

Dimensions of Immersive Audio Experience

Dr Hyunkook Lee
Applied Psychoacoustics Lab (APL)
University of Huddersfield, UK

What do you mean by “Immersive” audio?

Immersion



Immersive?



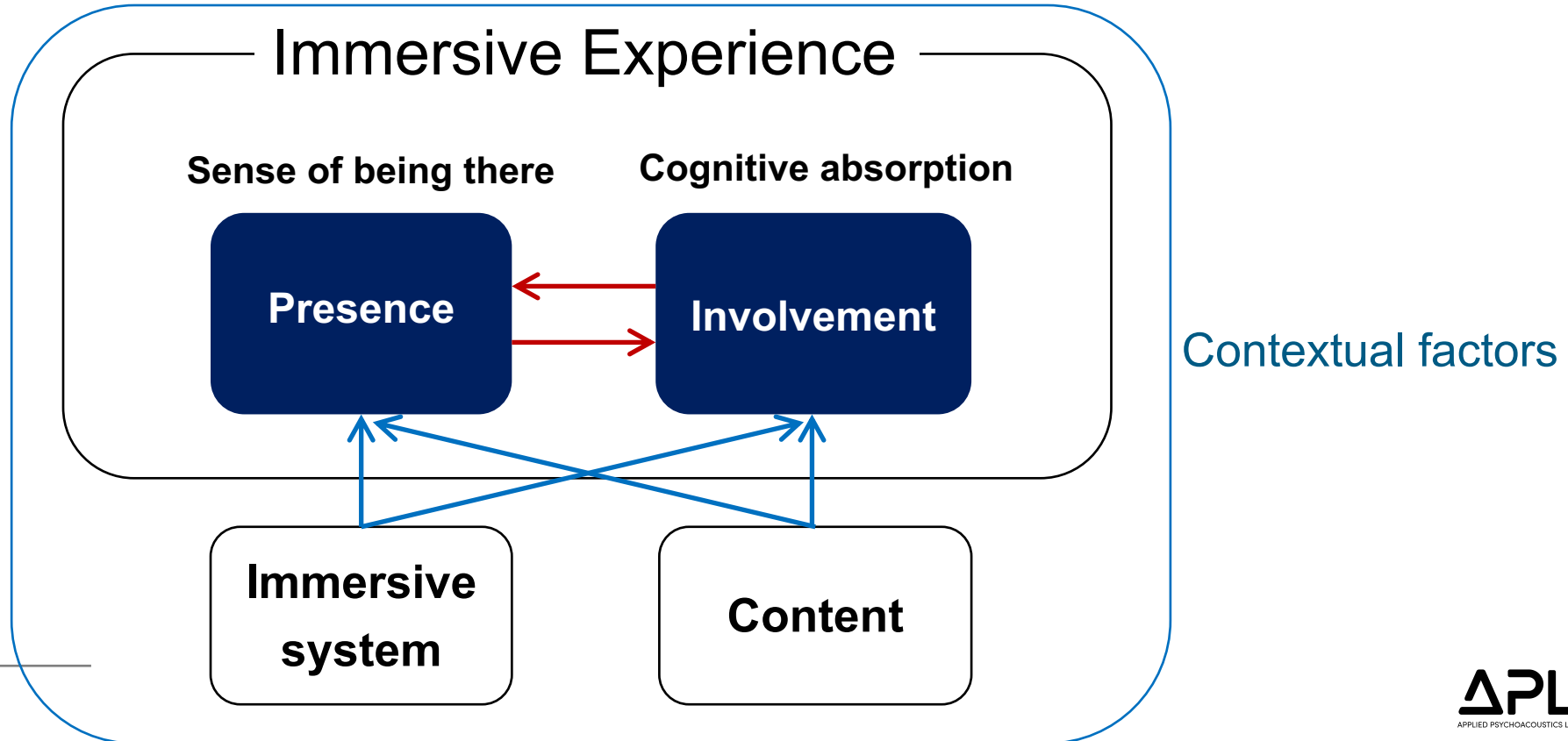
Immersive?

