SLIK
SIGN LANGUAGE INTERPRETER FOR KINECT
PHASE 2 | SYSTEM SPECIFICATIONS
INDEX

Problem Statement 3
Existing Solutions 4
Fundamental Design Goals 6
Functional Requirements 7
Quality Expectations 10
User Profile and Deployment Environment 11
Problem Statement

Based on information provided by the World Health Organization (WHO) website, about 360 million people worldwide suffer from a condition known as hearing loss. Those with hearing loss are considered by the WHO to be those who are not able to hear as well as someone with “normal” hearing [1]. Though the severity of this condition generally ranges from degrees that exhibit slight to profound hearing loss (Table 0), attempts at establishing a standard metric for hearing loss were made by research groups throughout the global community.

<table>
<thead>
<tr>
<th>Degree of hearing loss</th>
<th>Hearing loss range (dB HL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>~10 to 15</td>
</tr>
<tr>
<td>Slight</td>
<td>16 to 25</td>
</tr>
<tr>
<td>Mild</td>
<td>26 to 40</td>
</tr>
<tr>
<td>Moderate</td>
<td>41 to 55</td>
</tr>
<tr>
<td>Moderately severe</td>
<td>56 to 70</td>
</tr>
<tr>
<td>Severe</td>
<td>71 to 90</td>
</tr>
<tr>
<td>Profound</td>
<td>91+</td>
</tr>
</tbody>
</table>

Table 0 – Degrees of hearing loss and hearing loss range in dB
(Taken from http://www.asha.org/public/hearing/Degree-of-Hearing-Loss/)
Approximately 7.5 million people in the United States alone have trouble using their voices [1]. They are reliant upon sign language to communicate. But the majority of hearing individuals do not understand sign language, so communication between speech impaired people and those without such a disability can be challenging.

In the popular movie “The Miracle Worker” (1962), the story revolves around a girl called Helen who is shown to be deaf and blind. The movie shows to what extent a person can turn paranoid if they are not able to communicate with other people, especially with family.

The only means by which a speech impaired or a deaf person can communicate is by using sign language and for that reason it is an absolute must that the person on the receiving end should know sign language. However at most times, that’s not the case. A common inconvenience faced by those who use sign language, or signers, is the ability to communicate universally to others without the use of written medium.

The Sign Language Interpreter for Kinect, or SLIK, would help alleviate this by acting as the "translator" between nonverbal and verbal communication. Our primary objective is to create a prototype that understands the gestures of sign language, converts them to text, and then synthesizes that text to speech.

The system will have the capacity to capture the movements of the signer and then render a written and spoken translation of the signer’s gestures instantaneously. For this, it will use the sensors available within an Xbox Kinect.

Our aim is to build something that our users can rely upon all the time. It should be unobtrusive, easy to use and recognize a user’s gestures with a degree of accuracy, so that users don’t have to repeat themselves often, which can be frustrating. Of equal importance is the issue of practicality and affordability. The system should run on readily available commercial platforms, and should be cost effective.

**Existing Solutions**

A number of assistive technologies for the speech impaired have existed for a number of years, from very simple devices that play back previously recorded messages at the user's
behest, to touchscreen devices that display various pictures depicting commonly used phrases, which the user can tap to select what they want to say, and the machine speaks that aloud.

- **Big Talks [2]** – This is simply a device with a big button that the user can press to play back previously recorded messages.

  Apart from its obvious lack of flexibility, this solution requires the users to be dependent on someone to pre-record whatever they would want to say in advance. It is also very expensive.

- **DynaVox [3]** – DynaVox is a company that produces augmentative technology. One class of solutions they offer is designed specifically for people with speech impediments. A touchscreen portable device enable users to say what they want by tapping the screen which displays images associated with common phrases and objects.

