Final Year Proposal

Hand Gestures-to-Speech: An Android Application

Abdullah Khalil and Zia Ul Hassan
01-134212-007, 01-134212-199

Supervisor:
Dr. Arif Ur Rahman

Department of Computer Science
Bahria University, Islamabad

September 20, 2024

Abstract

Individuals with speech impairments often face significant challenges in communicating effectively, leading to social isolation, frustration, and difficulties in both personal and professional interactions. Current communication aids are often limited in functionality or accessibility, leaving a gap in providing real-time, intuitive solutions. This proposal presents the development of a mobile application that aims to address these challenges by converting hand gestures into speech. Leveraging machine learning algorithms for gesture recognition, the application utilizes TensorFlow and OpenCV for real-time video processing. By integrating Google's Text-to-Speech API, it translates recognized gestures into audible speech, enabling more seamless communication. The app will be built with a user-friendly interface compatible with Android using Flutter. Through extensive evaluation, including field tests with target users, the project seeks to ensure high accuracy and user satisfaction. This solution not only enhances communication for individuals with speech impairments but also fosters greater inclusivity in society.

Contents

List of Figures

1 Introduction

The ability to communicate effectively is a vital aspect of human interaction,effective communication is essential for human interaction, yet many individuals face barriers due to speech impairments. Our project, "Hand Gesture to Speech Conversion Application," aims to address these communication challenges by developing an innovative mobile application that translates hand gestures into speech. This technology will serve as a communication bridge for people who are unable to speak, thus providing an assistive tool that enhances their ability to interact with the world around them. The system will leverage advanced machine learning, computer vision, and natural language processing technologies to ensure the accurate and seamless conversion of hand gestures into audible speech.By leveraging these technologies, the project seeks to create an assistive tool that empowers users and facilitates better communication. The rising population of hearing-impaired individuals in Pakistan faces significant communication barriers that hinder their ability to interact with society[5]. Many of these individuals struggle to express themselves effectively, which limits their access to education, employment, and social opportunities. Our project, "Hand Gesture to Speech Conversion Application," aims to address these challenges by providing an innovative solution that translates hand gestures into speech. By leveraging advanced technologies, we seek to empower hearing-impaired individuals and enhance their ability to communicate, fostering greater inclusivity in society.

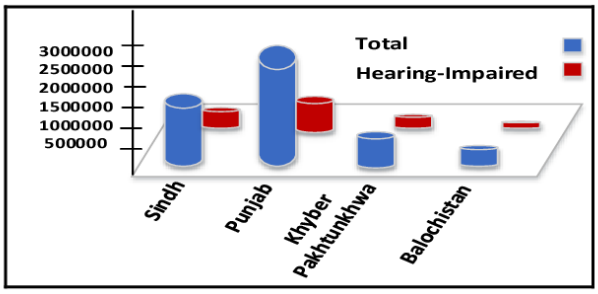

Figure 1: Hearing-impaired people in Pakistan

2 Literature Review

Effective communication is essential for human interaction, yet individuals with speech impairments often encounter significant barriers. Various studies have explored the development of systems that convert hand gestures into speech, highlighting the potential of gesture recognition technologies in facilitating communication for the speech-impaired community. Recent advancements in machine learning and computer vision have led to the creation of applications that utilize Convolutional Neural Networks (CNNs) for gesture recognition. These systems often employ real-time video processing techniques to identify hand movements and translate them into text or speech outputs. For instance, a study highlighted a two-stage deep CNN architecture that improves hand gesture recognition performance by integrating color images with pseudo-depth information. Moreover, the integration of Natural Language Processing (NLP) techniques allows for the formation of coherent sentences from recognized gestures, further enhancing communication effectiveness. The use of Google's Text-to-Speech API in these applications ensures that users receive audible feedback, making interactions more seamless. Despite these advancements, challenges remain, including achieving high accuracy in diverse environments and ensuring real-time processing on mobile devices. Addressing these challenges is crucial for the widespread adoption of gesture recognition technologies in assistive applications.

3 Objectives

The main objectives of the project are:

- Develop a mobile application capable of converting hand gestures into speech using machine learning models.
- Integrate real-time gesture recognition through computer vision (using OpenCV).
- Implement Google's Text-to-Speech API for converting recognized gestures into speech.
- Ensure the application is cross-platform compatible (Android and iOS) using Flutter.
- Provide a user-friendly interface tailored for individuals with speech impairments.

