

Application of Measured Directivity Patterns to Acoustic Array Processing

Mark R. P. Thomas
Microsoft Research, Redmond, USA

My Background

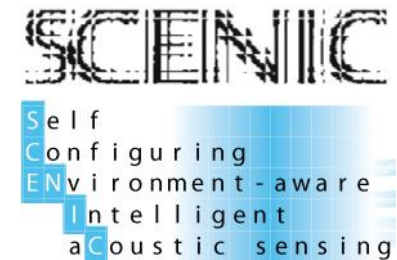
- 2011-present: Postdoctoral Researcher, Researcher (2013), Audio and Acoustics Research Group, Microsoft Research, Redmond, USA.
 - Microphone arrays (linear, planar, cylindrical, spherical).
 - Echo cancellation, noise suppression.
 - Head-related transfer functions.
 - Loudspeaker arrays.
- <http://research.microsoft.com>

Microsoft®
Research



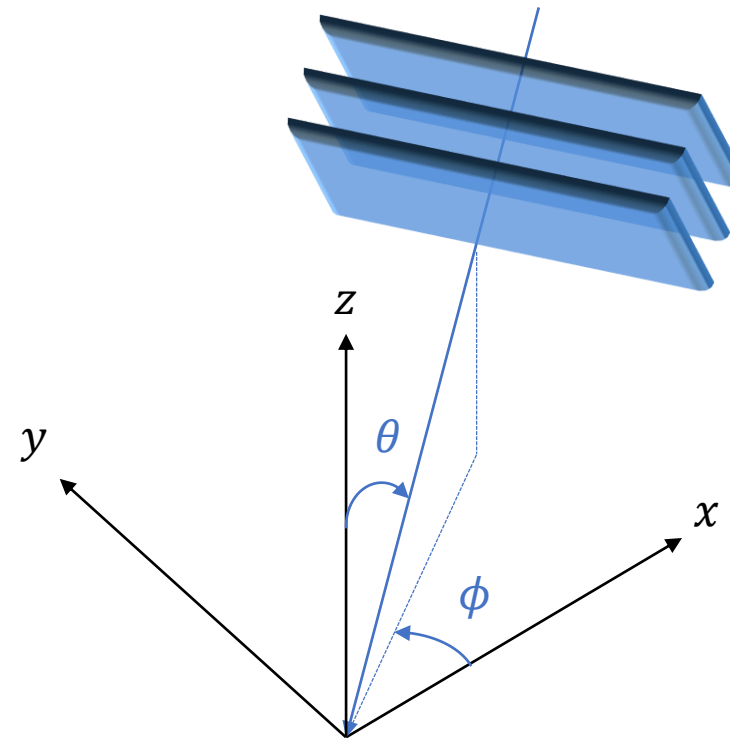
My Background

- 2001-2002: Pre-University/Vacation Trainee, BBC Research & Development, Kingswood Warren, Tadworth, Surrey.
 - DAB data protocols, audio signal processing for HDTVs, TV spectrum planning, hardware for live TV streaming.
- 2002-2010: MEng/PhD in Electrical and Electronic Engineering, Imperial College London.
 - MEng Thesis, “A Novel Loudspeaker Equalizer.”
 - PhD Thesis, “Glottal-Synchronous Speech Processing.”
- 2010-2011: Research Associate, Imperial College London
 - EU FP7 project Self Configuring ENVironment-aware Intelligent aCoustic sensing (SCENIC)
 - Spherical microphone arrays, geometric inference, channel identification & equalization.



Directivity Patterns: Background

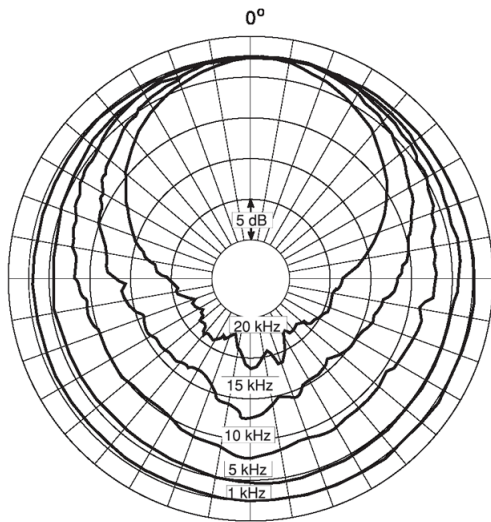
- Directivity pattern is the response to a **plane wave** emerging from a known direction relative to the device under test.
 - Function of azimuth ϕ
 - Function of elevation / colatitude θ
 - Function of frequency ω
- This is the 'farfield' response
 - Practically measured with a loudspeaker at a fixed distance of 1-2m.
- Independent of reverberation



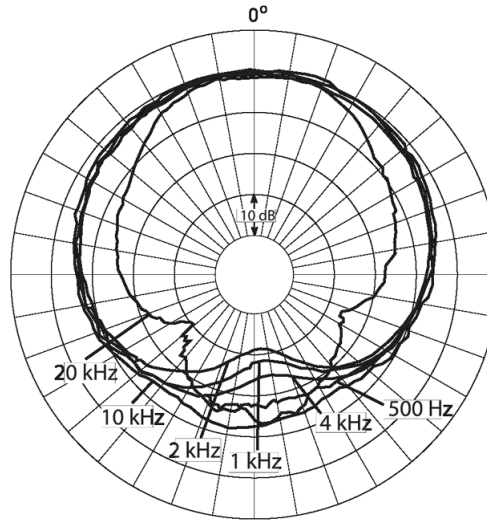
Directivity Patterns: Background

- All acoustic transducers exhibit some degree of directivity
 - Sometimes by design (e.g. cardioid microphone)
 - Sometimes parasitic (e.g. mounting hardware – example to come)

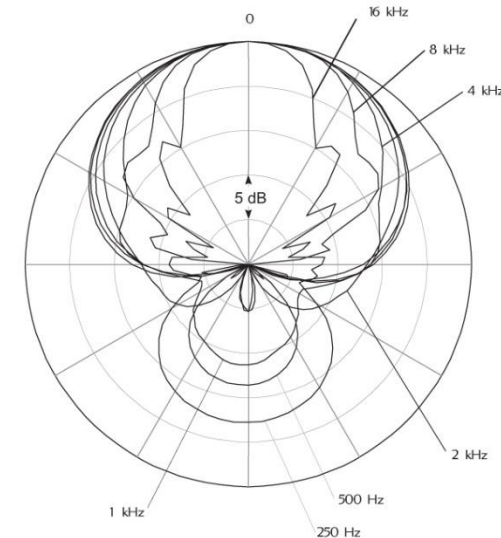
DPA 4006 Omni



DPA 4011 Cardioid

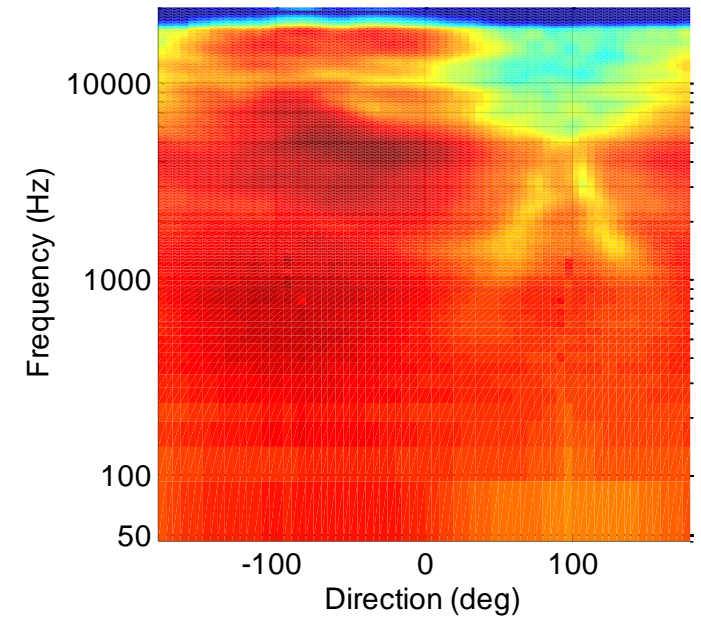
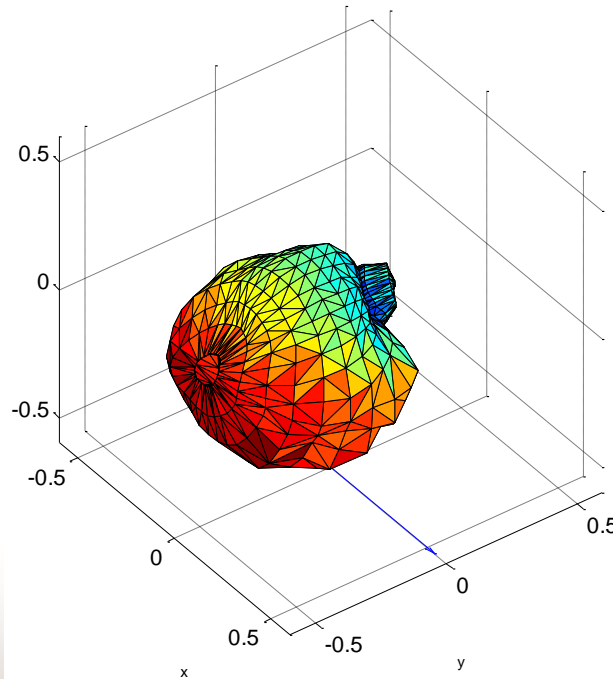


DPA 4017 Shotgun



Other Examples of Directional Behaviour

- Head-Related Transfer Functions (HRTFs)



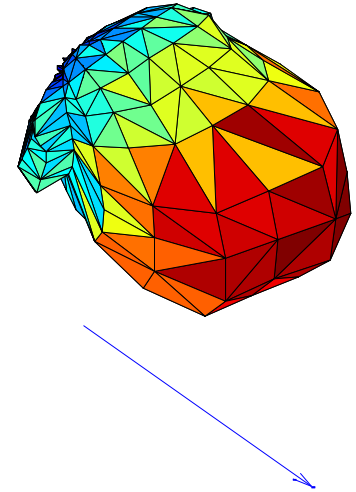
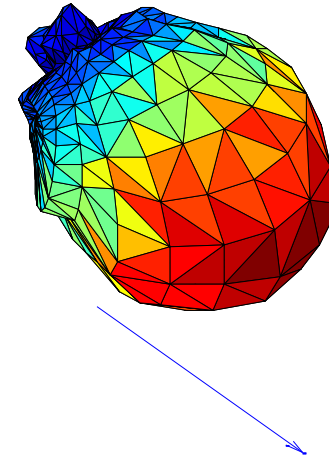
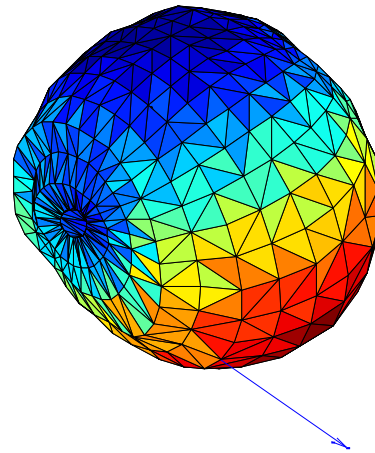
Other Examples of Directional Behaviour

