Neural Audio Codec Cheat Sheet

Fora Soft Learn

What each codec does, the 2026 numbers, and when Opus still wins - on one page.

The one idea: residual vector quantization (RVQ)

A neural codec learns compression from data instead of hand-built rules. RVQ rounds the sound in stacked layers - the first captures the coarse value, the next the leftover error, and so on, like making change with dollars, then dimes, then pennies. Drop the lower layers and the bitrate falls instantly, so one model serves many bitrates.

The codecs at a glance

Codec	Bitrate	Equal-quality note	Cost / status 2026
Opus (2012)	6-510 kbps	the baseline everyone compares to	microseconds; WebRTC default
SoundStream	3-18 kbps	3 kbps approx Opus 12 kbps	research blueprint (RVQ)
Lyra v2	3.2/6/9.2	9.2 kbps approx Opus 14 kbps	0.57 ms/frame on a phone; ships
EnCodec	1.5-24 kbps	3 kbps > Opus 12; +48 kHz music	high cost; fuels audio AI
Mimi	~1.1 kbps	speech-LLM tokenizer, not hi-fi	inside voice-AI models

The numbers that decide it

- Neural codecs reach Opus quality at 3-4x fewer bits - but a network must run every frame, costing far more CPU and battery than Opus.
- 50 speakers x 12 kbps Opus = 600 kbps; at 3 kbps neural = 150 kbps. A 4x audio saving - real where bandwidth is scarce, but the compute is not free.
- Lyra v2 is the exception that ships on phones: 0.57 ms to encode+decode a 20 ms frame on a Pixel 6 Pro, ~35x real time, 20 ms delay.
- The bigger 2026 use is voice AI: the codec turns sound into tokens an LLM can predict. Mimi splits them into semantic (what is said) + acoustic (how it sounds).
- Opus stays the live-calling default; the next traditional codec (AOMedia OAC, started 2026) is an Opus successor, not a neural one.

Before you reach for a neural codec - the gates

- [ ] Real-time calling? Default to Opus - near-zero decode, no licence, WebRTC-mandatory.
- [ ] Neural only where bandwidth is the hard limit AND the per-client compute budget exists.
- [ ] Check the worst client device: CPU per frame, battery drain, and added latency.
- [ ] Building voice AI (assistant, dubbing, on-device TTS)? The neural codec is the foundation, not optional.
- [ ] 'Beats Opus at low bitrate' is a lab result - re-test it on your devices before switching.
- [ ] Watch the codec/generator line: rebuilding a voice from 1 kbps is close to cloning it.