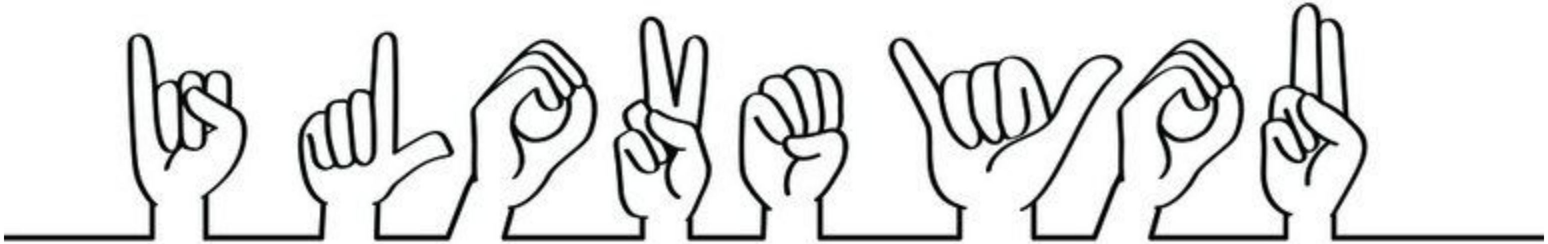
Team EI:

Give Me A Sign

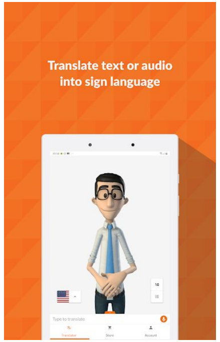
Ran Fang, Leia Park, Sejal Madan



Problem

There exists communication barriers between the deaf community and those who are not familiar with sign language.

Existing solutions: ASL translation apps, wearable devices, human interpreter apps



Hand Talk

BrightSign

Jeenie

Solution & Use Case

Our Goal: Create a real-time sign language translator - a compact, portable device to be seamlessly integrated into everyday conversations, particularly in social settings

ASL user signs into their phone camera → Translates to written English → Displays on dual screens for both users to read

ECE Areas:

Software Systems,
Hardware Systems,
Signals & Systems

Use Case Requirements

Requirement	Metrics & Values	Justification
Person must be near camera so gestures are visible and tracked	Distance: 1-3.9ft from iPhone front camera	iPhone front camera best resolution range[1] + normal conversation dist[2]
Gesture recognition should be accurate	Accuracy: >= 95% for gesture detection and recognition	Average accuracy rate proposed by existing research on MediaPipe and OpenCV gesture recognition [3][4][5]
Translation should be accurate	Accuracy: >= 95% for sign-language-to-English translation	RNN empirical accuracy: 97.76%[6] LSTM empirical accuracy: 95.21%[7]
Translation should be relatively immediate to work as “live subtitles”	Latency: <= 500ms	MarianNMT (Microsoft Translator): 8.9 - 13.9 ms in C++, based on GPU
Good accessibility for positive user experience for both parties involved	User Experience: ~90 % user satisfaction	Google Translate: ★★★★★ 4.3 • 75.2K Ratings Microsoft Translator: ★★★★★ 4.8 • 161.9K Ratings Apple Translate: ★★★★★ 2.3 • 5K Ratings

Technical Challenges

Challenge	Linkage to requirement	Risk mitigation
CV must ignore background distractions and only identify hands	Involve signing to occur 1-3.9ft High accuracy for gesture detection	Consistent testing of CV tracking + Maintain clean camera lens
ML must be well trained to recognize and translate gestures + enforce proper grammar (ex. Grammarly)	High accuracy/low error rate for translation Low latency	Inaccurate delivery of translation mitigated by replacement of previous sentence with new one (rest gesture)
Design of attachment must be easy to put on, take off, secure, and adjust	Good accessibility	3d printing plastic material to reduce any chance of injury from handling
Mobile & web application development	Good accessibility & portability	Debugging practices

