CIA\_1  
Unmute – Project Documentation

Avula Nitin Kumar

23112005

**Introduction**

Communication is something most people take for granted. A simple conversation, a quick question in class, or a discussion in the workplace happens effortlessly for those who can hear and speak. But for millions of people who are deaf or mute, every interaction is a challenge. They rely on sign language, yet most of society does not understand it. This gap leaves them excluded in classrooms, offices, and even daily social life.

Unmute is designed to close this gap. It is not just another app; it is an inclusive communication platform that translates sign language to text/voice and text/voice to sign language, in real time. By combining machine learning and cloud computing, Unmute aims to provide a scalable, reliable, and user-friendly solution that empowers people to communicate freely.

**Why This Problem Matters**

**Today, there are tools that partly solve this issue:**

Some apps convert gestures into text, but only for a limited set of static signs.

Speech-to-text apps exist, but they only help one side of the conversation.

A few avatar-based systems can show signs, but they look robotic and do not support natural, real-time conversation.

These solutions lack inclusivity, scalability, and flexibility. They don’t cover multiple sign languages, and they don’t allow smooth two-way communication. Unmute is different because it combines all these needs into one system.

**How Unmute Works**

**Input**

The deaf-mute user makes signs using their hands and facial expressions. A camera captures this.

Alternatively, a non-signing user can type or speak.

**Processing**

Machine learning models interpret gestures into text.

Speech recognition models convert voice into text.

Natural language models ensure that translation is contextually accurate.

**Output**

The sign language is shown as text or even spoken out loud.

The typed or spoken words are shown as animated sign language through a digital avatar.

This back-and-forth flow creates a real-time conversation bridge between both users.

**Role of Cloud Computing**

Cloud computing is at the heart of Unmute’s scalability:

Model Hosting: Instead of running complex ML models entirely on the phone (which would slow things down), models are hosted on cloud servers with GPU support. This ensures quick and accurate recognition.

**Data Storage:** User preferences, conversation history, and training datasets are securely stored in the cloud.

**Real-Time Processing:** Cloud servers handle the heavy lifting, so users with ordinary smartphones can still use the app effectively.

**Scalability:** Whether 100 or 1,000,000 users log in, the cloud ensures the system expands smoothly.

**Machine Learning Techniques**

Unmute relies on advanced ML techniques to make communication natural:

**Convolutional Neural Networks (CNNs**): To detect and classify hand gestures from video frames.

**Recurrent Neural Networks (RNNs) / LSTMs:** To capture continuous sequences in signing, since signs are not isolated but flow together.

**Transformers:** For natural language understanding, ensuring translations are meaningful and not just literal word conversions.

**Speech-to-Text and Text-to-Speech Models:** To support voice interactions.

**3D Animation Models:** To generate lifelike sign language avatars, making the output easier to understand.

**Challenges and How to Overcome Them**

**Diversity of Sign Languages:** American Sign Language (ASL), Indian Sign Language (ISL), British Sign Language (BSL), and others differ significantly.

**Solution:** Build multilingual datasets and use transfer learning to adapt one base model for different languages.

**Gesture Recognition Accuracy:** Lighting, camera quality, and different signing styles affect accuracy.

**Solution:** Train models on diverse datasets and include feedback loops where users can correct outputs.

**Real-Time Latency:** Conversations need instant translation.

**Solution:** Use cloud GPU acceleration and edge computing to minimize delay.

**Avatar Naturalness:** Existing avatars often look robotic.

**Solution:** Integrate advanced 3D modeling libraries and smooth animation techniques.

**Adoption Barriers:** People may be hesitant to use new tools.

**Solution:** Design a simple, intuitive interface and keep the app lightweight.

**What Makes Unmute Different**

**Two-Way Translation:** Most tools focus only on one side. Unmute supports sign → text/voice and text/voice → sign equally.

**Cloud-Powered:** Unlike offline-only apps, Unmute can scale globally without device limitations.

**Context-Aware Translation:** Thanks to ML models, the translation captures meaning, not just words.

**Multi-Language Sign Support:** Extending beyond a single sign language.

**Feedback-Driven Accuracy:** Users themselves can improve the system by providing corrections.

**Development Roadmap**

**Phase 1:** Research and design user-friendly interface.

**Phase 2:** Prototype gesture recognition with CNN and RNN.

**Phase 3:** Add speech-to-text and text-to-speech modules.

**Phase 4:** Build the sign language avatar system.

**Phase 5:** Deploy models on cloud infrastructure and enable real-time chat.

**Phase 6:** Conduct testing with deaf-mute communities and refine.

**Phase 7:** Launch MVP and expand features gradually.

**Conclusion**

Unmute is more than a technology project. It is an attempt to make communication equal for everyone. By combining cloud computing for scalability and machine learning for intelligence, it can overcome the limitations of current systems. The vision is not just to build another app, but to create a tool that brings inclusion, dignity, and independence to millions who have been excluded for too long.