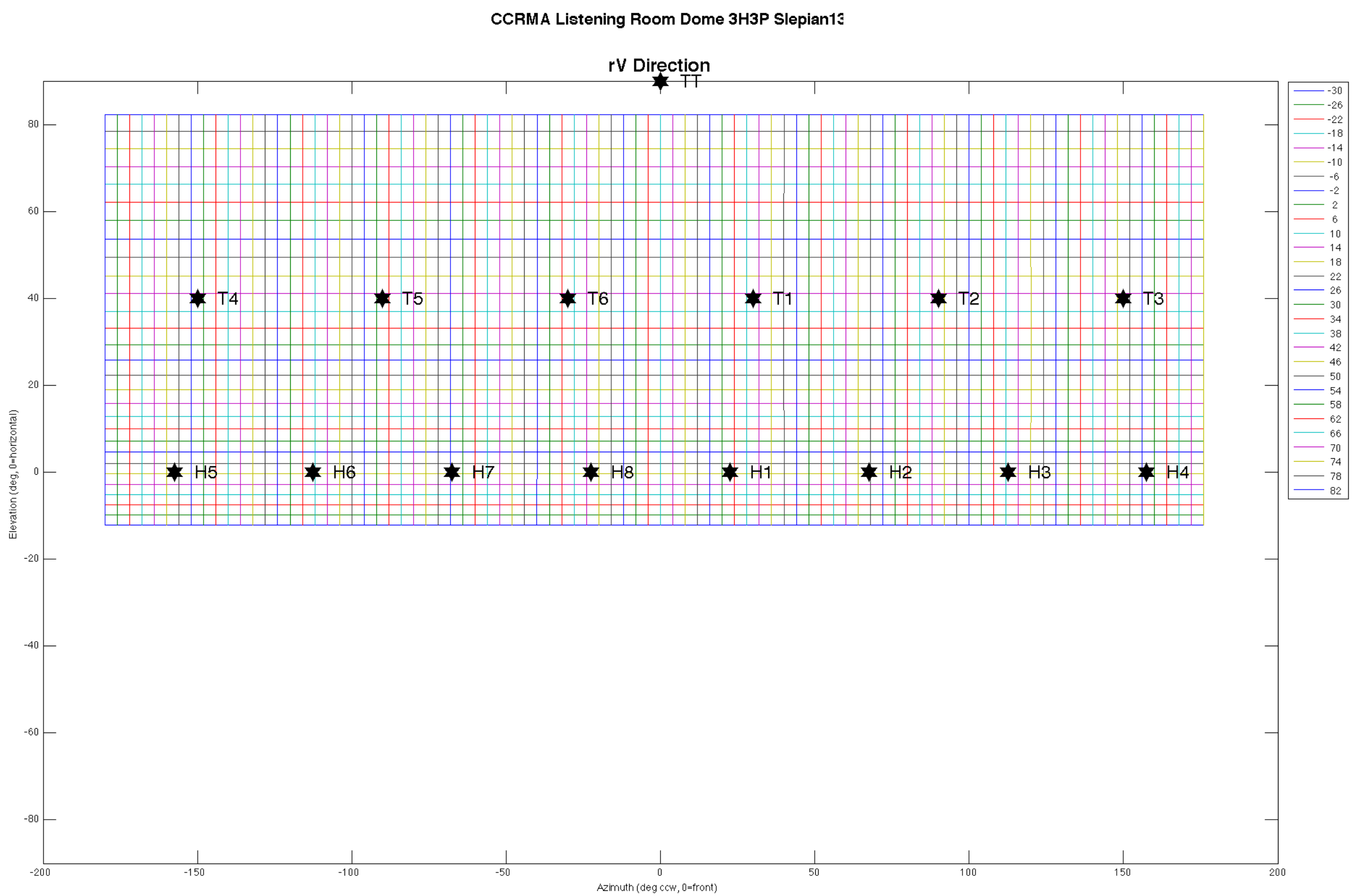






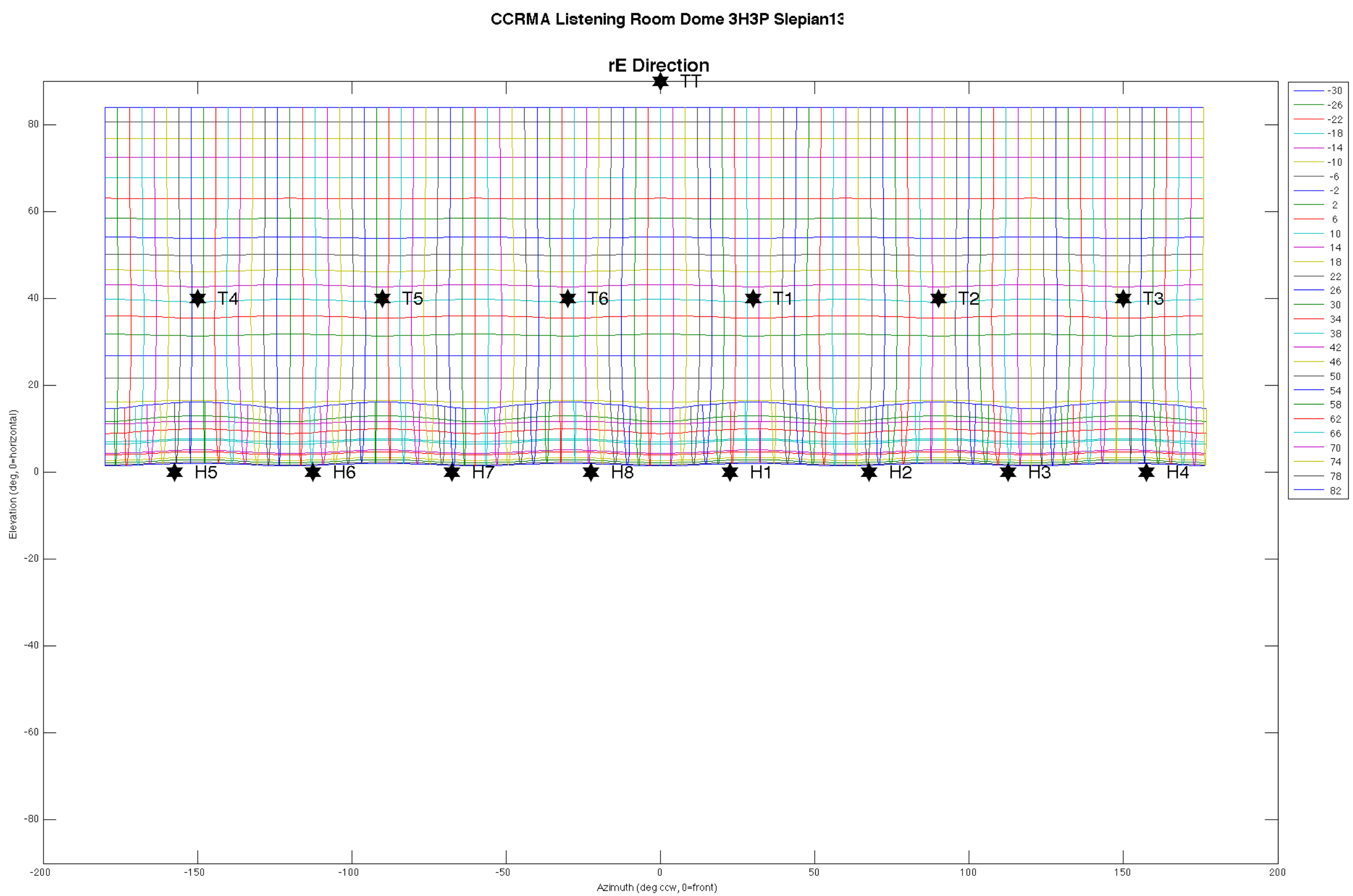
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