Solution Approach

Hardware

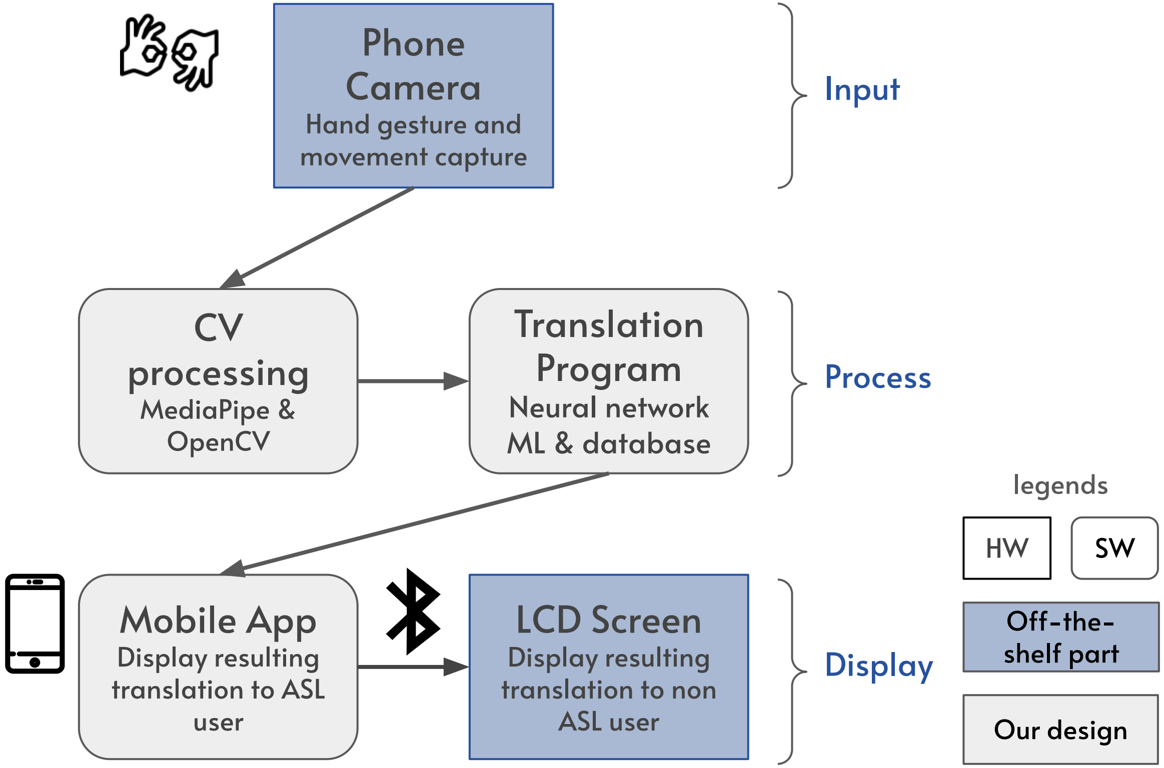
- Arduino
- LCD screen

Signals

- Computer vision

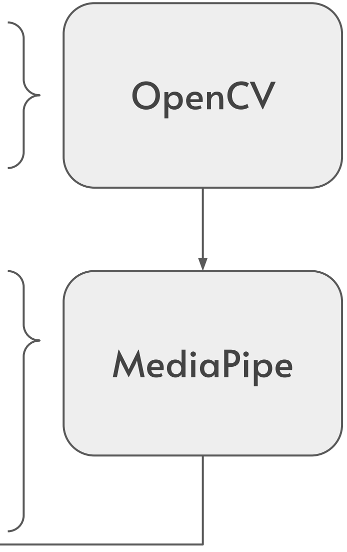
Software

- Mobile App
- ML

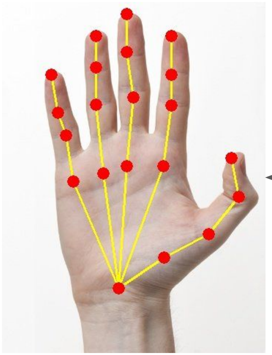


Computer Vision

MediaPipe/OpenCV



- Prepare the input data:
  - Capture frames from camera
  - Use OpenCV to load the video (30-500 frames/sec)
  - Convert the frame received from OpenCV to a MediaPipe's Image object.
- Run the task (MediaPipe)
  - Perform gesture recognition on the provided single image.
- Handle and display results
  - Landmarks: hand-knuckle coordinates (21 per hand)
  - Gestures (8 in library, can be customized)



Next step: ML translation...