- Dictionary definitions (Oxford Learner's)
 - “the act of putting somebody/something into a liquid, especially so that they or it are **completely covered**; the state of being covered by a liquid.”
 - Sensory simulation (Sense of being there, i.e. Presence)
 - “the state of being **completely involved** in something.”
 - Cognitive absorption

Immersive
System

Immersive
Experience

Conceptual Model of Immersive Experience

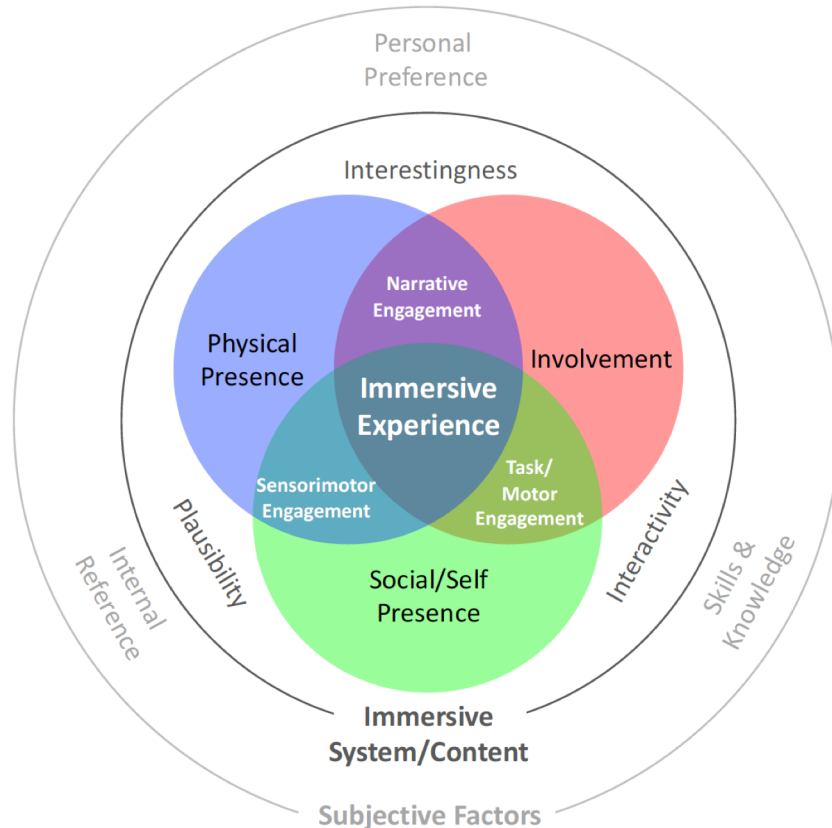


Hypothesis: Overall Level of Immersive Experience is a weighted sum of Presence and Involvement.

$$\text{OLIE} = a \times \text{Presence} + b \times \text{Involvement}$$

With a and b depending on the type of content and contextual factors

Conceptual Model of Immersive Experience



Lee (2020) “Conceptual model of immersive experience in extended reality” <https://psyarxiv.com/sefkh/>