4 Methodology

The methodology for developing this project will be broken down into the following stages:

4.1 System Design

The system will be designed with four core components:

1. **Gesture Recognition:** TensorFlow will be used to train models for identifying hand gestures from video input [1].
2. **Video Processing:** OpenCV will handle real-time video input, detecting and tracking hand movements.
3. **Speech Synthesis:** Google's Text-to-Speech API will convert recognized gestures into speech.
4. **Mobile Application Development:** The front-end will be developed using Flutter to ensure compatibility across Android and iOS platforms.

4.2 Evaluation

The performance of the system will be evaluated based on the accuracy of gesture recognition, the speed of real-time processing, and user satisfaction. Testing will involve both automated unit tests and field tests with individuals from the target user base (people with speech impairments).

5 Project Scope

5.1 In-Scope

- The primary focus of this project is to develop a mobile application that assists individuals with speech impairments in converting hand gestures into spoken language using Pakistani Sign Language.
- The application will leverage **Convolutional Neural Networks (CNNs)** for gesture recognition, recognizing hand-signs using **OpenPose's Pakistani Sign Language dataset** for accurate predictions.
- The application will utilize **Natural Language Processing (NLP)** techniques to convert recognized hand gestures into coherent sentences, facilitating meaningful communication.

- The app will use **Google's Text-to-Speech (TTS) API** to provide audio output for the translated sentences, ensuring a seamless conversation experience.
- The app will be developed using **Flutter** to support both **Android** and **iOS** platforms, offering a unified codebase and cross-platform compatibility.
- **OpenCV** will be used for the computer vision tasks involved in hand-sign detection and recognition.

5.2 Out of Scope

- The application will not include support for languages other than Pakistani Sign Language (PSL) in its initial version.
- The project will not focus on sign language interpretation beyond hand gestures, such as facial expressions or body posture, which also play a role in full sign language communication.
- Advanced features like offline mode for machine learning inference or TTS functionality may not be supported in the first release due to resource constraints.
- Any hardware development such as specialized cameras or gloves to enhance gesture recognition will not be included.

6 Feasibility Study

6.1 Technical Challenges

- **Gesture Recognition Accuracy**: Achieving high accuracy in recognizing gestures from the Pakistani Sign Language dataset using CNNs may be challenging, as variability in hand movements and environmental factors (lighting, background, etc.) can affect performance[3].
- **Real-time Processing**: Ensuring the application processes hand gestures in real-time with minimal latency is critical. Optimizing the CNN and handling heavy computations on mobile devices without affecting user experience will require attention.
- **Cross-platform Compatibility**: Ensuring smooth performance and identical user experience across both Android and iOS platforms may present challenges due to differences in hardware and system architecture.
- **Integration of NLP and TTS**: Seamlessly integrating gesture recognition with NLP for sentence formation and using Google's TTS API without significant lag will require efficient data flow and processing pipelines[4].

6.2 Market Adoption

- The app targets a niche market, particularly individuals with speech impairments who rely on sign language. Market adoption could be enhanced by the increasing accessibility of smartphones and a growing demand for assistive technology in healthcare and education.
- There is an opportunity to gain traction in countries where PSL is prevalent, as there are limited mobile solutions specifically catering to the speech-impaired population using this sign language.
- Adoption could be slower if the app lacks features such as support for multiple sign languages or advanced gesture recognition, which might limit its global applicability.

6.3 Computing Resources

- **Mobile Devices**: The app will rely heavily on the processing capabilities of mobile devices. The CNN model for gesture recognition will need to be optimized for on-device performance or cloud-based processing if necessary[2].
- **Cloud Services**: Google's Text-to-Speech API will be leveraged, requiring internet connectivity. Cloud services might also be required for training or updating the CNN model, though on-device inference will be prioritized to minimize latency.
- **Model Training**: Significant computing resources will be required to train and fine-tune the machine learning model using the PSL dataset. However, the app will primarily focus on using pre-trained models for inference to reduce computational demands on mobile devices.

