University of HUDDERSFIELD

Perceptual Signal Processing for 3D Sound Recording

Dr Hyunkook Lee

h.lee@hud.ac.uk

Applied Psychoacoustics Lab (APL)

University of Huddersfield, UK

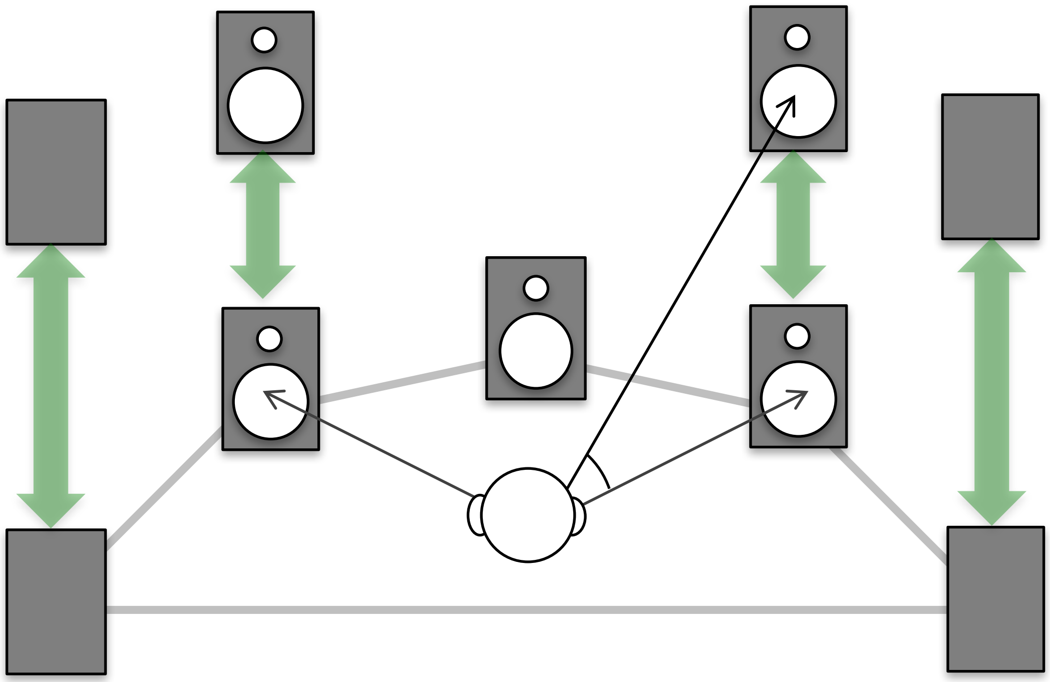
EPSRC

Engineering and Physical Sciences Research Council

THE QUEEN'S AWARDS FOR ENTERPRISE

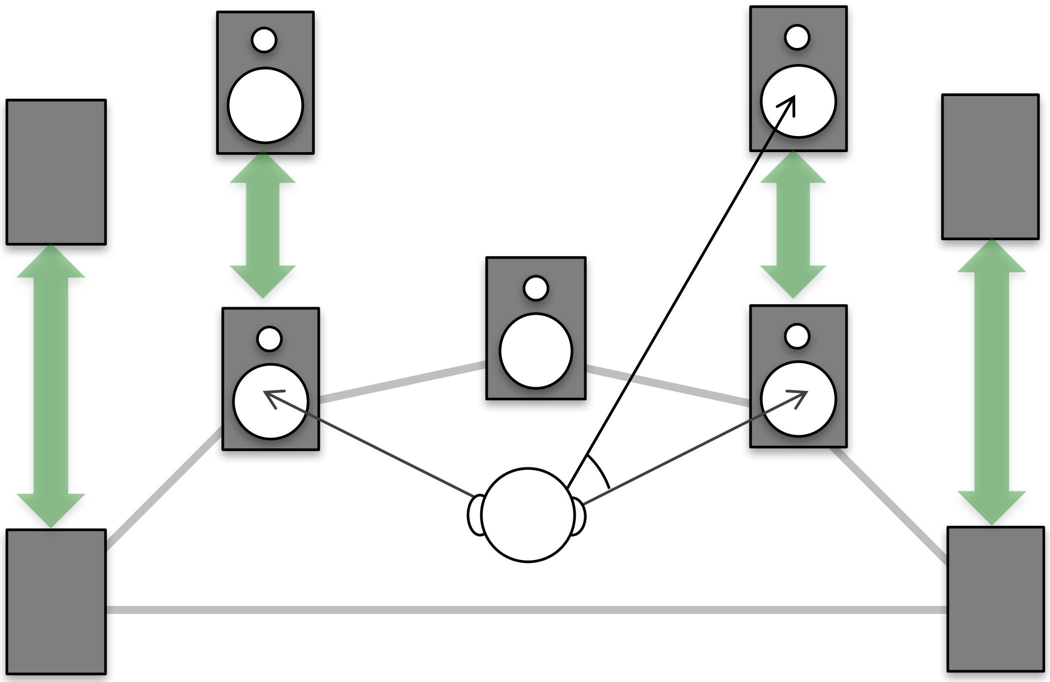
Introduction

- What's the optimal way to record for 3D?
- How can we enhance 3D recordings?
- How do we perceive sounds in vertical stereophony?



Introduction

- Purpose of this tutorial / demo
  - To discuss the psychoacoustics of vertical stereophonic perception.
  - To provide a link between psychoacoustic principles and practical techniques for capturing and enhancing 3D sound.



Introduction

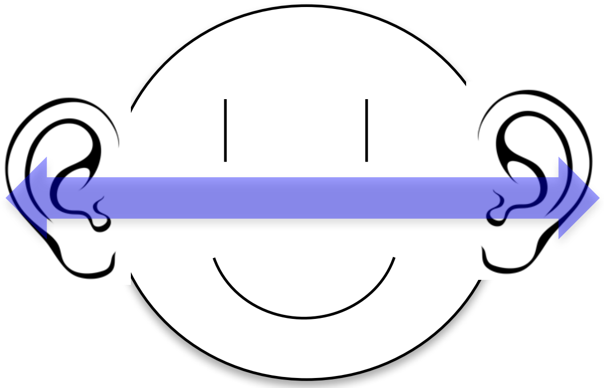
- Content
  - Vertical localisation & Phantom image elevation
  - Vertical interchannel crosstalk
  - Vertical image spread enhancement
  - 2D to 3D upmixing

Vertical localisation &

Phantom image elevation

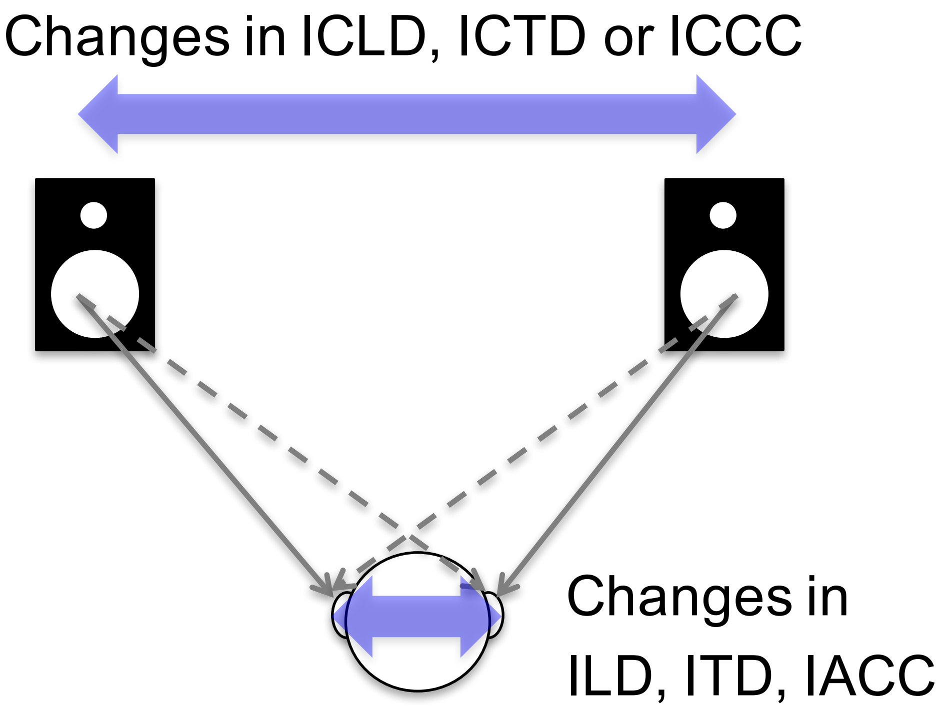
Horizontal vs. Vertical Stereo

- Vertical auditory perception is fundamentally different from horizontal perception.
  - Horizontal stereo: Interaural cues



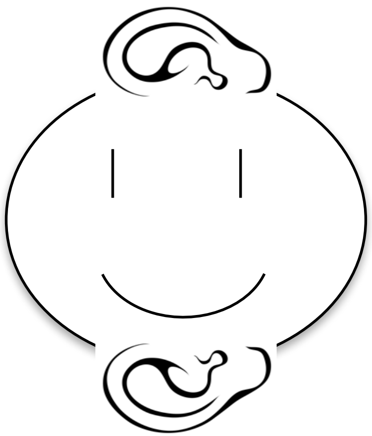
Horizontal vs. Vertical Stereo

- Horizontal spatial perception
  - Inter-Channel cues → Inter-Aural cues

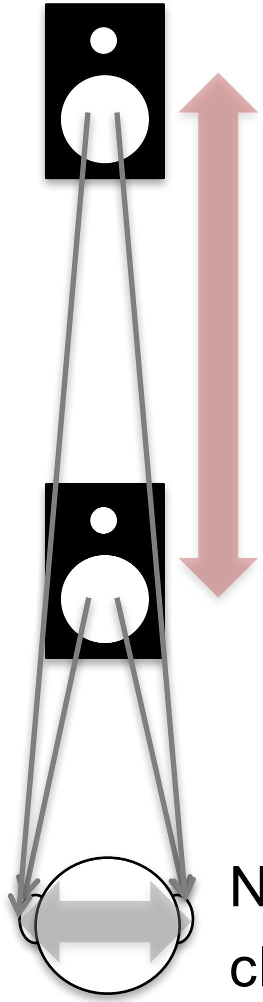


Horizontal vs. Vertical Stereo

- Vertical spatial perception



Vertical localisation relies on **spectral cues** and **torso reflections**.



Changes in ICLD, ICTD, or ICCC

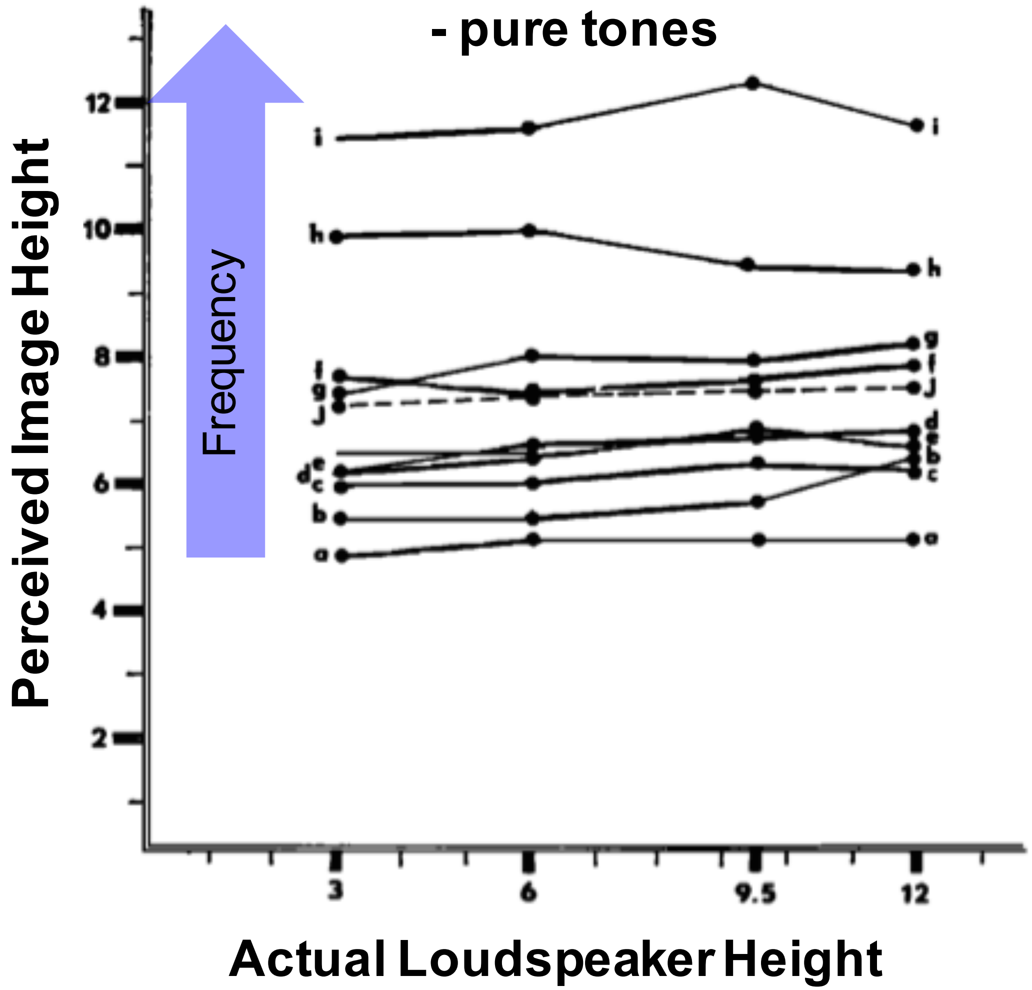
NO interaural changes

Pitch-Height Effect for “Real” Source

- The higher the frequency of a **pure tone** is, the higher the perceived image position is, **regardless of** the physical height of the loudspeaker. (Pratt 1930).

- Confirmed by Trimble (1934), Roffler and Butler (1968a), etc.

From Roffler and Butler (1968a) - pure tones

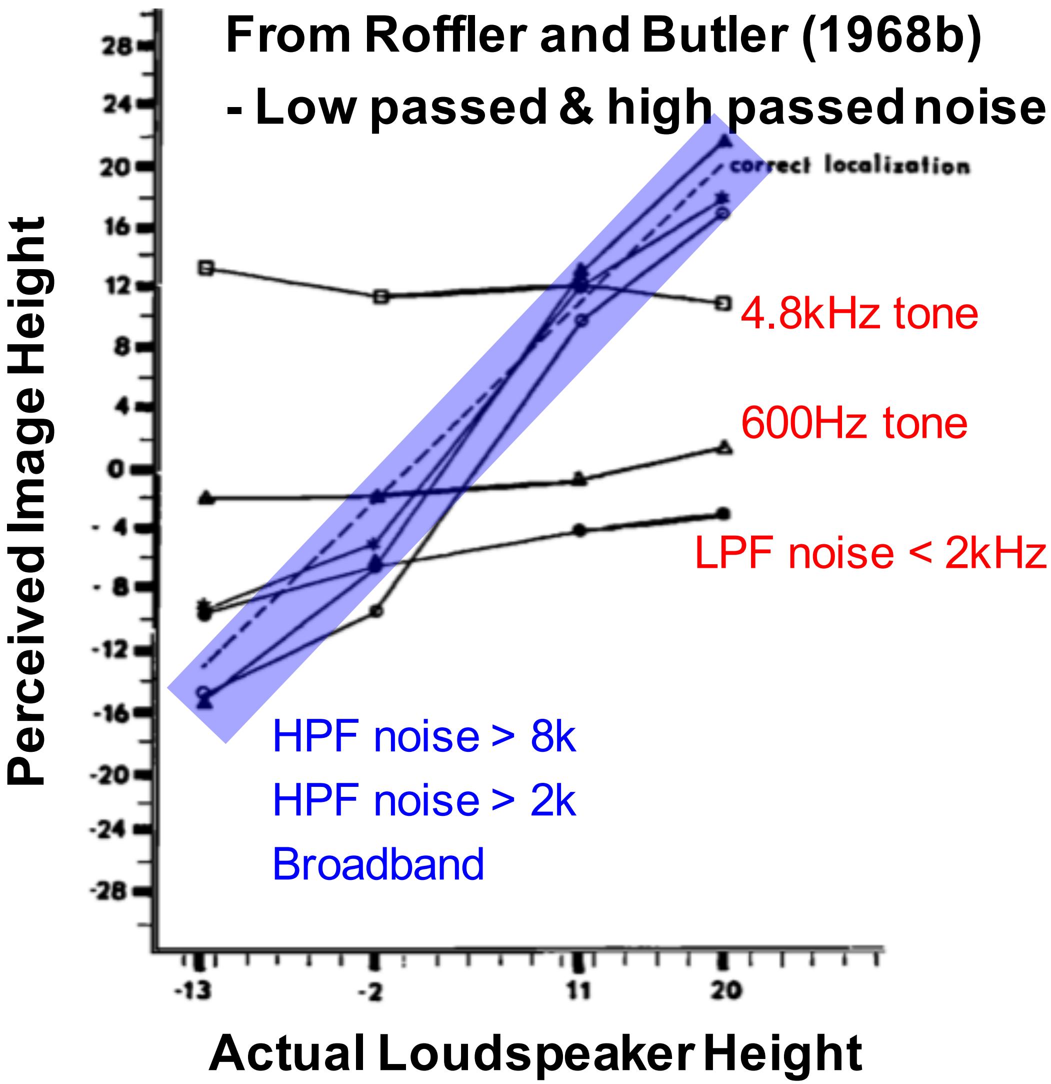


Pitch-Height Effect for “Real” Source

University of HUDDERSFIELD

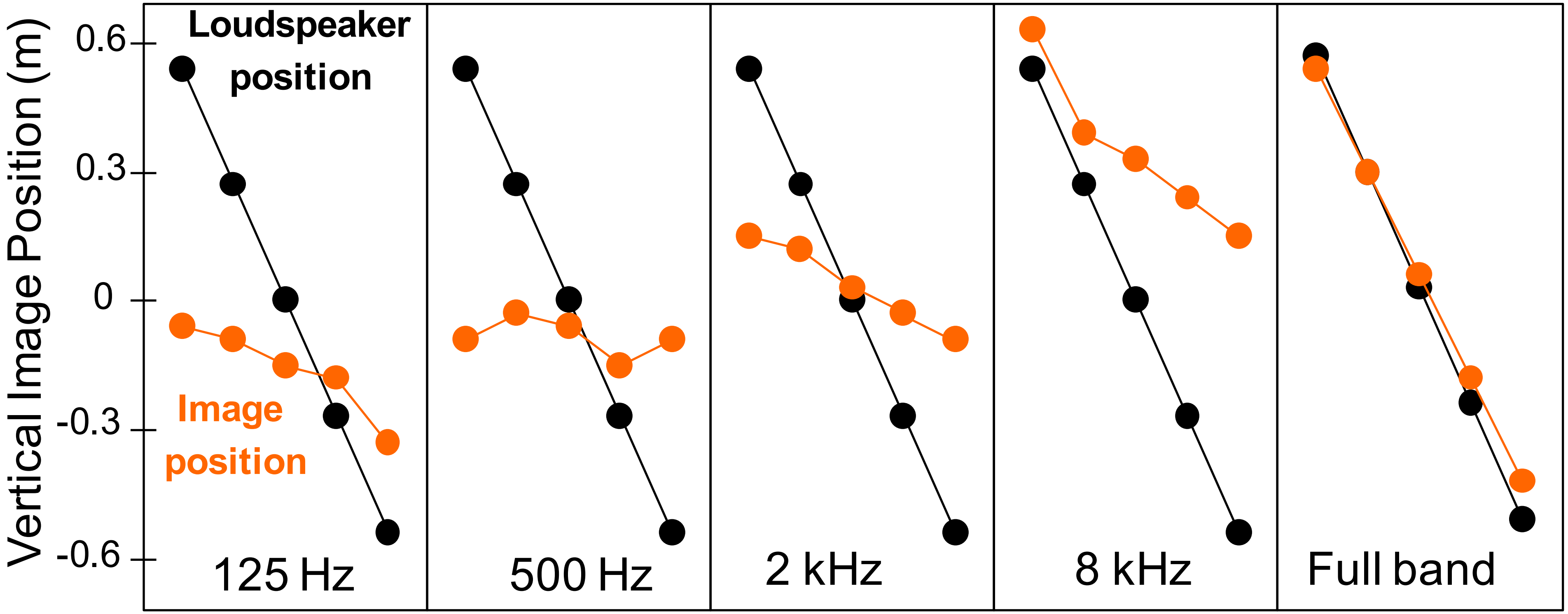
- For **band-passed noise** signals, high frequency components (above 7kHz) are essential for accurate vertical localisation.

(Roffler and Butler 1968b)



Pitch-Height Effect for “Real” Source

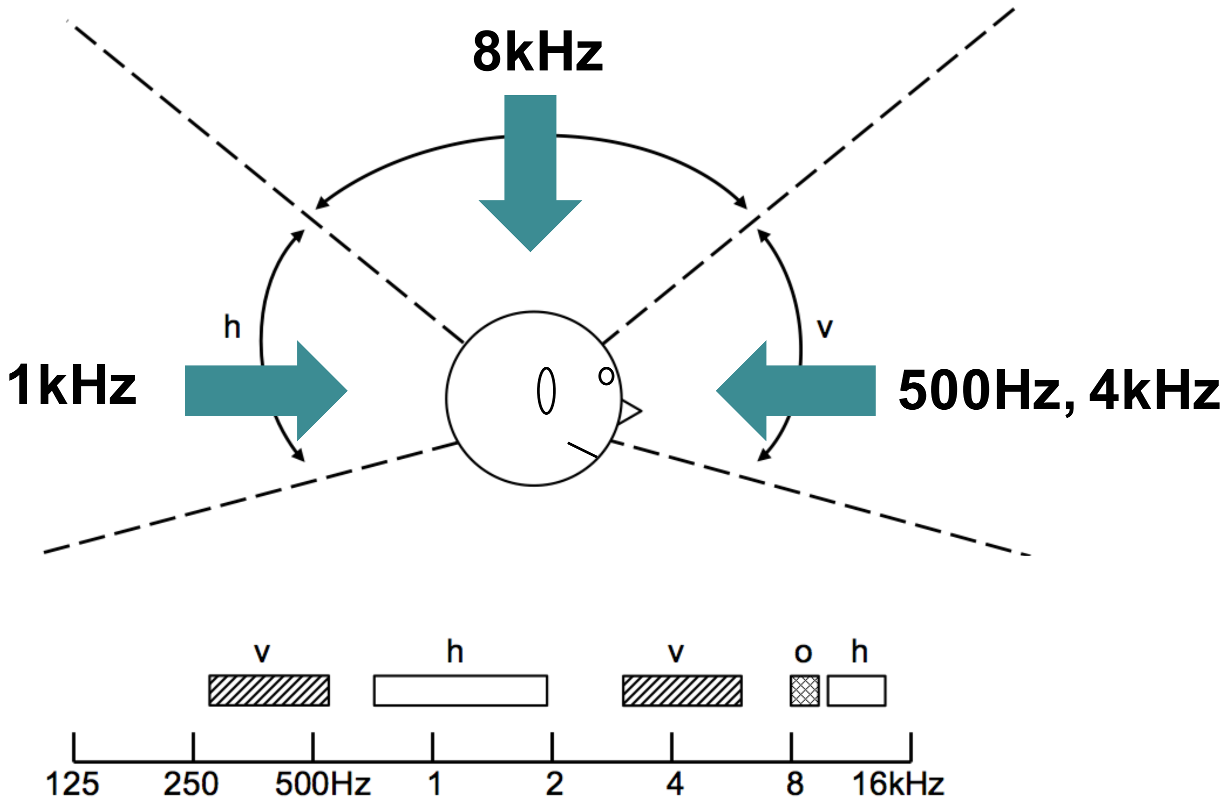
- Pitch height effect for **octave band pink noise**
  - after Cabrera and Tiley (2003); median plane results



Directional bands

University of HUDDERSFIELD

- Blauert (1968): physical mapping between frequency bands and their perceived positions in the median plane.

