Final Project Report

Project Title: VoiceBridge: Real-Time P2P Translation
Team: HackAz Team WildHackers
Course: INFO 511 Foundations of Data Science
Participants: Kanit Mann, Tanishk Singh
Submission Date: March 25, 2025

---

1. Introduction and Inspiration

Language barriers can significantly hinder daily interactions, especially on campus where international students often struggle to communicate effectively. Our personal experiences, such as volunteering at the campus pantry or communicating with maintenance staff, highlighted the gaps in existing translation tools. Conventional solutions often produce inaccurate or contextually inappropriate translations, leading to missed opportunities for meaningful human connection.

VoiceBridge was conceived to address these challenges by providing a real-time translation solution that fosters inclusivity and breaks down language barriers for international students and the broader university community.

Try It Out:

- **GitHub Repository:** https://github.com/kanitmann01/hackaz_team_wildhackers
- **Live Demo:** https://kanitmann01.github.io/hackaz_team_wildhackers/

---

2. Research Focus and Objectives

Primary Focus:
This report centers on the **Research/Formal Reporting** focus. We clearly state a research question, describe our methodology and data processing, and provide detailed analysis and conclusion. In addition, we incorporate supporting elements of data visualization (e.g., screenshots, figures) and data management (e.g., GitHub documentation and a live demo) to present a comprehensive picture of the project.

Research Question:
"How can on-device AI models be integrated into a peer-to-peer architecture to achieve low-latency, accurate, real-time speech translation across multiple languages?"

Objectives:

- Integrate state-of-the-art AI models for speech recognition, translation, and synthesis.
- Develop a modular, streaming architecture that maintains conversation context and minimizes latency.
- Create an intuitive, accessible user interface for real-time translation.

---

3. Methods and Analysis

3.1 System Architecture and Components

VoiceBridge is built with a modular design that splits the translation process into several components:

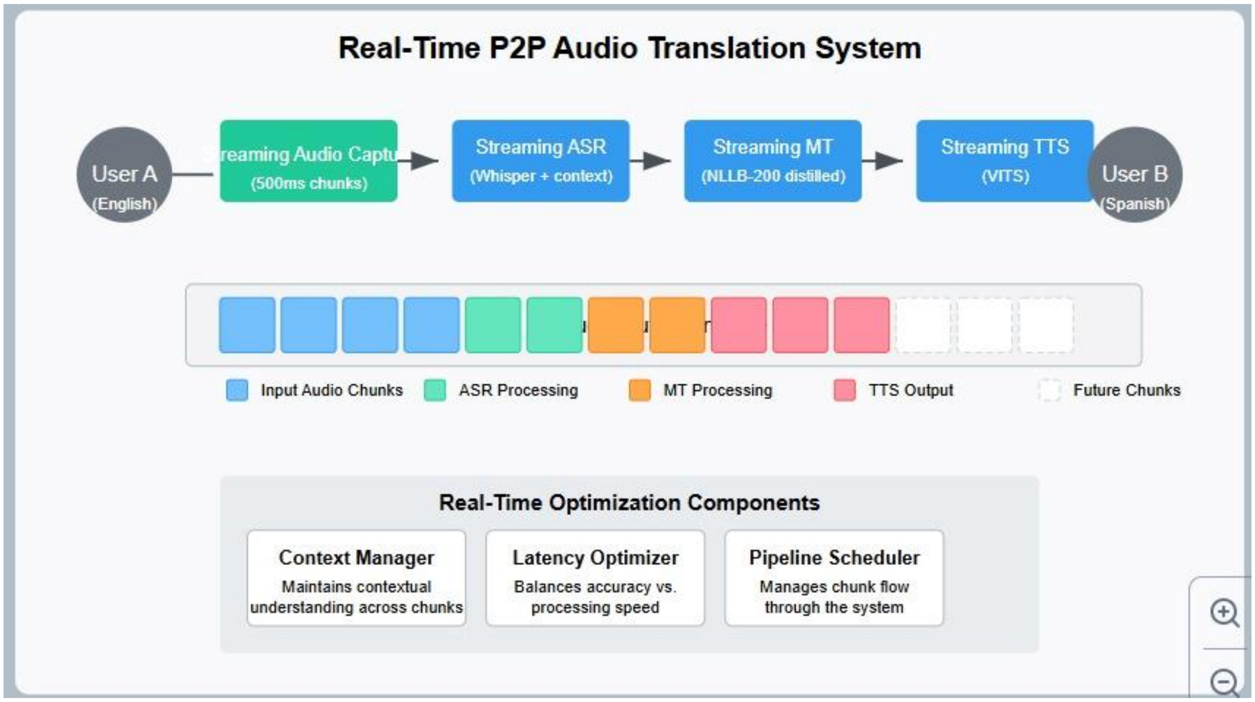
- **Speech Capture:** Audio is captured via a web-based interface using the device's microphone.
- **Automatic Speech Recognition (ASR):** Utilizes a lightweight version of OpenAI's Whisper model to convert spoken language into text. Two instances are configured for each conversation direction (e.g., Hindi to English and vice versa).
- **Machine Translation (MT):** Employs Meta's NLLB-200 model in a streaming translation setup to handle short audio chunks while preserving context.
- **Text-to-Speech (TTS):** Converts translated text back to speech using Meta's MMS-TTS model. Despite challenges with audio quality, multiple fallback approaches and audio amplification techniques were implemented.
- **Real-Time Communication:** A Flask server with Socket.IO handles peer-to-peer connections, ensuring data flows seamlessly between users without heavy server-side processing.

