Capturing 360° VR Audio Using Equal Segment Mic Array (ESMA)

Dr Hyunkook Lee
h.lee@hud.ac.uk
Applied Psychoacoustics Lab (APL)
University of Huddersfield, UK

University of HUDDERSFIELD

- Background
- Psychoacoustics for microphone array design
- Proposed technique
- Perceptual evaluations and Demo

Background

University of HUDDERSFIELD

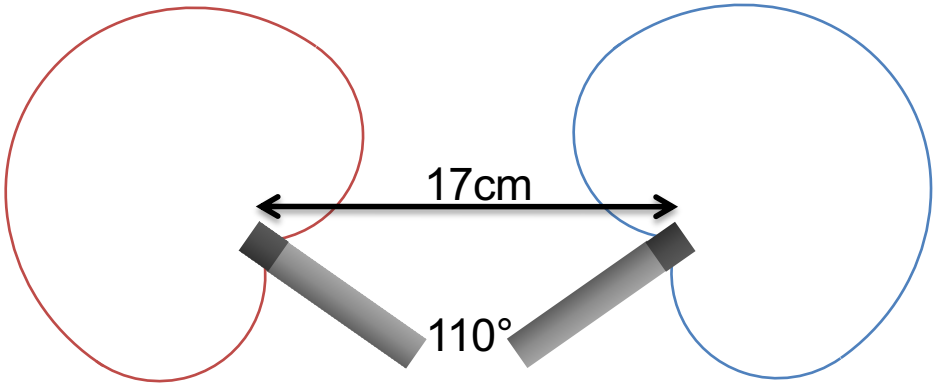
- Stereophonic microphone techniques

	Coincident pair	Near-Coincident	Spaced pair
Polar pattern	Uni-directional	Uni-directional	Omni-directional
Configuration	XY, Blumlein, MS	ORTF, NOS	AB, Decca Tree
Stereo cue	ICLD (Leve diff)	ICLD + ICTD	ICTD (Time diff)
Localisability	High	Mid-High	Low-Mid
Spaciousness	Low	Mid-High	High

Background

University of HUDDERSFIELD

- Near-coincident microphone techniques
  - ORTF, NOS, etc.
  - Popular microphone techniques among recording engineers.
  - Best of both worlds! (Localisability & Spaciousness).



ORTF configuration

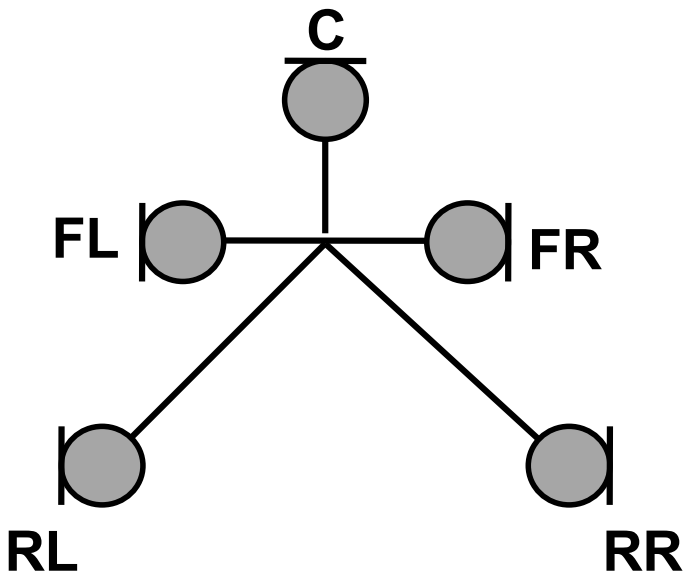
Background

- Brown or White?

Background

University of HUDDERSFIELD

- Near-coincident microphone techniques
  - Also popular for surround microphone techniques.
  - Generally, directional microphones are used for balanced localisation and spatial impression.
  - OCT [Theile/Wittek], ICA [Williams], Fukada Tree, etc.

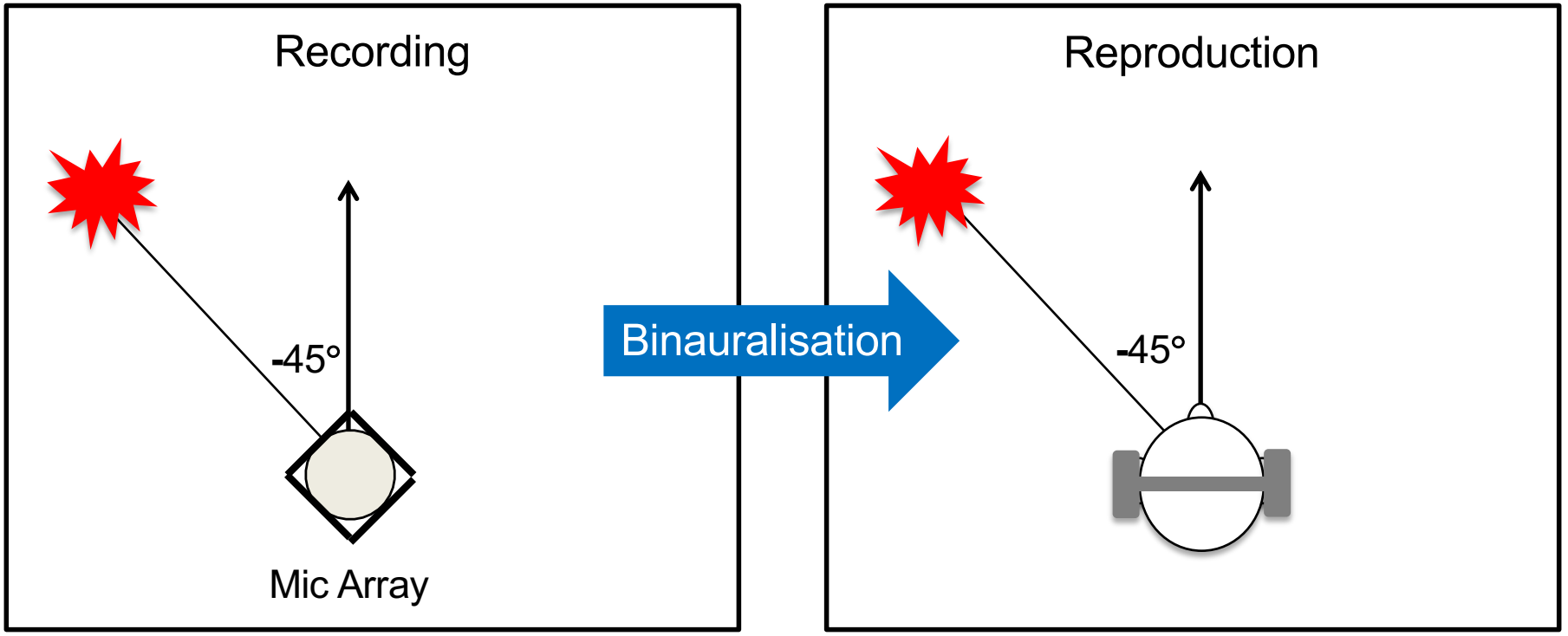


How to design a near-coincident mic array for 360VR?

Requirements for VR mic arrays

University of HUDDERSFIELD

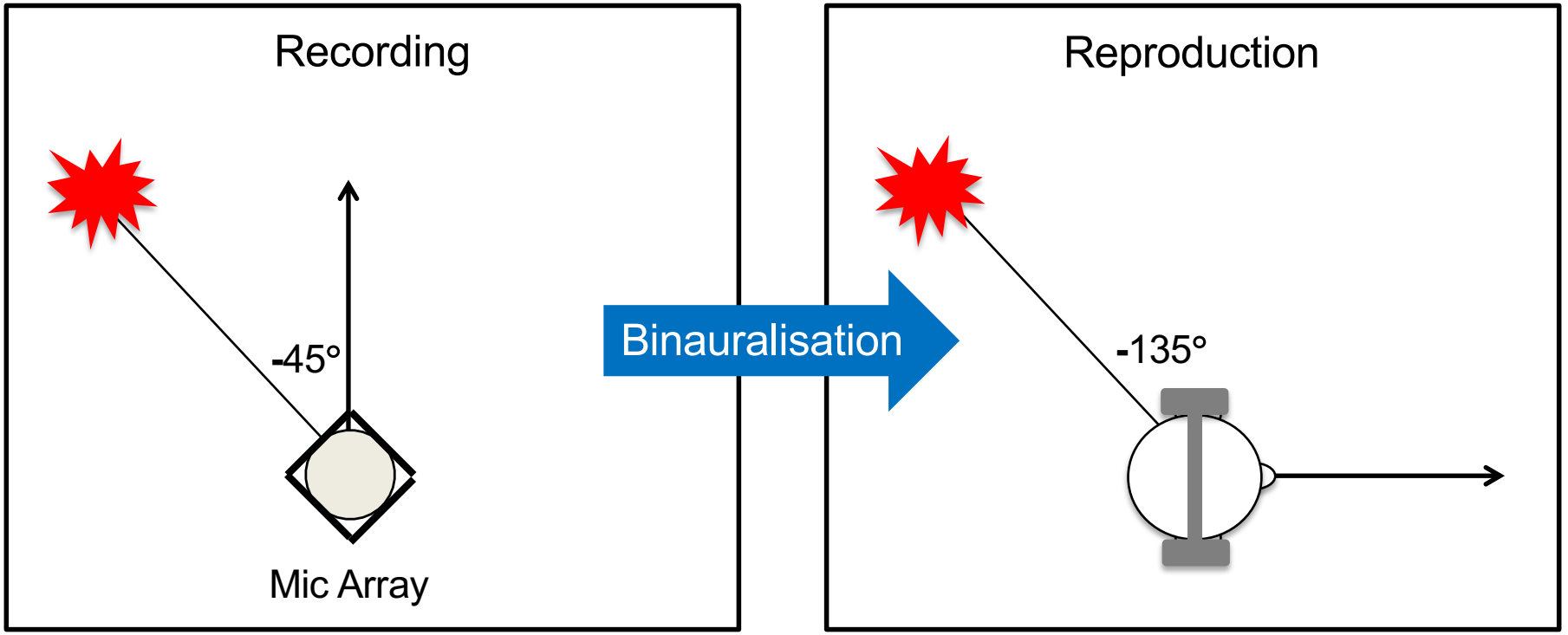
1. The actual and perceived image positions should match (ideally!).



Requirements for VR mic arrays

University of HUDDERSFIELD

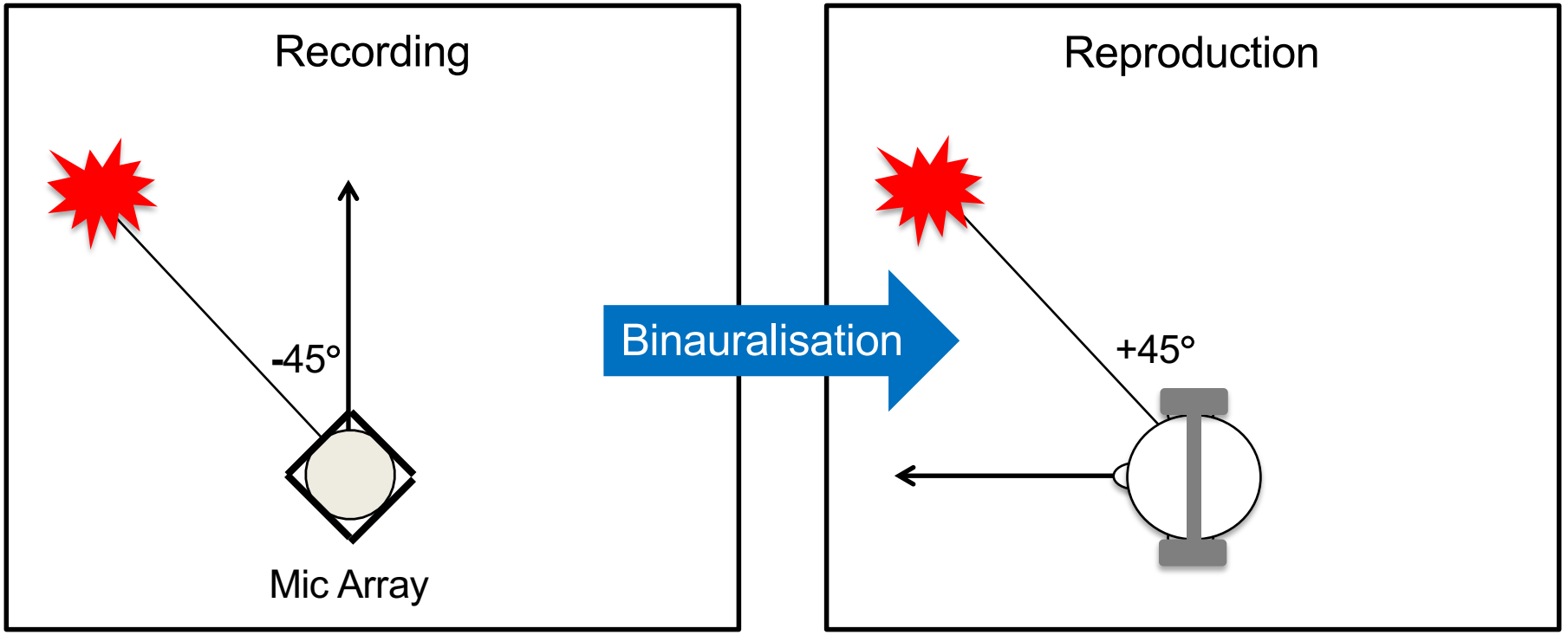
2. The perceived source position should stay the same when the head rotates.



Requirements for VR mic arrays

University of HUDDERSFIELD

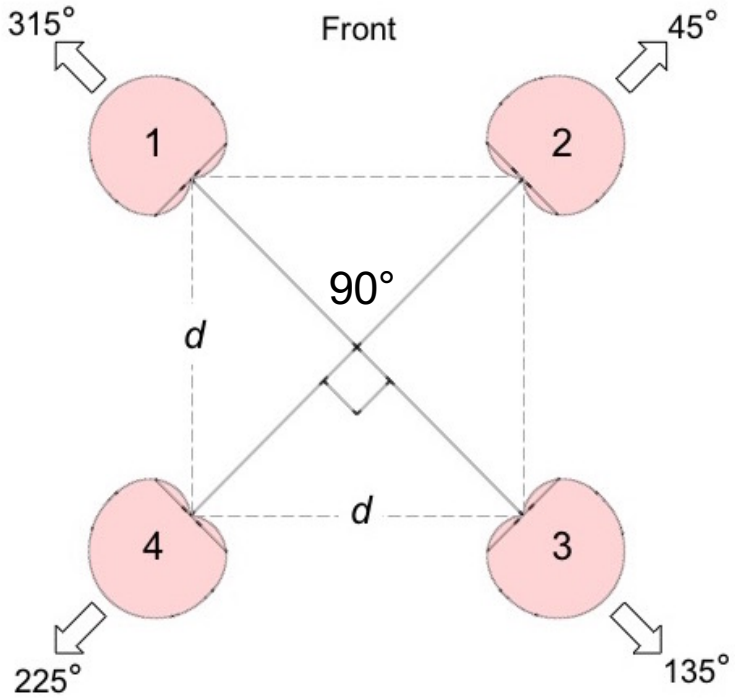
2. The perceived source position should stay the same as the head rotates.



Proposed technique

- Equal Segment Microphone Array (ESMA)
  - The original concept by Williams [1991] developed for quadraphonic surround sound.
  - IRT-Cross for ambience capture.

→ Adapted for VR here.