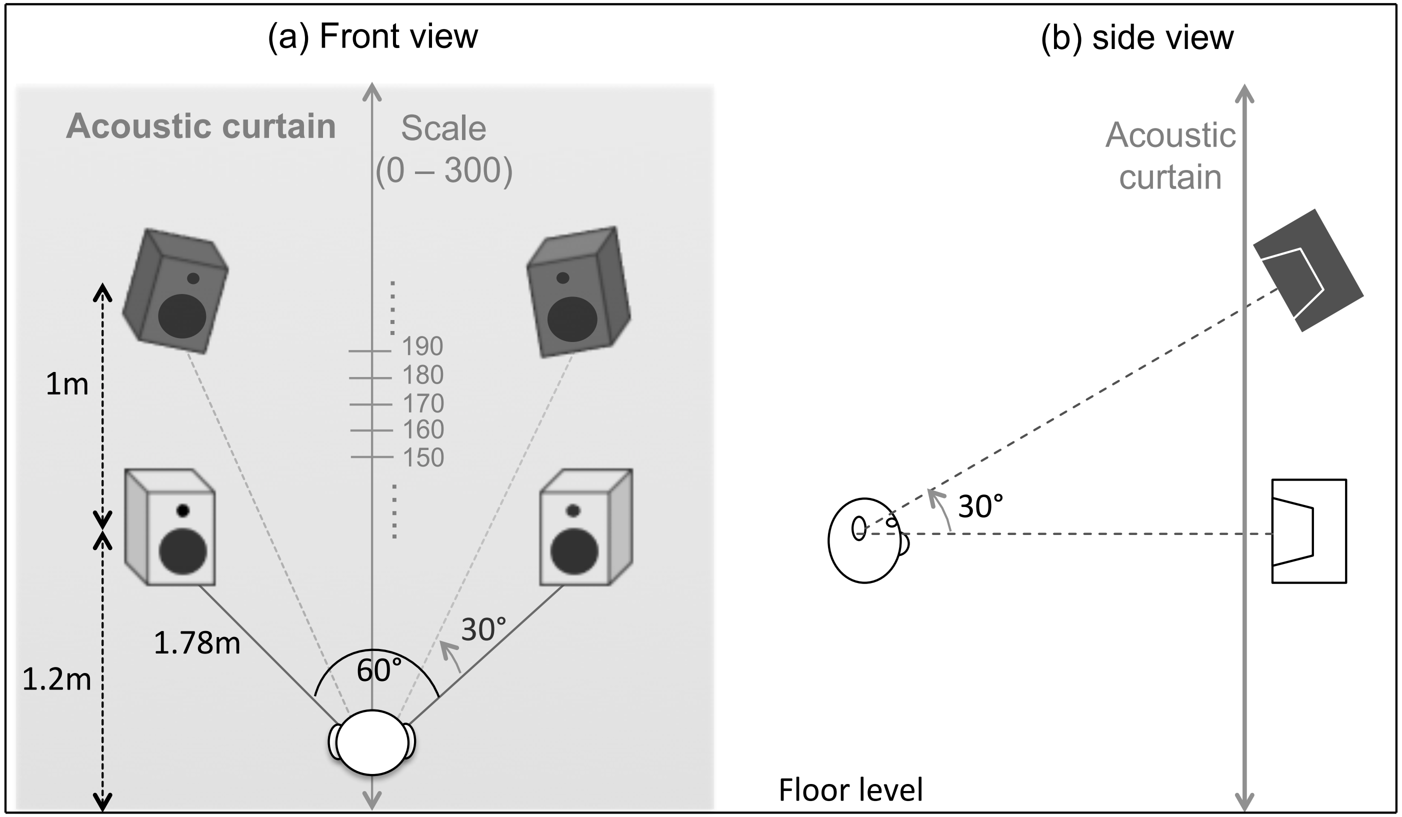


THE QUEEN'S AWARDS FOR ENTERPRISE

Pitch-Height Effect for “Phantom” Source

University of Huddersfield

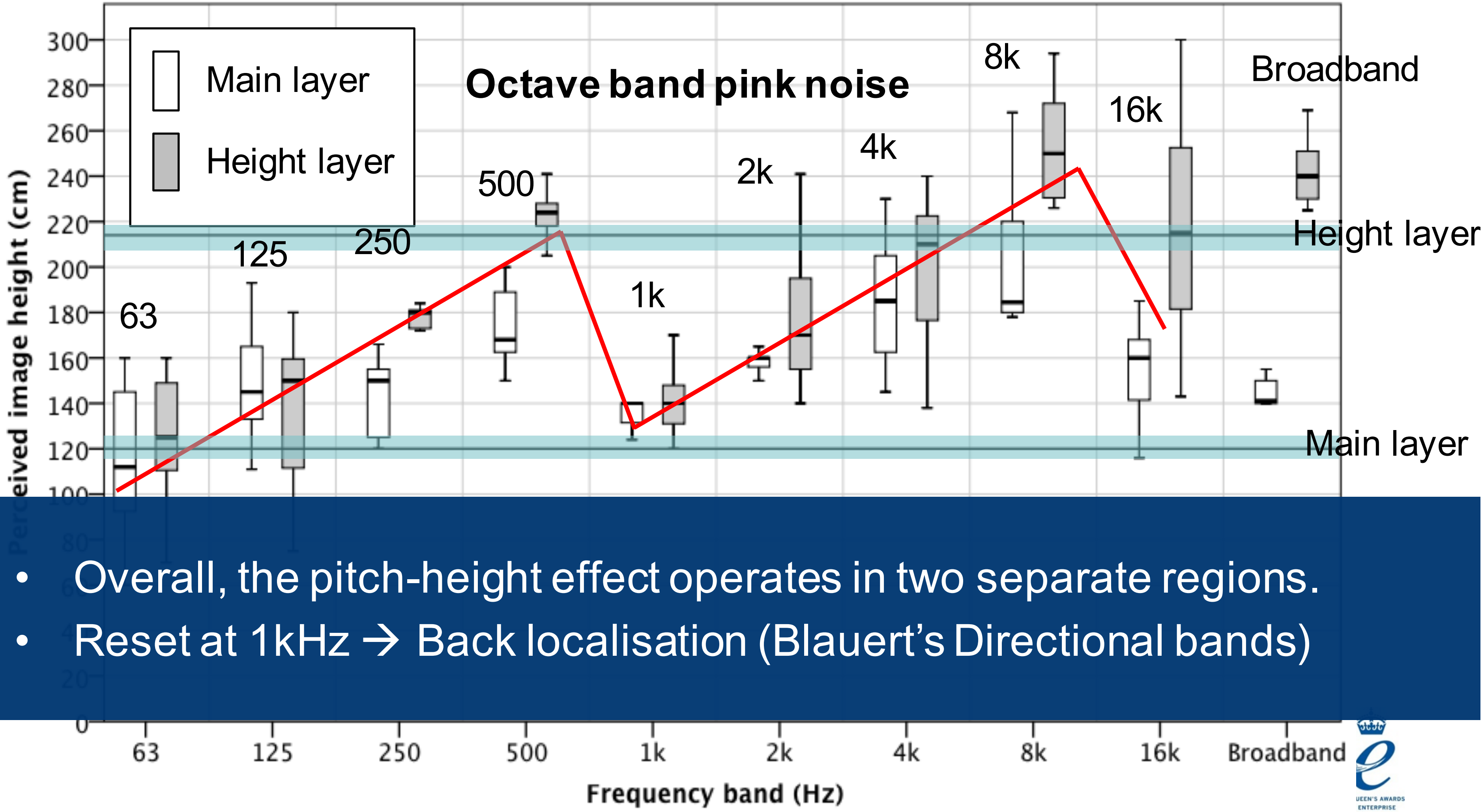
- Pitch-height effect for horizontal **phantom** images from main and height layers (Lee 2015)



H. Lee, “Perceptual Band Allocation (PBA) for the rendering of vertical image spread,” 138th AES, 2015. JAES 2016 under review.

Pitch-Height Effect for Phantom Source

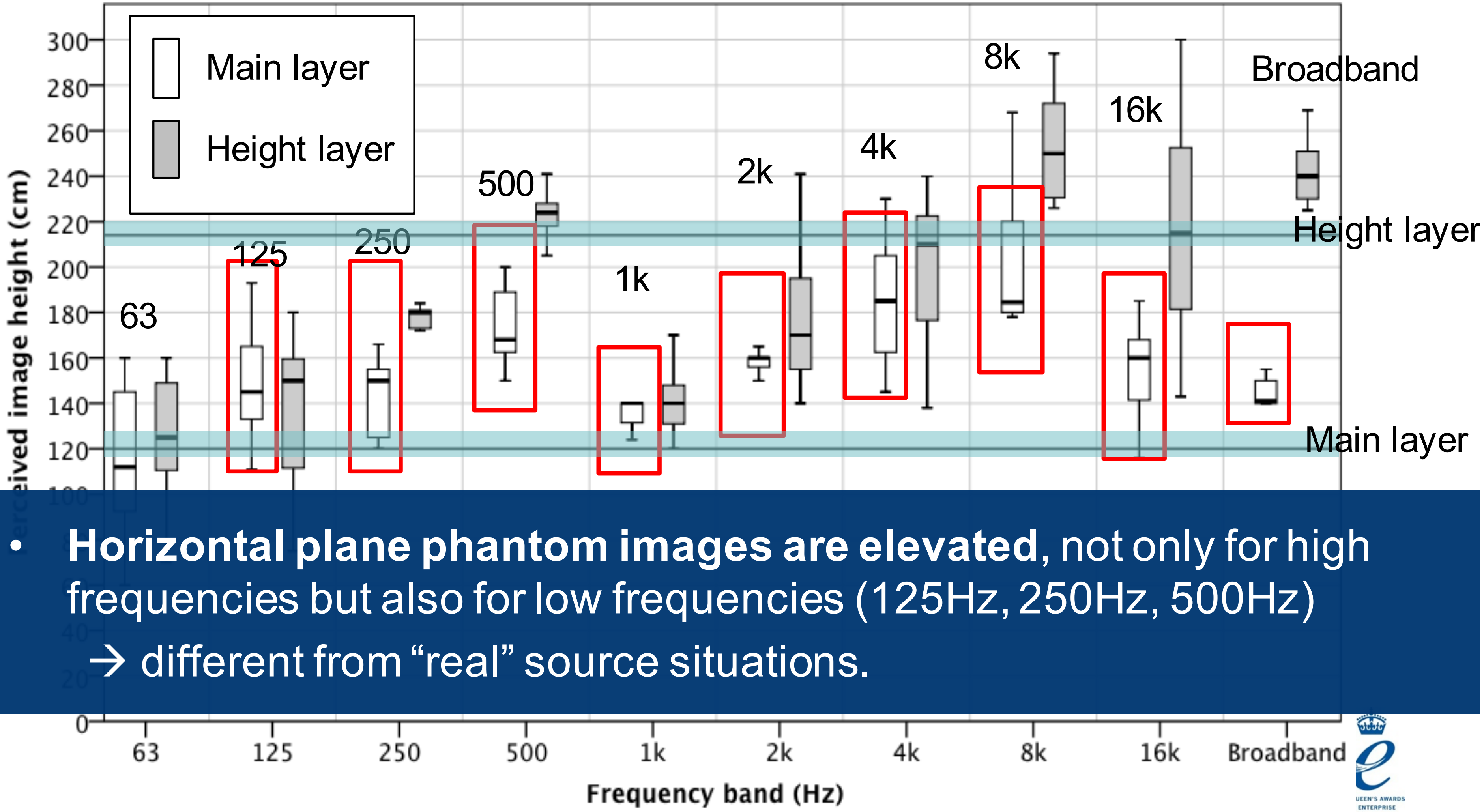
- Pitch-height effect for horizontal **phantom** image (Lee 2015)



- Overall, the pitch-height effect operates in two separate regions.
- Reset at 1kHz → Back localisation (Blauert's Directional bands)

Pitch-Height Effect for Phantom Source

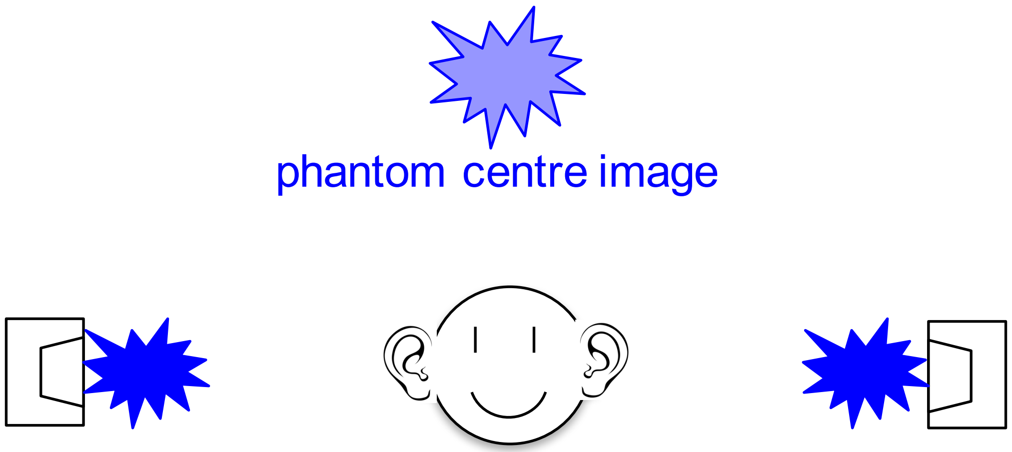
- Pitch-height effect for horizontal **phantom** image (Lee 2015)



- **Horizontal plane phantom images are elevated**, not only for high frequencies but also for low frequencies (125Hz, 250Hz, 500Hz)
  → different from "real" source situations.

Phantom Image Elevation

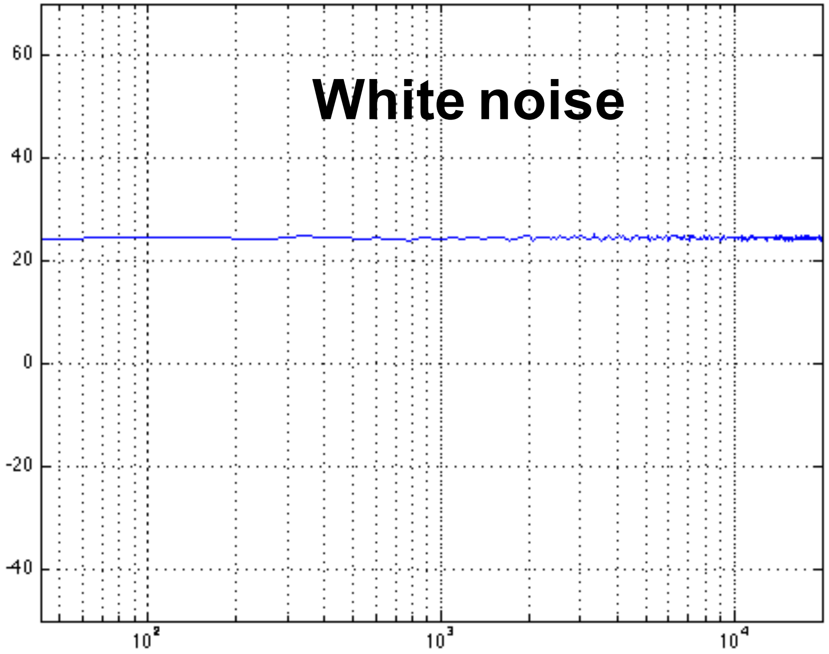
- de Boer (1947): Phantom centre image is perceived to be elevated, and the elevation angle increases as the loudspeaker base angle increases. (180° → overhead region)
- Confirmed by Damaske and Mellert (1969/1970).
  - But only with white noise (650Hz – 4.5kHz)

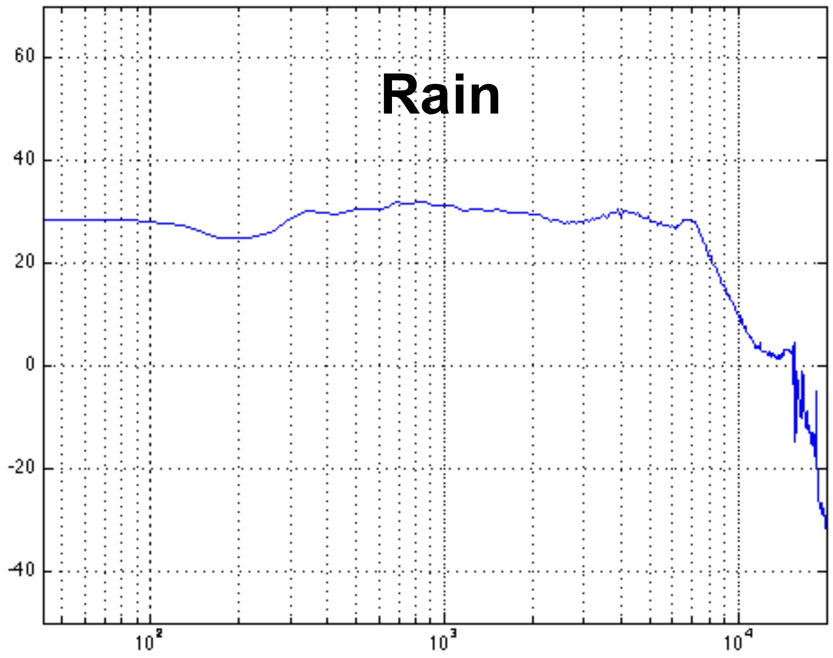


Phantom Image Elevation

University of HUDDERSFIELD

- Sound source dependency (Lee 2015)
  - Responses are most linear and consistent for source with a broad and flat spectrum.



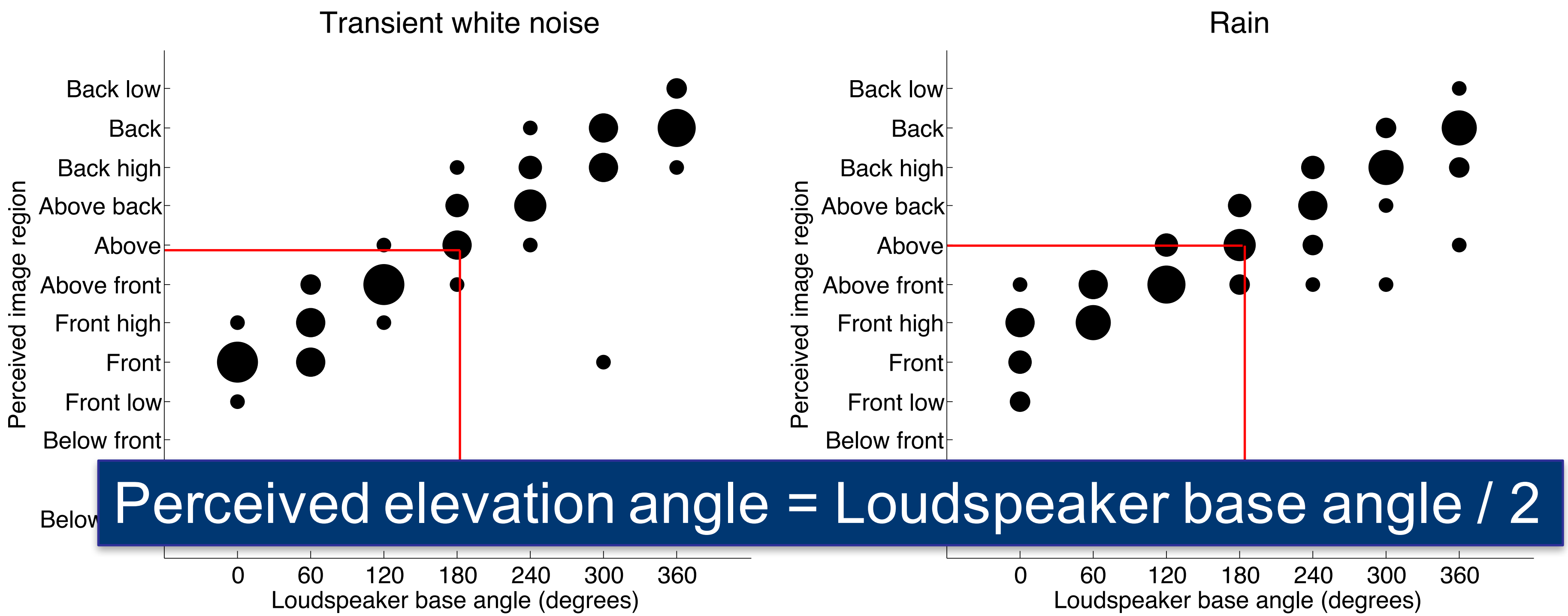


 H. Lee, "Investigation on the Phantom Image Elevation Effect," 139th AES, 2015

Phantom Image Elevation

University of HUDDERSFIELD

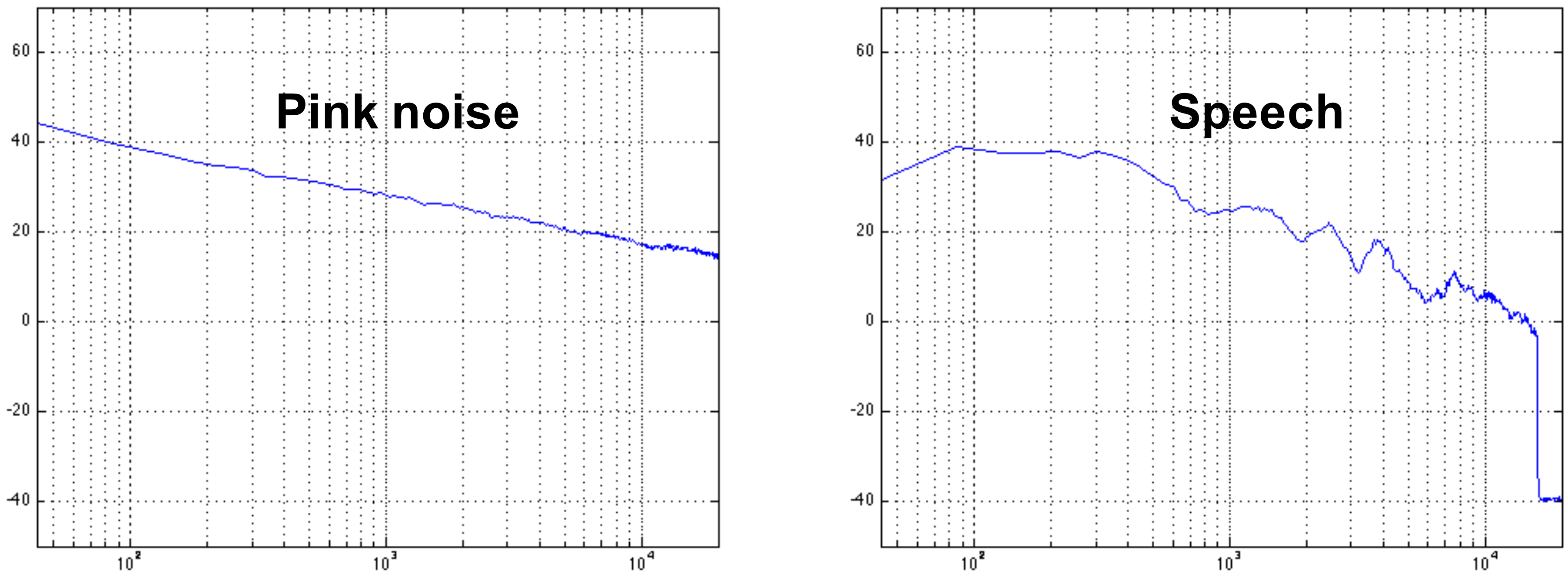
- Sound source dependency (Lee 2015)
  - Responses are most linear and consistent for source with a broad and flat spectrum.



