Department of Computer Science
Texas State University

Final Report: SIGNIFY

Bailey Dalton onx3@txstate.edu

Tristan Rech jjp176@txstate.edu

Jonathan Villarreal j_v609@txstate.edu

1 Introduction, Motivation, and Problem Definition

When searching for the perfect project, we aimed to choose a topic that not only lacked robust and engaging resources but also piqued our interest. We quickly realized there was a noticeable gap in accessible American Sign Language (ASL) learning applications. Popular platforms like Duolingo[1] and Preply[3] do exist, but they come with significant limitations. These include the requirement to pay for essential features and, in some cases, the reliance on users to identify signs displayed on screen rather than actively signing themselves. Additionally, most apps only quiz users on static signs without incorporating dynamic sign usage, offering a very constrained learning experience. There's also a distinct lack of real-time feedback, making it difficult for learners to measure their progress independently.

To truly enhance the learning journey, we believe it's crucial for learners to do more than merely replicate signs. Our app introduces innovative features like dual learning modes—a traditional learning track and an interactive game mode, that engages different cognitive processes. With real-time feedback and an expansive library of signs, our app doesn't just cover the basics; it pushes learners to immerse themselves fully in ASL. We're committed to redefining expectations and empowering users with the tools to succeed in learning ASL, if users find themselves struggling, it won't be for lack of resources or support. Our application aims to be a comprehensive, user-driven ASL learning experience, fostering both skill and confidence in new signers. Ultimately, our goal is to assist users in developing the ability to communicate in real-world ASL conversations.

2 Background and Development Context

2.1 Technology Used

SIGNIFY utilizes a combination of modern web technologies and innovative algorithmic approaches to enhance the learning experience for users aiming to acquire American Sign Language (ASL). At its core, the application employs plain JavaScript to ensure wide accessibility and compatibility across various devices and platforms, making it a versatile choice for learners worldwide

Web-Based Platform Developing the application as a web-based tool using JavaScript ensures that it is accessible on a broad range of devices, including desktops, laptops, tablets, and smartphones. This approach aligns with modern educational technology trends that prioritize accessibility and user convenience.

2.2 Existing Technology

In the initial phase of our research for a new project aimed at enhancing sign language learning, we carefully evaluated several existing applications to identify their strengths and weaknesses. Among the apps we reviewed, popular platforms like Duolingo[1] and Preply[3] were noted for their comprehensive language offerings. However, they tend to fall short in fostering active participation in signing, primarily because they focus more on language recognition rather than the production of signs. Additionally, many of their advanced features are restricted to paid memberships, which can be a barrier for users seeking extensive learning tools without financial commitment.

In contrast, our project, SIGNIFY, is designed to prioritize active participation. Unlike the traditional models employed by Duolingo and Preply, SIGNIFY introduces an innovative approach where users learn through actual signing practice rather than mere recognition. This hands-on method is intended to enhance muscle memory and improve sign language fluency.

While exploring potential models for real-time user feedback, we discovered fingerspelling.xyz[2], an application that excels in providing immediate feedback to learners. This feature is crucial as it helps users correct their signs in real-time, thereby accelerating the learning process. However, fingerspelling.xyz lacks mechanisms for users to assess their progress or to challenge their signing skills in a competitive or structured manner.

To bridge the identified gaps in existing sign language learning applications, SIGNIFY has successfully integrated a dynamic evaluation system. This system allows learners to monitor their progress through a series of increasingly challenging levels, effectively tracking their improvements over time. Furthermore, we have designed SIGNIFY to be a flexible and expandable platform. It supports the creation of new challenges, additional levels, and achievements, all aimed at testing and enhancing the user's abilities while fostering deeper engagement with the material.

By incorporating these features, SIGNIFY delivers a comprehensive and interactive learning experience, supporting users from the foundational basics to the more complex aspects of sign language proficiency. This approach ensures a strong framework for continual learning and skill advancement, making SIGNIFY a pioneering tool in sign language education.

3 Project Methodology and Implementation

3.1 Development Process

Our project originated from the desire to enhance existing technology and create a unique application. Recognizing the scarcity of intuitive and comprehensive apps for learning American Sign Language (ASL), we saw an opportunity to innovate. Our approach involved devising a strategic road map anchored by three pivotal objectives.

