Perceptual Soundfield Reconstruction and Coherent Emulation

Zoran Cvetković

joint work with

Enzo De Sena and Huseyin Hacihabiboglu

KING'S College LONDON

Acknowlegements

James Johnston

Francis Rumsey

EPRSC

School of Natural and Mathematical Sciences, KCL

Aims and objectives

Ovearching goal

Low-count multichannel systems capable of

- transposing a listener to the actual space of an acoustic event
- providing a convincing illusion of an event in a desired space

Specific objectives

- understand principles involved
- general scalable and reconfigurable framework
- provide practical solutions

State of the art

	WFS	Low Channel Count	Binaural
Channel count	50+	5-10	2
Equipment Load	High	Commercially viable	Low
Psychoacoustics	None	Required	Critical
Sweet Spot	Large	Medium, small group	Small, individual

Low channel count systems

yet to achieve spatial realism that is possible with the available channels

Commercial surround sound systems

- based on the legacy of sound production for the film industry
- focus on attention grabbing effects and a general ambiance feel

Drawbacks

- heavily mixed and inconsistent with the acoustics of a physical space
- complex empirical methods reliant primarily on tonmeister's skills

Ambisonics

- aims primarily at physical approximation in the centre
- very limited sweet spot

Towards achieving convincing spatial sound

Minimal number of channels?

- 3 channels - minimum for frontal perspective [Fletcher 1940s]
- 5 channels - minimum for diffuse sound [Ando and Kurihara 1986]
- 5 channels seems to be the minimum

Optimal playback scheme?

- mixing blurs the auditory perspective [Fletcher 1940s]
- each channel plays back the signal of the corresponding microphone

Optimal acquisition of sound field information?

- microphone array for acquisition of necessary perceptual cues
- Perceptual Sound Field Reconstruction [Johnston et al. 2000]

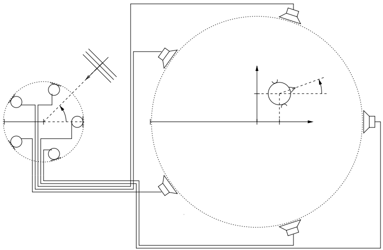
Johnston's system

Interaural Level Difference (ILD)

The difference in level between the signals at the two ears

Interaural Time Difference (ITD)

The time shift between the signals at the two ears



System specifications

- 5 channels
- 31cm array diameter - approximately the path length between the ears
- microphone directivity - -3dB at 72° effectively 0 at 144°

Johnston's system

Unresolved issues

- are the captured ITD and ILD cues really reproduced
- is the proposed array radius optimal
- are the proposed microphone polar patterns optimal
- generalisation to systems with more than five channels
- generalisation to irregular configurations

Departure from Johnston's original technology

- do not capture ITDs and ILDs, but required play-back

ICTD

inter-channel time differences

ICLD

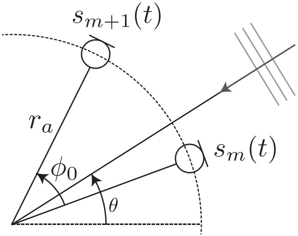
inter-channel level differences

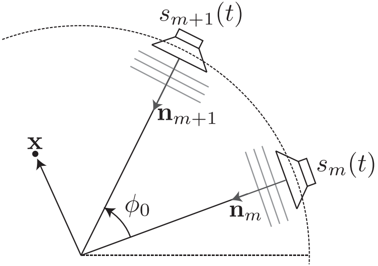
Generalization

- systems with more than 5 channels
- irregular circular configurations

Technical issues: perceptual sampling

- microphone polar patterns
- array diameter





Objective analysis

- active intensity fields for complex plane monochromatic waves

$$\mathbf{I}_c(\mathbf{x}) = \frac{1}{2}p(\mathbf{x})\mathbf{v}^*(\mathbf{x}) = \mathbf{I}_a(\mathbf{x}) + j\mathbf{I}_r(\mathbf{x}) \quad (1)$$