Spherical Slepian r_V direction grid



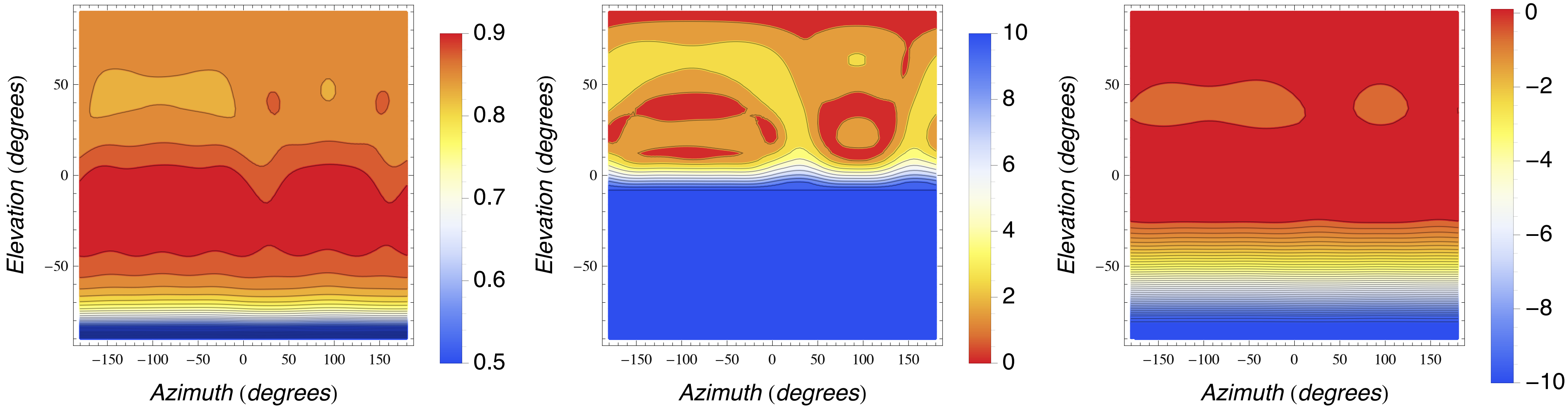
Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

Spherical Slepian r_E direction grid



Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

3rd order Hybrid Ambi-VBAP (AllRAD)

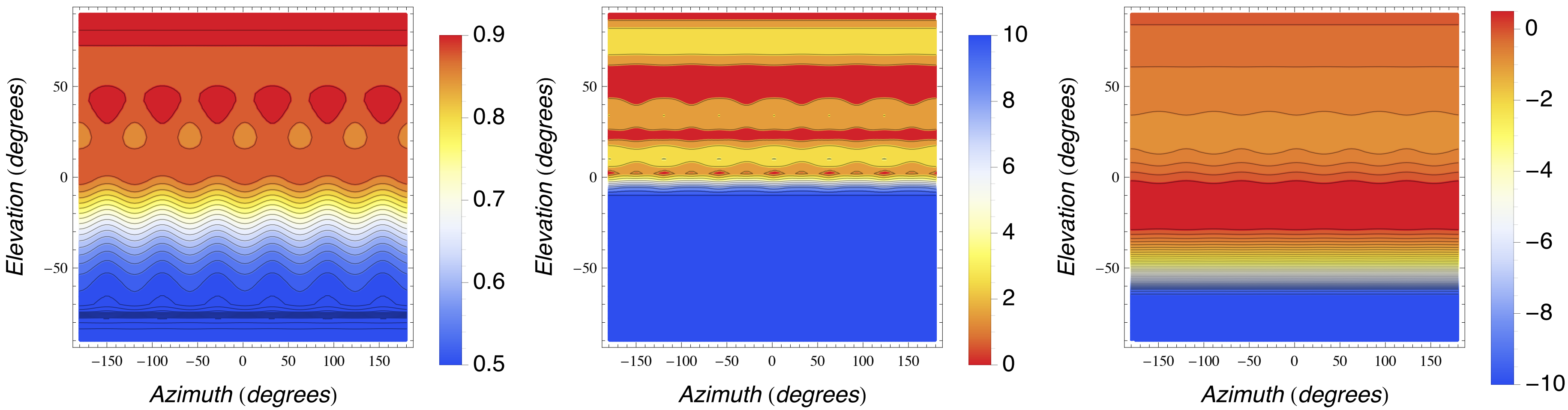


(a) r_E vs. Test Direction

(b) r_E Direction Error (degrees)

(c) Energy Gain (dB)

3rd order spherical Slepian function (EPAD)



