How High is Enough?
Strategies for Ambisonics Immersive Audio over Headphones

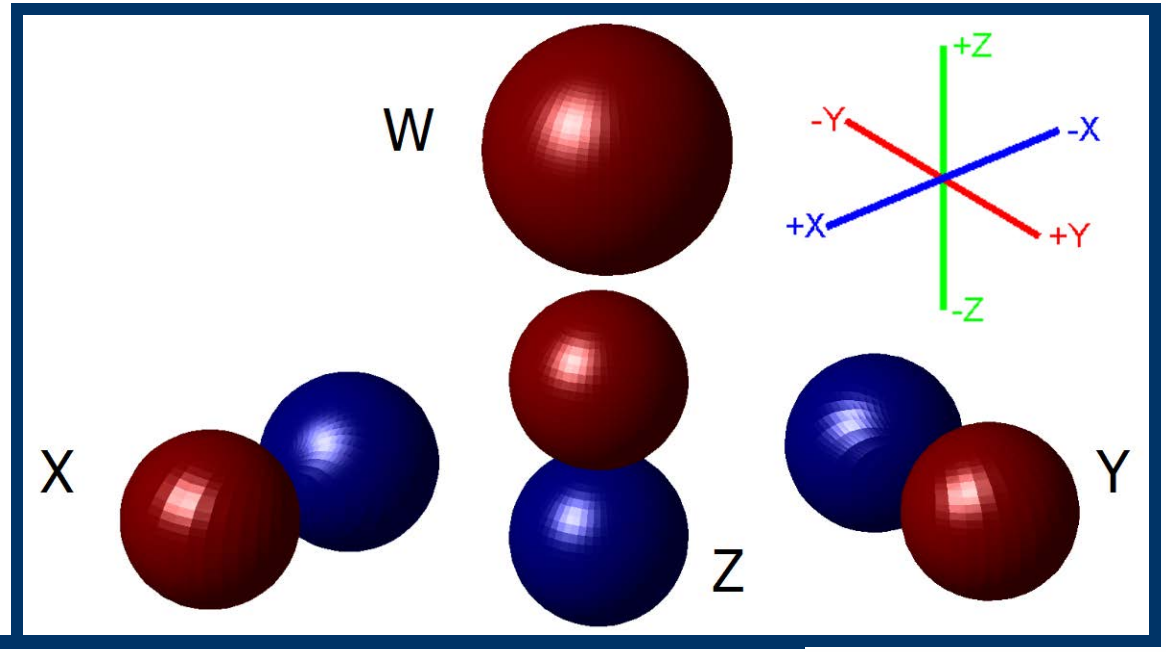
Dr Bruce Wiggins and Mark Dring
University of Derby
Sounds in Space 2018

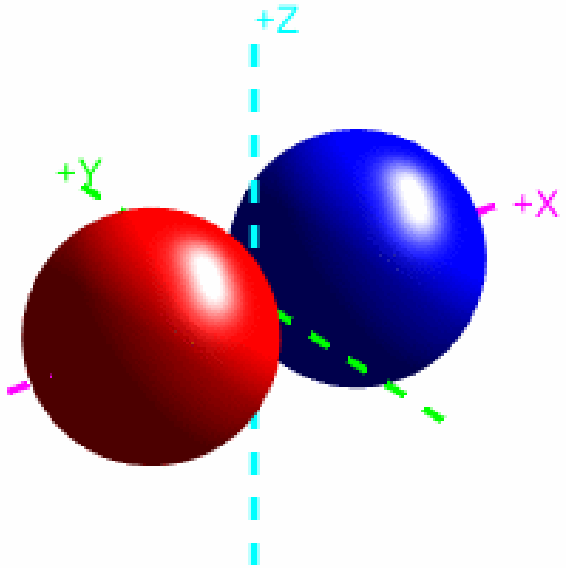
Spatial Audio for VR

- Google implemented Spatial Audio for 360 videos at the end of April, 2016
- https://github.com/google/spatial-media
  - A collection of specifications and tools for 360° video and spatial audio
- Currently implemented in the YouTube Android application (binaural) and on chrome on the desktop (virtual microphone)
  - And the off-line Jump Inspector (now discontinued!)
- Binaural Audio is delivered to the user…
- …via 1st to 3rd order Ambisonics
  - YouTube app (currently 1st order)
  - Off-line Jump Inspector App (currently 3rd order)

Why Ambisonics – Head Tracking?

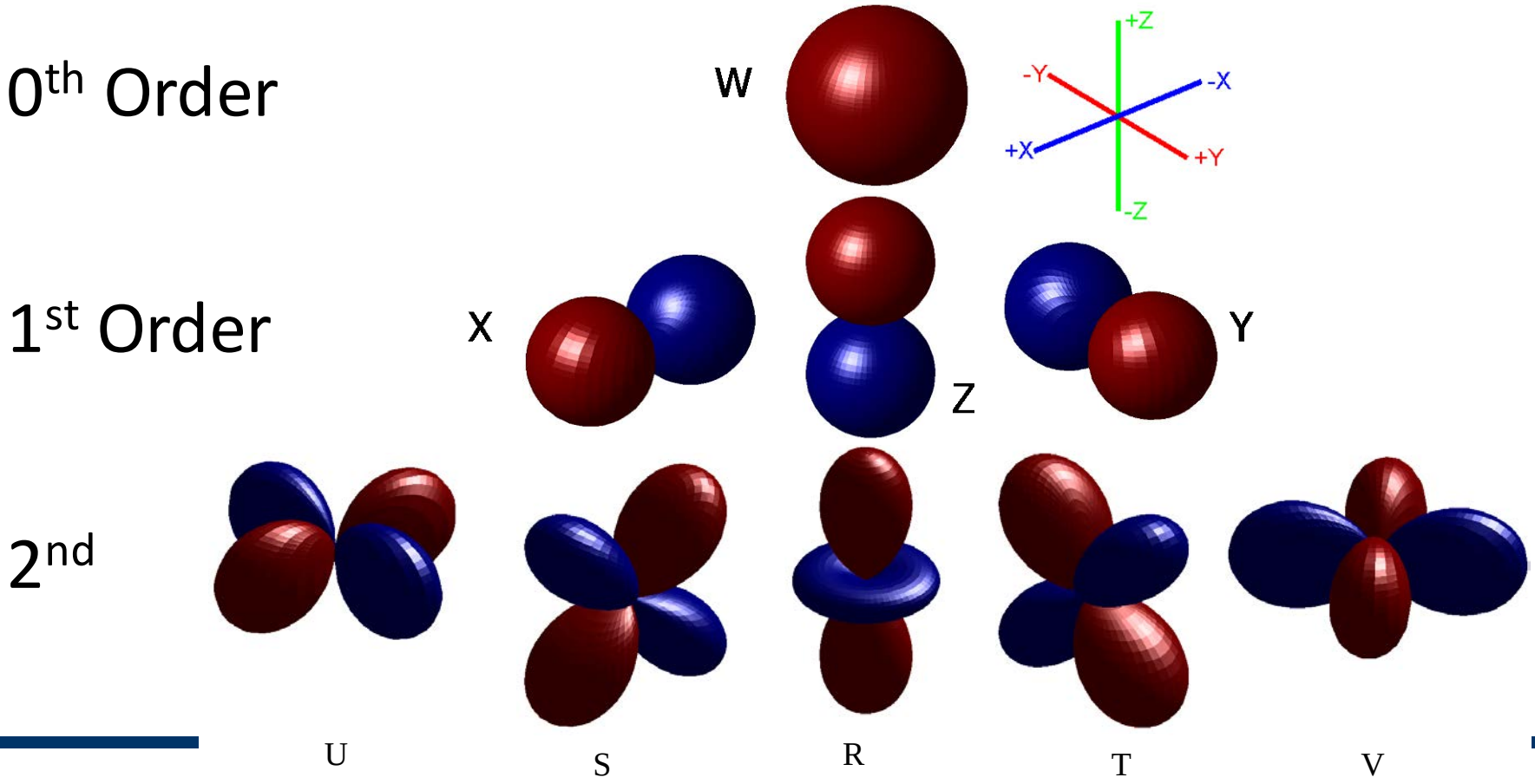
- Reducing the number of channels needed to represent a full sphere audio scene
- Straight forward to convert to a Binaural output
- Rotation of the whole scene is also straight forward (correct for head movement)



