Bilkent University

Department of Computer Engineering

Senior Design Project

Project Specifications Report

Project: Signify

Team Members: Ali Taha Dinçer, Çağlar Çankaya, İrem Ecem Yelkanat, Muhammed Naci Dalkıran, Sena Korkut

Supervisor: Ayşegül Dündar

Jury Members: Shervin Arashloo & Hamdi Dibeklioglu

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

In society, people who are hearing impaired and/or speech impaired have difficulty expressing themselves and communicating with other people because most people lack knowledge of sign language. Even though the improvements in technology have changed the way people live and made the lives of people easier by, for example, transforming mobile phones from sound devices into multi-functional devices, communicating with people who are hearing impaired and/or speech impaired continues to be a problem in many areas including social and technical contexts. These communication activities include social life, healthcare, career development, and education. Furthermore, as a result of the Covid19 pandemic that started in late 2019, obligations regulated by most countries, including wearing face masks and social and physical distancing, have increased the communication and social challenges for hearing impaired people. For example, wearing face masks has led to some negative impacts in communication with other people as it eliminates speech perception by visual features through lipreading. Additionally, a considerable amount of face-to-face communications have turned into virtual communications, which results in more hardship for the hearing impaired and/or speech impaired people as some of the most used virtual communication services like Zoom, Microsoft Teams or Skype do not support sign language.

According to the World Health Organization (WHO), 5% of the people on the earth are hearing impaired, which is more than 350 million people [1] and will exceed 700 million by 2050 [2]. Considering that, sustainability of the social lives of hearing-impaired and/or speech impaired people will be an essential issue in the future. Therefore, we propose a solution to this problem named Signify. Signify is a mobile application with the main aim of helping hearing and/or speech impaired people in their social lives by translating sign language into text along with speech and text-speech to sign language translation in real-time.

This report consists of the description of the proposed project, respective constraints, professional and ethical issues, and functional and non-functional requirements of this project.

1.1. Description

Signify is a mobile application that aims to solve social problems regarding communication, understanding, and expressing for the hearing impaired and/or speech impaired people. By this means, Signify will help these people to improve their quality of life. In that manner, the application can be seen as a communication tool.

There are already existing applications such as “Hand Talk Tradutor para Libras” [3], “ASL Translator” [4], and “S.L.A.I.T.” [5] for the hearing impaired and/or speech impaired people. However, “Hand Talk Tradutor para Libras” does not contain real-time translation over the video, which creates a one-way channel between communicators. “ASL Translator” is a paid application and works on pre-given text and video translations, in which translations are pretty limited, and for every new word or phrase, the app needs to be updated. Finally, “S.L.A.I.T.” can only be used during video calls, and it is still in the beta phase.

Signify combines these apps in terms of their translational and conversational capabilities with tools such as a warning system and two-way translation. With Signify, users can easily communicate through video calls and use their mobile phones to real-time translate to both text and sign language in a day-to-day conversation. With this, hearing impaired and/or speech impaired people can join conversations even if the people in the conversation do not know sign language. In addition to that, if activated, the warning system on the app can warn users of various types of vibrations when there is a nearby car horn, siren sound, dog bark, or any other urgency that needs to be heard. Along with these innovations, we aim for Signify to be a swiss-knife for hearing impaired and/or speech impaired people during their everyday lives.

1.2. Constraints

1.2.1. Implementation Constraints

- To ensure implementation integrity and version controls, Github and Git will be used.
- Python programming language will be used for machine learning models’ development.
- Flutter will be used for mobile application development.
- The machine learning model will be deployed to AWS.
- SageMaker API will be used.

- Google Colab will be used for utilizing extra computational power in the model training process.

1.2.2. Economic Constraints

- The application will be free for every user.
- Usage of AWS services will require a fee.
- Publishing mobile applications for both Android and IOS platforms will require a fee.
- Free development tools will be used.
- Open source datasets will be used for the machine learning models' development process.

1.2.3. Ethical Constraints

- User data will not be shared with any user, third-party company, or application.
- Any personal information that is required will be encrypted.

1.2.4. Social Constraints

- The application will be designed to maximize communication efficiency between individuals both in the physical world and the virtual meeting environment.

1.2.5. Language Constraints

- As sign language, the American Sign Language (ASL) will be used.
- For speech to sign language translation and captioning, English will be used.

