AES 147th International Convention
Paper 10309

The Influences of Microphone System, Video, and Listening Position on the Perceived Quality of Surround Recording for Sport Content

Aimee Moulson and Hyunkook Lee

h.lee@hud.ac.uk

Applied Psychoacoustics Lab (APL)
University of Huddersfield, UK

Research Question

- What technical factors significantly influence the perceived audio quality of 3D immersive sport content?

Independent Variables		Dependent Variables	
Microphone system	First-Order Ambisonics (FOA) vs. Equal Segment Mic Array (ESMA)	**Presence**	Sense of being there in the space.
Loudspeaker format	4-channel Quadraphonic vs. 8-channel double quad layers	**Listener Envelopment (LEV)**	Degree to which auditory scene is enveloping the listener.
Listening position	On-centre vs. Off-centre	**Robustness**	Degree to which the position of an auditory event changes with listener movements.
Content type	Indoor (squash) vs. outdoor (hockey)	**Depth**	Sense of perceived distance of in the auditory scene as a whole.
Video	With and without Video on TV	**Quality of Experience (QoE)**	The overall acceptability of the auditory experience as perceived subjectively.

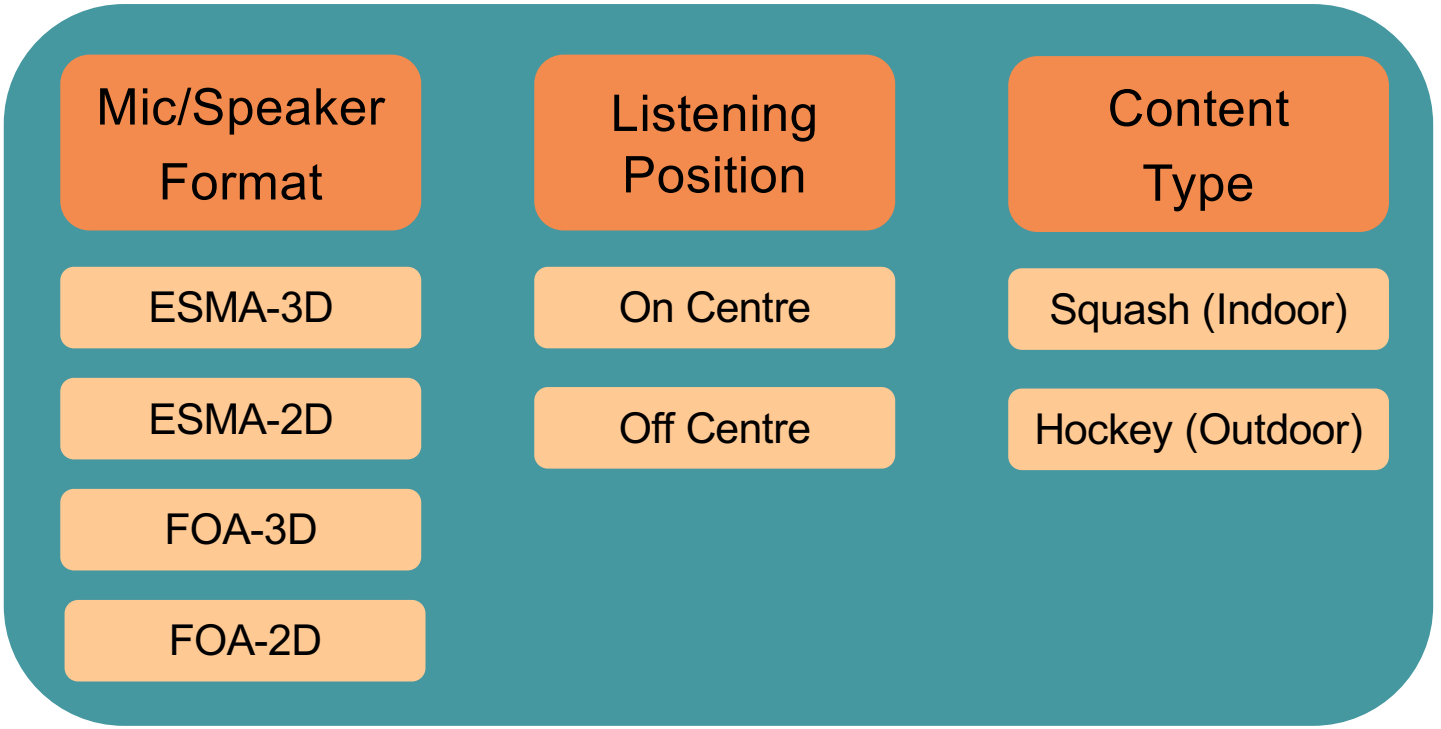
Experimental Design

- Mixed design
  - Between-Subject for Video (On/Off) - 2 subject groups
  - Within-Subject for Recording/Reproduction Format, Video and Listening Position.

Video On

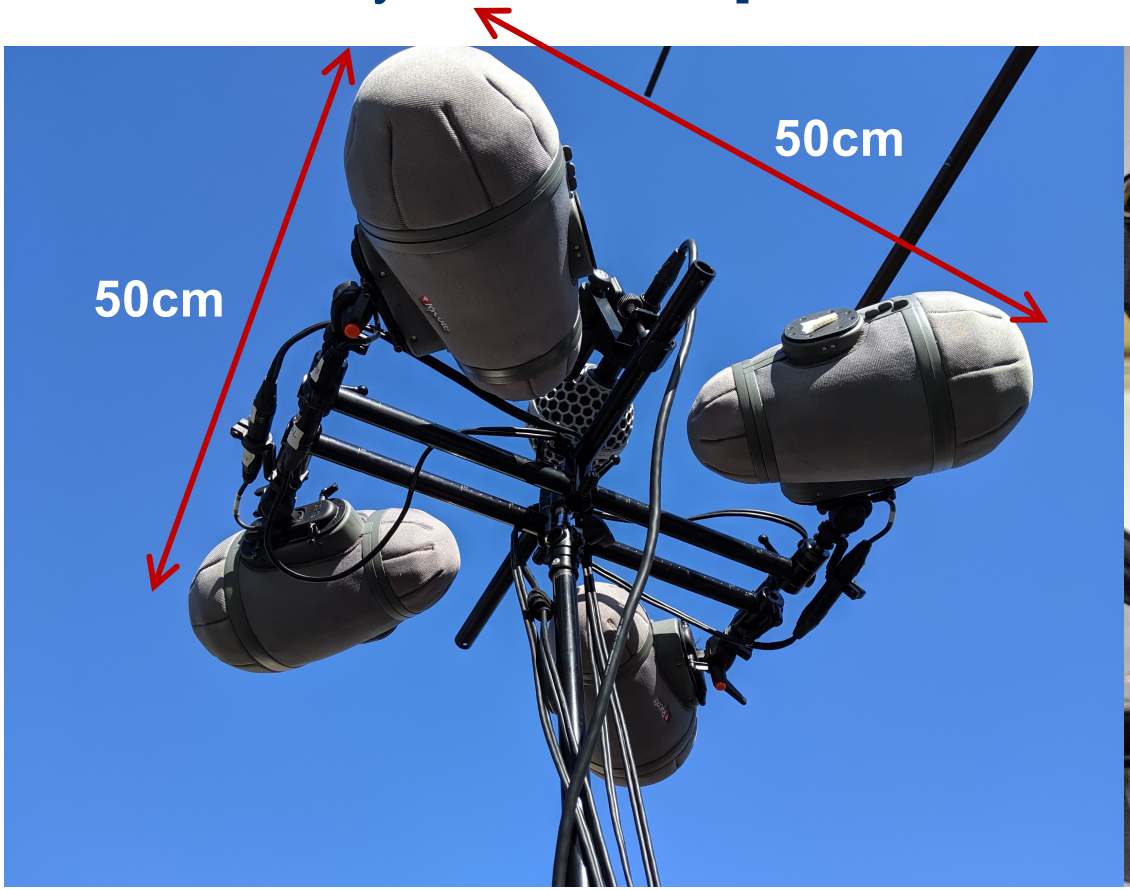
vs.

