

Wireless Online Real-time Language Expression Yielder

Ian Oberbeck & Yubo Cao & Anish Goyal
GOSA Governor's Honors Program 60 Engineering

Background

GOAL

The goal of **PROJECT WORLEY** is to create a wireless online real-time language expression yielder (WORLEY) that can be used to act out American Sign Language (ASL) live. Specifically, WORLEY will be able to:

- Recognize speech and translate it into ASL in real time.
- Act out the translation on the robotic hand through remote general-purpose input/output (GPIO) control.

MOTIVATION

Our friend Cory suffers from hearing loss, making communication difficult. We aimed to develop a device to improve communication and quality of life for Cory and others in similar situations.

This device can also be used for Pre-K and K-12 students. While it may seem easy for parents to learn ASL and teach it to their kids, ASL is a complex language with its own grammar, syntax, and nuances. WORLEY provides the necessary language exposure & support for hard-of-hearing children and their parents.



Figure 1: WORLEY's robotic hand.

Introduction

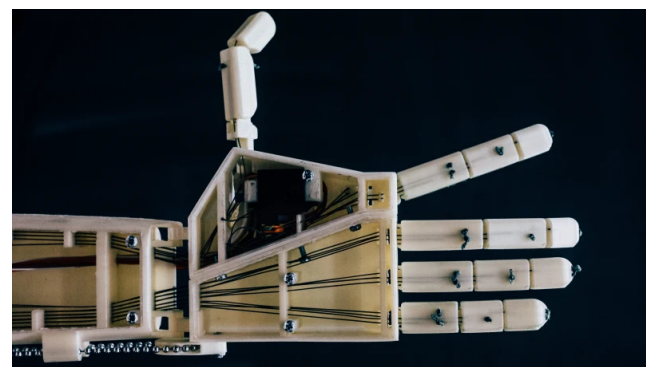
One of the key distinguishing features of this project is its real-time interpretation capability. Unlike existing systems that require a delay between input and output, our team developed a solution that could instantaneously translate spoken words into sign language.

Moreover, we designed WORLEY with the grammar and syntax of ASL in mind. ASL gloss is a shortened representation of traditional English with its own distinct syntax and grammar. Using transformers, WORLEY can translate spoken English into ASL gloss, which is then converted into ASL signs.

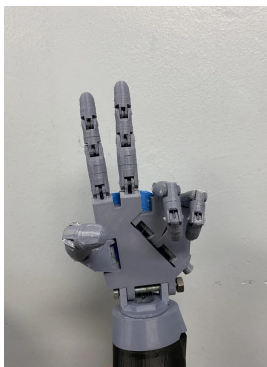
Another significant aspect of WORLEY lies in the design of the robotic hand. The hand does not have a reset stage between frames, which optimizes the signing process. Additionally, we designed an efficient rotation scheme for the thumb that reduced the DoF by 2, reducing the cost of the project and making it more accessible.

Significance

EXISTING SOLUTION



(a) Project ASLAN



(b) Instructable Hand

Despite the existing solutions, there are still many problems:

- They merely perform manually-coded English (MCE).
- They are incapable to act out ASL in real-time and require text input.

SIGNIFICANCE

- In comparison, 37.5 million people in the United States suffer from hearing problems.
- In the United States, 30 deaf children are born every day, and 90% of their parents do not know ASL.
- WORLEY** will help to solve this problem by providing a low-cost, portable, and easy-to-use solution for ASL interpretation.

