

MPEG-H 3D Audio - System Cheat Sheet

The scene model, the packet transport, the profiles, and where it ships in 2026 - on one page.

The scene model (Metadata Audio Elements)

AudioSceneInfo	Root of the scene - the whole program.
Group	Bundled signals treated as one unit (a stereo pair, a sub-mix).
Switch Group	Mutually exclusive options - pick exactly one (e.g. language).
Preset	Named one-tap combo, e.g. 'Dialogue Enhancement', 'Live Mix'.

MHAS packet types (the typed boxcars)

PACTYP_MPEGH3DACFG	Decoder config: layout, objects, core params.
PACTYP_AUDIOSCENEINFO	The MAE scene - must follow the config packet.
PACTYP_MPEGH3DAFRAME	The compressed audio frames.
PACTYP_USERINTERACTION	Viewer's knob turn, fed back into the stream.
PACTYP_AUDIOTRUNCATION	Shortens last frame for sample-accurate splice.

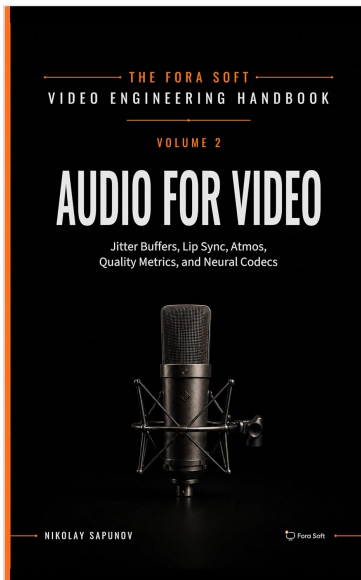
Profiles & levels (what a TV chip implements)

MAIN	(5 levels): full toolkit, up to 64 speakers / 128 core ch. Studio / reference - NOT in consumer devices.
LOW COMPLEXITY (LC)	channels + objects + HOA, ~50% lower decoder cost. ATSC 3.0 uses Levels 1-3.
BASELINE (BL)	subset of LC, channels + objects, simplified metadata, no advanced HOA. Added to ATSC 3.0 in 2025.

LC Level 3 caps loudspeaker output at 12 channels (enough for 7.1.4). Broadcast bitrate cap 1,200 kbps.

Six facts to remember

- ☐ MPEG-H is a SCENE, not a mix - the renderer assembles it for each device.
- ☐ Spec: ISO/IEC 23008-3 (codec + scene); ATSC A/342-3 constrains it for broadcast.
- ☐ Hybrid delivery: main stream over the air (mhm2), aux streams over broadband, merged by MAE.
- ☐ Loudness compensation auto-trims level when a viewer raises an element - stays CALM/EBU compliant.
- ☐ Binaural headphone output is a RENDER of the scene, not the master. Author the scene.
- ☐ Ships: Korea terrestrial UHD (sole codec, since 2017); Brazil DTV+ (required, 2026 World Cup); Sony 360RA.



GO DEEPER · VOLUME 5 OF 7

Audio for Video: A Complete Guide to Sound in Video Products

From Jitter Buffers and Lip Sync to Atmos, Quality Metrics, and Neural Codecs

By Nikolay Sapunov, CEO at Fora Soft · 1,288 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
amazon.com/dp/B0HH6XPY46

This sheet covers the MPEG-H scene model, MHAS packets, profiles and where the system ships in 2026. Volume 5 explains the delivery side of immersive audio: spatial and object audio, timestamps and clock domains, lip sync across hybrid streams and how to measure the result.

WHAT YOU WILL FIND INSIDE

- Object audio and interactive scenes
- Timestamps and clock domains in hybrid delivery
- Lip sync between broadcast and broadband streams
- Measuring immersive audio quality

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

Building an MPEG-H pipeline for broadcast or hybrid delivery?

We will review your authoring, packaging and hybrid delivery plan and list the gaps against ATSC 3.0 or DTV+. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)