Immersive system



University of
HUDDERSFIELD



Immersive system

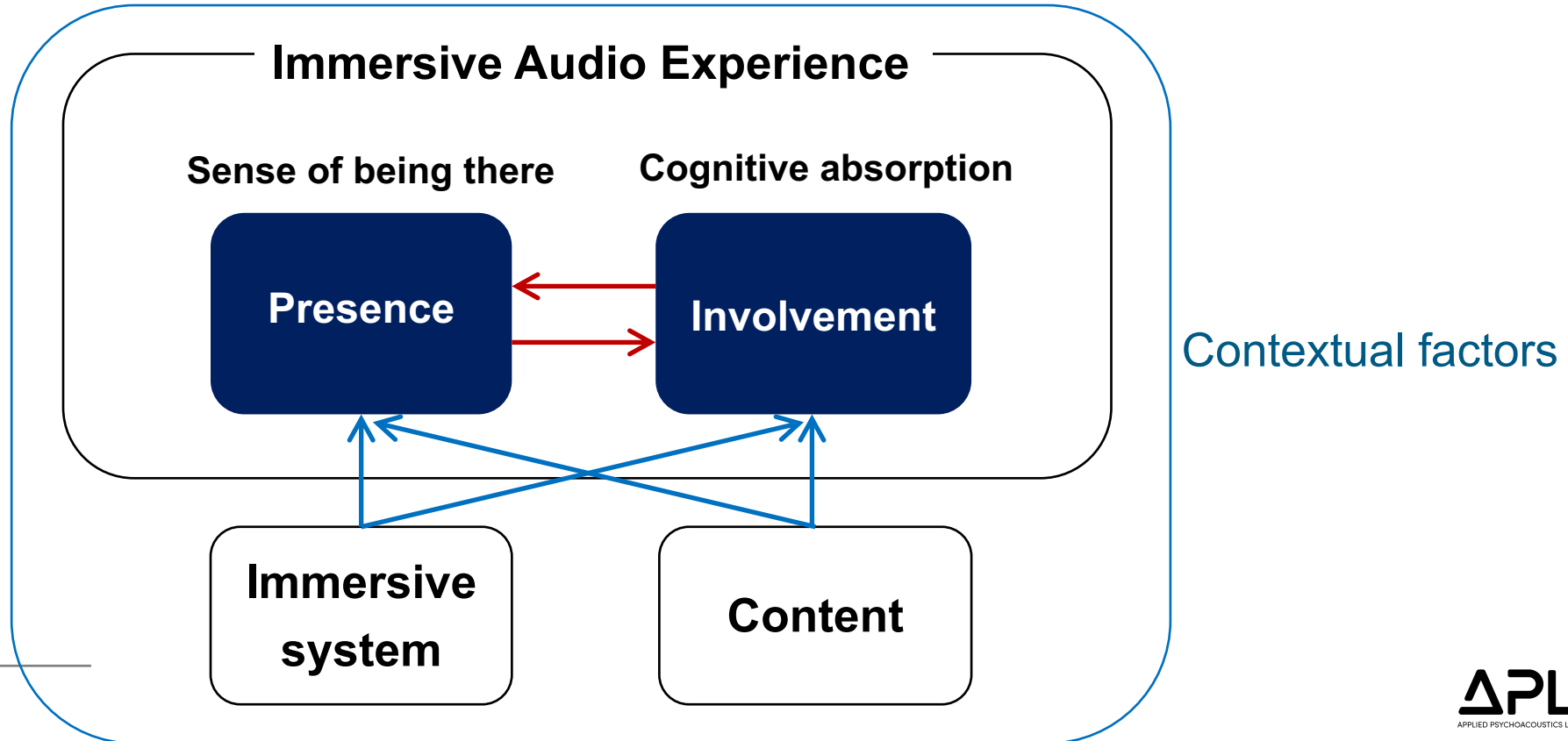


University of
HUDDERSFIELD



Immersive **Audio** Experience

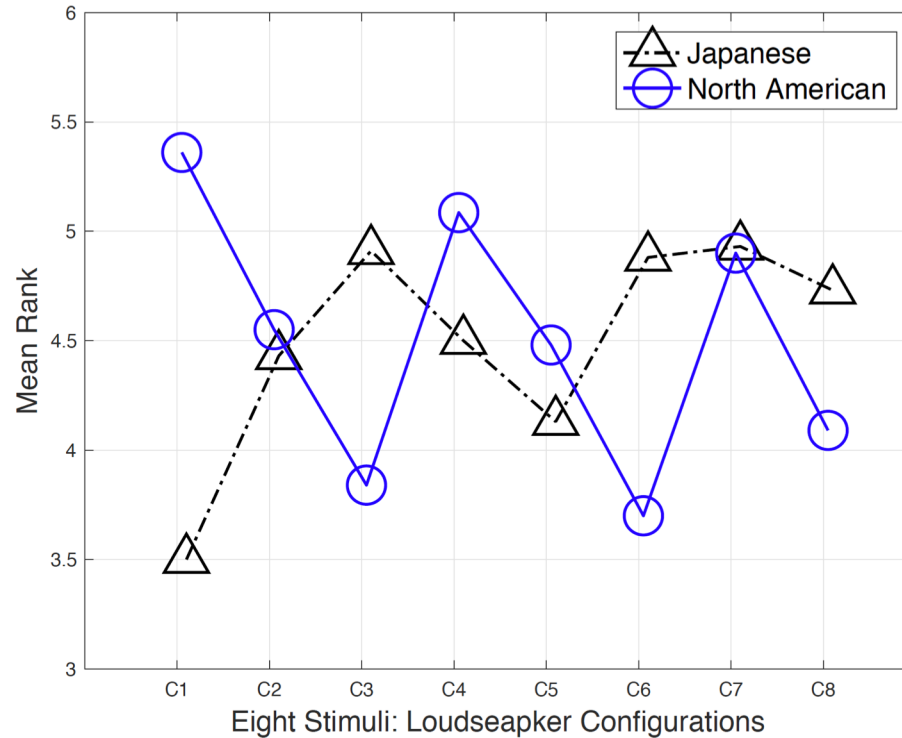
Conceptual Model of Immersive Experience





- Content is the key factor for involvement (cognitive absorption).
- Dependent on contextual factors
 - Subjective preference
 - Cultural background
 - Expectation

- Contextual factors
 - Culture
 - Gender
 - Age
 - Expertise
 - Environment
 - Etc.



