ETH zürich

Mixed Reality AS 2023
263-5905-00L

# Automatic Transcription & Translation

Alejandro Cuadrón Lafuente[1], Elisa Martínez Abad[1], Ruben Schenk[1], Sophya Tsubin[1]
[1]Institute for Visual Computing, ETH Zurich

## 1. Introduction

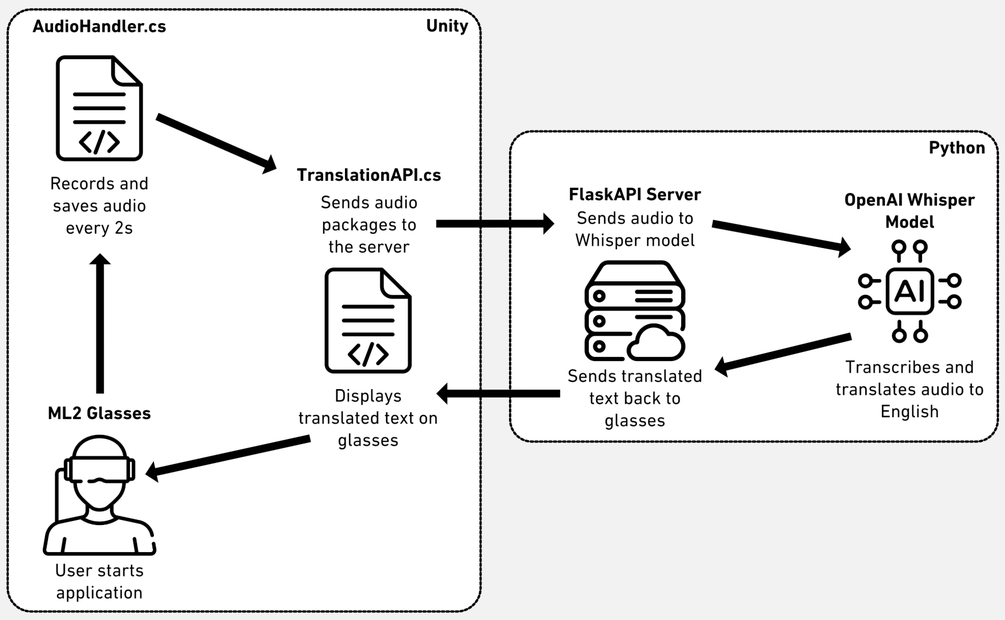
The realm of Mixed Reality (MR) presents a unique opportunity to enhance interpersonal communication across linguistic barriers. Our project aims to change the way we interact in a multilingual environments by integrating real-time translation and transcription into the MR experience.

Using the capabilities of the Magic Leap 2 (ML2) glasses, we have developed an application that captures spoken language, transcribes it, and translates it into English.

## 2. Background

MR applications are just beginning to explore their potential in enabling real-time multilingual communication. Toyama et al. (2014)'s innovation in MR, introducing a head-mounted display with eye gaze-driven text translation, is close to our project's vision of overcoming language barriers through advanced MR and translation technologies.

## 3. Method




Overview of the architecture of the Automatic Transcription & Translation MR application.

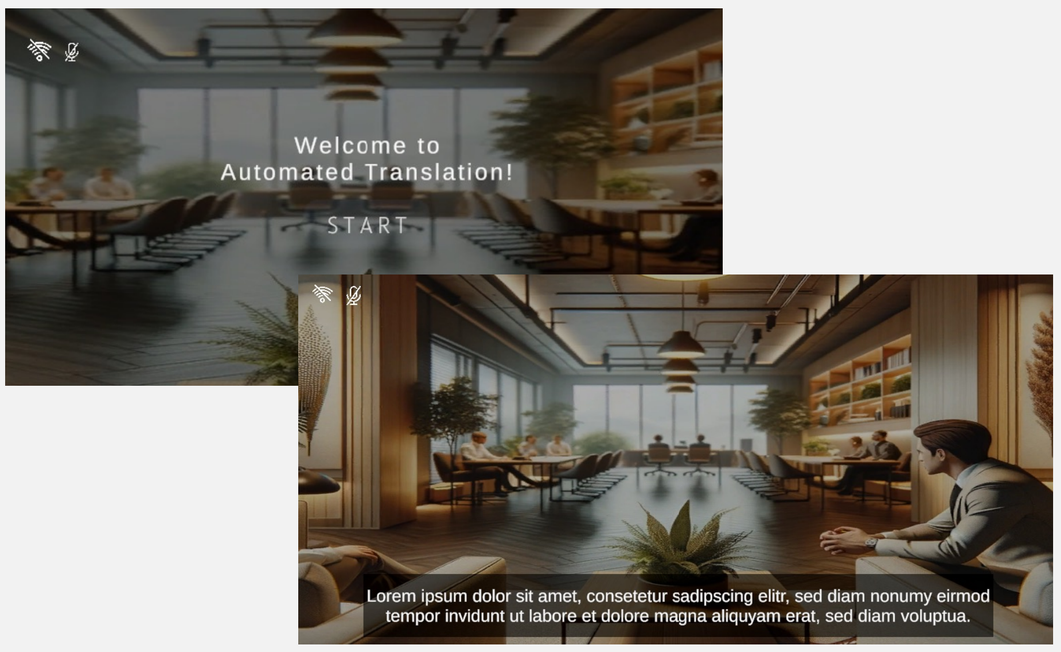
The core of our application's MR experience is built upon the ML2 glasses, chosen for their advanced MR capabilities, which include high-quality visual overlay and responsive user interaction. The integration of these glasses was achieved using **Unity**, a powerful engine for creating immersive MR experiences. Unity not only allowed us to design a user-friendly interface but also facilitated seamless interaction with our backend systems.