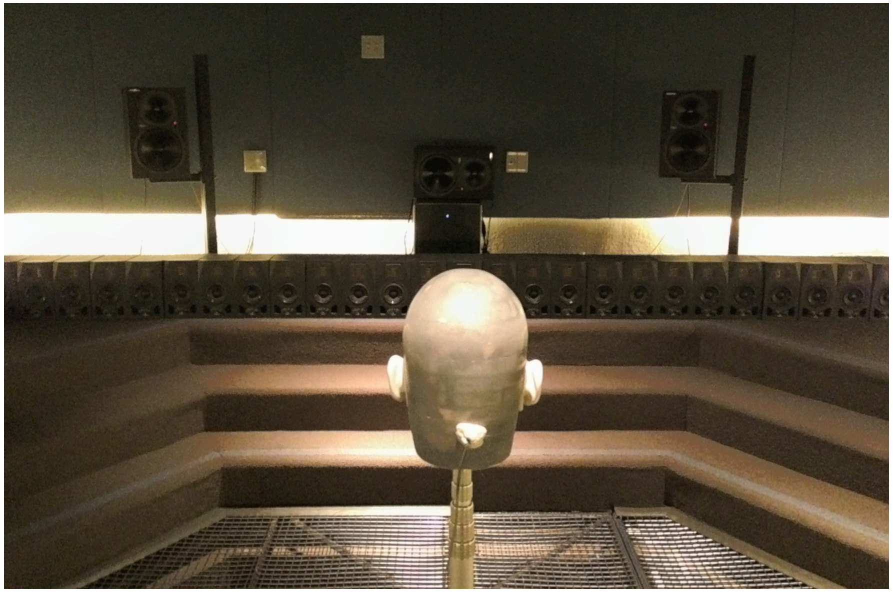
(a) r_E vs. Test Direction

(b) r_E Direction Error (degrees)

(c) Energy Gain (dB)

Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

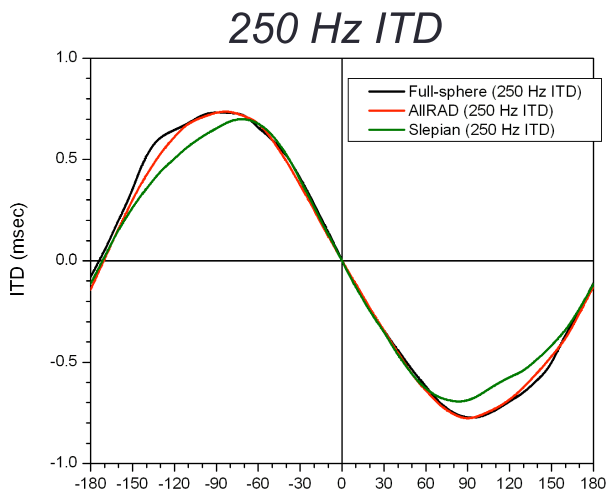
In situ performance measurements

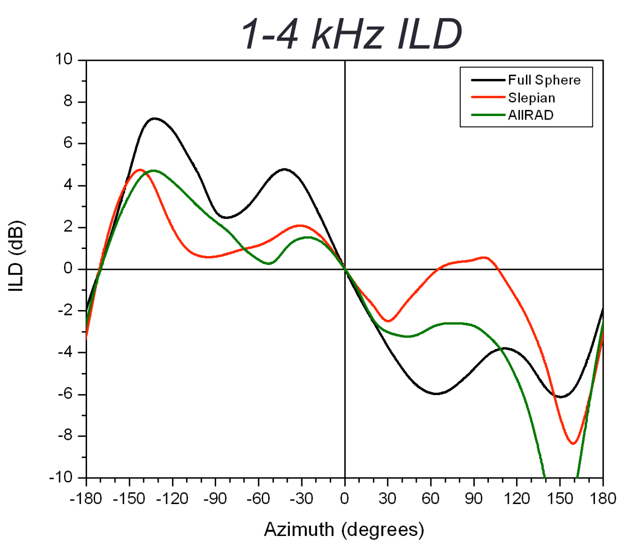


- Dummy head and reference omni
- Dome array using upper 15 speakers in CCRMA's listening room (8-6-1)

- Tested
  - AllRAD Dome
  - Spherical Slepian Dome
  - Full-sphere (from LAC2012)
- Collected
  - individual speaker IRs
  - Ambisonically panned IRs at 10° azimuth, 30° elevation intervals for each decoder
- Analyzed horizontal data
  - 250 Hz ITD (r_V)
  - 1 to 4 kHz ILD (r_E)

ITD and ILD measurements





Observations

- The measured ITDs were similar with the three decoders but ILDs were very different
- This supports the subjective observations that the three decoders sound different
- Detailed analysis is pending

Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

Informal listening tests

- 3rd-order test programs
  - Full-sphere mix of “Babel” by Allette Brooks (Jay Kadis)
  - *Chroma XII* by Rebecca Sanders (Jörn Nettingsmeier)
- Both dome decoders sounded good subjectively (but different!)
  - Compact and directionally accurate localization down to horizon
  - Faded below horizon
  - SSF decoder sounded brighter and more detailed than AllRAD
- Neither decoder sounded as good the full-sphere reference decoder
- 1st-order orchestral recording not reproduced well
  - Most of orchestra is below the horizon

