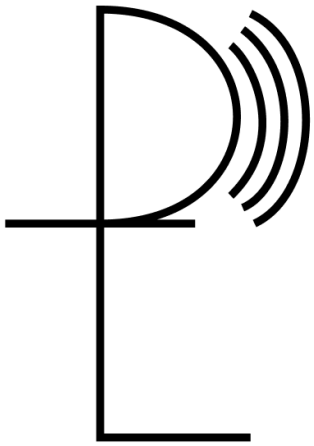
Boston University
Electrical & Computer Engineering

EC 463 Senior Design Project

First Prototype Report

Portable Language Translator



By

Team 23

Portable Language Translator

Team Members

Yohan Kim yohank@bu.edu
Ryan Liao ryanliao@bu.edu
Cristian Palencia cris0211@bu.edu
Andrew Nguyen aynguyen@bu.edu

Equipment and Setup:

For the language translation on computer, a react web app is used to control the base language and voice settings of the program, which it sends using a flask api. One device is used for the conversation. The users must wait for the translation and playback to fully finish before beginning the next phrase. The program captures audio, splits it into chunks and uses voice activity detection software to determine if someone is speaking. Once it detects no speech, it sends the audio file to the google cloud for the transcription, translation, and finally text-to-speech. The audio file is sent back to the program which it plays from the system's speaker.

For the ASL detection on computer, the external camera (USB 2.0 Camera) will record a live video stream and the script will initialize the video feed. Mediapipe Hands will detect the hand in the video feed and locate 21 key landmarks along the user's hand which are then normalized and prepared for classification. The keypoints are passed into the KeyPointClassifier, which is a pre-trained ML model that matches the keypoints to an ASL alphabet gesture. Each frame is checked to identify the letter being signed and the predicted letter is displayed on the laptop screen. If the signed label is held stable for .75 seconds then the detected letter will be put into the detected string. The user will consecutively sign letters until their word is complete. Once 2 seconds have passed, the detected string will be updated into the word list and the user may continue signing their next word if needed. Once the user has stopped signing their sentence for 3 seconds, the finalized word list/sentence is moved into a txt file.

For the hardware setup we have a Raspberry Pi 3 that interfaces with all the relevant components. For one, we use the GPIO pins to connect to the SPI interface for the OLED display. The OLED display interfaces with the adafruit_ssd1309 library to send over the appropriate signals over the GPIOs and uses PIL to allow for relevant drawing capabilities to the OLED. Additionally, we have our I2S MEMS microphone also connected to the Pi's GPIO pins. The microphone directly interfaces with the sounddevice library which handles reading from a connected I2S audio device such that we can generate a .wav file recording. Finally, we have a USB 2.0 Speaker that, using the pygame library, can output the audio recorded by the microphone, or any provided .wav file.

Measurements and Data:

Measurements and Data:

Language tests

Accuracy: We wanted to collect some data to demonstrate the accuracy and functionality of our language translation code. The data collected in Tables 1-6 below demonstrates different translation paths such as English - Korean and Korean - Spanish. For each row we notice: "Phrase Said" being the words spoken by the user, "Phrase Recorded" being the words picked up by the software, "Translated Phrase" being the final translation that is determined and outputted by the software. Under "Errors?" we write the degree to which the translation and transcription were incorrect. Transcriptions defines whether the actual translation was wrong while the Translation then describes whether given a transcription we were able to correctly output it.

Phrase Said	Phrase Recorded	Translated Phrase	Errors
Hello, how are you?	hello how are you	안녕하세요. 어떻게 지내세요	Transcription: no errors Translation: no errors
Can I get a large coffee with two sugars and no milk, please?	can I get a large coffee with Two Sugars and no milk	Two Sugars와 우유 없이 큰 커피 한 잔을 마실 수 있나요?	Transcription: no errors Translation: no errors
Excuse me, how do I get to the nearest train station?	excuse me how do I get to the nearest train station	실례합니다. 가장 가까운 기차역까지 어떻게 가나요?	Transcription: no errors Translation: no errors

Table 1: English - Korean Translation

Phrase Said	Phrase Recorded	Translated Phrase	Errors
Do you have fresh strawberries or are they out of season?	you have fresh strawberries or are they out of season	¿Tienes fresas frescas o están fuera de temporada?	Transcription: missing word Translation: no errors
Can I try this jacket on in a medium size?	can I try this jacket on in a medium sized	¿Puedo probarme esta chaqueta en talla mediana?	Transcription: minor error Translation: no errors
What time are we meeting tomorrow for lunch at the new	what time are we meeting tomorrow for lunch at the new	¿A qué hora nos reuniremos mañana para almorzar en el	Transcription: missing word Translation: no errors

restaurant?		nuevo?	

