Sign Out Loud

Phase 3

Team 25

Todd Locke

Eric Ernst

Isaac Reath

Jason Gregory

Gino Notto

Table of Contents

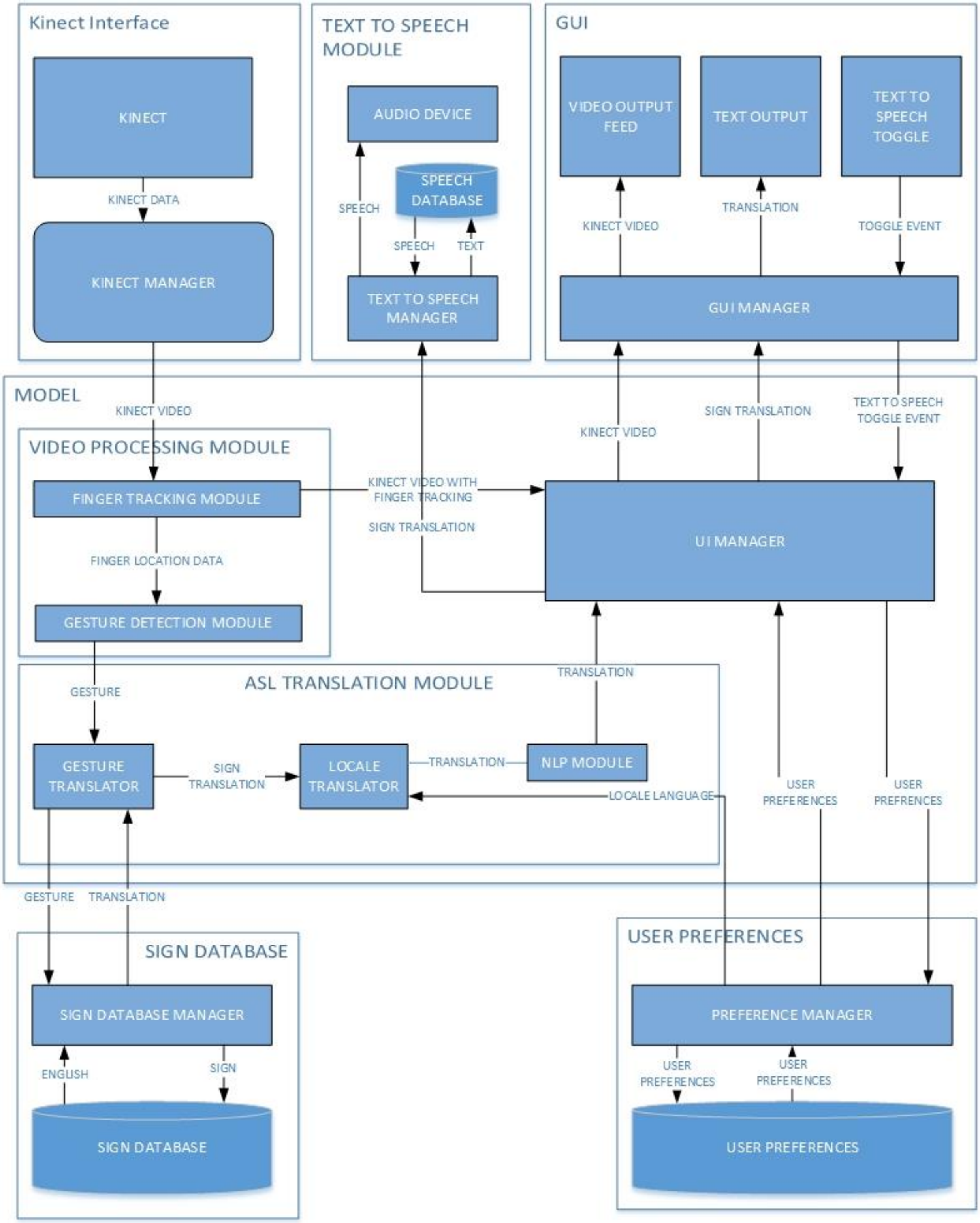
Problem Statement

The main problem this software will attempt to solve is the difficulty of communication between the deaf and hearing. The problem is that most people do not know ASL. This creates barriers for those who use ASL as their main form of communication. Sign Out Loud attempts to break down this barrier by giving the deaf a new way of communicating.

