Audio Frames — Cheat Sheet

Fora Soft Learn

Frame sizes, the samples-to-ms formula, the RTP packet, and where latency comes from.

The one rule

A frame is the smallest chunk a codec can decode alone. It is a slice of time, measured in samples.

frame duration = samples per frame / sample rate. e.g. 1,024 / 48,000 = 21.3 ms (AAC).

Frame size by codec (at 48 kHz)

Codec	Frame (samples)	Duration	Internal sub-unit
Opus	120-2,880 (you choose)	2.5 / 5 / 10 / 20 / 40 / 60 ms	none
AAC-LC	1,024	21.3 ms	8x128 short blocks on transients
MP3	1,152	24.0 ms	2 granules x 576
AC-3	1,536	32.0 ms	6 audio blocks x 256
G.711	1 (no transform)	often 20 ms packets	none

Mouth-to-ear latency stack (representative, 20 ms frame)

[frame-size] capture (one 20 ms frame must fill)	**20 ms**
[frame-size] encoder frame + look-ahead	**~25 ms**
network transit (one-way, good path)	**~30 ms**
[frame-size] jitter buffer (holds ~2 frames)	**~40 ms**
decoder + output device buffering	**~20 ms**
total	**~135 ms**

tinted rows shrink if you pick a smaller frame.

Frame-and-packet checklist

- [ ] Frame size matches the latency budget (20 ms is the default sweet spot).
- [ ] RTP: one frame per packet; sequence number and timestamp present.
- [ ] Opus RTP timestamp uses the fixed 48 kHz clock (+960 per 20 ms frame).
- [ ] Stream segments rounded to a whole number of audio frames (no drift).
- [ ] Gapless playback: encoder delay / end-padding metadata set correctly.