H. Lee, “Investigation on the Phantom Image Elevation Effect,” 139th AES, 2015

Phantom Image Elevation

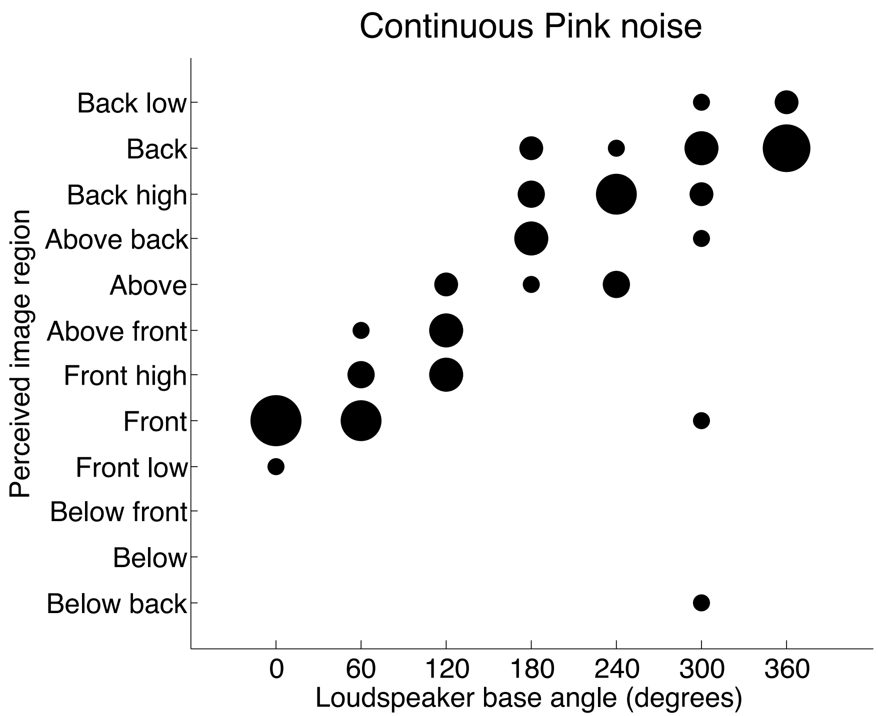
- Sound source dependency (Lee 2015)
  - The elevation effect is weaker for sources with more low frequency energy. (no strong “aboveness”)

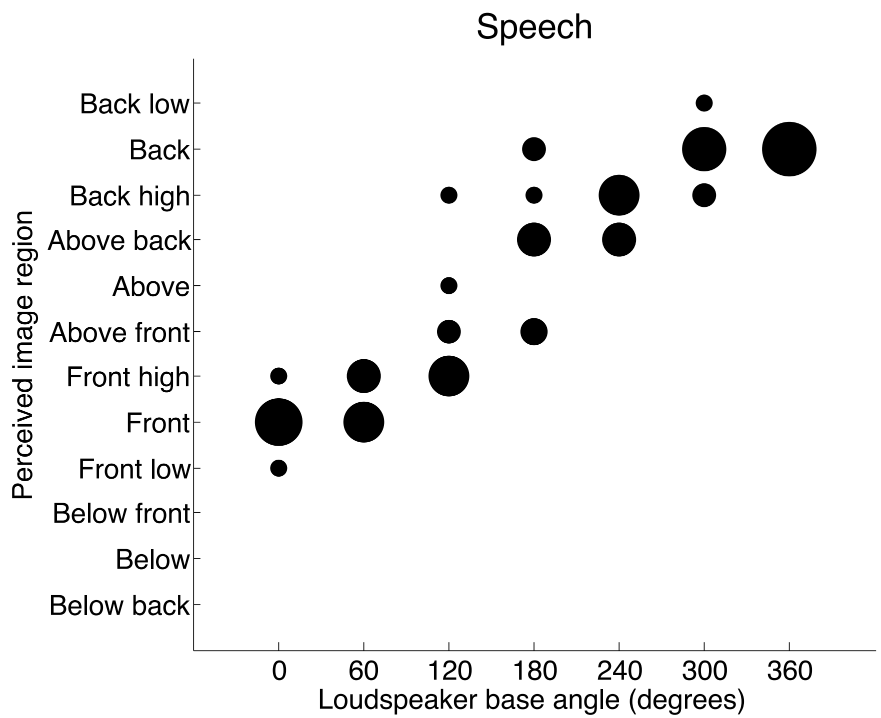


 H. Lee, “Investigation on the Phantom Image Elevation Effect,” 139th AES, 2015

Phantom Image Elevation

- Sound source dependency (Lee 2015)
  - The elevation effect is weaker for sources with more low frequency energy. (no strong “aboveness”)

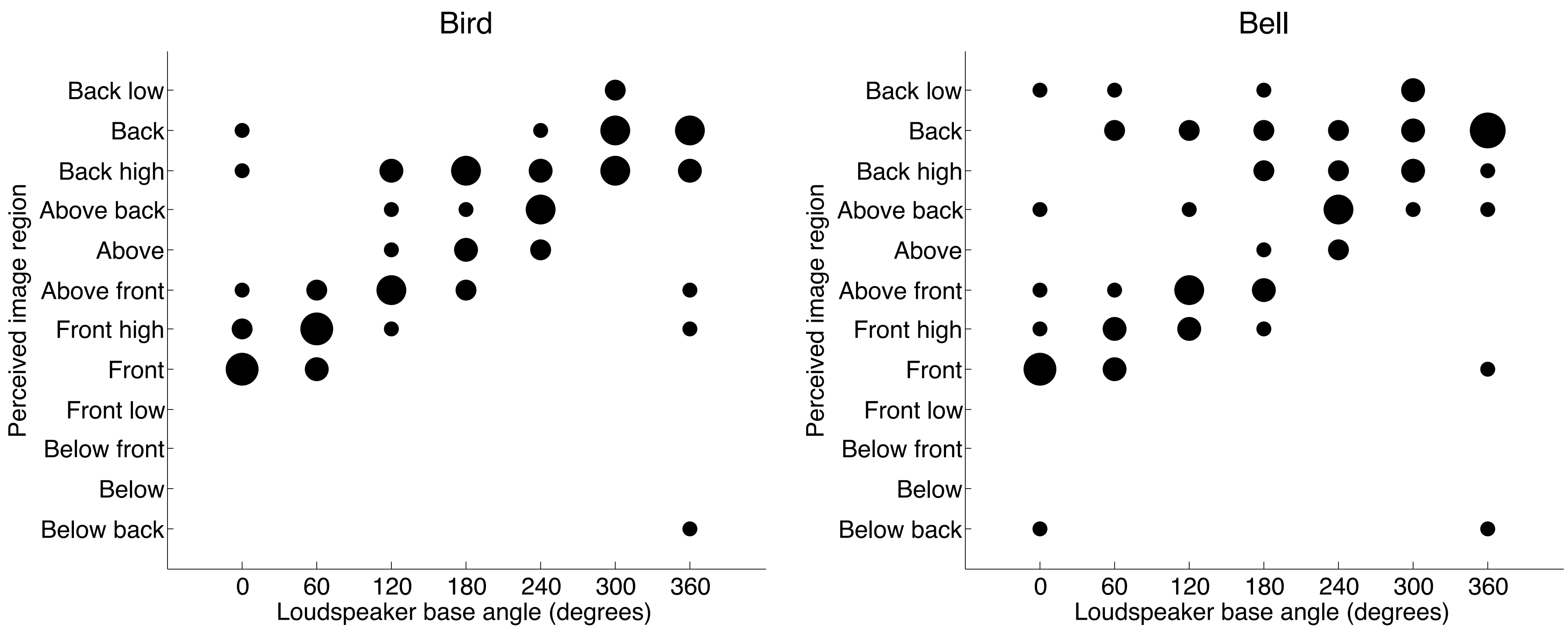




H. Lee, “Investigation on the Phantom Image Elevation Effect,” 139th AES, 2015

Phantom Image Elevation

- Sound source dependency (Lee 2015)
  - Responses are most inconsistent for sources with narrow spectrum or steady-state nature.

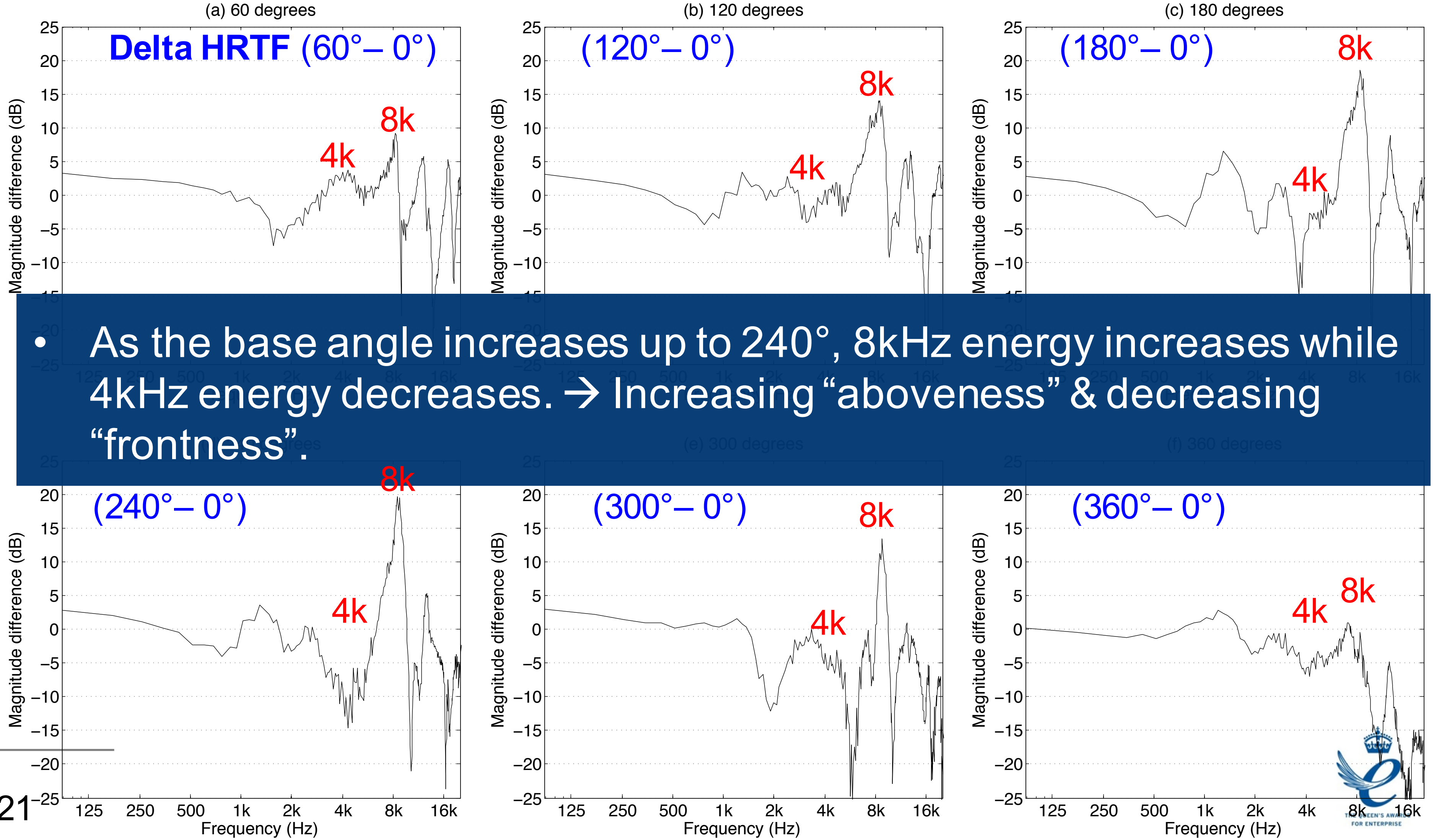


 H. Lee, “Investigation on the Phantom Image Elevation Effect,” 139th AES, 2015

Theoretical explanations

University of Huddersfield

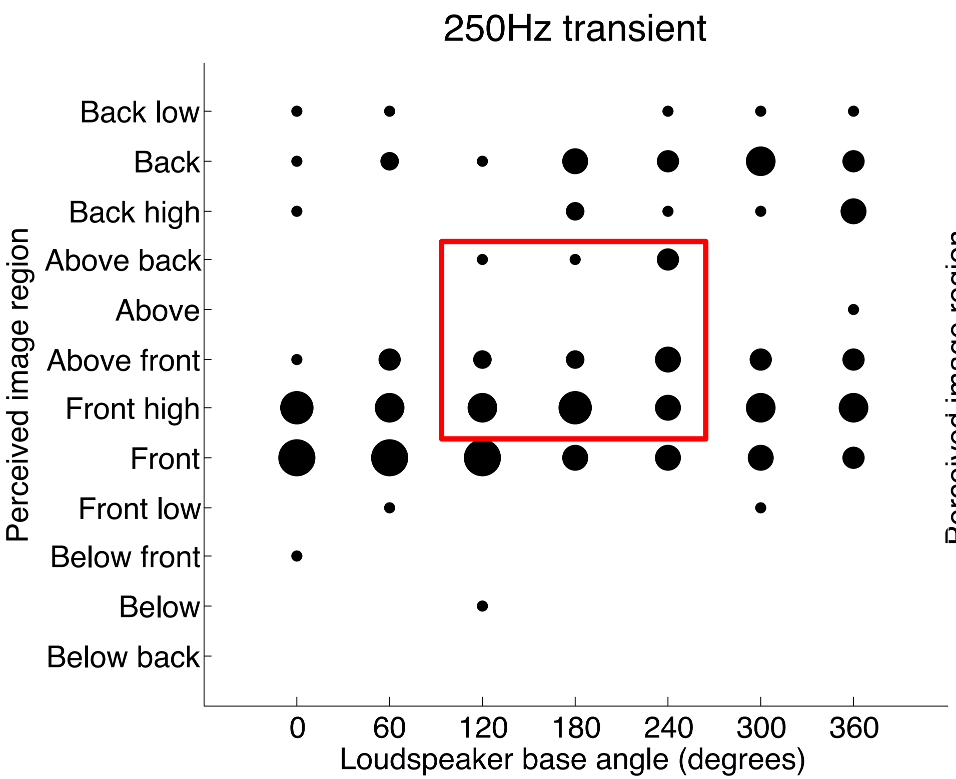
- Spectral energy distribution of ear signal

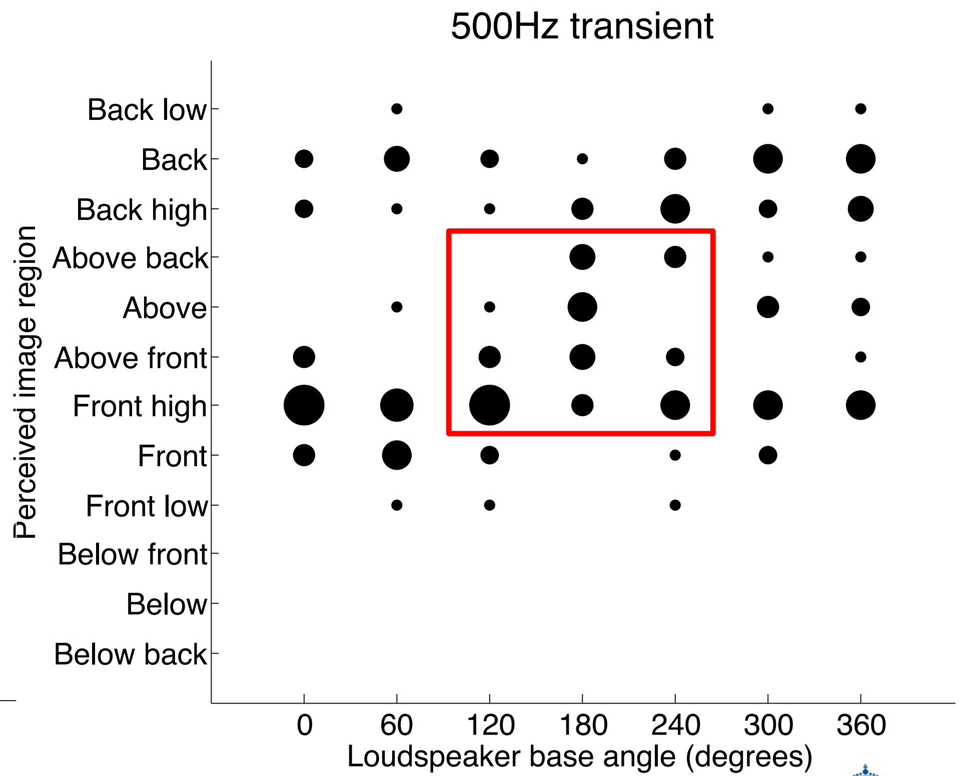


- As the base angle increases up to 240°, 8kHz energy increases while 4kHz energy decreases. → Increasing "aboveness" & decreasing "frontness".

Theoretical explanations

- HRTF does not explain the phantom image elevation for **low frequencies**! (Lee 2016)

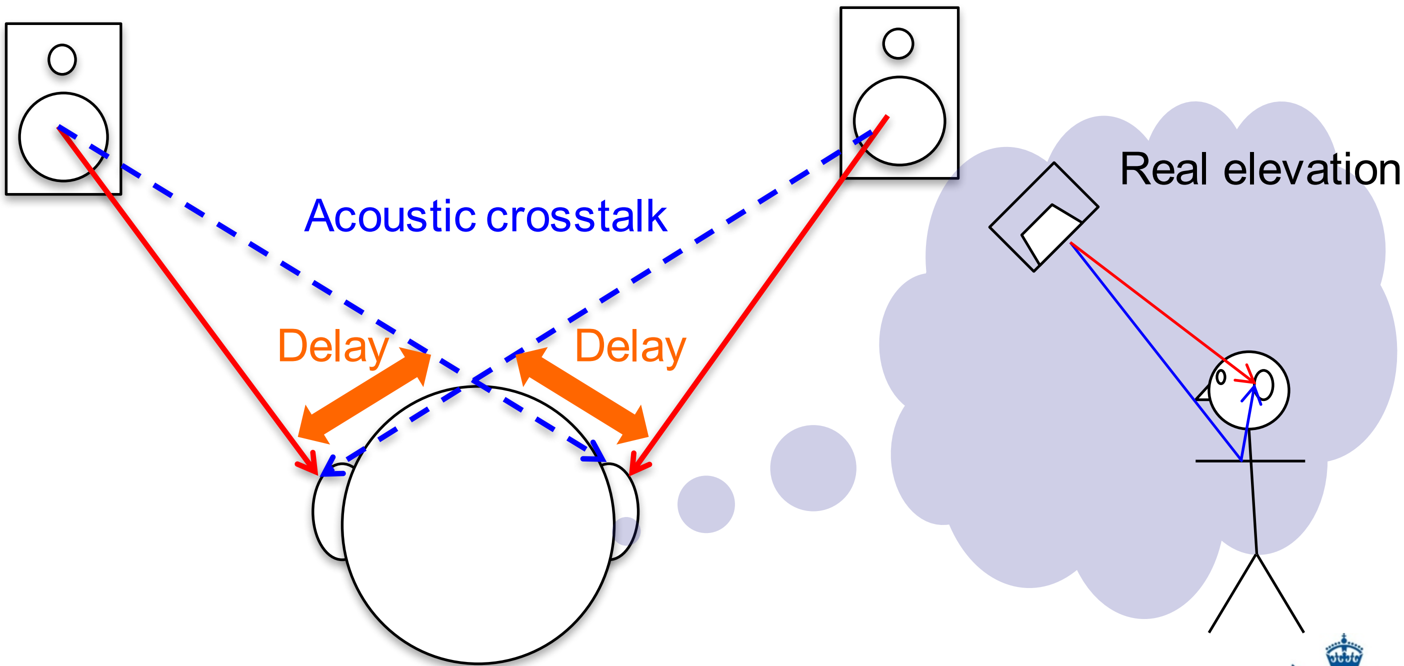




H. Lee, "Phantom image elevation explained," 141st AES, 2016

Theoretical explanations

- A new theory from a **cognitive** perspective (Lee 2015)
  - The brain interprets the **acoustic crosstalk delay** as a shoulder reflection delay for a real elevated source.
  - **Shoulder reflection delay** is the main cue for elevation perception for **frequencies < 3kHz** in the median plane (Algazi et al. 2001)

