SignNet

American University of Beirut

Department of Electrical and Computer Engineering

EECE 490 Proposal

by

Khaled Nasser

ID: 202304020

Karim Habbal

ID:202204641

Khaled Ammoura

ID:202304327

Hamza Atout

ID: 202301448

October 2024

Contents

1 Motivation

Communication is fundamental to human connection, yet for millions of people with hearing or speech impairments, the lack of accessible communication tools creates a significant barrier. Around 466 million people worldwide have disabling hearing loss, and this number is expected to rise to over 700 million by 2050, according to the World Health Organization (WHO). Many people who are deaf or mute rely on sign language, yet fewer than 2% of the global population is fluent in it. This language barrier can lead to feelings of isolation, exclusion, and difficulty in accessing everyday services.

Without effective ways to communicate, individuals with hearing or speech impairments may avoid social situations, find it challenging to integrate into the workforce, or struggle to engage with public services, contributing to a lower quality of life. Our project seeks to eliminate these barriers by providing a technological solution that bridges the communication gap, empowering people with disabilities to communicate freely and confidently with those who do not know sign language.

2 Existing Solutions

2.1 Competitors:

In recent years, several AI-driven tools have emerged to bridge the communication gap between sign language users and those unfamiliar with it. These technologies, while groundbreaking, often fall short in providing a comprehensive, personalized, and globally accessible solution.

- **Sign-Speak:**
  One of the most notable innovations is the Sign-Speak platform, which utilizes machine learning to translate American Sign Language (ASL) into spoken words and vice versa. The platform also incorporates an AI-powered avatar that can render spoken language back into ASL, allowing for two-way communication in real-time. This system is designed to facilitate communication in various everyday environments, such as restaurants or hospitals, and is scalable due to its cloud-based infrastructure hosted on AWS. However, Sign-Speak is primarily focused on ASL, leaving out many global users who rely on different sign languages. Furthermore, it lacks customization options that could allow users to retain personal elements, such as their own voice tone in translations.
  Amazon Web Services, Inc.

- **Silence Speaks:**
Silence Speaks takes a different approach, focusing on converting sign language into both text and speech. The platform uses advanced AI and 3D motion tracking to capture gestures and convert them into readable and audible communication in real-time. It also allows for the reverse, converting spoken language into sign language to facilitate communication. However, its current implementation is still growing, and it primarily serves British Sign Language (BSL) users, leaving a large portion of the global sign language population underserved. In addition, challenges remain in adapting the system for use in noisy environments or for users with unique signing styles, which can lead to inaccuracies in translation
Silence Speaks

2.2 Critical Gaps:

While these solutions have made strides in the field, several limitations still exist:

- **Limited Sign Language Support:** Both Sign-Speak and Silence Speaks are primarily focused on specific sign languages, leaving millions of users who communicate using other regional sign languages without access to these tools.
- **Customization and Emotional Connection:** Current solutions focus on functional translation but lack the emotional and personal touch of allowing users to maintain their unique voice. For those who have lost their ability to speak, the ability to preserve their voice through personalized samples would be a significant improvement.
- **Accuracy in Complez Settings:** Both platforms struggle with accuracy in complex environments where background noise or variations in signing can interfere with real-time translation.

2.3 Our Approach:

Our platform seeks to address these shortcomings by offering:

- **Support for Multiple Sign Languages:** Expanding beyond just ASL or BSL, we aim to create a solution that can serve a global audience, providing accurate translations for various sign languages.
- **Voice Personalization:** Our system will offer users the ability to integrate their own voice

samples, preserving personal identity and emotional connection, which is particularly meaningful for those who have lost their ability to speak.

- **Improved Real-Time Translation:** Leveraging advanced AI models and data refinement, our platform will improve accuracy even in noisy or complex environments, offering more reliable communication for all users.

By closing these gaps, we hope to create a more inclusive, accessible solution that enables seamless communication and empowers individuals with hearing or speech impairments to engage more fully in society.