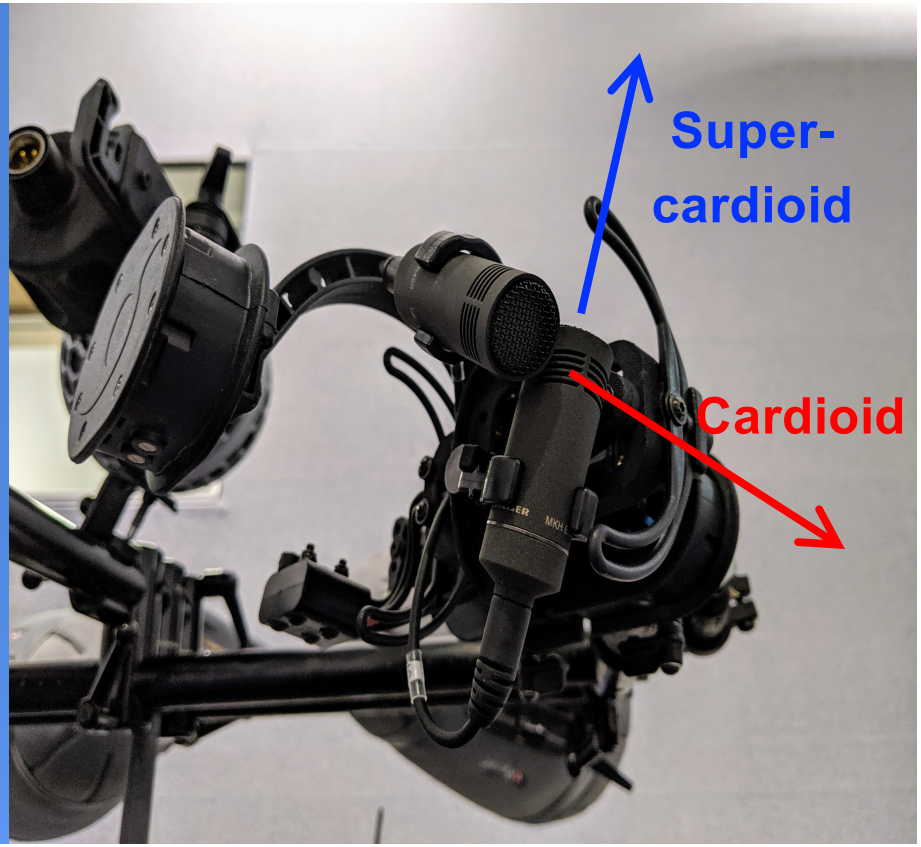
Video Off



ESMA-3D (Equal Segment Microphone Array 3D)

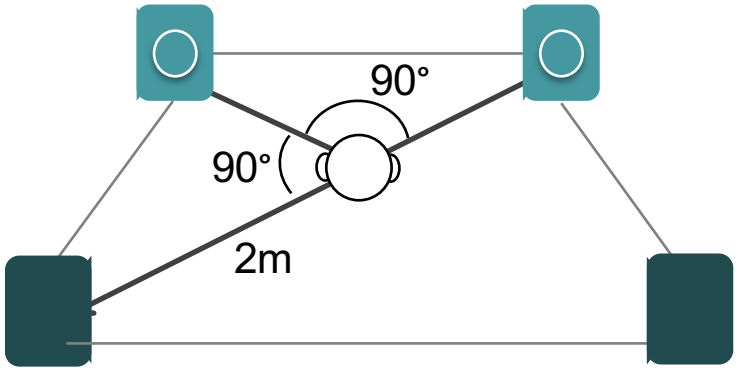
- A 50cm x 50cm square of 4 cardioids for the main layer and 4 supercardioids for the height layer [Lee 2018].
- Vertically Coincident [Lee and Gribben 2014].