1.2.6. Sustainability Constraints

- Further development of the application and features will be decided regarding the user feedback.
- The data obtained from users will be used for machine learning models to increase accuracy and understand the model's performance.

1.3. Professional and Ethical Issues

Our target users hear and speech impaired people; therefore, some ethical and professional responsibilities must be taken. The most significant moral obligation is to provide equal opportunities in the usage of the application.

To ensure the ethical responsibility mentioned above, the dataset used in training should not be biased. The dataset should be representative of all ethnic groups. The training data should appeal to all the users to not humiliate the users' identity and to provide a better user experience. Since the project aims to improve the standard of hearing and speech impaired people, it should be sufficient and representative to make correct and similarly accurate predictions without discriminating against any user. Furthermore, the output of the voice to text and sign language models should not be discriminative. The caption generated by the application should not be biased regarding any race, gender, or social class.

Furthermore, some conversations might include private information about the user; therefore, this information should be protected and should not be shared with other users or third-party companies. As the application grows, user feedback and some recordings of the meetings can be utilized to enhance the performance of the machine learning models. In these cases, any personal information, such as name, address, and job should not be revealed. Additionally, to serve the machine learning models, we will use APIs; however, any information about the user will not be sent to APIs to protect the user's data.

In conclusion, in the development of this project, a representative and unbiased dataset should be selected not to cause any discrimination. Moreover, the machine learning models' output should not include phrases that imply gender inequality, humiliation, and ethnic-based bias. Finally, users' personal information will not be shared with other users or third-party companies and will be protected.

2. Requirements

2.1. Functional Requirements

2.1.1. Two-Way Language Translation

The main purpose of the application is to provide a channel among hearing-impaired or speech-impaired people and society. For this intention, two-way language transformation is ensured. The application should be able to translate a speech or text in English to sign language through the instrumentality of animation for continuous visualization. For the reverse side, the application should also convert sign language from video to text and speech for the users who do not know sign language. For this, the user should be able to use the microphone and camera of their phones.

2.1.2. Real-time Language Translation

The application should work close to the real-time speed to achieve practical and effective communication, especially for online meetings and conferences. Users should be able to reach the translation synchronically with the initial communicator.

2.1.3. Notifications Through Vibration

The application should also help impaired people to identify urgent issues happening around them. The application should vibrate in different types for hearing/speech impaired people to become aware of the significant voices such as sirens and car horns.

2.1.4. Language-free Guidance

For the straightforward usage of the application, guidance on the features should be done without language constraints. Guidance should be both with text and sign language to resolve the discordance happening from various impairments.

2.2. Non-functional Requirements

2.2.1. Usability

The application should have a user-friendly interface for users to navigate and appreciate any feature within three seconds.

2.2.2. Extensibility

The application architecture should be easily extensible for new learning models and additional features that will be done in 8 months.

2.2.3. Performance

The server's response time should be less than 300 milliseconds.

2.2.4. Security

The application should not leak any non-encrypted personal information in case of a data breach.

2.2.5. Compatibility

The application should be available for all operating systems, including Android, iOS, and web platforms, to deliver the service to all users.

2.2.6. Privacy

Any personal information and video and audio records should not be used or shared with any user, a third-party company, or application.

3. References

[1] E. McPhillips , “World wide hearing loss: Stats from around the world,” *Audicus*, 14-Sep-2021. [Online]. Available: https://www.audicus.com/world-wide-hearing-loss-stats-from-around-the-world/. [Accessed: 05-Oct-2021].

[2] “Every 4th person to suffer hearing loss by 2050: Who,” *Down To Earth*. [Online]. Available: https://www.downtoearth.org.in/news/health/every-4th-person-to-suffer-hearing-loss-by-2050-who-75718. [Accessed: 05-Oct-2021].

[3] “Hand talk translator - apps on Google Play,” *Google*. [Online]. Available: https://play.google.com/store/apps/details?id=br.com.handtalk. [Accessed: 05-Oct-2021].

[4] “ASL translator - apps on Google Play,” *Google*. [Online]. Available: https://play.google.com/store/apps/details?id=com.asltranslator. [Accessed: 05-Oct-2021].

[5] “S L a I T. real-time sign language translator with AI.,” *S L A I T. Real-time Sign Language Translator with AI.* [Online]. Available: https://slait.ai/. [Accessed: 05-Oct-2021].