In Search of Better Audio Codecs

Requirements and constraints for modern audio codecs on the Internet

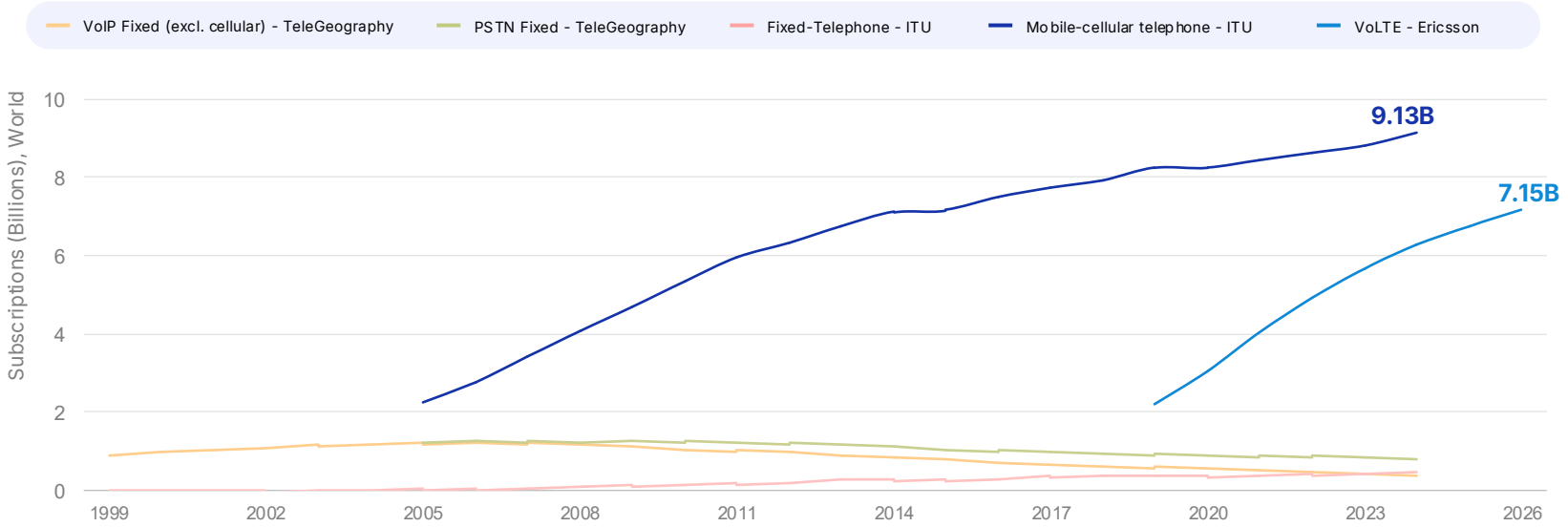
Cullen Jennings

CTO FOR AUDIO/VIDEO COLLABORATION AND AI, CISCO

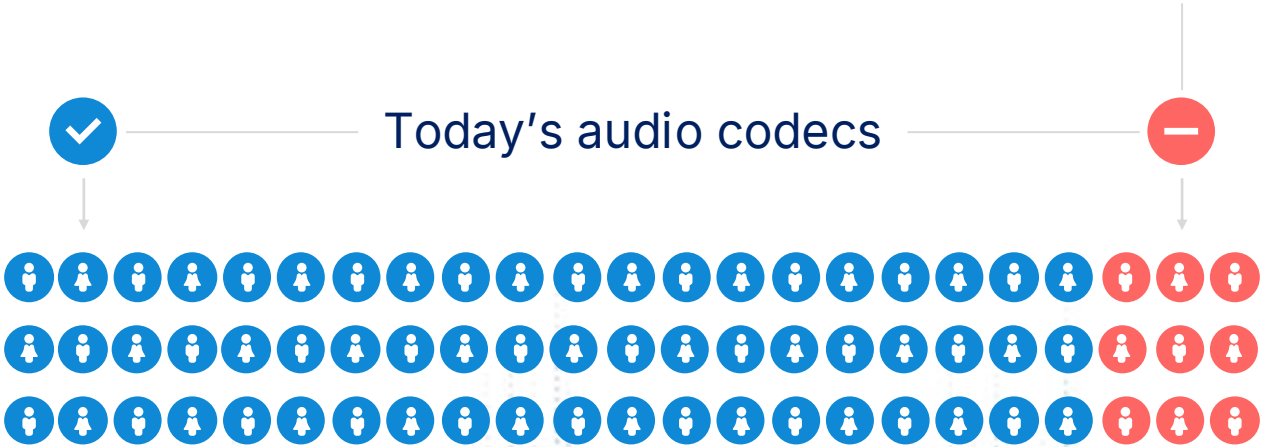
4 MAY 2026

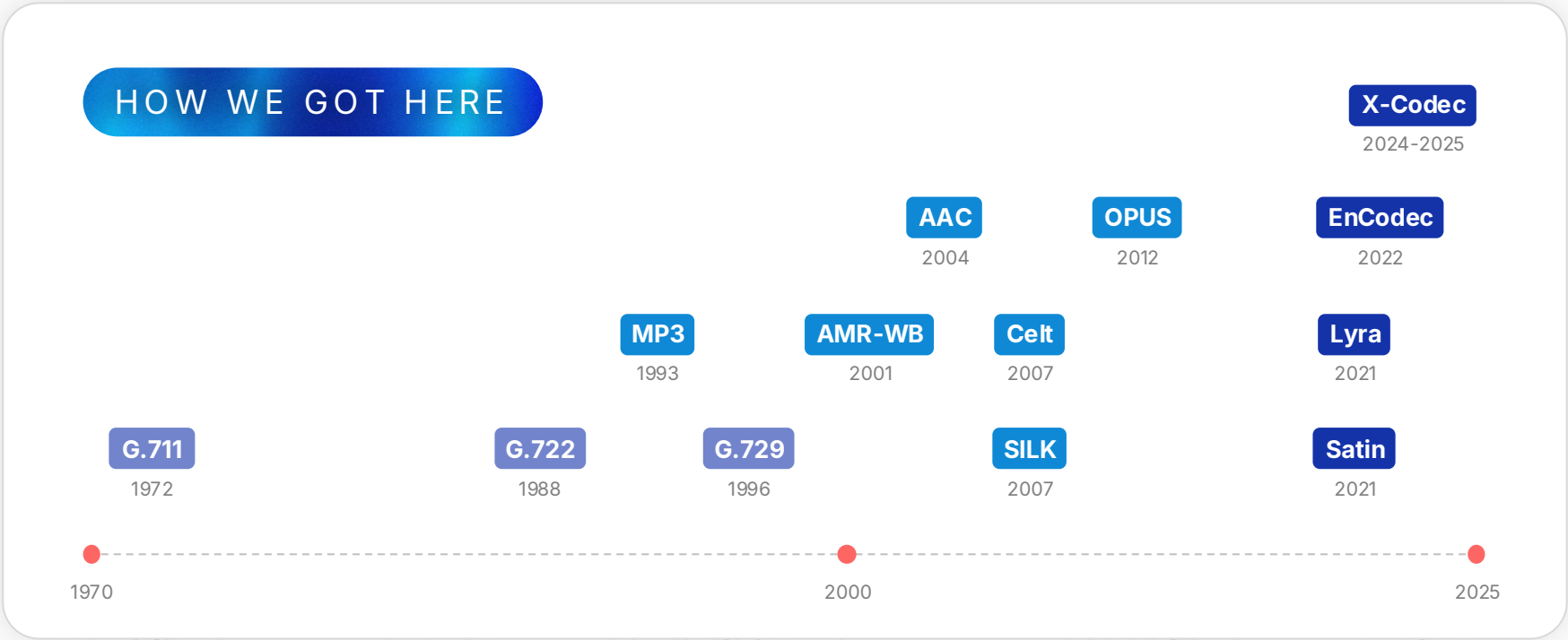
The best audio codec has yet to be built

Rise in VoIP subscriptions



How do we get real-time communication to the last billion?




HOW WE GOT HERE
X-Codec
2024-2025
AAC
2004
OPUS
2012
EnCodec
2022
MP3
1993
AMR-WB
2001
Celt
2007
Lyra
2021
G.711
1972
G.722
1988
G.729
1996
SILK
2007
Satin
2021
1970
2000
2025

REAL-WORLD CHALLENGES

Why we need better codecs

Infinite bandwidth is a myth.