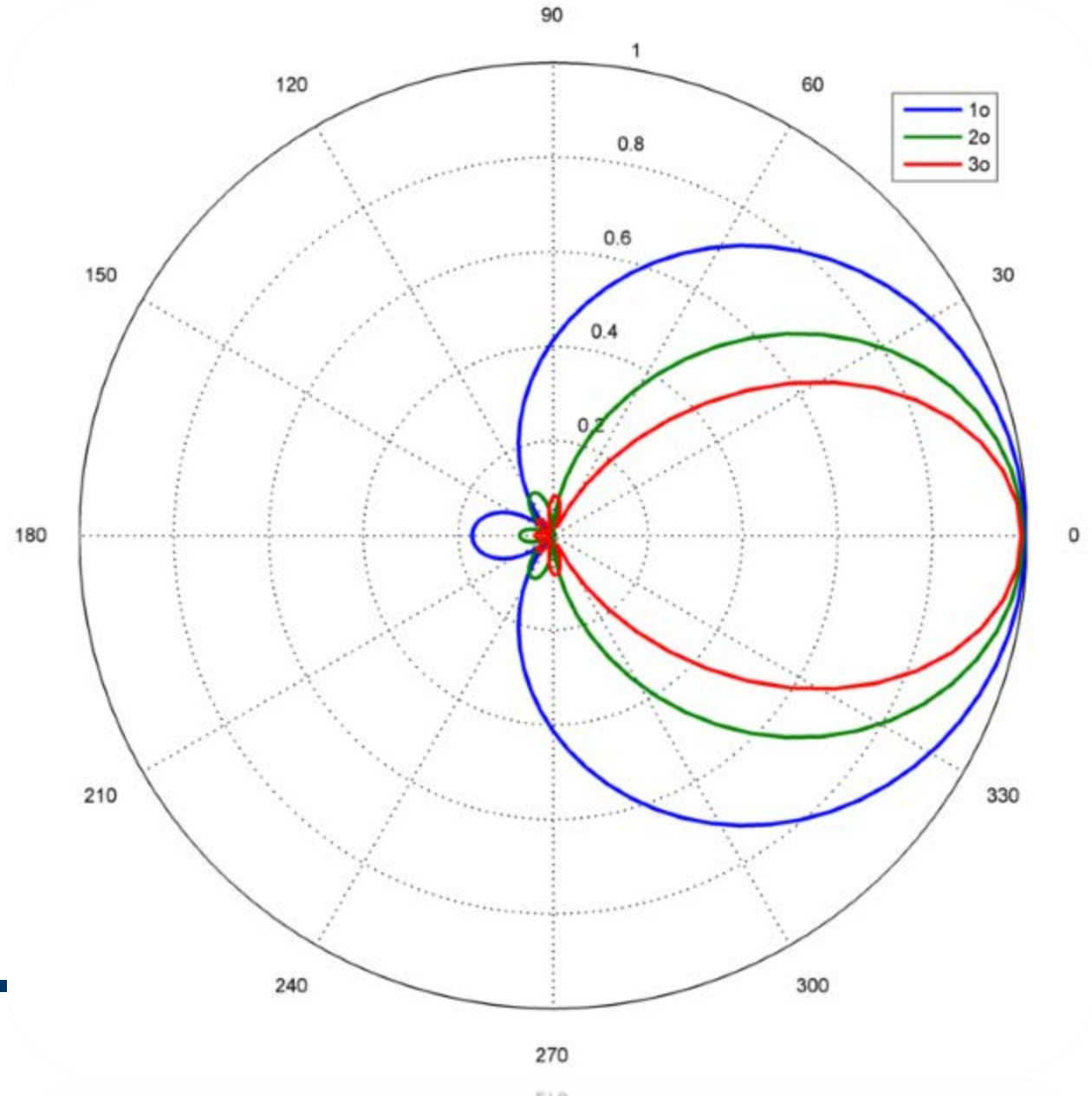


Higher Order Ambisonics

- Uses more input signals…
- …which can result in better control of the speaker feeds and, hence, reproduced sound field.



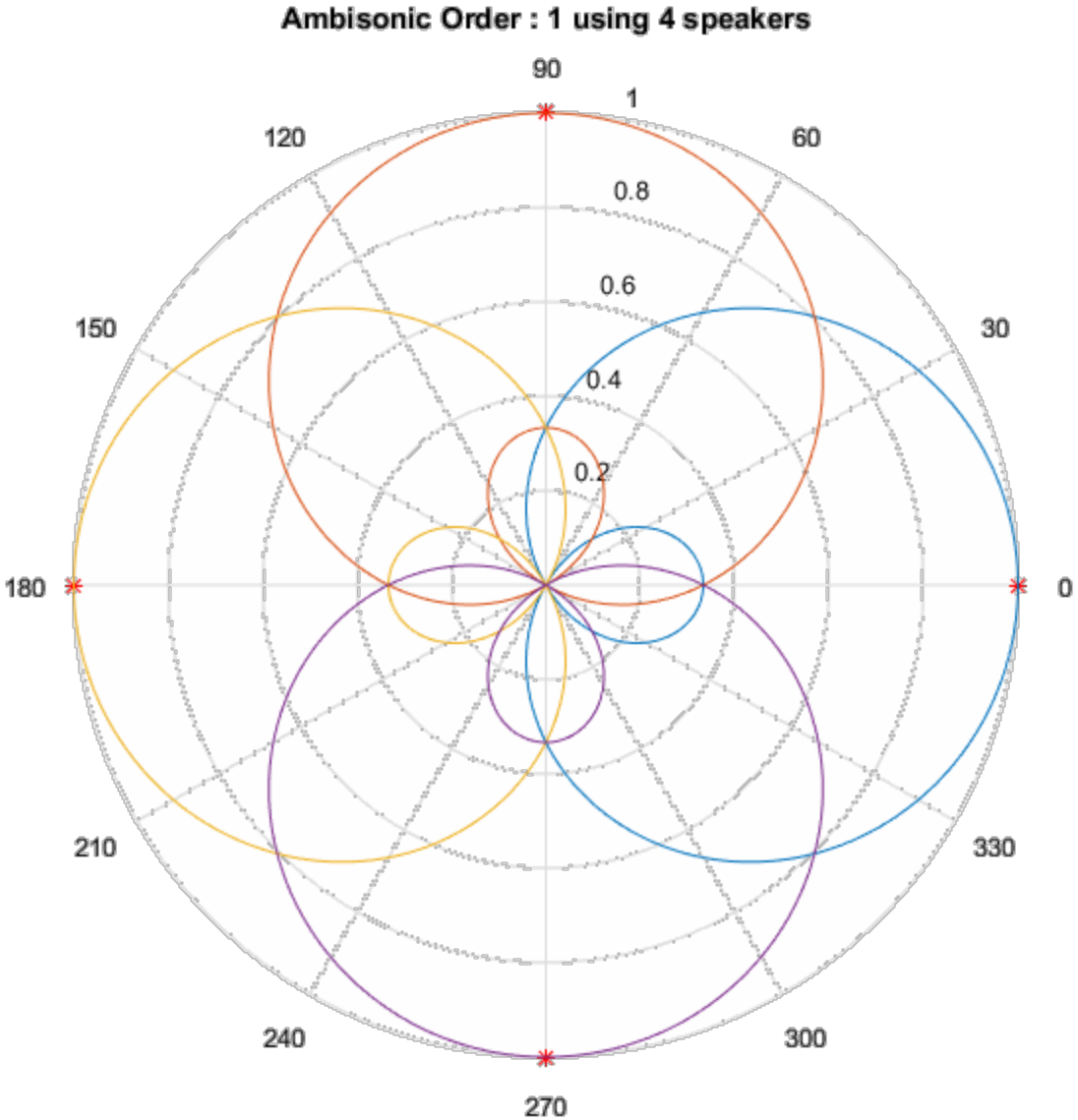
Polar Pattern Choice



Spatial Sampling

- In order to recreate, spatially, the sound field, it is 'sampled' from:
  - A number of different directions
  - Using spherical harmonics (see polar patterns in previous slide)
- The more the samples, or the higher the sampling rate (Ambisonic order)…
- …the higher the frequency of 'correct' spatial reproduction
- Radius of *correct* reconstruction can be approximated to:

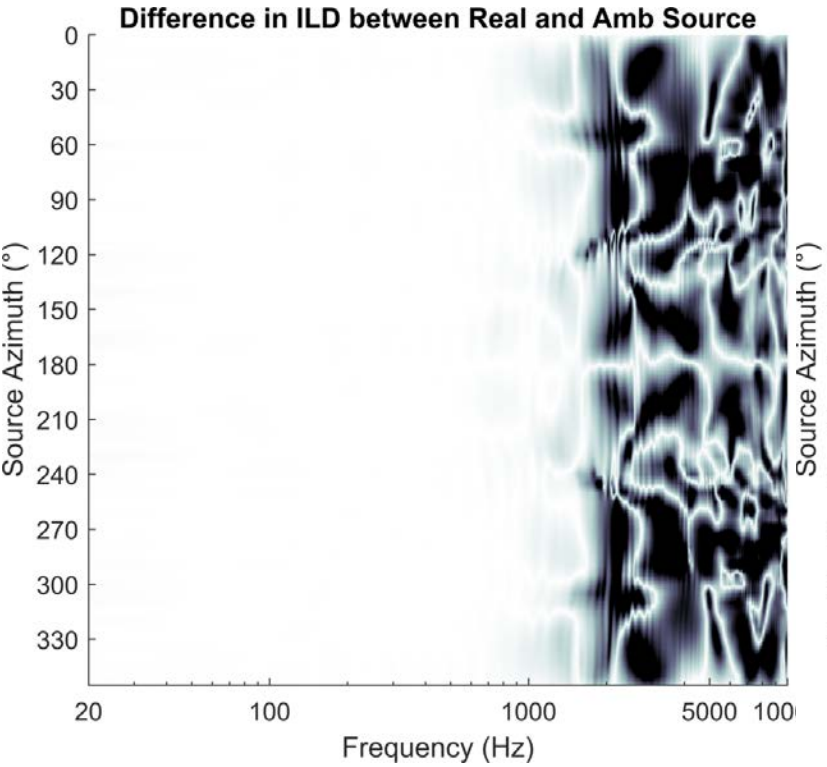
$$f \approx \frac{nc}{2\pi r}$$ where n=order, c=speed of sound, f=freq, r=radius



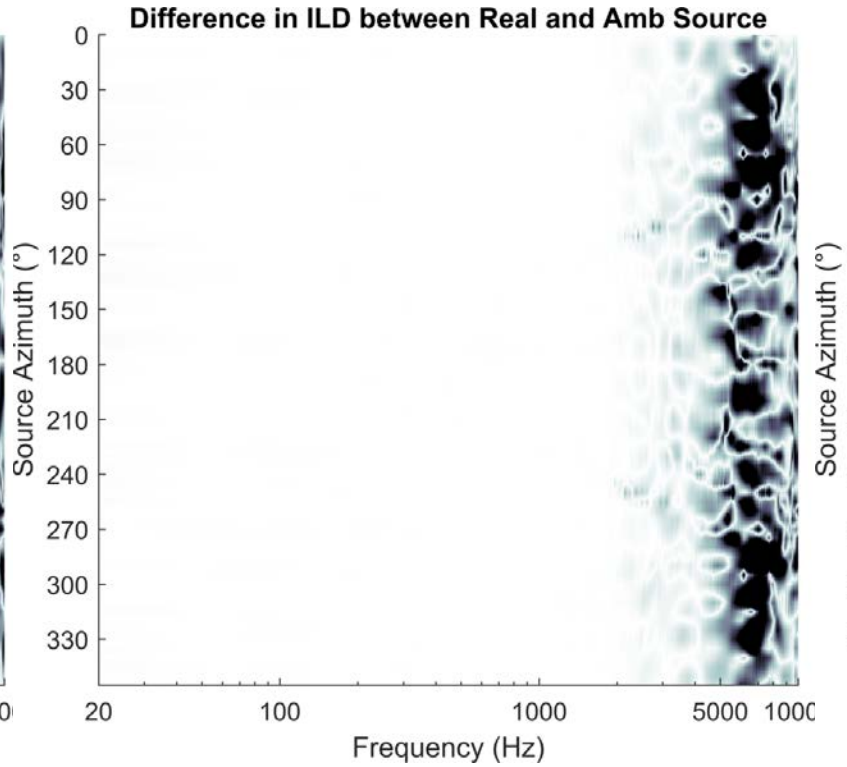
Higher Order Ambisonics

- Increasing the order of Ambisonics (n) will increase the frequency to which correct operation will occur (for both **I**nteraural **T**ime and **I**nteraural **L**evel **D**ifferences)
- Also increases the number of speakers/samples needed $\left(2 \times (n+1)\right)$ horizontal only