The first milestone was the development of a "Learning Module," designed to teach users finger spelling for word construction. This module not only demonstrates each letter but also allows users to learn through a practical, trial-and-error method. This hands-on approach marked a significant shift from our earlier examples focusing on mere sign recognition and laid the groundwork for more immersive learning experiences.

Building on the foundational skills acquired in the Learning Module, we introduced the "Guess the Word" game. This interactive game challenges users to apply their finger spelling knowledge to form words. Initially envisioned as a variant of hangman, our testing revealed challenges in user penalization, which detracted from the enjoyment. By eliminating penalties, we enhanced the game play, making it more enjoyable and fluid for users.

Our final objective was to surpass the conventional standard by integrating dynamic sign recognition. Utilizing dynamic time warping, we implemented a system that accurately tracks hand movements through a camera and assesses sign accuracy with a quantifiable score. This innovative feature sets our app apart by offering a dynamic, real-time learning tool increasing users progress and proficiency in ASL.

Overall, each element of our project was crafted to foster a deeper, more engaging learning experience, reflecting our commitment to making ASL accessible and enjoyable for everyone.

3.2 Algorithms and Methods

Hand Tracking and Gesture Recognition For accurate hand tracking and gesture recognition, SIGNIFY integrates Google's MediaPipe, a cutting-edge framework that provides real-time, high-fidelity hand tracking. This technology captures the nuanced movements of a user's hands, which is critical for teaching precise hand shapes and movements required in ASL.

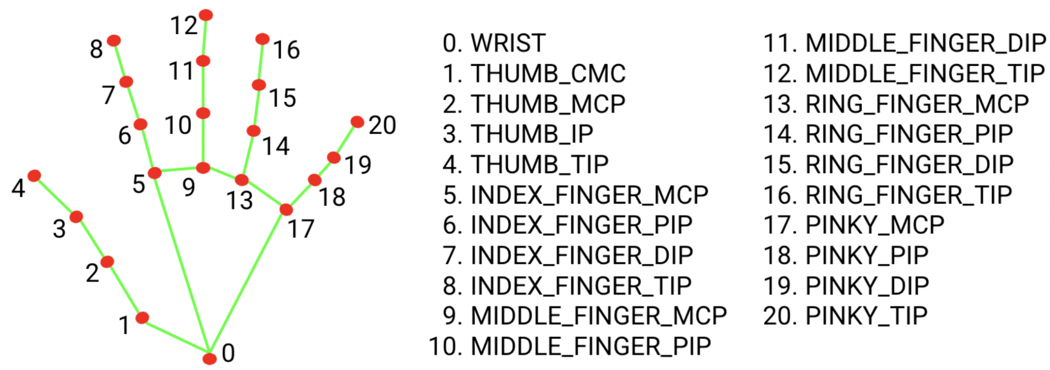

Figure 1: Diagram showing the 21 hand landmarks used in SIGNIFY, from Google's MediaPipe Hand Detection Model for gesture recognition. Each point corresponds to a strategic location on the hand, vital for accurately replicating ASL signs.

Data Collection Data collection in SIGNIFY was rigorously designed to create comprehensive templates for both left and right-hand signs. Utilizing Google's MediaPipe hand detection model, we captured a series of 21 coordinates for each hand gesture involved in ASL. These coordinates represent critical points on the hand that define each sign's unique shape and orientation.

To ensure that our templates could be universally applied to different users, normalization of these coordinates was necessary. This process involved adjusting the coordinates to have a consistent scale and translation across various hand sizes and positions by centering and scaling the landmarks based on their calculated mean and standard deviation. This normalization allows the static template matching algorithm to effectively handle variations in user hand sizes and positioning, making the learning experience more robust and user-friendly.

Using custom-developed plain JavaScript functions, we extracted and normalized the coordinates from the MediaPipe model. These normalized coordinates were then saved as individual templates for both left and right hands. To ensure efficient access and comparison of these templates during application use, they were stored within a structured dictionary in the application. This organization not only facilitates quick retrieval but also optimizes the matching process, significantly enhancing the user experience by reducing load times and improving response efficiency during sign language practice.

Static Template Matching For static gestures, the application calculates the Euclidean distances between the user's hand gesture vectors and pre-defined template vectors. This straightforward comparative method helps users learn and replicate the correct hand positions for various signs.

