

技术 · 2026 年 10 月

性能

以下数据于 2026 年 10 月 14 日在本地配置下测得，同时运行四种输出语言（阿拉伯语、英语、法语、德语）。

端到端延迟

测量内容：从一句话结束，到其传译音频发送到收听该语言的手机为止。通过会议室 WiFi 传输只增加几毫秒。每部手机也会显示自身的总延迟（“落后实时 x 秒”）。

指标	数值
典型延迟	落后发言人两到四秒
实测会议中位数	3.0 秒
句子	说出 12 句，送达 48 次（12×4 种语言）
丢失	丢弃 0 · 未转写 0 · 未翻译 0 · 未播报 0

实测会议使用的是长达约 25 个词的正式长句。更短、更口语化的句子会更快。

时间花在哪里（一句话）

环节	时间
语音识别	约 0.6 秒
翻译（同时四种语言）	约 2.2 秒
语音合成	约 0.4 秒

一句话说完之前无法翻译：语音检测器会等待短暂的停顿再交给下一步。这段等待加上句子长度，就是长句耗时略长的原因。

启动

应用窗口出现	约 1-2 秒
--------	---------

会议就绪	35-45 秒
------	---------

容量

- **听众：** 工作按语言进行一次，增加听众只增加网络流量。同一语言三十位听众与三位成本相同。实际上限取决于会议室的 WiFi 接入点。
- **语言：** 最多同时四种输出语言。
- **保持接近实时：** 如果某部手机落后超过 20 秒，会跳过排队中的句子 – 绝不跳过正在播报的句子 – 并提示听众。在此之前，它会稍微加快语速追赶。

硬件

显卡	最低 NVIDIA RTX 3090 24 GB · 推荐 RTX 5090 32 GB
磁盘	建议约 60 GB 可用空间
网络	每位听众在会议室 WiFi 上占用一路音频流