University of Waterloo

Faculty of Engineering
Department of Computer and Electrical Engineering

Project Specifications and Risk Assessment

Voice - Device to Enable Real-time Sign Language to Speech Translation

Group 037

Akshay Budhkar (20457593)
Eliot Chan (20462016)
Biraj Kapadia (20467909)
Amish Patel (20460828)

Consultant: Prof. Dana Kulic

May 30, 2016

Table of Contents

1. High Level Description

1.1. Motivation

There are 375,000 culturally deaf people in Canada [1], and an estimated 1,250,000 deaf in the United States [2], and in-person communication barriers between the deaf or mute and the general populous still remain high. As technology and society have advanced, accommodations for the culturally deaf have slowly improved, but members of the culturally deaf community can still only interact with others naturally in online or other text-based environments. In-person communications between a native sign language speaker and person who does not know sign language still requires a translator to be present, specialized hardware [3], or a reduction to a commonly accepted, but less personal medium - such as through text on a phone. With Voice, we aim to allow the culturally deaf and native sign language speakers to interact with others in a way that is natural and personal to them.

1.2. Project Objective

The objective of this project is to break down communication barriers between the culturally deaf and the general populous by designing and building a system to allow sign-language speakers to communicate with non-speakers in an unobtrusive and natural way. A user's sign language symbols should be translated in real-time and converted to an audio output to mimic speech.

The system should be composed of the fewest hardware components possible, ideally utilizing hardware that a person would normally have with them (i.e. A smartphone). Current sign language translation systems are bulky and obtrusive, either covering the majority of a user's arms or require additional hardware such as cameras or laptops. This is obviously cumbersome, and it is of utmost importance that a user has minimal interaction with the system itself.

Our solution aims to support multiple sign language dialects, namely American Sign Language (ASL) and Japanese Sign Language (JSL), a novel feature that has not been adopted by any other publically-revealed prototype. The language selection (and overall translation functionality) should be moderated by an accompanying smartphone application, allowing the user to control when and what language they want to speak in.

1.3. Block Diagram

The proposed solution involves building custom-made gloves fitted with flex sensors, an accelerometer, a gyroscope, and possibly other sensors which will be used to gather analog data during a user's signing motions. A microcontroller will convert these analog signals to digital for processing. The digital data will then be fed into a processing subsystem which will analyze the given data and make a prediction of what word(s) the sign corresponds to. A high-level overview of the product is given in Figure 1. It comprises of two main subsystems: the External Subsystem (ES), and the Processing Subsystem (PS).