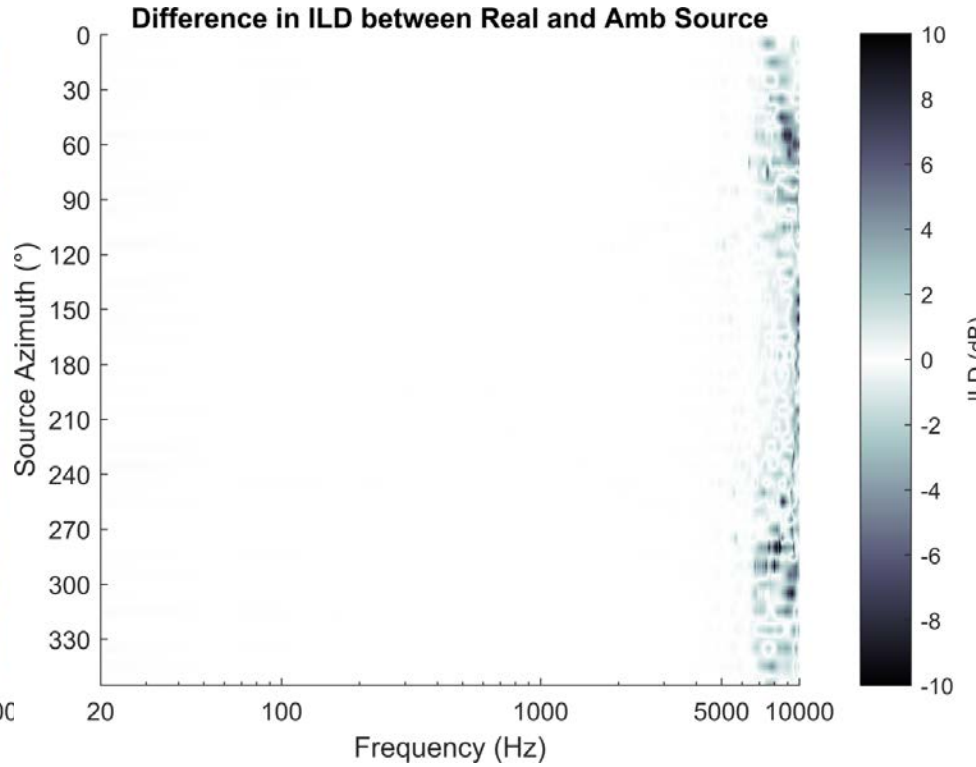
3rd Order



8th Order



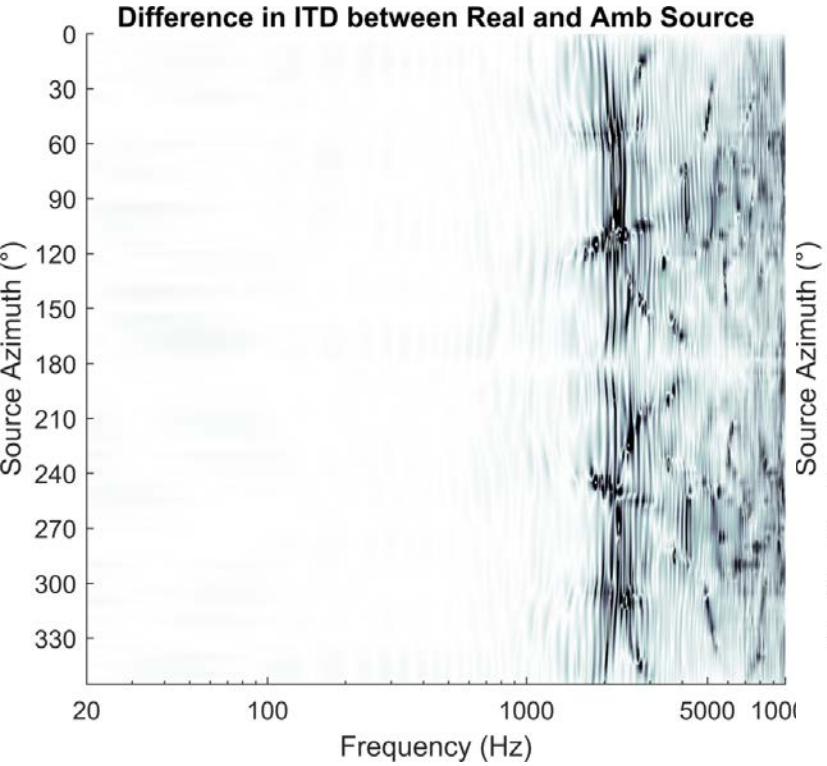
17th Order



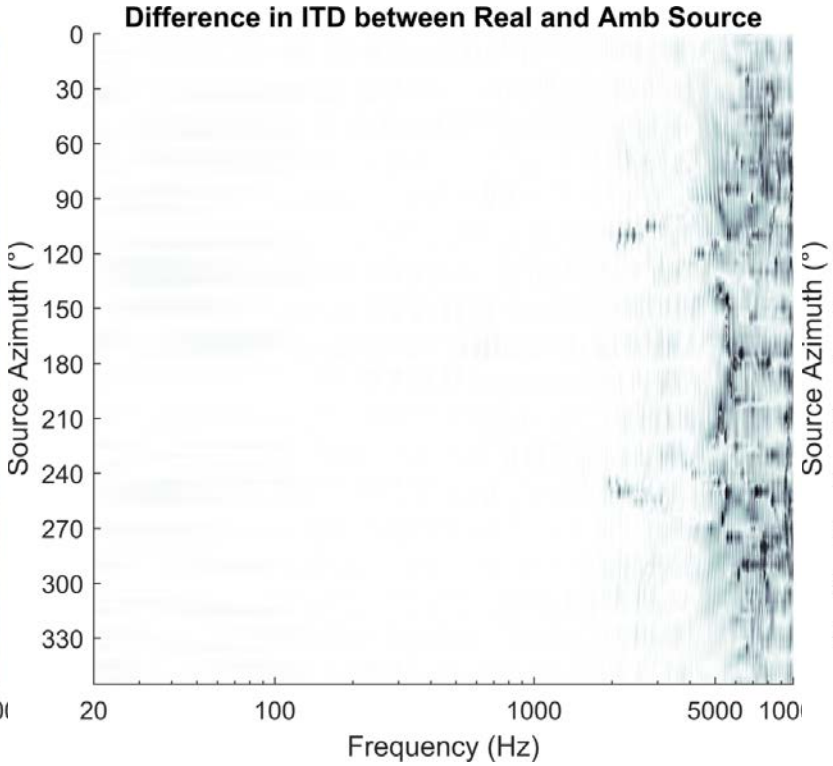
Higher Order Ambisonics

- Inter-aural Time Difference (ITD)