Machine Learning

- Gather dataset of sign language gestures and corresponding translation
- Train a neural network suitable for temporal data
  - TensorFlow, Keras
  - Experiment with best one (RNNs, 3D CNNS, hybrid etc)
- Assess the model's performance
  - Training/validation sets
- Model Optimization
  - Loss functions and optimization techniques

Next step: Integrate with generating subtitles...

Hardware & Mobile App

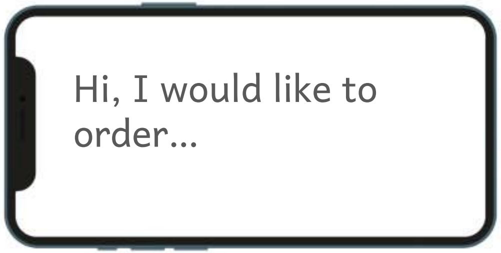
Product:

- Connect LCD screen with Arduino board
  - Screen for display; Arduino for control & bluetooth
- Design phone attachment
  - Adjustable & 3d printed
  - Integrate screen + board seamlessly

Phone application:

- Develop mobile app with Swift using Xcode
- Use CV & ML programs from cloud data storage
  - Receive translations to send to Arduino

Phone attachment + stand

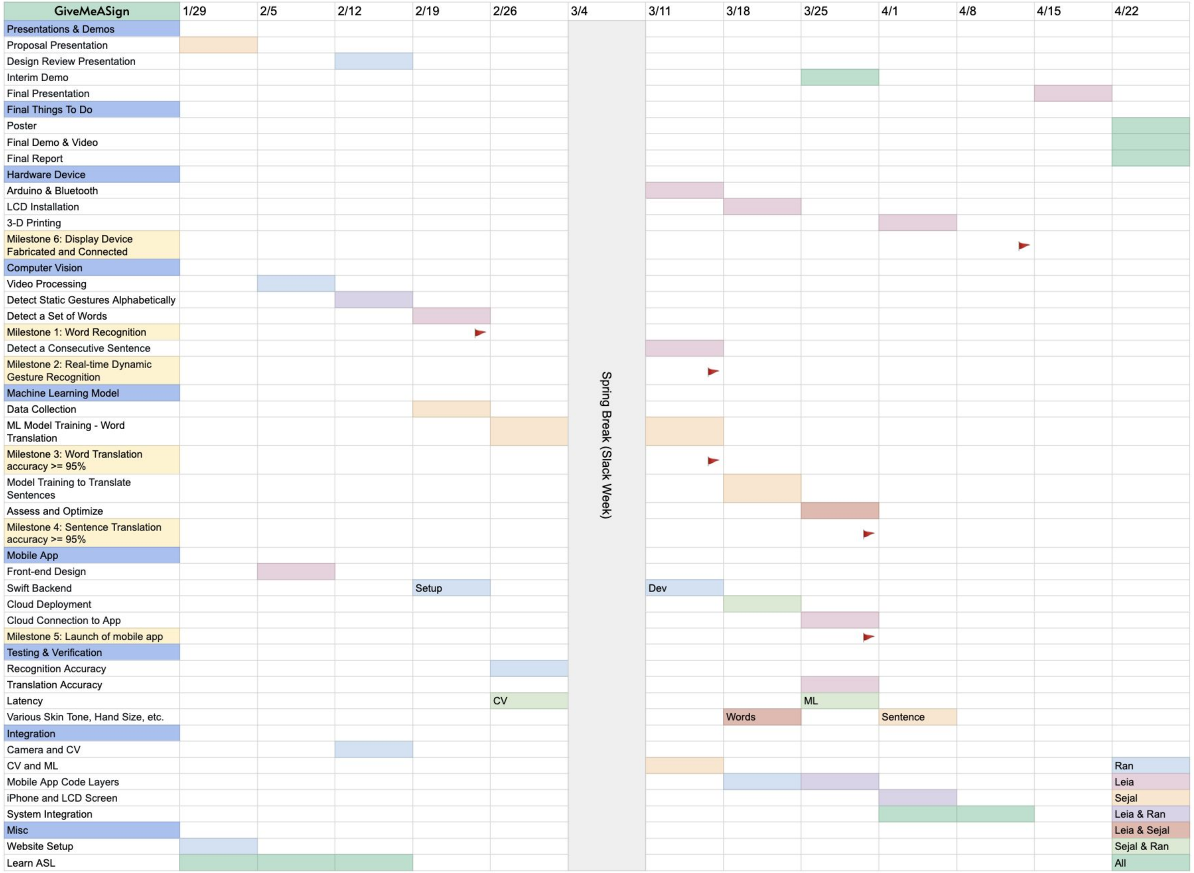


Displays subtitles

Testing, Verification and Metrics

Use-Case Metric	Method for Testing
Involve signing to occur 1-3.9ft from the camera	- Try different distances from camera - See if CV will make a “skeleton” of user
High accuracy (~95%) for gesture detection and recognition	- Use sign language in front of camera: does CV makes proper “landmarks” of the hands and forearms? - Try different room lightings + Add distractions in background
High accuracy (~95%) for sign language translation	- Use sign language in front of camera: Does translation appear? - Try singular words (ex. Hello, bye) - Positive reinforcement given when translation is correct - Sign sentences to observe if translator can handle complexity
Low latency (<500ms) in translation	- Time the speed at which the translation appears after a gesture
~90 % user satisfaction	- Invite friends and other people to try and receive their feedback

Schedule



🚩 Milestones 🚩

1: Word gesture recognition
2: Real-time dynamic gesture recognition
3: Word detection-translation accuracy >=95%
4: Sentence Translation accuracy >= 95%