- Loudspeakers

Loudspeaker Radiation Pattern at 200 Hz

Loudspeaker Radiation Pattern at 10 kHz

Loudspeaker Radiation Pattern at 1 kHz



Left image: <http://www.m-audio.com>

Contents

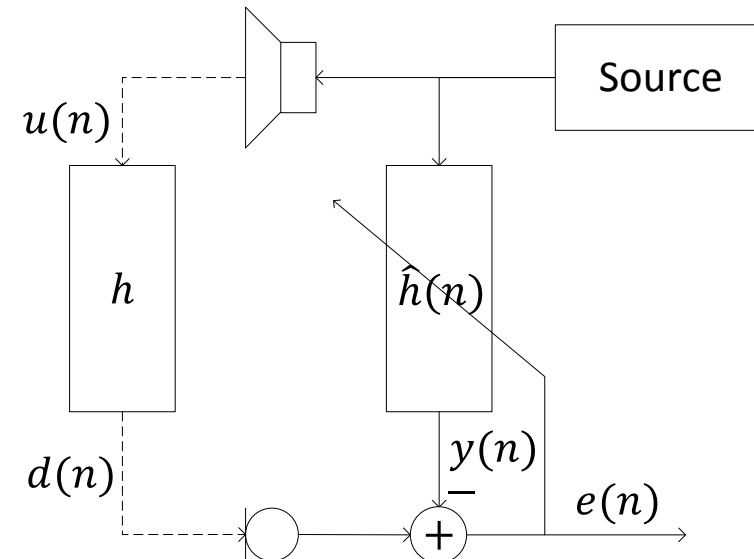
- Background on directivity patterns
- Part 1: Design of a measurement rig
 - Test signals
 - Loudspeaker placement
 - Extrapolation/interpolation of missing data
- Part 2: Practical Applications
 - Beamforming with Kinect for Xbox 360
 - Head-related Transfer Functions
- Conclusions

Design of a Measurement Rig: Requirements

1. Must be able to reliably measure the linear **impulse response** (transfer function) between a source signal and a test microphone.
2. Source signal must be **spectrally flat**.
 - Loudspeaker response may need compensating.
3. Sources must be able to be moved to a **precise location**.
4. Sources must be sufficiently far away to **avoid nearfield effects**.
5. Environment must be **anechoic** or sufficiently far away from acoustic reflectors.

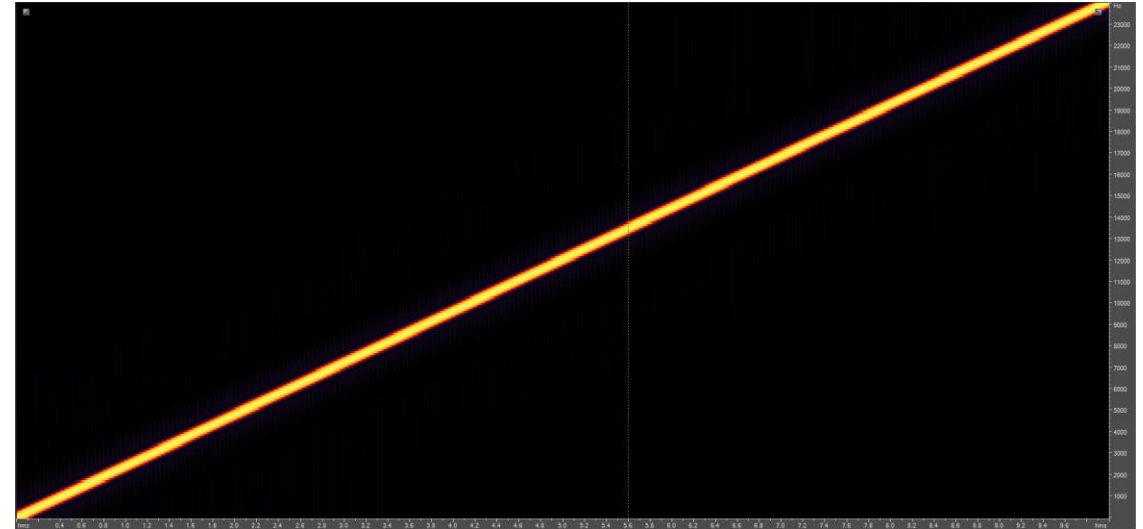
Test Signals

- Source is a known signal $u(n)$.
- Record signal $d(n)$
 - Has been filtered by unknown **finite impulse response** (FIR) system h .
- Estimate h by minimizing the difference between $y(n)$ and $d(n)$.
 - FIR system identification is a **convex problem**: always a **unique minimum**.
 - Most solutions are **closed form** (non-adaptive).
 - **Adaptive** solutions are useful for cases when h is constantly changing.



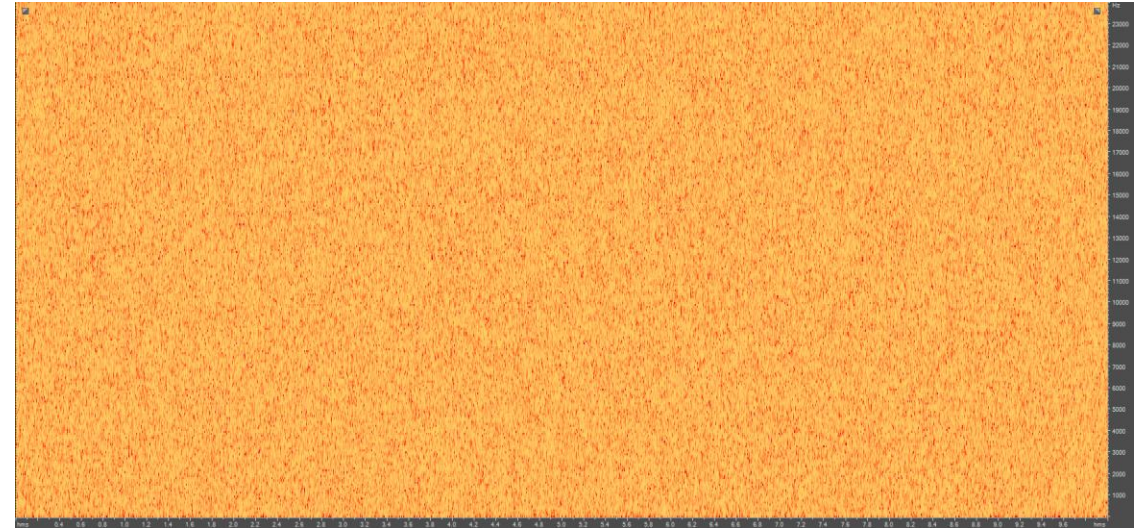
Choice of Test Signal: Chirp-Like

- Chirp-Like Signals
 - Linear chirp
 - + Easy to produce
 - + Intuitive
 - System ID requires generalized methods.
 - Time-stretched pulse (TSP)
 - + Pulse and its inverse are compact in support. Very low-complexity system ID.
 - + Robust to nonlinearities.
- Energy is concentrated in a narrow band; possibility of standing waves in cone material producing nonlinearities.



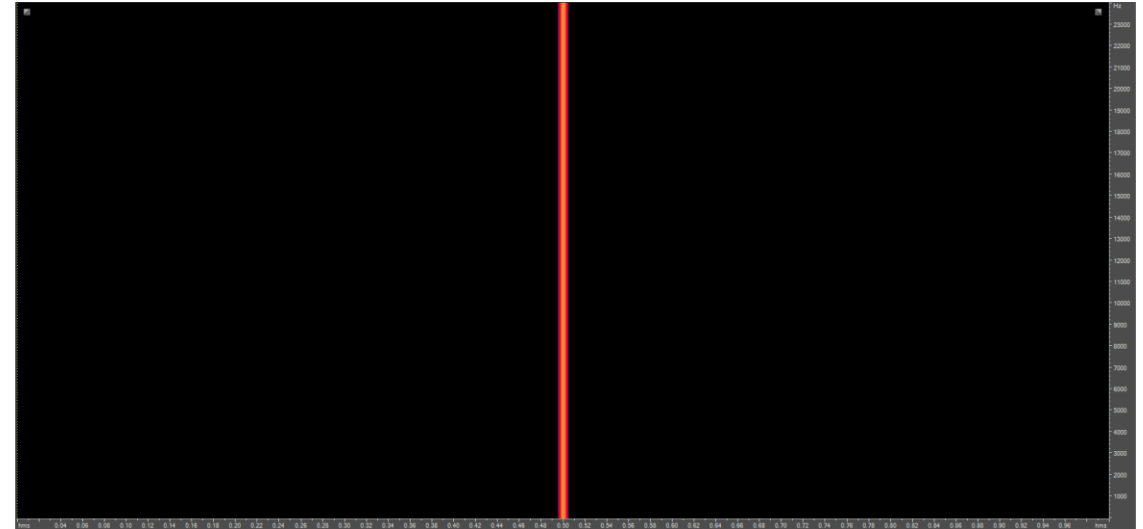
Choice of Test Signal: Pseudorandom Noise

- Maximum-length sequences (MLS) / perfect sequences
 - + Autocorrelation is a perfect impulse.
 - + Fast system ID with modified Hadamard transforms.
 - Sensitive to nonlinearities.
- Gaussian Noise
 - + Easy to generate
 - + Autocorrelation theoretical impulse with sufficiently long data
 - Several solutions for system ID, some inexact and/or computationally expensive.
- + Spectrally flat (energy not concentrated in a single spectral band).



Choice of Test Signal: Direct Impulse

- + Recorded signal is the system impulse response.
- + Straightforward to produce in the digital domain.
- In the analogue domain, gunshots, hammer blows and clickers have been used for room acoustics.



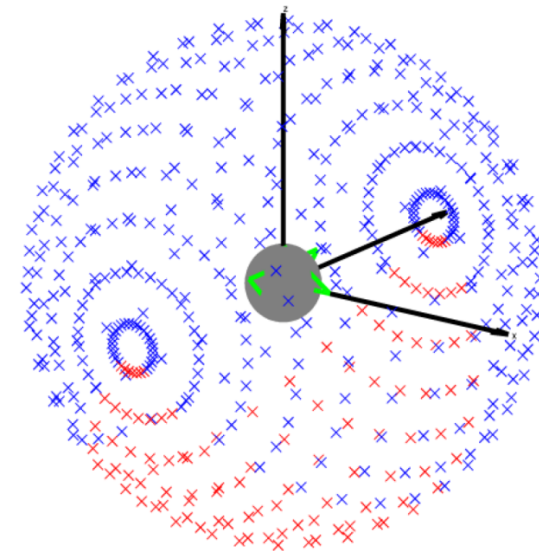
- Requires high amplitudes in order to provides good signal-to noise ratio (risk of nonlinearity).