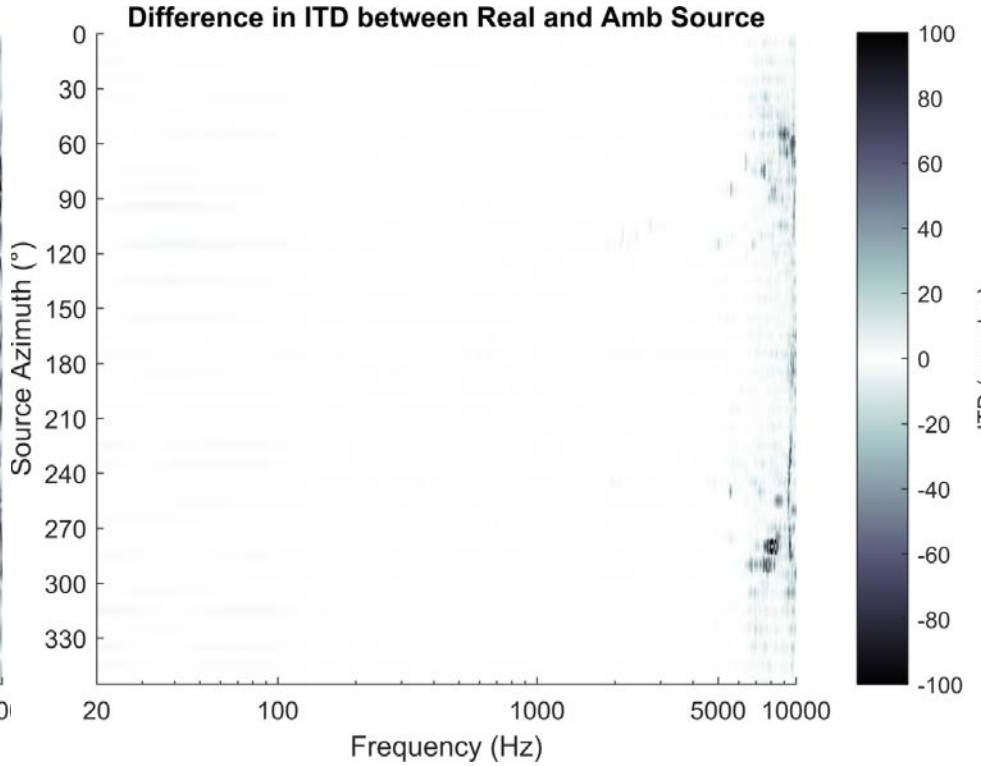
3rd Order



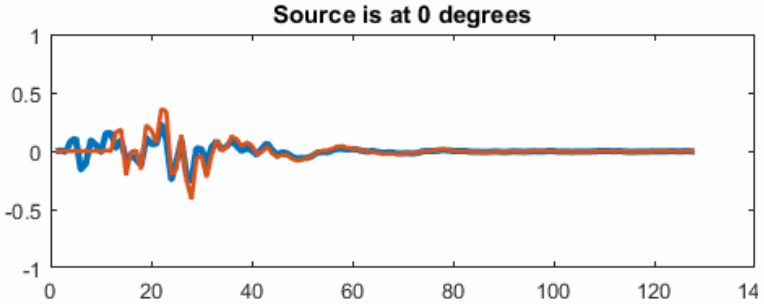
8th Order

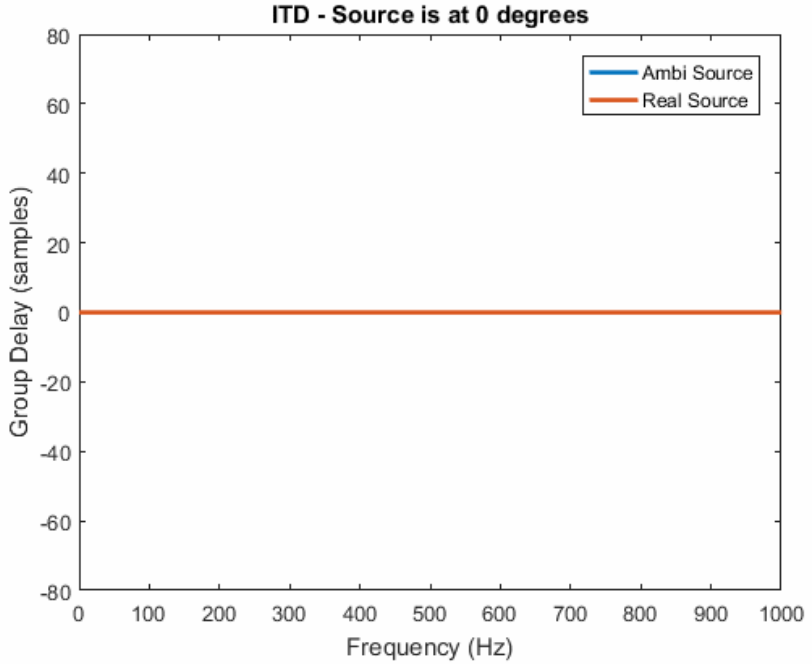


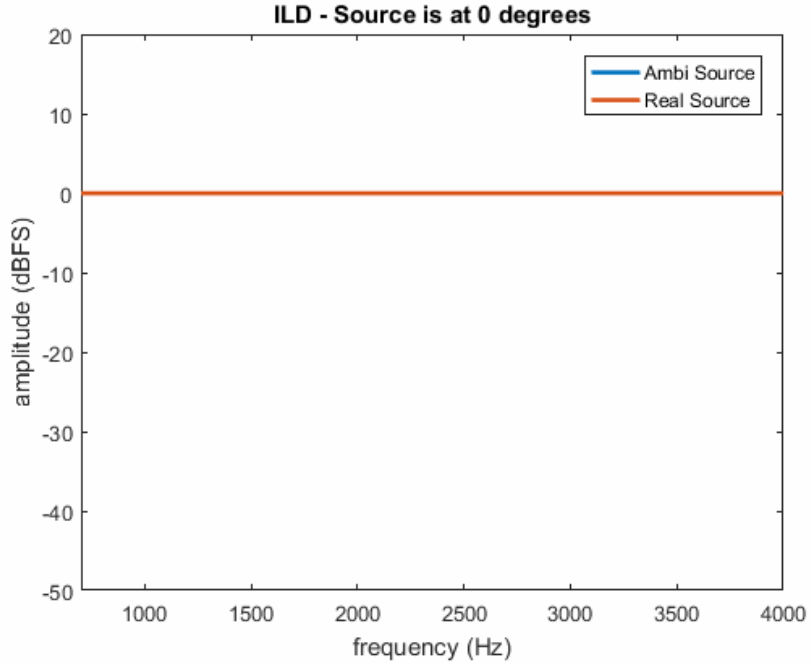
17th Order

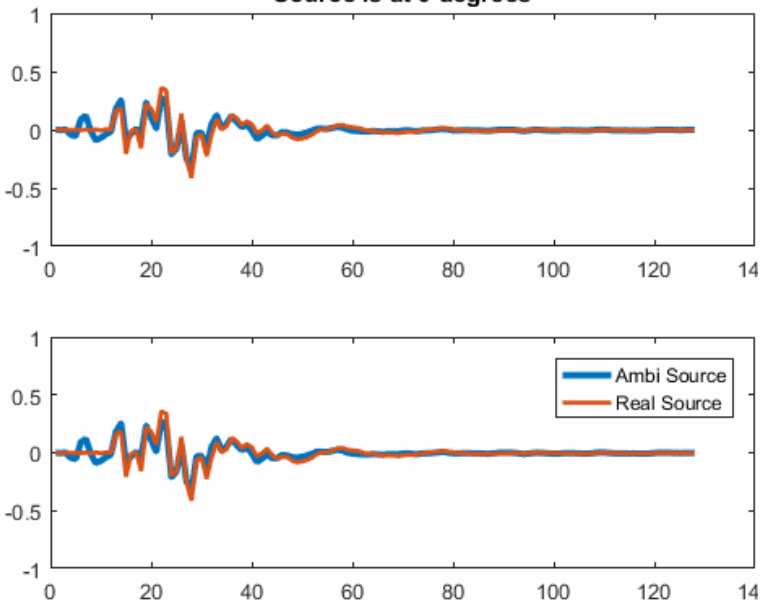


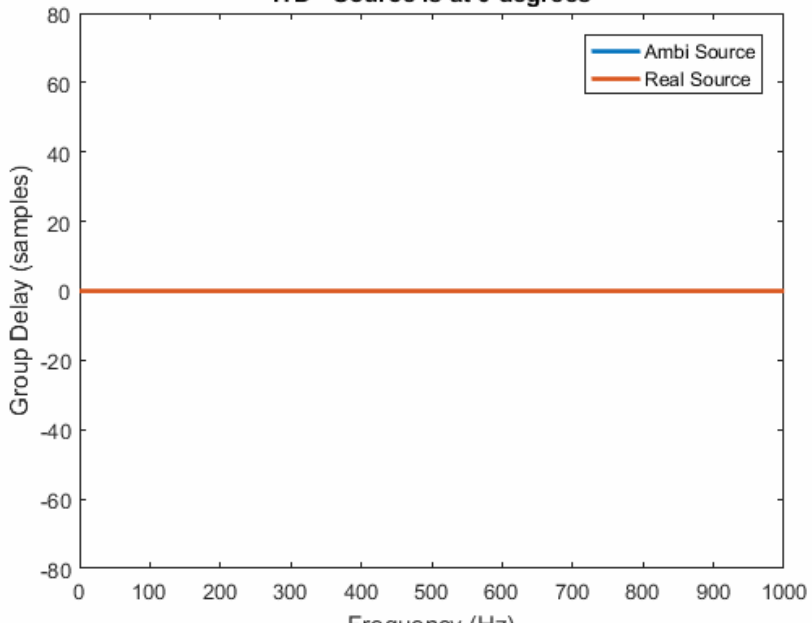
HRIR and HRTF data – 1st and 3rd order

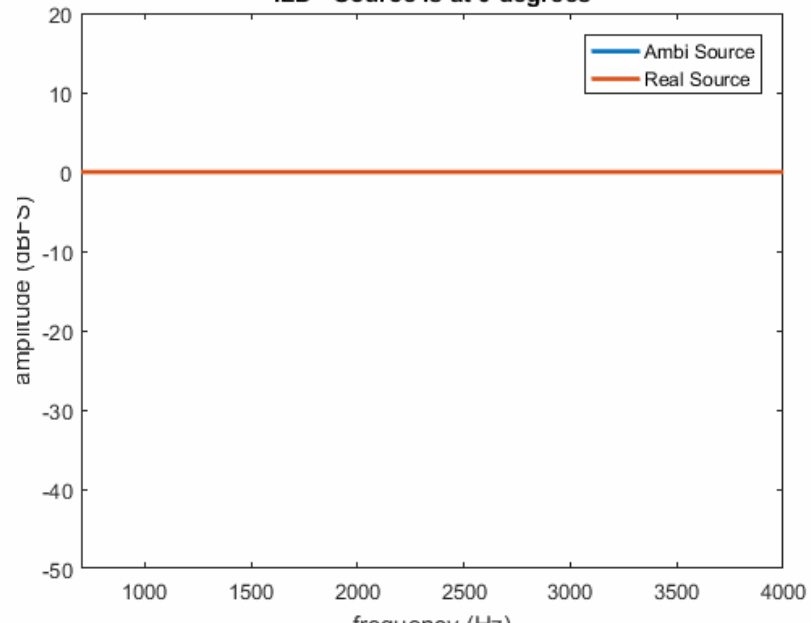




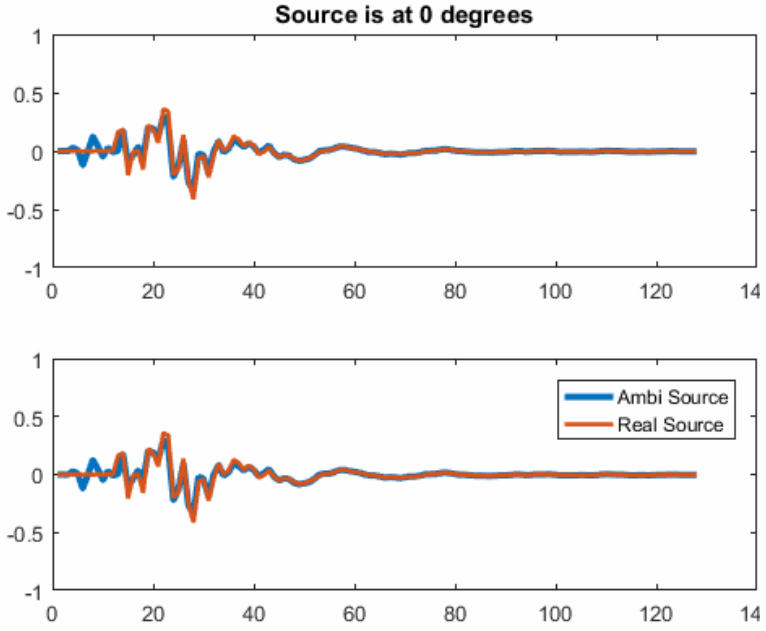


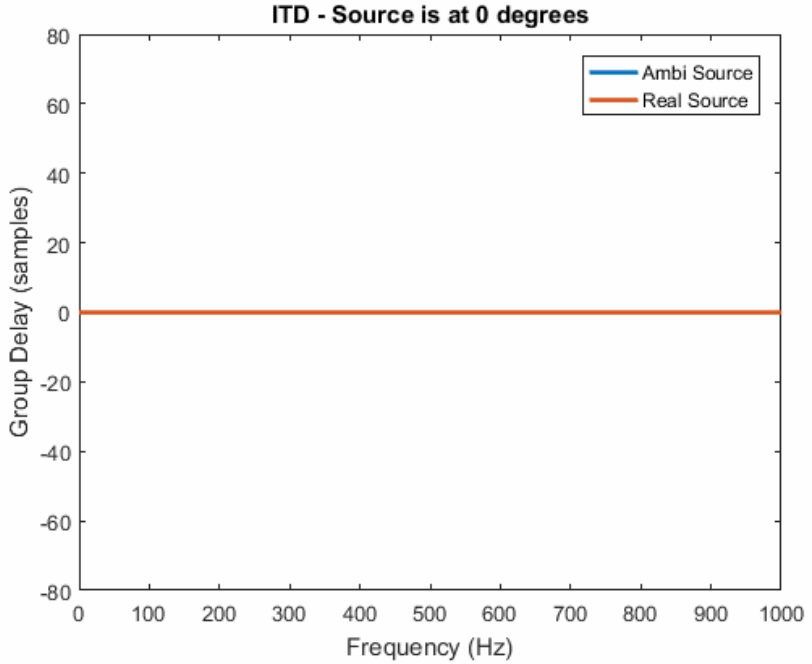


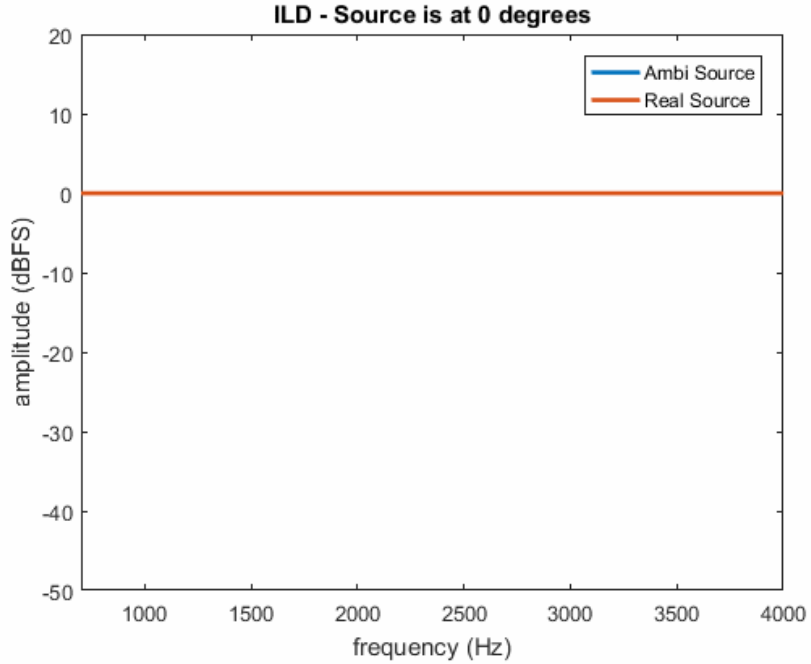