Satellite communication with mobile devices

THE PROBLEM

Power/bandwidth constraints

Limited data rates with small antennas

THE OPPORTUNITY

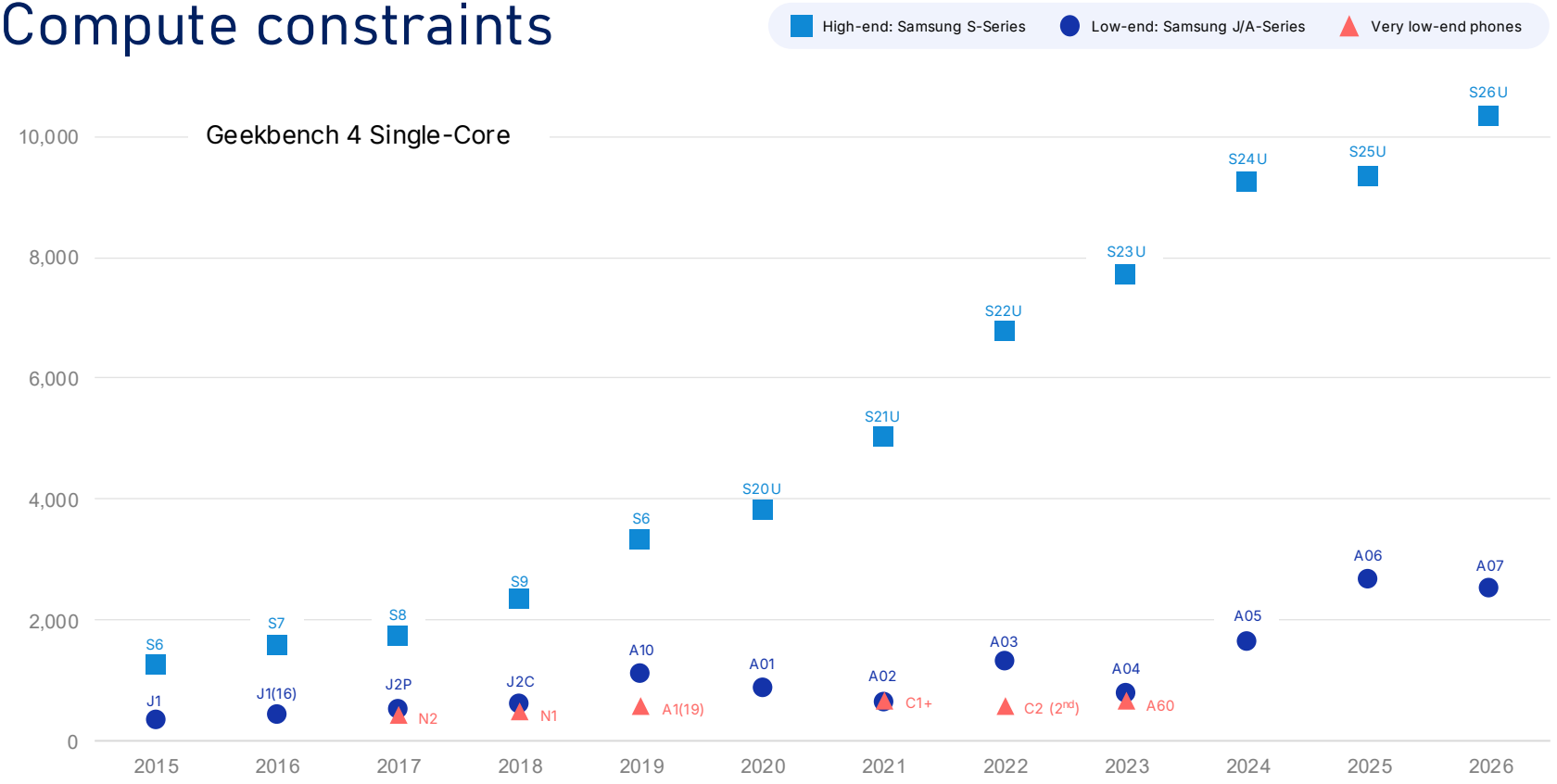
Design a codec that allows audio to work in these settings.