Kim et al. (2013)

- A property of Immersive System and Content, related to **Presence**.
- Slater (2009): “**Plausibility Illusion**”
 - *“determined by the extent to which the system can produce events that directly relate to the participant, the **overall credibility of the scenario being depicted in comparison with expectations.**”*
- Plausibility is about the system user’s **internal reference**.

- Exact replication of real life events vs. plausible representation?



Perceptual Realism



Social Realism

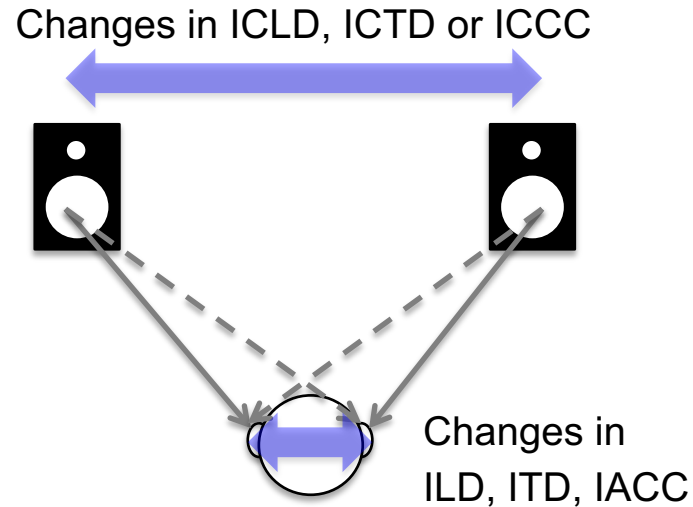
Lombard and Ditton (1997)

- Sensorimotor contingency (SMC)
 - Our sensory perception is linked with our motion (O'regan and Noe 2001).
 - Head / Motion tracking



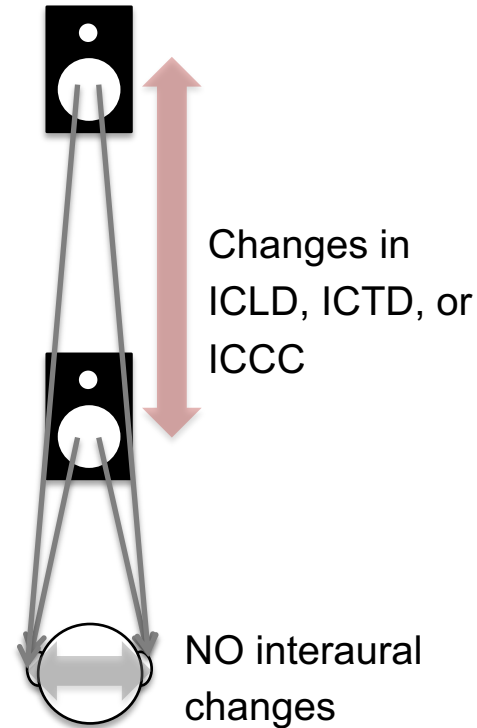
Auditory parameters for providing plausibility

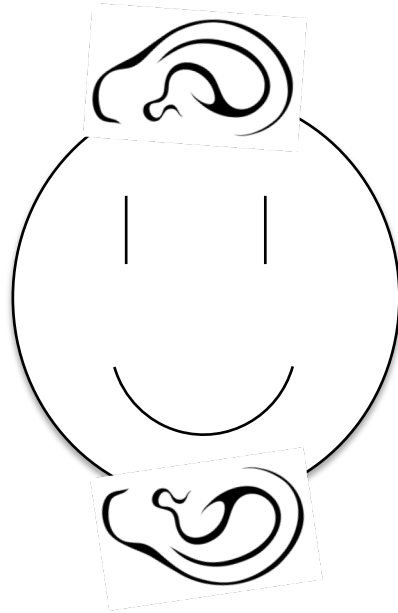
- Horizontal spatial perception
 - Inter-Channel cues translated into Inter-Aural cues



- Vertical spatial perception in the median plane.