Contents

- Background on directivity patterns
- Part 1: Design of a measurement rig
 - Test signals
 - Loudspeaker placement
 - Extrapolation\interpolation of missing data
- Part 2: Practical Applications
 - Beamforming with Kinect for Xbox 360
 - Head-related Transfer Functions
- Conclusions

Equiangular Sampling

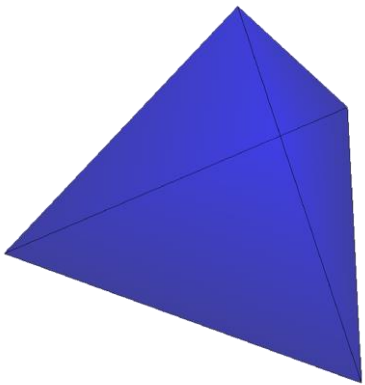
- Mount an array of loudspeakers on a semicircular arc and rotate about the device
 - Example: 16 loudspeakers spaced 11.25° , poles at the sides.
 - + Practically continuous azimuth.
 - Colatitude angles fixed at discrete locations.
 - Missing spherical wedge underneath.
 - Mechanically complicated
 - Nonuniform sampling
- Other variations on the theme
 - Rotate device relative to fixed loudspeaker.



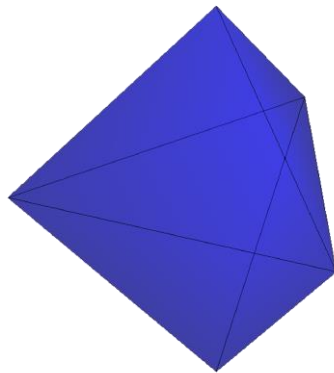
Uniform Sampling

- Place sources in **fixed locations** around the device under test
 - **Uniform distribution** of test points can be ensured.
 - + No moving parts
 - Only 4 truly uniform solutions in 3D! The points lie on the vertices of 4 regular polyhedra.

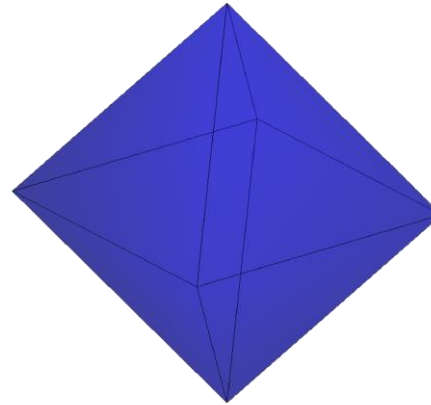
$N=4$: Tetrahedron



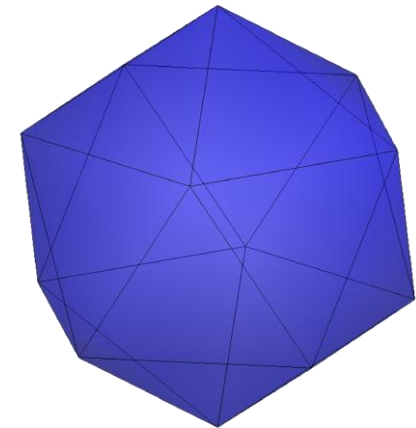
$N=5$: Triangular dipyramid



$N=6$: Regular octahedron



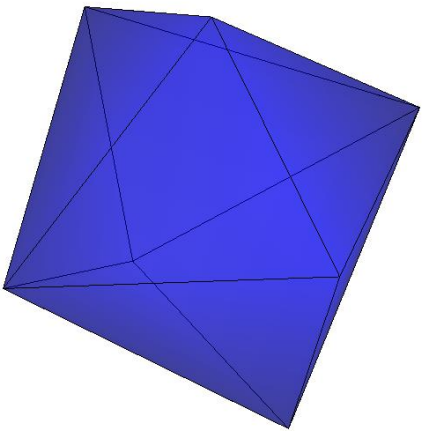
$N=12$: Regular icosahedron



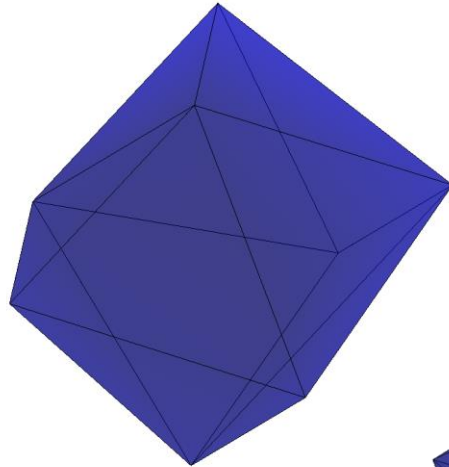
Near-Uniform Sampling: Geometric Solutions

- There are a few geometric solutions to the **near-uniform sampling** case.

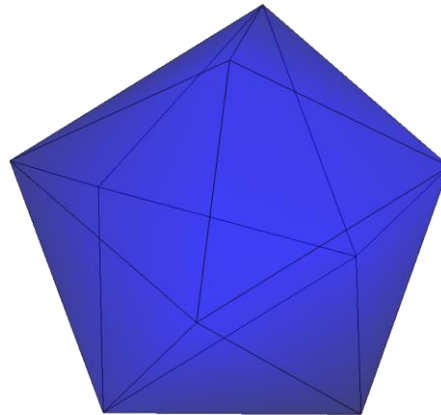
$N=7$: Pentagonal Dipyramid



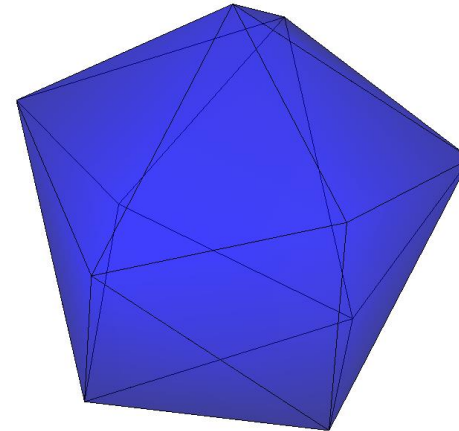
$N=8$: Square Antiprism



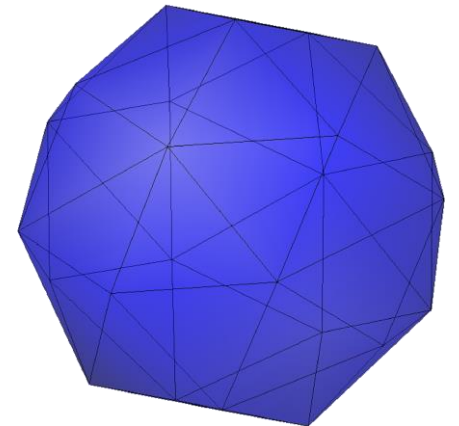
$N=9$: Triaugmented Triangular Prism



$N=10$: Gyroelongated square dipyramid



$N=24$: snub cube



Near-Uniform Sampling: Numerical Solutions

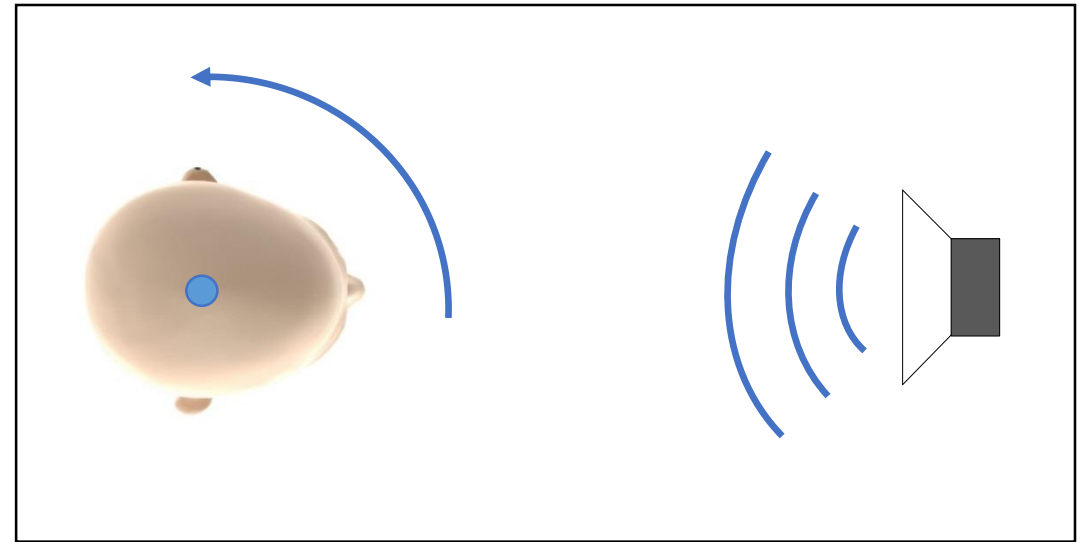
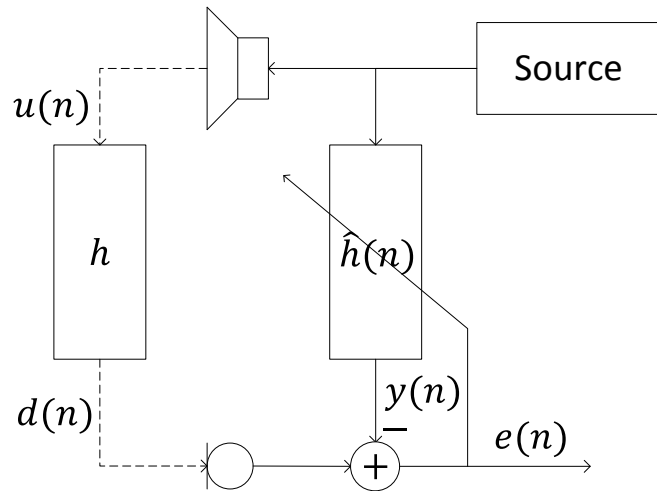
- For all other N , only **numerical solutions** exist
 - This is the **Thomson Problem**: determine the **minimum electrostatic potential** energy configuration of N electrons on the surface of a unit sphere.
 - I use the **Fliege** solution¹.
- Solutions have several other uses:
 - Spherical **microphone arrays**
 - **Geodesic domes**
 - Solutions in higher dimensions useful for **quantization** in coding schemes.



¹J. Fliege, "The distribution of points on the sphere and corresponding cubature formulae," *IMA J. Numer. Anal.* Vol. 19, no. 2, pp. 317-334, 1999.

Continuous Sampling

- Sound source is continuous Gaussian noise
- Device under test (in this case a human head) is **continuously rotated**.
- NLMS adaptive filter identifies **instantaneous** transfer function
 - Assumption: filter is **constantly converged** to correct solution.
 - Only suitable for horizontal plane.