Table 2: English - Spanish Translation

Phrase Said	Phrase Recorded	Translated Phrase	Errors
¿Hola, tienes fresas frescas o están fuera de temporada?	Hola tienes fresas frescas o están fuera de temporada	안녕하세요. 신선한 딸기가 있나요? 아니면 제철이 아니나요?	Transcription: no errors Translation: no errors
¿Puedo probarme esta chaqueta en talla mediana?	puedo probarme esta chaqueta en talla mediana	이 재킷을 미디엄 사이즈로 입어봐도 될까요?	Transcription: no errors Translation: no errors
¿A qué hora nos reuniremos mañana para almorzar en el nuevo restaurante?	a qué hora nos reunimos mañana para almorzar en el nuevo restaurante	우리 내일 몇 시에 새 식당에서 점심 먹으러 만나요?	Transcription: no errors Translation: no errors

Table 3: Spanish - Korean Translation

Phrase Said	Phrase Recorded	Translated Phrase	Errors
안녕하세요. 어떻게 지내세요	안녕하세요 어떻게 지내세요	Hola, cómo estás	Transcription: no errors Translation: no errors
우유 없이 설탕 그게 큰 커피 주세요	우유 없이 설탕 그게 큰 커피 주세요	Me gustaría un café grande con dos azúcares y sin leche.	Transcription: no errors Translation: no errors
실례합니다. 가장 가까운 기차역까지 어떻게 가나요?	실례합니다 가장 가까운 기차역까지 어떻게 가나요	Disculpe, ¿cómo llego a la estación de tren más cercana?	Transcription: no errors Translation: no errors

Table 4: Korean - Spanish Translation

Phrase Said	Phrase Recorded	Translated Phrase	Errors
신천한 딸기 있나요 아니면 제철이 아니나요	신천한 딸기 있나요 아니면 제철이 아니나요	Are there any fresh strawberries or are they out of season?	Transcription: no errors Translation: no errors

이 재킷을 M 사이즈로 입어봐도 될까요	이 재킷을 M 사이즈로 입어봐도 될까요	Can I try that jacket in size M?	Transcription: no errors Translation: no errors
우리 내일 몇 시에 새 식당에서 점심 먹으러 만날 거야	우리 내일 몇 시에 새 식당에서 점심 먹으러 만날 거야	What time are we meeting for lunch at the new restaurant tomorrow	Transcription: no errors Translation: no errors

Table 5: Korean - English Translation

Phrase Said	Phrase Recorded	Translated Phrase	Errors
Hola, cómo estás	Hola cómo estás	Hello how are you	Transcription: no errors Translation: no errors
¿Puedo tener un café grande con dos azúcares y sin leche por favor?	puedo tener un café grande con dos azúcares y sin leche por favor	can I have a large coffee with two sugars and no milk please?	Transcription: no errors Translation: no errors
Disculpe, come alcanzo llegar al estacion de tren mas cercano?	Disculpe cómo alcanzo a llegar a la estación de tren más cercano	excuse me, how do I get to the nearest train station?	Transcription: minor errors Translation: no errors

Table 6: Spanish - English Translation

Latency: Referring to Table 7 below, we were able to record some measurements for the latency of our software for different sized phrases. For context, the data was collected using English - Spanish translations (English phrase translated into Spanish). In our collected data, latency is measured by the time the untranslated audio is sent to the cloud to when the translated audio playback starts. This was measured by the python timer method to provide accurate times.

Specifically, as we move down the table we observe that we are testing with larger and larger phrases to test the correlation between different sized phrases and latency.

ASL Test

Accuracy: With our ASL software we wanted to test that if a user were to attempt to enter in a phrase, that both the initial ASL letter detection and autocorrect model would work appropriately to read input and correct it after the fact if needed. Table 8 below demonstrates an attempted phrase at different distances as input and demonstrates the output of the ASL model and then the output of the Autocorrection model.

Phrase	Distance from Camera	Before Autocorrect	After Autocorrect	Errors?
Where is that?	.33 meters	Where is that	Where is that?	No
Where is that?	.67 meters	Where is that	Where is that?	No
Where is that?	1 meter	Whexe i ss thaa t	Where is that?	No
The car is red.	.33 meters	The car is re d	The car is red.	No
The car is red.	.67 meters	The car is r e d	The car is red.	No
The car is red.	1 meter	The cab is r e d	The cab is red.	Yes, B was recorded instead of R
That sounds so exciting!	.33 meters	That sounds so exciting	That sounds so exciting!	No
That sounds so exciting!	.67 meters	That sounds sso exciting	That sounds so exciting!	No
That sounds so exciting!	1 meter	That soumdss so exciting	That sounds so exciting.	Yes, it did not add the exclamation mark
Hi my name is Yohan.	.33 meters	Hi my name is yohan	Hi, my name is Yohan.	No
Hi my name is Yohan.	.67 meters	Hi my name iss yohan	Hi, my name is Yohan.	No
Hi my name is Yohan.	1 meter	Hi my mame is yoham	Hi, my name is Yohan.	No
I went to school today and learned	.33 meters	I w e n t to school and lea r	I went to school and learned how	No