Proposed technique

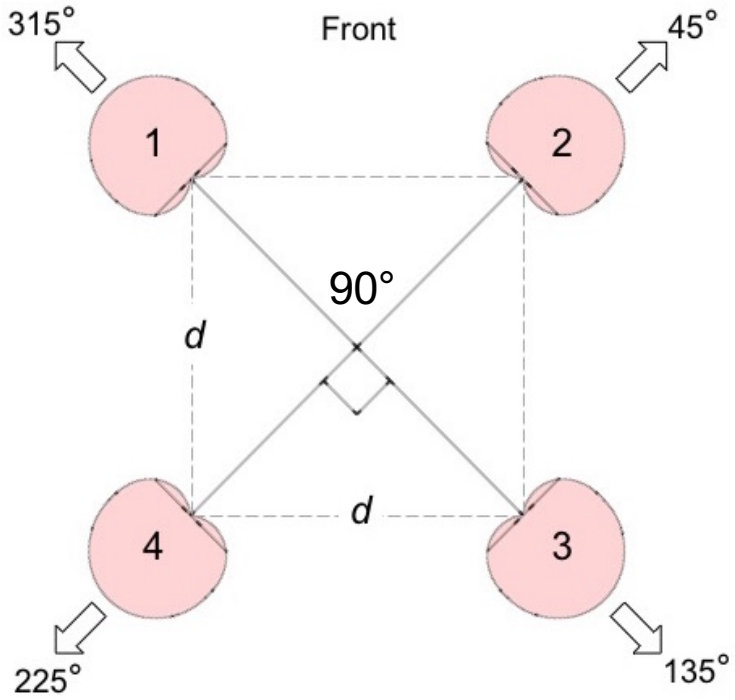
University of HUDDERSFIELD

- Equal Segment Microphone Array (ESMA)

Constraints

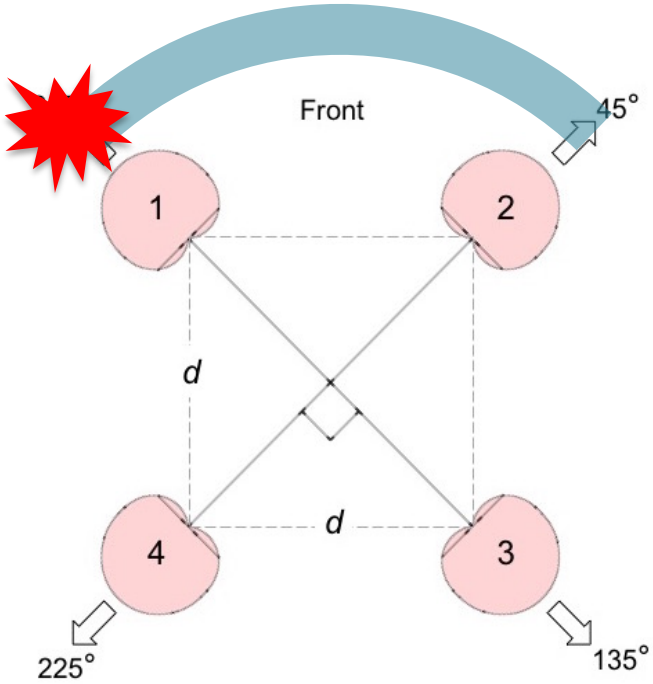
– Equal subtended angle (90°) and mic spacing for all stereo segments

– The stereophonic recording angle (SRA) for each segment should match the subtended angle of the segment (90°).

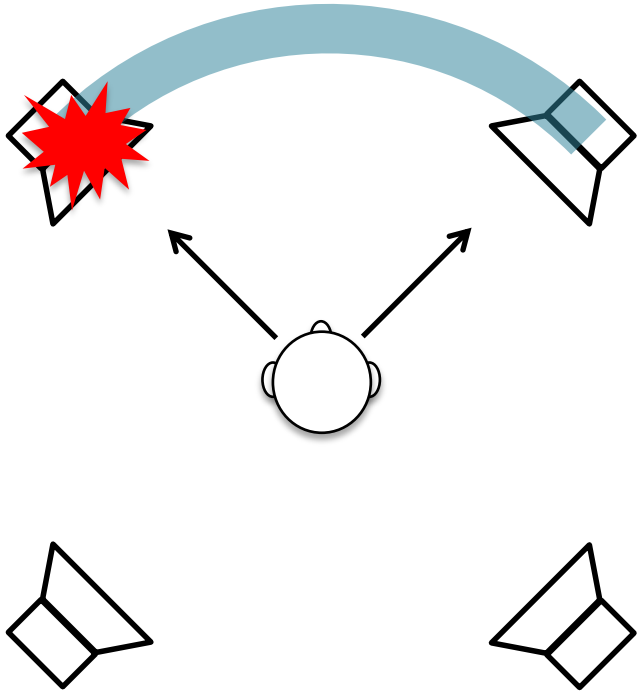


Proposed technique

- Stereophonic recording angle (SRA)
  - Coverage of a sound field that produces sufficient ICLD and ICTD for a full image shift between two loudspeakers.

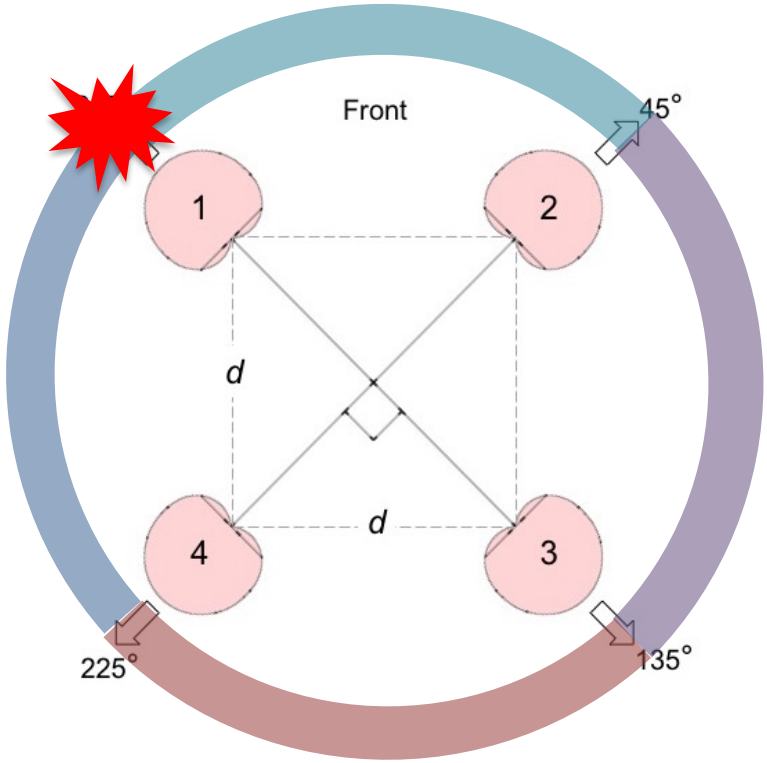


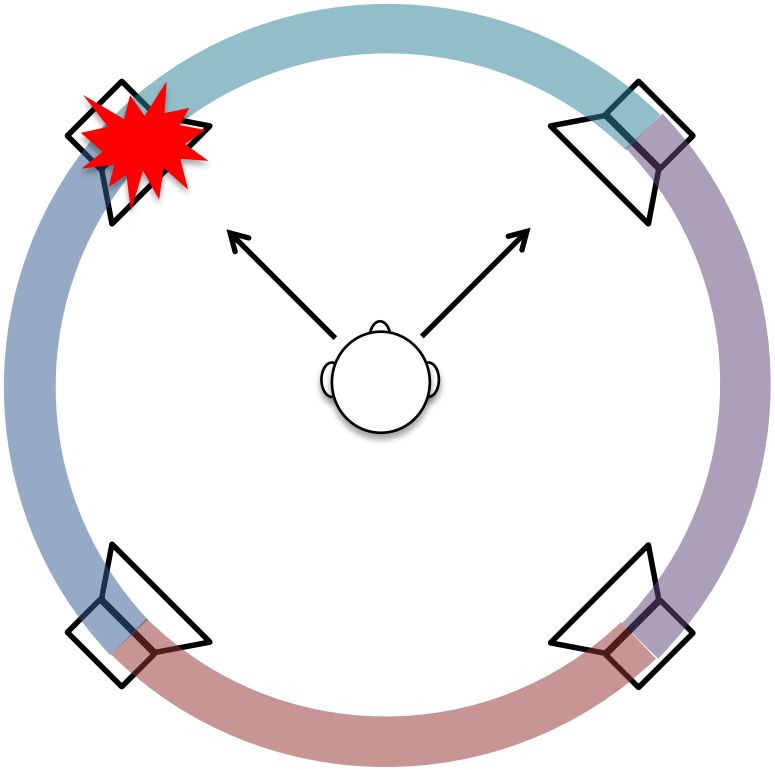
SRA = 90°



Proposed technique

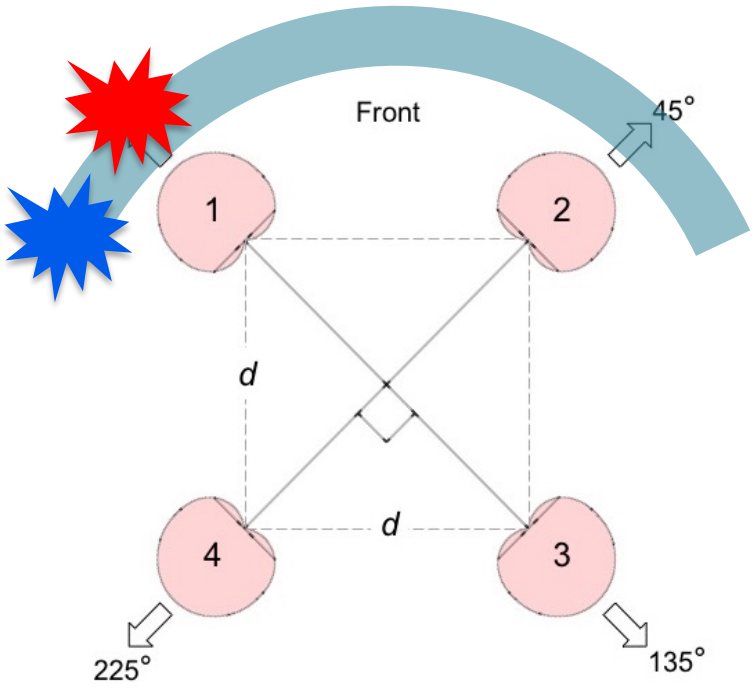
- SRA should not overlap between each stereo segment.
  - To maintain the perceived image position with head rotation.

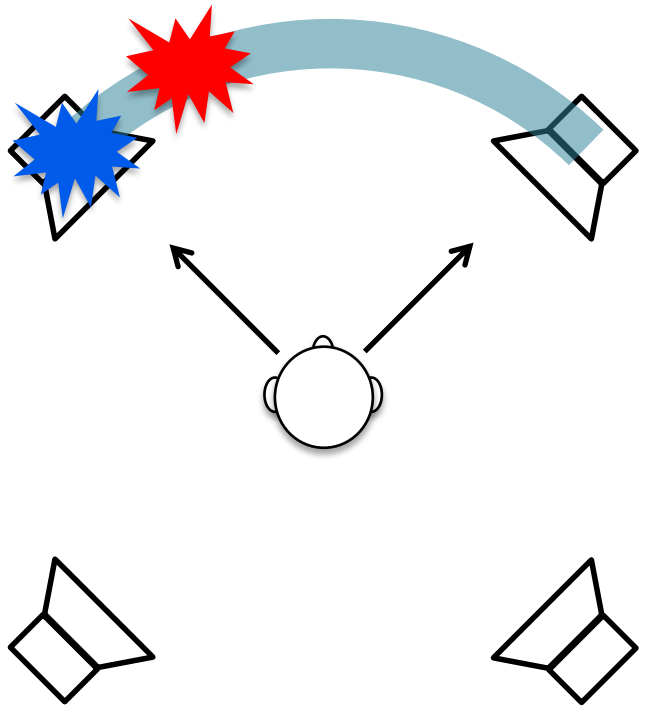




Proposed technique

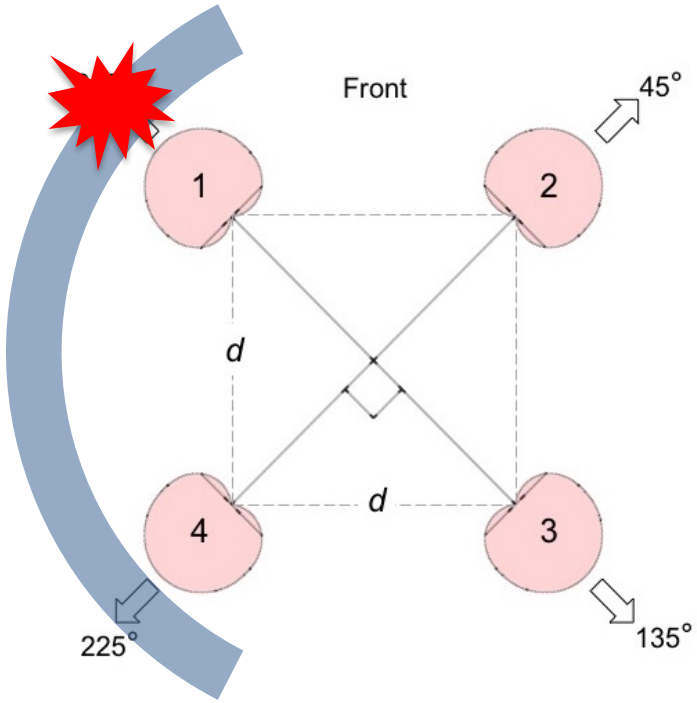
- Problem with an SRA overlap
  - Localisation of the target source at a narrower angle.
  - Image position moves as the head rotates.

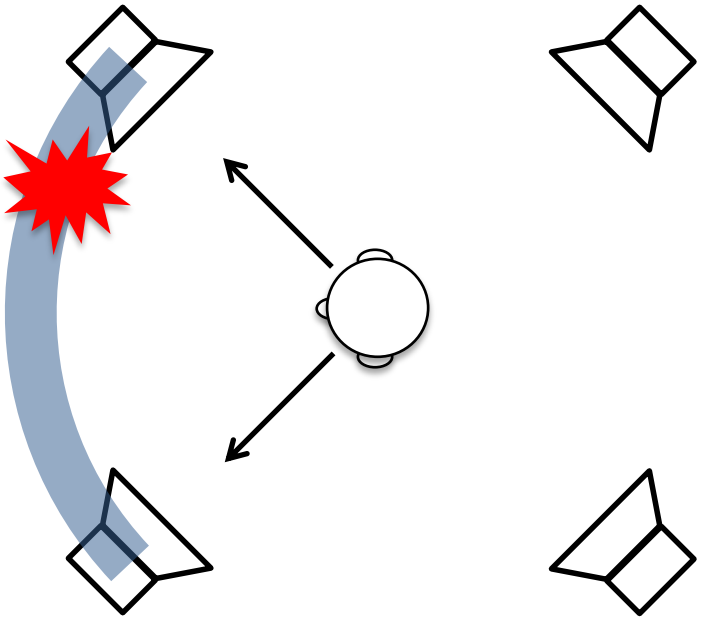




Proposed technique

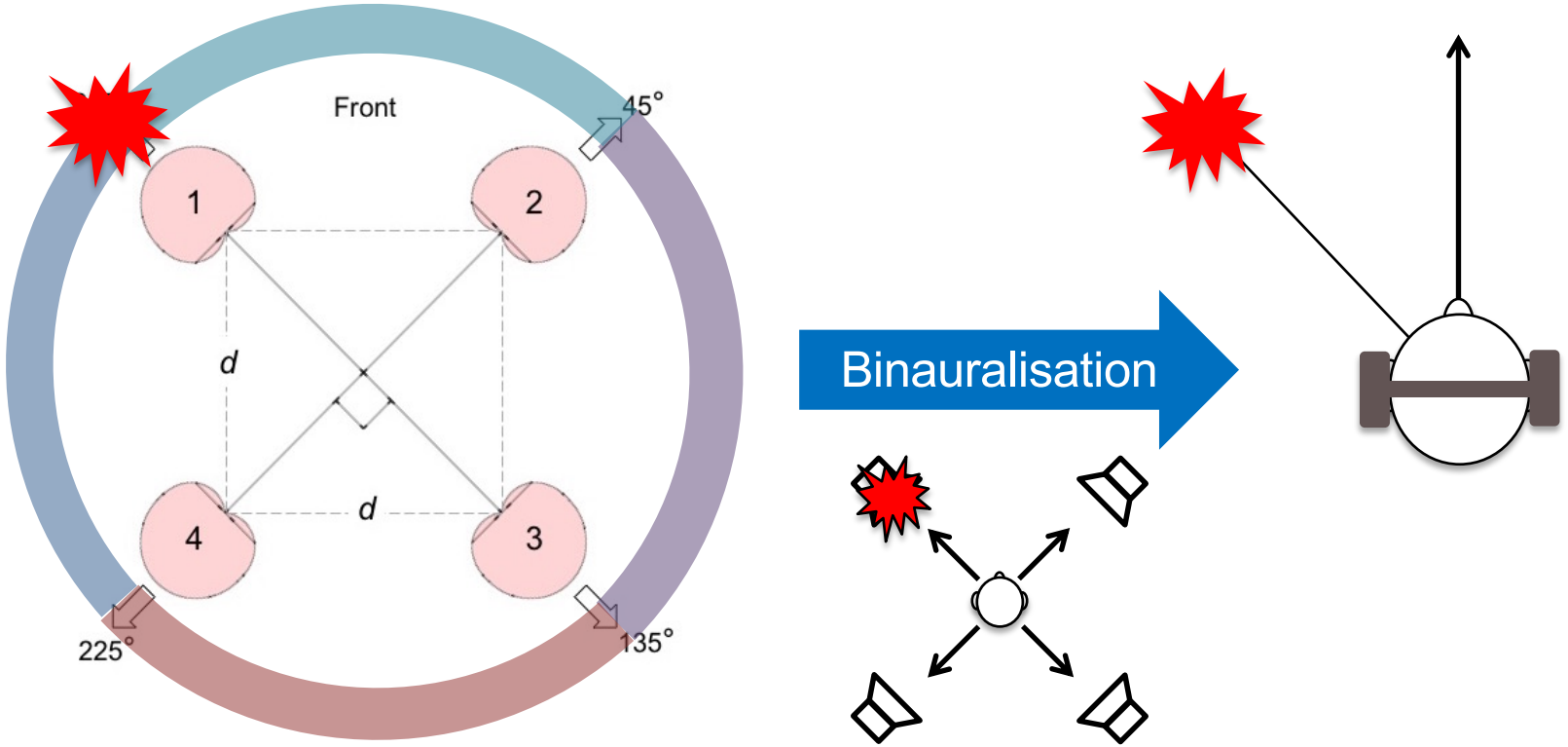
- SRA should not overlap between each stereo segment.
  - To maintain the perceived image position with head rotation.