5: Mobile App launch
6: Display device fabricated and connected

Ran

Video processing with openCV
Hand detection with mediaPipe

Sejal

Word translation ML model
Sentence structuring and optimization

Leia

LCD screen integration
Device fabrication
Mobile app front end design

Conclusion

Through a simple and sleek phone attachment and combined mobile app, we can break down language barriers and ensure accessibility for deaf and hard of hearing community

References:
[1] https://support.apple.com/en-us/105007
[2] Reader, T. M. P. (2019, December 23). Proxemics 101: Understanding personal space across cultures. *The MIT Press Reader*. https://thereader.mitpress.mit.edu/understanding-personal-space-proxemics/
[3] Amit, M. L., Fajardo, A. C., & Medina, R. P. (2022). Recognition of real-time hand gestures using mediapipe holistic model and lstm with mlp architecture. *2022 IEEE 10th Conference on Systems, Process & Control (ICSPC)*, 292–295. https://doi.org/10.1109/ICSPC55597.2022.10001800
[4] Zhu, H., Deng, C., & Zhu, Y. (2023). Mediapipe based gesture recognition system for english letters. *Proceedings of the 2022 11th International Conference on Networks, Communication and Computing*, 24–30. https://doi.org/10.1145/3579895.3579900
[5] J. T. Camillo Lugaresi. (2019). MediaPipe: A Framework for Building Perception Pipelines.
[6] Ren, B. (2020). The use of machine translation algorithm based on residual and LSTM neural network in translation teaching. *PLOS ONE, 15*(11), e0240663. https://doi.org/10.1371/journal.pone.0240663
[7] Li, S. (2018, June 25). Neural machine translation with python. *Medium*. https://towardsdatascience.com/neural-machine-translation-with-python-c2f0a34f7dd
[8] *Breaking the latency barrier for real-time neural machine translation.* (2022, January 7). https://community.intel.com/t5/Blogs/Tech-Innovation/Artificial-Intelligence-AI/Breaking-the-Latency-Barrier-for-Real-Time-Neural-Machine/post/1344750