Number of active channels

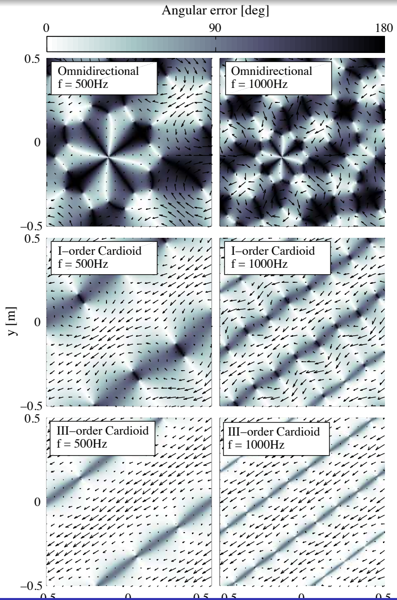
- more than two channels – reduced sweet spot
- three active channels – decreased locatedness [Lee and Rumsey 2005]

Array radius

- translation of the active intensity stripes

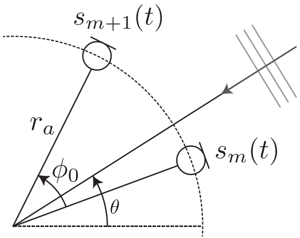
Number of channels

- the width of active intensity stripes



Inter-channel time difference – ICTD

- depends on the direction of incidence of sound wave
- the dependence set by array radius and microphone angular spacing

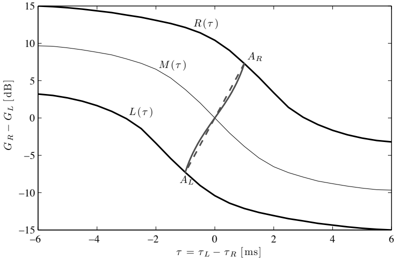


$$\tau(\theta) = \frac{2r_a}{c} \sin \frac{\phi_0}{2} \sin \left(\frac{\phi_0}{2} - \theta \right)$$

$$\tau_{\text{max}} = \frac{2r_a}{c} \sin^2 \frac{\phi_0}{2}$$

Inter-channel level difference – ICLD

- completely determined by microphone polar patterns
- design polar patterns so to ensure correct ICLD



Coincident array – ID

$$\Phi(\theta) = \frac{\Gamma_{m+1}(\theta)}{\Gamma_m(\theta)} = \frac{\sin(\theta - \phi_m)}{\sin(\phi_{m+1} - \theta)}$$

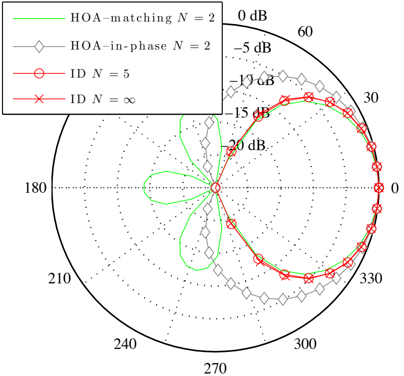
equivalent to tangent panning law

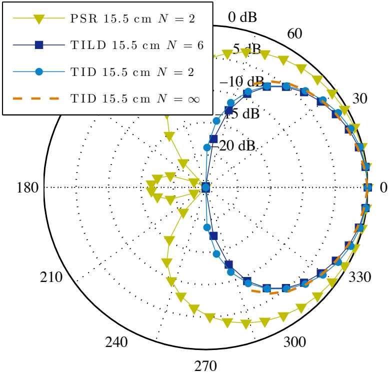
Non-coincident array – TID

$$\Phi(\theta) = \frac{\sin(\theta - (\phi_m - \beta))}{\sin((\phi_{m+1} + \beta) - \theta)}$$

Non-coincident array – TILD

$$\Phi(\theta) = 10^{\frac{\kappa_o}{20}\tau(\theta)}$$


HOA−matching $N = 2$
HOA−in-phase $N = 2$
ID $N = 5$
ID $N = \infty$
0 dB
−5 dB
−10 dB
−15 dB
−20 dB
0
30
60
180
210
240
270
300
330