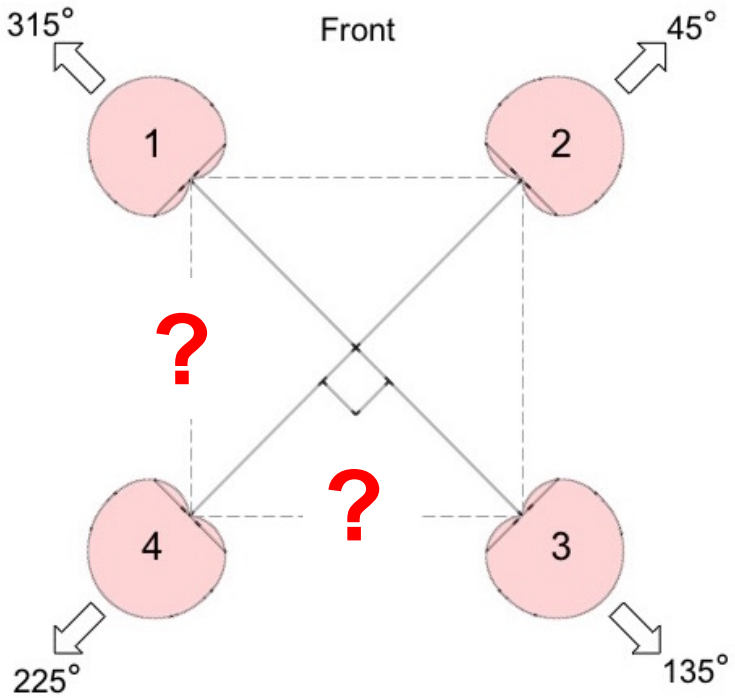


Proposed technique

- Binauralisation of ESMA
  - Convolve the mic signals with quadraphonic HRIRs.



Now, what should be the spacing between the microphones?

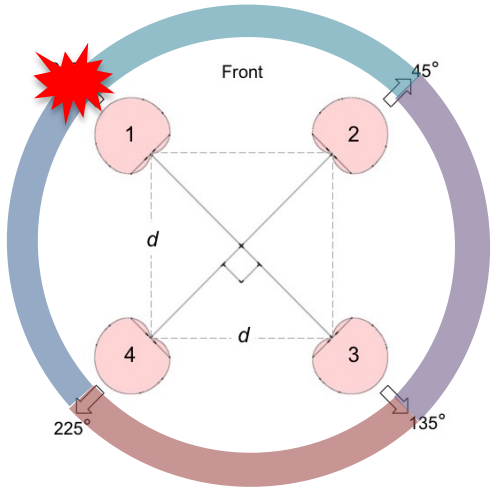


Psychoacoustic principles

University of HUDDERSFIELD

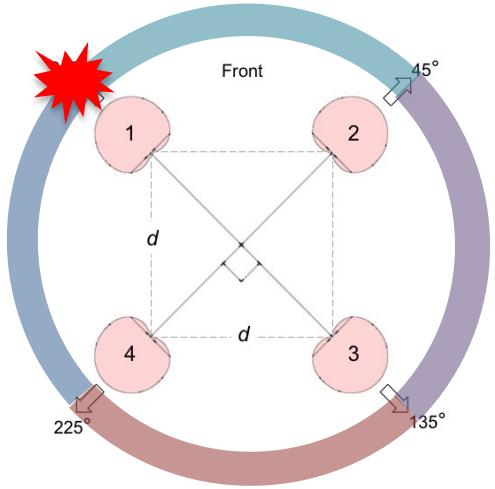
- ICTD and ICLD trade-off
  - The spacing and angle between two directional microphones determine interchannel time and level differences encoded between the signals.
  - ICTD and ICLD complement each other in shifting the perceived image [Theile 2001].

Total image shift = shift(ICTD) + shift(ICLD)



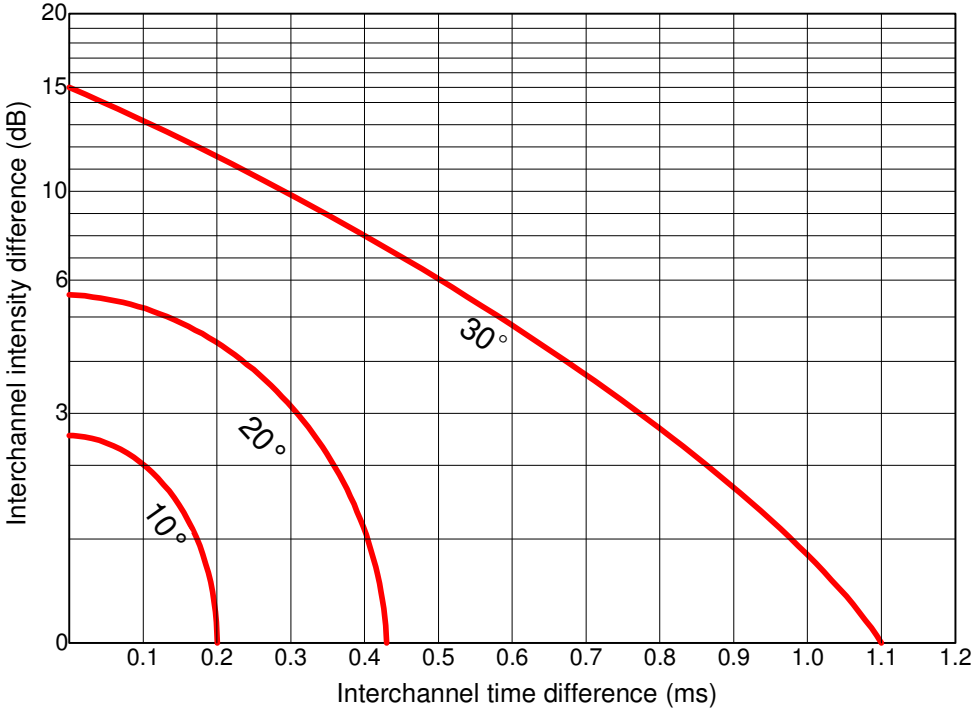
Psychoacoustic principles

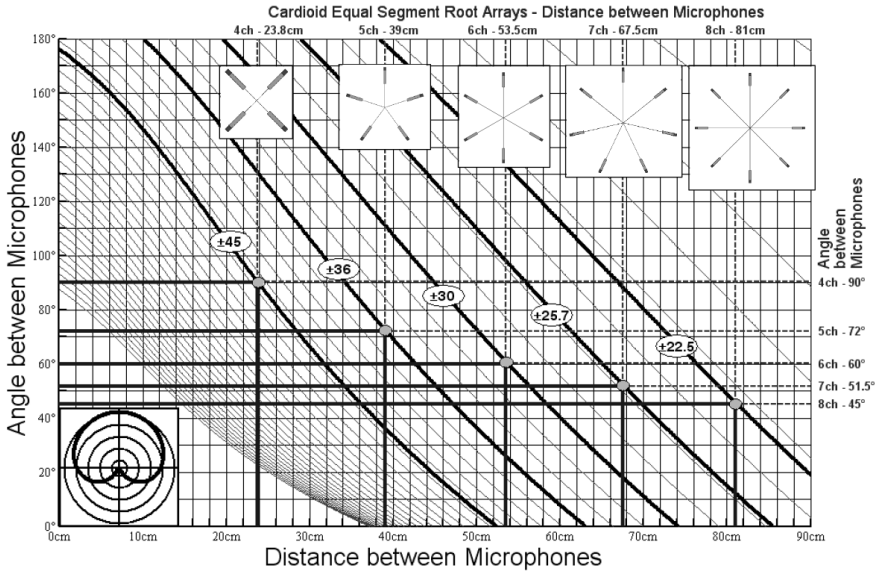
- ICTD and ICLD trade-off
  - With the 4ch ESMA design, the goal is to achieve the SRA of 90° for each segment.
  - The subtended angle is given as 90°.
  - The spacing needs to be determined so that the resulting ICTD and ICLD are just enough to produce the full ±45° image shift.
  - There exist several ICTD-ICLD trade-off models.



Psychoacoustic principles

- Williams [1987]
  - Interpolation of ICTD and ICLD data obtained for 10°, 20° and 30°.
  - Used to calculate the SRA for different mic techniques (so-called the Williams curves)



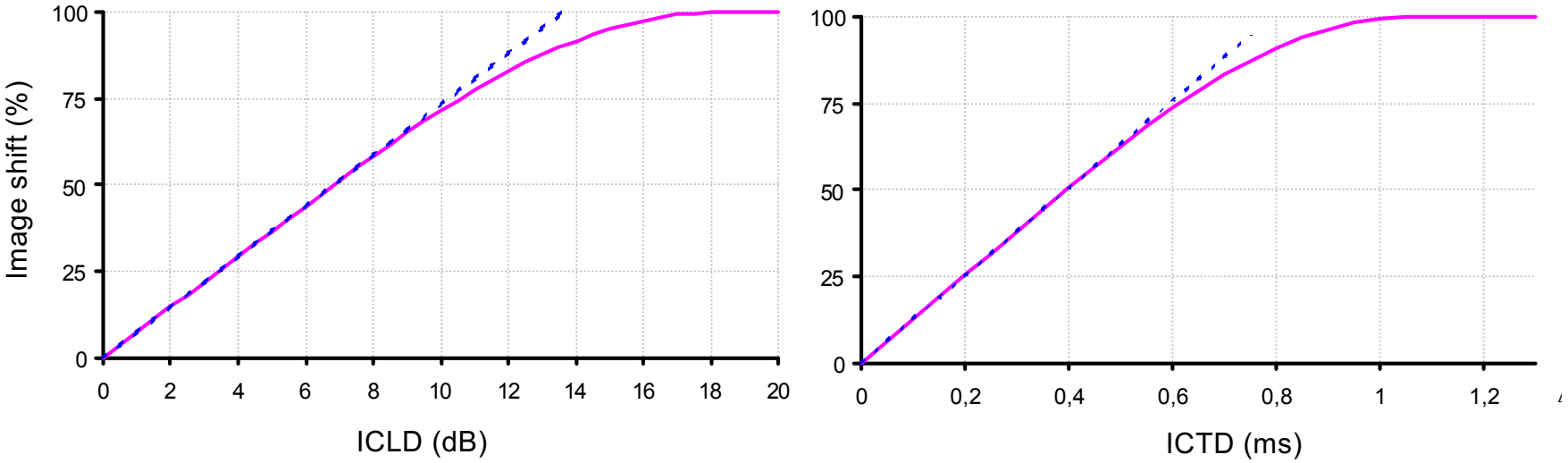


Williams [2008]

Psychoacoustic principles

University of HUDDERSFIELD

- Wittek/Theile [2001]
  - Total image shift = shift(ICLD) + shift(ICTD)
  - Up to the linear 75% of the whole shift region.
  - Basis for the “Image Assistant” tool for mic array design.



Wittek [2001]

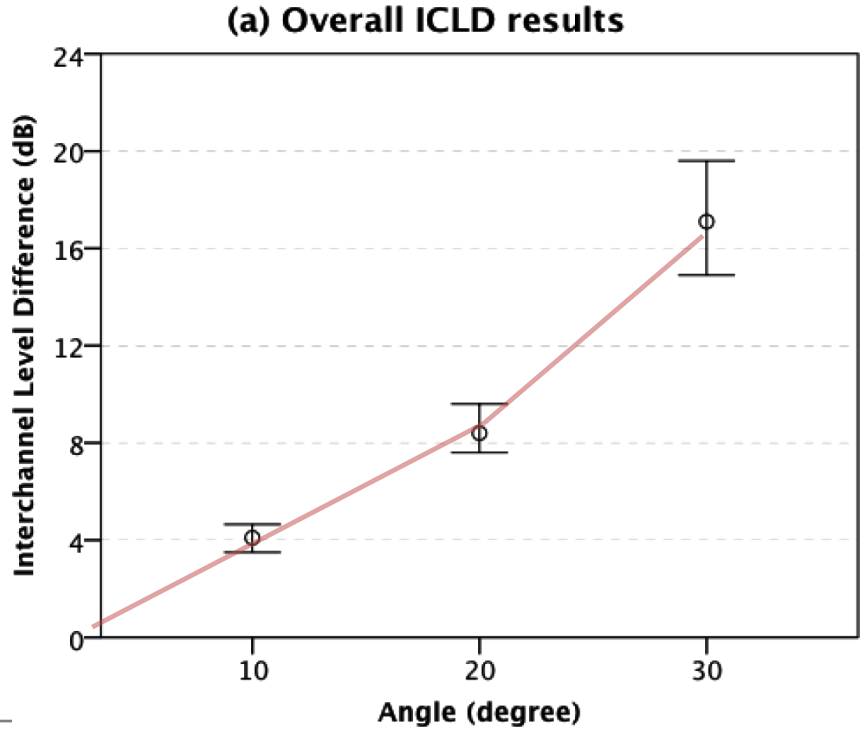
Psychoacoustic principles

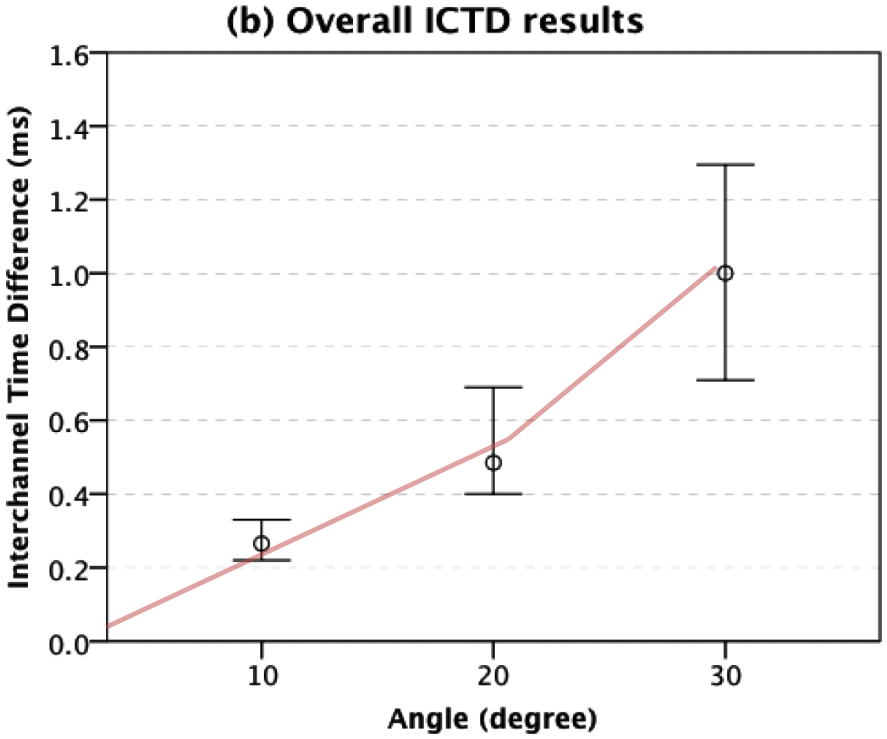
University of HUDDERSFIELD

- Lee and Rumsey [2013]
  - Two linear shift regions: 0° to 20° & 20° to 30°

$$ICLD(\alpha) = \begin{cases} 0.425\alpha & [dB],\ \alpha \leq 20 \\ 0.85\alpha - 8.5\ [dB],\ 20 < \alpha \leq 30 \end{cases}$$

$$ICTD(\alpha) = \begin{cases} 0.025\alpha & [ms],\ \alpha \leq 20 \\ 0.05\alpha - 0.5\ [ms],\ 20 < \alpha \leq 30 \end{cases}$$