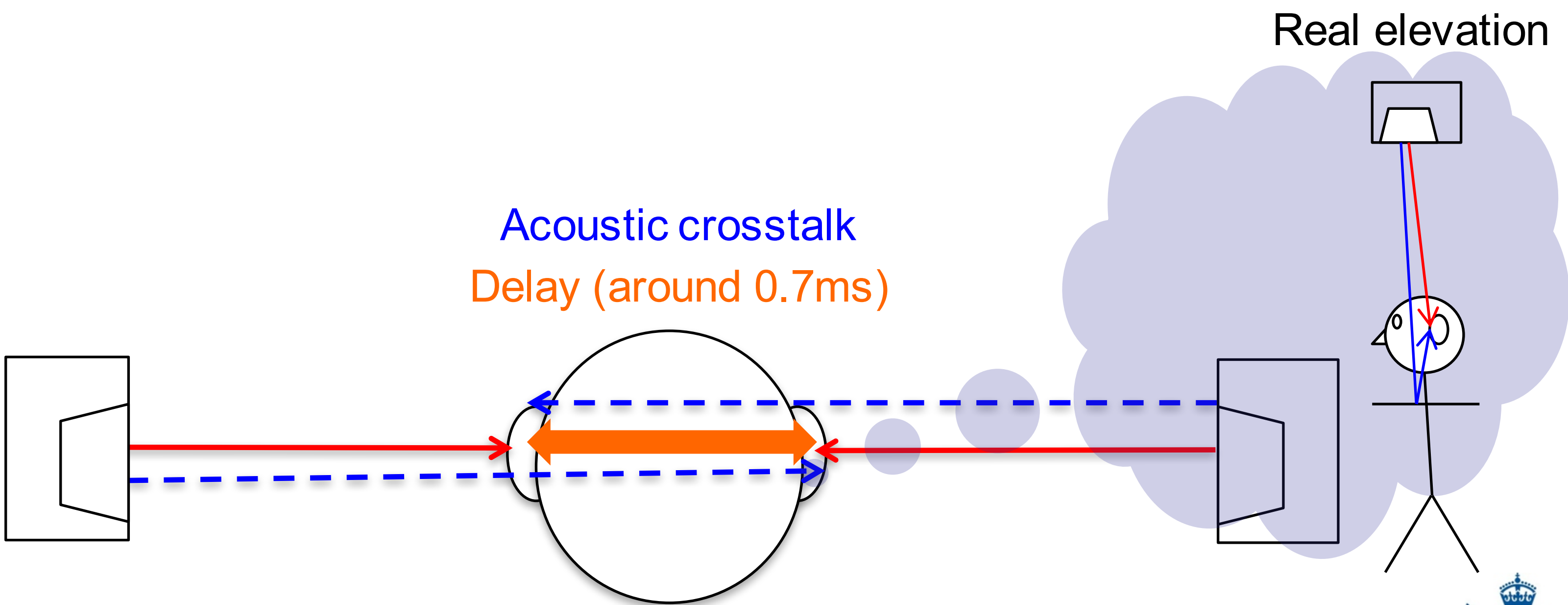


 H. Lee, "Investigation on the Phantom Image Elevation Effect," 139th AES, 2015

Theoretical explanations

University of HUDDERSFIELD

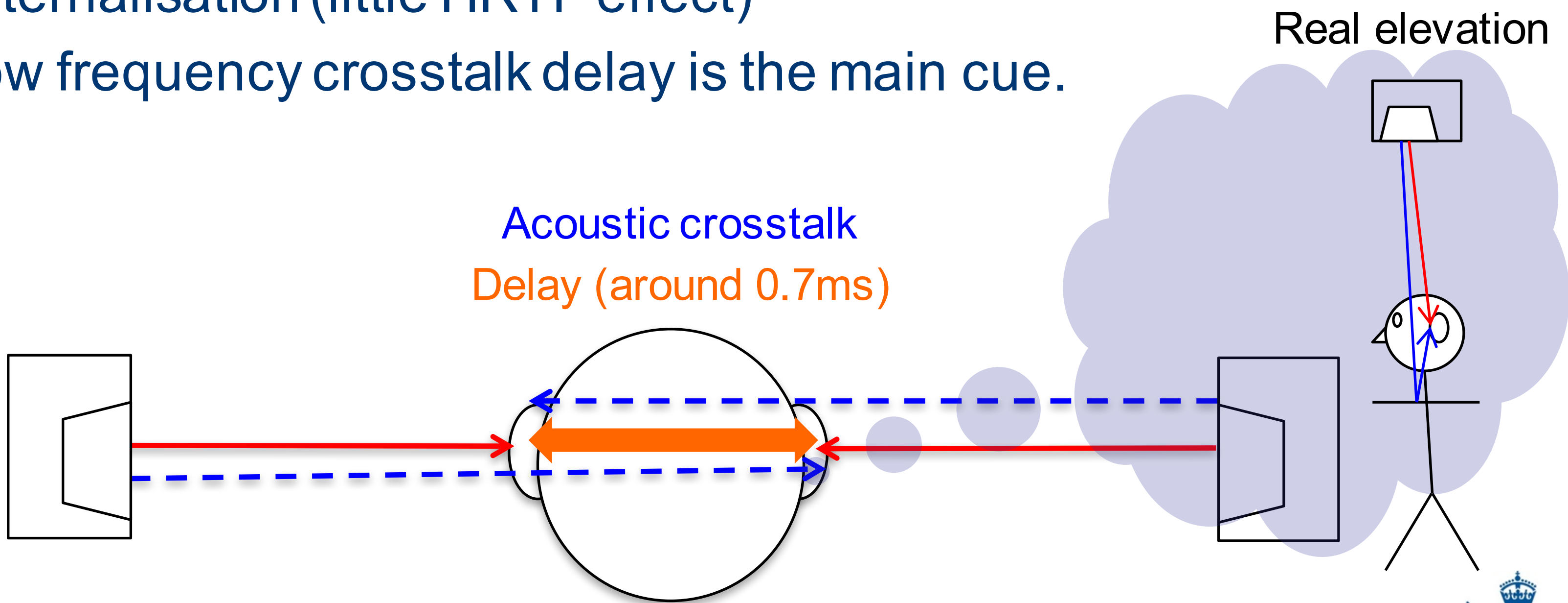
- A new theory from a **cognitive** perspective (Lee 2015)
  - As the **loudspeaker base angle** increases, acoustic crosstalk delay increases (max. around 0.7ms for 180°)
  - As the **real source elevation angle** increases, should reflection delay increases (max. around 0.7ms for a source right above).



 H. Lee, "Investigation on the Phantom Image Elevation Effect," 139th AES, 2015

Theoretical explanations

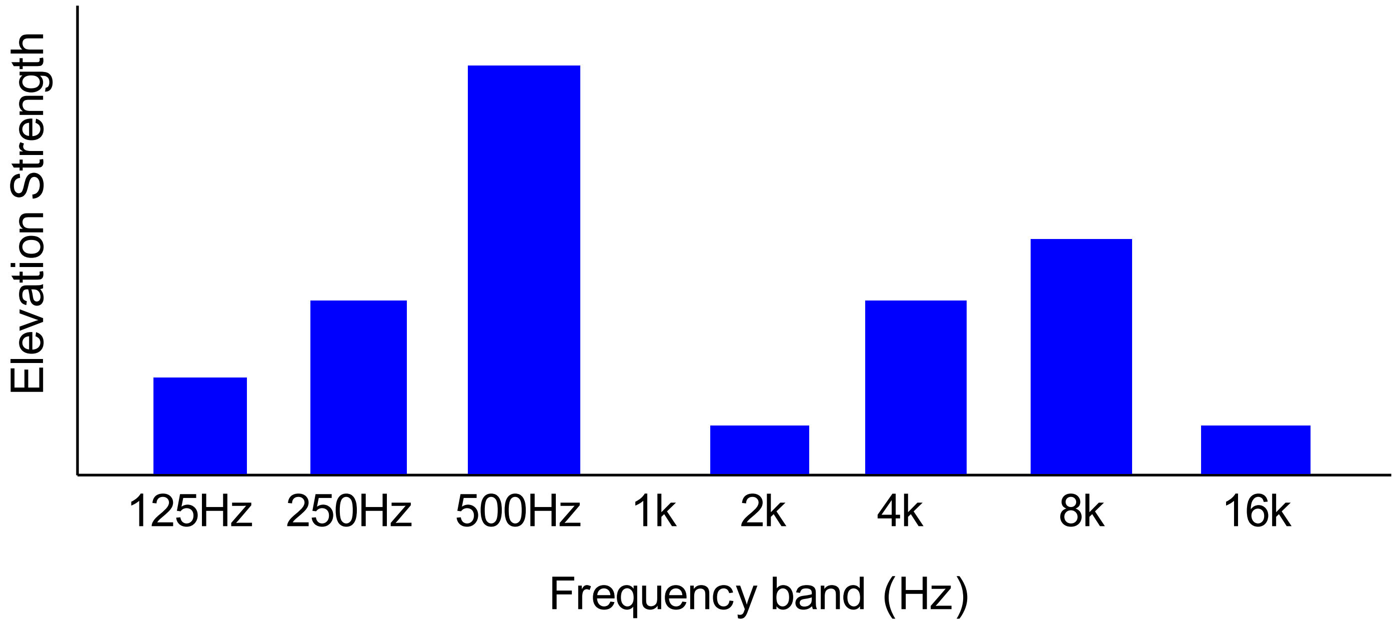
- A new theory from a **cognitive** perspective (Lee 2016)
  - Verified binaurally with BRIRs.
  - With crosstalks removed, no elevation is perceived.
  - With crosstalks delay is made as 0ms, no elevation is perceived.
  - With crosstalk < 3kHz removed, a slight elevation but little externalisation (little HRTF effect)
  - Low frequency crosstalk delay is the main cue.



H. Lee, "Phantom image elevation explained," 141st AES, 2016

Theoretical explanations

- A new theory from a **cognitive** perspective (Lee 2016)
  - Low frequencies: Cognitive effect (Crosstalk delay)
  - High frequencies: Hard-wired effect (Directional bands)

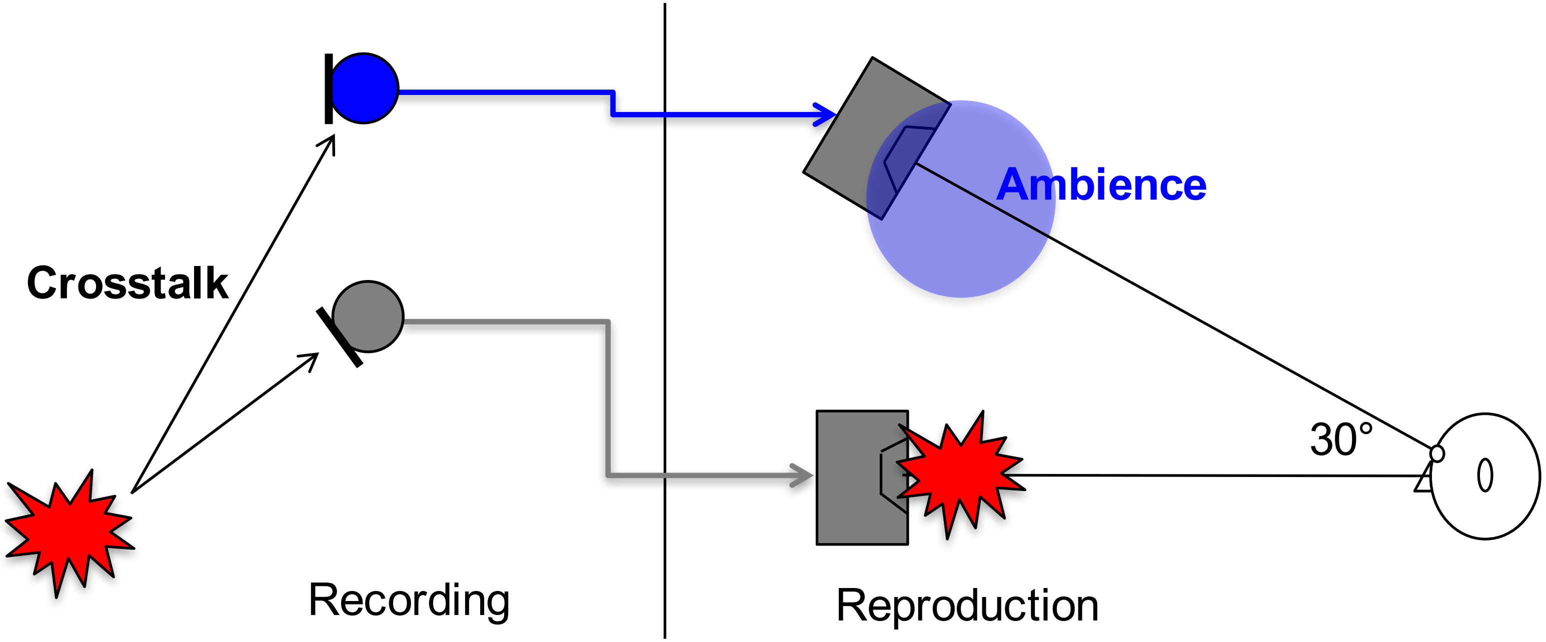


 H. Lee, “Phantom image elevation explained,” 141st AES, 2016

Vertical Interchannel Crosstalk

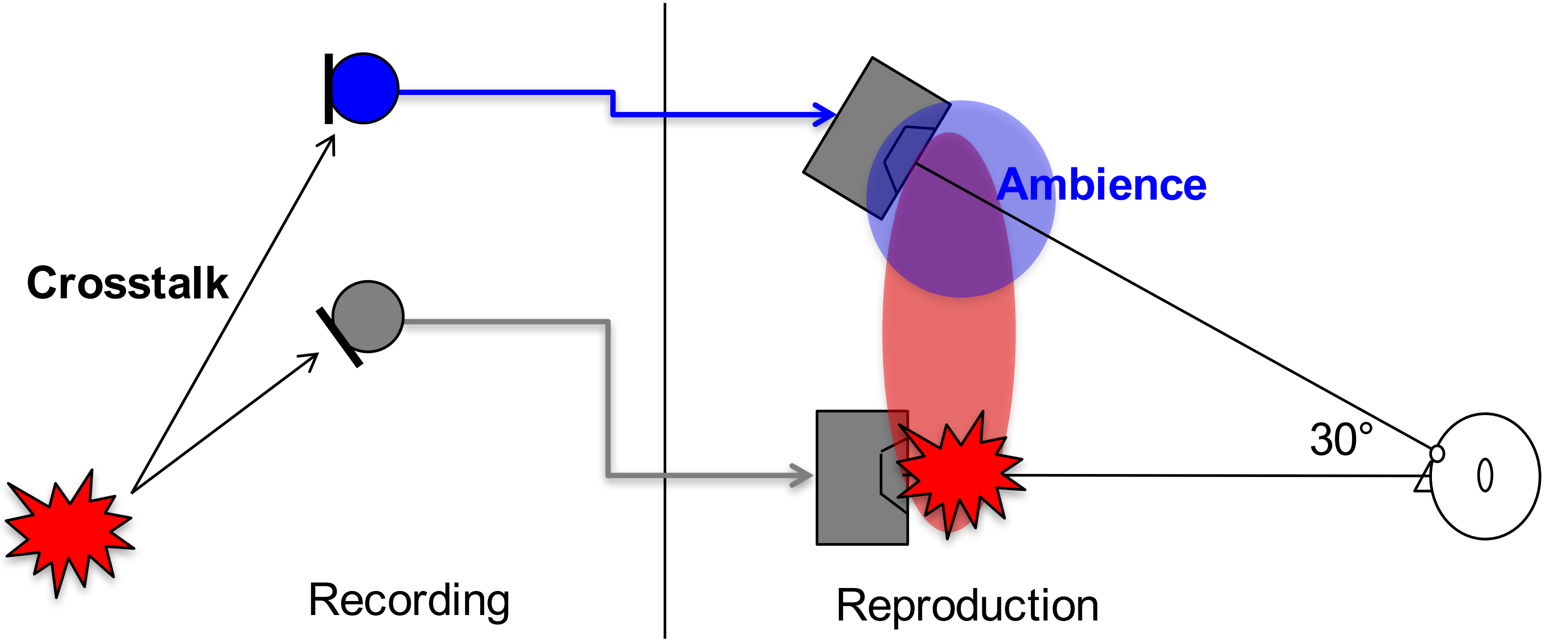
Vertical interchannel crosstalk

- What is vertical interchannel crosstalk?
  - A (delayed) direct sound captured by a height microphone that aims to capture ambience.



Vertical interchannel crosstalk

- What is vertical interchannel crosstalk?
  - A (delayed) direct sound captured by a height microphone that aims to capture ambience
  - Perceptual effects: Localisation shift, spatial & tonal effects, etc.



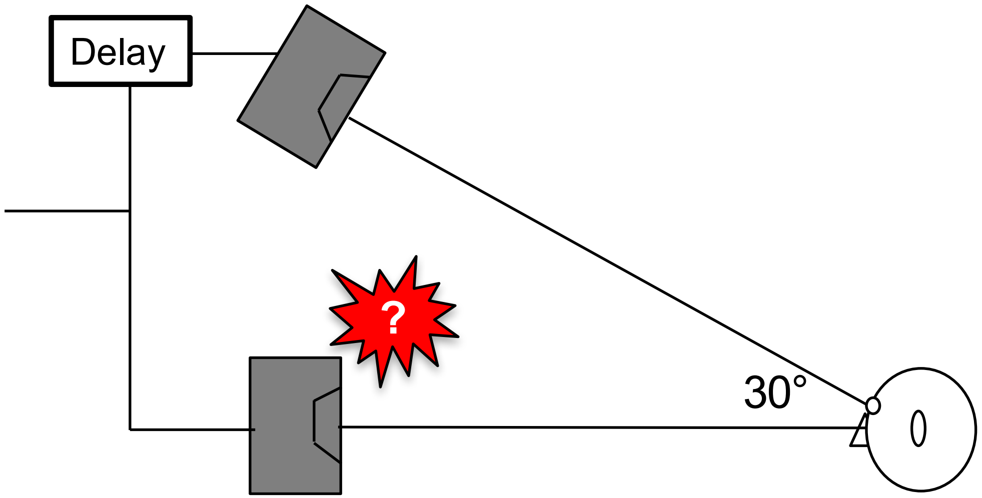
Vertical interchannel crosstalk

University of HUDDERSFIELD

- Vertical time delay (ICTD) effect on localisation (Wallis and Lee 2015)

  - No level reduction but only time delay to height channel

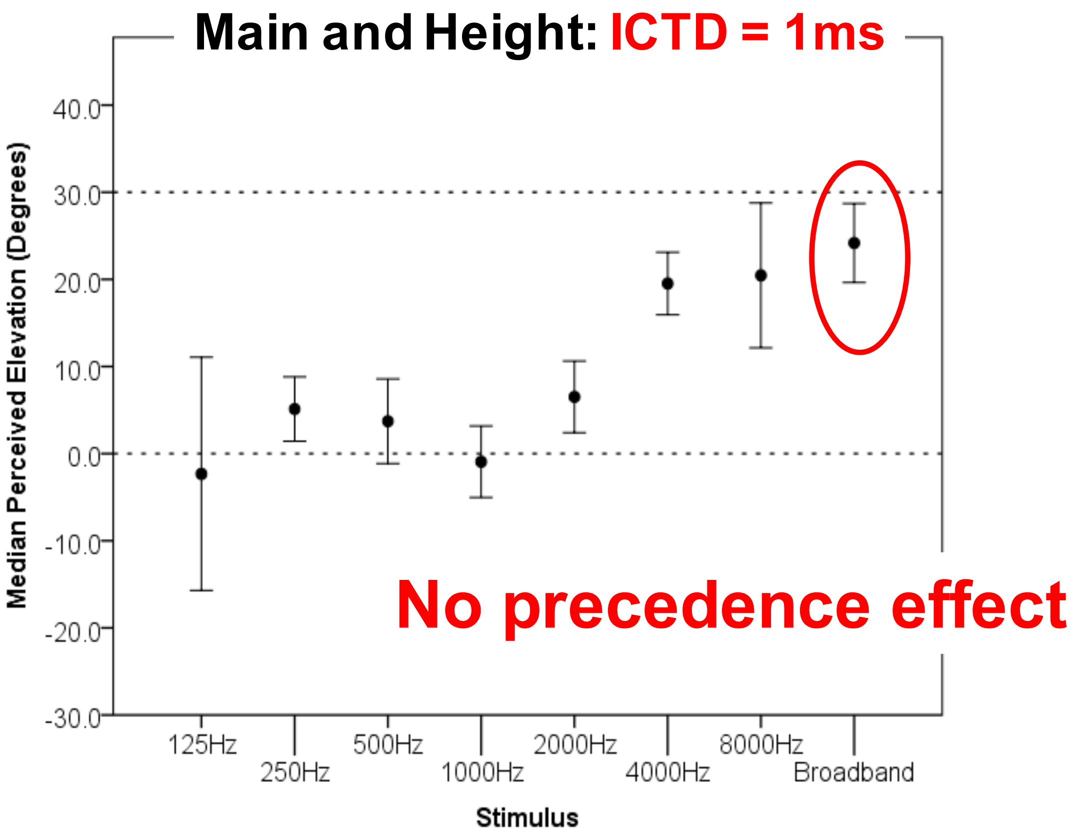
  e.g. Omni mic for height



R. Wallis and H. Lee, “The Effect of Interchannel Time Difference on Localisation in Vertical Stereophony,” JAES 2015

Vertical interchannel crosstalk

- Vertical stereo with ICTD = 1ms (Wallis and Lee 2015)

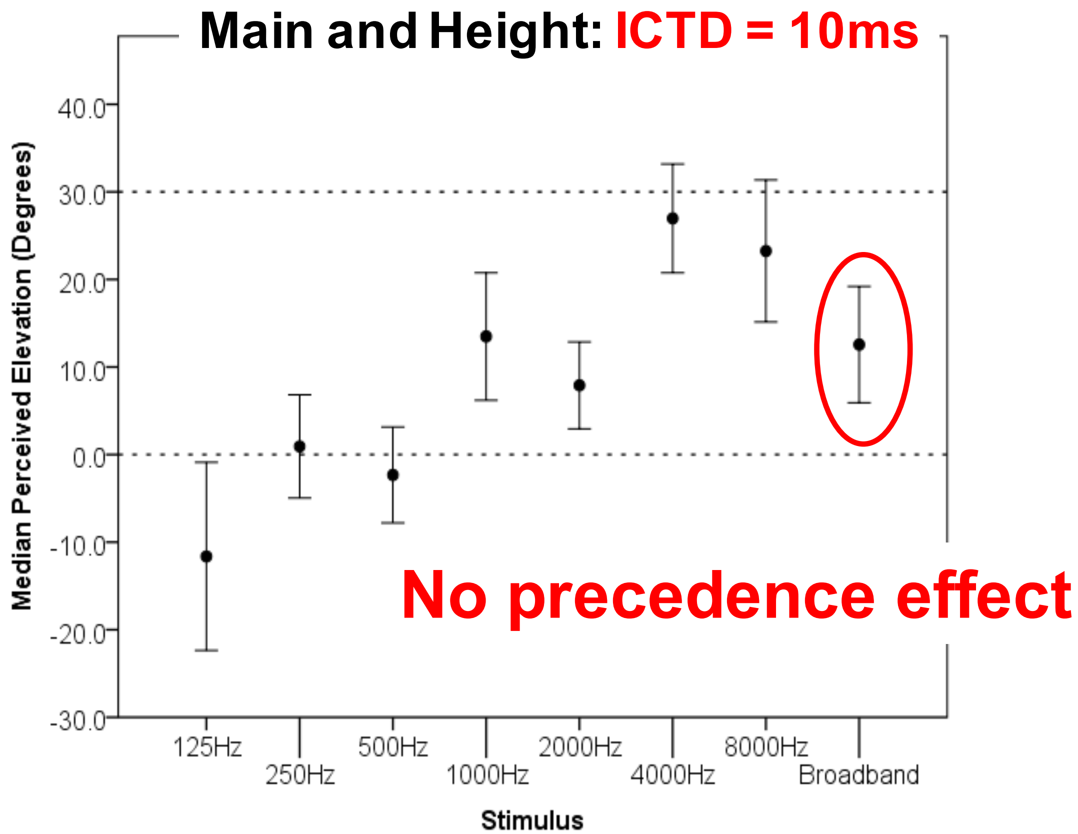


R. Wallis and H. Lee, "The Effect of Interchannel Time Difference on Localisation in Vertical Stereophony," JAES 2015

Vertical interchannel crosstalk

University of HUDDERSFIELD

- Vertical stereo with ICTD = 10ms (Wallis and Lee 2015)