3.2 Data Handling and Preprocessing

Although no new training data was collected (pre-trained models were used), the project involves several key processes:

- **Streaming Audio Processing:** Continuous capture and segmentation of audio for real-time analysis.
- **Context Management:** Maintaining separate text buffers per user to ensure coherent conversation flow.

- **Error Handling and Reset:** Robust procedures to reset contexts when network disruptions or session terminations occur.



- *Figure 1:* Figure Showing Incremental Translation Mechanism

3.3 Implementation and Testing

Implementation:

- **Backend:** Python (Flask, Socket.IO) is used to create the real-time server. Core logic is contained in files such as p2p_server.py.
- **Frontend:** HTML, CSS, and JavaScript form a responsive interface featuring a "hold-to-speak" function.
- **Component Testing:** A diagnostic tool built with Gradio and audio processing via the SoundDevice library ensures each module works as expected.

Testing:
Testing occurred during HackAz, simulating diverse network conditions and multiple language pairings. The streaming architecture allowed for incremental processing, ensuring a smooth and natural conversation flow.

---

4. Project Narrative, Challenges, and Learnings

4.1 What It Does

VoiceBridge enables real-time conversation between speakers of different languages by integrating advanced AI models to handle speech recognition, translation, and synthesis in a peer-to-peer format. The system is designed to be both fast and context-aware, making it a powerful tool for real-world applications.

4.2 Challenges We Encountered

- **Text-to-Speech (TTS) Issues:** The MMS-TTS model sometimes produced silent or low-quality output, prompting the use of fallback approaches and audio amplification.
- **Compute Constraints:** Limited compute power required the use of threading techniques to handle real-time processing.
- **UI Design:** Building a seamless and responsive user interface within tight time constraints proved challenging.

4.3 Accomplishments and What We Learned

Accomplishments:

- **Successful Integration:** We built a fully functioning end-to-end real-time speech translation system.
- **Streaming Architecture:** Our innovative design supports incremental processing, leading to a natural conversation flow.
- **User Experience:** The team developed an intuitive interface that received positive feedback during demos.

Learnings:

- Gained expertise in real-time speech processing pipelines and the integration of multiple AI models.
- Learned the importance of robust error handling and fallback mechanisms.
- Developed valuable skills in WebSocket programming and responsive UI design.

What's Next:

- **Enhanced Language Support:** Extend the system to support 25+ languages.
- **System Robustness:** Improve context management and error correction.