THE IMPACT

Emergency calls from anywhere

More accessibility around the globe

Compute constraints



Compute constraints

4-Core A53 Processor 1.4 GHz

- Benchmarking ≈ 1.5 GMACs per core

Neural Codecs

- DAC: 120 GMACs
- EnCodec: 3 GMACs
- LRAC Challenge Tracks
  - Transparency 0.35 GMACs
  - Enhancement 1.3 GMACs

Neural codecs are changing everything

TRADITIONAL CODECS

ADVANTAGES

Low compute

More uniform performance across input space

Relative robustness to noise and reverb

DRAWBACKS

Poor quality at very low bitrates

Limiting factor for redundancy

NEURAL CODECS

ADVANTAGES

High quality across board (including at ultra-low bitrate)

Universal tokenizers for downstream ML tasks

DRAWBACKS

High compute complexity

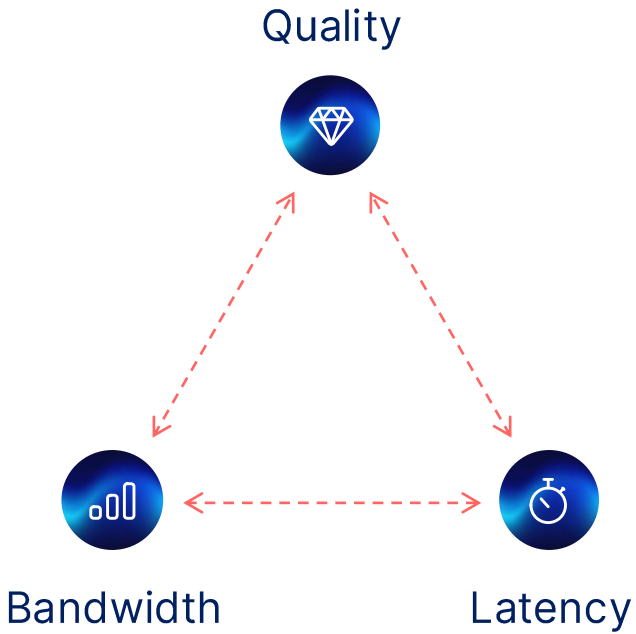
Sensitivity to noise/reverb

Black box (lacks explainability)

Data-driven biases

Training data concerns of neural codecs

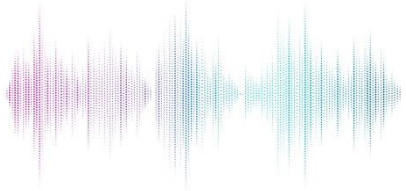
Quality of communication tradeoffs



Real-life consequences

Sources: Schoenenberg, Raake & Koeppe. “Why are you so slow? – Misattribution of transmission delay to attributes of the conversation partner at the far-end” International Journal of Human-Computer Studies (2014).
Brucks, Rifkin & Johnson. “Video-call glitches trigger uncanniness and harm consequential life outcomes.” Nature (2025).

What makes a really good codec?