System Architecture

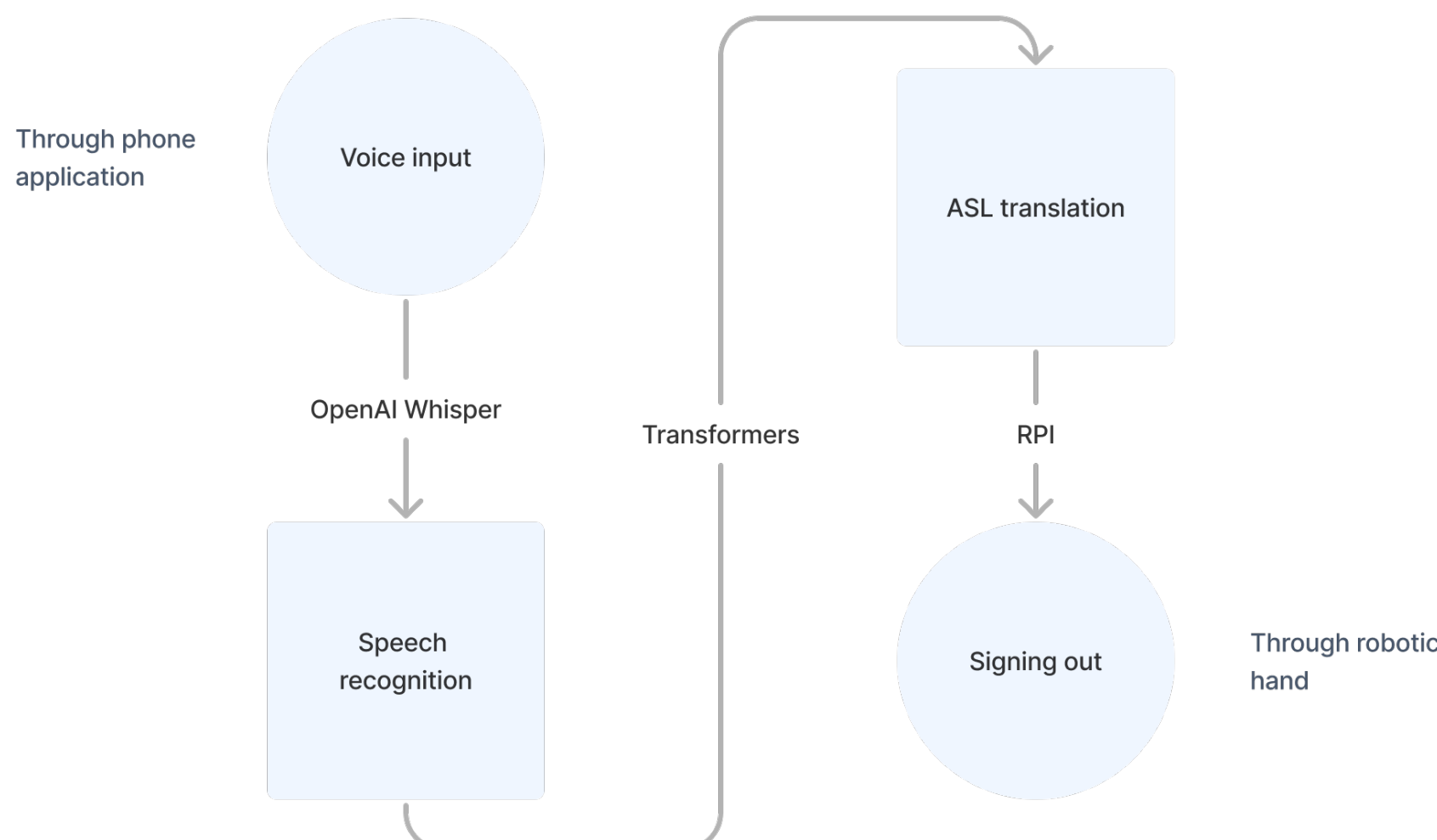


Figure 3: Flowchart of the system architecture

- The user speaks into the Worley application on their phone.
- The application establishes a Web real-time communication (WebRTC) connection with the server and streams the audio.
- The server runs a lightweight voice activation detection (VAD) model **sliero** detects speech. Such results are buffered into a queue along with the timestamps, and a 2-pointer technique is used to extract the longest continuous subsequence, *i.e.*, the speech. Such segments are sent to **whisper** for speech recognition.
- The text is translated to ASL language using a custom-built transformer model trained on the **ASLG-PC12** dataset.
- The ASL translation is sent to the Raspberry Pi, which controls the robotic hand, and the mobile application.