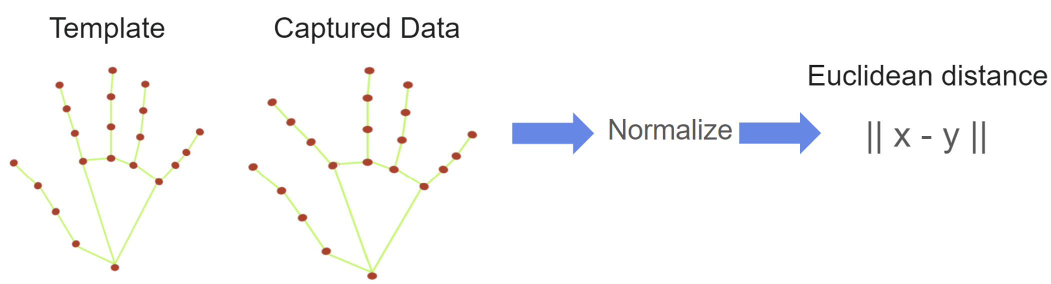

Figure 2: Here we show the general pipeline for detecting static hand gestures. Starting on the left we have a template of the target hand position and the current captured hand position. We then normalized the template and captured data to bring data back to a common scale. Lastly, the calculation of Euclidean distances between user-captured data and static gesture templates is used to detect once the user has made the correct hand pose. If this accuracy is within a threshold, we say the user made the correct sign.

Dynamic Time Warping (DTW) To handle dynamic gestures, particularly for signs that involve movement such as the letters 'J' and 'Z', SIGNIFY employs DTW Fig.3. This algorithm effectively measures the similarity between two temporal sequences, which may vary in speed. By implementing DTW, SIGNIFY allows users to practice and perfect more complex sign sequences in a forgiving and educational manner.

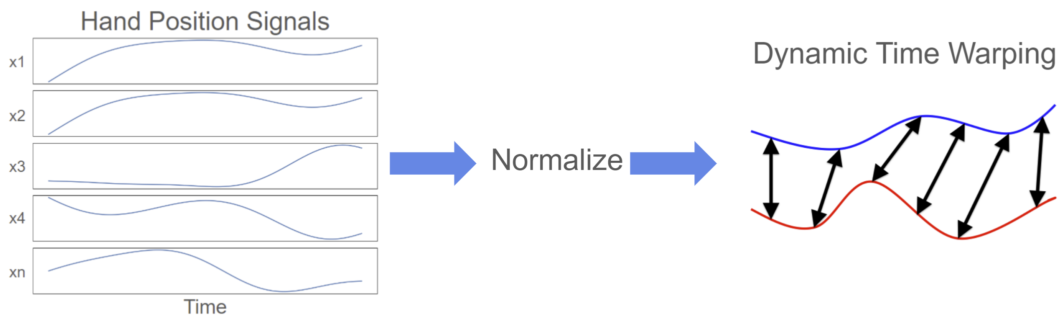

Figure 3: Here we show the general pipeline for detecting dynamic hand gestures. Starting on the left with the real-time collection of hand position signals. The signals are then normalized to bring data back to a common scale. Finally, we measure the signal similarity with he target sequence using Dynamic Time Warping, which quantifies similarity across dynamic ASL gestures irrespective of timing discrepancies.

3.3 Human Factors

We have incorporated human factors into our project by prioritizing contrast sensitivity, ensuring that text on screens is easily readable without causing eye strain. Additionally, significant effort has been dedicated to optimizing the app's performance, with a focus on accurately tracking hand movements and updating these movements on the screen promptly. This immediate feedback is essential for users to quickly recognize and correct their mistakes.

Our responsive model enhances the app's usability by ensuring swift updates. Beyond technical efficiencies, our design process includes a deep dive into the cognitive processes involved in acquiring new skills. The app is crafted to stimulate the brain with diverse training techniques, ranging from repetitive practice in our "Learning Module" to more intense exercises in our "Game Module," where users must recall and perform signs independently, thus creating external cognitive connections.

Furthermore, our application features a one-button start for the webcam, supports interaction with both hands, and provides immediate feedback on correct sign performance or letter guessing. This feedback loop keeps users well-informed of their actions. The app's user interface is intentionally simple, making it accessible and easy to navigate for most age groups with minimal guidance.

