Bilkent University

Department of Computer Engineering

Senior Design Project

Vox

Project Specification Report

Group Members: Furkan Başkaya, Atakan Sağlam, Berk Kerem Berçin, Oğulcan Çetinkaya, Sarp Ulaş Kaya

Supervisor: Shervin Rahimzadeh Arashloo

Jury Members: Dr. Shervin Arashloo and Dr. Hamdi Dibeklioglu

Project Specifications Report

October 11, 2021

This report is submitted to the Department of Computer Engineering of Bilkent University in partial fulfillment of the requirements of the Senior Design Project course CS491/2.

1 Introduction

Sign Language is a nonverbal language that people with hearing and speaking disabilities use to communicate with the rest of the world. It is built upon the visual use of hands, eyes, face and mouth. It is a complete, natural way of communication that shares similar linguistic properties as spoken languages. Sign language is adapted by deaf communities as a useful means of communication [1]. It is an integral part of their lives. According to an article on HRW there are more than 70 million people around the world that depend on sign language to communicate [2]. With the help of this technique these people are able to learn, work and be included in their local communities. It helps to ensure disabled people live a life on an equal basis with non-disabled people.

Even though deaf people use sign language to communicate with other people, the majority of the population don't know how to use sign language and this creates a communication barrier between the deaf community and the rest of the population. According to Human Rights Watch's research on deaf people's rights, lack of awareness on sign language causes deaf people to struggle while accessing public services all around the world. Deaf communities are excluded from essential parts of life such as: health, education, justice because of face to face communication problems [3]. Only a small fraction of these services have reasonable adjustments for disabled people to benefit from them. A recent study led by the British Deaf Association states some of the key areas affecting the deaf community with their relationship with the public. They are lack of awareness in booking, insufficient interpreters, and a lack of understanding about deaf people and their levels of written English [4]. As a team we are determined to meet the needs of the deaf community.

One of the most efficient ways to communicate with someone who doesn't speak the same language with you is to use an interpreter. You can pay someone to translate the language for you but always traveling with someone to interpret for you is not a sufficient solution. Instead you can use interpreter applications available on your phone. These kinds of applications are very common for spoken languages like English, Turkish etc. but it is not the case with sign language. Our project idea started out when we realized the communication gap between the deaf community and non disabled people, and the lack of sign language to spoken language interpreters. We would like to address this problem by providing a service where we will track the hand movements and gestures of the user and then process them to convert sign language to spoken language in real-time.

In our project specification report we are going to briefly describe the project, go over the constraints we will follow during development such as: economical constraints, timing constraints and social constraints. Moreover, the professional and ethical issues about the project will be discussed. Lastly, the functional and non-functional requirements will be introduced.

1.1 Description

Vox is a service which aims to provide real-time interpretation between deaf community and the rest of the world. Our name represents what we provide for our users and it is Latin for voice. Our application will become the voice of the deaf and the mute communities. The main goal of this application is to become an American Sign Language (ASL) recognition engine and capture sign language movements of the mute and translate them to spoken language in real-time using nothing but a simple camera. With this technology we can provide a face to face communication experience which can continue at the same pace as spoken communication.

This application will bring an ASL-to-English interpreter with the touch of a button. It will track the hand movements and facial gestures and assess English words to signs. Currently this service is not officially provided by any company on the market. There are prototypes which can currently translate up to 100-200 sign language sentences to English [5], but it is not enough yet to satisfy the needs of the deaf and mute communities.

We plan to use Google's MediaPipe for developing customized machine learning models for hand tracking, human pose detection and face gesture tracking. In addition to that we want to apply advanced neural network models to further improve efficiency of our service. As for the dataset we have found several open source datasets but we plan to further improve our system by training it with user data if the user gives permission. As we get more users, our model will improve on the fly. We will also apply computer vision techniques, image processing and deep learning.

Our mission is to radically improve the life standards of deaf and mute people via helping them be understood by non-disabled people. We want to become their voice.

1.2 Constraints

1.2.1 Development Constraints

- For version control, GitHub will be used.
- The application will be for mobile platforms, both android and iOS.
- Python will be used as the back-end language while implementing the project.
- Project will be implemented using an object oriented approach.
- For group meetings, Discord will be used.

1.2.2 Economic Constraints

- The application will be free for all users.
- Libraries that will be used while implementing the project are open-source and free.
- To publish the application, annual fee will be required for Google Play Store and Appstore.

1.2.3 Social Constraints

- Sign language is required in order to use the application so the main users of the application will be deaf people.

1.2.4 Timing Constraints

- A functioning application should be implemented in 2 semesters since it is a project to be presented in CSFair 2022.