  This is much more flexible than the Big Talk, it still doesn't allow the user full expressive power. Getting a DynaVox product is also a very complicated process, requiring speech therapists and DynaVox representatives to be involved. [4]

Other solutions similar to the DynaVox exist, with similar disadvantages and a general lack of widespread availability. Some of these are –
- Augmentative Communication and Assistive Technology (ACAT)
- RSLSTeeper AAC
- Autism Community Assistive Technology

There has been research investigating the possible use of the Kinect for use in applications similar to our proposed solution by academicians in China in collaboration with Microsoft [5]. However, there isn’t yet a publicly available working prototype version of it yet. We imagine that it must be a very complex system to design and test.

Therefore, we plan on releasing our application over three versions. The first version will have very basic functionality, intended to be a proof of concept. Future versions will add more functionality. We discuss this in the “Future Expansion” section.

**Fundamental Design Goals**

Our goal is to enable our customers to be able to interact and communicate with others independently. Since our customers are a part of the deaf and/or mute communities, we seek solutions that will utilize either well-established or easy-to-learn signing gestures, which are widely known and used. This is especially important for those within these communities that have not yet achieved mastery in signing, or are inhibited from attempting to do so. Users of an ideal system will thus be able to learn and practice gesture-based language on it, regardless of their proficiency, and communicate with others effectively.

Others within the deaf and mute communities may also very rarely be accommodated for, most notably those that must speak up before an audience without a voice to do so. Though not our primary goal, we would like this ideal system to also enable the deaf and mute that maintain obligations in public speaking to speak to large audiences without the need for an audible voice. This would not only help in empowering the deaf and mute populous, but could also lead to many new opportunities for those without the ability to readily communicate with words.

The primary purpose of our product is to enable people with a speech impairment condition the freedom to communicate. Such communication should be –

- Spontaneous, and without lag
- Enable the user to use the full expressive power of sign language

The system should also be
- Easily available
- User-friendly
- Affordable

## Functional Requirements

At a basic level, the system should be able to read a sign from the user, present it on the screen as captions and convert the text to speech in real time. Below is a section wise description of how the system should work and what is expected of it.

### Recognize Gestures and Generate Captions

The system should be able to recognize sign language gestures from a user and be able to generate captions based on the recognized gesture.

1. **Multiple Sign Language Dialect Support**
   1. The system should be able to handle all the sign language dialects spoken across the globe with an additional support feature to add new sign gestures.
   2. A list of sign language which must be present in the system may be found here: [http://en.wikipedia.org/wiki/List_of_sign_languages#cite_note-5](http://en.wikipedia.org/wiki/List_of_sign_languages#cite_note-5)

2. **Generate Captions in Multiple Languages**
   1. The system should be able to translate captions in all major languages and show it to the screen.

### Camera Preview and Speech Output

The system should show a live video of the user and should be equipped to speak aloud any generated captions.

### Configuration
There should be a configuration menu present in the system which allows user to make changes in caption language, speech language, voice types and sign language. It should also have an additional feature for the user to add custom signs and their respective translations.

![Configuration Menu]

**User Interface**

The User Interface should have the following components:

1. **Main Menu**: Basic layout of the Main Menu should have a Welcome Screen which welcomes a specific user as depending on the account. In addition to this, the screen should show the video of the user and should have the following buttons:

   1. **Start Recording**: This button initiates the conversion of sign language to text. The text on the button should change to **Stop Recording** when the conversion is underway. **Stop Recording** button ends the conversion of sign language to text.
2. **Speak Aloud**: This should be a toggle button which when pressed, speaks aloud the translated caption text.
   1. **Volume bar**: A volume bar must be present for regulating the speech volume inside the application.
3. **Settings**: This button should take the user to a configuration menu which has the following options:
   1. **Calibrate Kinect**: To calibrate the kinect recognition system.
   2. **Add Sign**: To add or generate new signs to be added to the system. There should be a mini screen which accepts sign recognitions and allows the user to add new text relating to the sign.
   3. **Text Language Selection Menu**: Allows user to change the language in which text captions are presented. Must include all major languages.
   4. **Sign Language Selection Menu**: Allows user to change the sign language dialect preferred by the user. Must include all major sign languages.
   5. **Speech Language Selection Menu**: Allows user to change the language in which speak aloud speaks the text. Must include all major languages.
   6. **Voice Style Selection Menu**: Allows the user to change the voice styles. Must include multiple voice types such as Male, Female and different dialects.
4. **Save Recording**: Saves the recorded translation and generates a text file which has the translation written inside.