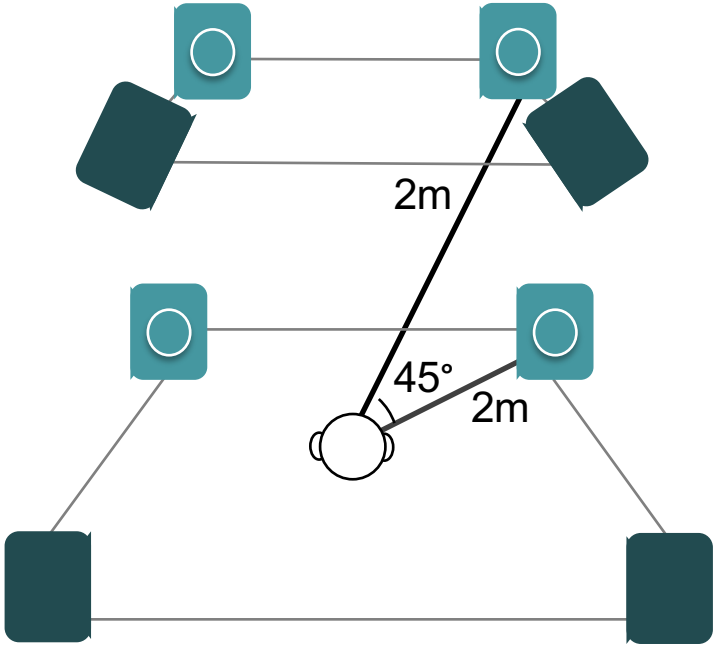


Reproduction Format

- 4ch quad (2D) vs. 8ch double quad layers (3D)



