WebRTC Audio Pipeline - Cheat Sheet

Fora Soft Learn

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.