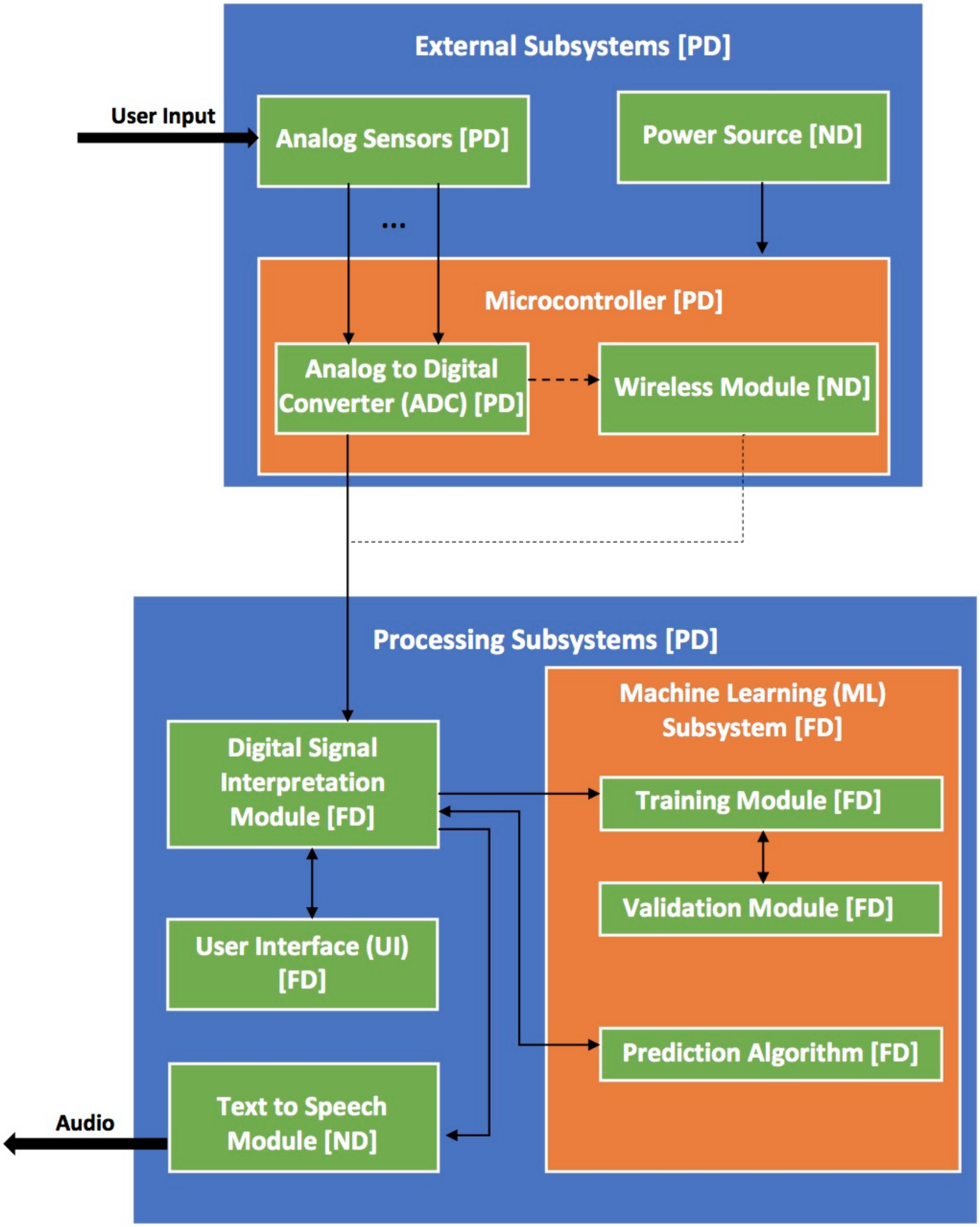

Figure 1: The block diagram of the system and the connections between each elements.

1.3.1 The External Subsystem

Flex sensors, accelerometers, and gyroscopes are possible considerations for the analog sensor array. Flex sensors would be fitted on the fingers of the glove to determine the curvature and movement of any given finger - a total of ten would be required, one for each finger. An accelerometer on each glove would measure the 3-dimensional motion of the hand, an integral part of "speaking" many signs. A gyroscope on each hand would also allow for the tracking of the orientation of the hands. Inputs from these analog sensors will be fed into a microcontroller, converting the signals to digital format. This data will then be streamed to the PS for interpretation. The ES will also require batteries to power the microcontroller.

1.3.2 The Processing Subsystem (excluding the Machine Learning Subsystem)

The Digital Signal Interpretation Module (DSIM) is the receiver of the ES's outputs, and transforms the data to prepare it for the Machine Learning Subsystem (MLS, further explained in 1.3.3.). Predicted results given as output from the MLS are then communicated to the DSIM, which communicates this information to the User Interface (UI), and Text to Speech Module (TTSM). The DSIM also receives input from the UI to determine which variant of sign language to interpret the signals with (ASL or JSL), and whether the entire system is currently active and translating. The UI can also display the transcript of the signs communicated by the user in real-time. The TTSM simply reads out the given text using the PS's speaker system (if available) to mimic speech.

1.3.3 The Processing Subsystem (Machine Learning Subsystem)

The MLS consists of three distinct portions – the Training Module (TM), the Validation Module (VM), and the Prediction Algorithm (PA). During the development of this prototype, we will manually create a set of training and validation data (two halves of the same set of data, randomly selected) once the ES and DSIM are verified as working. The training data will be given to the TM to teach the model what particular signs look like, and the validation data will be used in the VM to ensure that the model works. This process will be repeated multiple times to maximize the usage of the training set and to ensure there is little bias in the generated model. Once the model is determined to be correct and accurate, the generated model will become the PA, acting as a real-time translation model using inputs from the DSIM, and generating text outputs as a response.

2. Project Specifications

2.1 Functional Specifications

Table 1: Functional Specifications.

Specification	Type	Location	Description
Multi-Language Support	Essential	Processing Subsystem	Able to interpret 100 symbols in American Sign Language and 100 symbols in Japanese Sign Language.
Latency	Essential	Entire System	Maximum latency of <800 ms from the end of any given sign.
Accuracy	Essential	Machine Learning Subsystem	Sign should be recognized with an accuracy of > 85%.
Mobile Application	Essential	Processing Subsystem	Controls the variant of sign language translated, can turn on/off the translation itself, and has the ability to display the text and output audio for the corresponding signs.
Wireless Support	Non-Essential	External Subsystem	Connects to processing device using Bluetooth or other wireless mechanism
Transcript Exporting	Non-Essential	Processing Subsystem	Capable of exporting the transcription of a session into one or more file formats.

