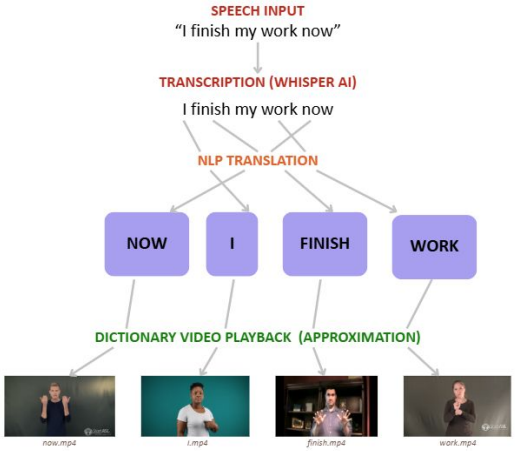
Speech-to-ASL-like Sign Output System

Q1: Engineering Problem & Project Objective

- English-to-ASL translation technology is limited.
- The objective is to design a system that can take speech as input and display ASL sign videos as output.

Q2: Project Design

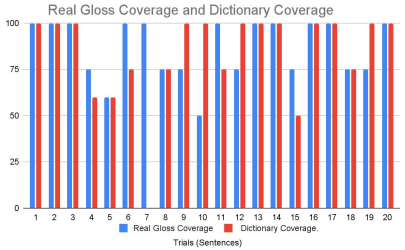
- Speech is transcribed using Whisper AI.
- The transcription is rearranged into ASL gloss.
- The model needs to have basic functionality, at least 70% accuracy, and pass 17/20 trials to be considered successful.



Q3: Data Analysis & Results

- **Overall Success:** The system passed 15 out of 20 trials.

Sentence	ASL Gloss Tokens	ASL Gloss (Realistic)
I finish my work now	['now', 'i', 'finish', 'work']	NOW I FINISH WORK
She opens the door slowly	['she', 'open', 'door']	SHE OPEN DOOR SLOW
We wait for the bus	['we', 'wait', 'bus']	WE WAIT BUS
He closes the window	['he', 'close', 'window']	HE CLOSE WINDOW
They build a small house	['they', 'build', 'house']	THEY BUILD SMALL HOUSE
I carry the bag home	['i', 'carry', 'bag', 'home']	I CARRY BAG HOME
She writes in her notebook	['she', 'write', 'notebook']	SHE WRITE NOTEBOOK
We watch the sun set	['we', 'watch', 'sun']	WE WATCH SUN SET



Q4: Interpretation & Conclusions

- The initial AI model did not perform well due to data scarcity.
- The alternative, rule-based NLP system exceeded the AI model's performance.
- The system shows promise as a Speech-to-ASL translator.
- Future additions to the project may include an expanded dictionary, or increasing the training data for the Text-to-Text Transfer Transformer (T5) model to simplify the pipeline, improve gloss accuracy, and accommodate complex sentences.

Abstract

This project develops a rule-based prototype system that translates spoken English into American Sign Language (ASL) gloss and corresponding sign videos. The system first converts speech to text using OpenAI's Whisper (medium model) and then normalizes the text by removing punctuation and standardizing formatting. SpaCy is used to parse the grammatical structure of each sentence, identifying subjects, verbs, objects, and time expressions. A rule-based algorithm rearranges these components into ASL-like gloss following the Time + Topic + Comment structure. Each gloss token is matched to a sign video in a JSON dictionary using fuzzy string matching with an 85% similarity threshold. Real Gloss Coverage measures the percentage of English words correctly converted to gloss tokens and matched to videos, while Dictionary Coverage measures the percentage of gloss tokens present in the dictionary. A trial is considered successful if both metrics are at least 70%. The system was tested on 20 simple sentences and successfully passed 15 trials. Results indicate that the prototype works well on simpler sentences but is limited by dictionary size and simplified gloss rules. Overall, the system demonstrates potential as an accessible communication tool for Deaf and Hard-of-Hearing individuals and provides a foundation for future improvements.