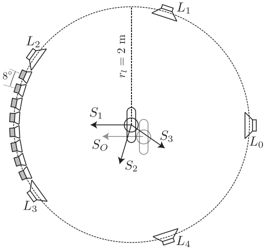

PSR 15.5 cm $N = 2$
TILD 15.5 cm $N = 6$
TID 15.5 cm $N = 2$
TID 15.5 cm $N = \infty$
0 dB
−5 dB
−10 dB
−15 dB
−20 dB
0
30
60
180
210
240
270
300
330

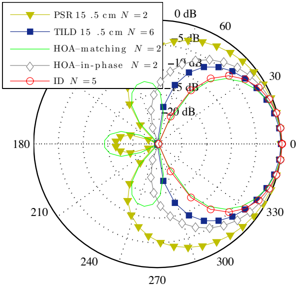
Array radius

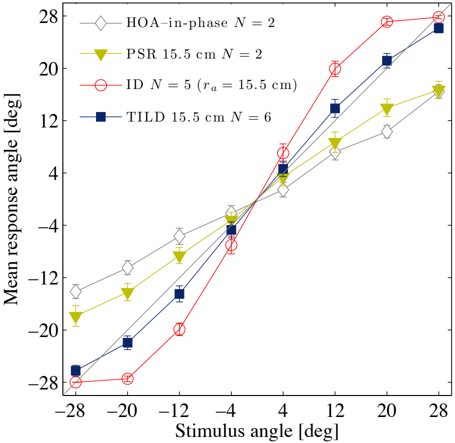
- given an arbitrary array radius, accurate auditory perspective is achieved by means of appropriate microphone polar patterns
- array radius is thus a free parameter
- it can be used to optimise some other qualities of reproduced sound or used in a creative manner to achieve some desired effects

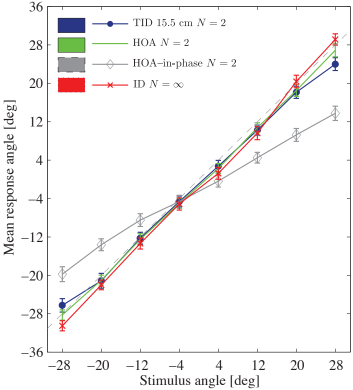
Example

- many ICTD/ICLD pairs which render a given source direction
- not all of these pairs are natural
- array radius can be used to optimise a naturalness measure
- 15.5 cm arrays provide natural ICTD/ICLD pairs

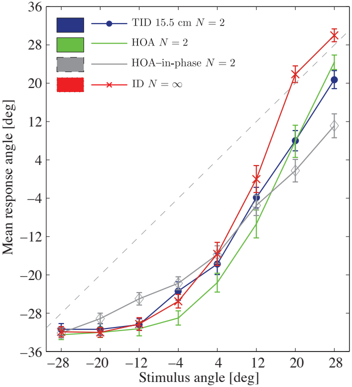

L_1
L_2
$8°$
$r_l = 2$ m
S_1
S_O
S_3
S_2
L_0
L_3
L_4


PSR 15 .5 cm N = 2
TILD 15 .5 cm N = 6
HOA–matching N = 2
HOA–in–phase N = 2
ID N = 5
0 dB
−5 dB
−10 dB
−15 dB
−20 dB
60
30
0
180
210
240
270
300
330


HOA–in-phase N = 2
PSR 15.5 cm N = 2
ID N = 5 (r_a = 15.5 cm)
TILD 15.5 cm N = 6
Mean response angle [deg]
Stimulus angle [deg]
28
20
12
4
−4
−12
−20
−28