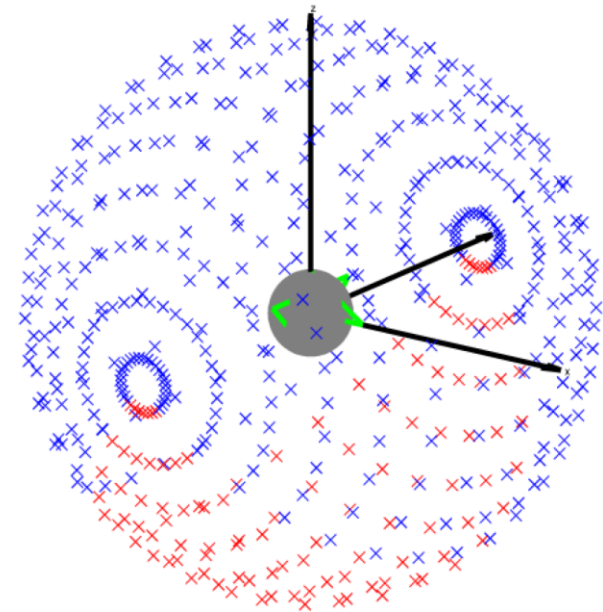


Contents

- Background on directivity patterns
- Part 1: Design of a measurement rig
 - Test signals
 - Loudspeaker placement
 - Extrapolation / interpolation of missing data
- Part 2: Practical Applications
 - Beamforming with Kinect for Xbox 360
 - Head-related Transfer Functions
- Conclusions

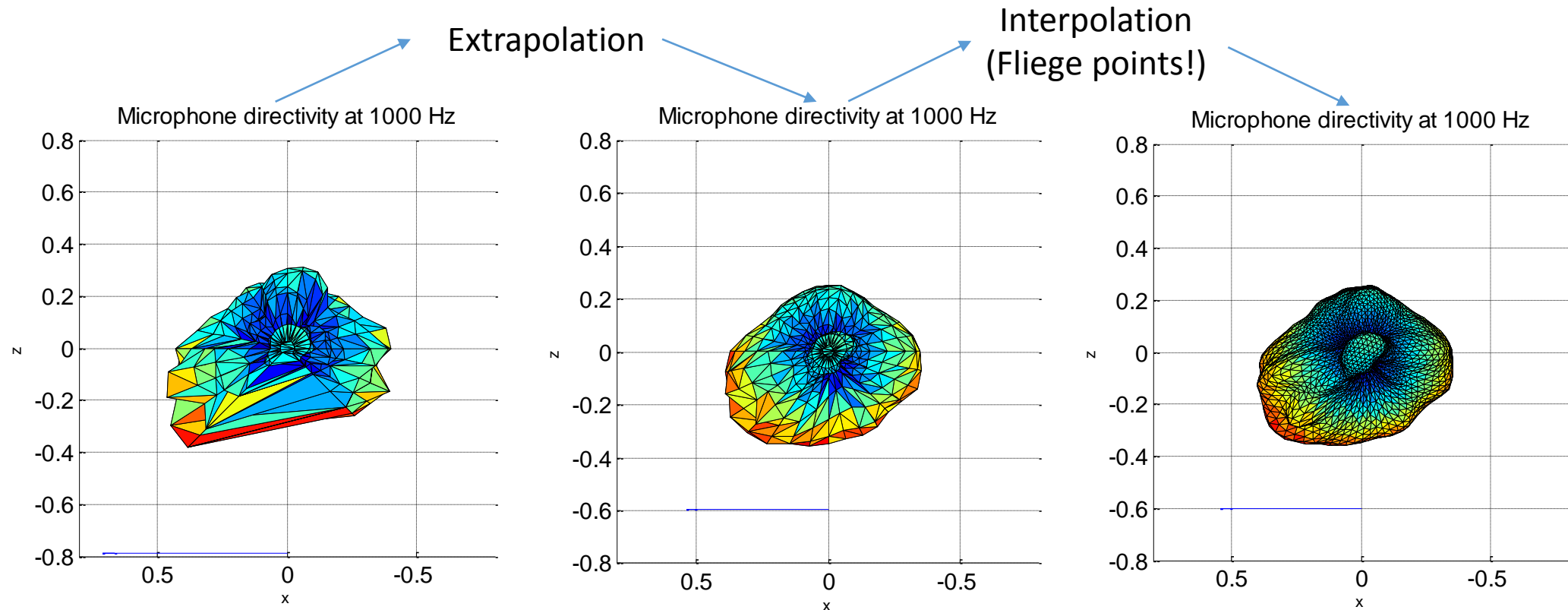
The Missing Data Problem

- A **spherical wedge** of data is missing beneath the test subject.
- Polynomial / spline interpolation do not work well
 - Do not exploit the **natural periodicity** of the data.
 - Do not account for **curvature** of the surface.
 - Solutions tend to be **numerically unstable**.
- Need an interpolation/extrapolation scheme better suited to data in spherical coordinates.



The Missing Data Problem

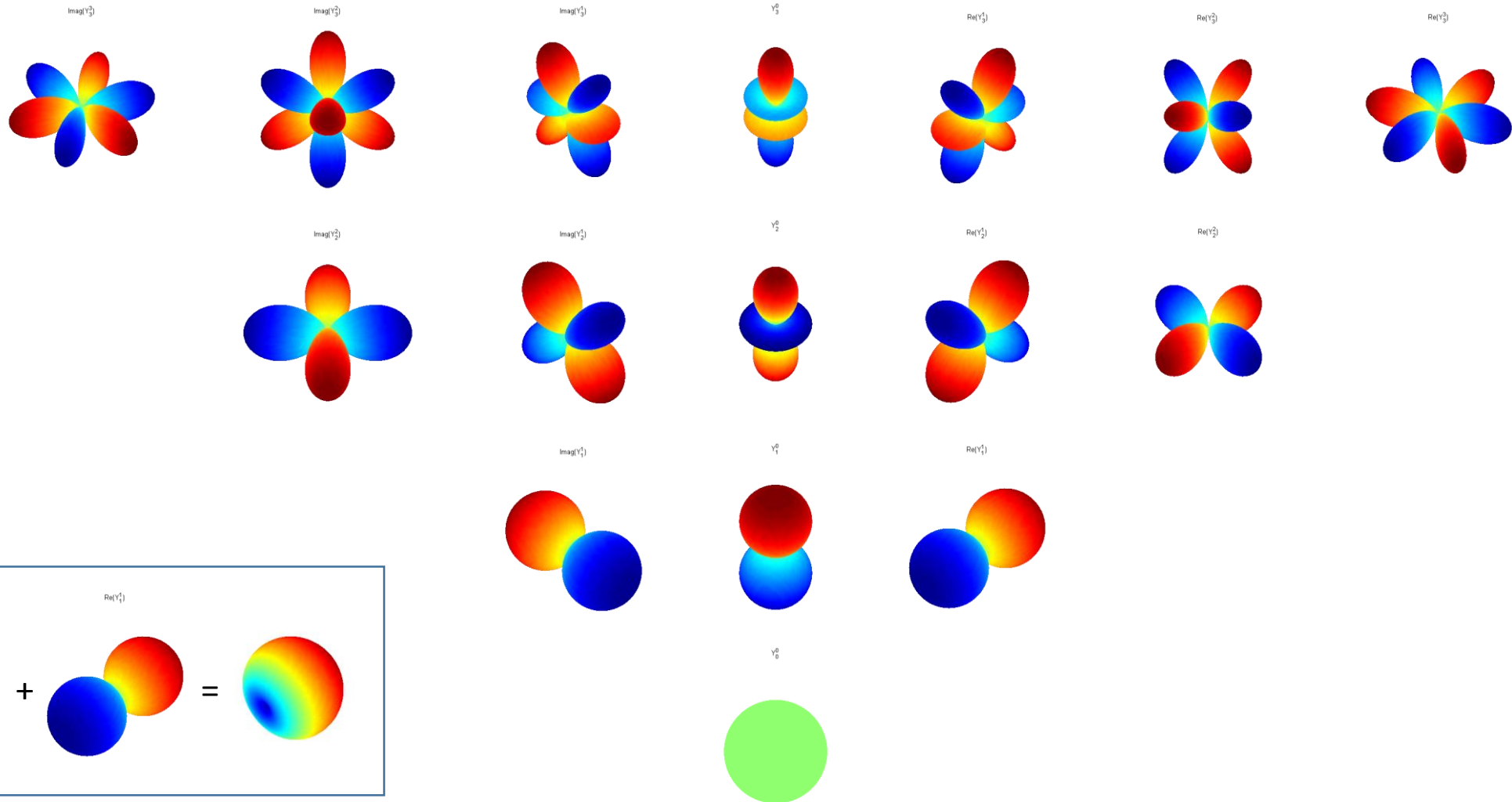
- Extrapolation: the missing spherical wedge underneath.
- Interpolation: the data between measurement points.



Spherical Harmonics

- Spherical harmonics are the **angular solutions** to the wave equation in spherical coordinates
 - They form an **orthogonal basis** for functions on the sphere.
 - Useful for analysis of orbital angular momentum of electrons.
 - Also useful for **wave field analysis** with spherical microphone arrays.
- They are to spherical space as the sine/cosine functions are to 1D space
 - They are the basis for a **spherical Fourier Transform**.
 - Think of it as a **spatial frequency domain**.
- Spherical harmonics have **discrete solutions** with degree n and order m .

Spherical Harmonics



Extrapolation with Spherical Harmonics

- If the sphere were complete, there would be **512 points** in total
 - 16 colatitude angles x 32 azimuth angles.
 - This permits a **15th order** model.
- The actual number of measured points is **400**
 - 16 colatitude angles x 25 azimuth angles.
 - We cannot compute a 15th order model in the unknown region.
- How do we perform a good fit?

