WebRTC Audio Pipeline — Cheat Sheet

Microphone to speaker: every stage, its component, and its typical one-way latency.

Pipeline stages

Stage	What it does	Latency
1. Capture	Microphone → 48 kHz samples (getUserMedia)	~10-20 ms
2. High-pass filter	Drop low rumble below speech	incl.
3. AEC3	Acoustic echo cancellation (uses reference)	incl.
4. Noise suppression	Attenuate steady background sound	incl.
5. AGC2	Bring every voice to a steady level	incl.
6. Opus encode	Compress 20 ms frames (RFC 6716)	~20 ms
7. RTP packetize	Header (seq/ts/SSRC) + SRTP encrypt	~1 ms
8. Network	Delay, reorder, loss, jitter	~10-150 ms
9. NetEQ	Adaptive jitter buffer + loss concealment	~20-100 ms
10. Opus decode	Frame → raw samples	incl.
11. Render	Samples → speaker / headphones	~10-30 ms

> **Capture clean-up runs in a FIXED order:**
> High-pass filter → AEC3 → Noise suppression → AGC2.

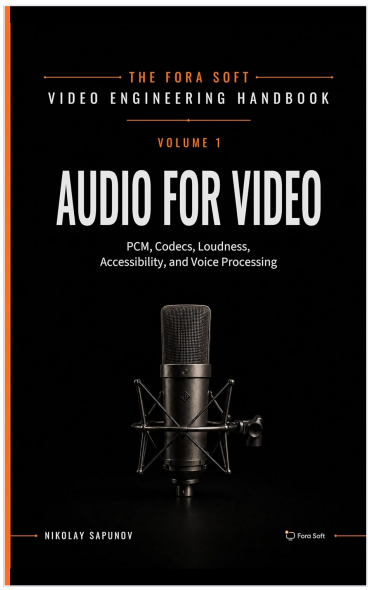
The six NetEQ operations

Normal	packet available, buffer fine → clean audio
Acceleration	buffer too full → play slightly faster
Preemptive exp.	buffer too empty → play slightly slower
Expand	packet late/lost → invent audio (PLC)
Merge	real packet after PLC → stitch together
Comfort noise	talker silent (DTX) → soft synthetic hiss

Remember

- [ ] Target one-way mouth-to-ear under ~200 ms (ITU-T G.114); aim ~70-120 ms.
- [ ] Opus FEC embeds a low-bitrate copy of the previous frame to survive loss.
- [ ] Opus DTX cuts a silent talker's bandwidth ~85-90% via silence descriptors.
- [ ] The network and the jitter buffer are the largest, least controllable terms.
- [ ] Use headphones to defeat echo; design AEC3 for the day users do not.

Companion to: 'The WebRTC Audio Pipeline End-to-End'
www.forasoft.com/learn/audio-for-video/articles-audio/webrtc-audio-pipeline-end-to-end



GO DEEPER · VOLUME 4 OF 7

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

From PCM and Codecs to Loudness, Accessibility, and Real-Time Voice Processing

By Nikolay Sapunov, CEO at Fora Soft · 1,229 pages · Kindle & paperback on Amazon

Read it on Amazon → amazon.com/dp/B0HH6PRZH1

This sheet maps every stage from microphone to speaker and gives each a latency figure. Volume 4 explains the capture side of that pipeline in depth: sampling and bit depth, Opus and speech codecs, and real-time voice processing with AEC, AGC, noise suppression and VAD.

WHAT YOU WILL FIND INSIDE

- AEC, AGC, noise suppression and VAD
- Sampling, bit depth and dynamic range
- Opus and speech codecs
- Loudness and LUFS

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

Is your WebRTC audio too late, too quiet or full of echo?

We will profile your pipeline stage by stage, tune the capture chain and Opus settings, and bring mouth-to-ear latency under 200 ms. 250+ video products shipped since 2005.

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