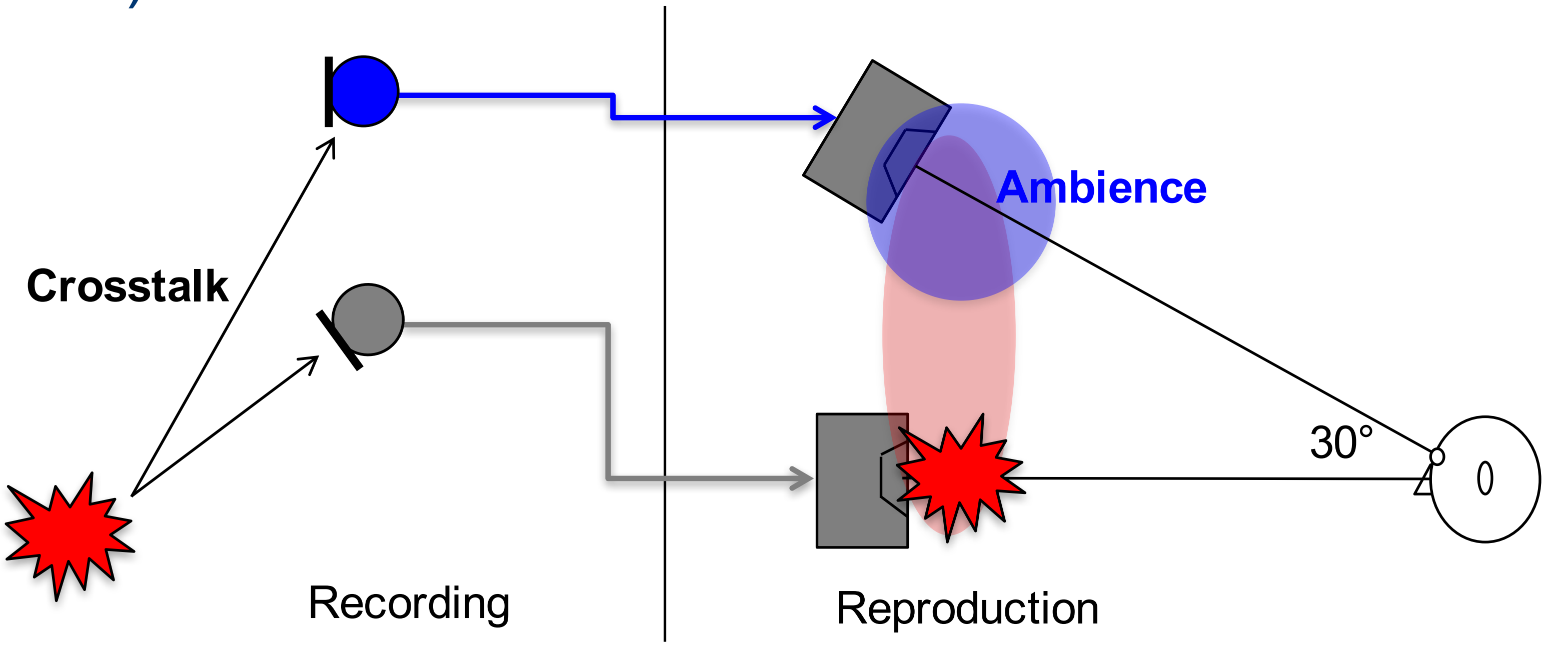
R. Wallis and H. Lee, "The Effect of Interchannel Time Difference on Localisation in Vertical Stereophony," JAES 2015

THE QUEEN'S AWARDS FOR ENTERPRISE

Vertical interchannel crosstalk

University of HUDDERSFIELD

- **6 to 9dB of vertical crosstalk reduction** is required for localisation at the perceived position of lower loudspeaker image (source dependent) (Lee 2011, Wallis and Lee 2016)

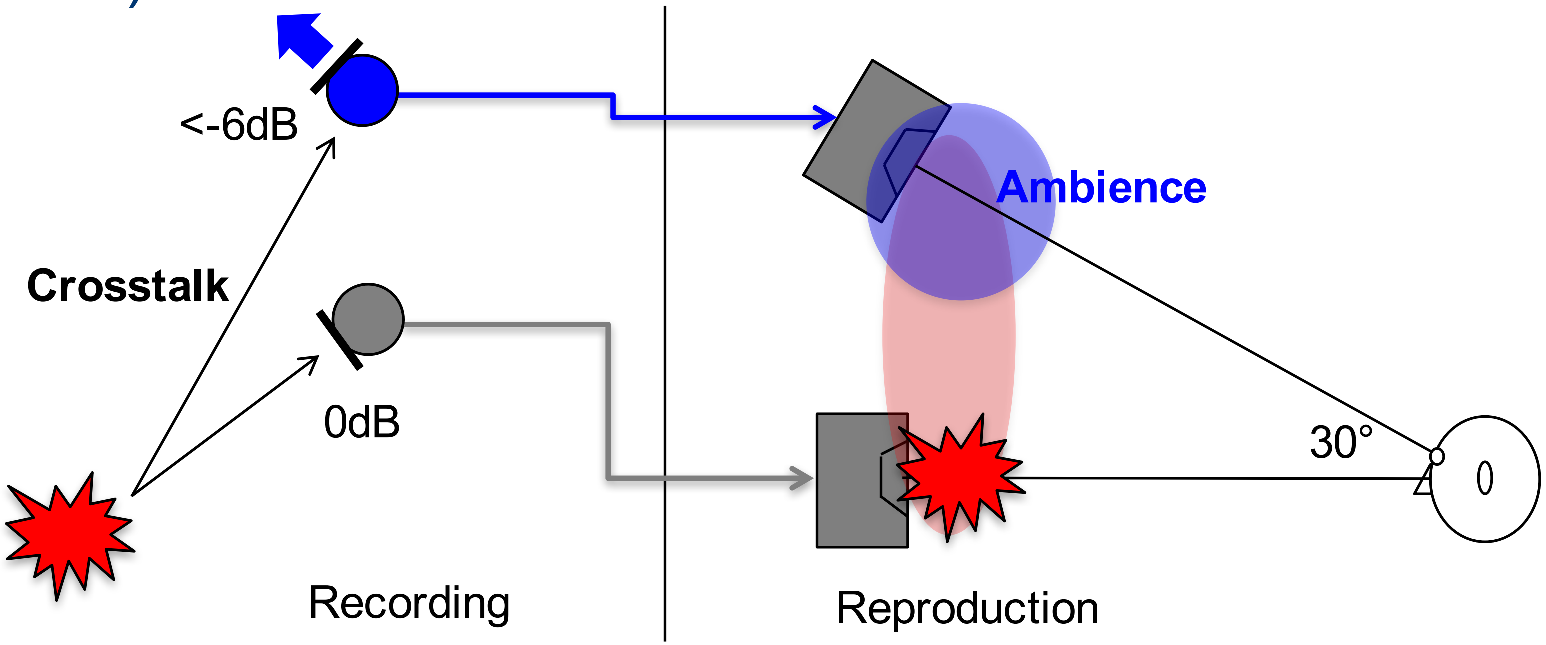


H. Lee, "The Relationship between Interchannel Time and Level Differences in Vertical Localisation and Masking," 131st AES, 2011.

Vertical interchannel crosstalk

University of HUDDERSFIELD

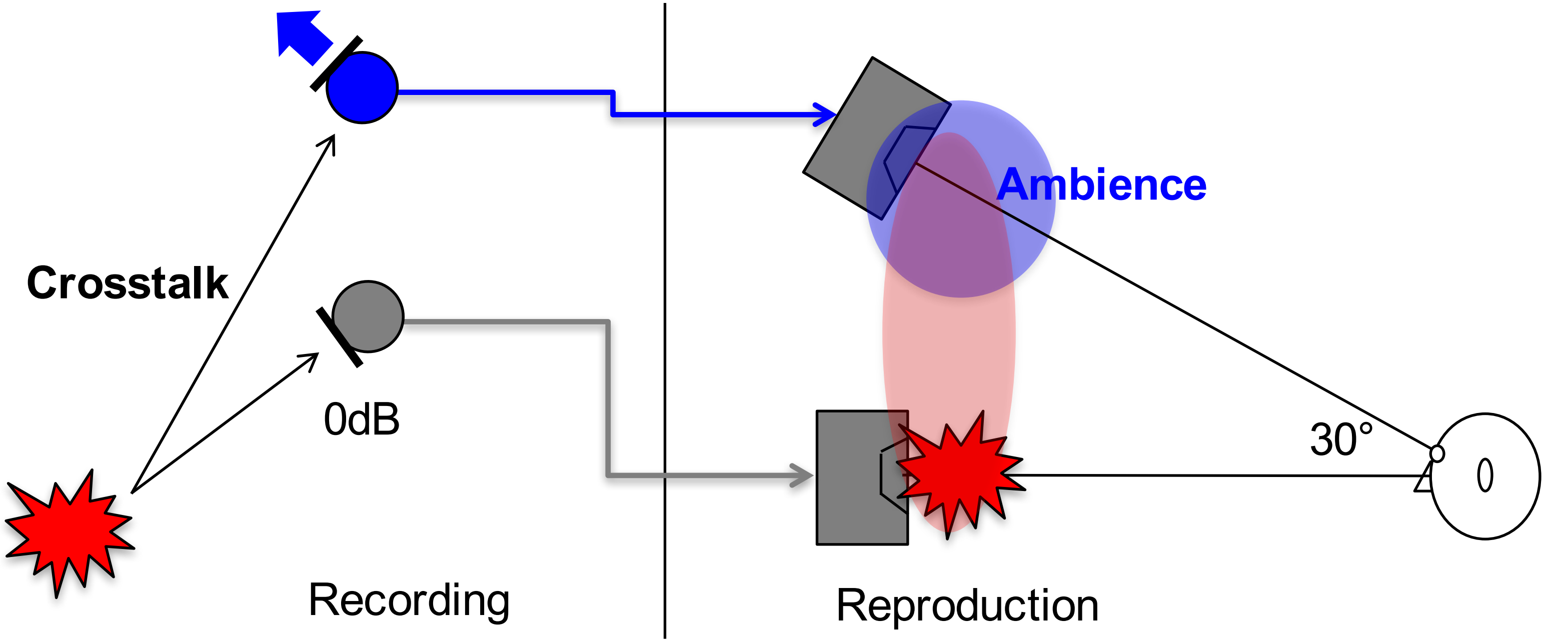
- **6 to 9dB of vertical crosstalk reduction** is required for localisation at the perceived position of lower loudspeaker image (source dependent) (Lee 2011, Wallis and Lee 2016)



R. Wallis and H. Lee, "The Reduction of Vertical Interchannel Crosstalk: The Analysis of Localization Thresholds for Musical Sources," 140th AES, 2016.

Vertical interchannel crosstalk

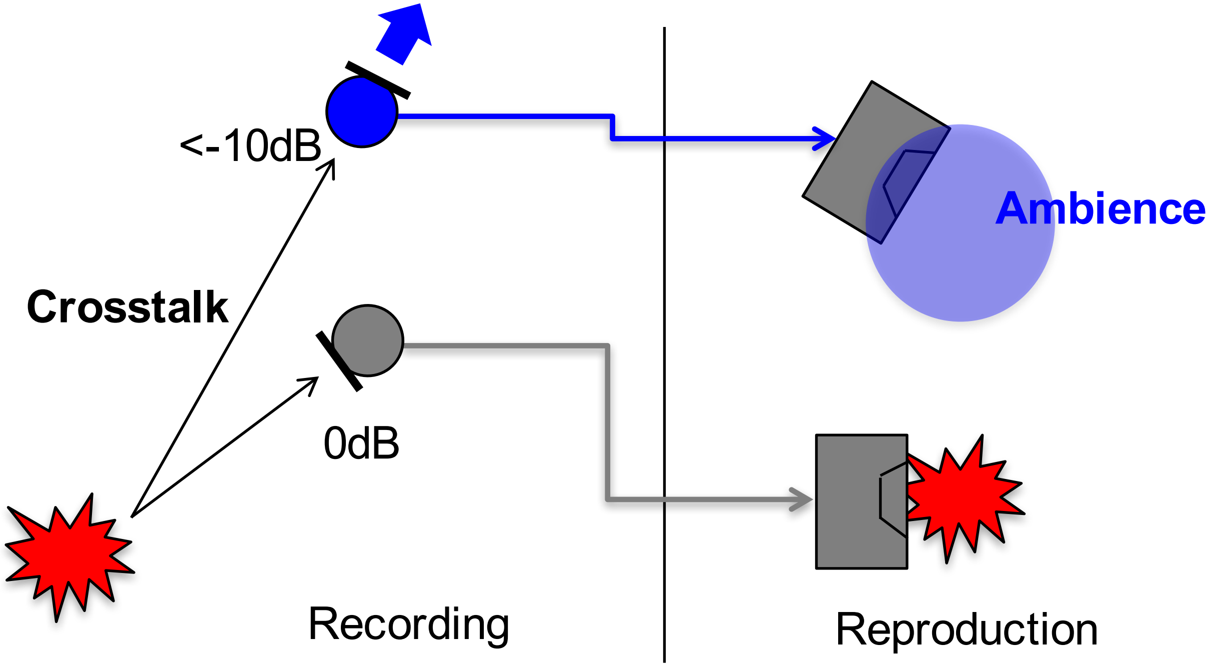
- How much level attenuation of direct sound is required for the perceptual effects of vertical crosstalk to be "**completely inaudible**"?



H. Lee, "The Relationship between Interchannel Time and Level Differences in Vertical Localisation and Masking," 131st AES, 2011.

Vertical interchannel crosstalk

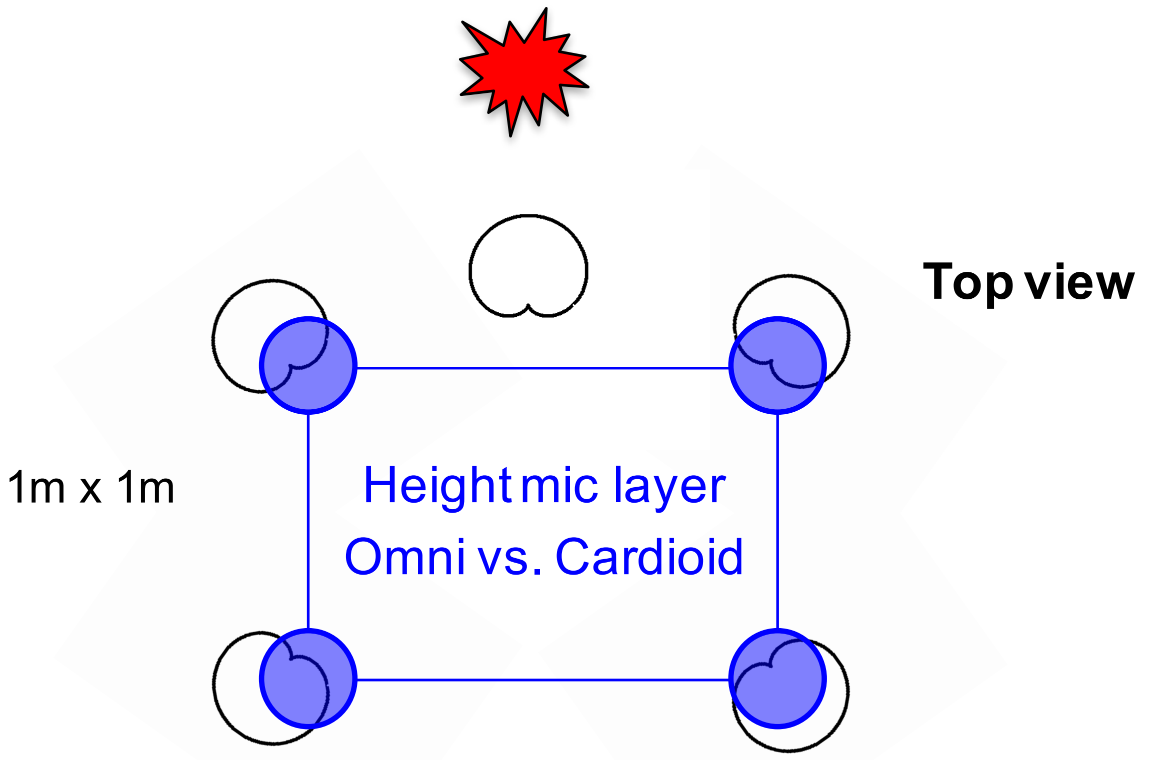
- **At least 10dB of direct sound attenuation** is required of the height microphone to make the vertical crosstalk completely inaudible (Lee 2011)



H. Lee, "The Relationship between Interchannel Time and Level Differences in Vertical Localisation and Masking," 131st AES, 2011.

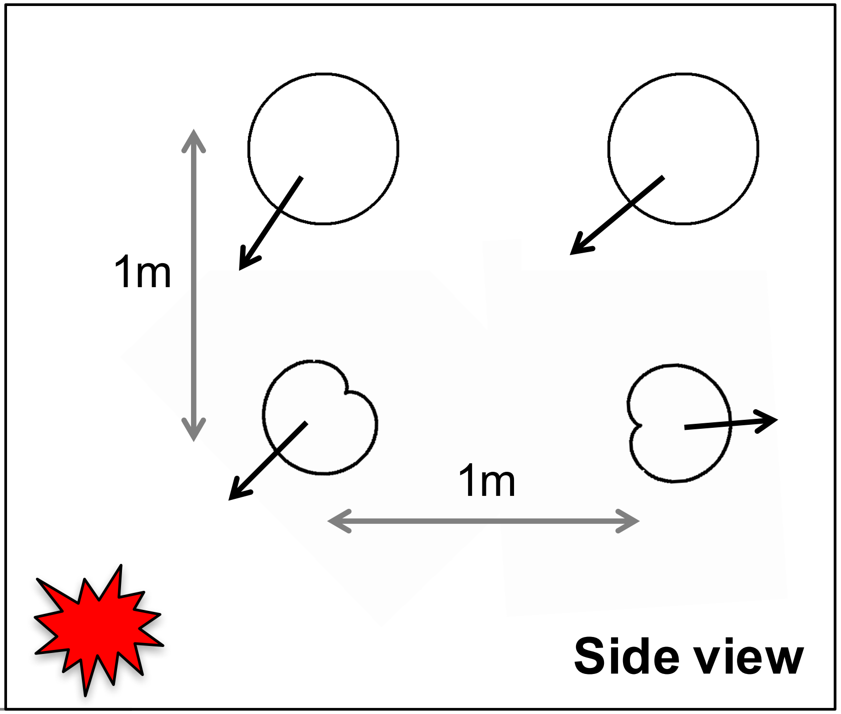
Demo: Omni vs. Cardioid for height

- Height mic polar pattern: **Omni vs. Cardioid**
- 9-channel 3D mic array
- Venue: St. Paul's concert hall (RT=2.1sec) in Huddersfield, UK

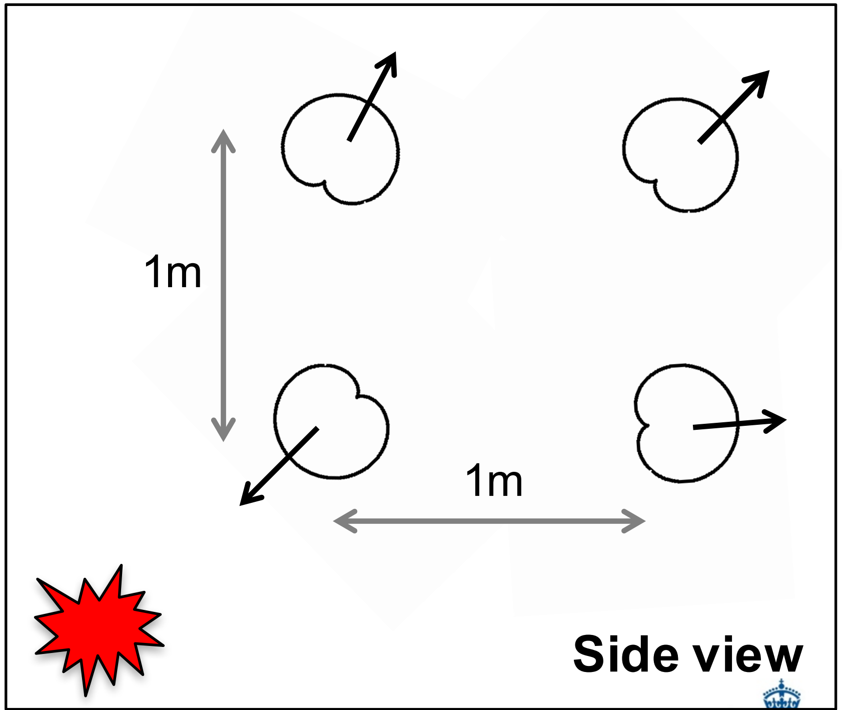


Demo: Omni vs. Cardioid for height

- Height mic polar pattern: **Omni vs. Cardioid**
- 9-channel 3D mic array
- Venue: St. Paul's concert hall (RT=2.1sec) in Huddersfield, UK



VS.



Demo: Omni vs. Cardioid for height

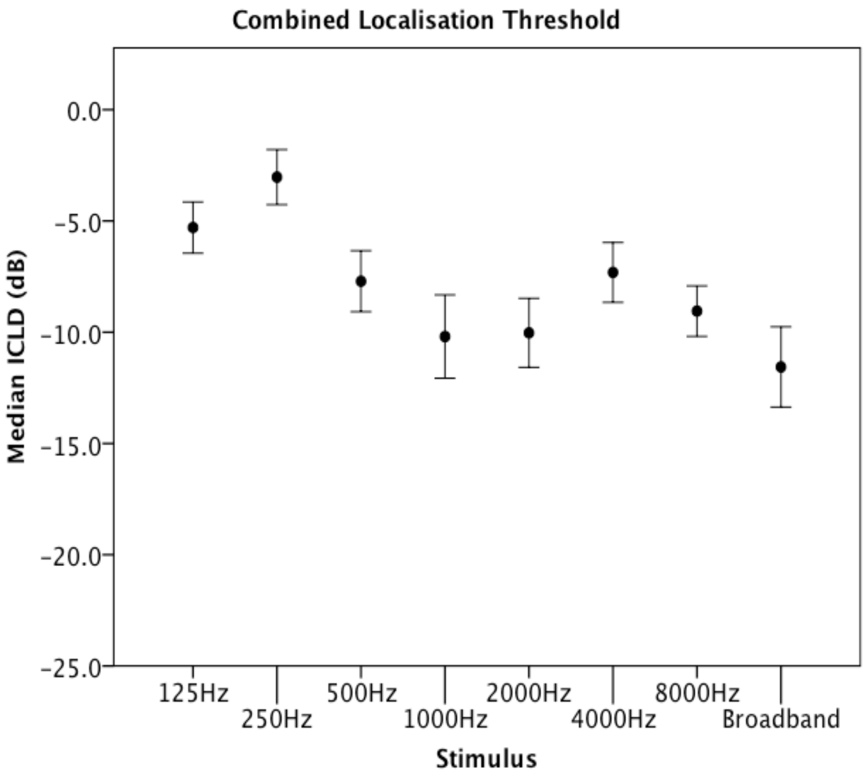
Demo: Omni vs. Cardioid for height

- Omni height: source-related effect (localisation shift and colouration due to comb-filtering)
  - Colouration gets worse as the source has more high frequencies.
- Backward-facing cardioid: environment-related effect (perceived source distance, listener envelopment)
- Backward-facing cardioid has more headroom to increase height ambience level without affecting localisation and tone colour.