how to read		ned how to read	to read.	
I went to school today and learned how to read	.67 meters	I we n t to scool a nd lear n ed how o read	I went to school and learned how to read.	No
I went to school today and learned how to read	1 meter	I we nt sto sc hool and learn e d how to read	I went to school and learned how to read.	No

Table 8: ASL Translation Data - Table demonstrates how an intended phrase is translated at different distances using our ASL and Autocorrection models.

Latency: We also wanted to record measurements for the latency or processing time for the Autocorrect model took to take a recorded phrase and correct it. For a given phrase, we recorded the time it took to determine if it required correction and to actually correct it. Latency is measured by the time the last letter is detected to when the final translated sentence is returned. The program uses the python timer method to provide accurate times. Table 9 demonstrates the different phrases and then the corresponding latency in seconds.

Phrase	Latency (seconds)
Where is that?	3.62
The car is red.	3.5
Hi my name is Yohan.	3.92
That sounds so exciting!	3.55
I went to school today and learned how to read	3.75

Table 9: ASL Latency Measurements - Demonstrates how long it took to run a given phrase through the autocorrection model (a given phrase could be mispelled as well, the latency timing is unaffected by small grammar or spelling errors)

Microphone Test

Functional Distance: In order to test the functional distance of the microphone, we first had to determine a method to rank a given audio recording. Referencing Table 10, we were able to come up with a scoring guideline for any given audio recording. With this scoring guideline we were then able to test the microphone at different distances and at varying audio levels of the phrase: “This is a test recording of the microphone”. Based on the audio recording we would then be able to give it a quality score between 0 and 4 (shown in Table 11).

Quality Score	Meaning
0	Inaudible
1	Muffled and distant
2	Language can be understood
3	Clear audio
4	Perfect audio recording

Table 10: Quality Score Table - Ranking given to a audio recording based on overall quality.

Distance (m)	Whisper	Conversational Speaking	Yelling / Loud Speaking
0.25	3	3	4
0.50	3	3	3
0.75	2	3	3
1	2	2	3
1.25	2	2	3
1.50	1	1	3
1.75	1	1	3
2	1	1	3

Table 11: Microphone Performance - Performance of I2S microphone at varying distances and audio levels.

Conclusions:

Language

The accuracy of the transcription and translation is very accurate with only minor errors. Even with these errors, any ASL individual reading the transcribed text or other party speaking a different language should be able to understand the meaning of each phrase. This accuracy is also above our minimum expected percentage accuracy. The translation latency, understandably, increases based on the size of the phrase it has to translate. The latency of the software will also increase with more languages added to its functionality, but we have specified English, Korean, and Spanish in our project description, so the latency will not change from this. Based on our data, even with a long phrase that is unlikely to be used during a conversation as it involves multiple sentences, the latency is significantly under our benchmark of 5 seconds. With most common phrases having a latency of under 1 second, our language software should be able to seamlessly translate bidirectionally without interrupting the flow of a conversation.

ASL

The raw output from the ASL recognition model shows occasional errors in spelling, spacing, and letter substitutions, particularly when the distance from the camera increases to 1 meter. This indicates a slight decrease in the model's accuracy as the distance grows. Despite these errors, the autocorrection effectively resolves the majority of issues, including spelling, punctuation, and spacing errors. Even at the furthest distance, the autocorrection successfully fixes most recognition mistakes, significantly improving the overall quality of the output. Regarding latency, the system takes an average of 3.57 seconds to process and correct a sentence. This processing time remains relatively consistent across different phrases, with minimal variation regardless of the phrase's length or complexity.

Microphone

One of our requirements of the project is defined by how far someone can be from the device for full translation functionality. The requirement we defined was a distance of one meter away from the device. From the collected data, we notice that at even a whisper we are able to understand language spoken to the microphone at a distance of one meter away. Additionally, we observe that, as expected, below the one meter requirement that the microphone performance is increased and is decreased at distances away from the one meter requirement. Although, we are able to obtain good audio that can be translated by the software, it would be helpful to guarantee even greater audio quality from the microphone since there is the room for improvement. For this reason, our group will most likely pursue improvements to the microphone whether it be implementing audio amplifiers or filters to the system.