2 Professional and Ethical Issues

- The project will be closed-source.
- There will be regular group meetings in order to keep up with the schedule.
- Important decisions about the project, such as adding or removing new functionalities will be decided with the whole team.
- The distribution of tasks among the team members will be fair and every team member will have specific tasks.
- Without any permission from the user, the user data won't be shared with a third party.
- The video footage provided by the user and the response from the server should be encrypted.

3 Requirements

3.1 Functional Requirements

3.1.1 Gesture Recognition

- The system should be able to recognize sign language with the device's camera using gesture recognition in real time.
- The system should be able to recognize gestures from various different skin colors.
- The system should be able to recognize gestures under different lighting conditions.
- The system should be able to display an alert if the footage provided does not contain valid sign-language gestures.

3.1.2 Sign Language Translation

- The system should be able to translate the recognized sign language gestures to English text.
- The system should be able to combine consecutive gestures to form larger phrases.
- The system should be able to determine when a word starts and when a word ends.
- The system should be able to only translate American Sign Language to the corresponding English text.

3.2 Non-functional Requirements

3.2.1 User Interface

- Everything in the user interface, especially the results of the translation, must be easily readable. Grammar errors should be avoided and contrasting colors should be used for texts or icons and their backgrounds.
- There shouldn't be any non-functional elements on the user interface that may falsely give the impression that they are functional pieces, i.e., there shouldn't be any solely decorative elements that look like buttons or text fields.

3.2.2 Performance Characteristics

- The application should handle the translation process at the server side in order to lower the battery consumption.
- The application should be able to translate the provided footage in real-time.
- The application should be able to send the real-time footage to the server in less than 3 seconds.
- The server should be able to respond in less than 3 seconds.

3.2.3 Error Handling and Extreme Conditions

- The system should display appropriate error messages whenever the user doesn't follow what they are supposed to do. The user must be properly informed about what they have done wrong and what the correct course of action would be.
- The system should be able to handle any error that would be the result of the user's own actions. Any functionality or design element of the application must not fail to work or load properly when a user doesn't follow the application's intended design.
- If any exceptions or crashes occur due to the technical errors originating from the flaws of the work done by the developers, such as logic errors in the source code, it must be apparent for the user that the error is due to a bug that must be fixed by the developers and not due to an error of their own.

3.2.4 Quality Issues

- The translation done by the system must be at least 80% accurate on average.
- The system should be able to recognize the appropriate gestures as ASL with at least a 90% success rate on average.

3.2.5 Security Issues

- The system should ask for permission by the user to use the camera or any similar recording unit on their device.
- The system must not share any information gathered by the recording with the user other than what is relevant, which is the translation of the sign language being used in front of the recording unit.

3.2.6 Supportability

- The application should be available on multiple of the most common mobile operating systems such as Android and IOS.
- The application should be runnable on almost every device that supports these operating systems regardless of its specific hardware components except for the camera, which should be mandatory. Performance of the program could increase or decrease with respect to these hardware specifications, but they should not restrict the user from being able to use the application on their phone.

3.2.7 Usability

- The users should be able to start the real-time translation with at most two clicks.

- The users should be able to access the tutorial video with a single click.
- The users should be able to send feedback with at most two clicks.

3.2.8 Accessibility

- The users should be able to use the program without creating an account.
- The users should be able to download the program for free.
- The users should be able to download the program from either https://s-guy-descend.github.io/vox/, App Store or Google Play.

4 References

[1] “American Sign Language”, *National Institute on Deafness and Other Communication Disorders*, 2019, [Online]. Available:https://www.nidcd.nih.gov/health/american-sign-language .[Accessed: 10- Oct- 2021].

[2] “Without Sign Language, Deaf People Are Not Equal”, *Human Rights Watch,* 2019. [Online]. Available:https://www.hrw.org/news/2019/09/23/without-sign-language-deaf-people-are-not-equal .[Accessed: 10- Oct- 2021].

[3] “Sign Language Key to Deaf People’s Rights”, *Human Rights Watch*, 2018. [Online]. Available: https://www.hrw.org/news/2018/09/23/sign-language-key-deaf-peoples-rights . [Accessed: 10- Oct- 2021].

[4] “Accessing Public Services: Issues for Deaf People”, *BDA*, 2014, [Online]. Available:https://bda.org.uk/wp-content/uploads/2017/03/BDA_Accessing_Public_Services-Issues_for_Deaf_People-London_Boroughs_12-2014.pdf .[Accessed: 10- Oct- 2021].

[5] “SLAIT’s Real-Time Sign Language Translation Promises More Accessible Online Communication”, *TechCrunch*, 2021, [Online]. Available:https://techcrunch.com/2021/04/26/slaits-real-time-sign-language-translation-promises-more-accessible-online-communication/ . [Accessed: 10- Oct- 2021].