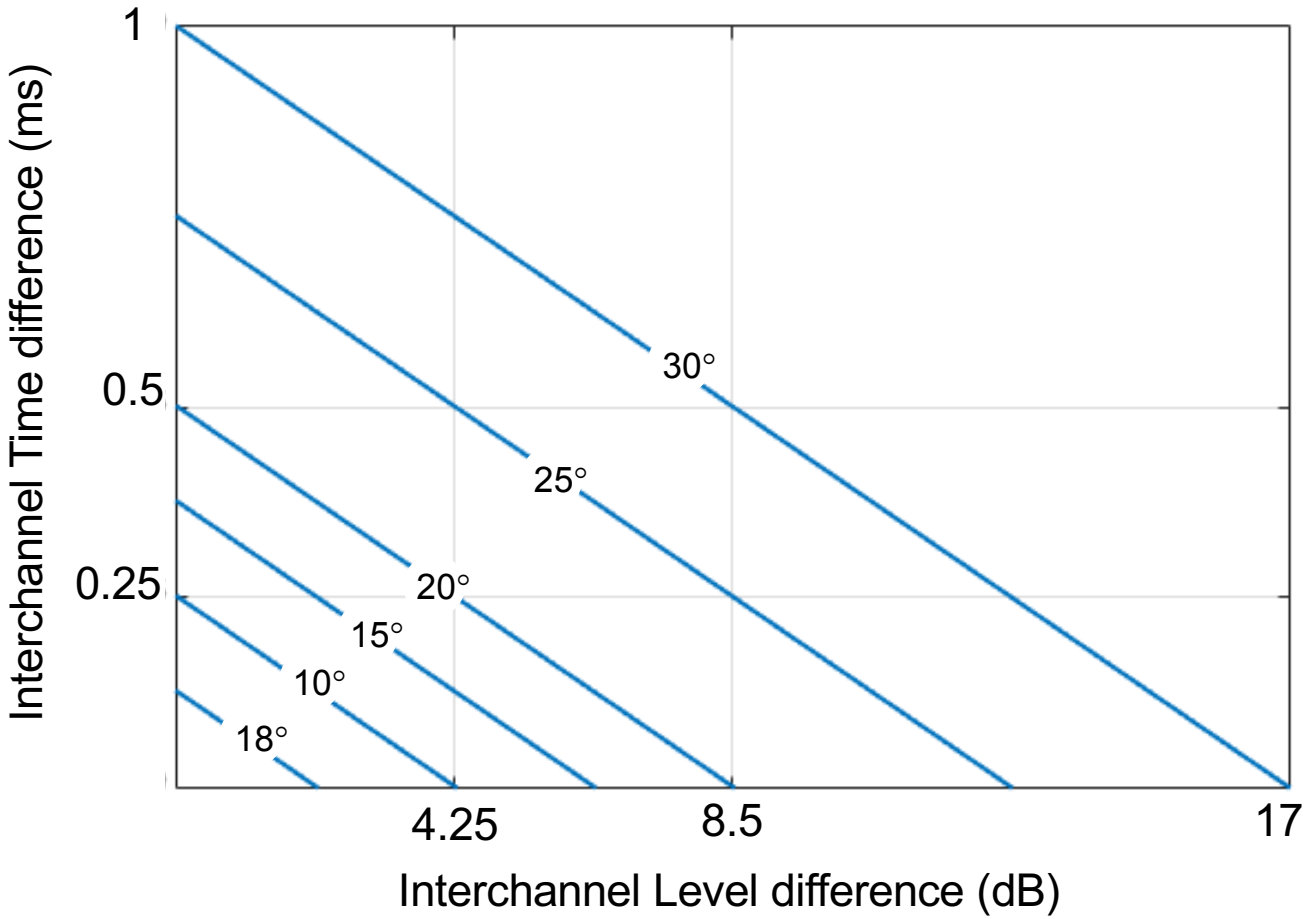




Psychoacoustic principles

University of HUDDERSFIELD

- Linear ICTD-ICLD Trade-off functions [Lee 2016]



Psychoacoustic principles

University of HUDDERSFIELD

- Limitations
  - The previous models are based on the **60°** loudspeaker base angle.
  - They assume that the same model can apply to other base angles also.
  - e.g., 17dB for 30° shift with the 60° base → 17dB for 45° shift with the 90° base.

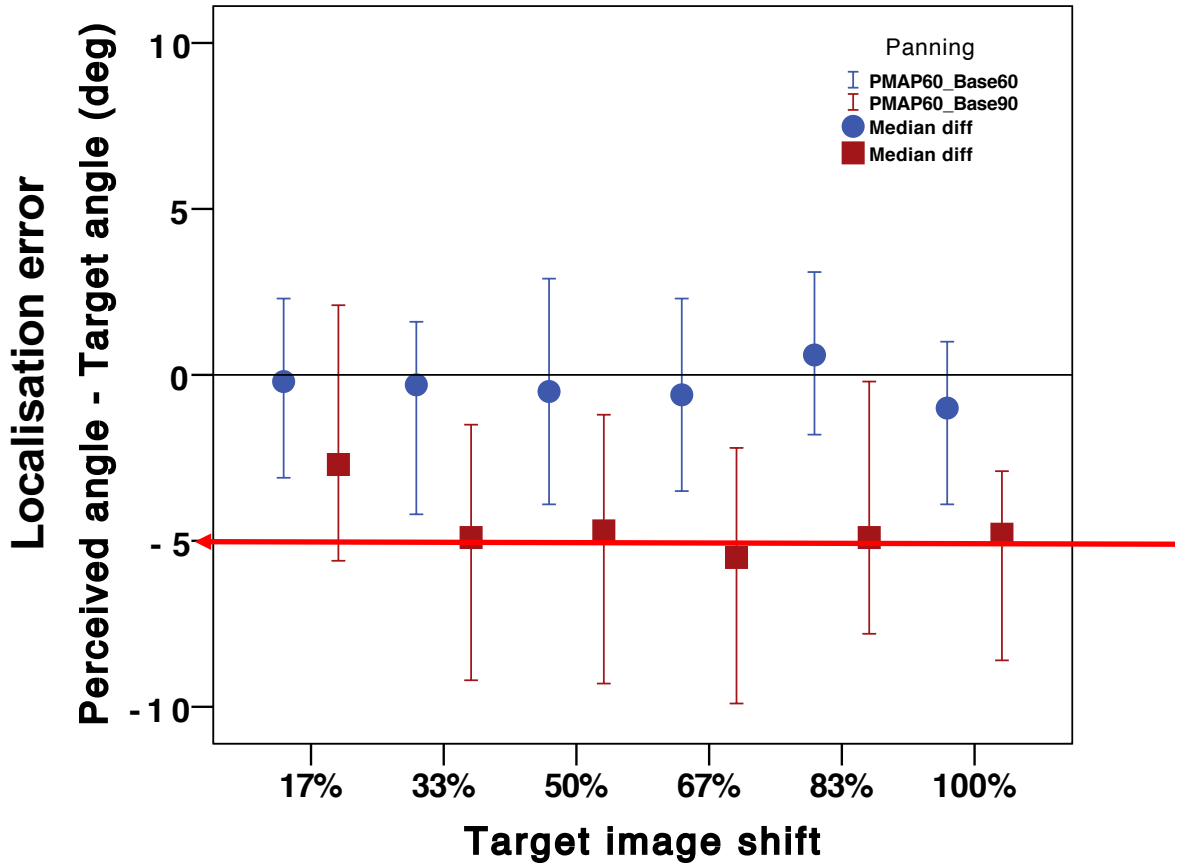
Psychoacoustic principles

University of HUDDERSFIELD

- Limitations
  - However, this does not work in practice!
  - More ICTD and ICLD values are required to achieve a full image shift with the 90° loudspeaker setup.

Psychoacoustic principles

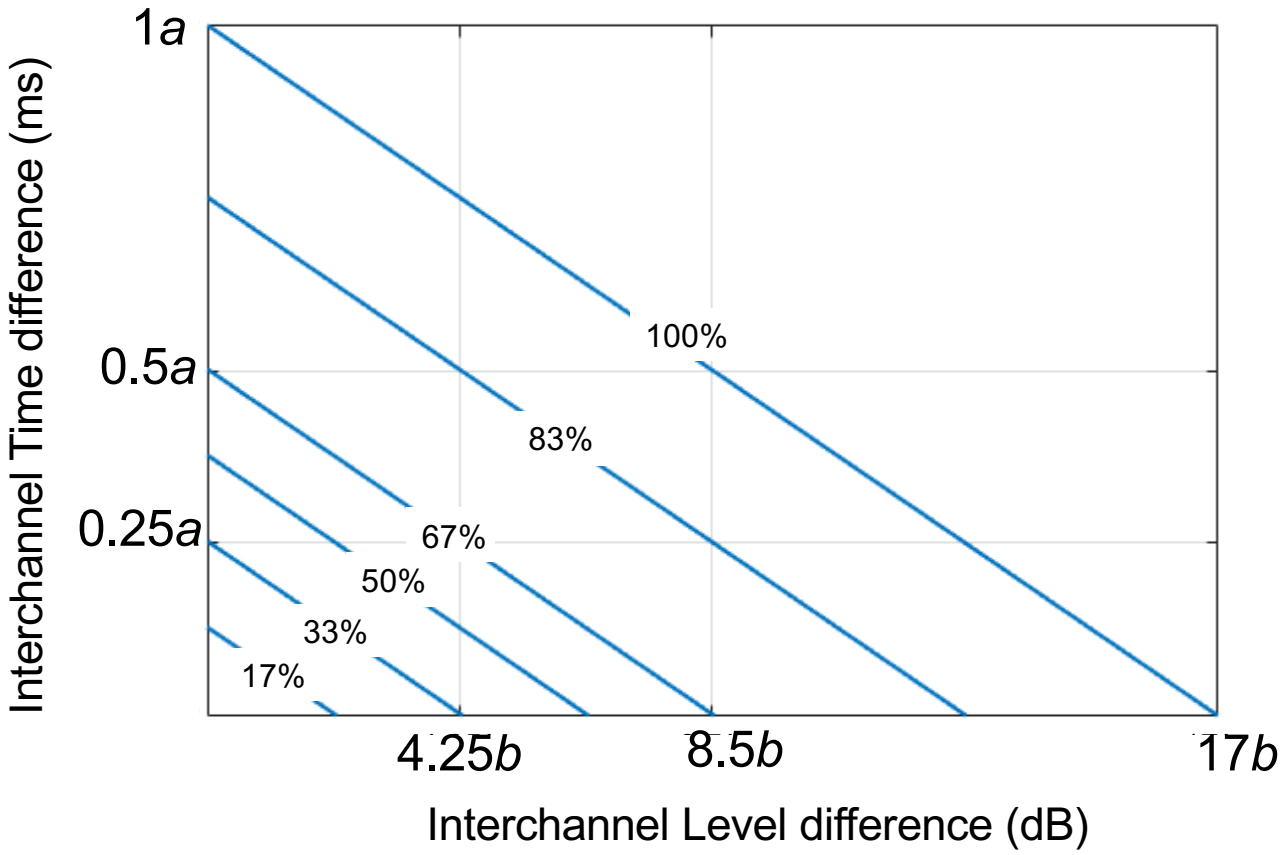
- PMAP: a new amplitude-panning law [AES142 2017]
  - Based on Lee and Rumsey [JAES 2013]
  - Accurate for the 60° base angle, but not for the 90°.



Around 5° localisation error

Psychoacoustic principles

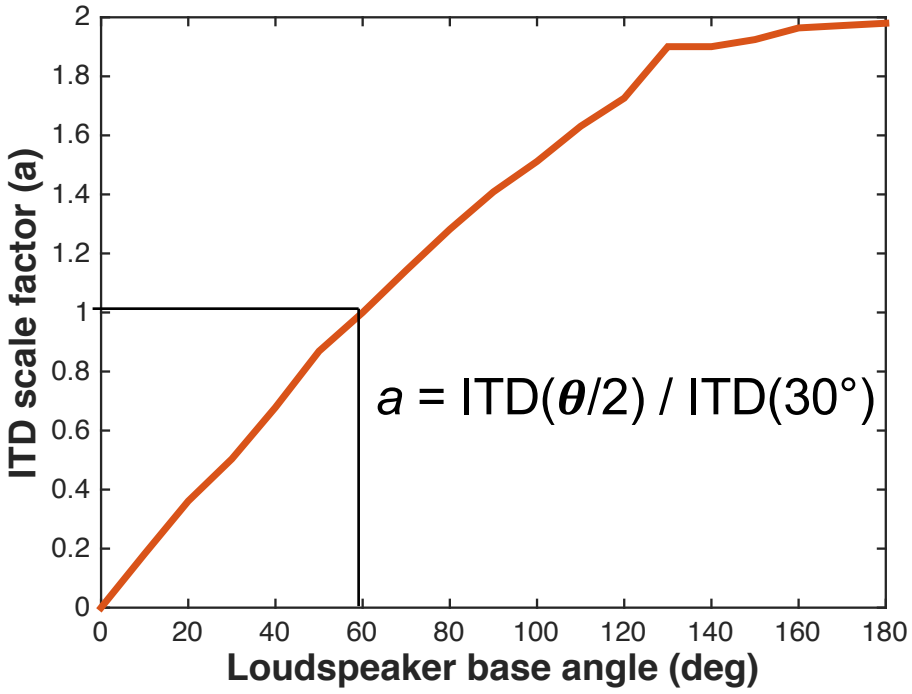
- Perceptual scaling based on ILD and ITD matching [Lee AES WIMP 2016]
  - Trade-off for an arbitrary base angle ($\boldsymbol{\theta}$).

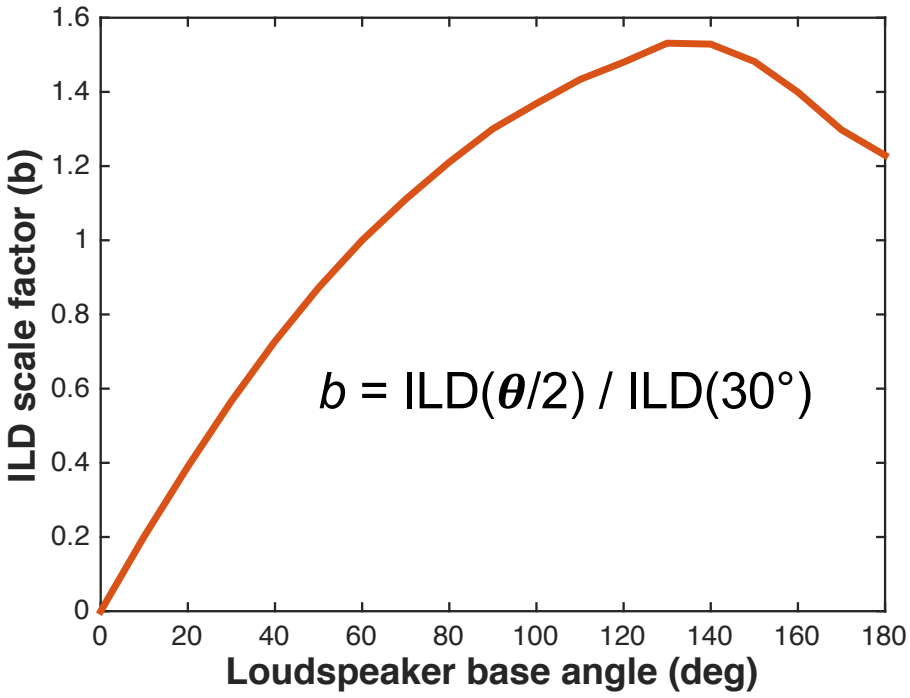


Psychoacoustic principles

University of HUDDERSFIELD

- Perceptual scaling based on ILD and ITD matching.
  - Scale factors a and b
  - Ratio of the ITD from real source to the ITD from phantom source at half the base angle.