Sign Out Loud will use a Kinect to track user's hand and finger motions to produce gestures. These gestures will then be handed off to be interpreted as language. Finally, the language will be outputted to the graphical user interface (GUI) as both text and if the user chooses speech. In real time the user can simply sign in front of a Kinect to have English or whatever the local language is outputted in both text and local speech. Further, the program will do it's best to correct and language syntax problems that arise from translating from ASL to a spoken language.

The usages of the program are numerous. This will be especially useful for talking to groups, however. A laptop could be hooked up to a projector and output speech and text in real time as the person signs. This eliminates the need for a translator and allows a deaf person to be totally self-sufficient and communicate with virtually anyone.

Software Architecture



Kinect Interface

The Kinect module provides a simple and accessible API for accessing the Kinect by the video processing module. It contains only the actual Kinect device and a manager that acts as a daemon for other modules to access.

Kinect – A device that captures video and audio. It is able to determine the position of the simple skeleton of the person such as hands, fingers, joints, etc.

Kinect Manager – A software module that will manage the Kinect for the rest of the program. It will encapsulate some of the functions of the Kinect and handle setting up the Kinect, detecting it, reading data from it and storing it, and allowing consumers to consume the data. This will provide a simple and buffered way for the rest of the program to read Kinect.

Text To Speech Module

The Text To Speech module is responsible for taking the translation of signs and synthesizing it as a voice. It consist of a manager, a database, and an output device. When part of the software would like to provide verbal feedback it must go through the manager. The manager will then queue up and speech and fetch the right audio files. This provides a simple interface that allows for easy access by any other part of the program.

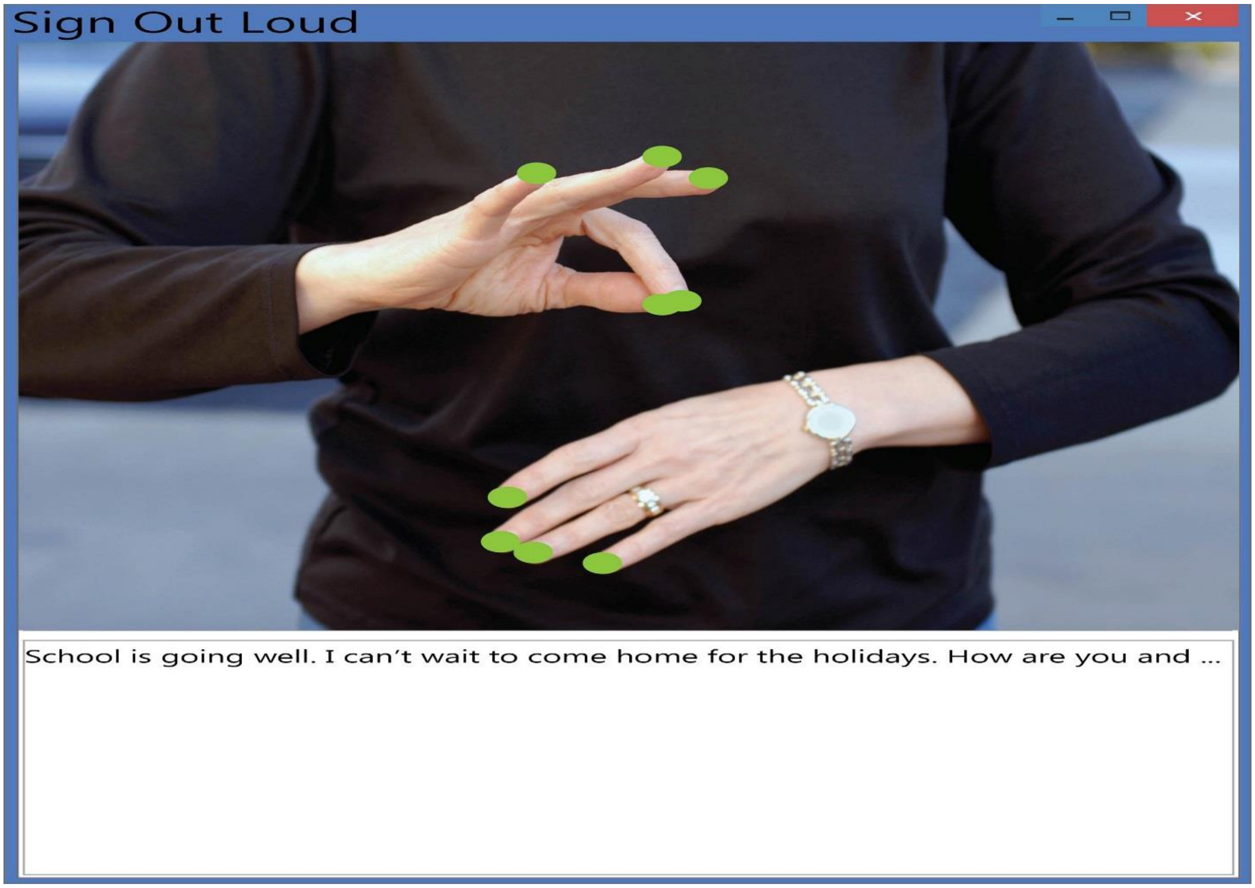
Audio Device – This can be any device that is capable of outputting audio. It may be a speaker or headphones plugged into the PC via AUX or perhaps a Bluetooth speaker or wireless speaker. The manager will find and manage them.

