

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.