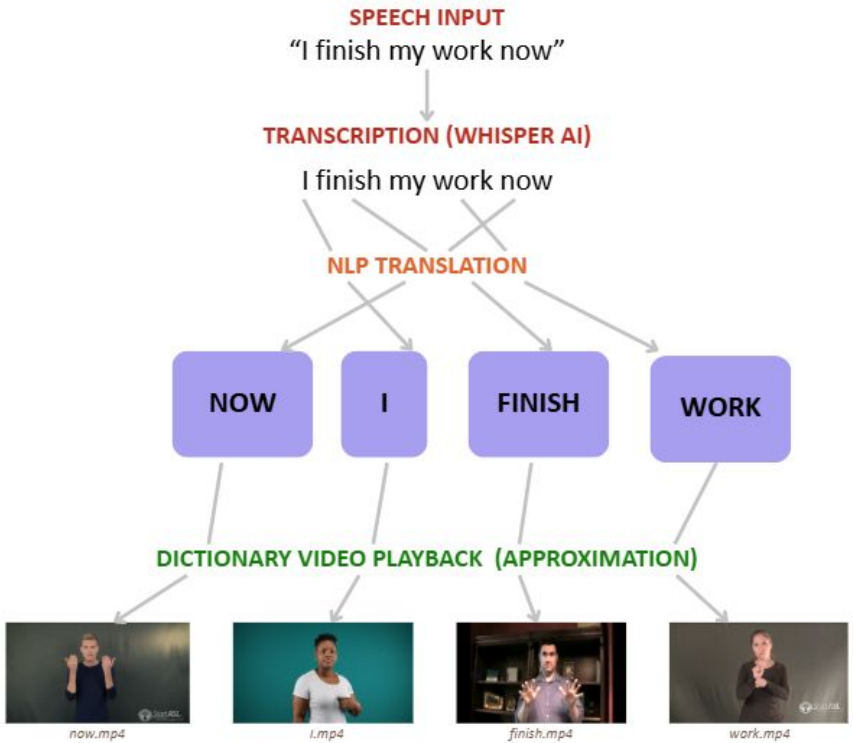
Introduction

- Millions of Deaf and Hard-of-Hearing individuals rely on American Sign Language (ASL) for communication, yet real-time translation between spoken English and ASL remains limited and often inaccessible.
- This project develops a rule-based prototype system that converts spoken English into ASL gloss and corresponding sign videos, aiming to provide a more inclusive and accessible communication tool.
- By combining speech recognition, natural language processing, and structured gloss generation, the system attempts to approximate ASL grammar and visually represent signed output through sequential sign videos.

Methods

- The system converts spoken English into text using OpenAI's Whisper (medium model). The text is normalized and analyzed with spaCy to identify grammatical components such as subjects, verbs, objects, and time expressions.
- A rule-based algorithm then rearranges the sentence into ASL-like gloss following the Time + Topic + Comment structure. ASL gloss is a simplified written representation of sign language concepts without unnecessary English grammar. Each gloss token is matched to a sign video stored in a JSON dictionary using fuzzy matching with an 85% similarity threshold.
- The videos are displayed sequentially to simulate a signed sentence. The system was tested using 20 simple sentences. Real Gloss Coverage is the percentage of key English words correctly converted into gloss tokens and matched to sign videos. Dictionary Coverage is the percentage of predicted gloss tokens that exist in the system's sign dictionary, and a trial was successful if both values were at least 70%.

Methods Cont.