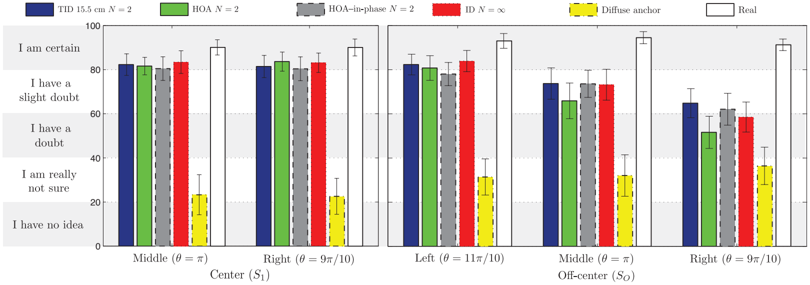
On-centre



30-cm off-centre

Locatedness

"The degree to which an auditory event can be said to be clearly in a particular location."



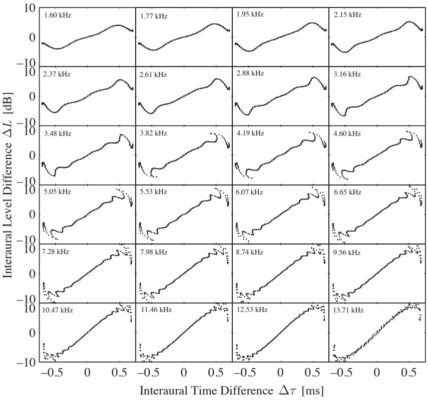
Computational model for predicting locatedness

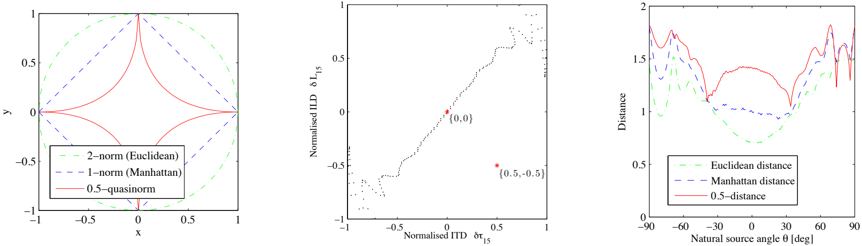
Observation

- ILD-ITD pairs come in specific pairs for natural (point-like) sources

Hypothesis

- When ILD-ITD pairs are far from natural curves result in lower locatedness



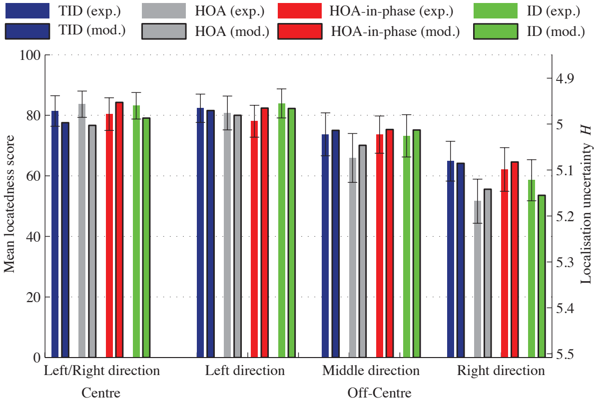


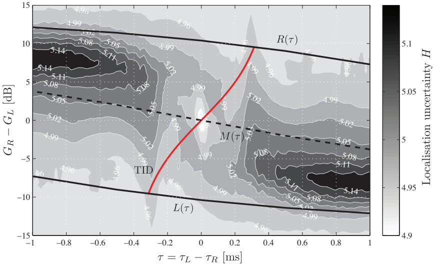
The model

1. Psychoacoustic filterbank (critical bands + neural transduction)
2. Score function: 0.5-distance from natural ILD-ITD pairs
3. Average score function over critical bands
4. Locatedness estimate: entropy of averaged score function