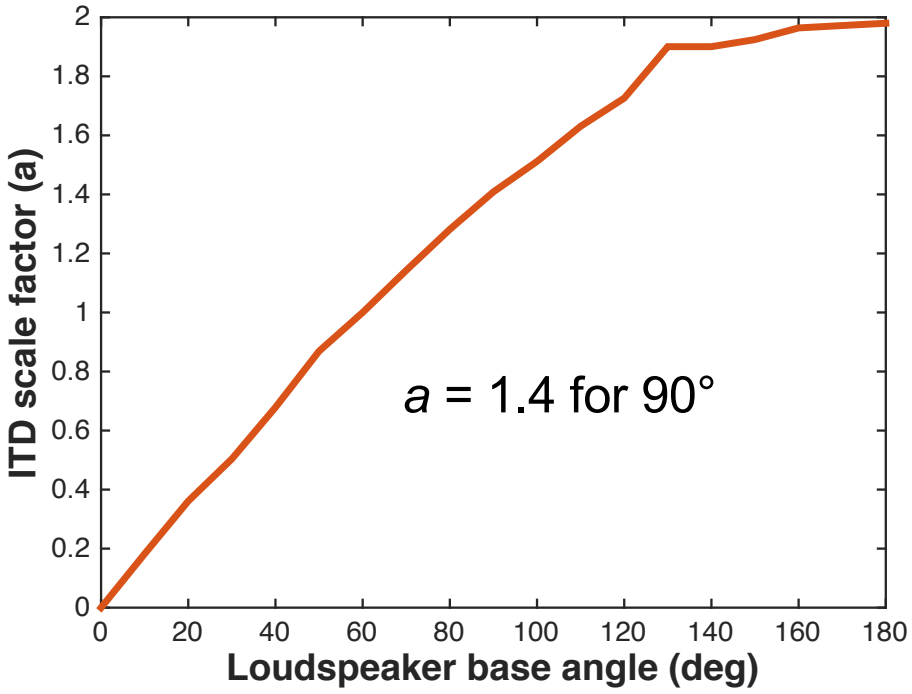


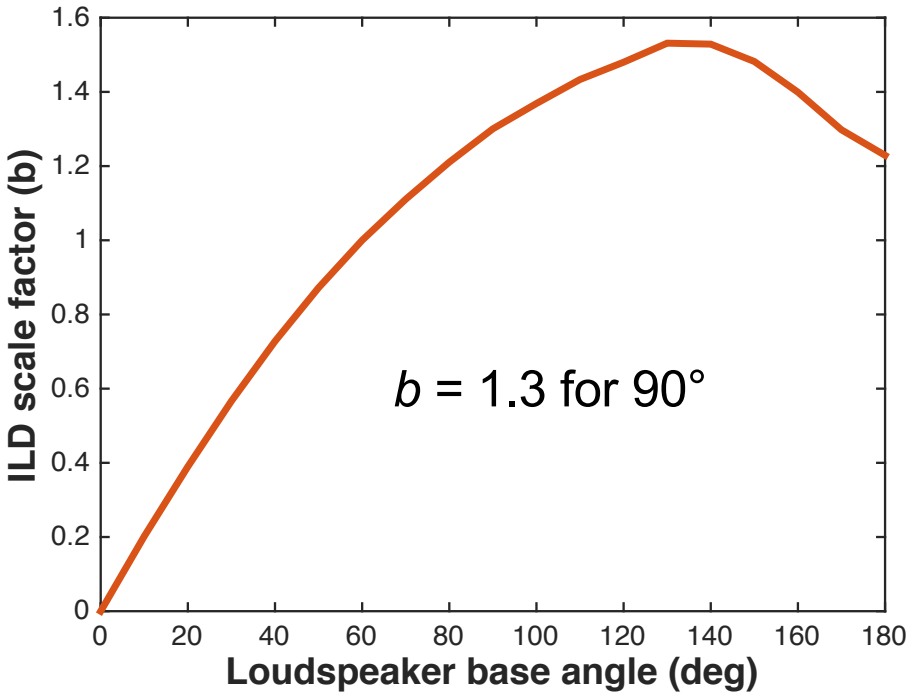


Psychoacoustic principles

University of HUDDERSFIELD

- Perceptual scaling based on ILD and ITD matching.
  - Scale factors a and b
  - Ratio of the ITD from real source to the ITD from phantom source at half the base angle.





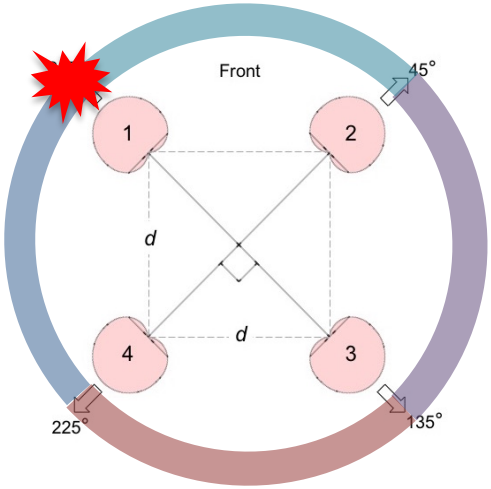
Psychoacoustic principles

- The appropriate spacing between microphones to produce the 90° SRA for the ESMA.
  - Depends on the model.

Model	Microphone spacing
Williams	23.8cm
Sengpiel	25cm
Wittek + Theile	24cm
Lee + Theile	30cm
Lee	50cm

Based on the 60° setup

Optimised for the 90° setup

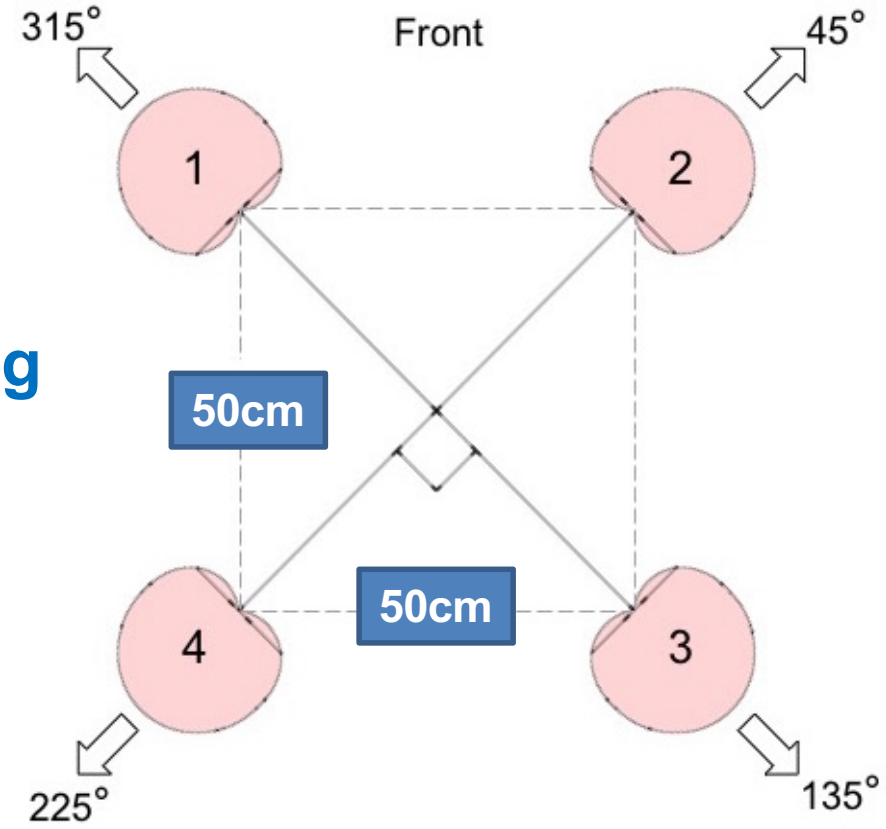


Proposed mic array

University of HUDDERSFIELD

- ESMA optimised for quadraphonic reproduction

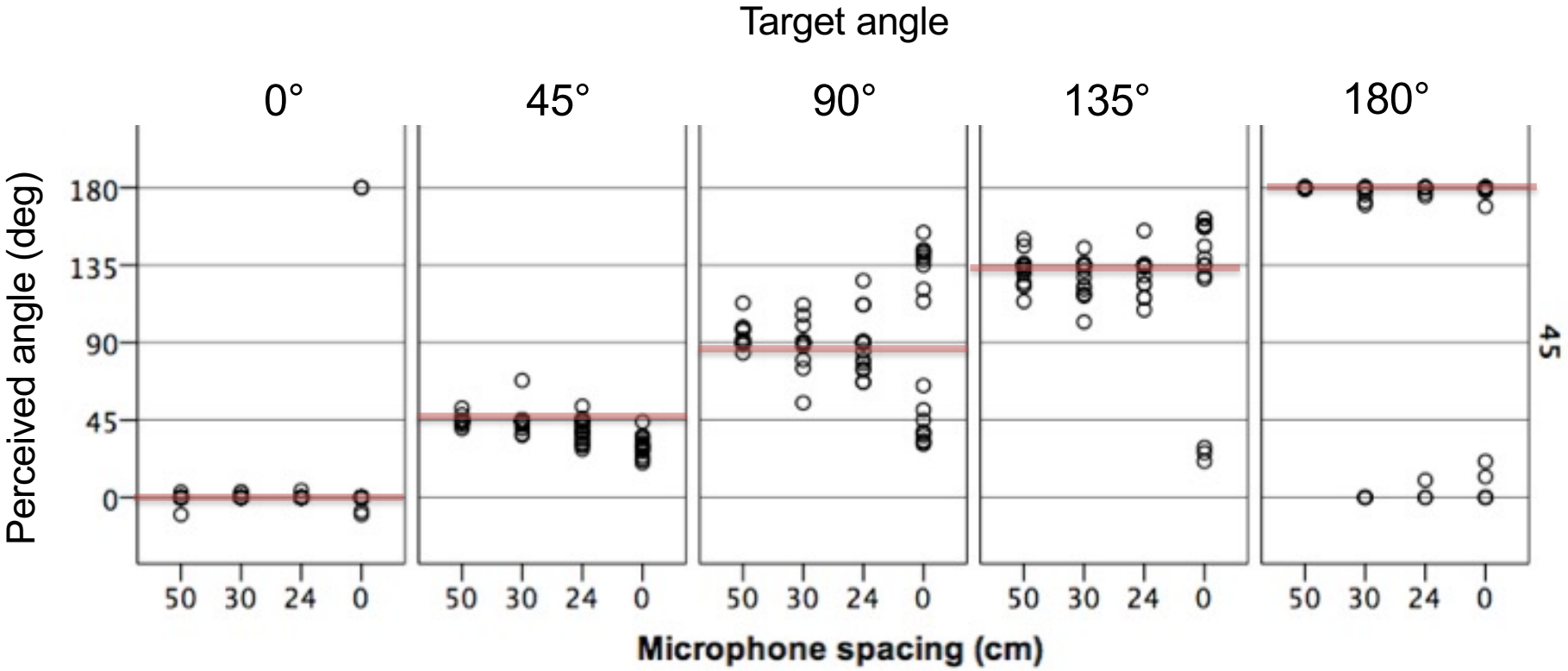
Optimal spacing = 50cm



Proposed mic array

University of HUDDERSFIELD

- Localisation accuracy test [Lee AES 2019]
  - Dry speech
  - ESMA 50cm, 30cm, 25cm & 0cm (FOA In-Phase)



H. Lee, "Capturing 360◦ Audio Using an Equal Segment Microphone Array (ESMA)," J. Audio Eng. Soc., vol. 67, no. 1/2, pp. 13–26, (2019 January/February.). DOI: https://doi.org/10.17743/jaes.2018.0068

Now let’s have a listen!

Perceptual Evaluations

University of HUDDERSFIELD

- Demo: *Recording of 3D 9.1 playback in a listening room*
  - Recording originally maded in Queens Elizabeth Hall, London.
  - 9.1 channel playback in the APL listening room.

Perceptual Evaluations

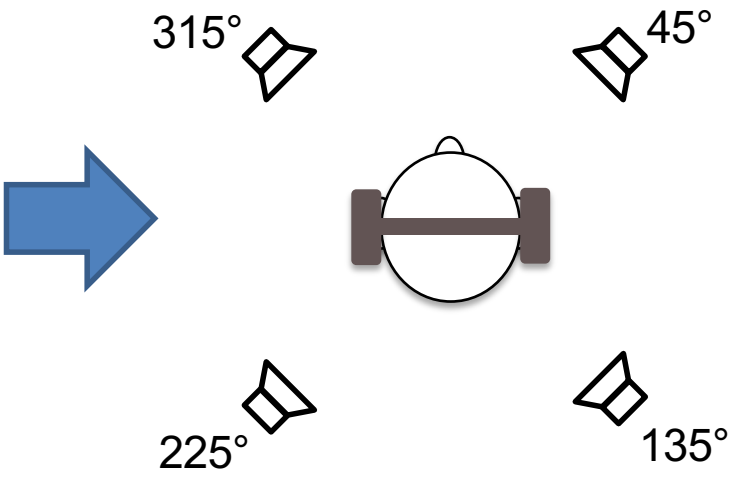
University of HUDDERSFIELD

- Demo: *Recording of 3D 9.1 playback in a listening room*
  - Reproduced sound was captured at the listening position, using
  - ESMA50cm (Nuemann KM184s), FOA (Soundfield SPS422b), Neumann KU100

9.1 Playback captured using mic arrays

Binauralised for Quad Reproduction

HRIR from www.sadie-project.co.uk

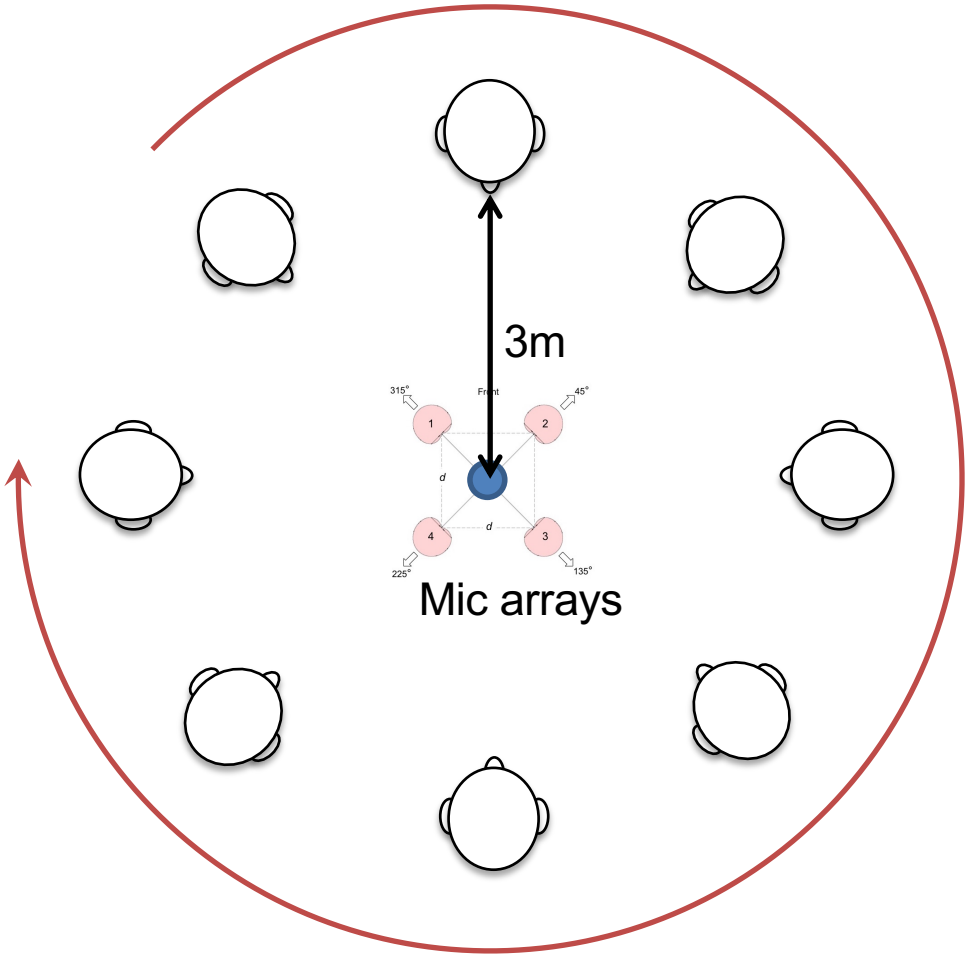


Perceptual Evaluations

University of HUDDERSFIELD

- Demo: *Early music choir*
  - St.Paul's concert hall, Huddersfield (RT = 2.1s).
  - ESMA 50cm (Neumann kk184 cardioid)
  - FOA (Sennheiser Ambeo); *MaxRe* decoder from www.sadie-project.co.uk

