Real-Time Translator with ASL Interpretation (RTAI)

University of Central Florida
Department of Electrical and Computer Engineering

Initial Project and Group Identification Document
Senior Design I
Dr. Samuel Richie
Group 14

Members

Gustavo Camero — Computer Engineering — gcamero@knights.ucf.edu

Jared Spinks — Computer Engineering — spinks.jared@knights.ucf.edu

Luis Hurtado — Electrical Engineering — luis.hurtado@knights.ucf.edu

Michael Loyd — Electrical Engineering — michael.loyd@knights.ucf.edu

Client/Advisor

Dr. Chung Yong Chan — chungyong.chan@ucf.edu

Table of Contents

Project Narrative

Translation provides the channel for maintaining a meaningful conversation between two individuals who speak different languages. Human based translations have been widely applied in many aspects of society including the financial, medical, legal, and travel industries [1]. The process of translating through human interpreters, however, proves to be erratic, difficult to arrange, and expensive [2]. In a hospital setting between a doctor and the patient, a human interpreter listens to the patient speak for several minutes and then provides a summary of the dialogue to the doctors. This abridged response that the interpreter provides the doctor may omit important information of symptoms, medical history, and other details that are critical. Therefore, this lack of information may “lead to misdiagnosis and improper or delayed medical treatment” [3]. Thus, the need for a more accurate, low-cost translation system is required for such applications. Neural machine translation (NMT) provides the next generation of real time translation with minimal errors, and it’s currently implemented by several tech giants such as Google and Amazon. The NMT algorithms effectively produce text-to-text translations that can then be processed into speech and vice-versa. However, these NMTs only perform the translations for spoken languages and don’t consider sign languages such as the American Sign Language (ASL). People who are congenitally deaf or have never developed an understanding of spoken language use ASL as their primary language. They don’t develop the same understanding of the language as an individual who is able to listen and speak the language [2]. Thus, a text-to-text translation is not sufficient for people that have the disability. This is where the NMT algorithms’ primary ability to translate text-to-text falls short of providing the service to ASL users. There is current research being conducted to train NMTs to produce image generation techniques, however, this is still far from commercialization [4]. In-person human interpreters used for English to ASL translation also proves to be challenging to implement. Hospitals require specialized translators who can interpret ASL accurately and this service requires significant upkeep. Other alternatives sought by hospitals for in-person translation include video conferencing with remote interpreters, however, this is still costly and difficult to implement [2]. In this document, we are proposing a real time translator with virtual ASL interpretation. This device will be designed to require minimal overhead, low budget, and be accurate on speech acquisition and translation delivery. The functionality of the translator is to provide two translation modes: the Speech-to-ASL Translation (SAT) mode, and the Speech-to-Speech Translation (SST) mode. The SAT mode will focus on translating spoken language to ASL, while the SST mode will focus on conventional translation.

The product will consist of a built-in speaker and microphone, a touch screen LCD, and peripheral LEDs. For the SAT mode, the built-in microphone will record the user’s speech and the speech will be translated into text and then back into speech that will output on the built-in speakers. The SAT mode will also display the original and translated text on the LCD monitor. For the SST mode, the product will utilize the touch-screen LCD monitor to display the sign language output using a 3D modeling software. The SST mode will also include functionality from the SAT such as displaying the original and translated text. For powering the product, we have several approaches. Our first approach would consist of a portable battery supply. This internal battery would supply power to the display, the GPU, etc. We are considering the use of a lithium-ion battery that will be able to power the console for at least three to four hours on a single, full charge. The battery will be rechargeable from a standard US AC wall outlet (120V at 60 Hz). Our other approach would to have the unit be plugged into the wall; that is, the power supply to the system would be from an external source, which would be a standard US AC wall outlet. This is possible since the unit is being developed for medical use, such as in hospitals. Most machines and stations inside of hospitals are installed into rolling carts; and when needed, they are rolled into the patient's room and plugged into the wall. Our product will utilize Wi-Fi to connect to online NMT service provided by companies such as Google, Amazon, or Windows. These services include speech recognition, real time translation, and speech to text. We decided that utilizing APIs for the text to text translation will allow us to focus more on the speech to sign language translation. It will also extend the functionality of the product to encompass several languages with the pretrained models available from these online services. It will ensure that the translated text is as accurate as possible. We intend to utilize several online/cloud translation algorithms in parallel; this way, we can combine common segments between the translations to produce a more accurate speech output. With the speech-to-speech translation processing offloaded to cloud services, we would have more data storage and processing power available for the speech-to-ASL translation.