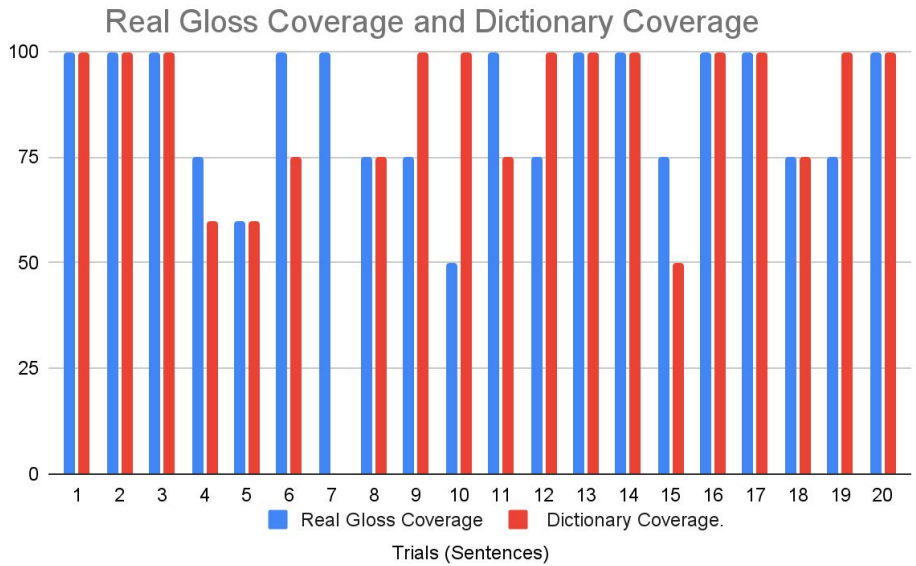


Results

- The system passed 15 out of 20 trials, resulting in a 75% success rate. A trial was considered successful only if both Real Gloss Coverage and Dictionary Coverage were at least 70%.
- The lowest Real Gloss Coverage recorded was 60%, while the lowest Dictionary Coverage was 0%. Real Gloss Coverage reached 100% in 11 trials, and Dictionary Coverage reached 100% in 12 trials.
- Eight trials achieved 100% in both metrics simultaneously.
- Among the failed trials, 60% had Real Gloss Coverage above 70%.

Results Cont.

Sentence	ASL Gloss Tokens	ASL Gloss (Realistic)
I finish my work now	['now', 'i', 'finish', 'work']	NOW I FINISH WORK
She opens the door slowly	['she', 'open', 'door']	SHE OPEN DOOR SLOW
We wait for the bus	['we', 'wait', 'bus']	WE WAIT BUS
He closes the window	['he', 'close', 'window']	HE CLOSE WINDOW
They build a small house	['they', 'build', 'house']	THEY BUILD SMALL HOUSE
I carry the bag home	['i', 'carry', 'bag', 'home']	I CARRY BAG HOME
She writes in her notebook	['she', 'write', 'notebook']	SHE WRITE NOTEBOOK
We watch the sun set	['we', 'watch', 'sun']	WE WATCH SUN SET



Discussion

- The results indicate that dictionary limitations were a major factor in unsuccessful trials. Even when gloss conversion performed well, missing vocabulary reduced overall success.
- The slightly lower frequency of 100% Real Gloss Coverage suggests the rule-based gloss generation could be improved. Expanding the sign dictionary would likely increase Dictionary Coverage scores.
- Overall, the system performs reliably on simple sentences but requires refinement to handle more complex or varied inputs.

Conclusions

- This project demonstrated that a rule-based speech-to-ASL prototype can achieve functional translation accuracy, passing 75% of test trials.
- While the system shows strong potential, its performance is limited by a restricted sign dictionary and simplified gloss-generation rules. The absence of non-manual markers and advanced ASL grammar features also reduces linguistic accuracy.
- Expanding the dictionary and improving the gloss conversion algorithm would significantly enhance performance. Future development could include training a machine learning model to generate more accurate and flexible ASL structures.
- Despite its limitations, this prototype provides a strong foundation for accessible, offline communication tools that could support Deaf and Hard-of-Hearing individuals in real-world situations.

Key References

- OpenAI. (2022). Introducing Whisper. Openai.com. https://openai.com/index/whisper/
- Mitchell, D. (2013). American Sign Language ASL Dictionary. Signasl.org. https://www.signasl.org/
- Provided by ASL Grammar Guide The Academic Center for Excellence 1 ASL Grammar Guide. (2023). https://germanna.edu/sites/default/files/2023-07/ASL%20Grammar%20Guide%20%28edit%207-24-23%29.pdf