The main User Interface
In addition to all these, the system should be running continuously until the recording is stopped by the user and the interpreter should allow for pauses in conversations. The system should ignore things that could be considered interference, like movement in the background. The interpreter should have an ability to keep running in the background, until the user chooses to stop the application.

The quality should be optimal, with minimal errors in interpretation of sign languages and translation of text and speech. This is explained further in the section Quality Expectation.

Finally, the system should be able to run on a variety of machines such as desktops, laptops and tablets with Windows installed on it. More information about environment may be found below in the section User Profile and Deployment Environment.

**Quality Expectations**

There is a certain amount of quality and professionalism that is expected out of this project. A completely finished project is expected, meaning extremely few or no software errors. Any errors should be relatively easy for the user to understand and fix.

For example, if the program spontaneously closes due to a user error, it should be clear to the user why the program shut down, and how this problem can be avoided in the future. General things such as this that make a system easier for an end user to use and understand are extremely important for the first deployment of the system.

Quality, or accuracy is extremely important for this system. If the translation makes too many mistakes the consumer will move on to other products that perform better. An accuracy of at least 95% is required, signifying that at least 19 out of every 20 gestures the user makes should be interpreted correctly. This accuracy requirement is mandatory for the first deployment. If a consumer uses our product, and it does not meet these quality requirements, then it won’t matter if they are fixed or not, that consumer has moved on. Other solutions have quality standards that are as strict if not stricter. We need this level of quality to stay viable and competitive in the market.
User Profile and Deployment Environment

User Profile

The system should be designed to be versatile in terms of deployment and user profile. It should be able to handle multiple sets of users with a minimal background in computers.

Since this system is expected to act as a communication link between general population and hearing or speech impaired individuals, the user base in addition to deaf and/or mute individuals would also include therapists, audiologists and a subset of the general population in need of a medium to communicate with the hearing and speech impaired community. This system will be used to eliminate the language barrier between such communities, facilitating regular conversation over video or audio chat using mostly consumer-grade readily available and affordable technology. The system may also be used by two individuals habituated to two different protocols of sign language, as is the case for people located in different countries.

In general, the primary user base of this group is speech impaired individuals and the respective audiologists or family members, who would help set it up for such individuals.

Deployment Environment

We plan to deploy SLIK in three stages –

1. Version 1 – The first version will only have basic functionality. It is intended to be a proof of concept. The User Interface needs to be complete, and users will be able to add simple custom gestures that correspond to some text that they can type in. Whenever they make that gesture, the system will recognize it and output the text associated with the gesture.

2. Version 2 – The second version will add support for gestures standard in different dialects of sign language. The user will no longer need to define their own gestures, they can simply use the protocol they are already familiar with. The system will have different databases for the different sign language protocols.

3. Version 3 – This version will add support for multiple languages for speech output.
As mentioned in our fundamental objectives section, the SLIK system needs to be affordable and easily available. Keeping these objectives in mind we have decided to base our solution as a Windows Store application, which can be downloaded and used for free. All a user needs to avail of the system is a Windows PC hooked up to an Xbox Kinect sensor.

The Kinect is available for $149.99 [6], and has applications beyond just assistive technology. This, in our opinion, makes it a very affordable and viable technology for a wide range of users.

**Usage Scenarios**

Listed below are some of the scenarios where the SLIK system can be useful for communicating –

- Speech impaired person uses SLIK, which converts sign language to speech, to converse with a normal person
- Speech impaired person uses SLIK, which converts sign language to text, to communicate with a hearing impaired person
- Normal person who knows sign language can use SLIK, which converts sign language to text, to communicate with hearing impaired people too
- With version 2, signers from different countries can use different sign language protocols to converse with each other
- With version 3, signers can converse with people from all over the world, because SLIK will integrate the functionality to output speech in different languages