Text To Speech Manager – This is the software interface that allows the program to output speech from strings of text. The manager then handles pulling the appropriate audio files from the database for the user's locale (or whatever language the user suggested) and then queueing the audio file for playback on the appropriate channel of the output device.

Speech Database – This database will hold the necessary files for each string of text or individual characters and for each installed language.

GUI

The GUI consist of video panel and a text panel. The video panel displays the feed from the Kinect with the overlaying information from the video processing module. This gives feedback to the user that the system is in fact detecting all ten fingers. The text box is a scrolling feed of what the user is signing. Either of them can be resized by dragging the top or bottom of the respective panels. The interface can then be projected to show a mass audience the translation of the signs.



School is going well. I can't wait to come home for the holidays. How are you and ...

Example user screen mockup

Video Output Module – This will encapsulate the panel that displays the video. It will manage outputting the video and resize it accordingly.

Text Output – This will wrap the textbox and handle scrolling, resizing, etc as other modules send text to it.

Text-to-speech Toggle – This is a main button that will toggle whether the text is to be sent to the speech output module. (Not shown)

GUI Manager – This manager will handle updating the appropriate views and modules accordingly as well as fetching and sending the appropriate GUI options and preferences such as resizing components.

Gesture Recorder – This allows for the user to record his own sign to the database.

Video Processing Module

The video processing module is responsible for taking the video feed from the Kinect and pulling out the important information such as the position of the user's fingertips, orientation of the user's hand and aggregating the information.

Finger Tracking Module – The finger tracking module is responsible for tracking the user's fingertips over time as well as keeping information such as speed, direction, starting and stopping points, etc. This information will then be handed off to the gesture detection module for possible translation. This module also modifies the video feed by overlaying the fingertip position on the video and then sending it to the GUI to display for visual affirmation that the system is properly tracking the user.

Gesture Detection Module – This module is responsible for taking in information from the finger tracking module and converting it into a series of gestures that can then be handed to the translation module for translation. It will take in the starting points and end points of fingers and then translate that into a series of gestures such as "INDEX FINGER: UP MOTION: VERTICAL…" This information will then be sent to the translation module in an attempt to match it to a sign.

ASL Translation Module

The translation module is responsible for taking in gesture data and translating it into the appropriate language. Due to sign language not being phonetic and having a different syntax structure, it will also attempt to reorganize sentences or phrases into the appropriate and syntactically correct language. It will interface with the sign database manager to send gestures and receive translations. Finally, after making the appropriate alterations to the translation it will hand it off to the GUI for display.

Gesture Translator – The gesture translator takes in gestures and attempts to retrieve translations from the sign database. It will then take the translations and send them to the locale translation module.

Locale Translator – This module's job is to take the translation and then in turn translate that into whatever the user's language is set to.