Other products on the market are catered for portable, general-purpose use. The ONE Mini, for example, has features such as voice recording and AI translation with live translation input [5]. Since the use of voice recording is integrated onto the device, and these recordings can be edited and sent via Bluetooth through its mobile application. This device would not be useful in private settings, such as hospitals, where voice recording would be a HIPAA violation. Another product is the WT2 Plus from Timekettle, which consists of two earpieces and a mobile application. This device has three modes of functionality. One mode provides continuous translation between the two earpieces. The second mode allows for segments of speech to be translated while holding a button. The third mode allows for one user to use the earpiece while the receiver uses the microphone and speaker of the phone. There is an alternative product from this company called the ZERO that attaches to the Lightning port of an Apple device. ZERO utilizes the microphone and speaker of the mobile device for two-way communication [6]. The ZERO and WT2 by Timekettle are still in their developmental stages but use recent translation algorithm technologies. Both the ZERO and the WT2, despite their portability, are too small, are easily lost, and are not traceable since there is no docking station for these items. Our design is intended to be bulkier but be more durable for a medical environment. Another final competitive product on the market is services by Stratus Video. This company provides iPads attached to a metal stand, which are lent to customers under a right-to-use policy. The customers only pay for a monthly subscription per stand. The major downside of this method is the constant overhead; i.e. the monthly payment for use of the translation services. Another downside is the use of human translators: although the use of human translation provides a more natural and consistent translation, the presence of another human for translation is limited by their availability; that is, not all languages are provided 24/7 by Stratus. Also certain languages are limited to only a few translators, so certain translators may be busy providing services to other customers [7]. Our design is intended to be a translation device that can be used for several languages that are available at any time.

Our proposed real time translator with ASL interpretation will bridge the gap between universal machine translation and physical human translation. This is critical in modern society where the greater population, including those with disabilities, rely on advanced artificial intelligence to enrich their daily lives. The goal of our RTAI is to provide a user-friendly experience, effective sign language rendering and a high-level of accuracy. Ultimately, RTAI aims to provide patients with hearing disabilities the capability to have effortless and natural communication with medical professionals.

Requirement Specifications

- Use of Wi-Fi connectivity to access cloud based services.
- Microphone capable of capturing human voice from 3-5 feet away.
- Microphone with frequency response best suited for voice (80Hz-15kHz).
- Speakers capable of amplifying sound enough for an average room size of 320 sq. ft.
- Custom housing 12in by 12in by 6in, large enough to accommodate the GPU, Memory, Display, Wi-Fi Module, Speakers.
- Back lit display with sufficient pixel density in order for user to interpret the sign language from a distance of 1-3 feet.
- Graphics processing unit (GPU) capable of processing real time rendering of the sign language models.
- Sufficient memory to store sign language models. Be able to display said models smoothly.
- Be able to support both power outlet and battery back-up capabilities
- Battery technology that can allow the device to operate for 3-4 hours while constantly translating.
- Be able to translate from English text to a text based ASL

Possible Project Constraints

Our group may face different types of limitations due to the technology we are trying to use. One of the biggest challenges, we might face along the road, is the fact that our device is always on the need of Wi-Fi capabilities in order to use the cloud base translation services. Moreover, finding the technology required to bring an accurate and robust system will be expensive as usually the accurate peripherals require a large upfront investment. Finally, since we are trying to display 3D models of the ASL on a display. This will require both a large storage capacity to store said models and a high density pixel display to show the motion of the models smoothly which adds more costs to our project.

Our group may also face the difficulty of translating text to a text-based sign language. To do this we will need to utilize an NMT that will allow us to teach the computer how to translate english text to our chosen notation. We will face possible limitations of accuracy and time when it comes to training the model. This will require a large amount of resources and time to train the model and the processing power on board to translate the text on board the system.

