Spatial Conferencing Decision Sheet

Fora Soft Learn

When to build it, who can hear it, and the Web Audio recipe - on one page.

What it buys you (and what it doesn't)

Win	Each voice sits at its on-screen tile. Angular separation lets the brain unmask overlapping talkers - less listening effort on long, crosstalk-heavy calls.
Not	Pure 'immersion' is not the pitch. No benefit in 1:1 calls, town-halls with one speaker, or for users on a single mono speaker.
Rule	Strongest when competing voices are at similar levels - normalize mics first.

Step 1 - can your users even hear it? (device gate)

Needs two channels reaching two ears: wired stereo headphones, wired stereo speakers, or built-in stereo speakers.

Classic Bluetooth drops to mono when the mic is live - no spatial effect. Teams disables it on Bluetooth.

LE Audio Bluetooth can carry stereo on a call - supported as devices ship it.

Detect the output device. If it can't do stereo, fall back to the mono mix silently.

Step 2 - how the platforms do it (2026)

Teams: tile position -> stereo pan. Wired stereo only; no 1:1, no meetings > 100.

Zoom: gallery/immersive seat -> stereo field. Zoom Rooms / Workplace desktop; no wireless.

Google Meet: stereo for presentations; Beam anchors each voice to the life-size 3D person.

FaceTime: HRTF + head tracking via AirPods; optional personalized HRTF from a head scan.

Step 3 - build it on WebRTC (Web Audio API)

One PannerNode per remote voice. panningModel = HRTF (binaural) or equalpower (cheap).

Position = remote tile minus my position. Glide with setTargetAtTime (~20 ms), never snap.

Map gallery columns to azimuth: x = sin(angle), z = -cos(angle).

Use an SFU (each voice = its own track); spatialize per client, not on the server (MCU).

Watch 3GPP IVAS (Release 18): carries 3-4 placed voices at 24.4 / 32 kbit/s in the codec itself.

Pre-ship checklist

- [ ] Output device detected; clean mono fallback when stereo isn't available.
- [ ] Spatial audio off for 1:1 calls and single-presenter town-halls.
- [ ] Positions glide, not snap, when tiles re-lay-out.
- [ ] Microphones normalized so competing voices sit at similar levels.
- [ ] Per-client spatialization on an SFU; not a per-listener server mix.