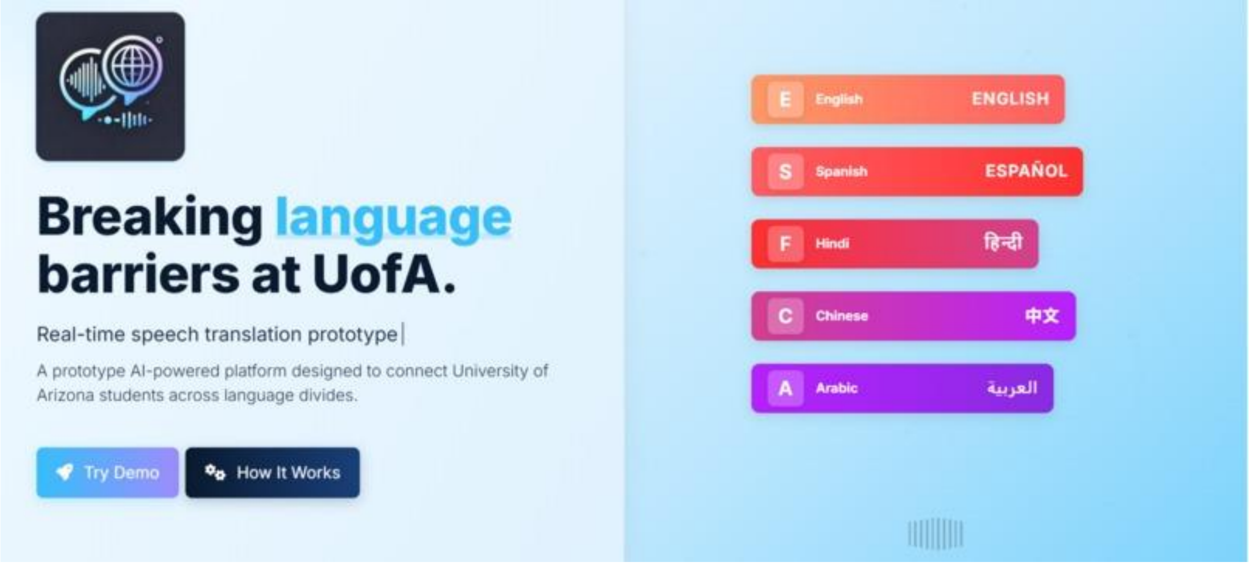
- **Mobile Optimization:** Develop native mobile applications.
- **Offline Mode:** Implement downloadable models for offline use.
- **Security Enhancements:** Introduce end-to-end encryption for increased data privacy.

5.1. Results and Discussion

VoiceBridge successfully demonstrates:

- **Accuracy:** The system provides accurate recognition and translation across multiple languages.
- **Low Latency:** The peer-to-peer architecture ensures minimal delay, preserving the natural rhythm of conversation.
- **Positive User Experience:** Features like "hold-to-speak" and real-time feedback contribute to an engaging translation experience.