- Demo: *Early music choir*

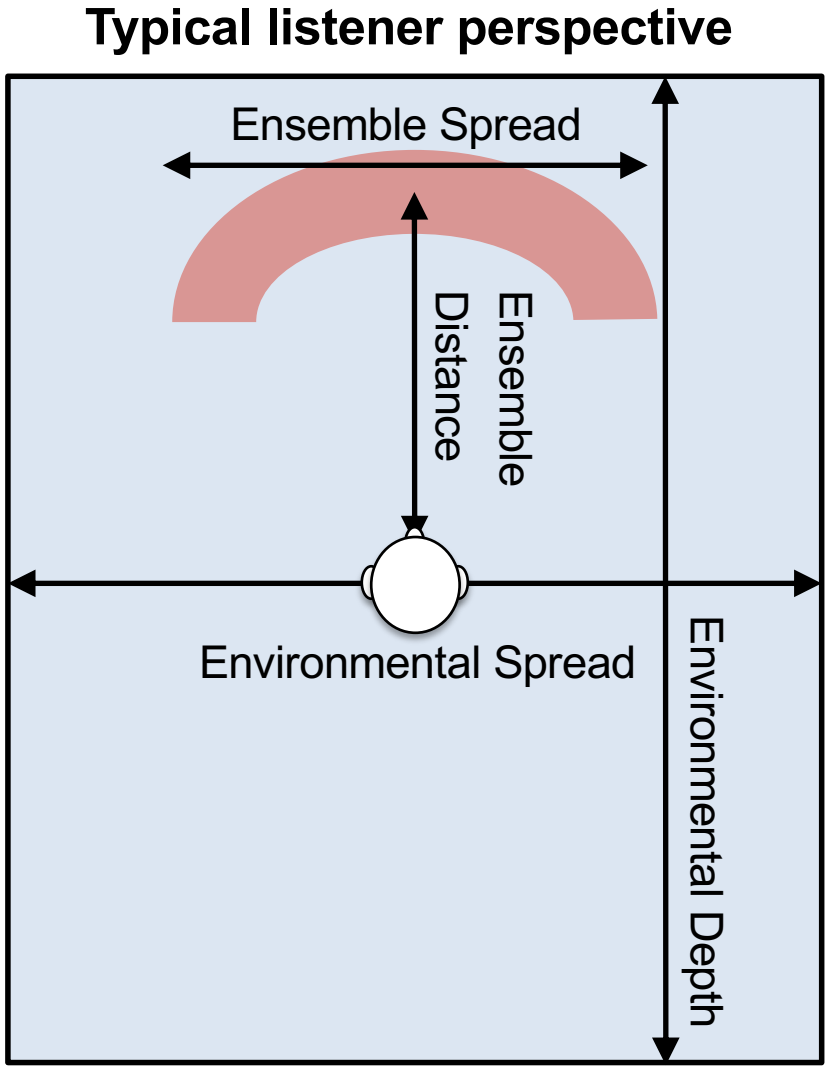


Perceptual Evaluations

University of HUDDERSFIELD

- Spatial attribute evaluations [Millns and Lee 2017]
  - Ensemble Spread
  - Ensemble Distance
  - Environmental Spread
  - Environmental Depth

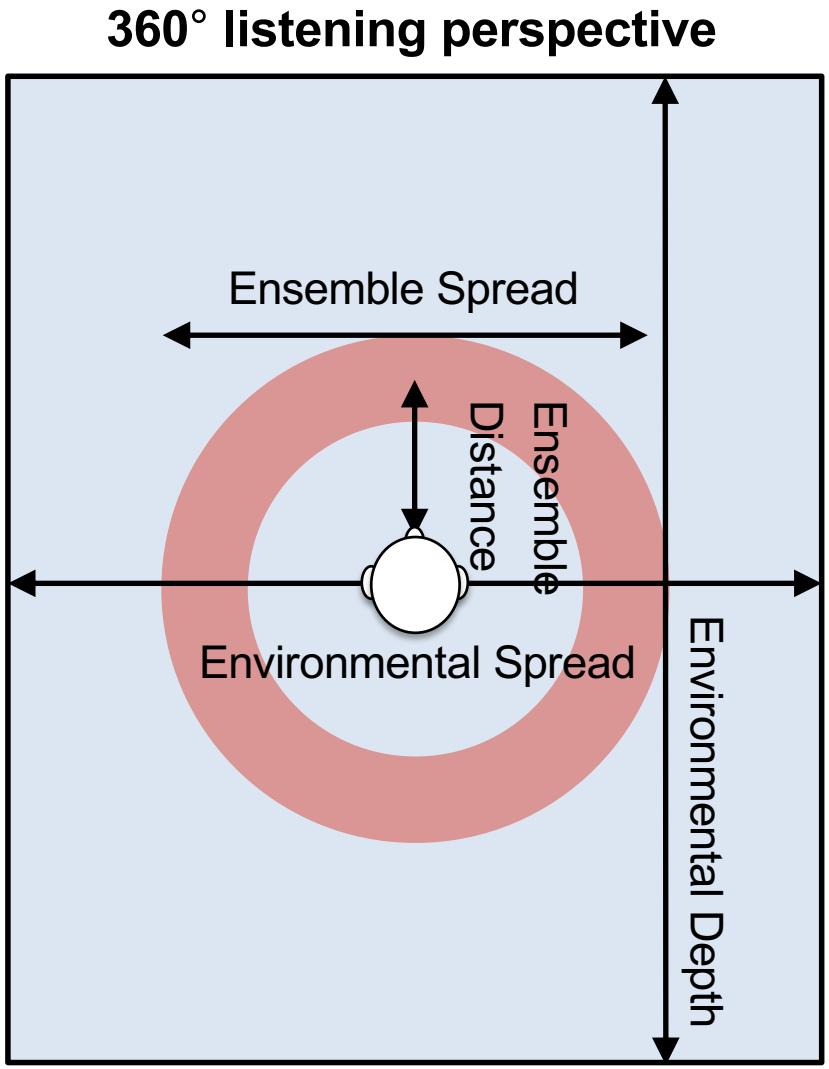
- Scene Based Paradigm
  - Rumsey [2001]
  - Source & Environment-related attributes for surround sound reproduction



Perceptual Evaluations

- Spatial attribute evaluations [Millns and Lee 2017]
  - Ensemble Spread
  - Ensemble Distance
  - Environmental Spread
  - Environmental Depth

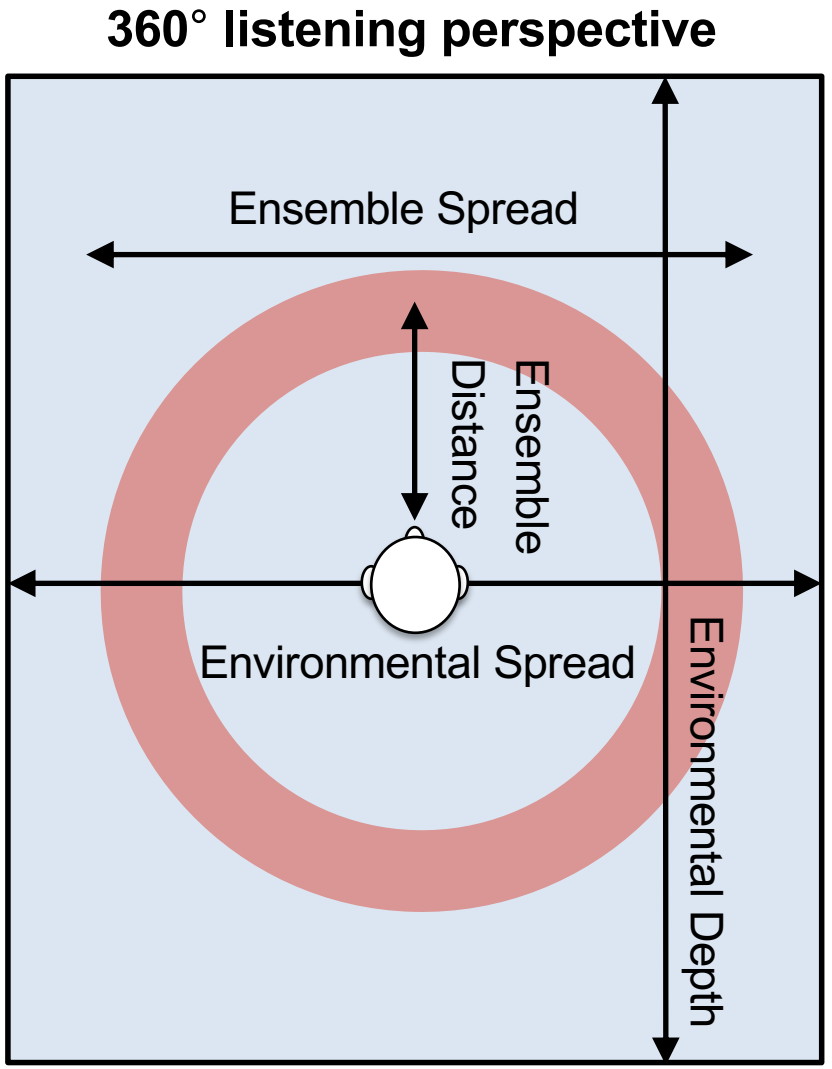
- Scene Based Paradigm
  - Rumsey [2001]
  - Source & Environment-related attributes for surround sound reproduction



Perceptual Evaluations

- Spatial attribute evaluations [Millns and Lee 2017]
  - Ensemble Spread
  - Ensemble Distance
  - Environmental Spread
  - Environmental Depth

- Scene Based Paradigm
  - Rumsey [2001]
  - Source & Environment-related attributes for surround sound reproduction



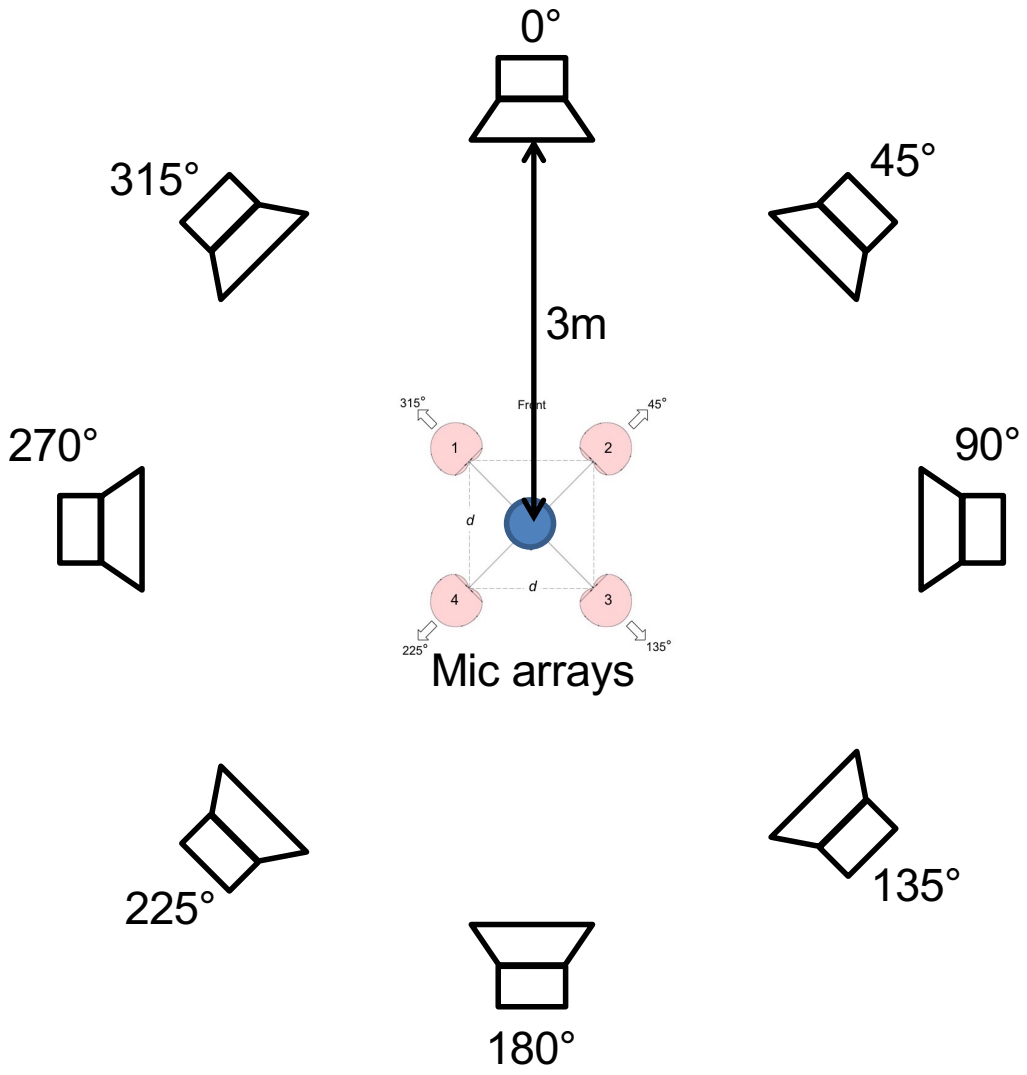
Perceptual Evaluations

University of HUDDERSFIELD

- Recording simulation using mic array impulse responses

Perceptual Evaluations

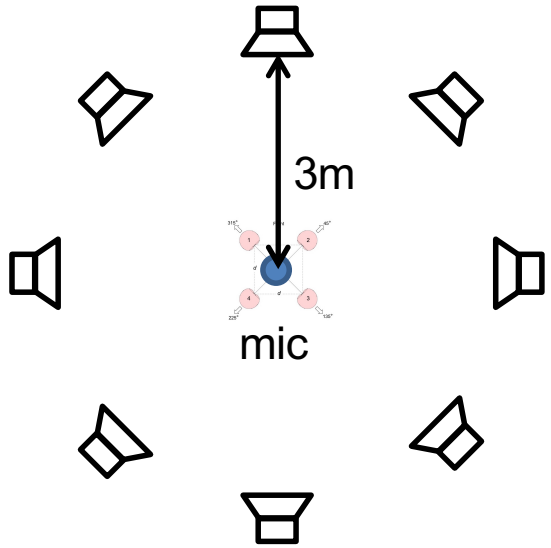
- Room impulse response capture using mic arrays
  - ESMA 50cm, ESMA 25cm (Neumann kk184)
  - FOA (Sennhesier Ambeo)
  - Dummy head (Neumann KU100)



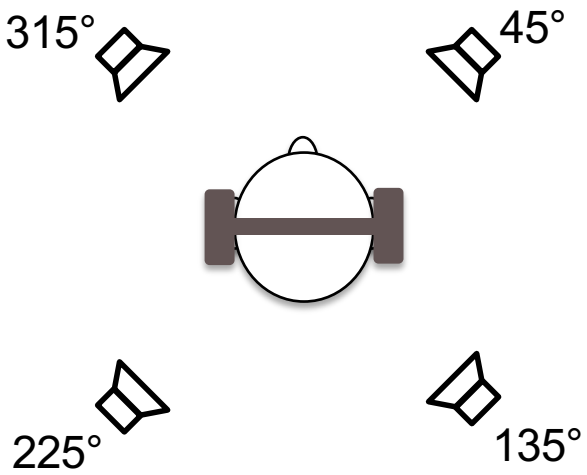
Perceptual Evaluations

- Binaural synthesis
  - Mic array RIRs were convolved with dry mono and multichannel recordings.
  - Then binauralised for the quadraphonic setup using the SADIE HRIR database www.sadie-project.co.uk.

4ch Mic Array RIRs * Dry Sources



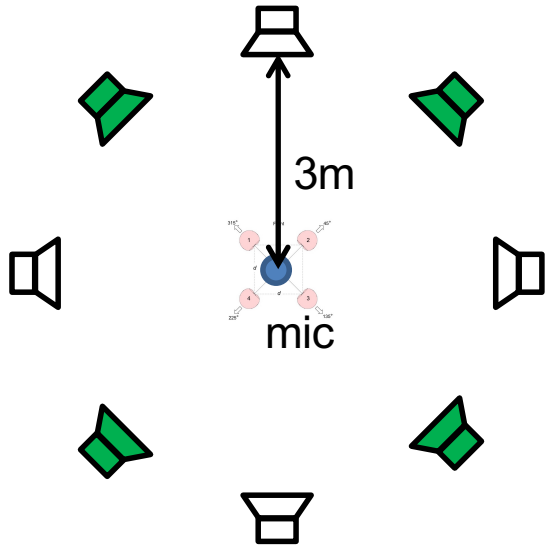
Binauralised for Quad Reproduction



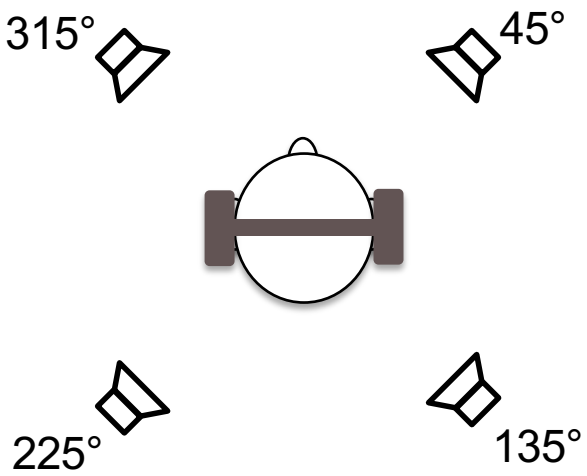
Perceptual Evaluations

- Demo: *Rounders 1* (4 singers at ±45° and ±135°)
  - KU100
  - ESMA 50cm, ESMA 25cm
  - FOA (MaxRe), FOA (Mode-Matching)