Decoding Engine

- New decoding engine written in FAUST
  - No inherent limit on order
  - Dual band, NFC filters, distance compensation, …
- Toolbox writes out configuration section, appends implementation
- Compiles to LADSPA, LV2, Pd, Supercollider, VST, AU …
- Can be used independently of toolbox
- Drawback: Configuration “baked into” plugin

- Toolbox also writes out configuration files for
  - Kronlachner’s ambiX plugin suite
  - Adriaensen’s Ambdec

Implementation

- Toolbox runs in MATLAB and GNU Octave
  - Implements all known channel ordering and normalization conventions; both mixed-order conventions (HP and HV)
  - No inherent limit on Ambisonic order
  - Actively in use by a few beta testers
  - Mixed results for graphics output in Octave
  - Moving graphics output code to Python with MayaVi
  - Interface to IDHOA optimizer
- GNU Affero General Public License
- Faust decoder engine BSD 3-Clause License
- Git repo at https://bitbucket.org/ambidecodertoolbox/adt

Summary and Conclusions

- Extensions to Ambisonic Decoder Toolbox to handle speaker configurations that do not cover full sphere
- New decoder engine in written in Faust
- Ability to generate decoders quickly has proven valuable in performance settings

- Plans
  - Dual-band AllRAD and Slepian decoders
  - Optimizer to refine decoders

- Open question:
  - What to do when sources move into areas of poor coverage.
    - Current implantation fades them out.
    - Decorrelate and mix into other speakers?
    - Should transmission standards include “rendering hints”?

Thanks!

- Fernando Lopez-Lezcano for helping with the listening tests and in-situ measurements, and overall feedback and encouragement.
- Andrew Kimpel, Marc Lavallée, and Paul Power who are active users.
- Richard Lee, Jörn Nettingsmeier, and Bob Oldendorf who read early drafts and provided feedback.
- LAC 2014 reviewers and organizers

Human Auditory Localization

- At low frequencies (up to about 800 Hz) works by Interaural *Time* Differences (ITDs)
- At middle frequencies (800 Hz to 5 kHz) works by Interaural *Level* Differences (ILDs)
  - Transition is fairly sharp
  - due to the ITDs becoming ambiguous once the wavelength become smaller than ear spacing.
- 2-channel stereo doesn't get it right
  - ILD cues are such that the images tend to stick to nearest speaker
- Ambisonics was designed from the beginning to get this correct with modest resources.
  - Small number of program channels and loudspeakers

Gerzon's Theory of Auditory Localization

- Early workers in stereo did theoretical analysis showing how stereo did (or didn't) provide proper localization cues
- Gerzon's contribution was to integrate those theories and came up with a theory that defined
  - $\mathbf{r}_V$, the vector sum of the signals from the loudspeakers
  - $\mathbf{r}_E$, the vector sum of the squares of the signals from the loudspeakers.
- By providing a simple mathematical encapsulation, we can use these to
  - design decoders
  - prove theorems, e.g., polygonal decoder theorem
  - help understand what various spatial sound reproduction systems can and cannot do

Localization Vector Theory

- $\mathbf{r}_V$ predicts low-frequency localization almost perfectly.
  - If $\mathbf{r}_V$=1, then low-frequency sounds will be precisely located.
- $\mathbf{r}_E$ predicts mid-frequency localization moderately well.
  - If $\mathbf{r}_E$=1, then mid-frequency localization will be good
  - BUT… $\mathbf{r}_E$ is always less than1, unless the sound is coming from a single point source.
  - At best $\mathbf{r}_E = \cos(\theta/2)$, where θ is the angle between the loudspeakers, so for a square array $\mathbf{r}_E \leq 0.707$.
  - In general, $\mathbf{r}_E$ is low in directions with few loudspeakers
  - Best we can do is have it change smoothly in performance from dense areas to sparse areas.

Heller, Benjamin, The Ambisonic Decoder Toolbox, Linux Audio Conference 2014, ZKM, Karlsruhe, Germany

Energy Localization Vector

- Maximizing $\mathbf{r}_E$ *and* getting it to point in the right direction is the crux of the decoder design problem.
  - Easy with regular arrays
  - Irregular arrays always involve tradeoffs
  - Virtually all real world arrays are irregular!
    - Arrays need to fit in real rooms
    - ITU 5.1 is the dominant domestic standard, rear speakers 120° apart.
- Because it is a non-linear function of speaker position, we currently need to use numerical optimization methods.