

The three localization cues (what you must reproduce)

- ITD** Interaural time difference - timing gap between ears. Dominates below 1,500 Hz.
- ILD** Interaural level difference - head shadow. Dominates above 1,500 Hz.
- Pinna notches** Spectral dips in the 4-9 kHz band; elevation needs energy above ~7 kHz.

Ambisonic order vs channels (channels = (N+1) squared)

- Order 1 (FOA): 4 channels. Coarse field. YouTube 360, low-bandwidth VR.
- Order 2: 9 channels. Better. Mid-tier VR.
- Order 3 (TOA): 16 channels. Sharp localization. High-end VR, immersive production.
- Order 7: 64 channels. Reference-grade. Research / studio monitoring.
- Rule: TOA (16 ch) = 4x the bits of FOA (4 ch). Pay it only when localization is the product.

Format hygiene (get this wrong and the field scrambles silently)

- Standardize on AmbiX: ACN channel ordering + SN3D normalization.
- Convert legacy FuMa (W,X,Y,Z) at the boundary - never mix conventions.
- Store HRTFs as SOFA / AES69 (.sofa) so you can swap in better or personalized sets later.
- YouTube needs AmbiX FOA: W,Y,Z,X at 48 kHz (or 6-ch with head-locked stereo).

Generic vs personalized HRTF

- Generic HRTF: fine left-right for most people; weaker elevation / front-back.
- Personalized HRTF (phone scan / photos): better elevation, externalization. Apple, Dolby ship it.
- Symptom 'sound is inside my head' = HRTF mismatch or none. Fix the chain, not the reverb.

Pre-ship checklist

- Real HRTF applied (not just panned stereo) - verify externalization.
- Head-tracking active and motion-to-sound latency kept low (tens of ms).
- Ambisonic order matched to the product, not maxed out by reflex.
- AmbiX/ACN/SN3D confirmed end to end; FuMa converted, not fed raw.
- Head-locked stereo routed for narration / music that must not rotate.
-