4ch Mic Array RIRs * Dry Sources



Binauralised for Quad Reproduction

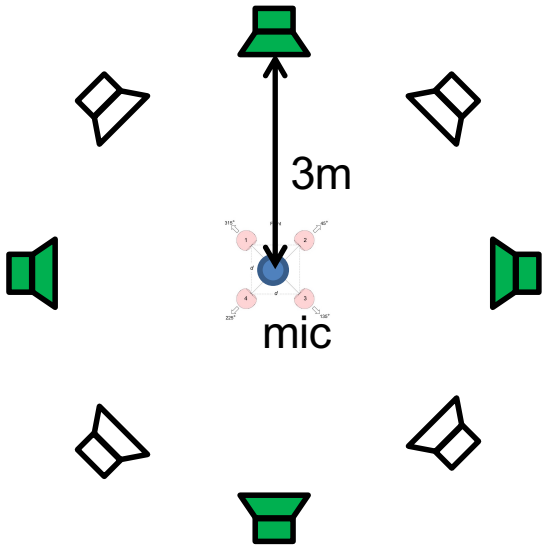


Perceptual Evaluations

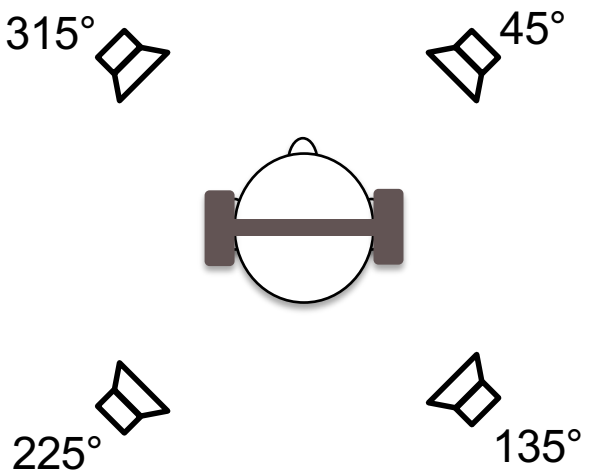
University of HUDDERSFIELD

- Demo: *Rounders 2* (4 singers at 0°, ±90° & 180°)
  - Dummy head
  - ESMA 50cm, ESMA 25cm
  - FOA (MaxRe), FOA (Mode-Matching)

4ch Mic Array RIRs * Dry Sources



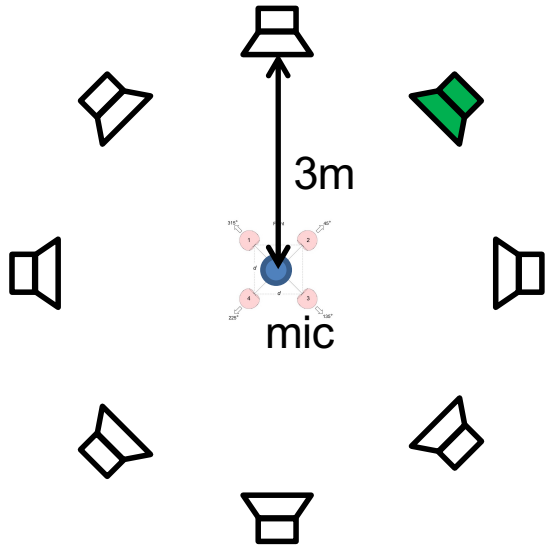
Binauralised for Quad Reproduction



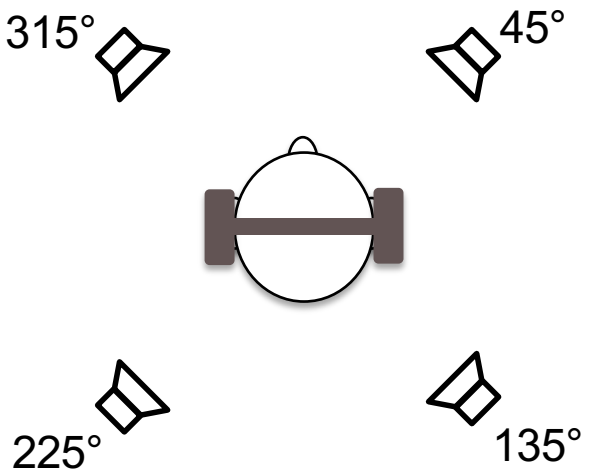
Perceptual Evaluations

- Demo: *Rounders 3* (Female speech at -45°)
  - Dummy head
  - ESMA 50cm, ESMA 25cm
  - FOA (MaxRe), FOA (Mode-Matching)

4ch Mic Array RIRs * Dry Sources



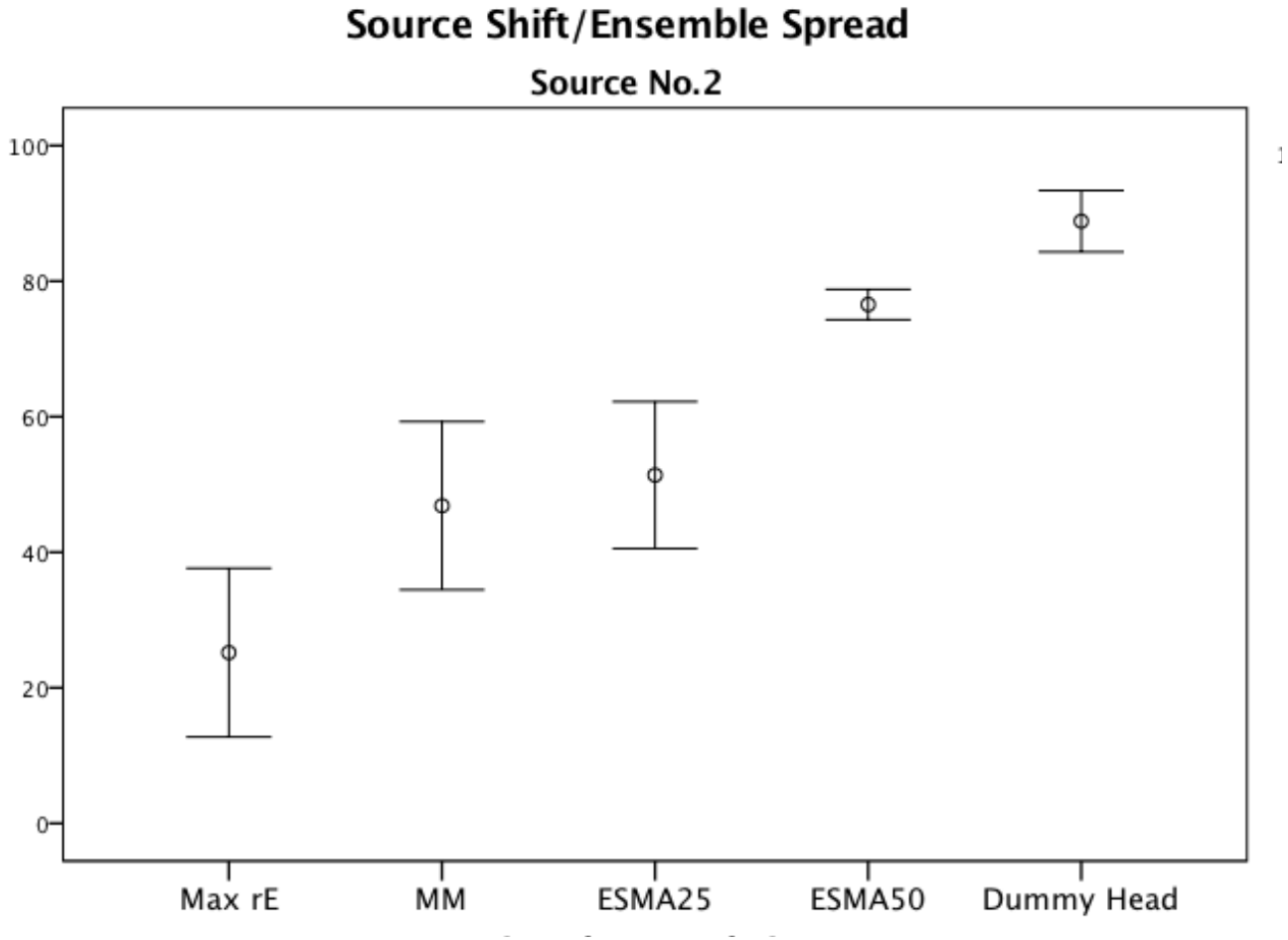
Binauralised for Quad Reproduction



Perceptual Evaluations

University of HUDDERSFIELD

- Results – Ensemble Spread

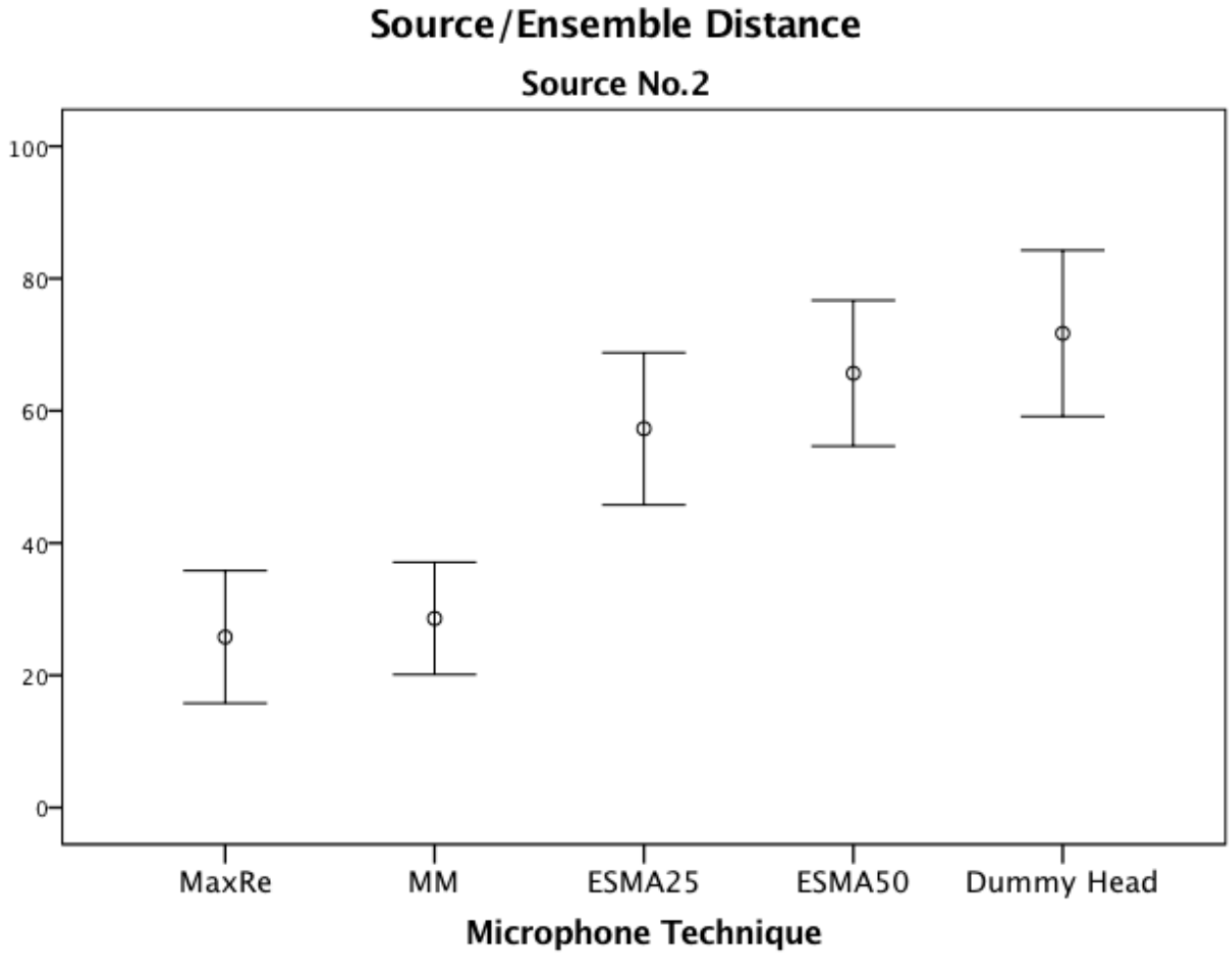


Millns, C. and Lee, H (2018) 'An Investigation into Spatial Attributes of 360° Microphone Techniques for Virtual Reality'. *In: AES the 144th International Convention, 23 – 26 May 2018*, Milan, Italy.