Importantly, we prioritize user privacy—no personal data from learners is stored, ensuring that users' information remains confidential. This commitment to privacy and the inclusion of adaptive features enhance human-computer interaction, allowing a broad range of users to effectively engage with the technology, regardless of their skill levels or personal preferences.

4 Results and Improvements

4.1 Learning Interface

The final iteration of the SIGNIFY learning interface has been developed to offer users the ability to engage with three progressively challenging sets of American Sign Language (ASL) words. These categories, ranging from beginner to advanced levels, incorporate a selection of words that gradually introduce more complex signs, culminating in those that require dynamic gestures. For the purpose of learning and practice, users are presented with static images that serve as templates for each sign. They are then tasked with mirroring these templates with their hands in real-time. Through rigorous testing, it was found that while the system performs commendably in recognizing static signs with high accuracy, the recognition of dynamic signs shows variability in consistency.

The platform currently exhibits certain limitations, particularly the fixed set of words, which does not allow for customization by users in adding or removing words. Moreover, the system's real-time feedback is binary in nature; it indicates whether the user is correctly mimicking the sign without providing nuanced guidance on improving the sign execution. An attempt was made to enhance the feedback mechanism by analyzing the accuracy of individual coordinate matching between the user's hand position and the template. However, this approach yielded inconsistent results across different signs, leading to its exclusion from the final product.

The ambition to expand SIGNIFY's capabilities beyond single-hand gestures to include dual-hand coordination, upper body posture, and even facial expressions remains unfulfilled. These elements are critical for a comprehensive gesture recognition system that can more accurately reflect the intricacies of ASL.

As for the user interface, it is suggested that future improvements could involve dynamic visual aids, such as animations or 3D models, to better demonstrate the gestures. This would not only aid in understanding the motions for static signs but also provide a multi-angular view for complex signs, significantly enhancing the learning experience. Such improvements could facilitate a more interactive and supportive environment, ultimately contributing to a more robust and user-friendly ASL learning tool.

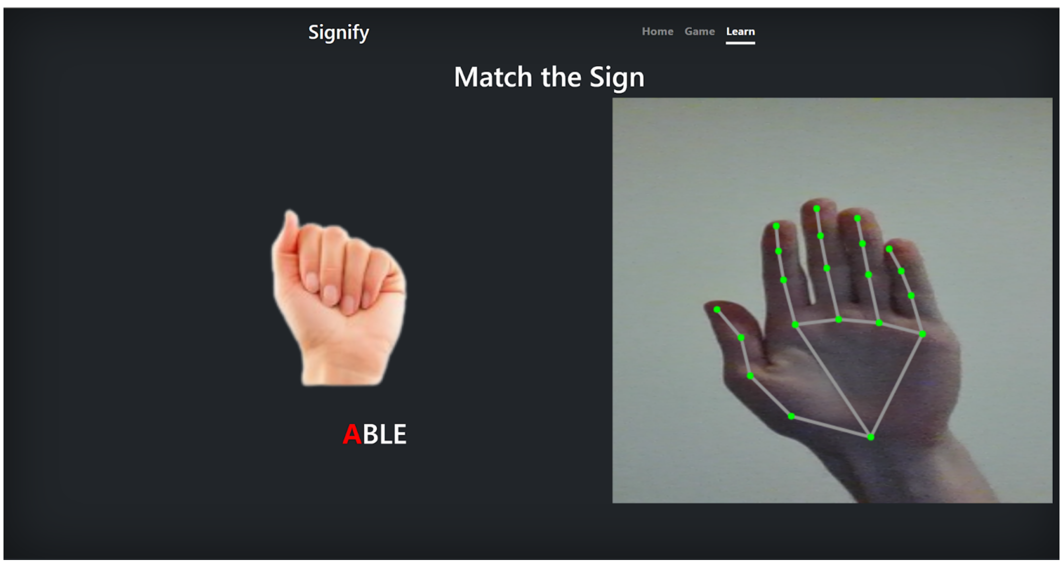

Figure 4: The learn page shows the users camera with the plots pointed over their hand. It also displays the word attempting to be spelled, the current letter and the correct way to perform the sign given the hand you selected at the beginning.

4.2 Game Interface

Initially, we envisioned the game with a hangman-style approach, where users would guess letters and, if incorrect, would lose a life before trying again. However, we encountered challenges in implementing a mechanism to penalize incorrect guesses. The options we considered, such as adding a button to capture the camera input for point determination or setting a limit on the number of guesses per letter, detracted from the fun and fluidity of game play, negatively impacting the user experience.