Demo: Band-adaptive level reduction

University of Huddersfield

- Localised thresholds for octave-band pink noises (Wallis and Lee 2016)



Band-adaptive level reduction (Wallis and Lee 2016)

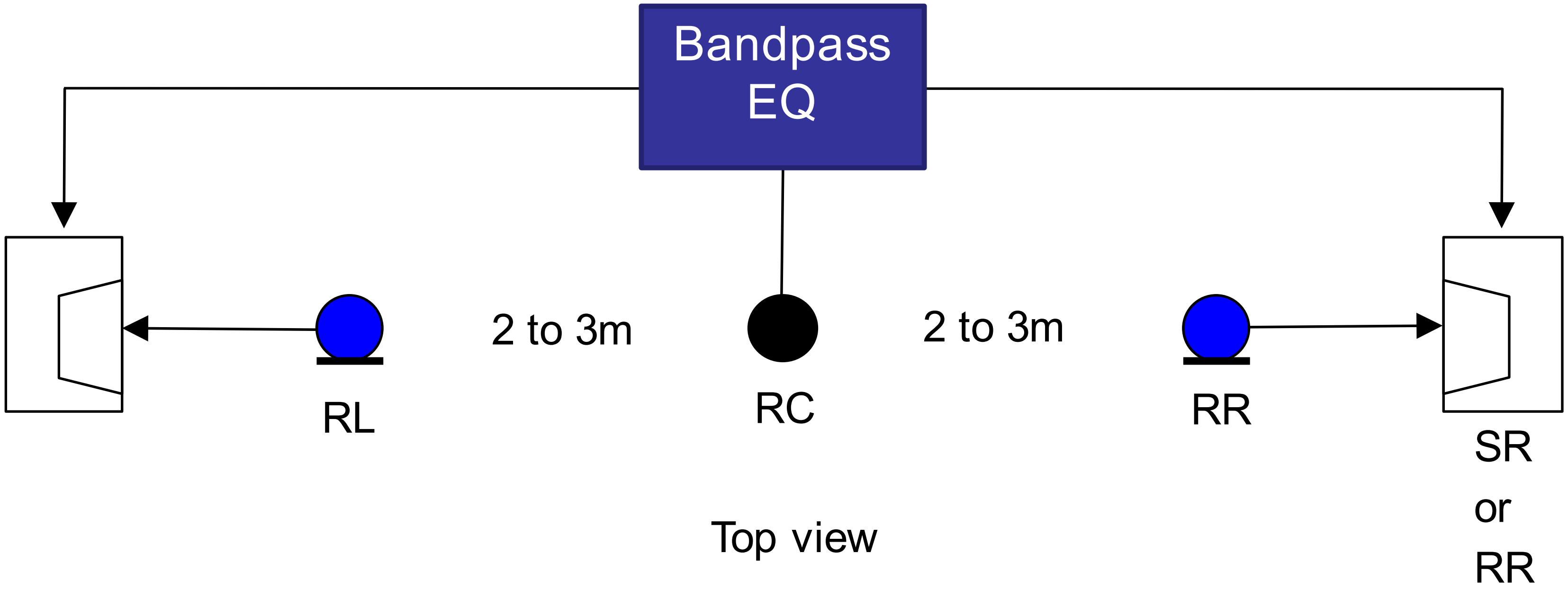
R. Wallis and H. Lee, "Vertical Stereophonic Localisation in the Presence of Interchannel Crosstalk: the Analysis of Frequency-Dependent Localisation Thresholds," JAES 2016 under review.

Demo: Organ recording

- Capturing direct sounds with height microphones can be beneficial for physically high instrument, e.g. Organ, or elevated sources, e.g. Choir on stands.

Demo: Organ recording

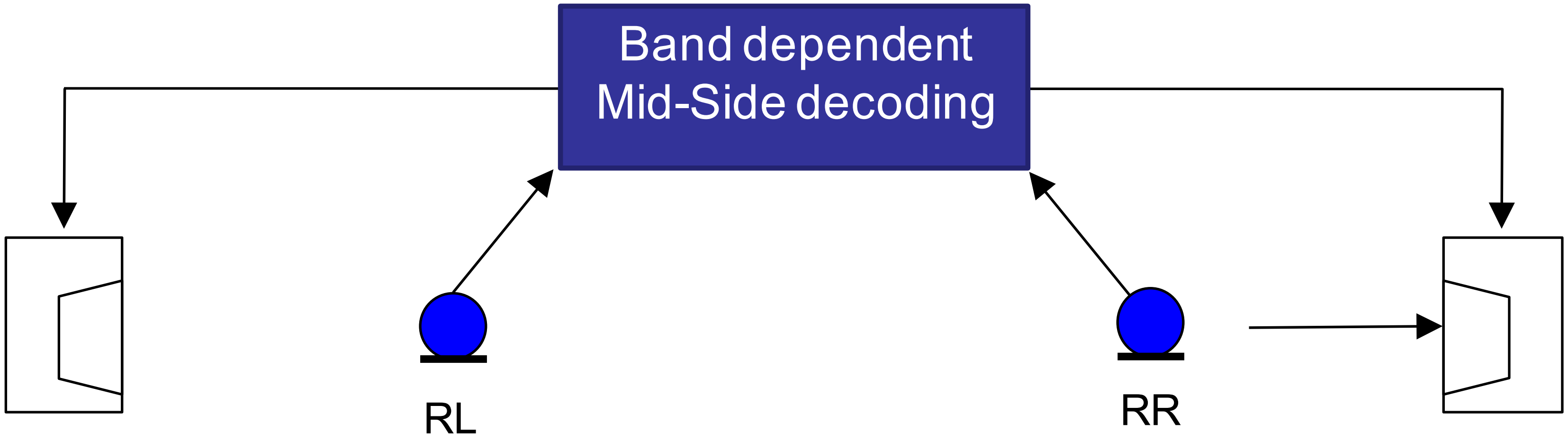
- Exploiting the phantom image elevation effect (Lee 2016)
- A rear centre ambience microphone to add "aboveness"



H. Lee, "Phantom image elevation explained," 141st AES, 2016

Demo: Organ recording

- Exploiting the phantom image elevation effect (Lee 2016)
- Band-dependent MS decoding for side or rear channels.



Top view

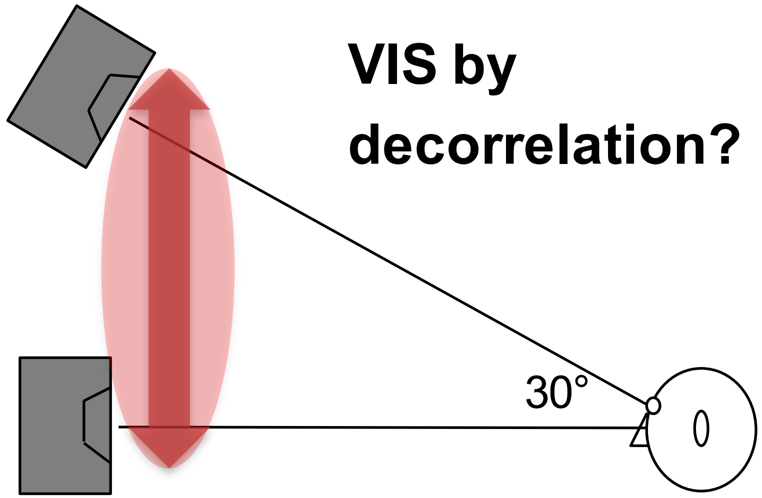
H. Lee, “Perceptually motivated 3D diffuse field upmixing,” AES SFC, 2016.
H. Lee, “Phantom image elevation explained,” 141st AES, 2016.

Vertical Interchannel Decorrelation & Vertical Microphone spacing

Vertical decorrelation

University of HUDDERSFIELD

- Vertical decorrelation on vertical image spread (VIS) (Gribben and Lee 2014, 2016)
  - The decorrelation effect on VIS is only slight.
  - Correlated source could be perceived more spread than decorrelated source in the vertical plane.



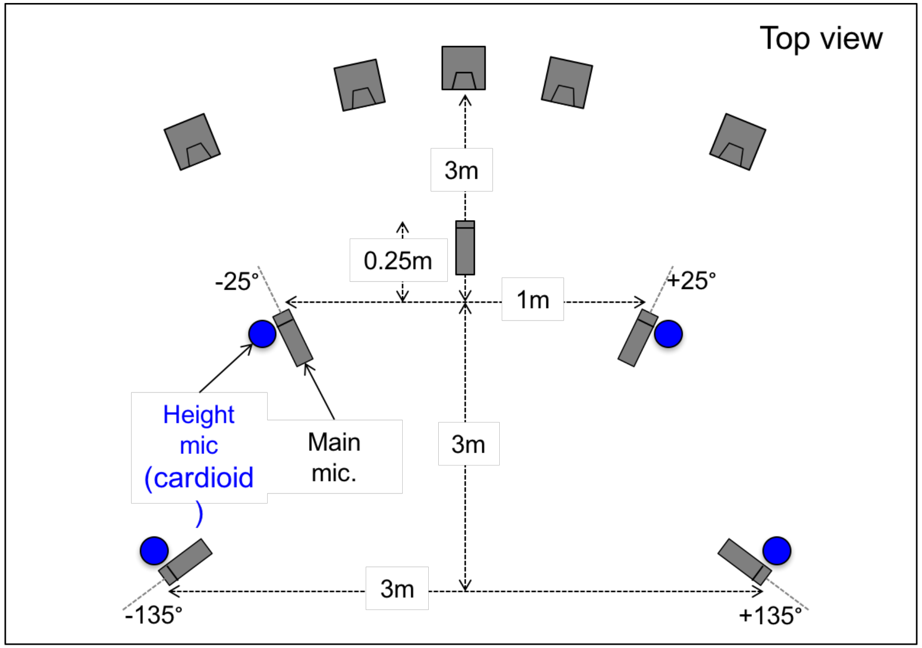
C. Gribben and H. Lee, "The Perceptual Effects of Horizontal and Vertical Interchannel Decorrelation, using the Lauridsen Decorrelator," 136th AES, 2014.

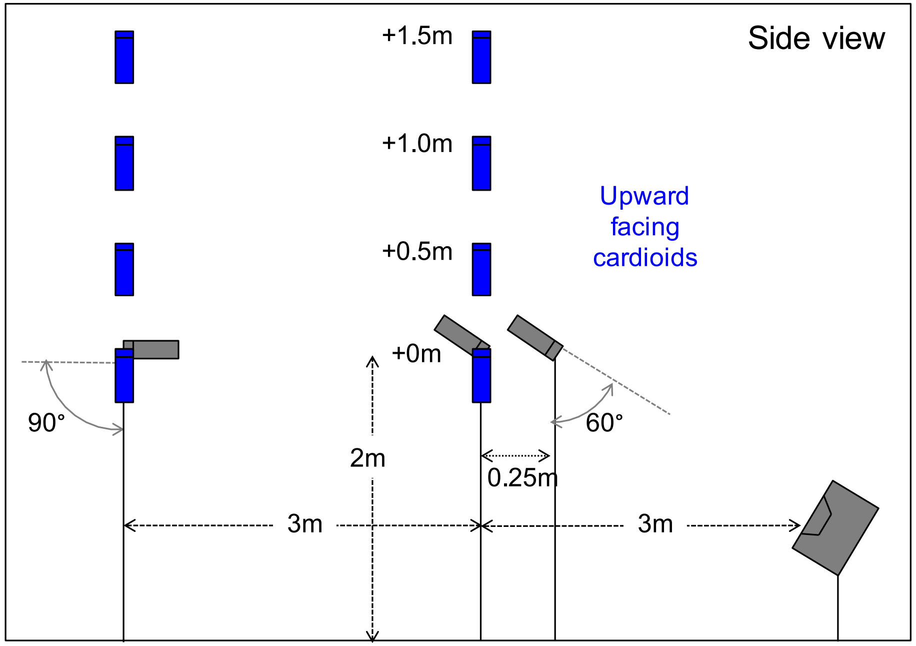
C. Gribben and H. Lee, "The Perception of Vertical Image Spread by Interchannel Decorrelation," 140th AES, 2016.

Vertical microphone spacing

University of HUDDERSFIELD

- The effect of vertical microphone spacing on spatial impression
  - NOT significant. (Lee and Gribben 2014)

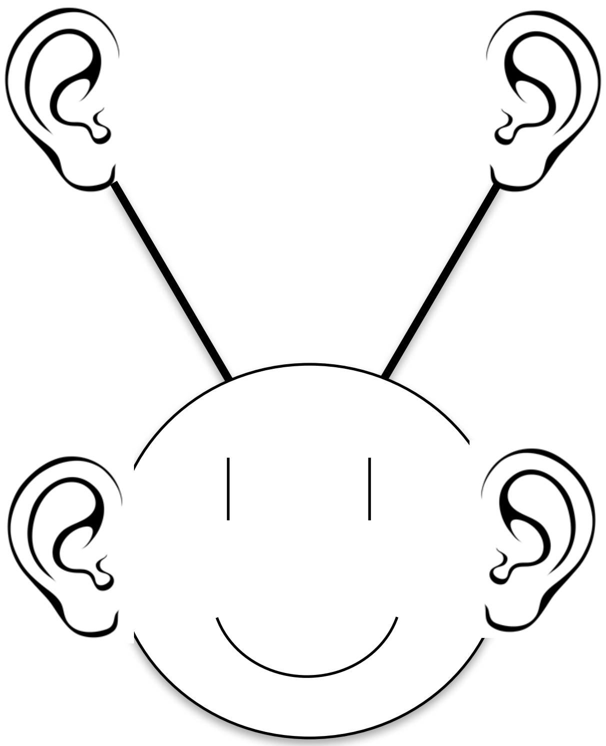




H. Lee and C. Gribben, “Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array,” JAES, 2014

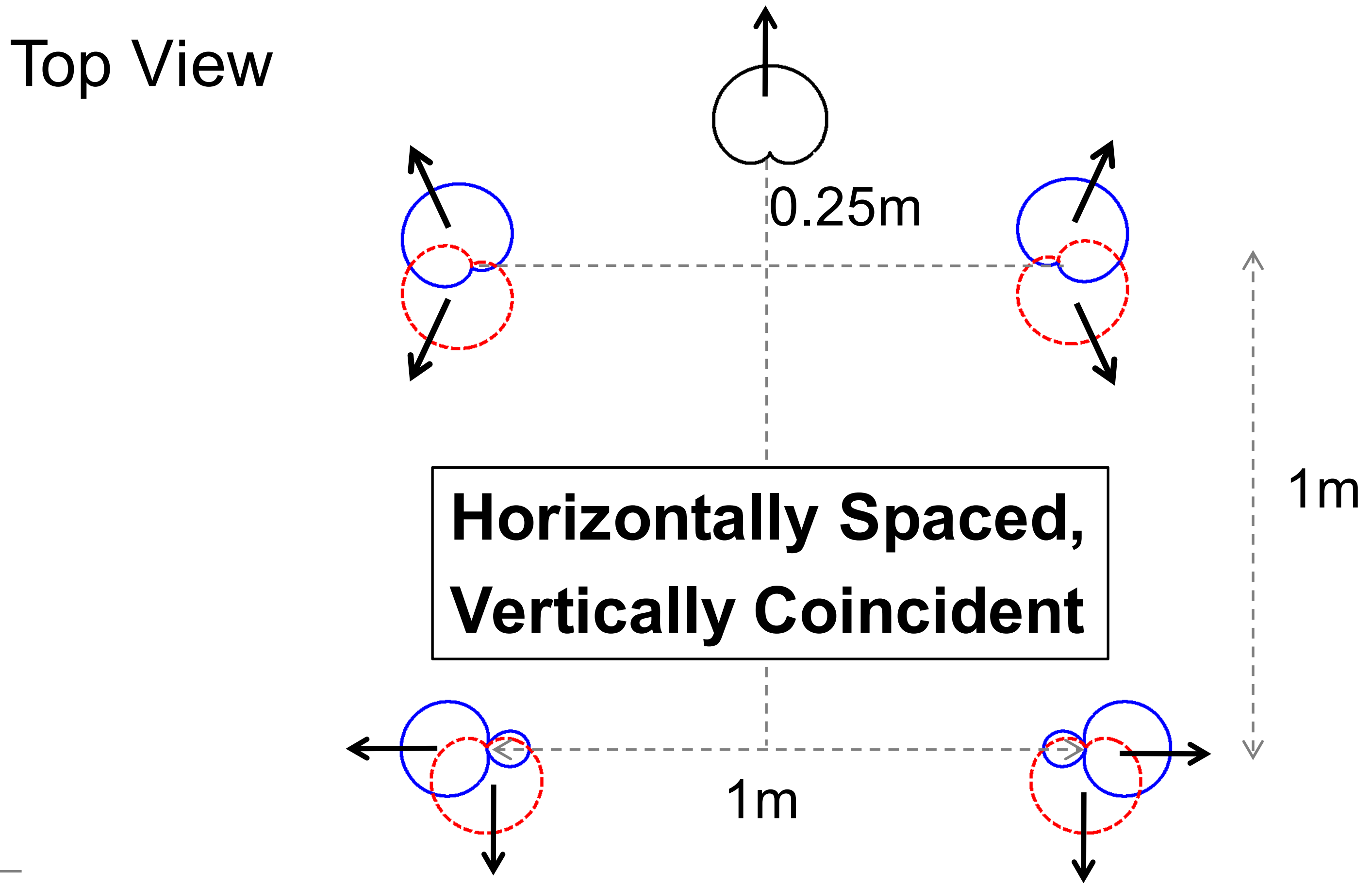
Vertical microphone spacing

- The effect of vertical microphone spacing on spatial impression
  – NOT significant. (Lee and Gribben 2014)



3D main mic array design

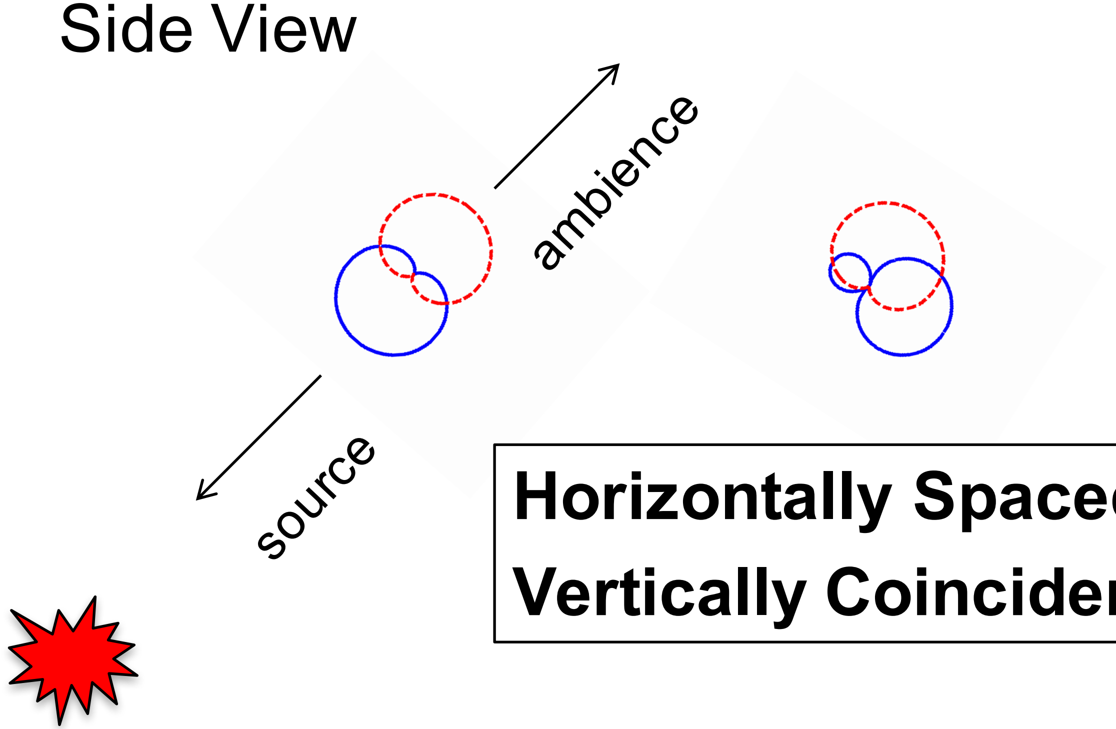
- PCMA - Perspective Control Microphone Array (Lee 2012)



H. Lee, “Subjective Evaluations of Perspective Control Microphone Array (PCMA),” 132nd AES, 2012.

3D main mic array design

- PCMA - Perspective Control Microphone Array (Lee 2012)



Horizontally Spaced, Vertically Coincident

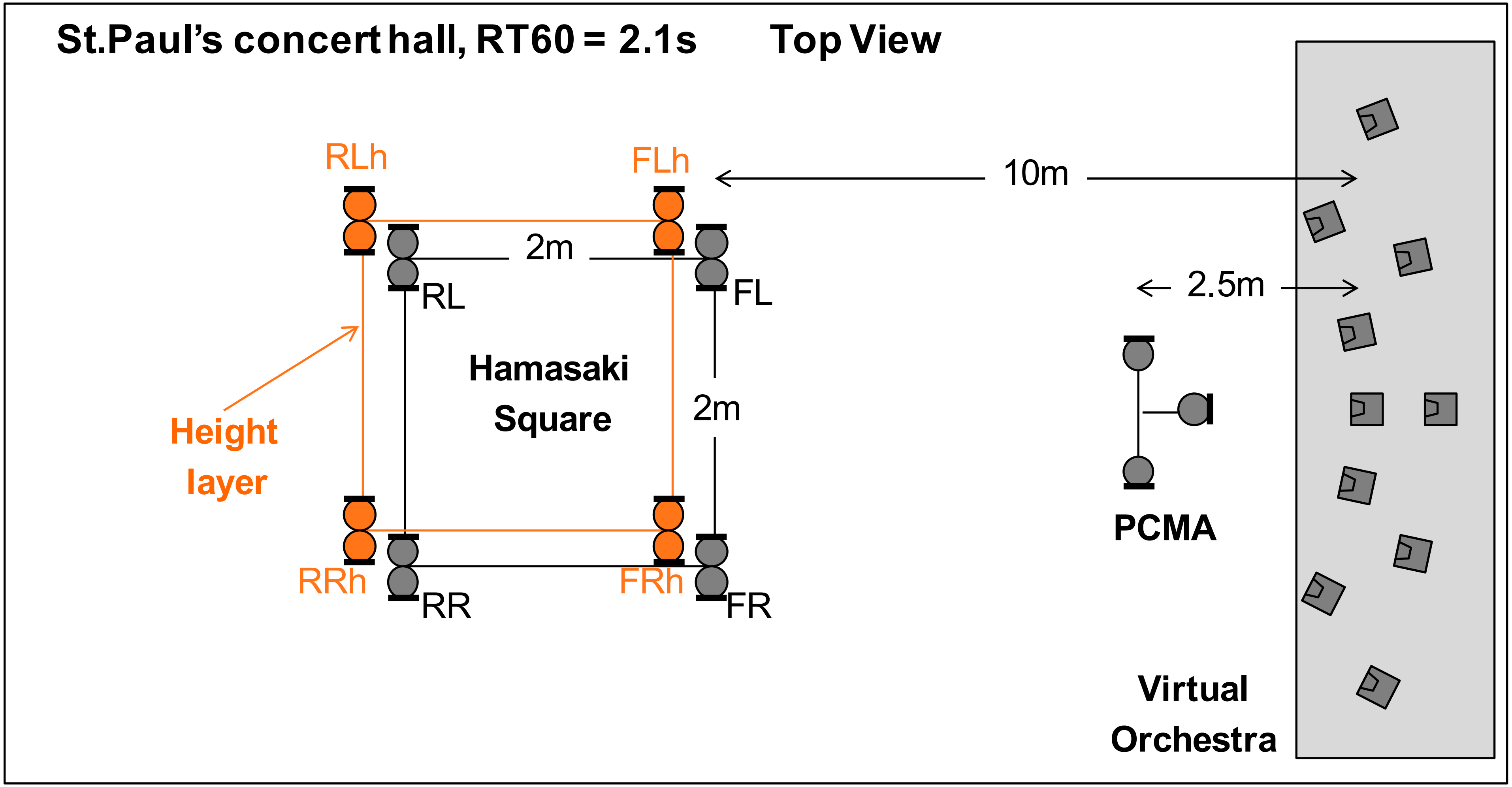
H. Lee, “Subjective Evaluations of Perspective Control Microphone Array (PCMA),” 132nd AES, 2012.

Demo: Vertical mic spacing effect

- Ambience captured by "Double Layered Hamasaki Square"
- Diffused field ambience recorded in St.Paul's hall, Huddersfield.