Perceptual Evaluations

University of HUDDERSFIELD

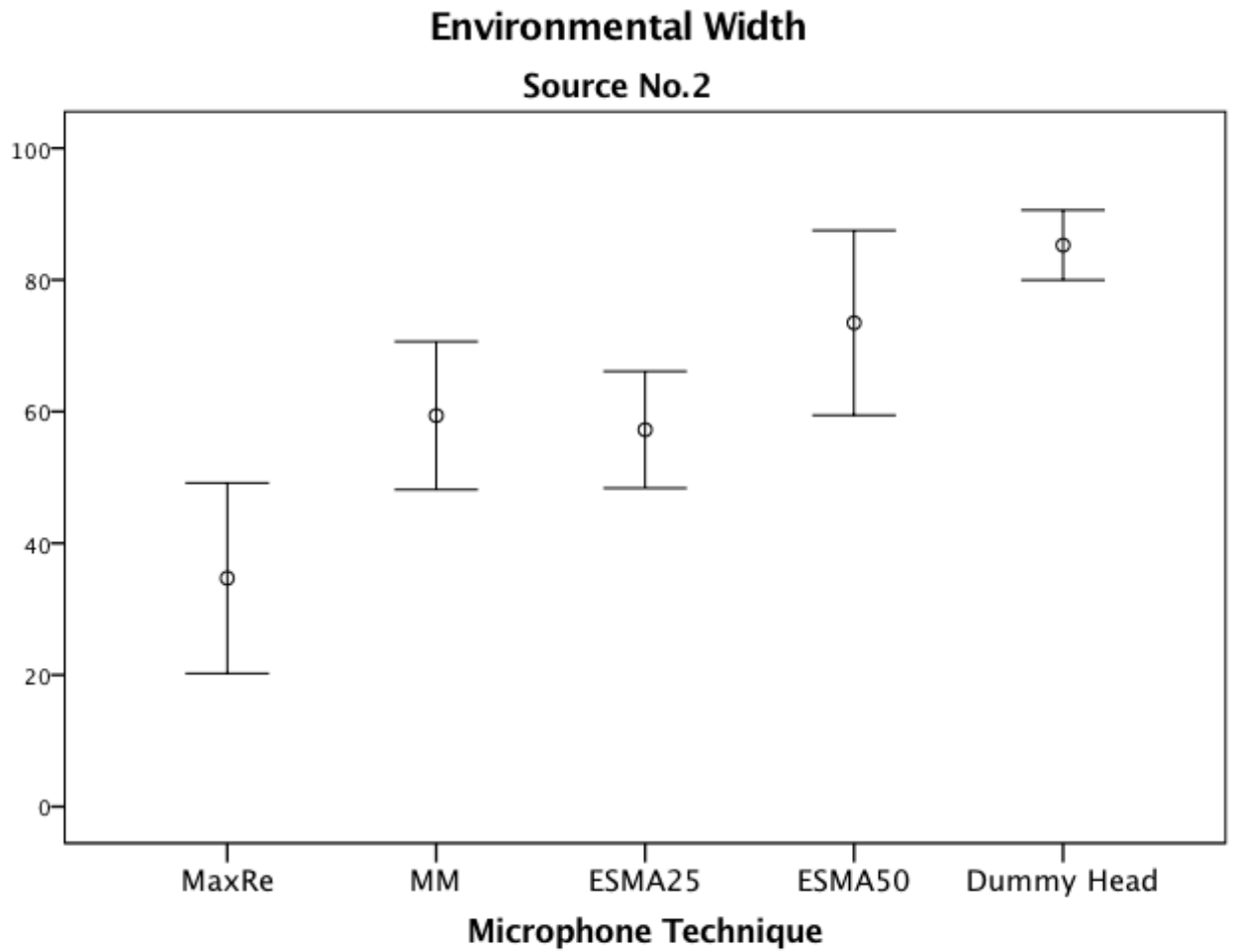
- Results – Ensemble Distance



Perceptual Evaluations

University of HUDDERSFIELD

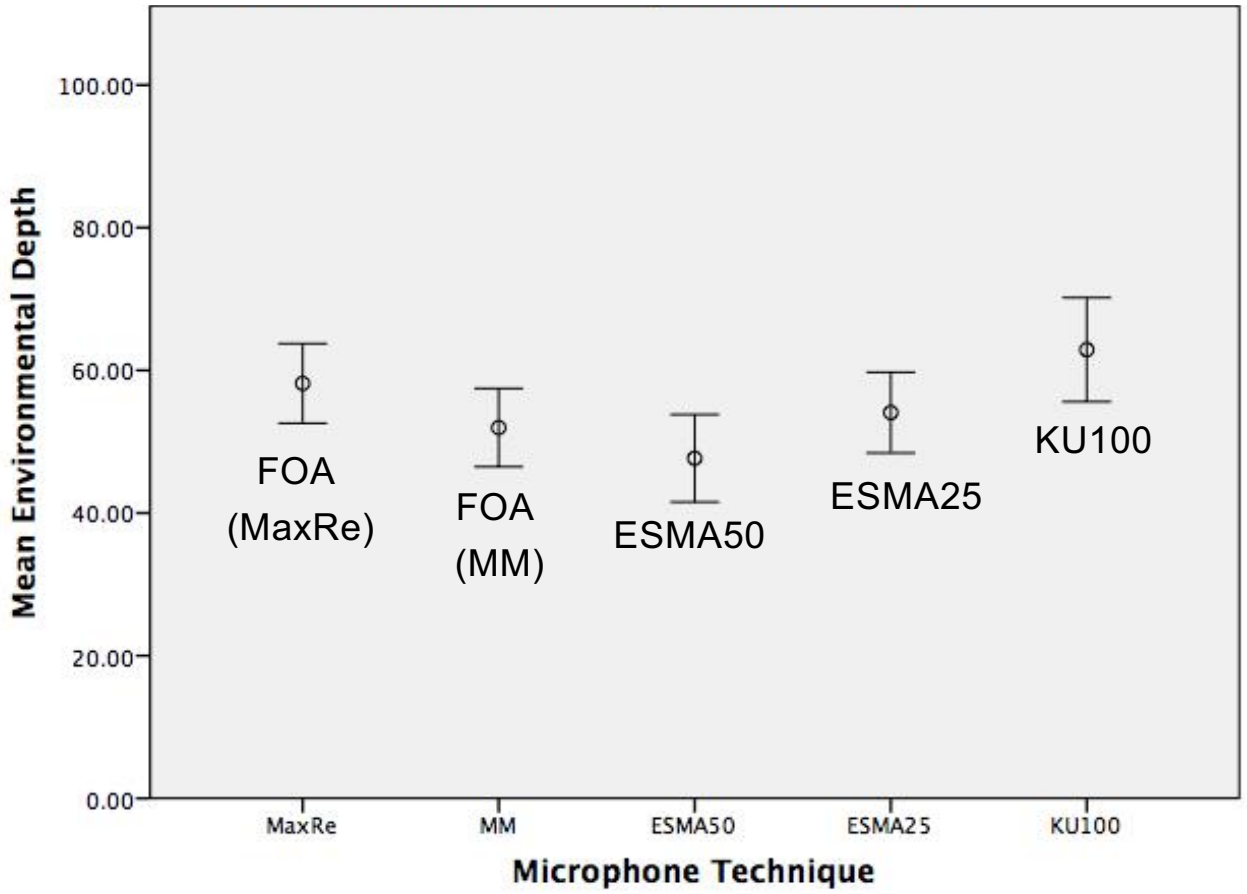
- Results – Environmental Width



Perceptual Evaluations

University of HUDDERSFIELD

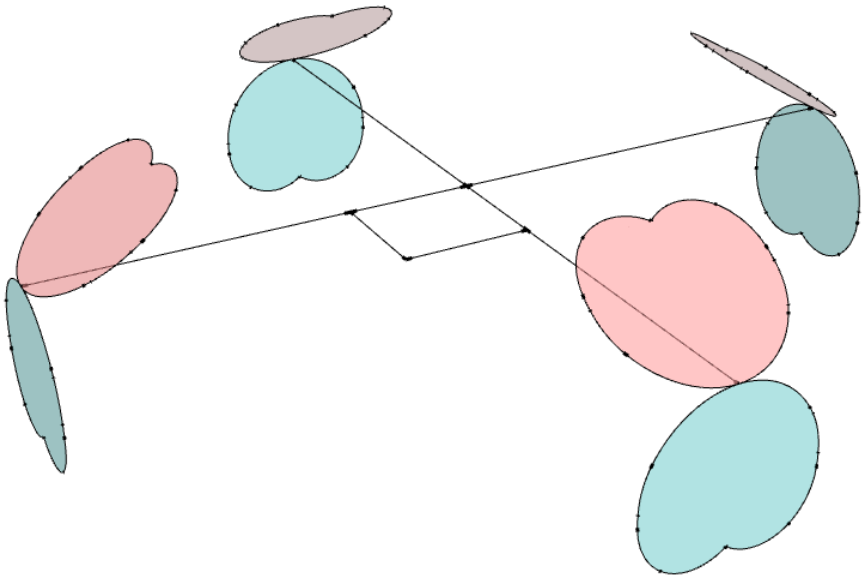
- Results – Environmental Depth



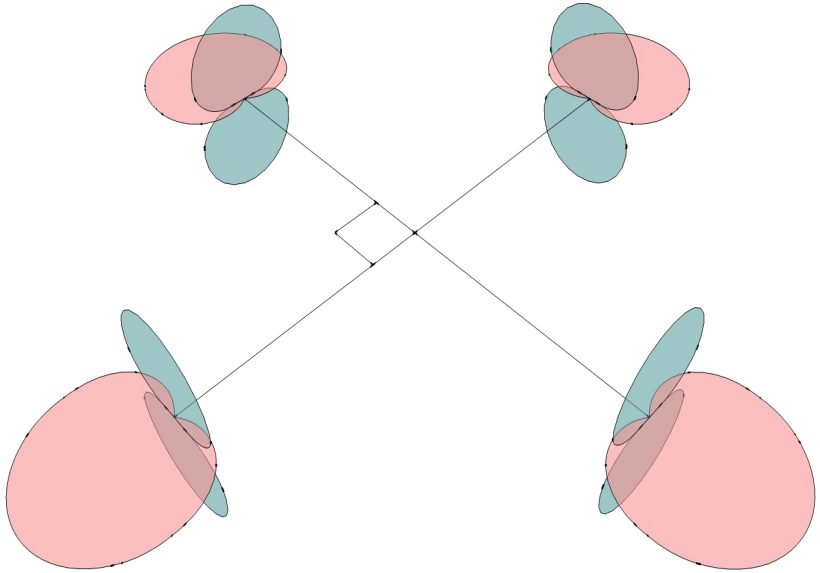
ESMA-3D (with height)

University of HUDDERSFIELD

- Augmented with upward-facing supercardioid or figure-of-8 microphones.



1 to 1 routing to speakers



Vertical Mid-Side decoding

ESMA-3D

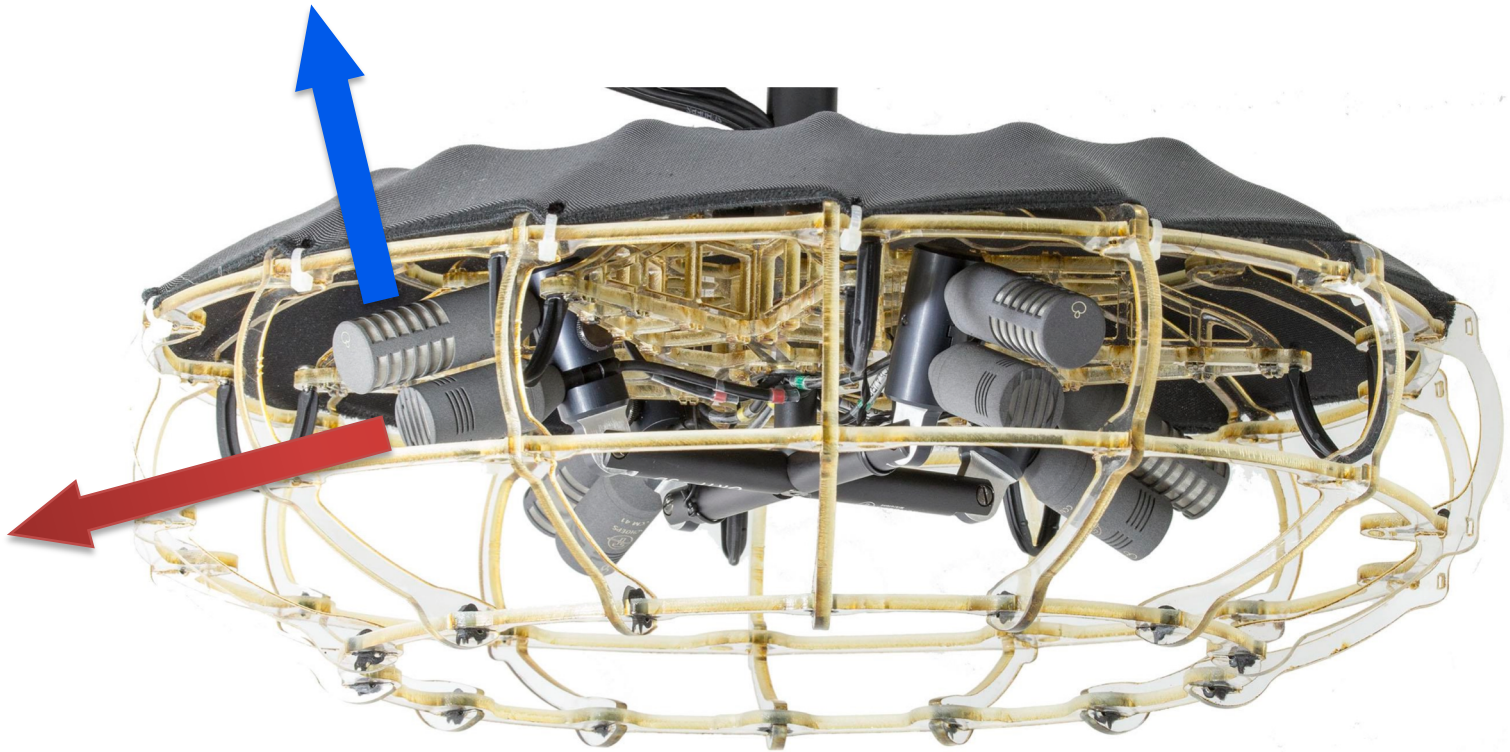
University of HUDDERSFIELD

- Vertical time panning is highly unstable [Wallis and Lee JAES 2015].
- Vertical microphone spacing has little effect on LEV [Lee and Gribben JAES 2014].
- Vertical level panning can still work with a limited resolution [Barbour 2003, Mironovs and Lee 2016].

ORTF-3D

University of HUDDERSFIELD

- Schoeps ORTF 3D
  - Share the same design concept.
  - Based on Lee and Gribben [JAES 2014]



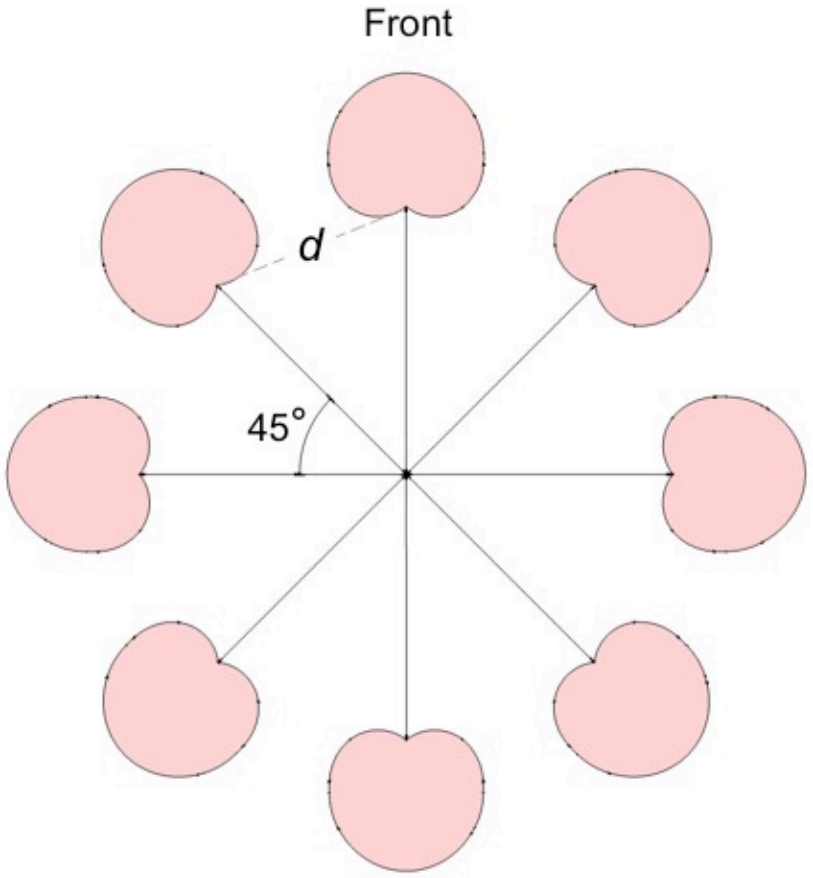
ESMA-3D

University of HUDDERSFIELD

- 4 supercardioid height microphones facing upwards.
- The height mic should have at least 7 to 10dB less direct sound than the main mic to avoid vertical image shift [Lee 2011; Wallis and Lee 2017]

Higher Order ESMA

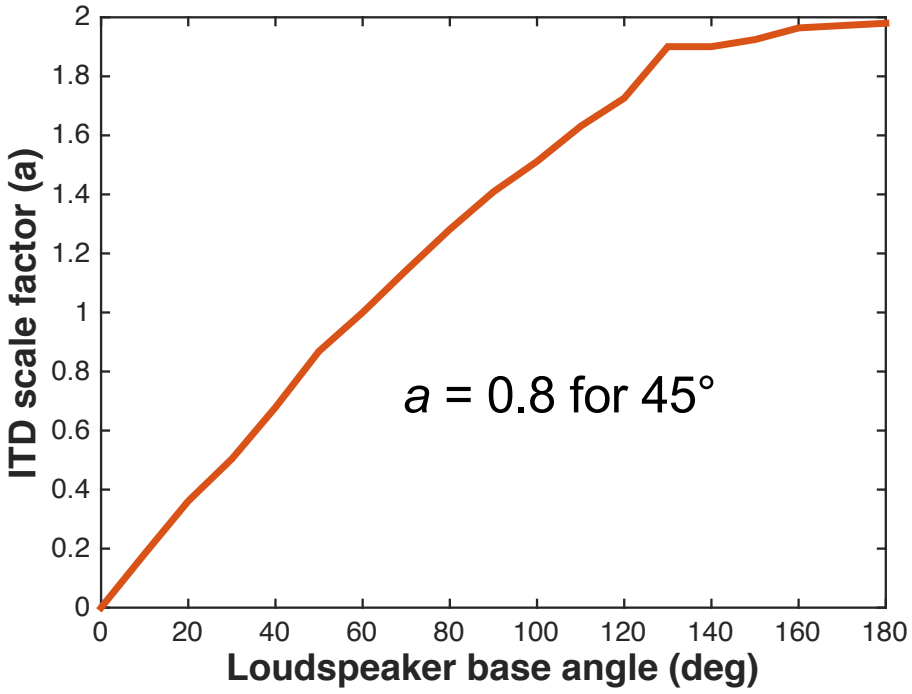
- Higher Order ESMA
  - For an octagonal setup, each segment should have the SRA of 45°.
  - Can potentially solve the problem of unstable side image localisation.
  - Mic spacing *d*
    - *Williams: 82cm*
    - *Lee: 55cm*

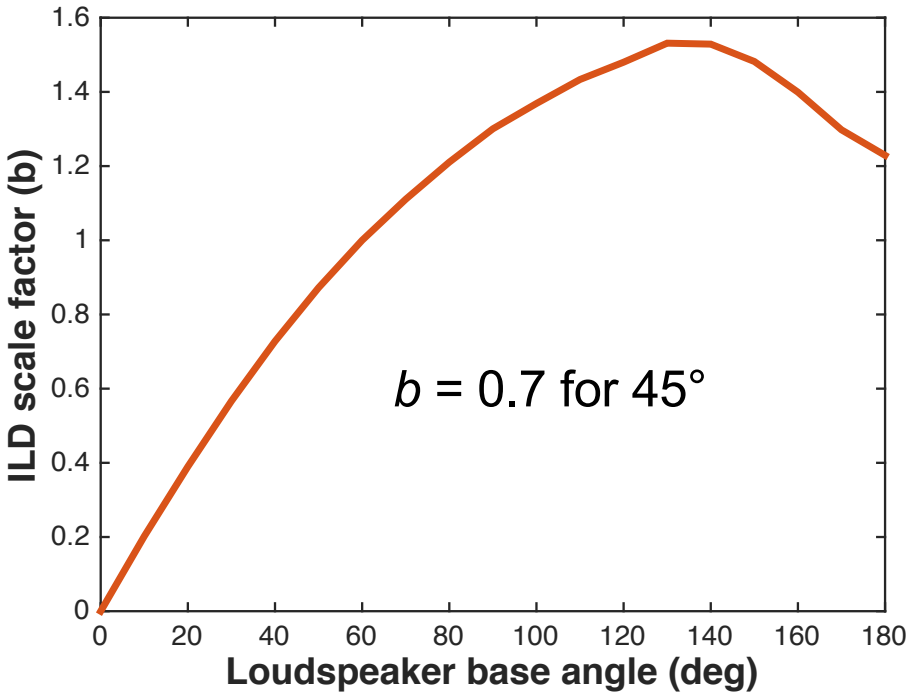


Higher Order ESMA

University of HUDDERSFIELD

- Perceptual scaling based on ILD and ITD matching.
  - Scale factors a and b
  - Ratio of the ITD from real source to the ITD from phantom source at half the base angle.





Thanks for listening.

Questions?