5.2. Application Preview

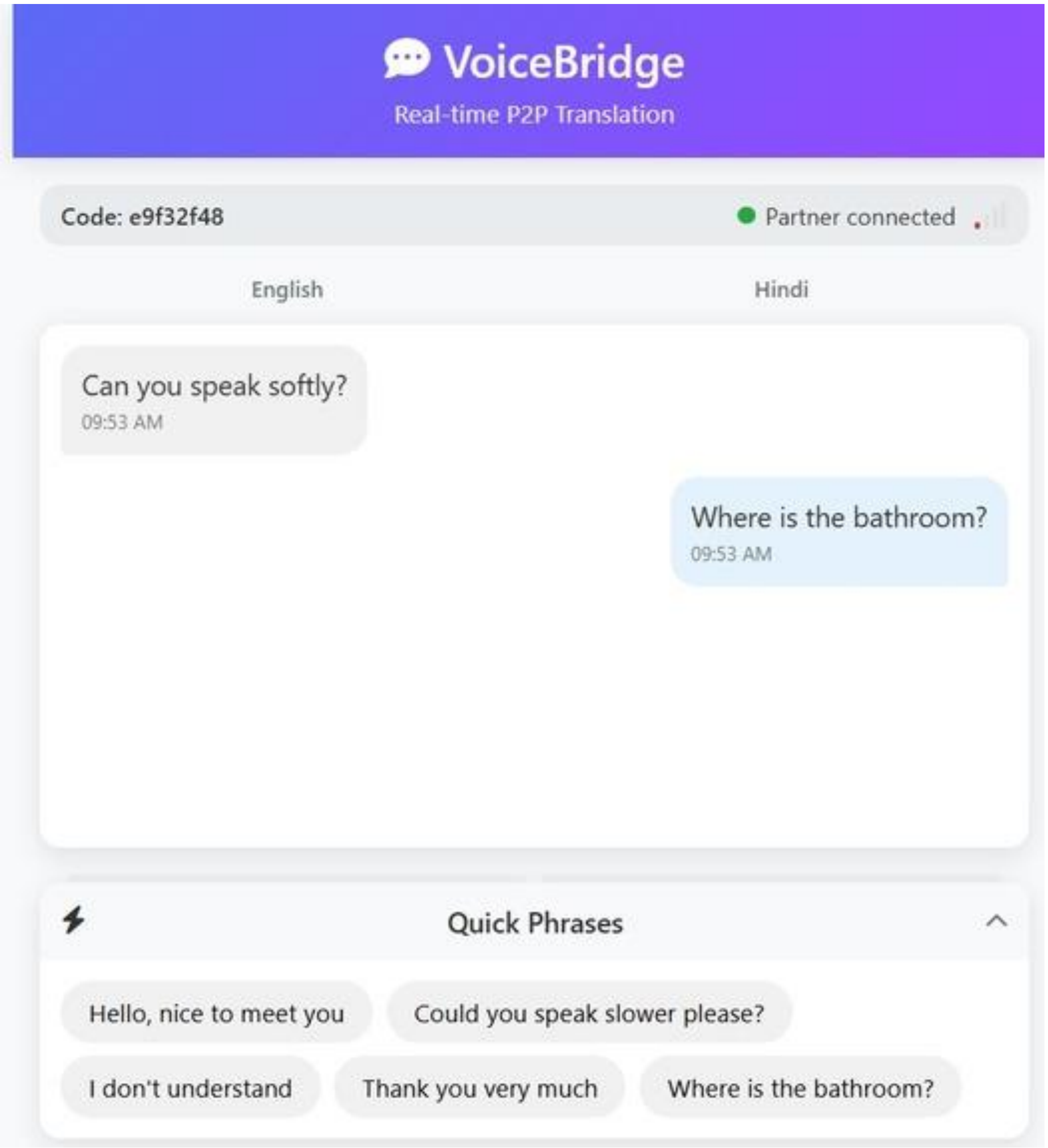


Breaking language barriers at UofA.

Real-time speech translation prototype

A prototype AI-powered platform designed to connect University of Arizona students across language divides.

- *Figure 2:* Figure Showing Landing Page for VoiceBridge



Can you speak softly?
09:53 AM

Where is the bathroom?
09:53 AM

- *Figure 3:* Figure Showing Live Translation Screen

---

6. Conclusion

VoiceBridge bridges language gaps in real time by leveraging state-of-the-art AI models. The project validates the feasibility of on-device, low-latency translation, demonstrating both technical prowess and a commitment to fostering inclusivity. With further enhancements, VoiceBridge has the potential to evolve into a versatile communication tool for a wide range of real-world applications.

---

7. Resources and Appendices

Resources:

- **GitHub Repository:** https://github.com/kanitmann01/hackaz_team_wildhackers
- **Live Demo:** https://kanitmann01.github.io/hackaz_team_wildhackers/

---

8. References

- OpenAI, “Whisper: Robust Speech Recognition via Large-Scale Weak Supervision.”
  GitHub Repository: https://github.com/openai/whisper
- Meta AI, “NLLB-200: Open-Source Machine Translation.”
  GitHub Repository: https://github.com/facebookresearch/fairseq/tree/nllb
- Meta AI, “MMS-TTS: End-to-End Text-to-Speech Synthesis.”
  GitHub Repository: https://github.com/facebook/fairseq/tree/main/examples/mms
- Socket.IO.
  Website: https://socket.io/