Vertical localisation solely relies on **spectral** cues.

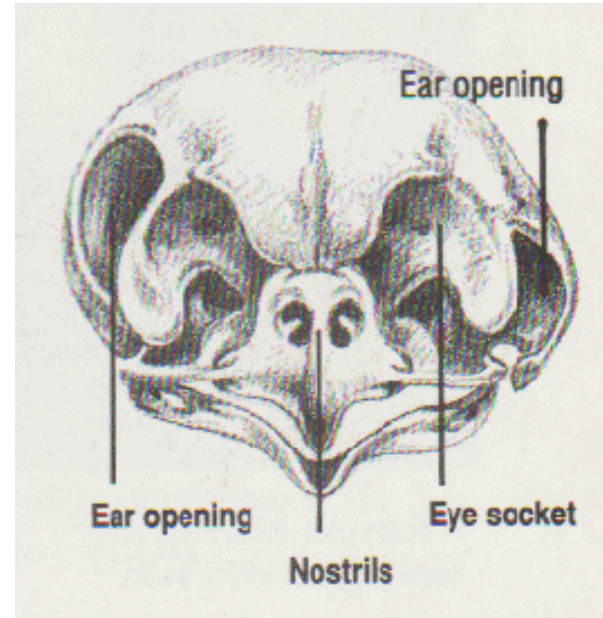




Barn owl



Asymmetrical ears

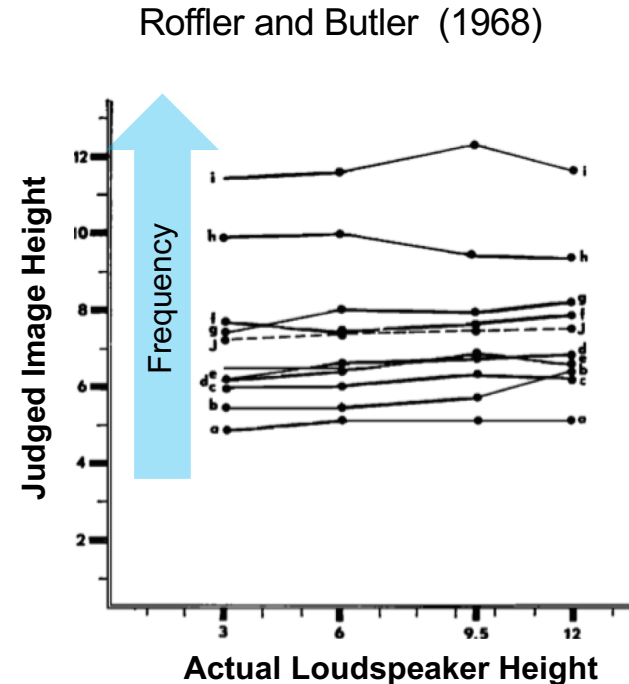


Human localisation ability using auditory cues alone is inherently limited.

