

TECHNIK · OKTOBER 2026

Leistung

Messwerte vom 14. Oktober 2026 im Profil vor Ort, mit vier gleichzeitig laufenden Ausgabesprachen (Arabisch, Englisch, Französisch, Deutsch).

Verzögerung von Anfang bis Ende

Was gemessen wird: vom Ende eines Satzes bis zum Versand seines gedolmetschten Audios an die Smartphones, die in dieser Sprache zuhören. Die Auslieferung über das Raum-WLAN fügt Millisekunden hinzu. Jedes Smartphone zeigt zusätzlich seinen eigenen Gesamtwert („x s hinter live“).

Messung	Wert
Übliche Verzögerung	Zwei bis vier Sekunden hinter dem Sprechenden
Median über die gemessene Besprechung	3,0 s
Sätze	12 gesprochen, 48 Auslieferungen (12 × 4 Sprachen)
Verluste	0 verworfen · 0 nicht transkribiert · 0 nicht übersetzt · 0 nie gesprochen

Die gemessene Besprechung bestand aus langen, förmlichen Sätzen mit bis zu etwa 25 Wörtern. Kürzere, gesprächsnahe Sätze gehen schneller.

Wohin die Zeit geht (ein Satz)

Stufe	Zeit
Spracherkennung	~0,6 s
Übersetzung (vier Sprachen gleichzeitig)	~2,2 s
Sprachsynthese	~0,4 s

Ein Satz kann erst übersetzt werden, wenn er zu Ende ist: Die Spracherkennung wartet eine kurze Stille ab, bevor sie ihn weitergibt. Diese Wartezeit plus die Satzlänge erklären, warum lange Sätze etwas länger brauchen.

Start

Anwendungsfenster erscheint	~1–2 s
Bereit für die Besprechung	35–45 s

Kapazität

- **Zuhörende:** Die Arbeit fällt einmal pro Sprache an; zusätzliche Zuhörende erzeugen nur Netzwerkverkehr. Dreißig Zuhörende in einer Sprache kosten so viel wie drei. Die praktische Grenze ist der WLAN-Access-Point des Raums.
- **Sprachen:** bis zu vier Ausgabesprachen gleichzeitig.
- **Nah an live bleiben:** Fällt ein Smartphone mehr als 20 Sekunden zurück, überspringt es wartende Sätze – nie einen, der gerade gesprochen wird – und informiert die Zuhörenden. Vorher beschleunigt es die Sprache leicht, um aufzuholen.

Hardware

Grafikkarte	Mindestens NVIDIA RTX 3090 mit 24 GB · empfohlen RTX 5090 mit 32 GB
Festplatte	~60 GB frei empfohlen
Netzwerk	Ein Audiostream pro Zuhörenden im Raum-WLAN