Perceptual Signal Processing for 3D Sound Recording

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

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

Changes in ICLD, ICTD or ICCC

Changes in ILD, ITD, IACC
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
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.
Pitch-Height Effect for “Real” Source

- For **band-passed noise** signals, high frequency components (above 7kHz) are essential for accurate vertical localisation. (Roffler and Butler 1968b)

From Roffler and Butler (1968b) - Low passed & high passed noise

- Low passed noise < 2kHz
- High passed noise > 2k
- Broadband
- 4.8kHz tone
- 600Hz tone

![Graph showing perceived image height vs actual loudspeaker height with different noise conditions.](image-url)
• Pitch height effect for octave band pink noise
  – after Cabrera and Tiley (2003); median plane results
Directional bands

- Blauert (1968): physical mapping between frequency bands and their perceived positions in the median plane.
Pitch-Height Effect for “Phantom” Source

- Pitch-height effect for horizontal \textit{phantom} images from main and height layers (Lee 2015)

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 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)
  - But only with white noise (650Hz – 4.5kHz)

phantom centre image
Phantom Image Elevation

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

---

Phantom Image Elevation

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

\[
\text{Perceived elevation angle} = \frac{\text{Loudspeaker base angle}}{2}
\]

H. Lee, “Investigation on the Phantom Image Elevation Effect,” 139\textsuperscript{th} 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”)

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

Phantom Image Elevation

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

Theoretical explanations

- Spectral energy distribution of ear signal

(a) 60 degrees

(b) 120 degrees

(c) 180 degrees

(d) 240 degrees

(e) 300 degrees

(f) 360 degrees

Delta HRTF (60° - 0°)

(120° - 0°)

(180° - 0°)

(240° - 0°)

(300° - 0°)

(360° - 0°)

- 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)
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)

---

Theoretical explanations

• 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).

---

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.

---

Theoretical explanations

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

![Frequency band graph]

Elevation Strength

Frequency band (Hz)

125Hz 250Hz 500Hz 1k 2k 4k 8k 16k

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

- 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

---

Vertical interchannel crosstalk

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

Vertical interchannel crosstalk

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

Vertical interchannel crosstalk

- **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)

Vertical interchannel crosstalk

- **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)

Vertical interchannel crosstalk

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

• At least **10dB of direct sound attenuation** is required of the height microphone to make the vertical crosstalk completely inaudible (Lee 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=2.1 sec) in Huddersfield, UK
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

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

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”

Top view

Bandpass
EQ

RL  2 to 3m  RC  2 to 3m  RR

or
SR
RR

H. Lee, “Phantom image elevation explained,” 141st AES, 2016
• Exploiting the phantom image elevation effect (Lee 2016)
• Band-dependent MS decoding for side or rear channels.

Vertical Interchannel Decorrelation & Vertical Microphone spacing
Vertical decorrelation

- 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.


Vertical microphone spacing

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

Top View

Horizontally Spaced, Vertically Coincident

3D main mic array design

- PCMA - Perspective Control Microphone Array (Lee 2012)

Side View

Horizontally Spaced, Vertically Coincident

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

St. Paul’s concert hall, RT60 = 2.1s

Top View

Height layer

Hamasaki Square

PCMA

Virtual Orchestra
Demo: Vertical mic spacing effect

St. Paul's concert hall, RT60 = 2.1s

Side View

Height layer

18m (floor to ceiling)

Vertical Channel Correlation

0.5m spacing

2m spacing

ICCC Low

ICCC Mid

ICCC High
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
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)

Main vs. Height in HRTF

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

Delta HRTF (RLh – RL)

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.

Vertical Upmixing
Conventional methods

- Interchannel decorrelation
  - All pass filters
  - Complementary Comb Filter (Lauridsen decorrelator)
**Perceptual Band Allocation (PBA)**

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

**Pitch-Height Database from Subjective Measurements**

**Spectral Band Decomposition & Grouping**
- Octave bands

**Band to Loudspeaker Mapping**
- One to one mapping

**Vertical image spread rendering**
- No frequency overlapping

---

Demo: PBA upmixing

- Recording with 4 ambience microphones
Demo: PBA upmixing

- 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

- PBA scheme used

<table>
<thead>
<tr>
<th>Channels</th>
<th>Layer</th>
<th>Allocated octave-bands (centre frequency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Front</td>
<td>Main</td>
<td>63 Hz, 1 kHz, 2 kHz, 4 kHz</td>
</tr>
<tr>
<td></td>
<td>Height</td>
<td>125 Hz, 250 Hz, 500 Hz, 8 kHz, 16 kHz</td>
</tr>
<tr>
<td>Rear</td>
<td>Main</td>
<td>63 Hz, 125 Hz, 500 Hz, 1 kHz, 2 kHz</td>
</tr>
<tr>
<td></td>
<td>Height</td>
<td>250 Hz, 4 kHz, 8 kHz, 16 kHz</td>
</tr>
</tbody>
</table>
Demo: PBA + VOS upmixing

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

- 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