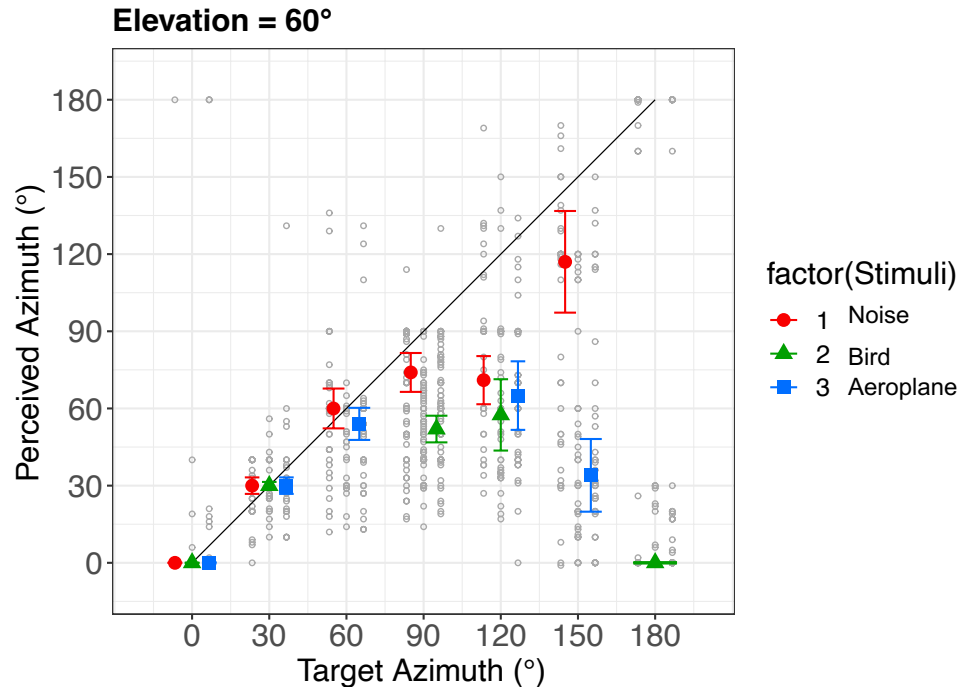
Needs for contextual factors

- Motion/head tracking
- Acoustic congruency

- Vertical localization mainly relies on spectral cues (e.g. HRTF).
- **The Pitch-Height Effect**
 - Higher frequencies tend to be localised at a higher position, regardless of the physical height of the source.
 - Low frequencies below about 200 Hz will be localisation at the ear height or below.



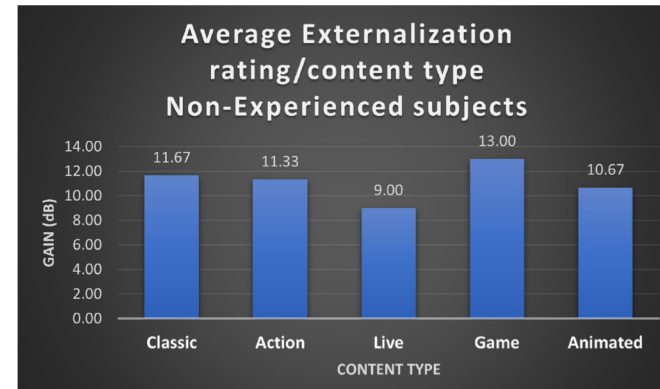
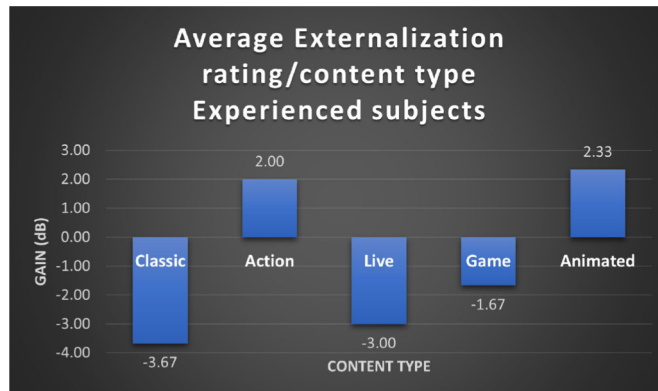
- Poor localization when head rotation is not allowed. Even with personal HRTF!



Lee and Mironovos (2022)

- Externalisation is a cognitive process. Plausibility!
- Adaptation process: the brain learns the acoustic pattern of the room you are in very quickly and uses internal references to perceive sound distance.
 - Stimulus (binaural audio) vs. Template (internal reference of the room acoustics).
- Room divergence effect [Klein et al 2017, Gil-Carvajal et al 2016, etc.]
 - E.g. Binaural audio captured in a large hall would not be externalized very well in your bedroom.
 - Externalization improves through training.
- Influence of visual cue
 - E.g. Ventriloquist effect

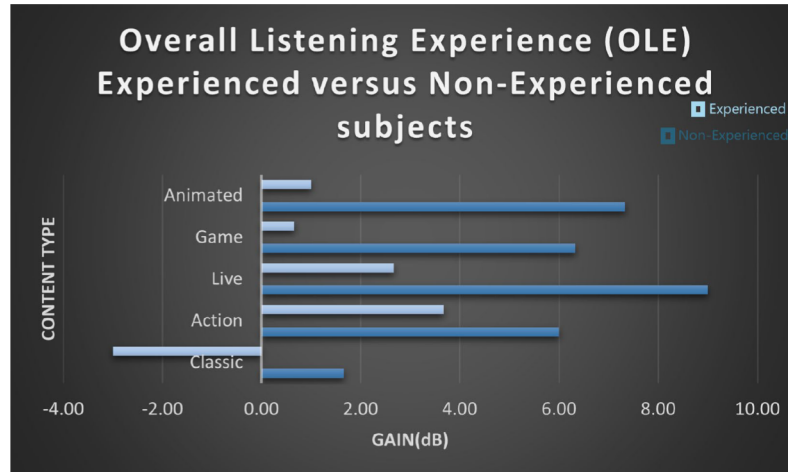
- How much DRR is required for matching binaural and real source distances for Atmos audio-visual content?
 - Experienced vs. Naïve listeners