2.2 Non-Functional Specifications

Table 2: Non-Functional Specifications

Specification	Type	Location	Description
Hand Flexibility	Essential	External Subsystem	The gloves should allow a range of motion such that the user can complete any supported sign.
Portability	Essential	External Subsystem	90% of the surface area of the glove-mounted components should be on the glove itself. A singular modified glove should weigh less than 500 grams.
External Subsystem Enclosed in Case	Non-Essential	External Subsystem	Install a case to protect the microcontroller from environmental damage.
Battery Life	Essential	External Subsystem	Last one hour without exchanging/charging battery
Cost	Non-Essential	Overall System	The price of the entire system should not exceed $500.

3. Risk Assessment

Table 3: Risk assessment with possible mitigations

Potential Risk	Probability	Impact	Description
Machine Learning Model Not Accurate	Medium	High	**Situation**: The machine learning model might not predict the correct text for a given symbol. Given the limited information the group currently has on sensor accuracy and noise, it is difficult to decide on the type of machine learning models that will be used to achieve the prediction. It is possible that it is difficult to interpret trends in the digital data, which would limit the ability to create an accurate model. **Mitigation**: Reaching out to our consultant, who specializes in machine learning, can mitigate the risk of creating a bad model. If the model is inaccurate despite those attempts, the prediction algorithm can be modified to use a heuristics-based approach (pattern matching, closest neighbours) instead.
Sensor Data Noisy/Inaccurate	Medium	High	**Situation**: There are hundreds of signs that need to be detected and distinguished from each other. If the sensor data is too noisy, then it may be challenging to detect which hand gesture corresponds to what sign. **Mitigation**: This device is using multiple sensors (flex sensors, accelerometer, gyroscope) to detect hand motion. For example, if one of the sensors such as flex sensor data is unreliable, then the data received from the more accurate sensors can be given more importance in the machine learning model. Another mitigation strategy is to reduce the set of supported signs to ones that can be differentiated despite noise.
Issues in Transmitting Data from Microcontroller to Processing Subsystem	Low	High	**Situation**: Transmission of data between two completely separate systems can run into errors (loss/corruption/lack of bandwidth), especially if the transmission is through a wireless medium. **Mitigation**: Some of the less essential sensors could be removed to reduce the amount of data that is being transmitted to

			the processing subsystem. Wireless transmission can be exchanged for the more reliable wired transmission.
Insufficient Time	Medium	Medium	**Situation**: Because the concept of this device has not been explored previously, there is a possibility that some of the required information to complete this project will be in the research phase and it may be beyond the scope of 4^{th} year ECE students. **Mitigation**: This can be resolved by consulting with professors whose research interests are in this field. To accommodate for insufficient time, the group can work on the project during the co-op term. If the additional work does not resolve possible issues, the scope can be reduced (fewer supported signs, less accurate model, etc.).
Technical Skill	Medium	Medium	**Situation**: At the beginning of the project, only one of the team members has experience in designing machine learning models. In addition, only one of the team members has experience interfacing with microcontrollers. All group members have limited knowledge of the various sensors that exist in the market, and there is potential to pick the incorrect parts, interface with the microcontroller incorrectly, or design the machine learning model incorrectly. **Mitigation**: The team members will study machine learning through external courses in their free time. The group will spend 2-3 meeting sessions researching various sensors available in the market and spend time gathering knowledge about using the sensors with the selected Arduino microcontroller.
Insufficient Sign Language Knowledge	Low	Low	**Situation:** None of the members of the team know sign language so while training the machine learning model if the lack of knowledge proves to be an issue then it could hinder the success of the project. **Mitigation:** The members of the team will consult with someone that knows sign language or spend more time learning sign language using online resources.

References

[1] "Statistics on deaf Canadians," Canadian Association of the Deaf, 2015. [Online]. Available: http://cad.ca/issues-positions/statistics-on-deaf-canadians/. Accessed: May 30, 2016.

[2] T. Harrington, "Deaf population of the U.S. - Local and regional deaf populations," Gallaudet University Library Library, 2010. [Online]. Available: http://libguides.gallaudet.edu/content.php?pid=119476&sid=1029190. Accessed: May 30, 2016.

[3] sComm, "UbiDuo 2 wired," 2015. [Online]. Available: https://www.scomm.com/product/ubiduo-wired/. Accessed: May 30, 2016.