Test Subject: B&K Head and Torso Simulator

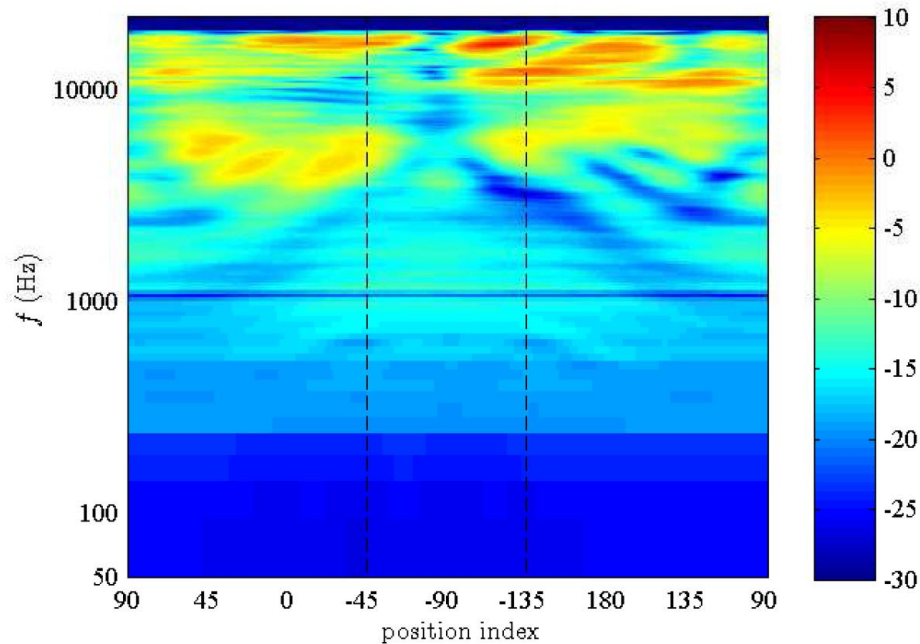
- B&K Head and Torso Simulator (HATS)
 - Simulates **anthropometry** of average human.
- Baseline measurements
 - Taken both right way up and upside down
 - Data combined to form complete sphere.



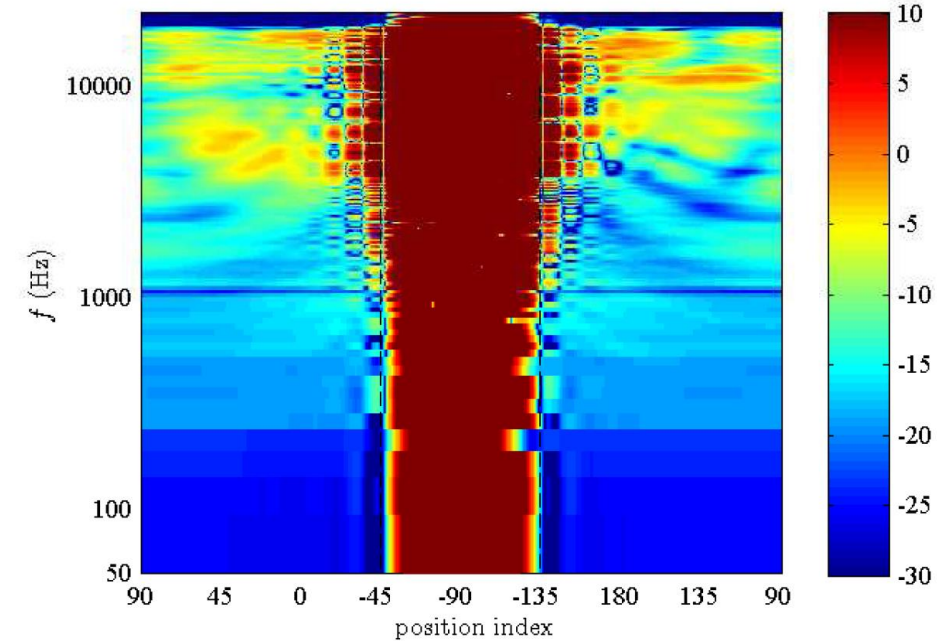
Fitting Problem

- Spherical harmonics aren't enough!
 - 15th order model is very poorly behaved when reconstructing missing region.

15th order model on complete data



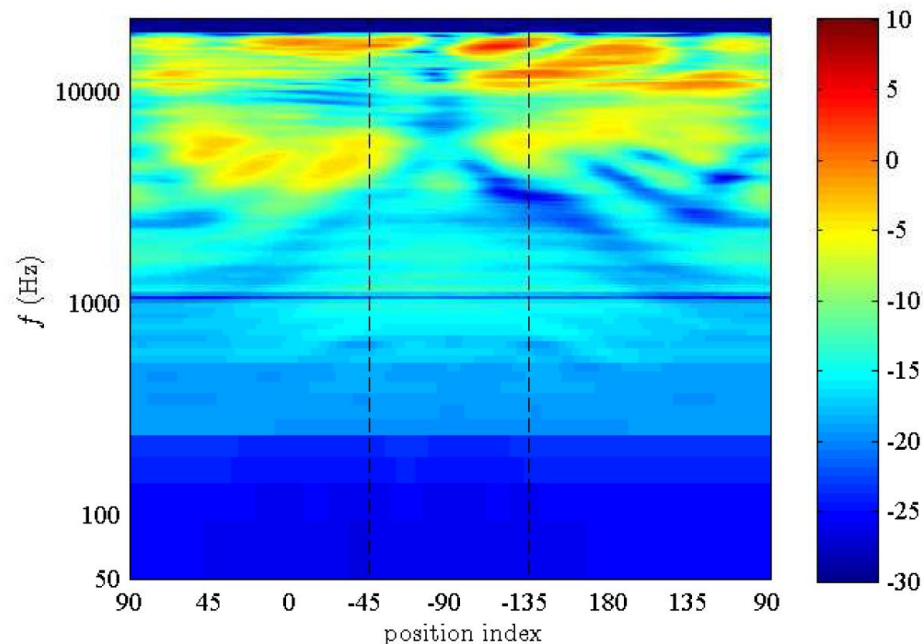
15th order model on incomplete data



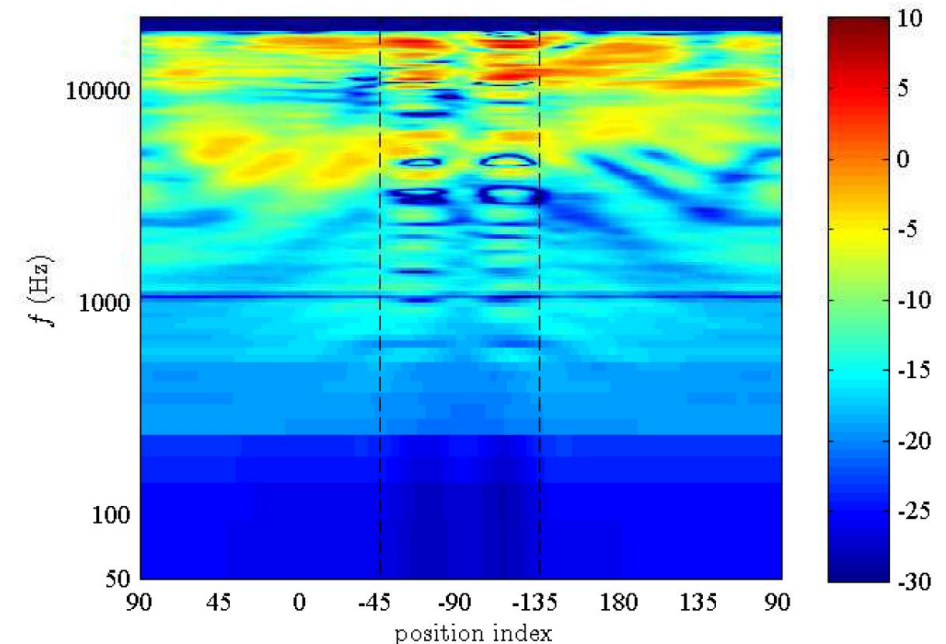
Regularization to the Rescue?

- Regularized least-squares fit can help prevent blow-up.
 - Regularization hurts the data we know to be correct.
 - Still a poor fit: artificial periodicity creeps in.

15th order model on complete data



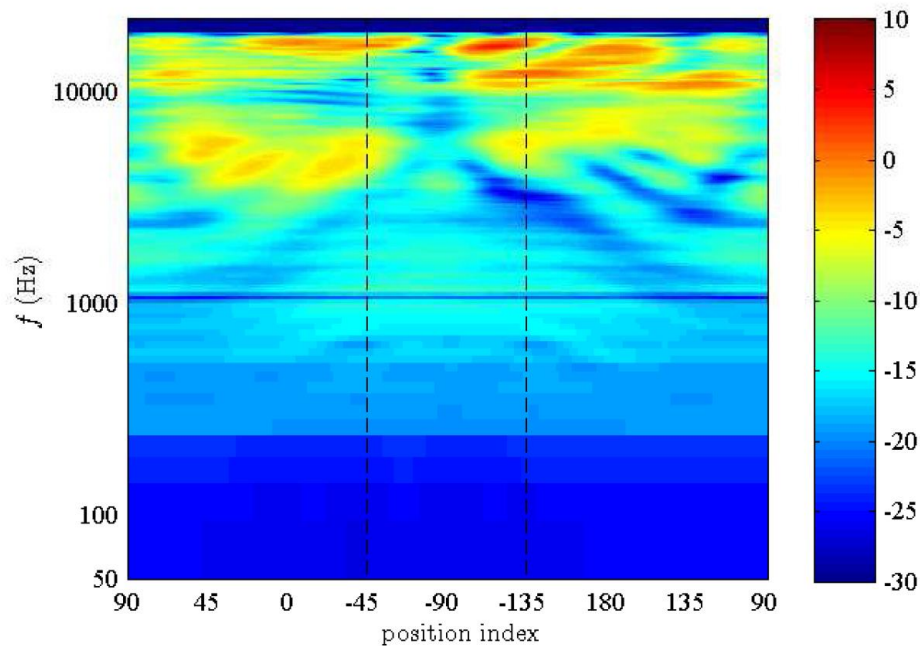
Regularized 15th order model on incomplete data



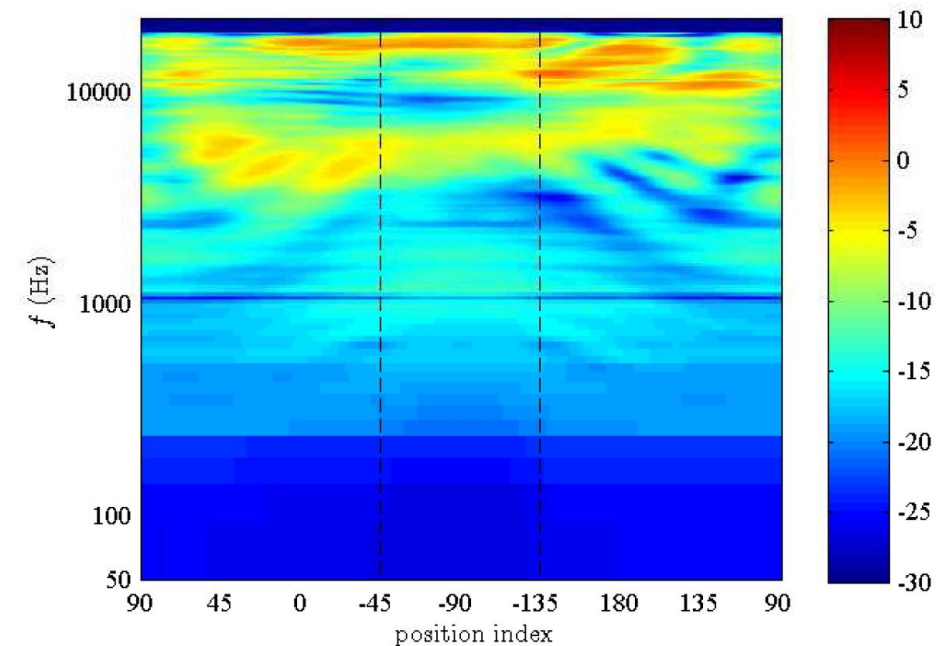
Non-Regularized Least-Squares Fit

- There is theoretically enough data for a 3rd order model.
 1. Extrapolate unknown data using a 3rd order model.
 2. Combine with the original data (leaves this unharmed).
 3. Perform 15th order non-regularized fit over the entire sphere.