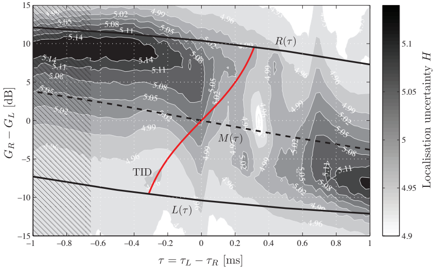
Validation

- Comparison with locatedness experiment
- Very strong correlation (0.94) with subjective listening tests





on-centre



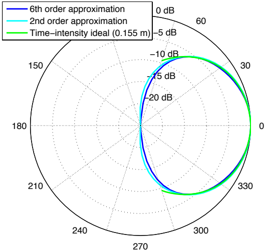
10 cm off-centre

Conclusion

- Coincident arrays lie on a line with high variation of locatedness
- Non-coincident arrays have higher and more uniform locatedness in off-centre listening positions

Recording in the real world

- PSR requires new types of higher order polar patterns

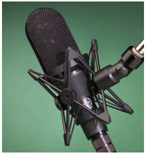


Microphone directivity pattern

$$\Gamma(\theta) = a_0 + a_1 \cos(\theta) + a_2 \cos^2(\theta) + \ldots + a_N \cos^N(\theta)$$

- $N \geq 2$ – still research grade, but growing interest

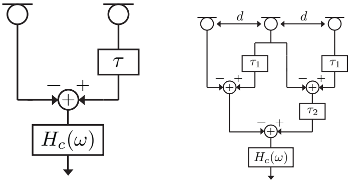
- Soundfield: $N = 1$ – hi-fi but very restricted
- Eigenmike: $N \leq 4$ – high cost
- Differential microphones – low-cost alternative

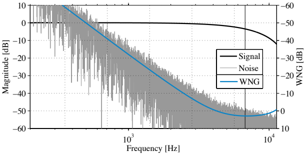

Soundfield ($N = 1$)

Eigenmike ($N \leq 4$)

Higher order diff. mics. and operational bandwidth

- 1-st order structure: difference between two omni outputs, plus a correction filter
- N-th order differential microphones: cascades of 1st order sructures
- New 2nd-order general structure





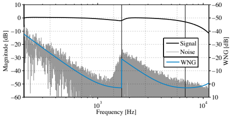
Limits of operational bandwidth

- high frequency bounded by spatial aliasing (the smaller d, the better)
- low frequency bounded by noise sensitivity (the larger d, the better)

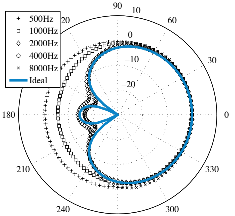
Example: 3rd order differential microphones

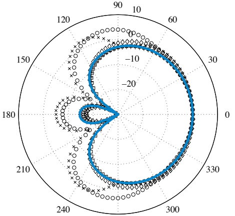
Solution to operational bandwidth limitations

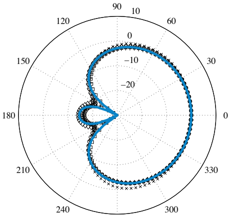
interleave structures with different distances between individual omnidirectional elements



- 3rd order structures with 7 lavalier omnis
- 2 structures with $d_1 = 0.5$ in and $d_2 = 2$ in.
- Good performance over 5 octaves







Coherent sound field simulation

Applications

- sound for production for film and TV
- sound production for music industry
- video games
- virtual reality
- architectural design

A major challenge

- efficient simulation of different spaces
- real-time rendition of dynamic scenes

Room simulators

Many methods – trade-off between complexity and accuracy

DSP effects easy, fast, but inaccurate

Statistical very fast, but not directly related to room properties

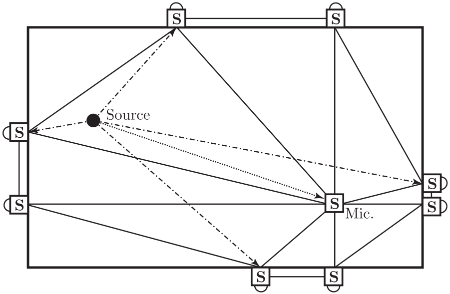
FDN fast, but tuning is indirect and trial-and-error

