RTP / RTCP Sync Reference Card

Fora Soft Learn

Timestamp clock rates, the sender-report fields, the wall-clock math, and the failure modes - on one page.

The two pieces

RTP ts	Counts media sampling ticks from a random start. Not packets, not wall time.
RTCP SR	Pairs one RTP reading with the NTP wall-clock time at the same instant.

RTP media clock rates

Opus audio	48,000 Hz - 960 ticks per 20 ms packet (RFC 7587).
G.711 audio	8,000 Hz - 160 ticks per 20 ms packet (RFC 3551).
Video	90,000 Hz - 3,000 ticks per frame at 30 fps (RFC 3551).

Media clock to wall clock

NTP format	64-bit fixed point: seconds since 0h UTC 1 Jan 1900; 32 whole + 32 fraction.
SR pair	RTP ts R0 maps to NTP time T0 (same instant).
Convert	wall(R) = T0 + (R - R0) / clock rate. Do it per stream.
Lip-sync	Both streams on one wall clock; render matching audio + frame together.

When sync breaks

Clock drift	**ppm crystal mismatch**	Lip-sync drifts steadily; worse over a long call.
Network jitter	**uneven arrival**	Sync wobbles, unstable, not one-directional.
Missing / wrong SR	**no valid anchor**	Streams fine alone, permanently out of sync.
Sync lock delay	**RTCP rate limit**	First ~1-5 s off, then snaps in (RFC 6051 fixes).

Remember

- [ ] Audio and video are separate RTP streams with separate random-start clocks.
- [ ] The RTP timestamp counts media sampling ticks, not packets and not wall time.
- [ ] Audio clocks at 48,000 Hz (Opus) or 8,000 Hz; video at 90,000 Hz.
- [ ] The RTCP sender report pairs one RTP reading with an NTP wall-clock time.
- [ ] One pair per stream puts both on one timeline - that is lip-sync.
- [ ] Sender reports must keep arriving; each fresh one re-anchors against drift.
- [ ] RFC 6051 and WebRTC abs-capture-time carry the mapping in-band for fast sync.
- [ ] A middlebox that regenerates SRs on its own clock silently breaks sync.