15th order model on complete data



Proposed model on incomplete data

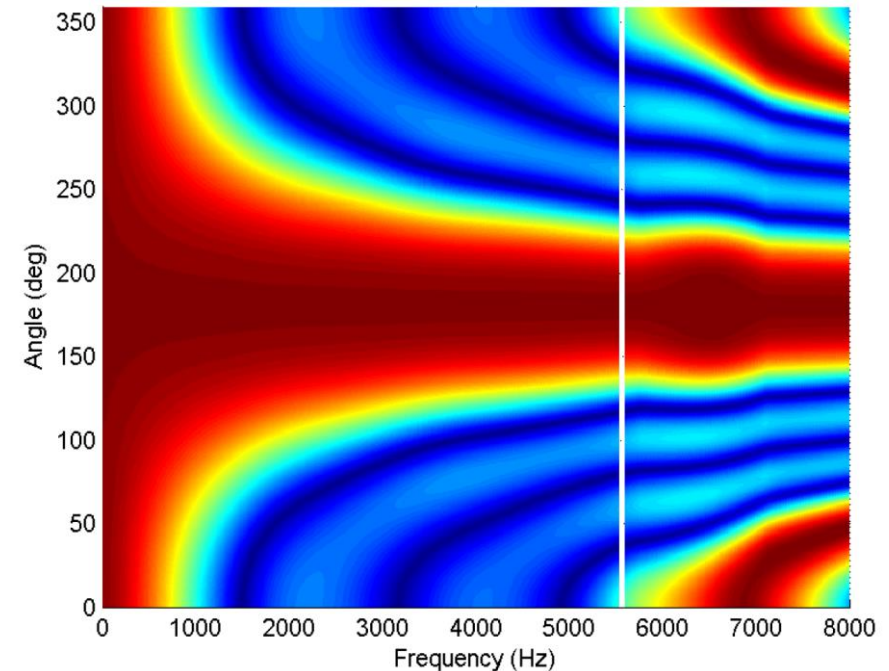
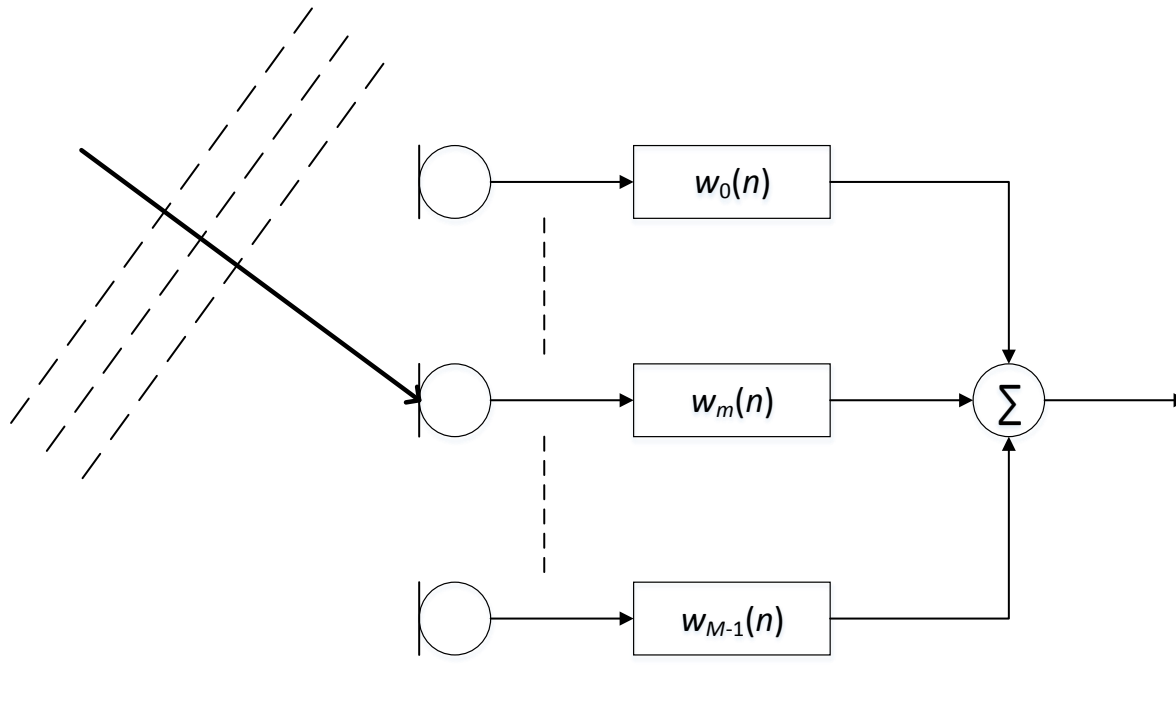


Contents

- Background on directivity patterns
- Part 1: Design of a measurement rig
 - Test signals
 - Loudspeaker placement
 - Extrapolation\interpolation of missing data
- Part 2: Practical Applications
 - Beamforming with Kinect for Xbox 360
 - Head-related Transfer Functions
- Conclusions

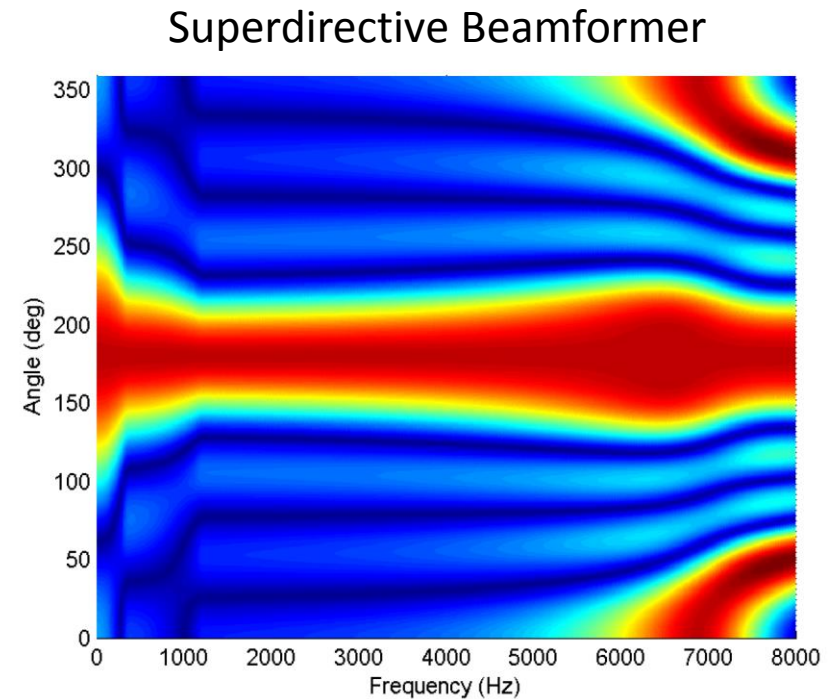
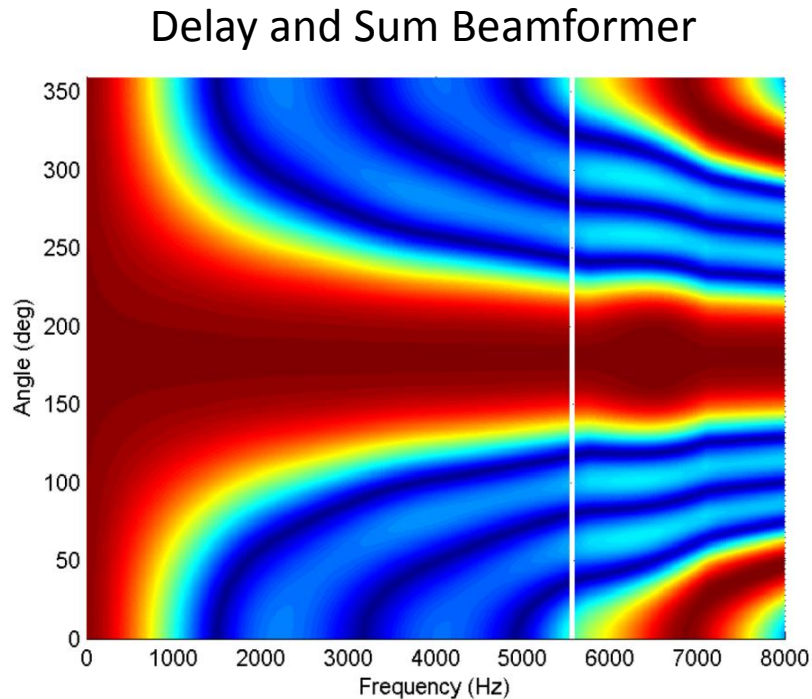
Delay and Sum Beamformer (DSB)

- DSB temporally **aligns wavefronts** so that they sum **coherently** in a given direction.



Optimal Superdirective Beamformer

- Calculate weights by **optimization** to synthesize a **frequency-independent** pattern.
 - Can exploit true directivity pattern.
 - Practical limitations with **white noise gain** at low frequencies.



