

Group Call Scale - Field Guide

The three regimes, the arithmetic, and the speaker-cap rule - on one page.

Three regimes: 50 / 500 / 5,000 at a glance

PROPERTY	<= 50	~500	5,000+
Architecture	SFU, forward all	SFU, forward active	MCU mix few; cascade
Server decodes audio?	Never	Never	Only active few
Each listener receives	Every voice	Active speakers	One mixed stream
Dominant cost	Client decode	Server fan-out	Server mix CPU
Streams downloaded	1 per talker	1 per talker	Exactly 1 (mix)
Added latency	Lowest	Low	Higher (mix)
Per-speaker control	Full	Full	Lost once mixed
Single-file recording	Separate mix	Separate mix	Mix already exists

Why it explodes: $K \times (K-1)$ voice deliveries

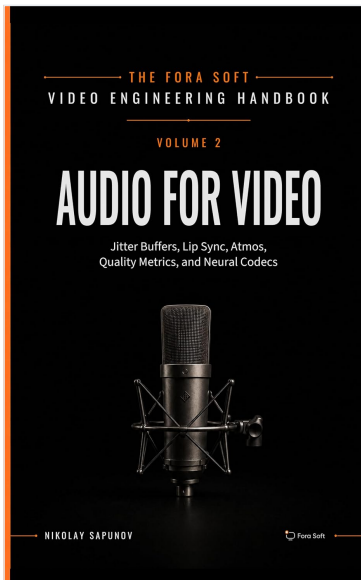
5 people	$5 \times 4 = 20$ voice deliveries.
50 people	$50 \times 49 = 2,450$ - about 122x the 5-person load.
500 people	$500 \times 499 = 249,500$ deliveries if you forward everything.
5,000 people	$5,000 \times 4,999 = 24,995,000$ - quadratic growth: doubling ~4x the work.

The rules that make scale possible

Audio != video	Last-N forwards only top videos; ALL audio is forwarded in small/medium rooms.
DTX	A silent Opus mic drops from ~24-32 kbit/s to ~1 kbit/s - cost follows talkers, not total.
RFC 6464	Loudness stamped per packet (0-127 -dBov); server reads it without decoding.
N-1 mix	Each speaker hears everyone EXCEPT themselves; mix only the loudest 3-6 channels.
Cascade / hybrid	Beyond one server: relay across a mesh, or stream the mix over LL-HLS to passives.

Remember

- ☐ <= 50 people: SFU forwards all audio to everyone; DTX keeps muted mics nearly free.
- ☐ 50-500 people: forward only the active speakers, ranked by RFC 6464 loudness levels.
- ☐ 5,000+ people: mix the loudest 3-6 speakers into ONE stream; never sum all channels.
- ☐ Beyond one server's ceiling: cascade across a mesh, or go hybrid (mix + LL-HLS to passives).
- ☐ Size for the PEAK simultaneous-talker count, not the average - the 'any questions?' spike.
- ☐ Cap simultaneously forwarded/mixed speakers at 3-6; back it with a raise-hand or stage UI.
- ☐ Active-speaker UI flickers when people pause? Suspect Opus DTX gaps, not the algorithm.
- ☐ If almost nobody needs to talk back, it's a broadcast - serve the passive tier over streaming.



GO DEEPER · VOLUME 5 OF 7

Audio for Video: A Complete Guide to Sound in Video Products

From Jitter Buffers and Lip Sync to Atmos, Quality Metrics, and Neural Codecs

By Nikolay Sapunov, CEO at Fora Soft · 1,288 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
amazon.com/dp/B0HH6XPY46

This field guide tells you which regime your room is in and when to stop forwarding and start mixing. Volume 5 picks up where the arithmetic ends: bitrate and topology decisions for audio, jitter buffers and NetEq, loss concealment with FEC and RED, and recording and transcription of large calls.

WHAT YOU WILL FIND INSIDE

- Bitrate and topology decisions for voice
- Loss concealment, FEC and RED
- Diagnosing audio incidents by symptom
- Jitter buffers and NetEq under load
- Recording and transcription of calls

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

Planning a call product that has to hold 500 or 5,000 people?

We will size your SFU and mixer tiers for peak talkers, pick the forwarding rules and show the server cost per hour at each scale. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)