On the backend we opted for a **Flask-based API** due to its simplicity and effectiveness in handling HTTP requests. This API serves as the intermediary between the MR glasses and our chosen language processing technology.

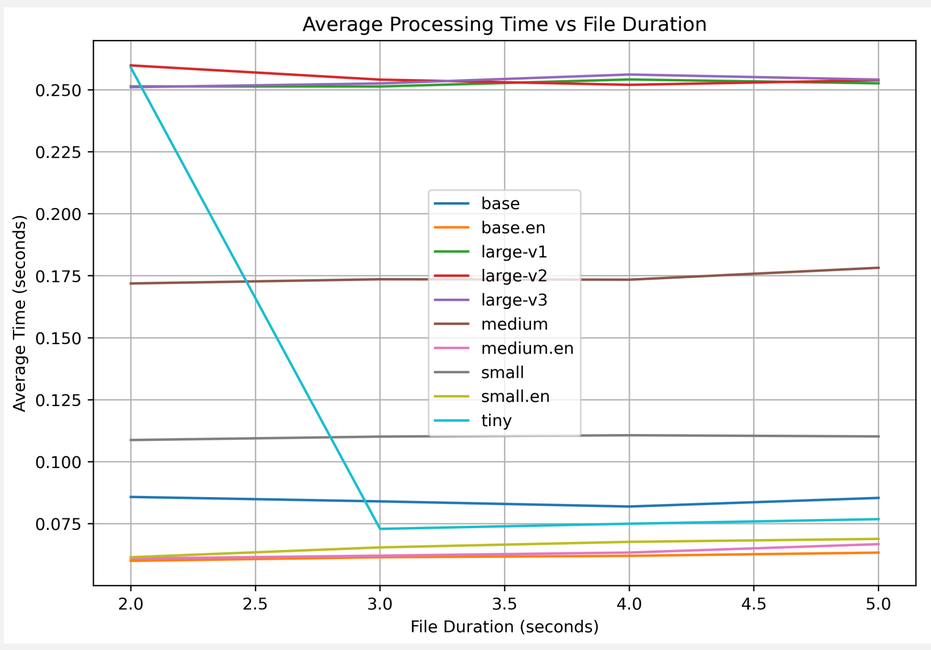
For the crucial task of speech recognition and translation, we integrated **OpenAI's Whisper** model. Selected for its state-of-the-art performance in processing diverse languages and accents, Whisper provides the backbone for our transcription and translation capabilities.

## 4. Results and Discussion

The key result of our project was the successful integration of the translation interface with the ML2 glasses. This achievement demonstrates the feasibility of real-time language translation within an MR environment.




To facilitate the application, we have tested 10 different-sized OpenAI Whisper models on their latency, whereby we noticed that the package size does not significantly increase the translation latency. Meaningful and accurate translations were consistently achieved with models that had a latency of >0.17s.




## 5. Conclusion and Future Work

Our project successfully demonstrates how MR technology can be used to bridge language gaps in real-time. Through the fusion of ML2 glasses and OpenAI's Whisper model, we can translate spoken language directly into an MR environment. This achievement is a step towards more inclusive and accessible communication across different languages.

Future improvements might include:

- Designing a spatially responsive UI
- Implementing speaker localization
- Handling multi-speaker environments

Partner/Sponsor: Magic Leap