4-channel
Quadraphonic



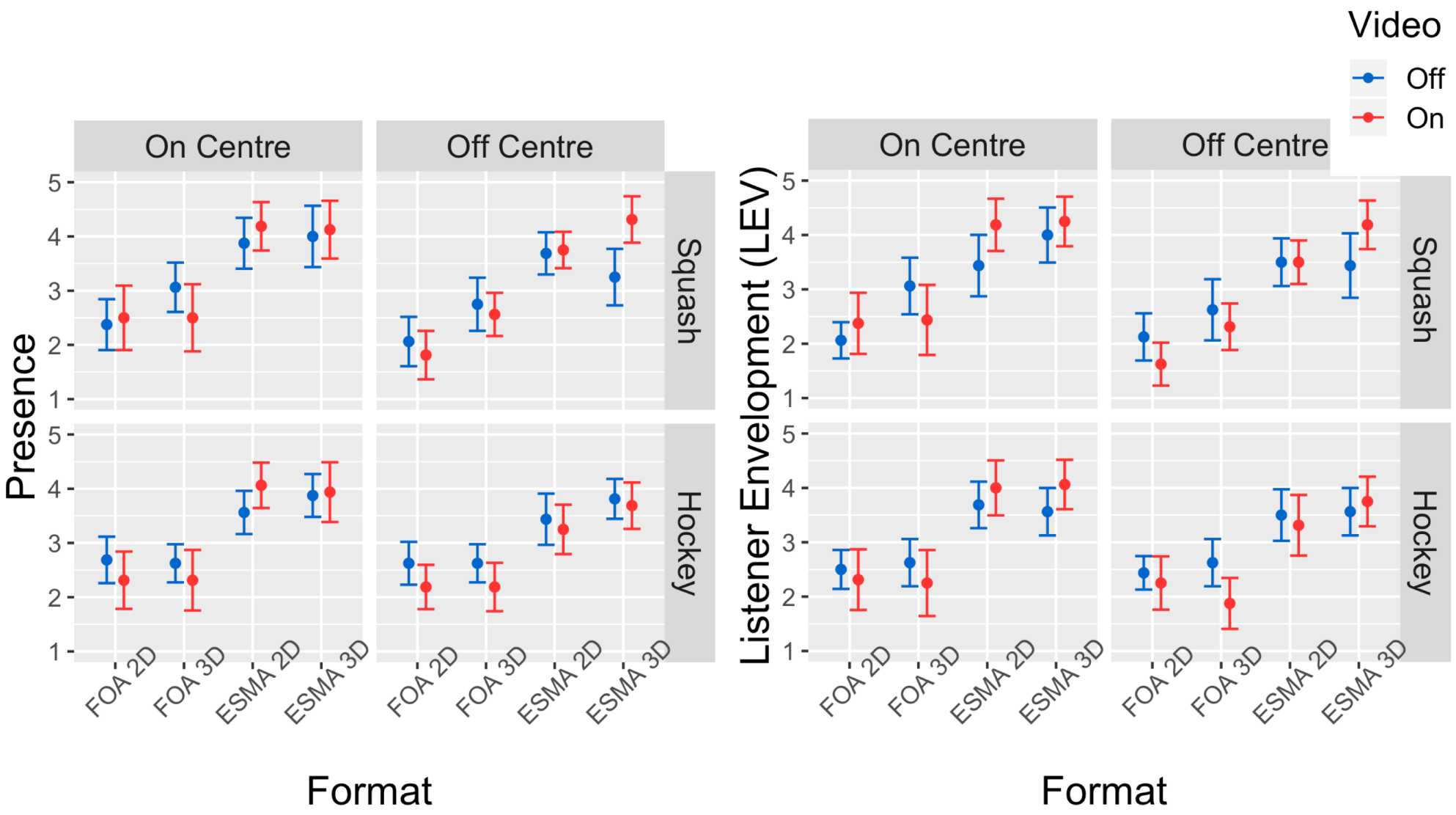
8-channel
Quad + Quad

Listening Test

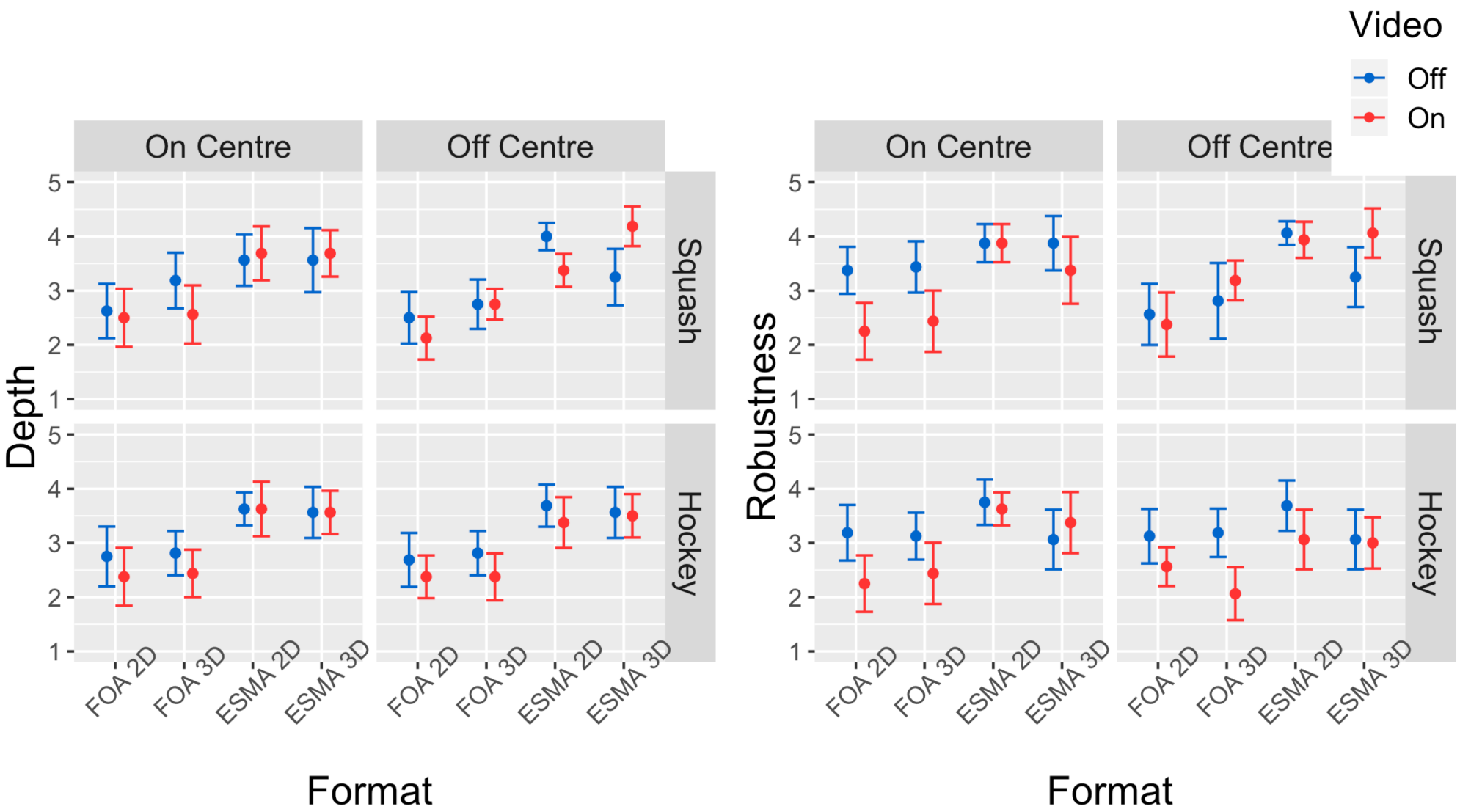
- Conducted in an ITU-R BS.1116-compliant listening room.
- 16 Subjects (2 repetitions for each condition)
  - 8 with Video on 50-inch TV
  - 8 without Video
- 16 stimuli for each group
- Single stimulus presentation in a randomised order.
- 5-point Absolute Categorical Rating (ACR)