1. Robustness
2. Faithfulness
3. Bitrate scalability and versatility
4. Redundancy
5. Low latency
6. Standardized

1 Robustness

2 3 4 5 6

Different speech characteristics

Various environmental conditions

Multiple speakers

Consistent results

1 **2 Faithfulness** 3 4 5 6

A codec should sound good,
but it should also sound **right.**

Identity of speaker

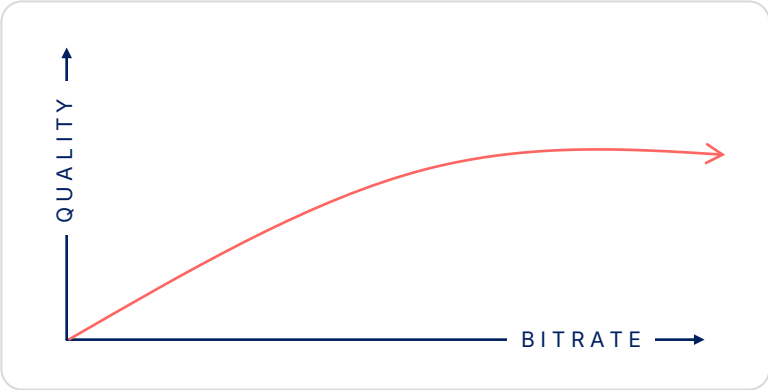
Emotion captured

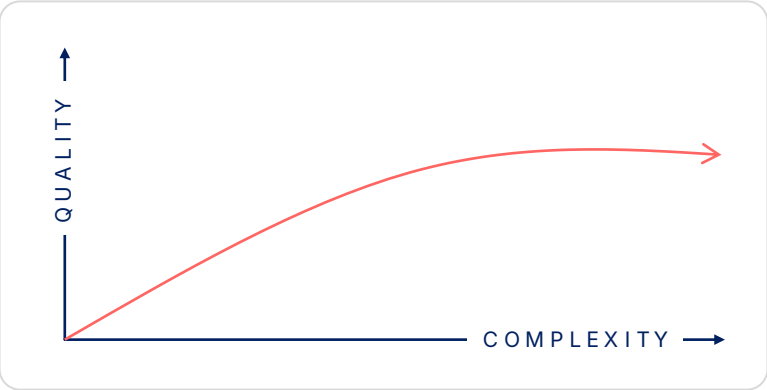
Context of sound

What violates faithfulness?

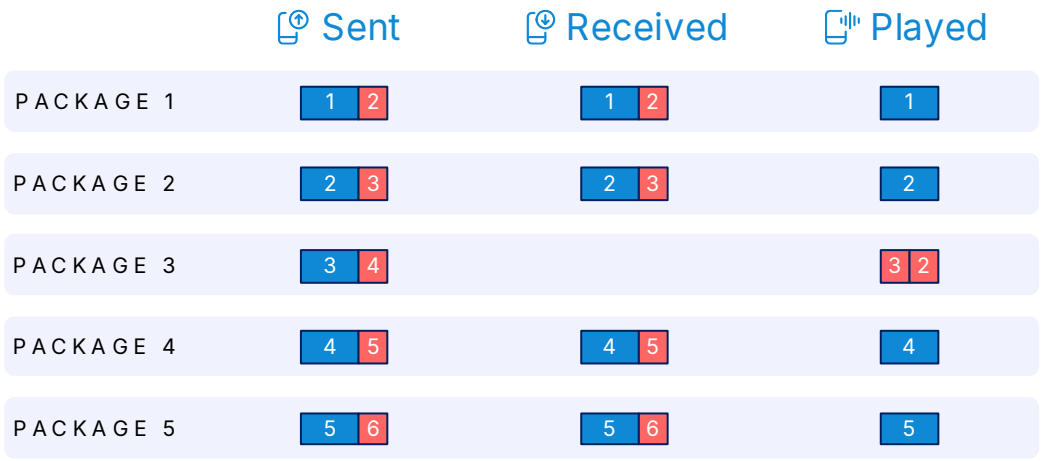
Can it be better than real life?

1 2 **3 Bitrate scalability and versatility** 4 5 6





1 2 3 **4 Redundancy** 5 6



Most successful solution to low-latency packet loss

Low-bitrate codecs can be used for concealment

Need minimal audio artifacts when inserting

1 2 3 4 **5 Low latency** 6

Latency vs. compression is a tradeoff (and a good place for research).

More data enables compression

But more data = more delay

20 ms latency is common

1 2 3 4 5 **6 Standardized**

Patents vs. goal of royalty-free

Rise in proprietary neural codecs

Value of codec is proportional to reach

How do we evaluate codecs?