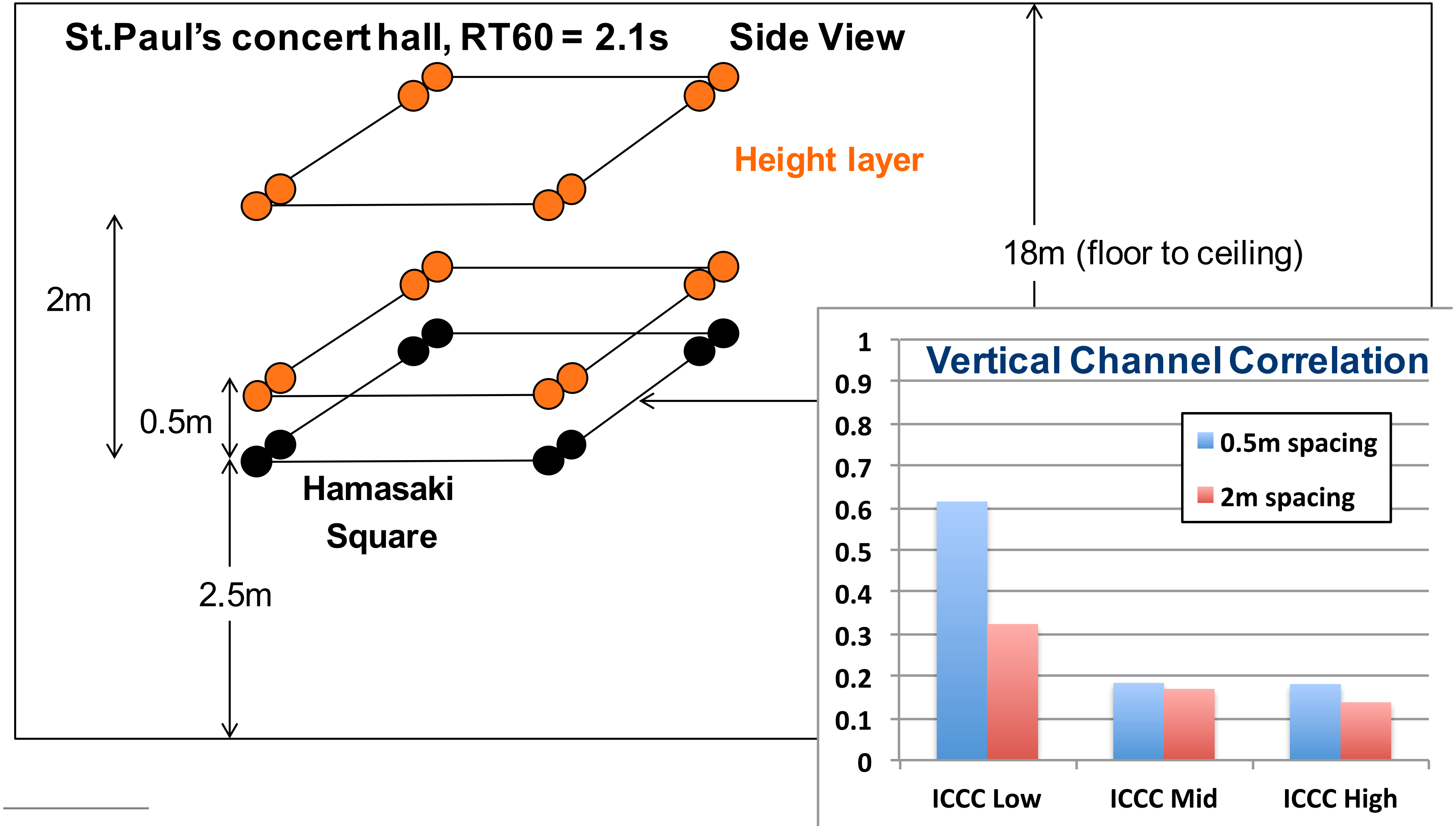
Demo: Vertical mic spacing effect

University of HUDDERSFIELD



THE QUEEN'S AWARDS FOR ENTERPRISE

Demo: Vertical mic spacing effect



Vertical Enhancement for 3D Recording

Front height vs. Rear height

- Front to Back Ratio for LEV measurement (Morimoto and Iida 1998)
  - The more ambience from the back, the more enveloping.

- Front height contributes to Front Depth/Distance.

- Rear height is for LEV

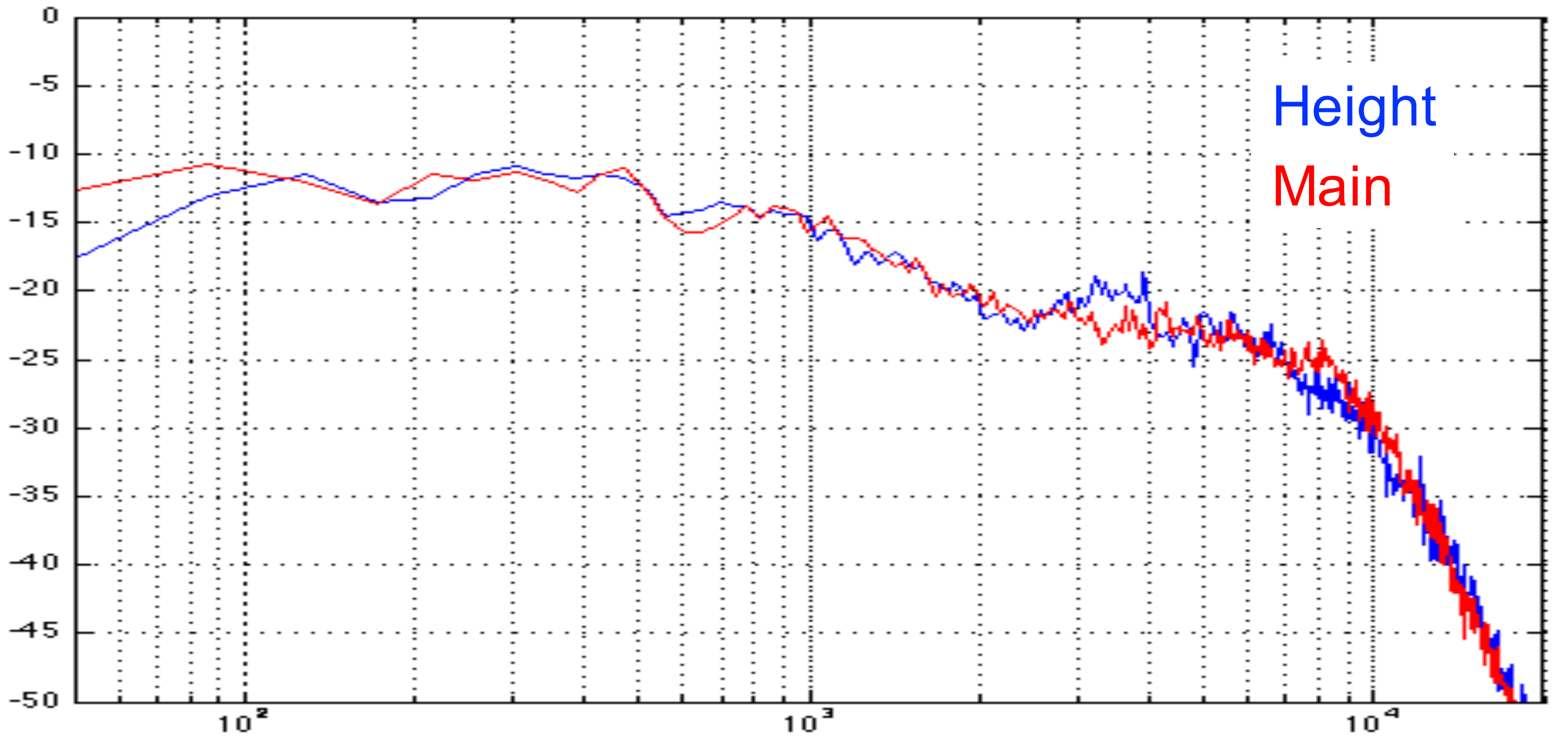
Demo: Front height vs. Rear height

University of HUDDERSFIELD

THE QUEEN'S AWARDS FOR ENTERPRISE

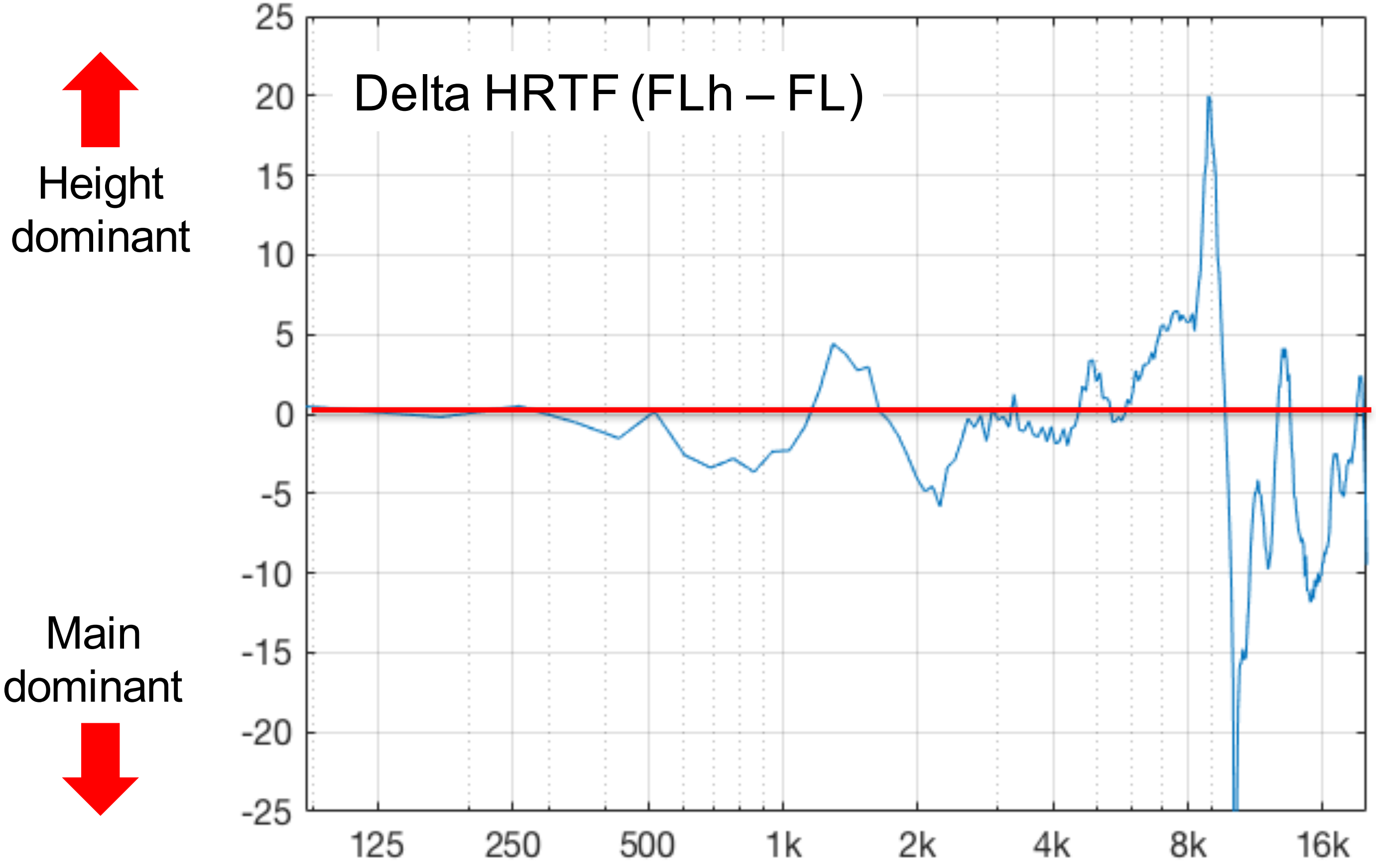
Typical reverb spectrum for music

- Reverberation spectrum
  - High frequency rolled off



Main vs. Height in HRTF

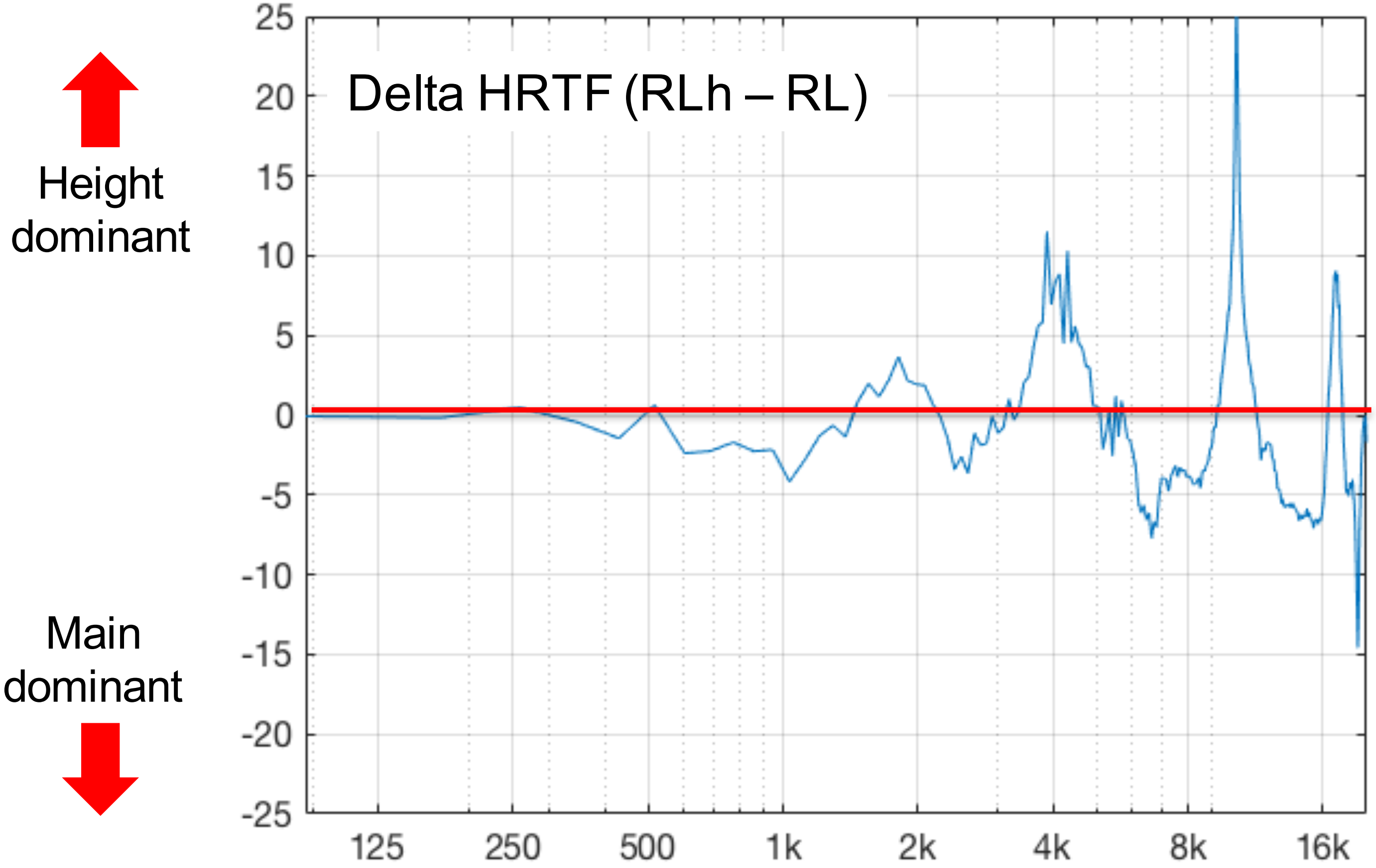
- HRTF difference between Front Left and Front Left Height (Lee 2016 AES SFC)



H. Lee, "Perceptually motivated 3D diffuse field upmixing," AES SFC, 2016.

Main vs. Height in HRTF

- HRTF difference between Rear Left and Rear Left Height (Lee 2016 AES SFC)



H. Lee, “Perceptually motivated 3D diffuse field upmixing,” AES SFC, 2016.

VIS Enhancement

University of HUDDERSFIELD

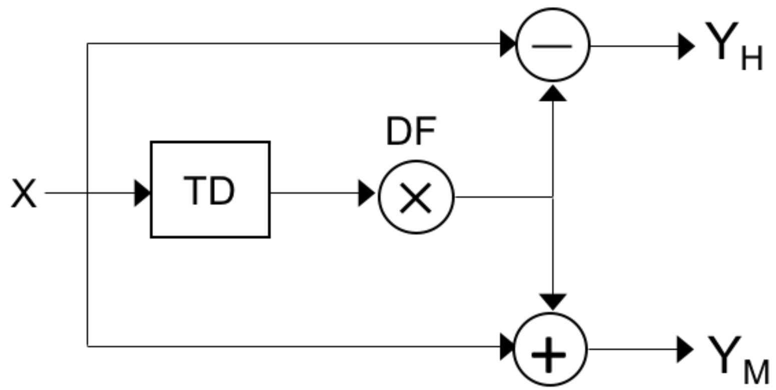
- Complementary perceptual equalisation (Lee 2016: AES SFC)
  - For the height channel, emphasize frequencies that are more dominant in the height speaker HRTF, while deemphasizing those that more dominant in the main speaker HRTF.
  - The same process for the main channel.
  - VIS and spectral clarity enhancement
  - The SPL and spectrum of the resulting signal at the listening position does not change.

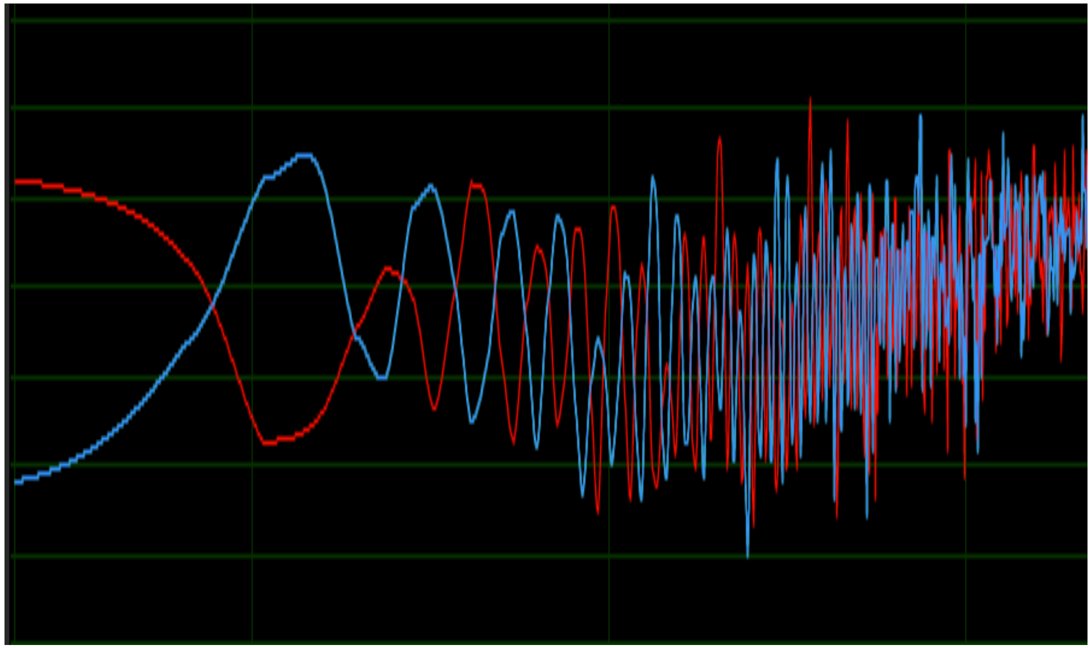
H. Lee, “Perceptually motivated 3D diffuse field upmixing,” AES SFC, 2016.

Vertical Upmixing

Conventional methods

- Interchannel decorrelation
  - All pass filters
  - Complementary Comb Filter (Lauridsen decorrelator)





Perceptual Band Allocation (PBA)

University of HUDDERSFIELD

- A novel vertical upmixing method that exploits the pitch-height effect (Lee 2015, 2016)

Pitch-Height Database from Subjective Measurements

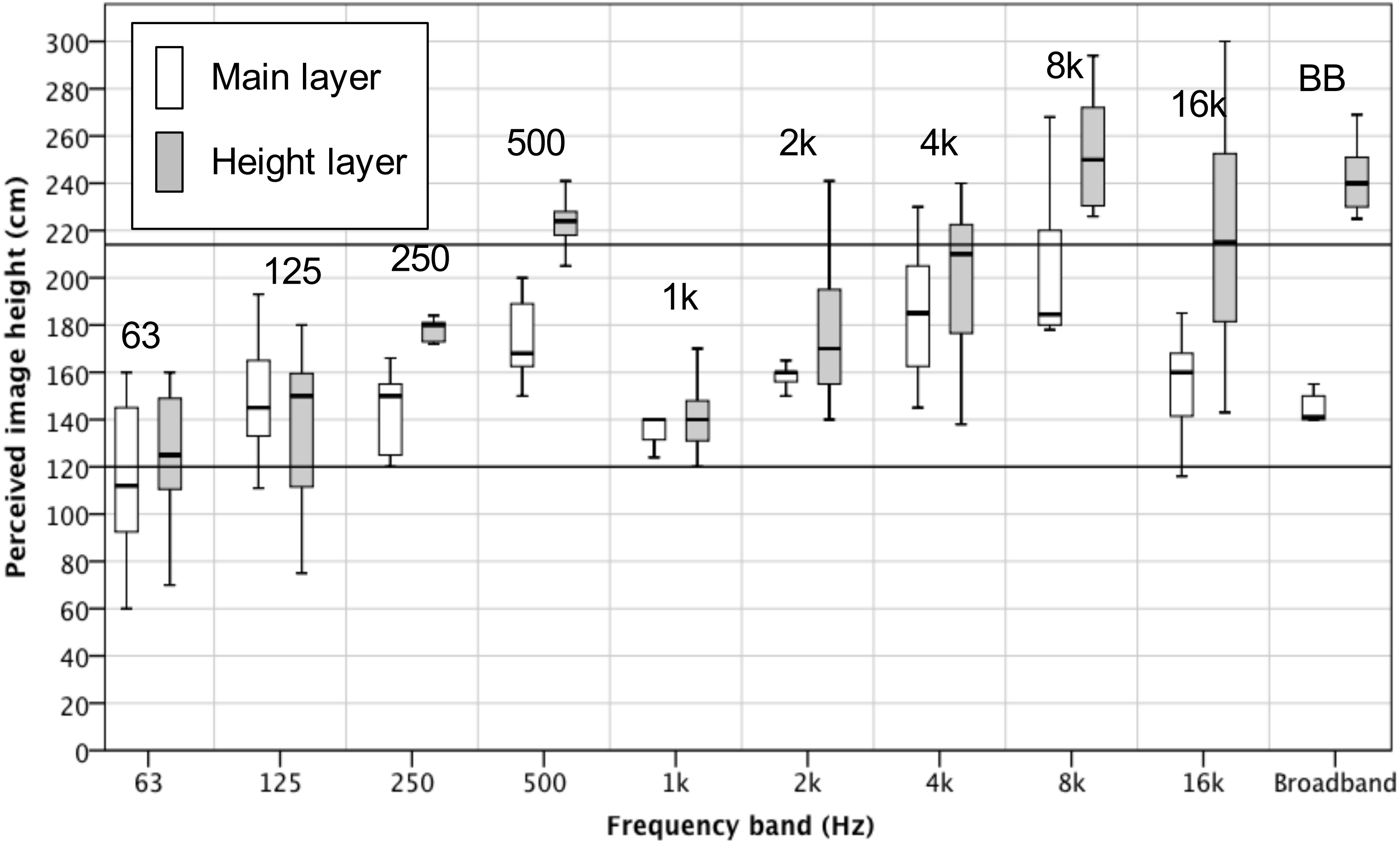
Spectral Band Decomposition & Grouping → Band to Loudspeaker Mapping → Vertical image spread rendering

- Octave bands
- One to one mapping
- No frequency overlapping

H. Lee, "Perceptual Band Allocation (PBA) for the rendering of vertical image spread," 138th AES, 2015, JAES, 2016 under review.