  (5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, 1 = Bad)

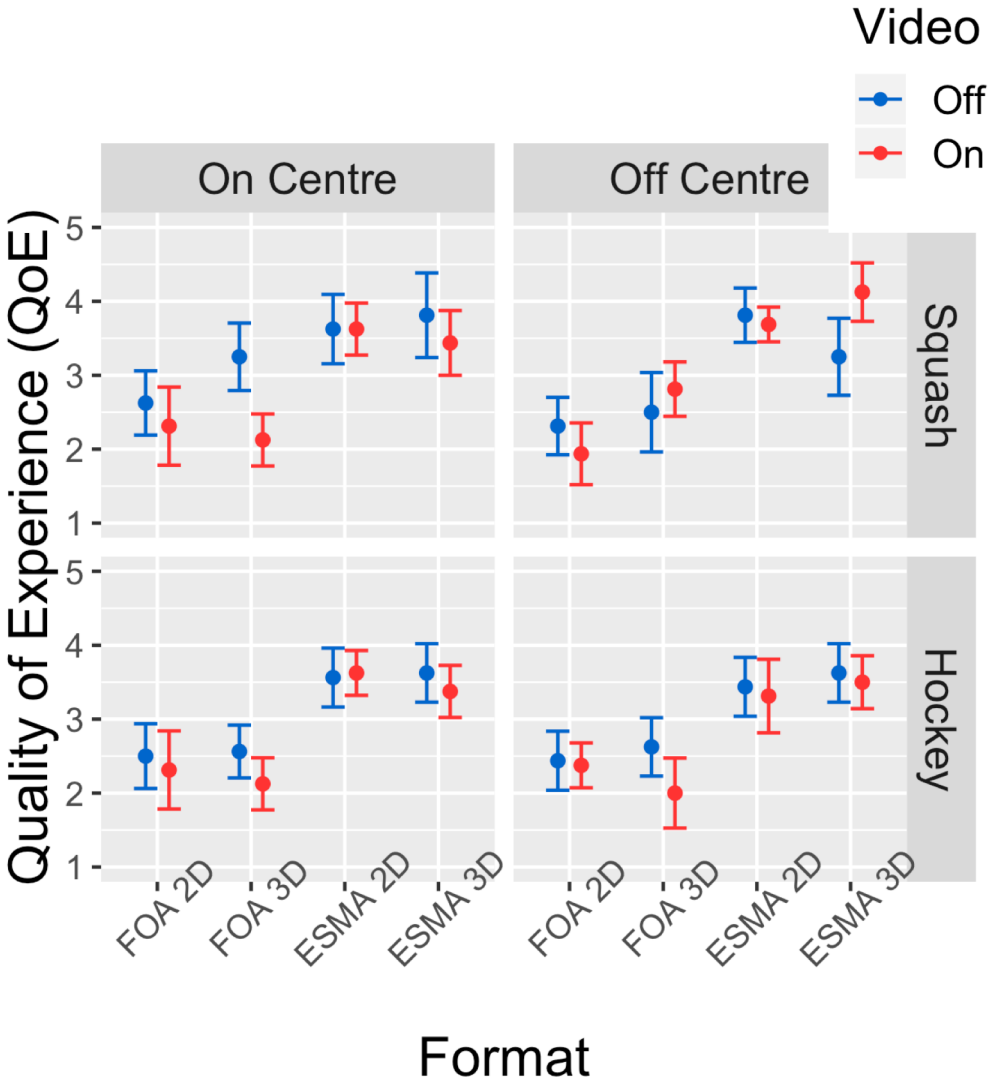
Results

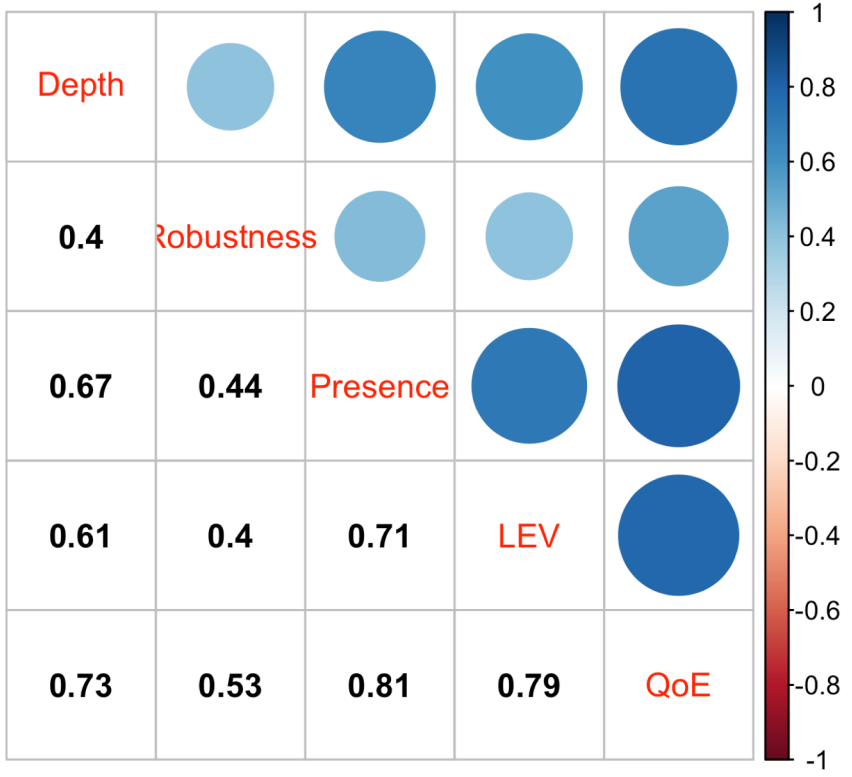

Video
Off
On
On Centre
Off Centre
Squash
Hockey
Presence
Listener Envelopment (LEV)
1
2
3
4
5
FOA 2D
FOA 3D
ESMA 2D
ESMA 3D
Format
Format

Results


Video
Off
On
On Centre
Off Centre
Squash
Hockey
Depth
Robustness
5
4
3
2
1
FOA 2D
FOA 3D
ESMA 2D
ESMA 3D
Format
Format