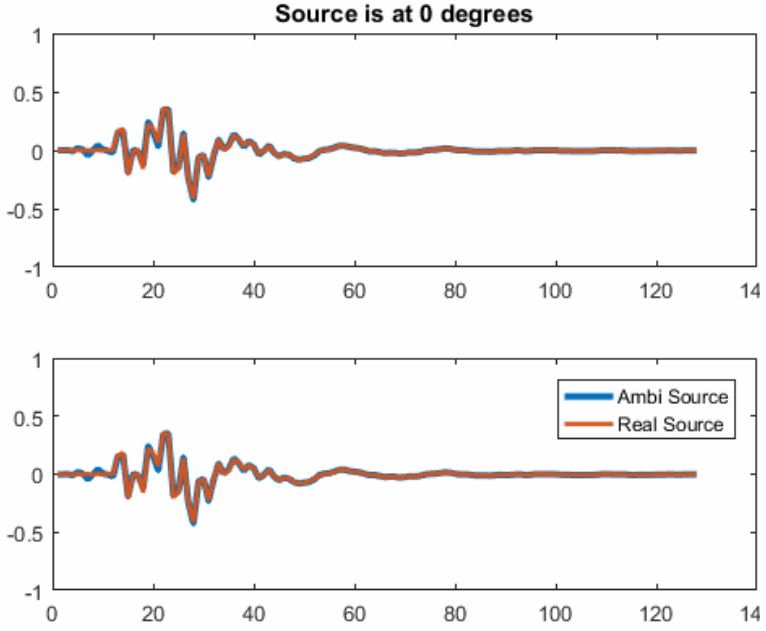


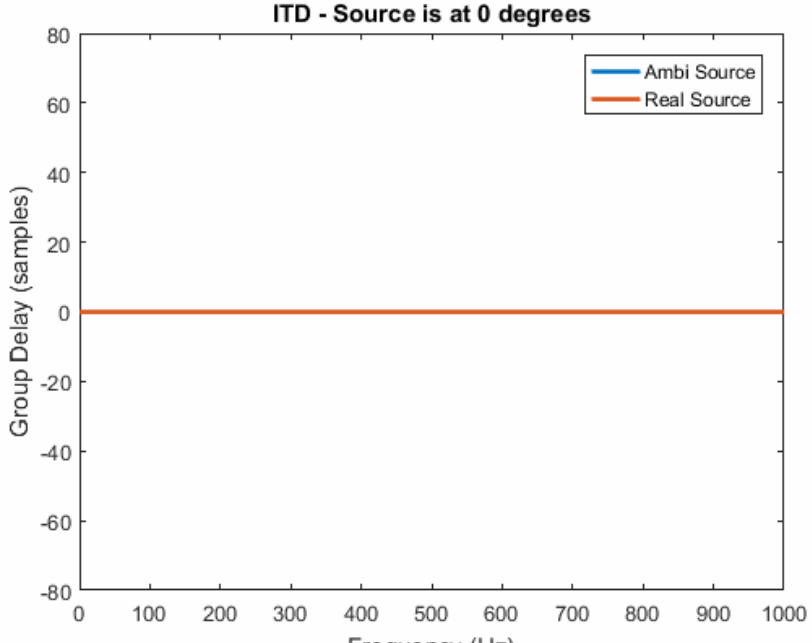
HRIR and HRTF data – 5th and 8th Order

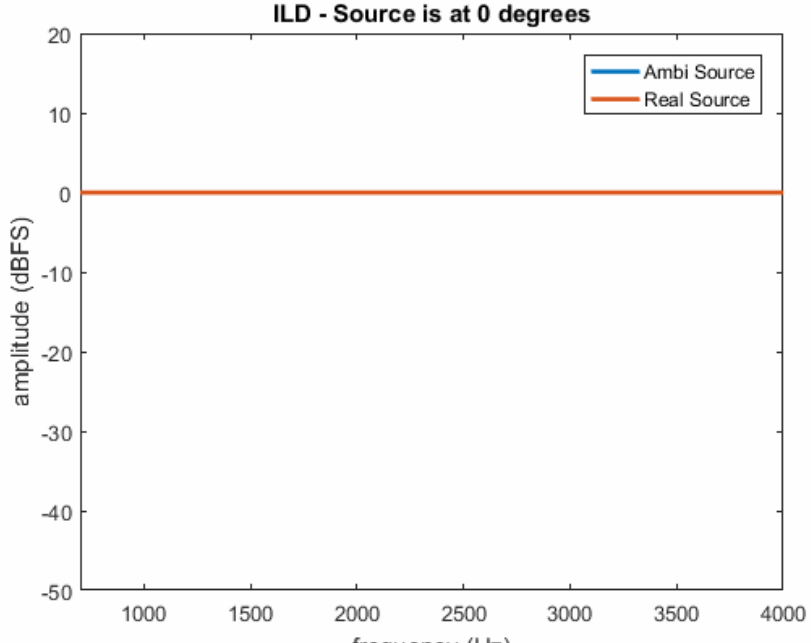






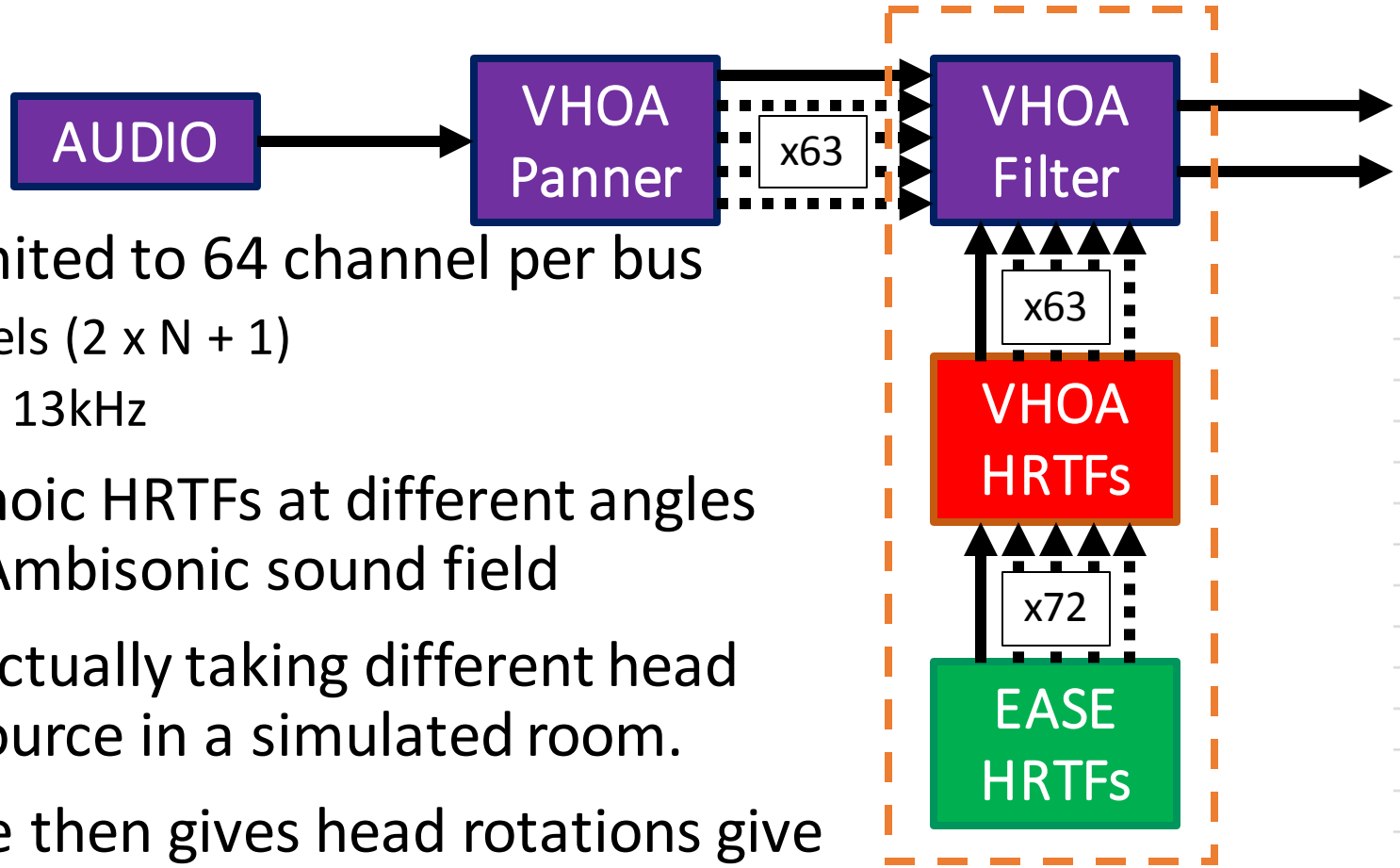






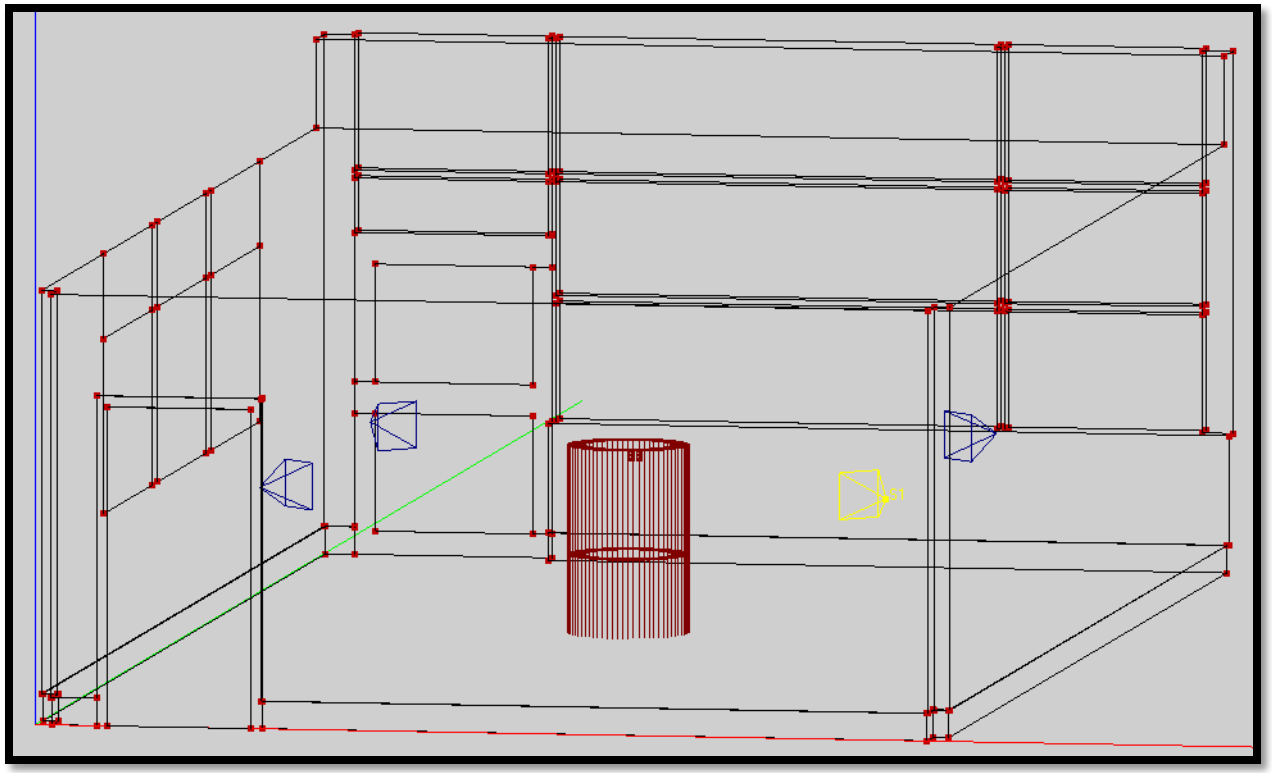
Method

- Originally working up to 35th order (71 channels)
- However, Reaper is limited to 64 channel per bus
  - 31st Order is 63 channels (2 x N + 1)
  - 'Correct' up to around 13kHz