Synth. Reverb. fast, sounds good, but careful tuning of parameters

Conv. Reverb. sounds real, but require actual recordings

Ray Tracing accurate, but heavy load (especially for later parts of RIR, which is also the less important perceptually)

DWM wave equation solution, but very heavy load



Karjalainen-type DWN

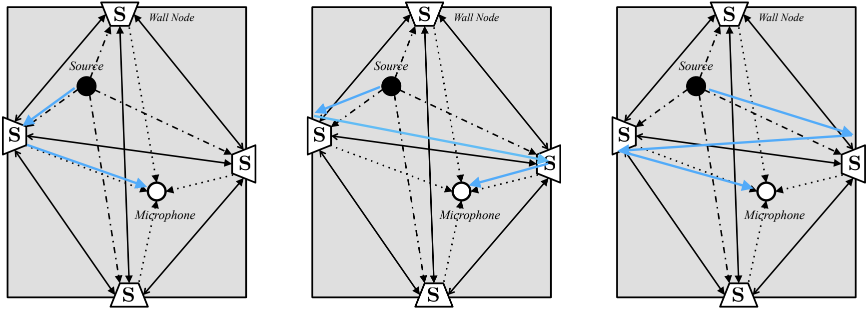
- As fast as FDN
- Low memory requirement

Drawbacks

- Careful heuristic tuning of model parameters to achieve naturally-sounding results
- Why this network? Unclear optimality criterion

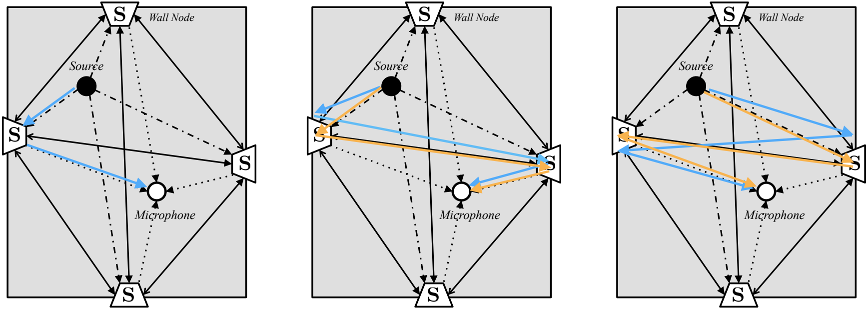
Scattering Delay Network

- Network is set to render correctly LOS and first-order reflections (most important for perception of size and shape of enclosure), while small approximation of second- and higher-order reflections.
- $\Rightarrow$ **perceptual approximation of ray-tracing/image-source method**

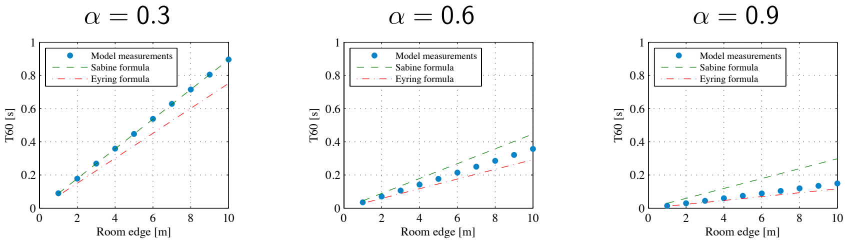


Scattering Delay Network

- Structure set to render correctly LOS and first-order reflections (most important for perception of size and shape of enclosure), while approximating second- and higher-order reflections.
- $\Rightarrow$ **perceptual approximation of ray-tracing/image-source methods**



Preliminary objective assessment



- Sabine and Eyring: popular formulas for reverberation time.
- Reverberation time always between the two formulas.
- Apply wall filters at output of scattering nodes.
- Good approximation of low-frequency room modes.

Summary

Perceptual sound field reconstruction

- a general systematic framework
- applicable to any number of channels
- applicable to irregular configurations

Current and future work

- render most natural ICTD/ICLD pairs
- automatic methods for generating psychoacoustic curves
- extensions to 3D systems