7 Tools & Technology

- **Flutter**: Chosen for its ability to develop cross-platform mobile applications for both Android and iOS using a single codebase, reducing development time and costs.
- **TensorFlow**: Will be used for training the CNN model to recognize hand gestures from the PSL dataset. TensorFlow Lite will be explored for efficient on-device machine learning inference.
- **OpenCV**: For image processing and computer vision tasks, specifically for recognizing hand gestures in real-time via the mobile camera.
- **Google Text-to-Speech API**: This will convert the recognized gestures, once translated into sentences, into speech output, providing the core functionality of the app.
- **Firebase**: Considered for backend services like authentication, real-time data storage, and cloud-based machine learning model updates if needed.
- **OpenPose**: Provides the dataset for Pakistani Sign Language, essential for training the hand-sign recognition model.
- **Dart**: The programming language used in Flutter for building the mobile application.

8 Timeline

Figure 2 outlines the timeline for the completion of the project.

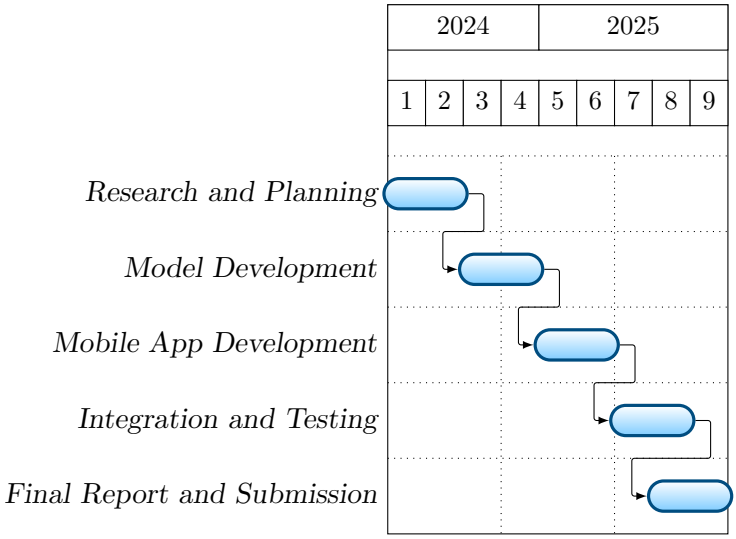

Figure 2: Time Line

9 Conclusions

this project aims to develop a mobile application that significantly enhances communication for individuals with speech impairments by translating hand gestures into audible speech. Through the integration of advanced machine learning, computer vision, and natural language processing technologies, the app is designed to recognize hand gestures in real-time and convert them into coherent sentences using Google's Text-to-Speech API. The project demonstrates a strong potential to address the communication challenges faced by individuals relying on Pakistani Sign Language, fostering inclusivity and providing a practical, assistive tool. While technical challenges related to gesture recognition accuracy and real-time processing remain, careful design and optimization strategies will ensure the application's success. This innovative solution stands to make a meaningful impact in improving communication accessibility for the speech-impaired community, contributing to a more inclusive society.

References

[1] Deep R. Chavan and Maheshkumar H. Kolekar. Conversion of hand gesture to speech for speech impaired people. In *2022 IEEE 19th India Council International Conference (INDICON)*, pages 1–5, 2022.

[2] Bo Guan, Jin Cao, Xingqi Wang, Zhuoyue Wang, Mingxiu Sui, and Zixiang Wang. Integrated method of deep learning and large language model in speech recognition. In *2024 IEEE 7th International Conference on Electronic Information and Communication Technology (ICEICT)*, pages 487–490, 2024.

[3] Ahmed Kasapbaşi, Ahmed Eltayeb Ahmed Elbushra, AL-HARDANEE Omar, and Arif Yilmaz. Deepaslr: A cnn based human computer interface for american sign language recognition for hearing-impaired individuals. *Computer methods and programs in biomedicine update*, 2:100048, 2022.

[4] Catalin Ungurean and Dragos Burileanu. An advanced nlp framework for high-quality text-to-speech synthesis. In *2011 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 1–6. IEEE, 2011.

[5] Muhammad Wasim, Adnan Ahmed Siddiqui, Abdulbasit Shaikh, Lubaid Ahmed, Syed Faisal Ali, and Fauzan Saeed. Communicator for hearing-impaired persons using pakistan sign language (psl). *International Journal of Advanced Computer Science and Applications (IJACSA)*, 9(5), 2018.