Mechanical Design

CORD ACTUATION

The fingers are curled by servo-driven pull cords. By reducing the available length of the cord, the surfaces of the finger segments are pulled closer to one another. The surfaces are angled 90 degrees; so the finger rotates around the back (where there is no gap) to curl. We utilized ripoff SG90 servos for this task, as they're cost-effective, space-efficient, and have the minimum sufficient torque required to actuate the fingers.

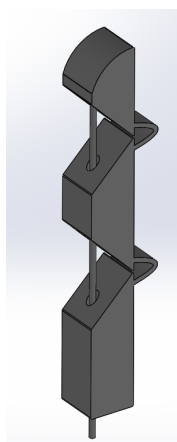


Figure 4: CAD of cord actuation design

COMPLIANT HINGES

Compliant hinges connect the finger segments, allowing the finger to curl as a single part. This is accomplished by 3D printing the fingers in polypropylene (PP); a polymer ideally suited for compliant hinges due to its fatigue-resistant and semi-rigid properties. The hinges are designed using splines constrained by their connection points and length, such that when the segment is rotated 90 degrees it has the proper length to temporarily deform into a circle. The energy stored in this deformation allows the fingers to passively return to their original position when the cord tension is released.

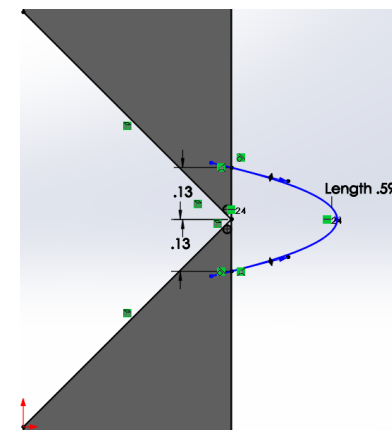


Figure 5: CAD of compliant hinge design

Challenges

- More tolerance would be needed for the servo motors to control the index finger to move behind the palm.
- Whisper model, even the CTranslate2 version, would require $1735 \text{ ms} \pm 139 \text{ ms}$ to translate a 3-second speech, which is not real-time. Deployment of the servo independently from the HTTP server on KServe would be needed to ensure performance and scalability.
- Polypropylene failed to print successfully on the 3D printer; further, the printed version has structural defects & fatigue issues, rendering the hand to be less realistic.

Material

- 5 SG90 servo, servo horns, and screws
- 4 3D printed fingers + 1 3D printed palm
- Polypropylene filament & Polylactic acid filament
- 1 Raspberry PI 4B
- faster_whisper**, **pytorch**, **asyncio**, **pyav**, **aiortc**, **aiortc**, Expo Framework.

Conclusion

FUTURE WORK

Despite WORLEY's current limitations, we believe it has the potential to be further developed and improved. Future work could implement the following:

Speaker diarization The current implementation cannot distinguish between multiple speakers.

Advanced model deployment The current implementation is not *truly* real-time, since the model is not deployed independently from the HTTP server.

Enhance realism We want to make the hand more realistic. This can be done by adding ball joints to the thumb and wrist. We could also add a layer of silicone to simulate skin.

Wire management To make the hand more aesthetically pleasing, we will position the servos inside of a compartment in the upper arm, and run the wires through the arm to the hand.

RESULT

Remarkably, WORLEY came to life in a mere nine days and a budget of just 150 USD. Moving forward, our aspirations lie in crafting a comprehensive prototype given ample time and resources. We triumph in knowing we successfully developed a robotic hand that can act out the ASL alphabet and be controlled using a mobile application with automatic speech recognition.