Results



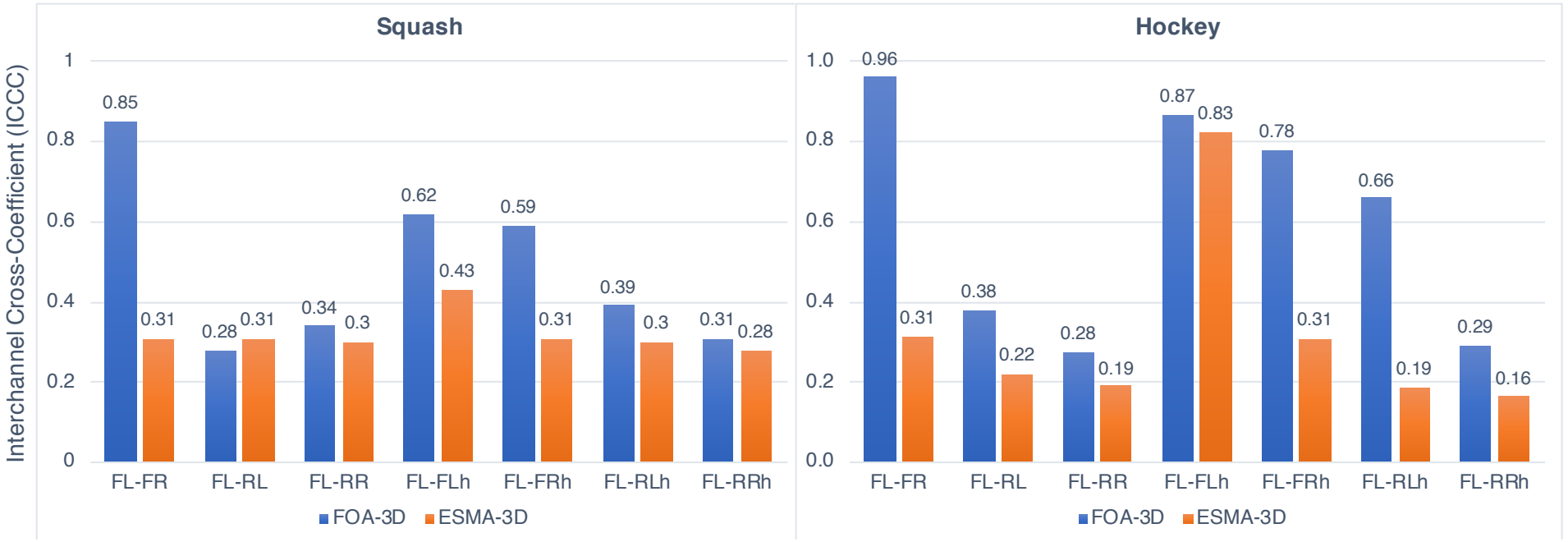


QoE is most correlated with Presence and LEV.

Discussion

- The results showed that ESMA was rated higher than FOA overall.
- 2D and 3D speaker formats did not have a significant difference.
- ESMA has lower interchannel correlation than FOA.

→ Greater LEV and Robustness → Better Presence and QoE



Discussion

- The video accompaniment mainly worked negatively on Robustness and QoE with the FOA, whereas it was a positive factor on Presence, Depth and QoE with the ESMA.

- For FOA, Video tends to degrade QoE at On-Centre but improve it at Off-Centre.

→ Discrepancy between audio and visual references due to the FOA's failure to physically reconstruct the sound field.

→ ESMA provides more plausible representation of the sound field.

References

- Lee, H, “Capturing 360° Audio using an Equal Segment Microphone Array (ESMA)”, *J. Audio Eng. Soc.,* vol. 67, no. 1/2, pp. 13–26, (2019).
- Millns, C & Lee, H (2018), “*An Investigation into Spatial Attributes of 360° Microphone Techniques for Virtual Reality”*, In *Audio Engineering Society 144th International Convention* (2018).
- H. Lee and C. Gribben, “Effect of Vertical Microphone Layer Spacing for a 3D Microphone Array,” J. Audio Eng. Soc., vol. 62, pp. 870–884 (2014 Dec.).
- Odya, P, Czyzewski, A & Kostek, B, “Determination of Influence of Visual Cues on Perception of Spatial Sound”. In *Audio Engineering Society 110th International Convention* (2001).
- ITU, “Definition of quality of experience” https://www.itu.int/md/T05-FG.IPTV-IL-0050/en, (2007)