NLP Module – The Natural Language Processing module is what is responsible for ensuring the text that is produced makes sense to the person the signer is signing to. It makes sure that the translation syntax is correct and fixes any problems that may pop up from translating from a non-written language to a written one. After manipulating the text it will then send it to UI manager to finally send it to the GUI.

Sign Database Module

The sign database module is responsible for maintaining a database of signs that can be accessed via gestures. It has a few responsibilities. It ensures the database is always up to date by pinging a central database and downloading new signs or replacing signs if need be. It also turns the gesture data into a format that can be used to access the database and retrieve signs.

Sign Database Manager – This is the module that is responsible for allowing communication between the database and any other modules. This guarantees the integrity of the database by ensuring no other modules modify the database. It is also responsible for keeping the database up to date.

Sign Database – The database that holds signs and translations.

UI Manager

The UI manager is responsible for bridging the data model and the GUI together. It manages all of the user's preferences and handles data accordingly. Depending on stored user preferences, it will format data accordingly.

User Preferences

The user preferences module is responsible for storing user preferences during execution and then storing them when the user exits the application. It consist of a manager and a database. The manager ensures that no module modifies the preferences unknowingly.

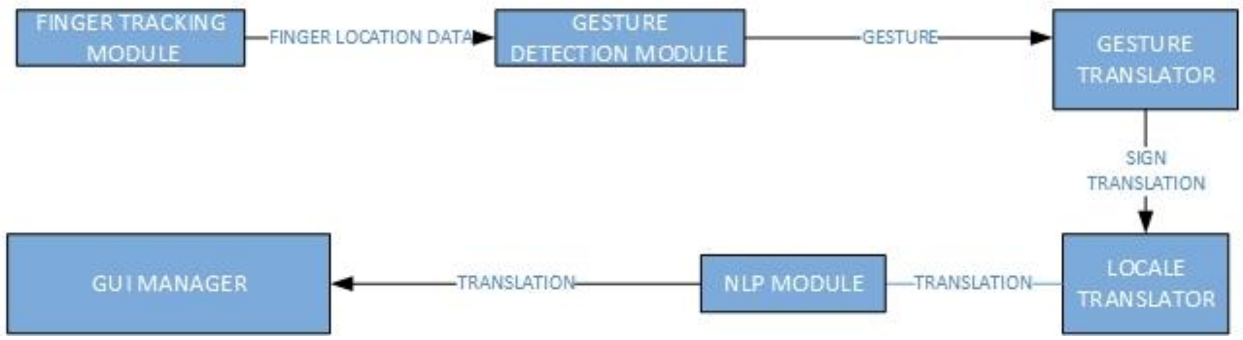
Preference Manager – The manager that allows read and write access to the database. It ensures the database stays consistent as well as wraps database functions to make it easy for other modules to change and retrieve preferences.

User Preferences Database – The database that holds persistent user information such as GUI settings and user settings.

Cross Reference Listing

Reference Number	Function	System Specification	System Requirements Specification
1	Kinect Support	Kinect Interface	Kinect Interface
2	Track user’s fingers	Video Processing Module	Finger Tracking Module
3	Detect Gestures	Video Processing Module	Gesture Detection Module
4	Custom Gesture Additions	GUI	GUI Manager
5	User Interface	GUI	GUI Module
6	Store User Preferences	User Preferences	User Preference Module
7	Translate Gestures to Local Language	ASL Translation Module	Gesture Translator
8	Support of multiple languages	ASL Translation Module	Locale Translator
9	Format language to local syntax	ASL Translation	NLP Module
10	Display Text	GUI	Text Output
11	Text-to-Speech	Text-to-Speech Module	Text to Speech Manager
12	Local Speech	Text-to-Speech Module	Speech Database
13	Support Output to Audio Devices	Text-to-Speech Module	Audio Device
14	Display Video to User	GUI	Video Output
15	Display Translation as Text	GUI	Text Output
16	Keep Sign Database Up to Date	Sign Database Module	Sign Database Manager

Integration Thread



The integration thread is the bare basic modules of the system. Each module provides an interface that allows for additional modules to be built upon it. Each one encapsulates one phase of the entire video capture to graphical output of the system. Due to each step being isolated, it allows modules to capture the data and interact with it at any step.