House of Quality

							-				
									-		
				-			-			-	
		-		-			-	-	-	-	-
		Engineering Requirements									
	Polarities	-	+	-	+	-	-	+	+	+	,
Polarities	Customer Requirements	Dimensions	Battery Life	Cost	Signal Strength of WiFi	Storage Capacity	NMT Limitations	Speaker Range	Microphone Accuracy	Display Pixel Density	Power Output
+	Real Time Translation					▲					
+	Powerful GPU	▼	▼▼	▼▼		▲					▼▼
+	Multi-Language Support			▲	▲▲	▲▲	▲				
+	Accuracy of Translation				▲▲		▼▼		▲		
-	Cost	▼		▼▼	▲▲		▲	▲	▲		
+	ASL Models					▲▲					▼
+	Speaker Clarity			▲							
+	Screen Clarity	▲	▼	▲		▼				▲▲	▼▼
	Targets for Engineering Requirements	12in by 12in by 6in	Between 3-4 hours	Approximately 500 USD	Approximately 3 to 5 feet	Approximately 500GB	Around 128 languages	320 sq-ft	80Hz - 15kHz Range	16-Inch Display	30 Watt-Hour

Polarity	
Positive	+
Negative	-

Correlations	
Strong Positive	▲▲
Positive	▲
Neutral	
Negative	▼
Strong Negative	▼▼