Perceptual Band Allocation (PBA)

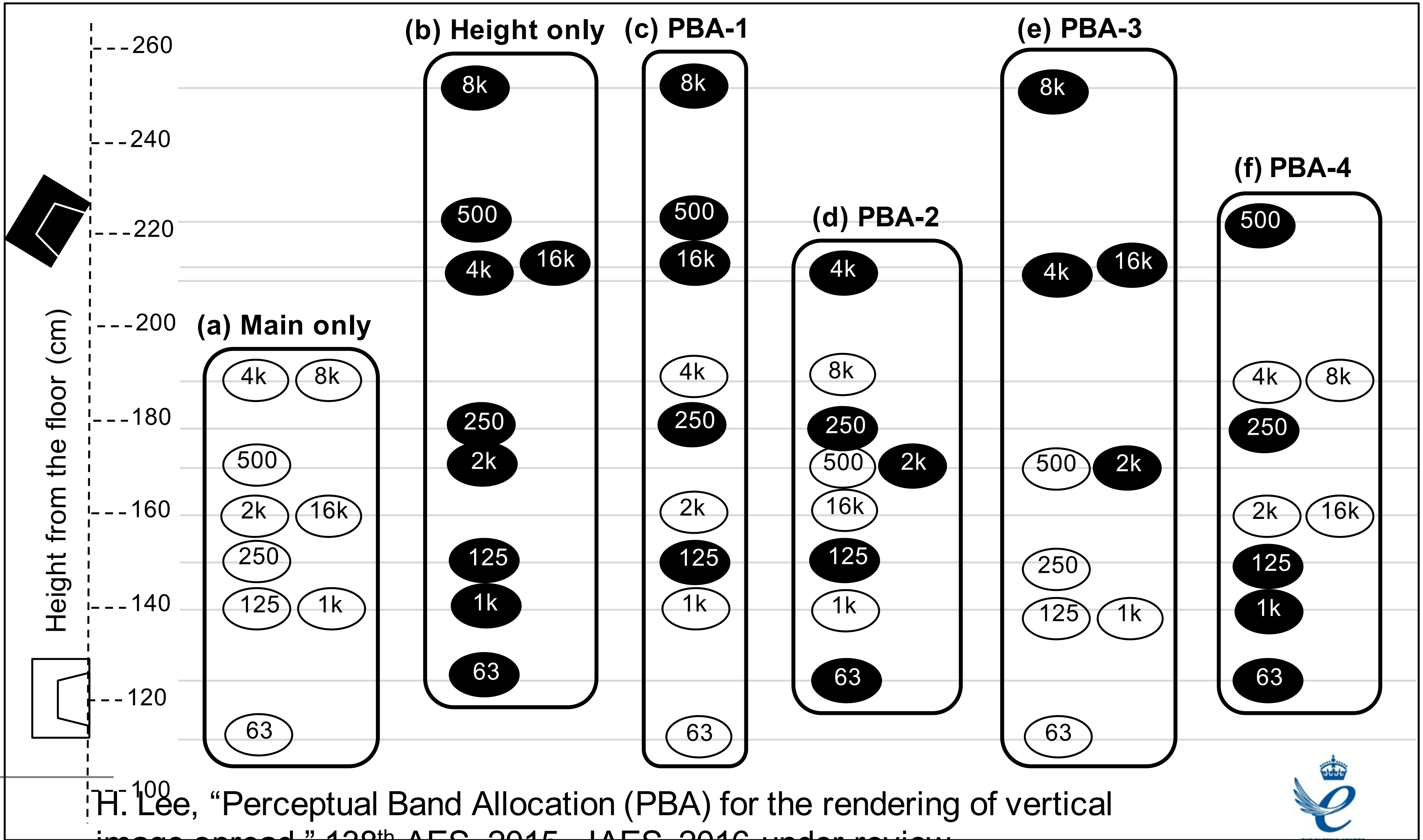
University of HUDDERSFIELD



H. Lee, "Perceptual Band Allocation (PBA) for the rendering of vertical image spread," 138th AES, 2015, JAES, 2016 under review.

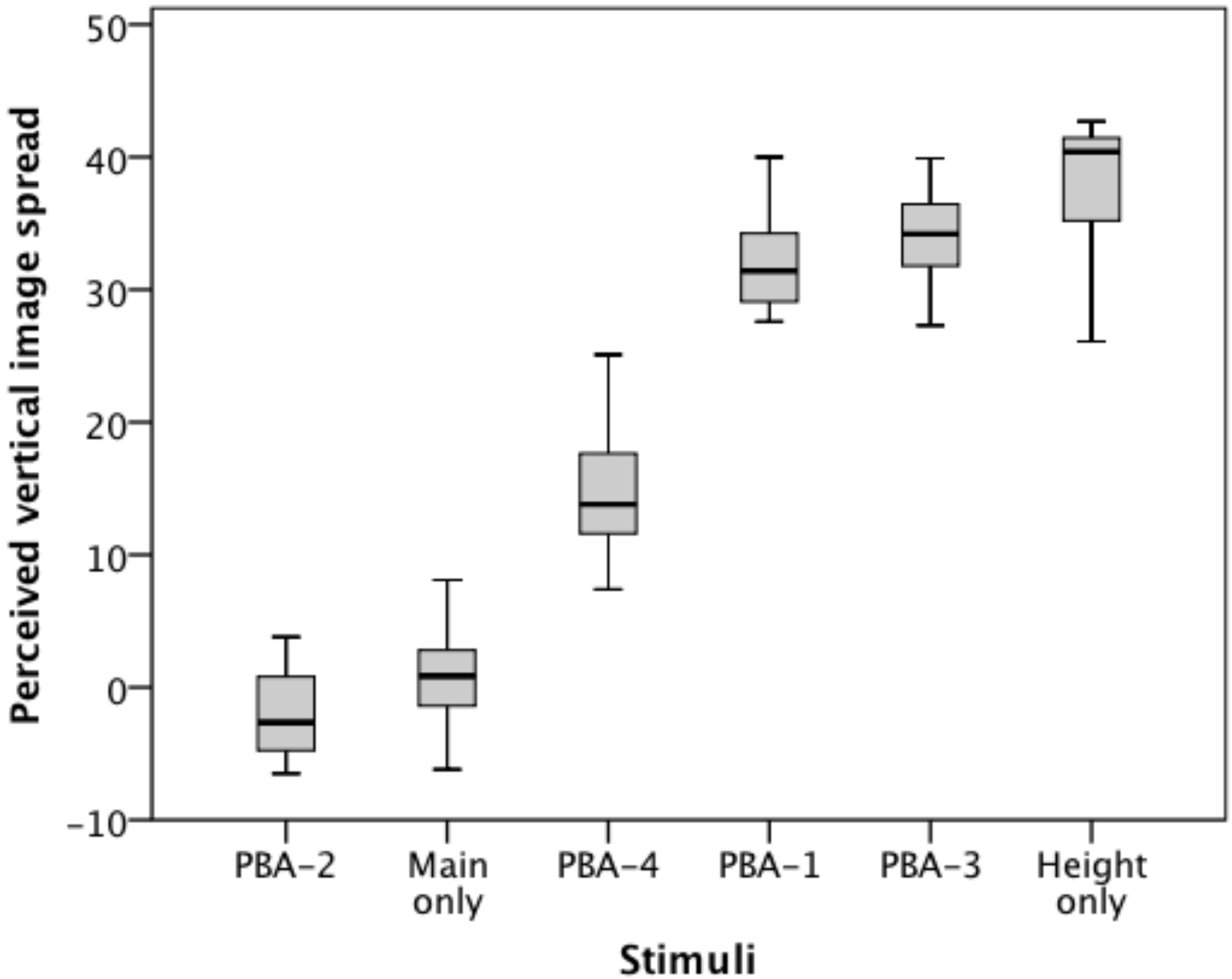
Perceptual Band Allocation (PBA)

University of Huddersfield



H. Lee, "Perceptual Band Allocation (PBA) for the rendering of vertical image spread," 138th AES, 2015, JAES, 2016 under review.

Perceptual Band Allocation (PBA)



H. Lee, "Perceptual Band Allocation (PBA) for the rendering of vertical image spread," 138th AES, 2015, JAES, 2016 under review.

Demo: PBA upmixing

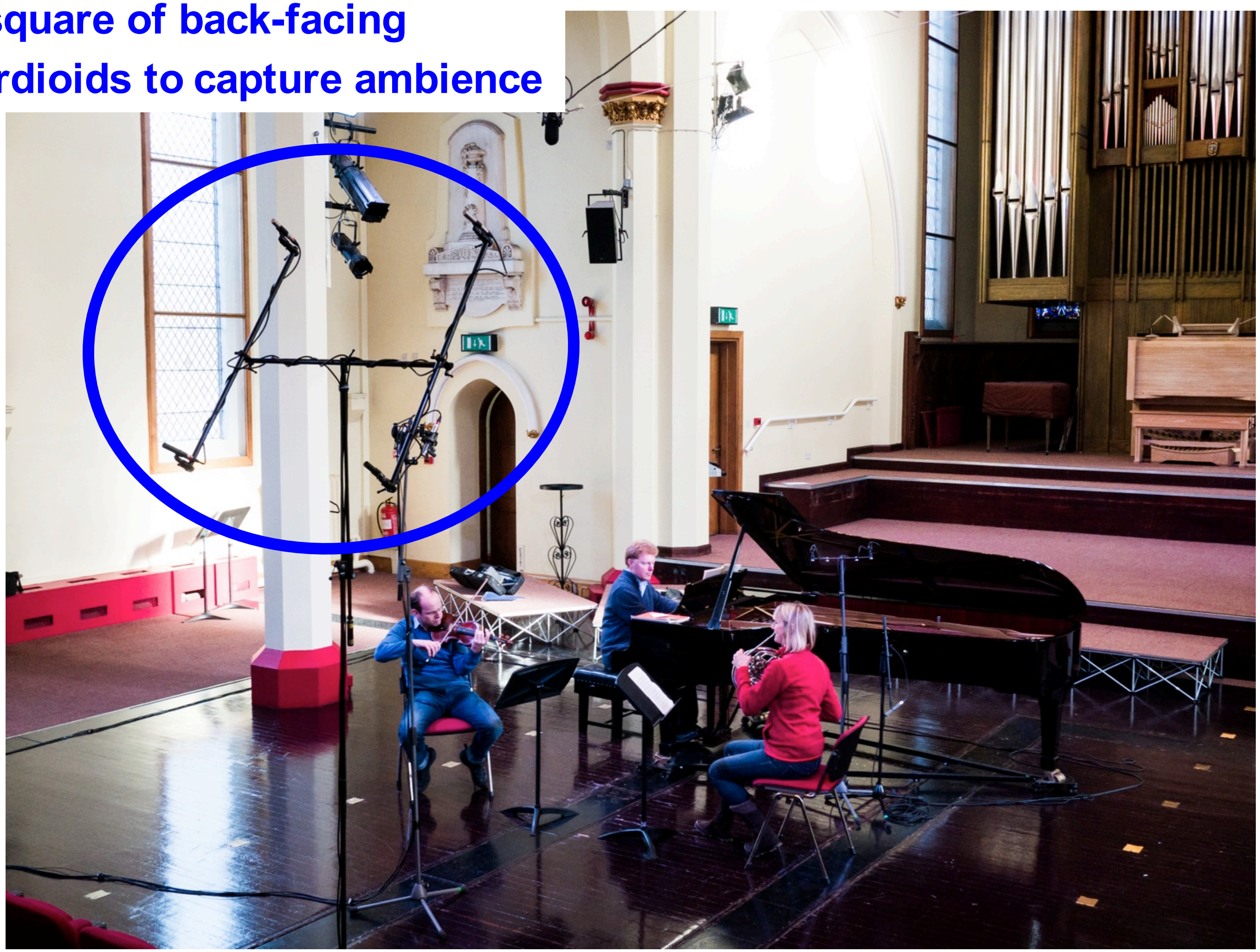
- Recording with 4 ambience microphones

Demo: PBA upmixing

University of HUDDERSFIELD

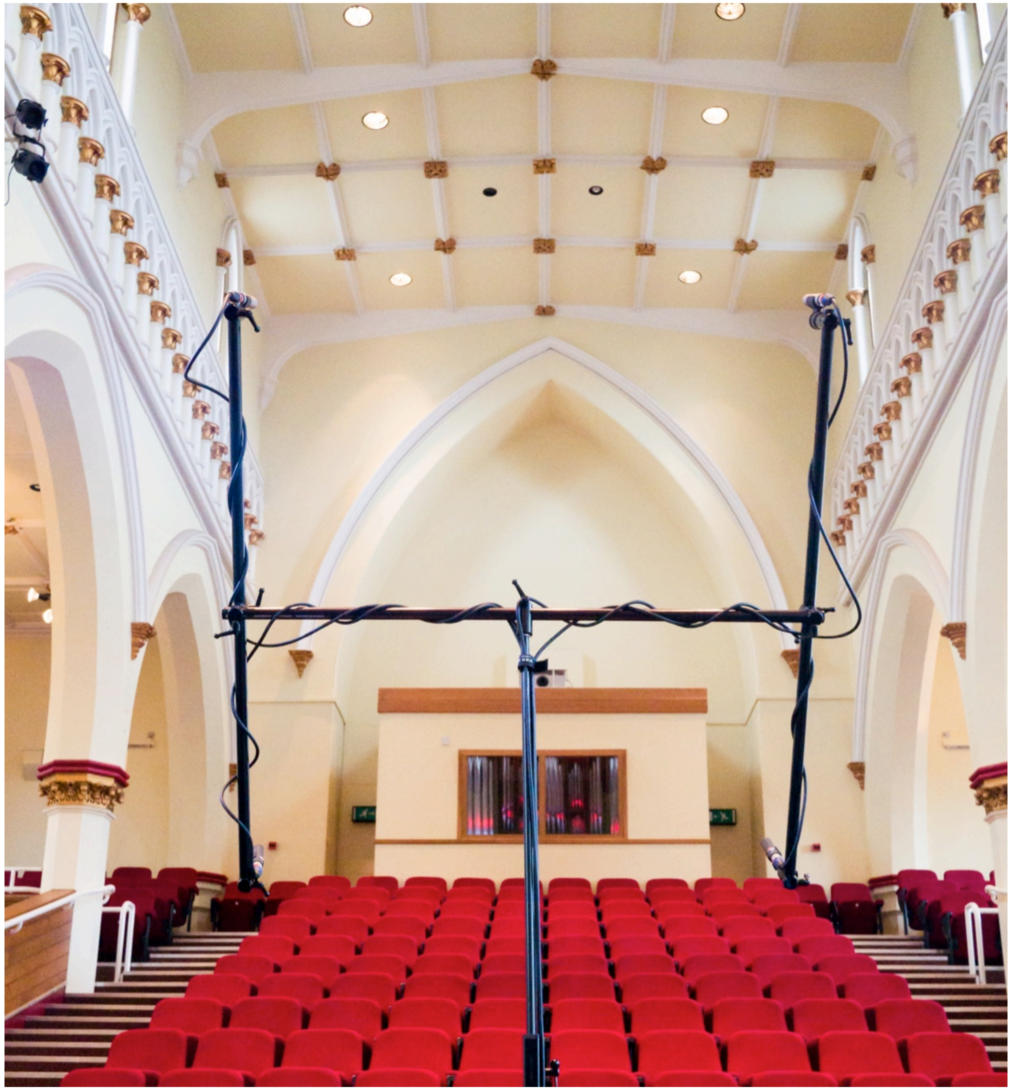
- Recording with 4 ambience microphones

A square of back-facing Cardioids to capture ambience



Demo: PBA upmixing

- Recording with 4 ambience microphones



Demo: PBA upmixing

University of HUDDERSFIELD

- PBA scheme used

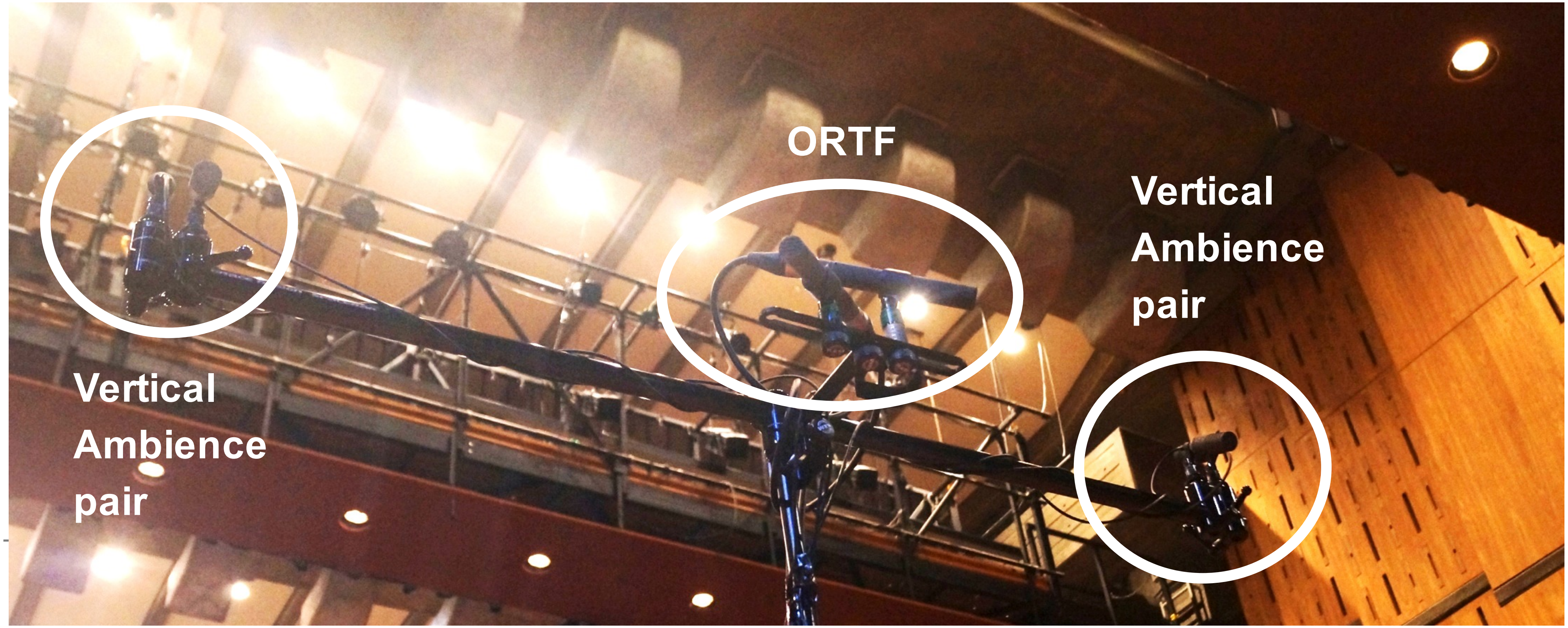
Channels	Layer	Allocated octave-bands (centre frequency)
Front	Main	63 Hz, 1 kHz, 2 kHz, 4 kHz
	Height	125 Hz, 250 Hz, 500 Hz, 8 kHz, 16 kHz
Rear	Main	63 Hz, 125 Hz, 500 Hz, 1 kHz, 2kHz
	Height	250.Hz, 4 kHz, 8 kHz, 16 kHz

THE QUEEN'S AWARDS FOR ENTERPRISE

Demo: PBA + VOS upmixing

University of HUDDERSFIELD

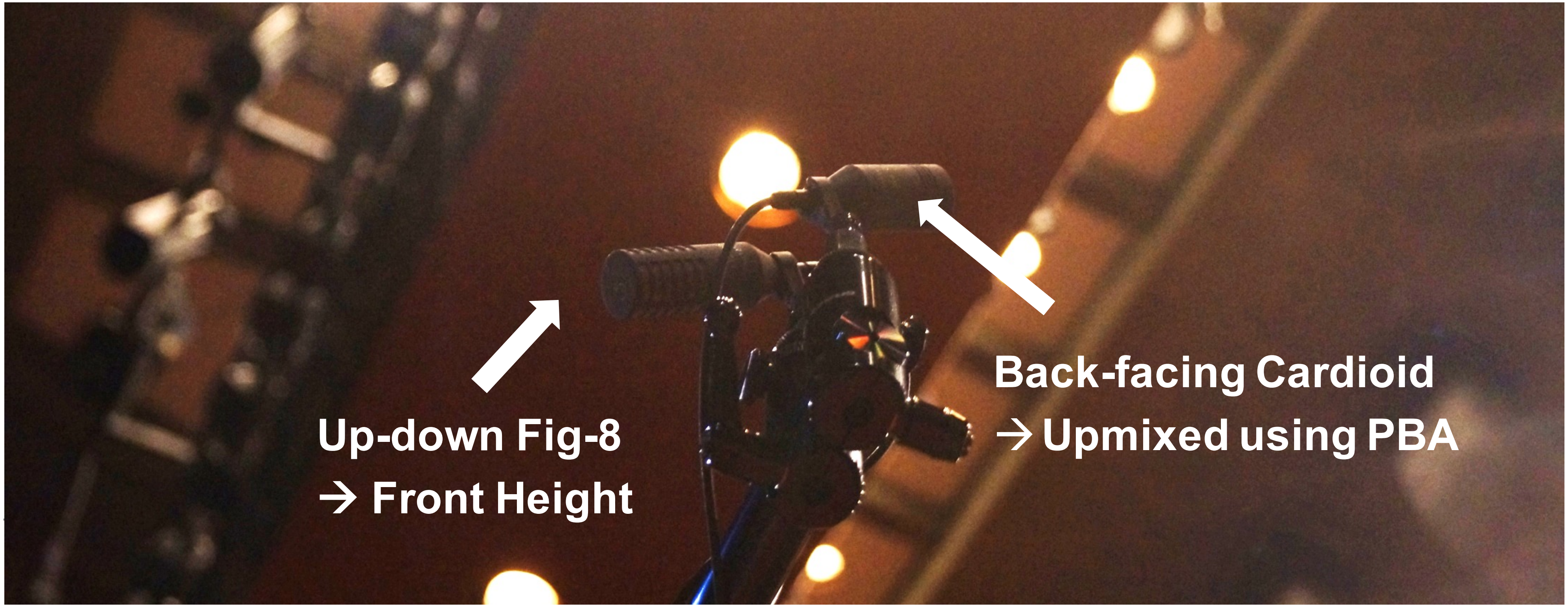
- 2+2+2 Recording
  - Recorded at Queen Elizabeth Hall, London
  - Live recording limitation: the size of mic array



Demo: PBA + VOS upmixing

University of HUDDERSFIELD

- Rear channel signals were vertically upmixed using a 2-band PBA.
- The 3rd reverb signal was equalised and routed to both side channels for the VOS (virtual overhead speaker) effect.



Download links for useful tools by APL

- IAR: http://eprints.hud.ac.uk/25547
- HULTI-GEN: http://eprints.hud.ac.uk/24809
- HAART: http://eprints.hud.ac.uk/24579

Please contact us for more information:

Hyunkook Lee

h.lee@hud.ac.uk