

The three ways to describe a sound scene: what each stores, where it wins, what to ship.

The three approaches

CHANNEL-BASED

Stores: Fixed speaker feeds, one per loudspeaker.
Examples: Stereo, 5.1, 7.1, 7.1.4, 22.2
Wins: Simple; zero playback compute.
Watch: One file per layout; sounds wrong if room differs from studio.

OBJECT-BASED

Stores: Each sound + its 3D position metadata.
Examples: Dolby Atmos, DTS:X
Wins: One master plays headphones -> 64-speaker cinema.
Watch: Atmos: up to 128 studio inputs, spatially coded to ~12-16 clusters for delivery.

SCENE-BASED

Stores: The whole sound field as layers (W, X, Y, Z...).
Examples: Ambisonics; YouTube / VR / 360 audio
Wins: Rotates with the head; great for VR/AR.
Watch: FOA = 4 channels; HOA = $(order+1)^2$ (order 3 = 16).

Reading the notation

5.1 = 5 full channels + 1 LFE (bass-only). The .1 is the subwoofer.

7.1.4 = 7 ear-level + 1 LFE + 4 OVERHEAD. Third number = ceiling speakers.

FOA = First Order Ambisonics, 4 channels (W omni + X/Y/Z axes).

HOA = Higher Order Ambisonics, $(order+1)^2$ channels: 9 (2nd), 16 (3rd), 64 (7th).

How to choose (2026)

- Head tracked (VR / AR / 360)? -> SCENE-BASED (ambisonics).
- Need precise moving sounds or device adaptivity? -> OBJECT-BASED (Atmos / MPEG-H / DTS:X).
- Neither? -> CHANNEL-BASED fallback (stereo / 5.1) is fine.
- MPEG-H (ISO/IEC 23008-3) carries all three at once + viewer interactivity.
- Tier-1 anchors: ITU-R BS.2076-3 (ADM, the 3 types), ISO/IEC 23008-3 (MPEG-H).