

Lip-Sync Test Methodology Checklist

Four ways to measure, one tolerance window, and the time-of-flight correction teams forget.

The four methods

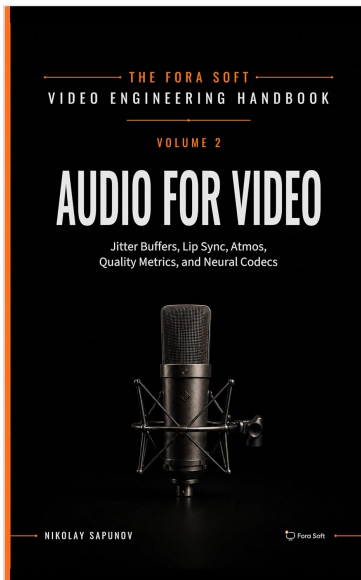
1 Clapper / flash-and-beep	One flash + one click in a clip; scrub frame by frame. ~1-frame precision. Cheapest absolute number.
2 Light + mic probe	Light sensor on screen, mic at speaker. ~0.1 ms over hundreds of ms. MUST subtract time-of-flight.
3 Automated detection	Onset correlation or SyncNet-style matching. No test card; works on real content; scales to library/live.
4 Perceptual check	Human sign-off in a defined room. The tolerance windows came from subjective panels in the first place.

The numbers you measure against

Sign	Positive = audio leads (voice early). Negative = audio lags (voice late).
Acceptability	Audio +90 ms lead to -185 ms lag (ITU-R BT.1359-1).
Detectability	Tighter: +45 ms to -125 ms. ATSC headroom target: +15 ms / -45 ms.
Time-of-flight	Sound = ~343 m/s = ~2.9 ms per meter. Subtract distance / 343 from every probe reading.

Make it repeatable: six checks

- ☐ Verify the source clip: confirm the flash is exactly one frame in the FILE, not the captured output.
- ☐ Record the capture point in the pipeline so two runs measure the same path.
- ☐ Measure and log the microphone-to-speaker distance; subtract distance / 343 m/s before reading.
- ☐ Re-run the same clip twice; if numbers differ beyond your method's precision, fix the rig, not the system.
- ☐ Trace the offset to a stage: constant lag = codec/priming path; growing lag = clock drift; live-only = RTP timing.
- ☐ End every release with a perceptual check in a fixed room: monitor, distance, speakers, material all stated.



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 checklist gives four ways to measure A/V offset and the time-of-flight correction to apply to every probe reading. Volume 5 explains where the offset comes from: lip sync, timestamps and clock domains, jitter buffers, and incident analysis by symptom, so you can trace a measured number to the stage that caused it.

WHAT YOU WILL FIND INSIDE

- Lip sync, timestamps and clock domains
- Diagnosing audio incidents by symptom
- Jitter buffers and NetEq behaviour
- Quality metrics: PESQ, POLQA, ViSQOL

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

Measuring lip-sync drift in your player or streaming pipeline and need a fix?

We will set up a repeatable A/V offset test, find the stage that adds the delay, and fix it in your pipeline. 250+ video products shipped since 2005.

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