Figure 1: House of Quality of RTAI

Block Diagrams: Software Flowchart

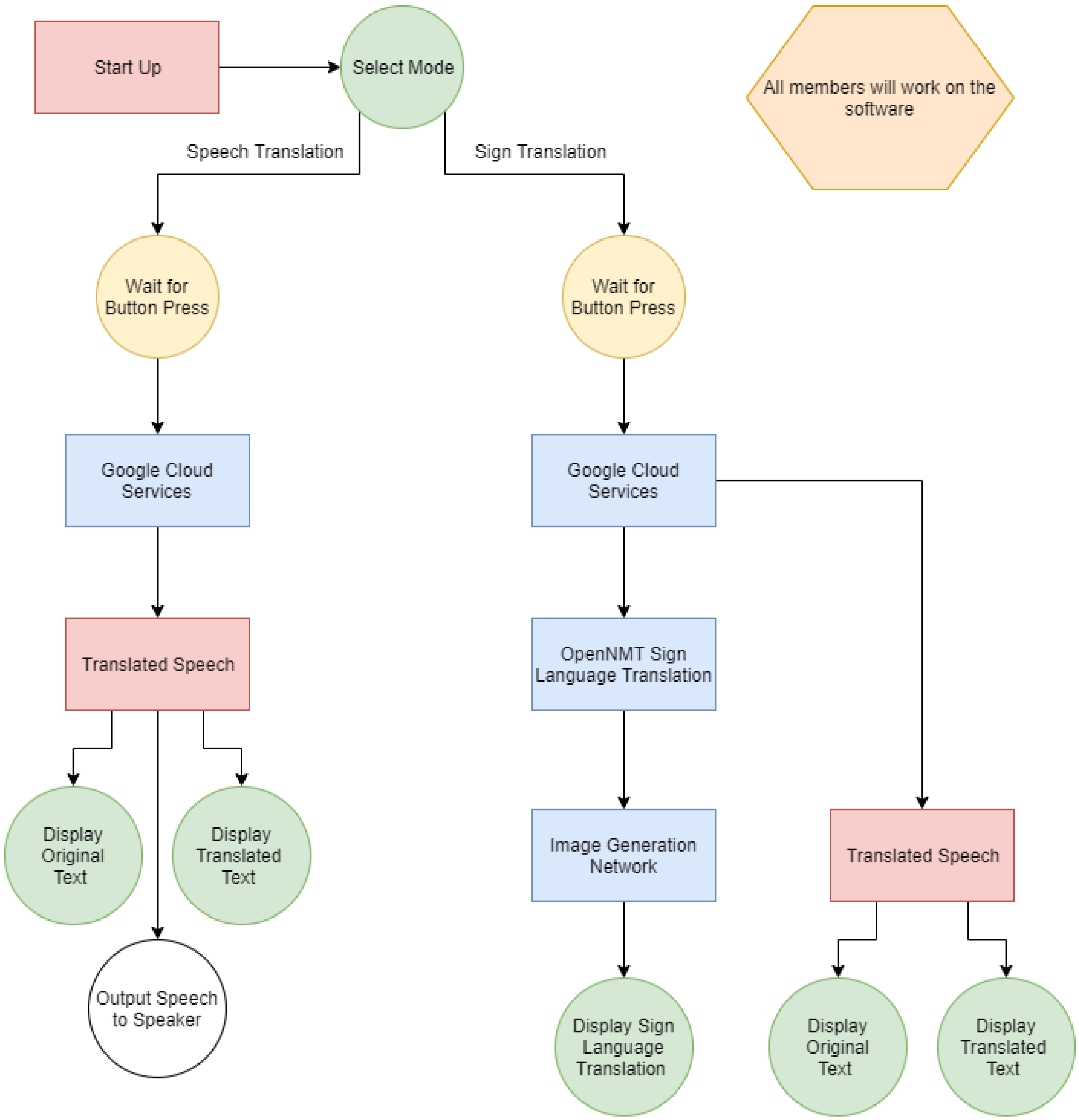

Figure 2: Software Flowchart of RTAI

Hardware Block Diagram

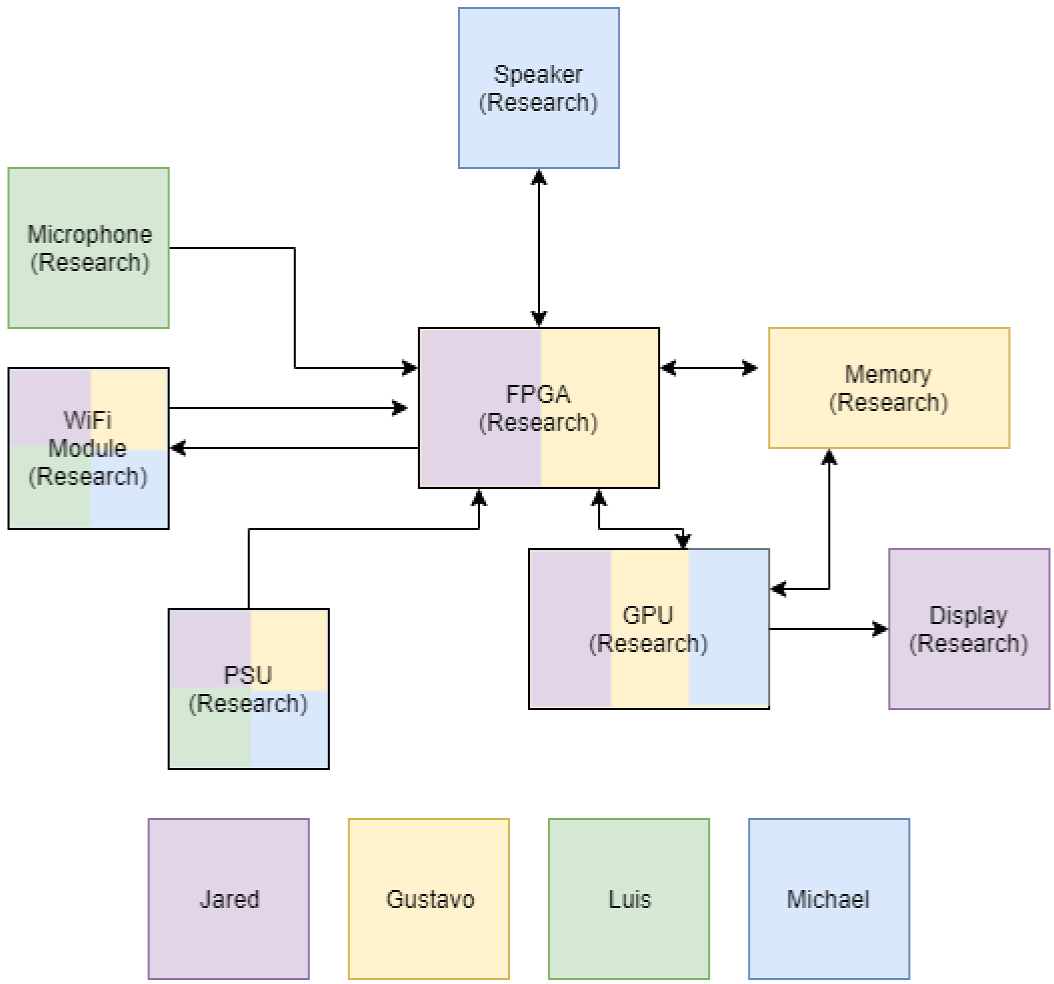

Figure 3: Block Diagram of RTAI

Budget and Funding

Table 1: Budget Approximations for each Component

GPU	$50.00 - $200.00
Display	$80.00 - $200.00
Speaker(s)	$10.00 - $20.00
Microphone	$10.00 - $20.00
Battery	$50.00 - $100.00
Wi-Fi Module / Antenna	$20.00 - $50.00 / $10.00 - $20.00
PCB	$10.00 - $100.00
Miscellaneous	$10.00 - $50.00

Given our above numbers, our finalized project will cost a minimum of $200 and and a maximum cost of $800. We are planning to have a budget of $1100. We plan to have a maximum budget of $800 for the finalized project, with an extra $300 for prototyping.

Project Milestones

Table 2: Senior Design I - Milestones

Process Description	Duration	Dates
Brainstorming & Project Selection	2 weeks	Aug. 30 - Sept. 14
Initial Research / Divide & Conquer	1 week	Sept. 14 - Sept. 20
Continuation of Research	3 weeks	Sept. 20 - Oct. 12
Possible Hardware & Software Selection, Purchase Demos For Preliminary Testing	3 weeks	Sept. 20 - Oct. 12
Cover page/Executive Summary/Technical Contents	4 weeks	Oct. 12 - Oct. Nov. 1
Administrative Content/Project Summary/Conclusion/Appendices	2 weeks	Nov. 1 - Nov. 15
Final Document Revision + Initial Prototyping	2 - 3 weeks	Nov. 15 - Dec 2

Table 3: Senior Design II - Milestones

Process Description	Duration	Dates
Complete Prototype	4 weeks	Jan. - Feb.
Prototype Testing and Debugging	4 weeks	Feb. - March
Product Finalization	4 weeks	March - April