- Usually, VR uses anechoic HRTFs at different angles and then ***rotates*** the Ambisonic sound field
- In this system, we're actually taking different head rotations to a single source in a simulated room.
- Re-***panning*** the source then gives head rotations give correct room response.

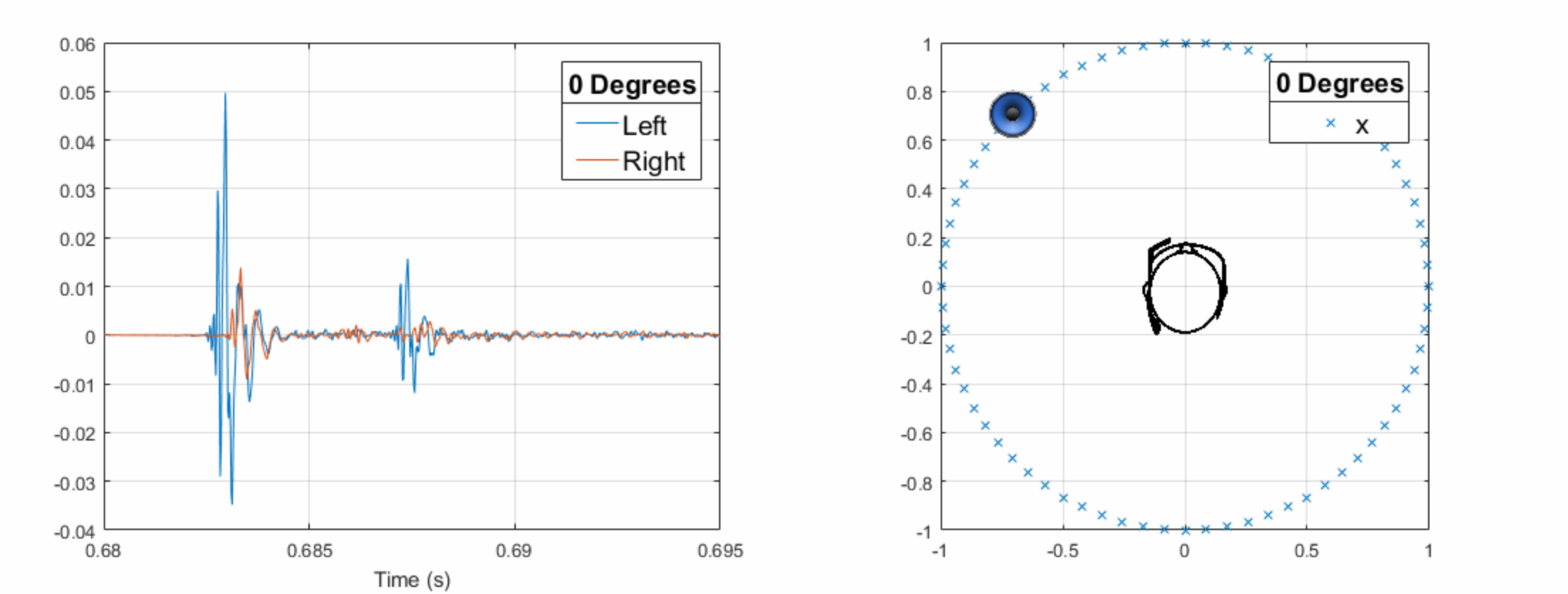


Order	Freq
1	419
3	1256
5	2093
7	2931
9	3768
11	4606
13	5443
15	6280
17	7118
19	7955
21	8793
23	9630
25	10467
27	11305
29	12142
31	12980
33	13817
35	14654