3 Project Goal

Our goal is to develop an AI-driven communication platform that uses computer vision to interpret sign language and translate it into audible speech, and vice versa. By doing so, we aim to make communication between individuals who are mute or deaf and the hearing population seamless. This will not only ease daily interactions but also open doors for those with disabilities to engage more fully in society—whether it be in education, work, or social settings.

In addition to translating sign language, our model will allow users to integrate their own voice samples, helping those who once had the ability to speak maintain their unique voice even after losing it. This personalized feature not only enhances the user experience but also adds an emotional connection for the individual and those communicating with them. The potential applications are broad, ranging from daily conversations to more sensitive uses, such as hearing the voice of a loved one who has passed.

By achieving these goals, we aim to enhance accessibility and ultimately ensure that no one feels isolated due to communication barriers.

4 Methodology

4.1 Data

The data we use may vary, and we may need to incorporate multiple sources due to its fragmented nature. Some datasets are exclusive to numbers, and others are exclusive to alphabets. However, our main source of data will be the WLASL (Word-Level American Sign Language) dataset which is a public dataset built for research and non-commercial purposes. This specific dataset has been used by many for sign-language recognition, Papers and Conferences, and even benchmarking between systems. We specifically selected this dataset due to its **scale and variety**, as it includes over 21,000 unique videos performed by multiple signers. This introduces variation in style and motion, all in all to train on 2,000 different ASL signs.

4.2 Models and Techniques

We plan to utilize the following machine learning models and techniques for our system:

- **Convolutional Neural Networks (CNNs)**: We will use CNNs for their ability to effectively extract spatial features from video frames, which is essential for understanding the visual aspects of sign language. Their success in image classification tasks makes them suitable for recognizing hand shapes and facial expressions.
- **Recurrent Neural Networks (RNNs)**: Specifically, we will incorporate Long Short-Term Memory (LSTM) networks to model the temporal dynamics of sign movements. LSTMs excel in sequence prediction tasks, allowing us to capture the continuity and context of signs over time.
- **Two-Stream Networks**: This architecture will help us process spatial and motion information separately—one stream for static hand shapes and another for temporal changes. This dual approach enhances our model's capability to interpret both the form and movement of signs, making it more robust in recognizing variations.
- **Attention Mechanisms**: Implementing attention mechanisms will allow our model to focus on the most relevant parts of the video frames during prediction. This selective focus improves the accuracy of recognizing important features in sign language.
- **Evaluation Metrics**: We will evaluate model performance using metrics like accuracy, precision, recall, and F1 score to ensure our system meets the necessary standards for practical use.

These methods are chosen for their proven effectiveness in similar tasks and their ability to address the unique challenges of translating sign language to text and vice versa. By combining these approaches, we aim to create a robust system capable of handling the complexities of sign language recognition.

5 Plan and Timeline

In this section, we outline a plan covering the expected progress of the project over the semester.

1. **Project Planning & Research (October 21 - October 27):**
   - Finalizing project scope & requirements
   - Researching datasets
   - Studying relevant techniques
2. **Dataset Acquisition & Preprocessing (October 28 - November 3):**
   - Acquiring data sets needed
   - Preprocessing the data for training
3. **Model Development for Sign Recognition & Voice-to-Text (November 4 - November 17):**
   - Implementing and training computer vision model for sign recognition (2 members)
   - Implementing and training voice-to-text model (2 members)
4. **Model Development for Text-to-Voice & Custom Voice Generation & Text-to-Sign Language Generation (November 18 - December 1):**
   - Implementing text-to-voice models and custom voice generation (2 members)
   - Implementing text-to-sign language models using animations,avatars,etc... (2 members)
5. **System Integration (December 2 - December 10)**:
   - Integrating all components of the project
6. **Final Testing & Submission Preparation (December 11 - December 15)**:
   - Conducting final testing to ensure robustness and accuracy
   - Finalizing project report and documentation