Kinect for Xbox 360

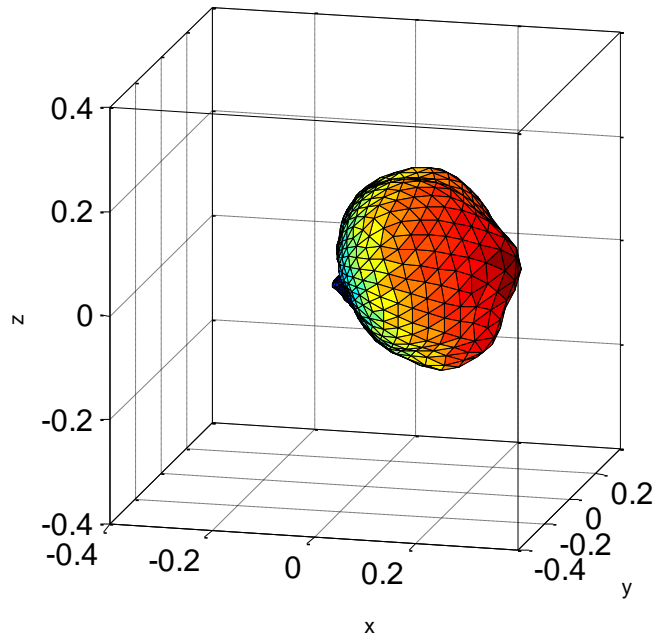
- RGB camera
- Depth camera
- Motorized tilt
- 4 cardioid microphones



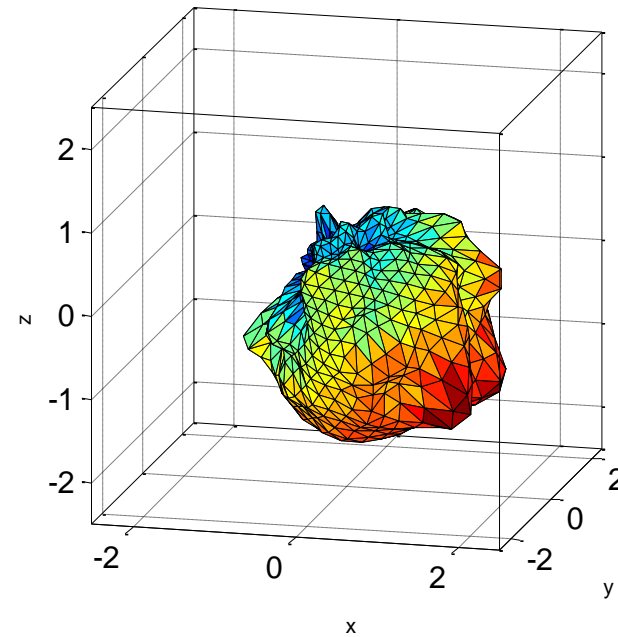
Kinect for Xbox 360

- Classic cardioid at 1 kHz, points to the floor at 5 kHz.
- Need to design with measured directivity patterns in mind.

Microphone Directivity at 1 kHz



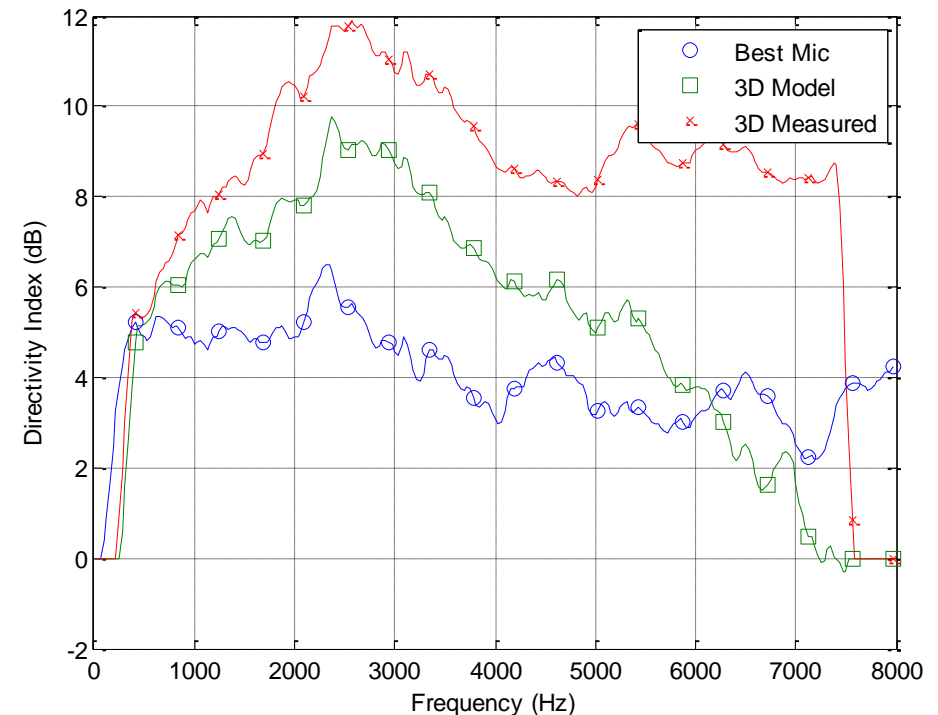
Microphone Directivity at 5 kHz



Performance Metrics

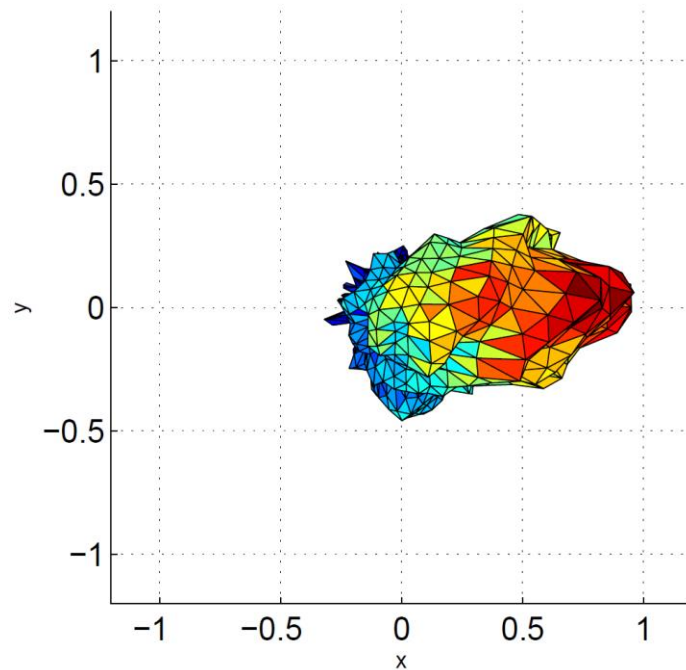
- Directivity index: log ratio of wanted signal to unwanted signals
 - Examples: Single cardioid: 4.8 dB, single omni: 0 dB.
 - **~4-6 dB improvement** over best microphone
 - Cardioid model OK up to ~3 kHz
 - Cardioid model suffers > 6 kHz
- Speech recognition task:
 - **50% relative** (5% absolute) in word error rate over 3D model.

	PESQ (1-4.5)	WER (%)	SER (%)
Best Mic	2.13	18.47	31.67
3D Model	2.64	9.79	15.00
3D Measured	2.66	4.92	9.17

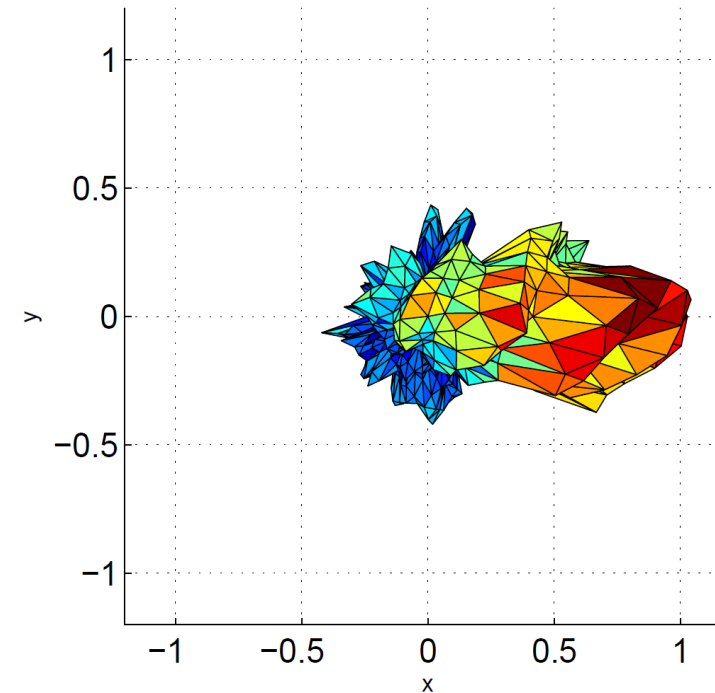


Optimal Beamforming with Kinect

Beamformer with 3D cardioid model at 1 kHz



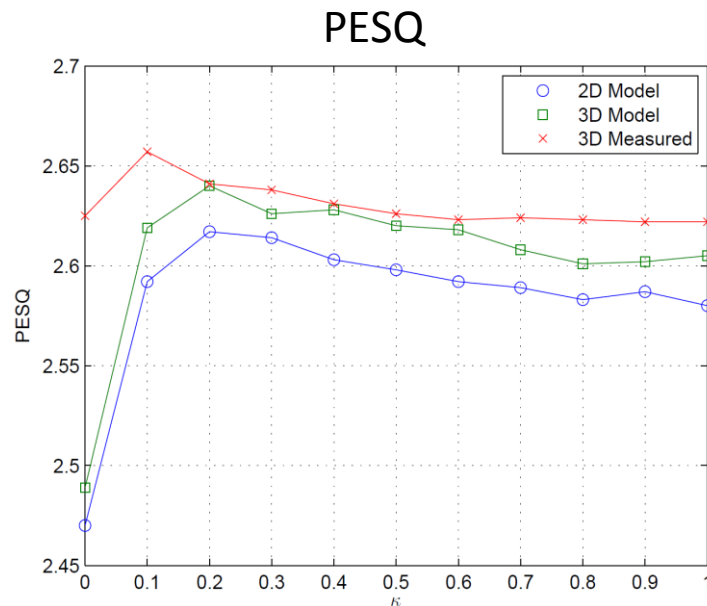
Beamformer with 3D measurements at 1 kHz



M. R. P. Thomas, J. Ahrens, I. J. Tashev, "Optimal 3D beamforming using measure microphone directivity patterns," *IWAENC* 2012.

Beamforming with Kinect – Regularization

- Problem: danger of becoming **too device-specific**
 - Account for manufacturing variations by adding **regularization** – becomes closer to delay-and-sum (lower performance).
 - Solution: a) **calibrate** during manufacture (expensive), or b) determine necessary regularization.

