

TECHNICAL · OCTOBER 2026

Performance

Figures measured on 14 October 2026 on the on-premise profile, with four output languages running at once (Arabic, English, French, German).

End-to-end delay

What is measured: from the moment a sentence ends to its interpreted audio being sent to the phones listening in that language. Delivery over the room WiFi adds milliseconds. Each phone also shows its own total ("x s behind live").

Measurement	Value
Typical delay	Two to four seconds behind the speaker
Median over the measured meeting	3.0 s
Sentences	12 spoken, 48 deliveries (12 × 4 languages)
Losses	0 dropped · 0 not transcribed · 0 not translated · 0 never spoken

The measured meeting used long, formal sentences of up to about 25 words. Shorter, conversational sentences are faster.

Where the time goes (one sentence)

Stage	Time
Speech recognition	~0.6 s
Translation (four languages at once)	~2.2 s
Voice synthesis	~0.4 s

A sentence cannot be translated until it has ended: the speech detector waits for a short silence before handing it on. That wait, plus sentence length, is why long sentences take a little longer.

Start-up

Application window appears	~1-2 s
Ready for the meeting	35-45 s

Capacity

- **Listeners:** the work is done once per language, so adding listeners adds only network traffic. Thirty listeners on one language cost the same as three. The practical limit is the room's WiFi access point.
- **Languages:** up to four output languages at once.
- **Staying close to live:** if a phone falls more than 20 seconds behind, it skips queued sentences – never one already being spoken – and tells the listener. Before that, it speeds speech up slightly to catch up.

Hardware

Graphics	NVIDIA RTX 3090 24 GB minimum · RTX 5090 32 GB recommended
Disk	~60 GB free recommended
Network	One audio stream per listener on the room WiFi