Peer Presentation		TBA
Final Report		TBA
Final Presentation		TBA

Important Deadlines

- September 20th - Initial Divide & Conquer Document
- October 4th - Updated Divide & Conquer Document
- November 1st - 60 Page Draft Senior Design I Documentation
- November 15th - 100 Page Submission
- December 2nd - Final Document Due

References

[1] "6 Big Industries that can benefit from language translation services." Pangeanic. Boston, MA. Feb. 24, 2015. Accessed Sept. 18, 2019. https://www.pangeanic.com/knowledge_center/6-big-industries-can-benefit-language-translation-services/#

[2] Leila Miller. "'I was panicked": Deaf patients struggle to get interpreters in medical emergencies." STAT. Boston, MA. May 22, 2017. Accessed Sept. 18, 2019. https://www.statnews.com/2017/05/22/deaf-patients-interpreters/

[3] U.S. Department of Justice. **"**Communicating with People Who Are Deaf or Hard of Hearing in Hospital Settings." Civil Rights Division: Disability Rights Section. Washington, DC. August 11, 2005. Accessed Sept. 18, 2019. https://www.ada.gov/hospcombr.htm

[4] Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, and Richard Bowen. "Sign Language Production using Neural Machine Translation and Generative Adversarial Networks." Centre for Vision, Speech, and Signal Processing. University of Surrey: Guildford, UK. Aug. 16, 2018. Accessed Sept. 19, 2019.

[5] "A pocket multilingual assistant." onemini. Wuhan, Hubei, China. May 5, 2019. Accessed Sept. 19, 2019. https://www.transnone.com/

[6] "AI Translator Earbuds." Shenzhen Timekettle Technologies, Ltd.

[7] "About Stratus Video." Stratus Video. https://www.stratusvideo.com/