How Audio Compression Works — Cheat Sheet

Fora Soft Learn

The four ideas behind every codec, and which idea each codec leans on.

The four ideas

1 · Psychoacoustic masking
Deletes sounds hidden behind louder ones — the ear can't hear them.
Heavy lifter for MUSIC. Lossy.

2 · MDCT (frequency transform)
Turns a slice of waveform into frequencies, with no storage penalty.
The tool that makes masking possible. Lossy setup.

3 · Linear prediction (LPC/CELP)
Models the voice, stores only the small prediction error.
Heavy lifter for SPEECH. Lossy.

4 · Entropy coding
Short codes for common values, long for rare. Throws nothing away.
Shared final stage. LOSSLESS — in every codec.

Which idea does each codec use?

Codec	Masking	MDCT	Lin. pred.	Entropy	Tribe
MP3 / MP2	Yes	Yes	No	Huffman	**Music**
AAC-LC	Yes	Yes	No	Huffman	**Music**
AC-3 (Dolby Dig.)	Yes	Yes	No	Custom	**Music**
Vorbis	Yes	Yes	No	Codebook	**Music**
G.711	No	No	No	None	**Speech**
CELP speech	Light	No	Yes	Varies	**Speech**
Opus — SILK mode	Light	No	Yes	Range	**Speech**
Opus — CELT mode	Yes	Yes	No	Range	**Music**
xHE-AAC (USAC)	Yes	Yes	Yes	Arithmetic	**Both**
FLAC / ALAC	No	No	Yes	Yes	**Lossless**

Read it this way

Transform + masking = a MUSIC codec. Linear prediction = a SPEECH codec.
Both at once = a unifier (Opus, xHE-AAC). Entropy coding is always the last step.
Lossy throws away the inaudible; lossless (FLAC) keeps everything and tops out near 2:1.