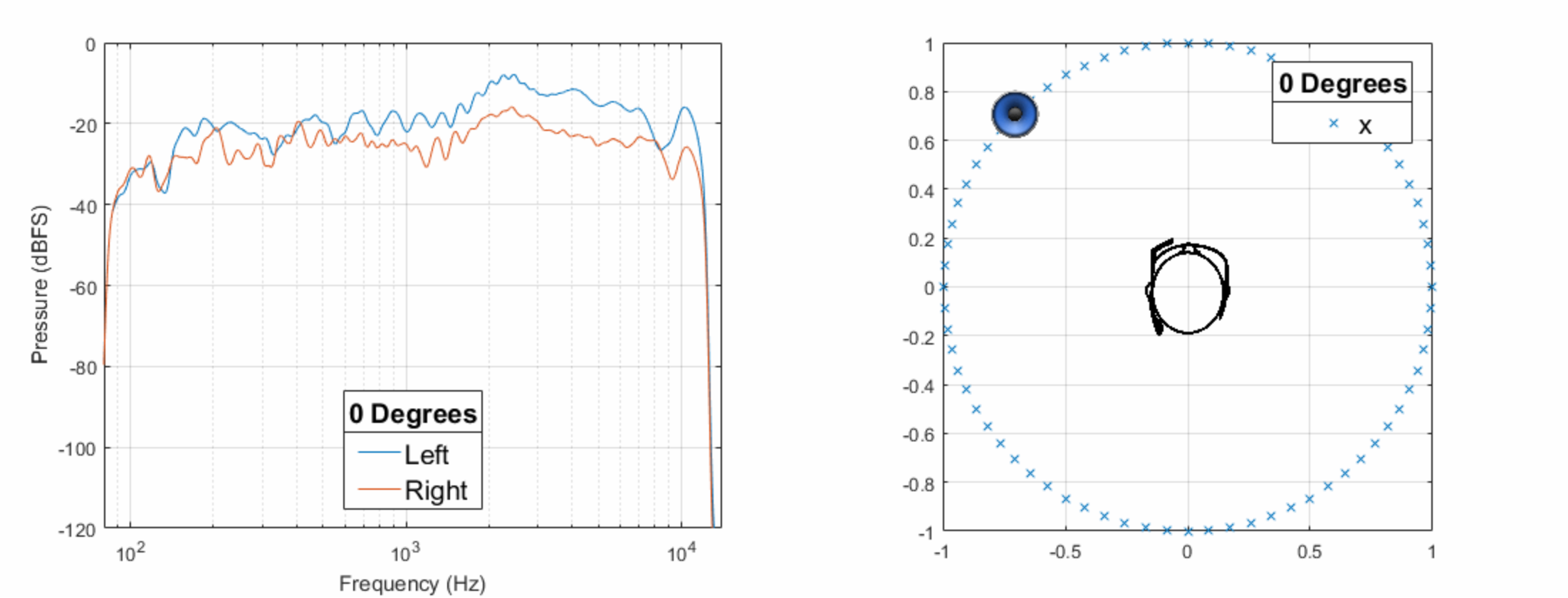
Room Modelling in EASE



EASE Generated HRTFs

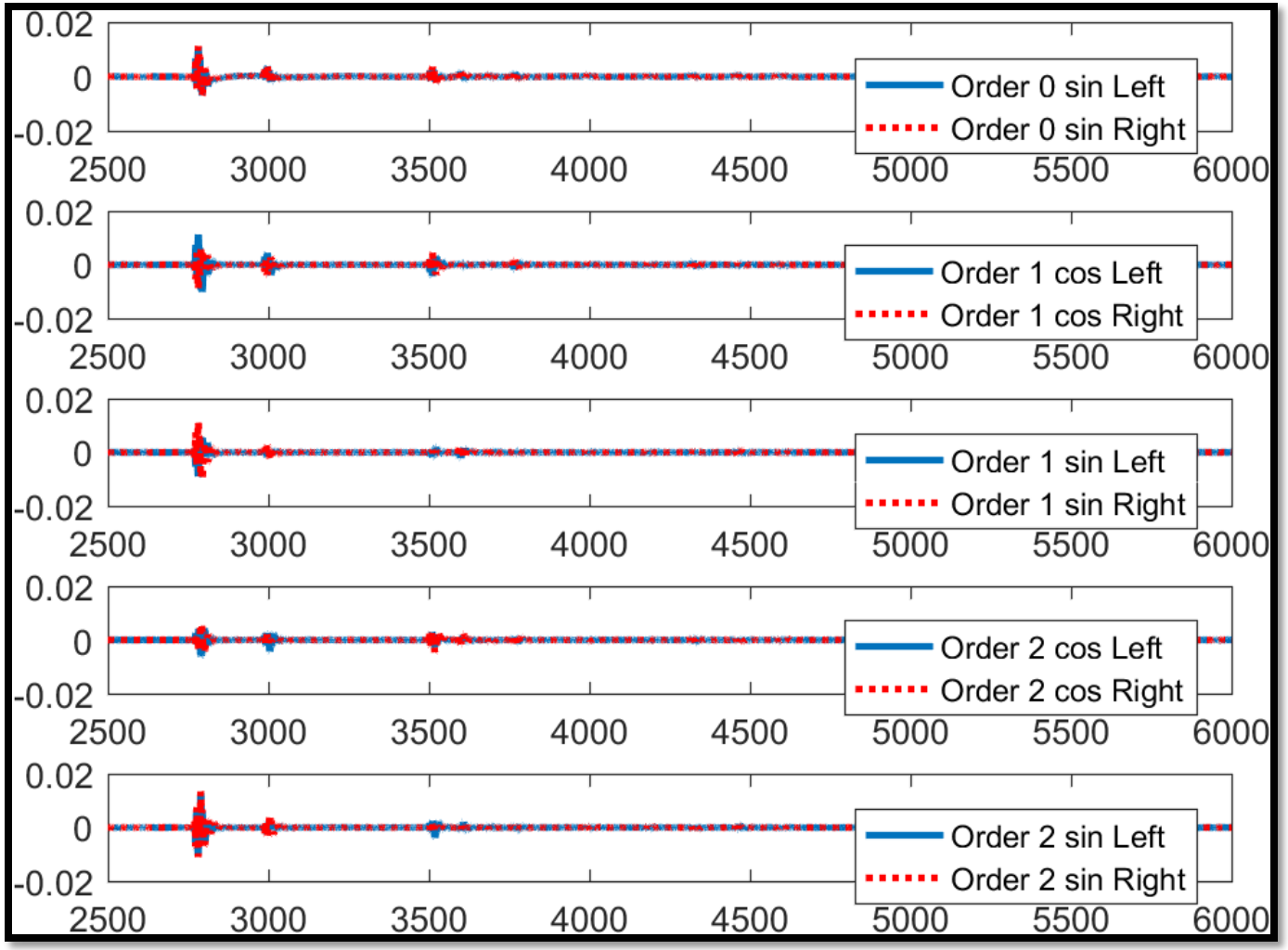


EASE Generated HRTFs



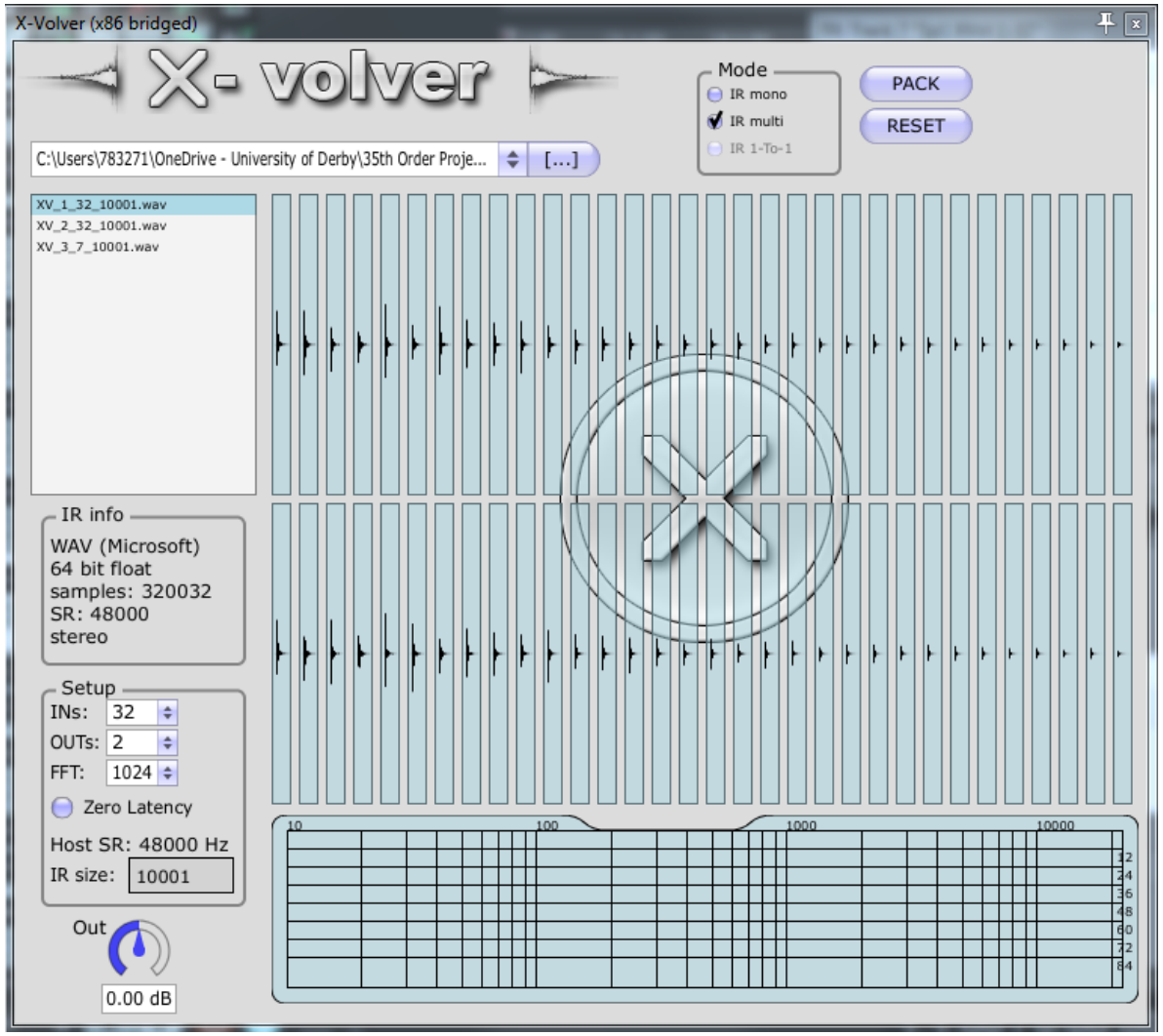
Ambisonic HRTF Generation

- Matlab Script.
- Receive 72 binaural HRTFs for a single speaker location.
- Calculate required spherical harmonic decoder values for every 5 degrees head rotation.
- Determine horizontal (X and Y) ambisonic HRTFs up to 35th order.
- Concatenate left and right signals independently into single impulse response to be used in…

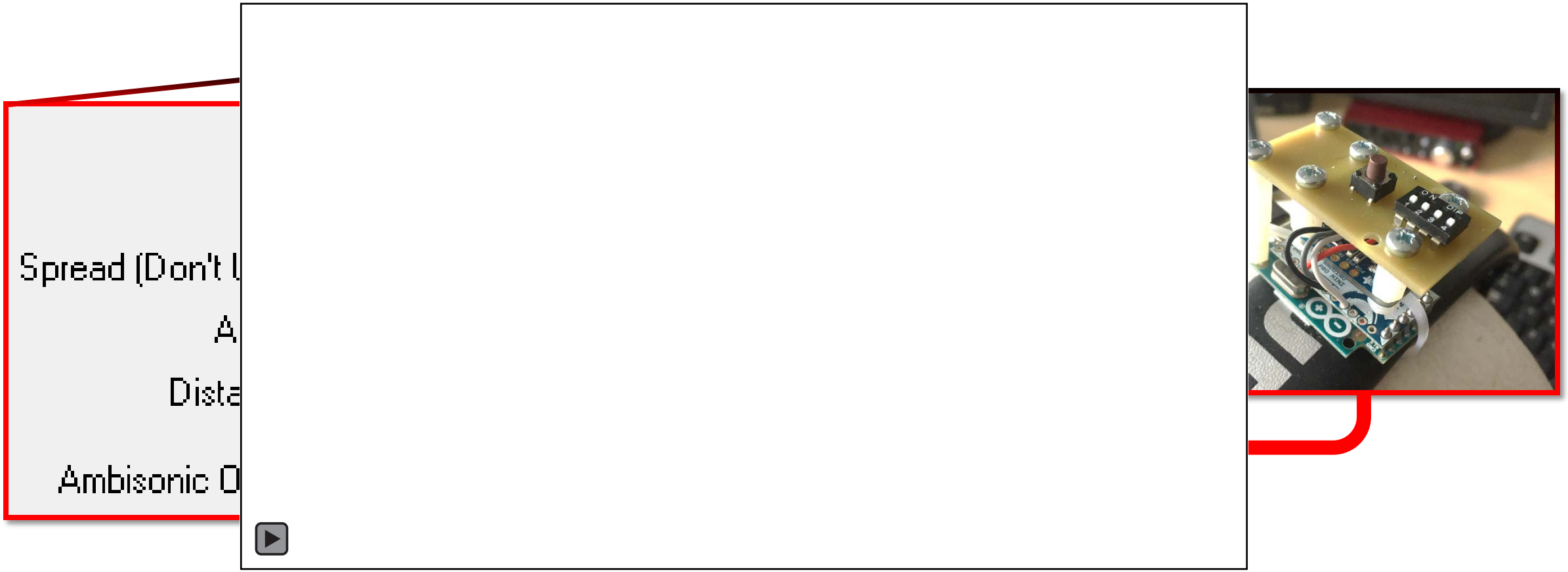


X-volver VST Plugin (Farina, 2017)