Moving forward, we aim to enhance the game by developing a more engaging and intuitive penalty system. Our goal is to provide a compelling reason for users to strive for accuracy, thus adding depth and challenge to the game. By refining this aspect, we hope to improve not only the game play experience but also the educational value of the app.

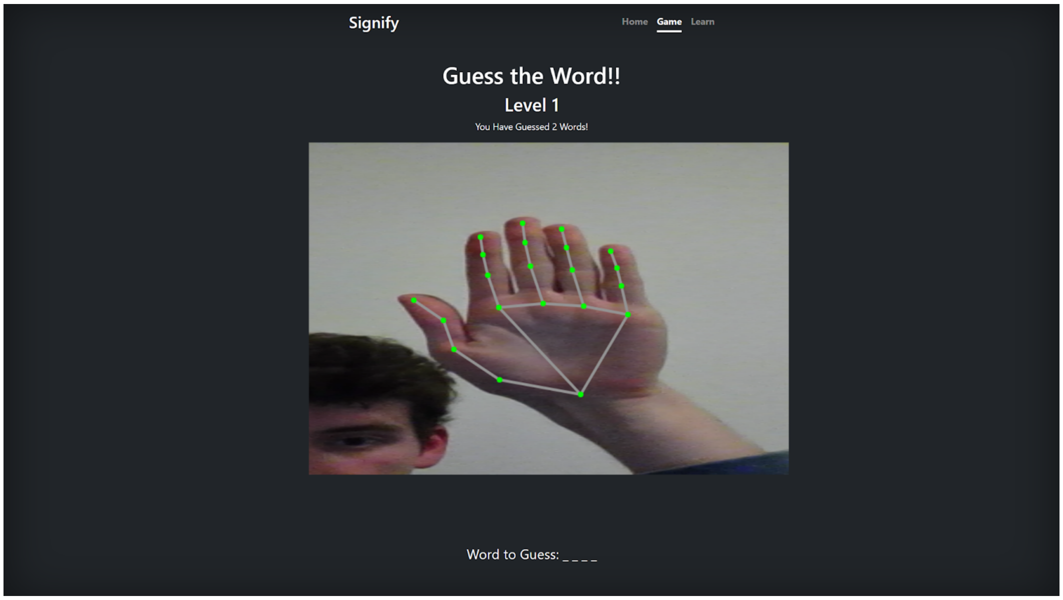

Figure 5: The game page shows the users camera, and the points plotted onto their hand. It also displays the level, how many words they've guessed and the word their attempting to guess, with filled spaces of letters guessed correctly.

5 Conclusion

Our project has been an enlightening journey, teaching us extensively about the intricacies of developing an app that utilizes hand-tracking technology. We encountered numerous challenges, from the technical complexities of implementing the model to understanding the diverse learning curves encountered by users. We realized that transitioning from no knowledge to fluency in sign language is a formidable task.

Acknowledging the varied learning preferences and paces of users, we designed the app with flexibility in mind. By offering multiple learning modes and scalable levels, we cater to individual needs, making the learning process more accessible and effective. Additionally, we've built the app to be adaptable and scalable, allowing for the easy integration of new features, such as additional modes, levels, or signs. This forward-thinking design ensures that our application remains a valuable resource for learning ASL, now and in the future.

References

[1] Duolingo. Homepage, 2024. Accessed: 2024-04-15.

[2] Fingerspelling XYZ. Homepage, 2024. Accessed: 2024-04-15.

[3] Preply. Homepage, 2024. Accessed: 2024-04-15.

6 Appendix

6.1 Additional Materials

The following materials are submitted with this report to provide further insight into the development and functionality of the SIGNIFY project:

- **Source Code:** All source code for the SIGNIFY application is available in our GitHub repository, which includes all modules used for hand tracking, gesture recognition, and UI management. The repository can be accessed here: SIGNIFY GitHub Repository.

- **User Manual:** The user manual for the SIGNIFY application is provided as a README.md file in the root directory of the GitHub repository. This manual includes installation instructions, usage guidelines, and troubleshooting tips to assist users in setting up and using the application. The README.md also outlines the structure of the repository and provides a comprehensive guide to the application's features.