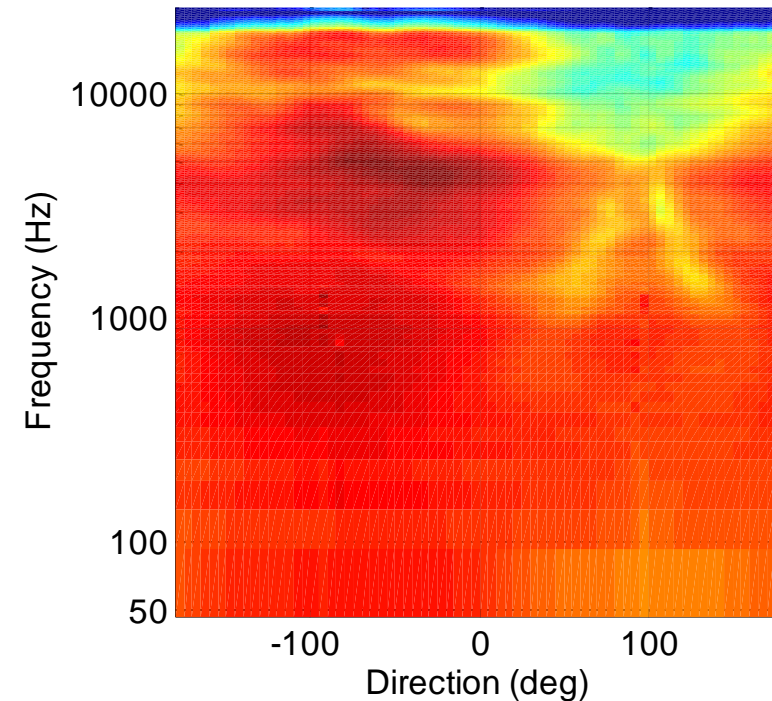
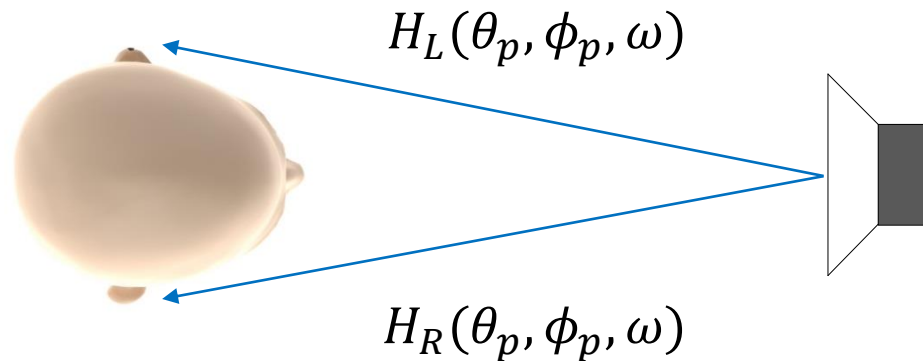


Contents

- Background on directivity patterns
- Part 1: Design of a measurement rig
 - Test signals
 - Loudspeaker placement
 - Extrapolation\interpolation of missing data
- Part 2: Practical Applications
 - Beamforming with Kinect for Xbox 360
 - Head-related Transfer Functions
- Conclusions

Head-Related Transfer Functions

- HRTFs capture acoustic properties of the head
 - Enables rendering of 3D audio over headphones



Personalizing HRTFs

- HRTFs are **highly personal**
 - Function of **anthropometric features** (head width, height, ear position, size etc.).
- HRTFs provide **temporal** and **spectral** cues for source localization
 - Inter-aural time differences (**ITD**)
 - Inter-aural level differences (**ILD**)
 - Pinna resonances
 - ITD and ILD insufficient: they help localize to within a **cone of confusion**.
 - Introduce subtle **spectral cues** to help resolve elevation and front/back.
- Should be used in conjunction with real-time **head tracking**
 - **Head rotations** provide additional information for source localization.

Measuring and Estimating HRTFs

1. Anechoic chamber and measurement rig

- Accurate
- Expensive

2. Finite-element modelling

- **Less accurate** than measurement
- **Slow**: can take a single machine several days

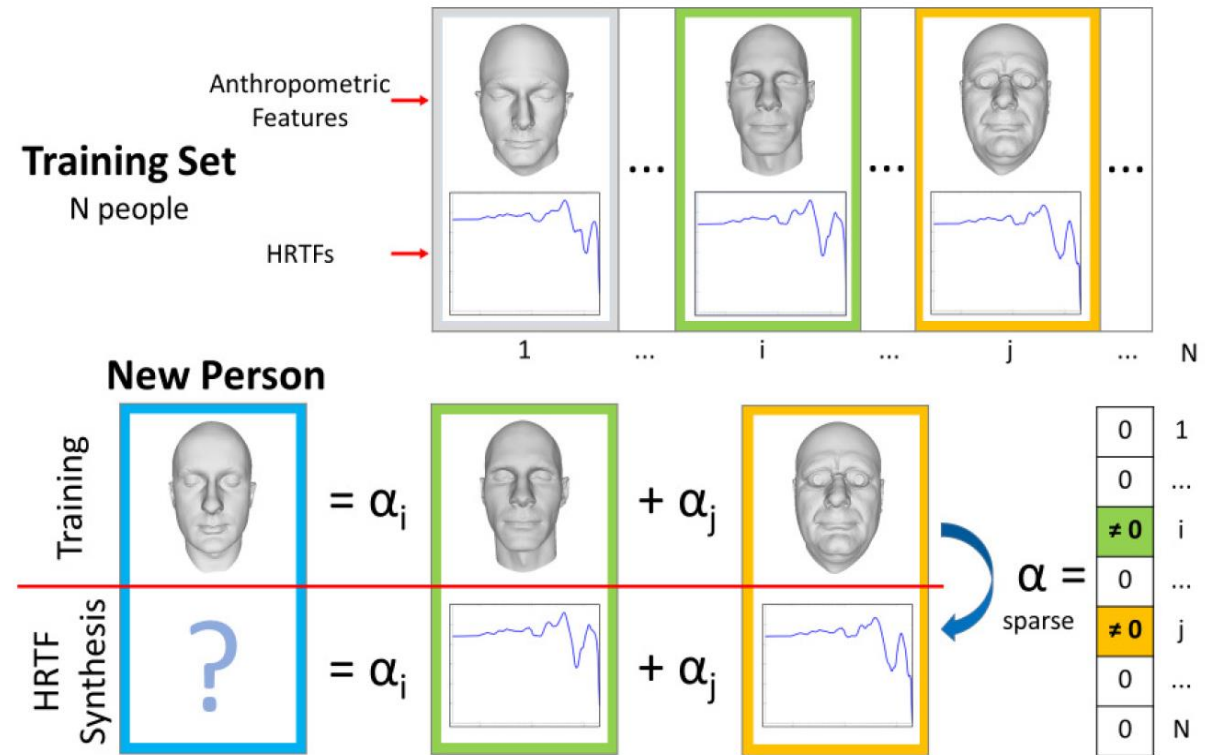
3. Estimate from anthropometric data

- **Less accurate** than measurement
- Requires **no invasive measurements**



HRTF Magnitude Synthesis

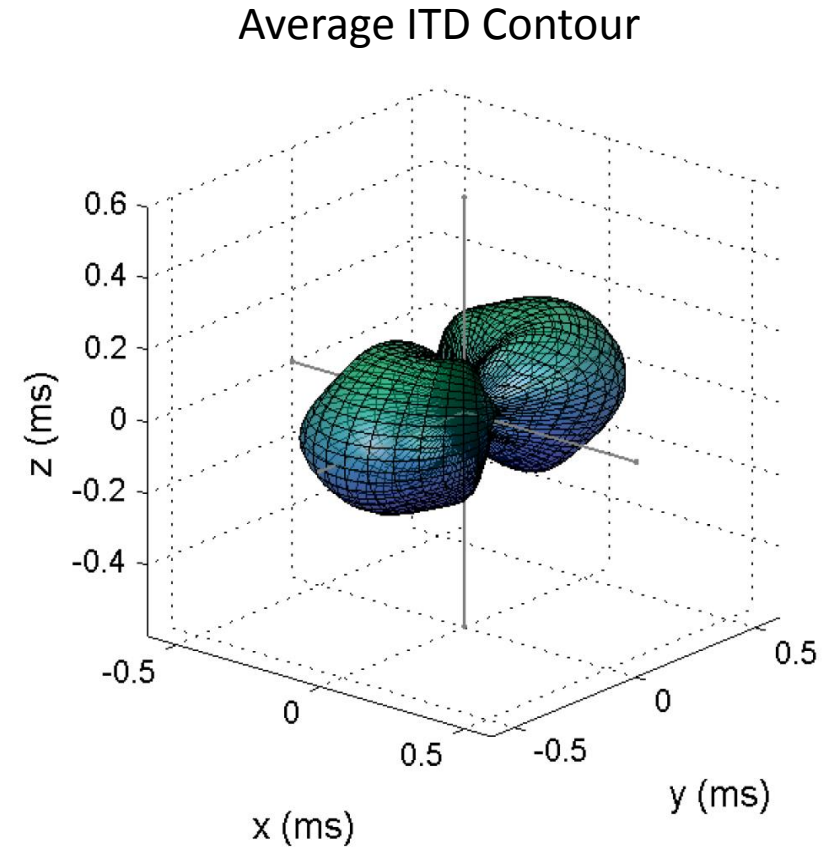
1. Measure anthropometric features on a large database of people.
2. Represent a new candidate's anthropometric features as a **sparse combination** α of people in the database.
3. Combine HRTF magnitude spectra with **same weights** α to synthesize personalized HRTF.



P. Bilinski, J. Ahrens, M. R. P. Thomas, I. J. Tashev, J. C. Platt, "HRTF magnitude synthesis via sparse representation of anthropometric features," *ICASSP*, 2014.

HRTF Phase Synthesis

- Most **ITD contours** have near figure-of-8 shape.
- Phase synthesis by scaling average ITD contour.
 - Also estimated with anthropometric features.
 - Appears to be perceptually sufficient with informal testing.



I. J. Tashev, "HRTF phase synthesis via sparse representation of anthropometric features," *ITA Workshop*, 2014.

Objective Evaluation of HRTF Magnitude Synthesis

- Very difficult to evaluate **perceptual quality** of HRTFs
 - Many more **degrees of freedom**: both spatial localization and perceived quality.
 - Not necessarily correlated.
 - Risk of '**uncanny valley**' effects: as realism increases, so too do the standards by which we judge the rendering quality.
- **Log spectral distance** used as an **objective measure** of magnitude response fit:

Direction	Frequency [Hz]	Best Classifier	Sparse Representation	HATS	Worst Classifier
Straight	50 – 8000	2.46	3.53	6.13	7.86
	0 – 20000	4.20	5.58	7.97	10.25
All	50 – 8000	4.32	4.49	7.35	7.85
	0 – 20000	9.48	9.88	13.77	14.93

HRTF Synthesis – Conclusions

- This is by no means a solved problem!
- First step is **reliable** and **consistent** measurement of HRTFs.
- Subjective testing for HRTFs is a big research problem
 - How is **perceived quality** linked to **localization accuracy**?
 - How soon does **listener fatigue** set in?
 - What is the nature of the **uncanny valley**?
- Objective measures equally in their infancy
 - Classic measures (PESQ, LCQA, LSD, MSE etc.) do not measure spatial component.

Contents

- Background on directivity patterns
- Part 1: Design of a measurement rig
 - Test signals
 - Loudspeaker placement
 - Extrapolation\interpolation of missing data
- Part 2: Practical Applications
 - Beamforming with Kinect for Xbox 360
 - Head-related Transfer Functions
- Conclusions

Conclusions

- Directivity patterns are everywhere!
- Many practical methods for measurements with real devices including:
 - Microphone (arrays)
 - Loudspeakers
 - Head-related transfer functions
- Some degree of choice on source signal, loudspeaker configuration and interpolation / extrapolation of missing data.
- Practical uses in:
 - Beamformer design (improved weights synthesis adds no overhead at runtime).
 - Personalization of HRTFs.
 - Also loudspeaker enclosure design.

Thank you!
Questions?