- Matrix convolution of audio signals (up to 32 in and 32 out).
  - Columns - inputs
  - Rows – outputs
- Two instances running to convolve 63 IRs, output to left (1) and right (2).
- Reduced from 35th to 31st order.
- Not all IRs are active at all times. Appropriate activation is made depending on the output from…

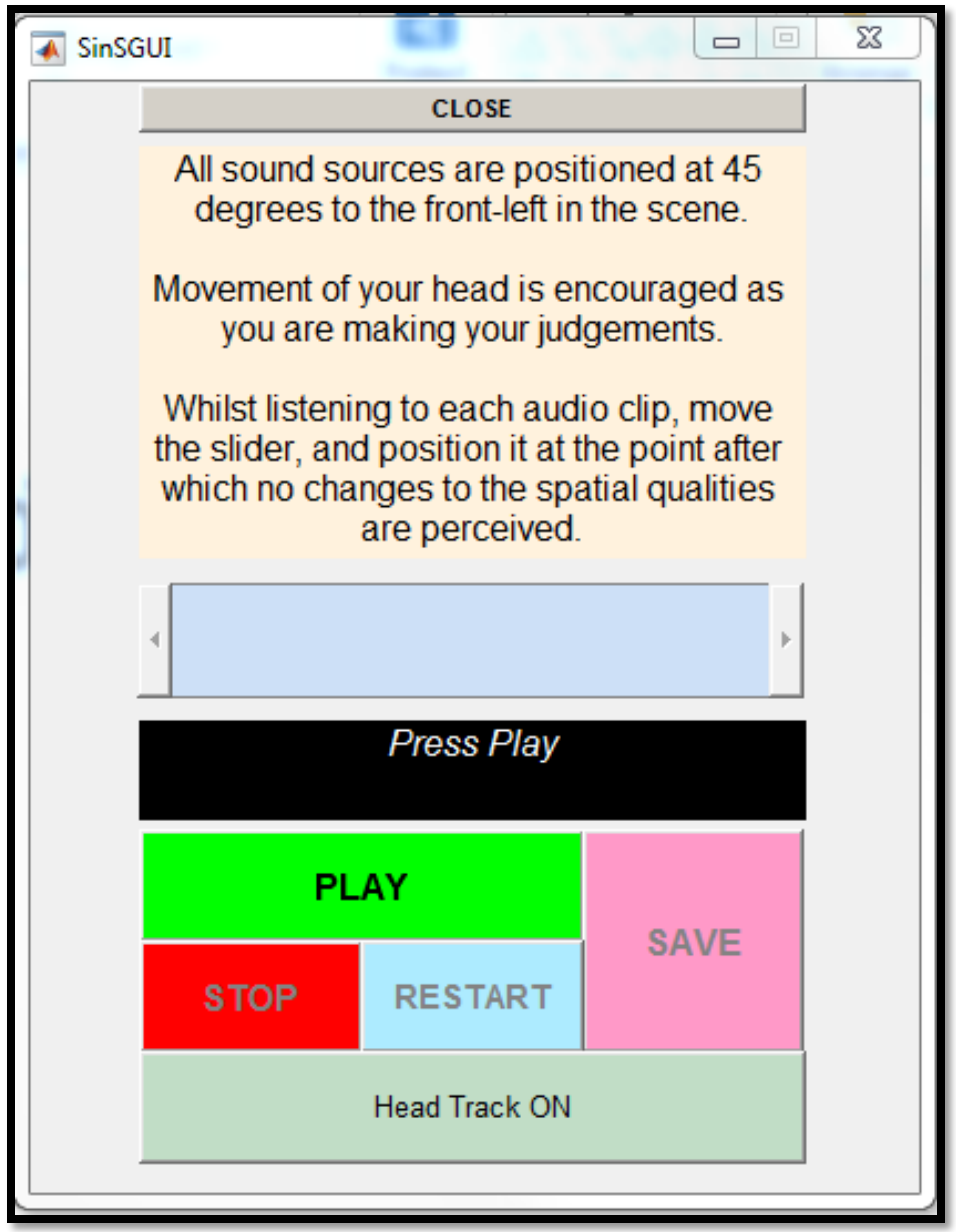


WigWare VHOA Panner and Mr Head Tracker



Subjective Testing

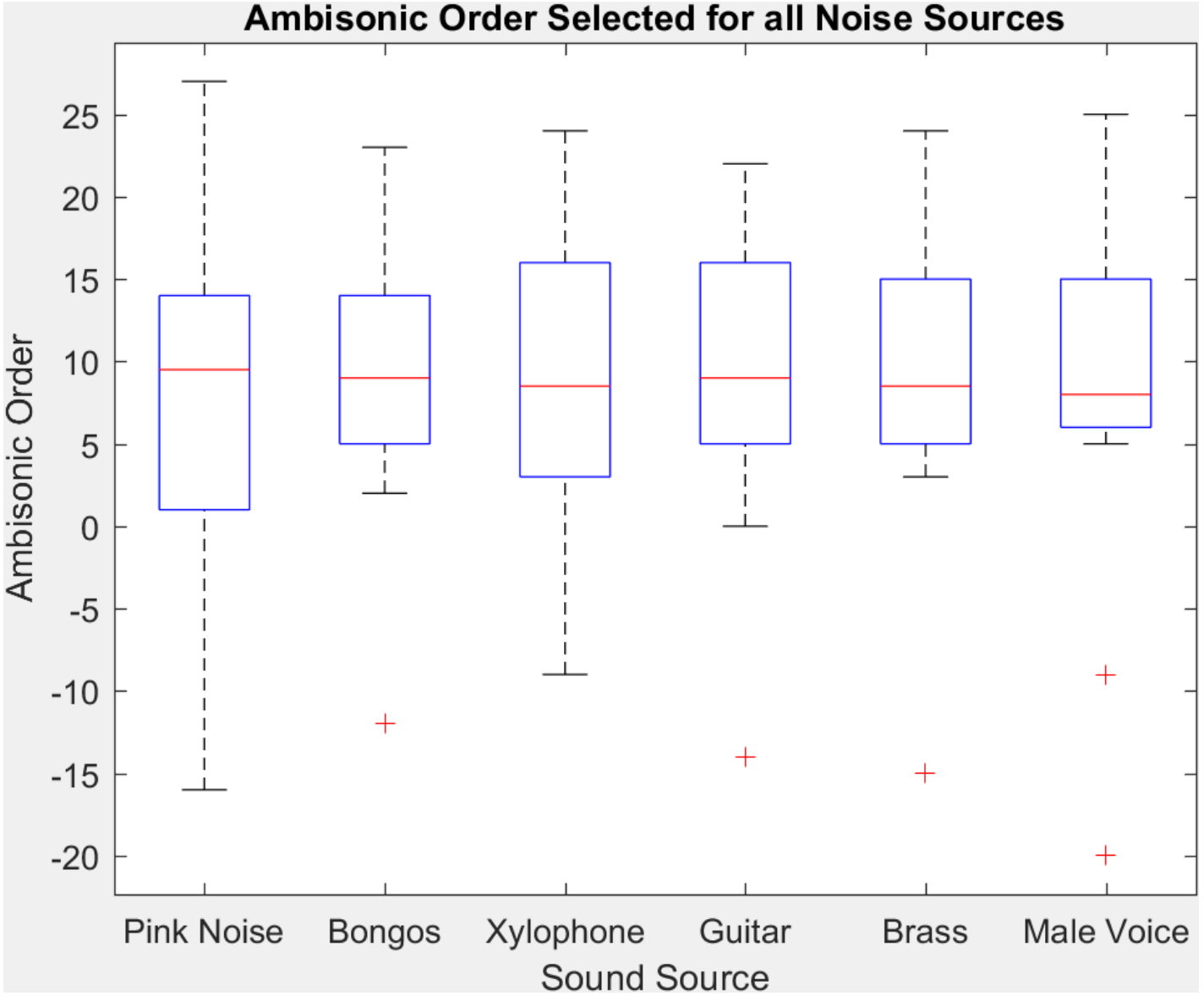
- Reaper controlled by Open Sound Control (OSC) via Matlab GUI.
- Six anechoic sound sources with differing tonal and dynamic characteristics.
- Single slider adjustment affects ‘Ambisonic Order’ position in VHOA Panner.
- Spatial quality judgements (Rumsey, 2002):
  - Source Focus
  - Source Stability
  - Scene Skew



Results

- 18 Participants.
- Strong similarities across differing sound sources.
- Pink noise shows greatest range – ‘not real sound’?
- Mean values between 7 - 9.6
- Median values between 8 – 9.5
- No responses at 31st order.

	Pink Noise	Bongos	Xylophone	Guitar	Brass	Male Voice
Mean	7.0	9.4	9.4	9.6	9.1	8.3
Median	9.5	9.0	8.5	9.0	8.5	8.0
Min	-16.0	-12.0	-9.0	-14.0	-15.0	-20.0
Max	27.0	23.0	24.0	22.0	24.0	25.0
IQR	13.0	9.0	13.0	11.0	10.0	9.0



Conclusions and Further Work

- 'No change' occurs in the majority around 9th order – *approximate point of transparency*.
- 31st order does not have any effect on the spatial qualities tested for these sound sources.
- Some participants latching onto tonal changes – "filtering effect"
- Effective presentation outside of the head – "I thought you could hear it too."
- Preliminary testing to ascertain the focus of our studies in future work – ABX?