[Lee and Hadjilambri and 2022]

- Naïve listener required about 11dB more reverb than the actual room!

- How much DRR is required for the Overall Listening Experience (OLE) for Atmos audio-visual content?

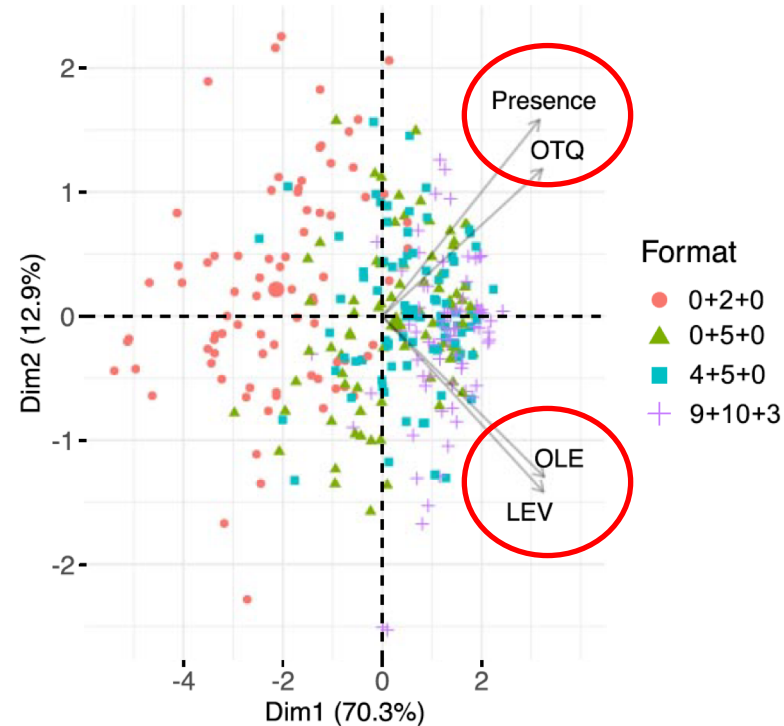


- Naïve listeners tend to like more reverb than experienced ones.

Presence (sense of being there) is not just about spatial attributes.

Presence & Overall Tonal Quality
(Plausibility?)

Overall Listening Experience &
Listener Envelopment

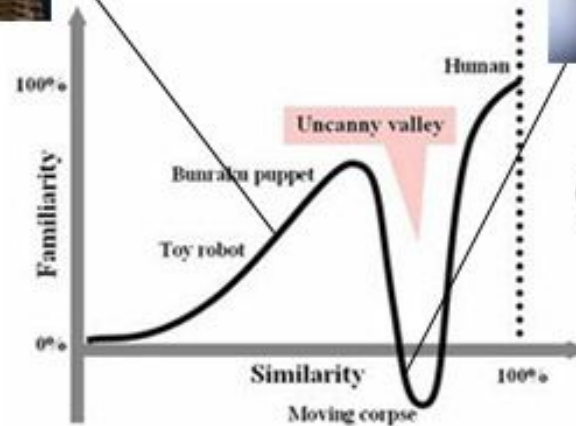


Tonal quality

- Uncanny valley



CUTE



CREEPY

- Let's distinguish between Immersive System and Immersive Experience when we talk about immersive.
- Immersive experience has two main attributes: Presence and Involvement
- Content and contextual factors are the first important factors for immersive experience.
- Plausibility is a key property of immersive system. 100% replication of real event is not necessary.
- Plausibility in binaural audio is not just about HRTF, but acoustic congruency and sensory motor contingency.
- Tonal quality is as important as spatial quality.

Thanks for listening.

h.lee@hud.ac.uk
www.hud.ac.uk/apl