Two evaluation approaches

OBJECTIVE

Automated algorithms

Fast and repeatable

Use-case dependent

Misses human POV

ML games the metric

vs

SUBJECTIVE

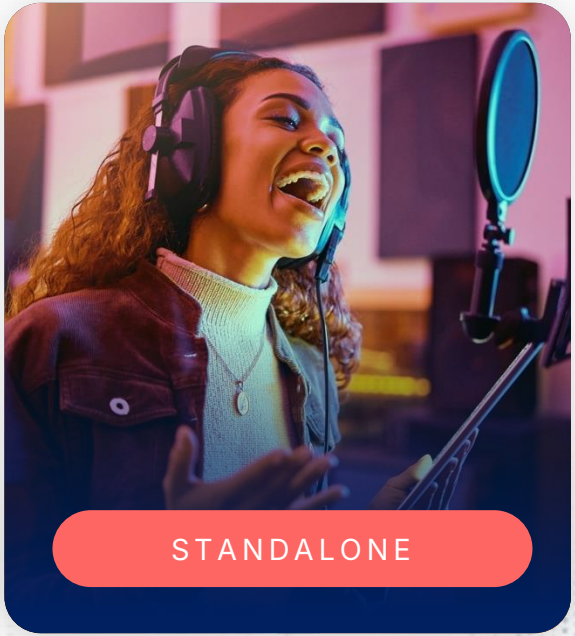
Human listeners

Crowdsourced

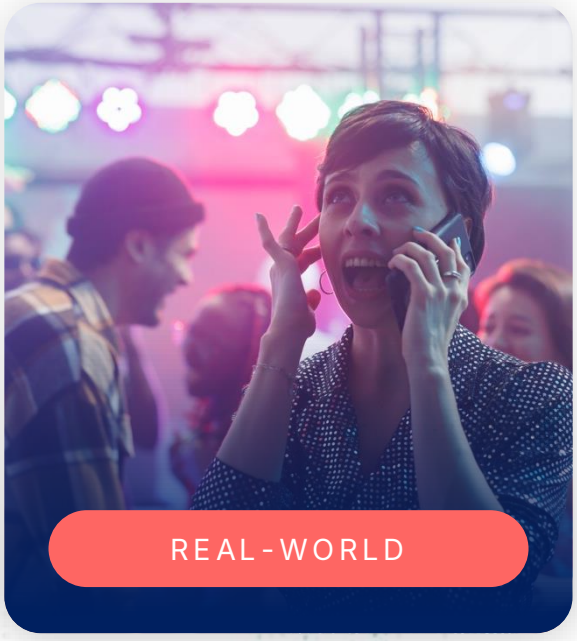
Ideal conditions only

Depends on listener skill

Preference over accuracy



vs



What does good mean?

