Audio & Subtitle Track Worksheet

Plan the non-video tracks before you ask a vendor for a quote: list audio languages and codecs, set one loudness target, choose caption formats, plan the accessibility tracks, and confirm audio is demuxed. Engineering guidance — confirm vendor and loudness numbers live, they change.

1 · AUDIO LANGUAGES & CODECS (demux, don't mux)

- ☐ **Demux audio from video.** Keep audio in separate files so one video ladder serves every language. Muxed = each language re-stores the whole picture (up to ~10x). HLS EXT-X-MEDIA / DASH AdaptationSet are built for this.
- ☐ **Ship AAC-LC stereo as the floor.** Every HLS/DASH device decodes it. Add a surround alternate (E-AC-3 / Atmos) in the same audio group for living-room devices; verify the device matrix.
- ☐ **Write your audio languages:** _____ · **surround codec:** _____ (E-AC-3 / AC-4)

2 · LOUDNESS TARGET (pick one, enforce at ingest)

- ☐ **Measure with ITU-R BS.1770** (unit LUFS = LKFS). Pick ONE catalog-wide target and normalize every title to it — or viewers ride the volume between programs.
- ☐ **Common targets:** on-demand streaming ~ -27 LKFS dialogue-gated, true-peak -2 dBTP; broadcast -23 LUFS (EBU R128) or -24 LKFS (ATSC A/85); AES TD1008 streaming -18 LUFS speech / -16 music.
- ☐ **Write your loudness target:** _____ LKFS · **true-peak ceiling:** _____ dBTP

3 · CAPTION & SUBTITLE FORMATS (sidecar vs embedded)

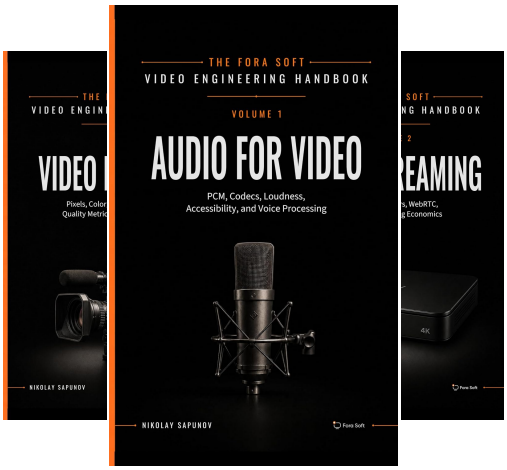
- ☐ **Default to WebVTT** (wvtt) for web-grade text; use **IMSC 1.1 / TTML** (stpp) for broadcast-grade styling — the TTML profile CMAF allows. Both are sidecar files, cheap to add per language.
- ☐ **CEA-608 / 708** (CTA-608/708) are embedded inside the video stream, not sidecar files — they ride along but can't be added or styled without re-touching the video. Mostly for broadcast-origin content.
- ☐ **Subtitles are not captions.** Subtitles translate dialogue; SDH captions add non-speech sound for deaf / hard-of-hearing viewers. Plan SDH, not just subtitles. Format: _____

4 · ACCESSIBILITY TRACKS & THE LAW

- ☐ **Described video** (audio description) is a separate audio track narrating visual action — for blind / low-vision viewers. **Forced-narrative** subtitles always show, to translate foreign dialogue and signs.
- ☐ **US law (CVAA, 47 CFR 79.4):** full-length internet video that aired on US TV with captions must be captioned online. **WCAG 2.1 AA:** captions on all content (1.2.2/1.2.4) + audio description on prerecorded video (1.2.5).
- ☐ **Confirm your obligations:** aired on US TV? _____ · WCAG 2.1 AA target? _____

THE ONE SANITY CHECK BEFORE YOU LOCK THE TRACK PLAN

Demux audio, then count tracks, not features. Muxed: storage = (sum of video rung bitrates) x hours / 8, multiplied by the number of languages, because the picture is welded to each language. Demuxed: store the video once and add a small audio track per language. Worked: a 17.5 Mbps video ladder for a 2-hour film with 10 audio languages at 128 kbps each costs 17.5 x 10 = 175 Mbps muxed, versus 17.5 + (0.128 x 10) = 18.78 Mbps demuxed — about 9x less storage and egress, and the CDN caches one video instead of ten. If a single-language pilot muxed audio into video, fix it before the second language ships. Subtitles, captions, and accessibility tracks are tiny by comparison — the cost is the audio-mux decision and the compliance gap, not the text.



GO DEEPER · THE 7-VOLUME HANDBOOK

Fora Soft Video Engineering Handbook

Seven volumes, 9,000+ pages: encoding, streaming, audio and AI for video products — from the pixel to production.

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

Browse the handbook on Amazon →

This worksheet plans audio languages, one loudness target, caption formats and accessibility tracks before a vendor quote. Our seven-volume handbook covers the full video stack; start with Volume 4 for codecs, loudness and accessibility, then Volume 3 for subtitles and players and Volume 1 for encoding.

WHAT YOU WILL FIND INSIDE

- Loudness, LUFS and audio description (Vol. 4)
- Codecs and channel layouts (Vol. 4)
- Subtitles, captions and players (Vol. 3)
- Encoding the video ladder (Vol. 1)

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

Planning languages, captions and accessibility for an OTT launch?

We will draft your track plan, check it against CVAA and WCAG duties and estimate the storage saved by demuxing audio. 250+ video products shipped since 2005.

Book a 30-min call → **WhatsApp →** **Email us →**