THEN

Sounds like
the original

vs

NOW

Sounds natural, authentic,
and easy to understand

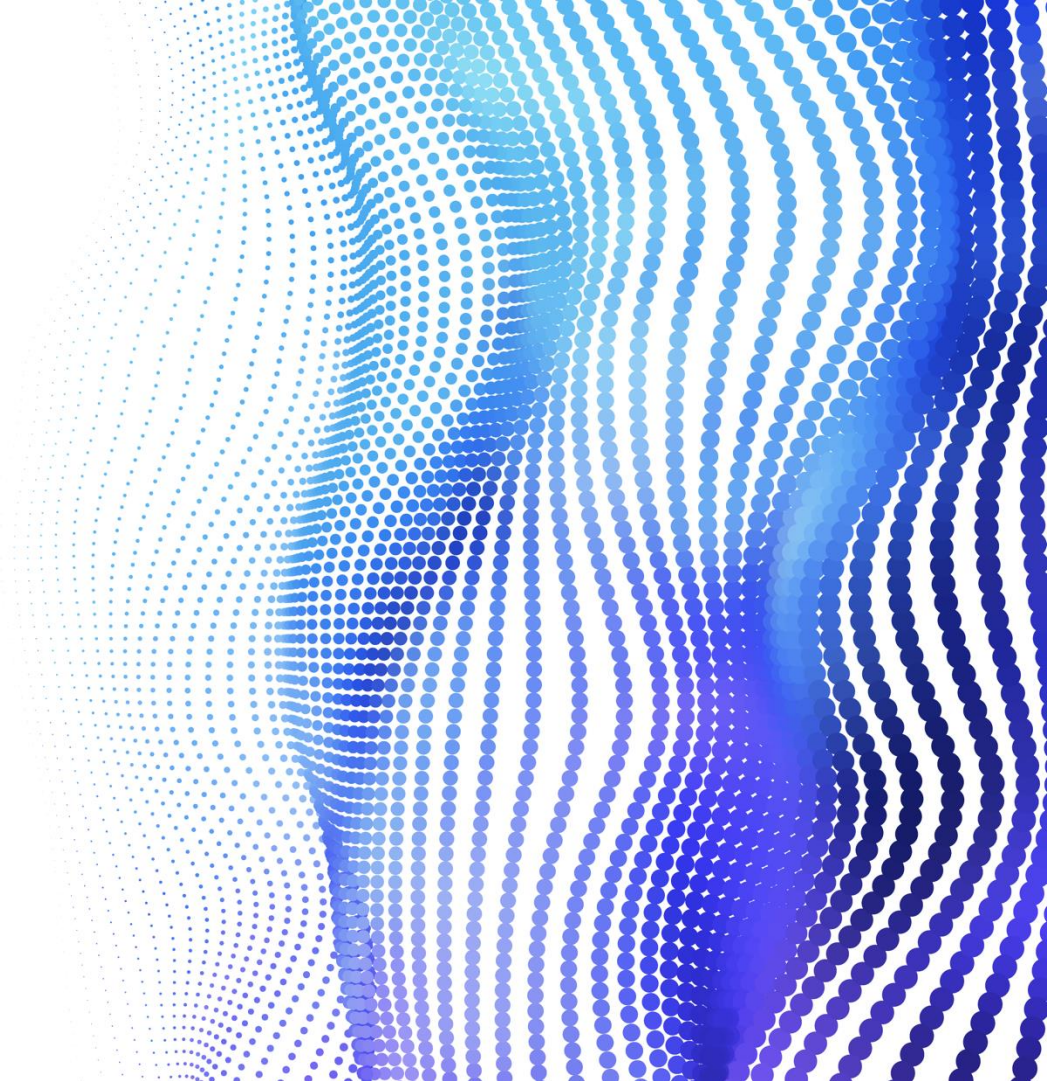
QUESTIONS TO ASK

Did it accomplish the job?

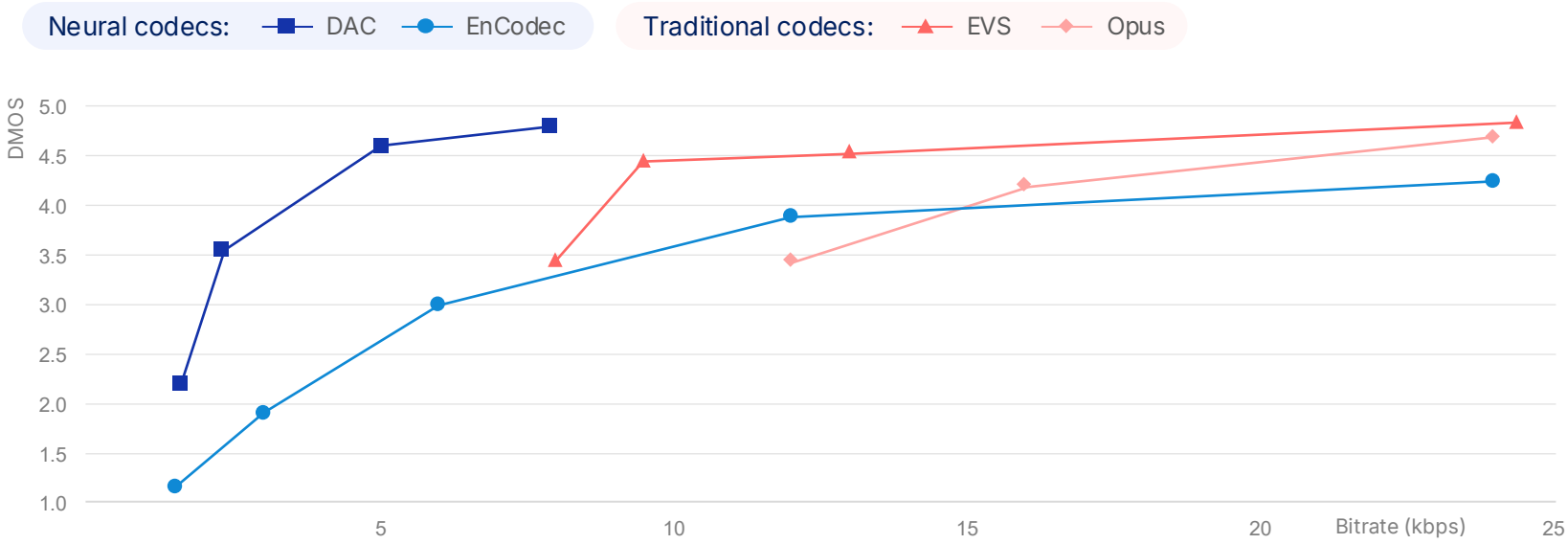
Were all parties heard?

Were there artifacts?

The path forward



Neural vs. traditional codecs



Adapted* from T. Muller et al. "Speech quality evaluation of neural audio codecs," Interspeech 2024.

*Note: effective speech bandwidth of the codecs may differ across data points. Refer to original study for details.

Pursue multiple paths to find the right fit for the challenges ahead.

THE LRAC CHALLENGE

Bringing these issues to the forefront:

- Low compute, bitrate, low latency
- Constrained compute, packet loss
- Shared benchmarks and datasets
- Background noise and reverberation
- Enable codec output to directly drive AI
- Industry and academia collaboration

CISCO

Challenge Results


Q&A