XR Co-signing: an Innovative Tool for Learning American Sign Language

Project team: James Spann, Yamin Zheng & Caitlin Sharp
Faculty advisor: Dr. Zhen Bai

Background

- **Need for ASL Learning Tools**
  - most hearing parents and educators lack formal ASL training (Bouzid et al., 2016; El Ghoul & Othman, 2022)
- **Pedagogical advantages of XR in Sign Language Learning**
  - supports embodied learning, allowing users to interact and practice signs in a spatially immersive way (El Ghoul & Othman, 2022; Aristizábal et al., 2017; Zirzow, 2015)
- **Importance of Non-Manual Signs in ASL**
  - facial expressions, head movements, and eye gaze, are essential to ASL as they add grammatical, emotional, and syntactic information. Those are vital for proficiency in ASL (Quandt et al., 2020; Wolfe et al., 2022)

Non-manual Markers in ASL

("Non-manual Markers in ASL / NMM's", n.d.)

1. **Eyebrows Raised**:
   - **Yes/No Questions**: Raised eyebrows are used when asking yes/no questions.
   - **Conditional Clauses**: Raised eyebrows are used to indicate the conditional part of a sentence.

2. **Eyebrows Furrowed**:
   - **Wh-Questions**: Furrowed eyebrows are used when asking questions that begin with words like who, what, where, wh
   - **Negation and Doubt**: Furrowed eyebrows can indicate negation or doubt.

3. **Head Nods**:
   - **Affirmation**: A nodding head indicates agreement or affirmation.

4. **Head Shakes**:
   - **Negation**: A shaking head indicates negation or disagreement.

5. **Mouth Shapes**:
   - **"oo"**: Indicates something small or diminutive.
   - **"mm"**: Indicates something regular, normal, or moderate.
   - **"cha"**: Indicates something large or big.
   - **"puffed cheeks"**: Can indicate something is very large, bulky, or intense.
   - **"th"**: Indicates carelessness or a lack of precision.
   - **"pursed lips"**: Can indicate intensity or seriousness.
   - **"gritted / barred teeth"**: Can indicate effort under pressure, pain, or intensity.

6. **Eye Gaze**:
   - **Directional Verbs**: The direction of the gaze can indicate the direction of an action or the subject of the conversation.
   - **Focus and Attention**: Eye gaze can show where the signer is directing their attention.

7. **Cheek Puffing**:
   - Indicates intensity, size, or effort.

8. **Nose Wrinkling**:
   - Can indicate distaste or disapproval.

9. **Shoulder Raising**:
   - Indicates uncertainty or questioning.
   - single cheek-to-shoulder indicates "very" or "just" as in "very recent" or "just happened"

10. **Head Tilting**:
    - Used to emphasize a point or to indicate a question.

Research Gaps

- Few tools effectively address non-manual components like facial expressions and body posture, a gap that limits full ASL comprehension for novice learners (Quandt et al., 2020; Wolfe et al., 2022)

- In early ASL learning, once phrases with multiple signs are used, manual and non-manual signs are learned independently (Reilly & Anderson 2002)

- Studies indicate that avatar-based ASL instruction can improve user engagement and retention compared to traditional methods (Bouzid et al., 2016; Beal, 2020; Wolfe et al., 2022), but research integrating avatar-based non-manual signs is rare

Research Question

Can we develop a fun XR Avatar game that helps novice learners grasp non-manual signs in ASL?

Main Idea

The hand movements from one user and facial expressions of another user are separately mapped to the same avatar

Co-signing Rationale

- **Shared embodiment theory** (Merleau-Ponty, 2024)
  - Individuals learn and communicate by engaging with others in a physical space.
    - Co-signing demands a synchronized effort
    - Dual-input experience fosters a sense of unity and shared purpose.

- **Reducing cognitive load through collaboration** (Robbins & Aydede, 2008)
  - The game distributes the cognitive load of learning complex ASL signs.
  - Each player focusing on only one aspect of ASL reduces the overall cognitive demand on each learner.

Co-signing Benefits

Collaborative Learning Theory

- Shared Responsibility in Co-Signing (David & Roger Johnson; Morton Deutsch)
  - Players rely on one another to complete each sign accurately, each player's success is tied to the other's contribution
- Interactive Learning through Real-Time Feedback (Lev Vygotsky; Albert Bandura)
  - Allow users to adjust and improve based on shared outcomes, enhancing mutual understanding and reinforcing learning
- Building Social and Communication Skills (Lev Vygotsky; David & Roger Johnson)
  - Require coordination and communication between players to synchronize their movements
- Mutual Scaffolding for Learning ASL Components (Lev Vygotsky; Jerome Bruner)
  - Observing how their part complements the other's, they gain a holistic understanding of ASL they might not achieve alone.
- Enhanced Engagement and Motivation (Robert Slavin; Albert Bandura)
  - Game-based collaborative environment, interactive nature and shared goals add a level of excitement.

Project Goals

- Create a two-person XR working prototyping that can allow two users to coordinately reproduce ASL signing using manual and non-manual signs
  - Tech
    - Tracking: non-manual sign (Facial Expression, Head Movement), manual sign (two hands)
    - Avatar mapping
  - Research
    - Theories that support the benefits of the collaborative signing activity
- Allow ASL instructors to try out the technology, and co-design learning activities to promote learning non-manual ASL sign

Prior Work: Teach Qatari Sign Language

- Learners interacted with a human-sized avatar to observe signs from various angles using VR controllers, and mimicked hand movements guided by transparent virtual hands, with the system tracking and calculating accuracy.
- A head-mounted camera for recording facial expressions.
- HTC Cosmos headset and Metaverse Oculus headset
- The challenge with facial expressions arose because the VR prototype lacked advanced facial animation features to fully replicate the nuances of human expressions. As a result, 67% of participants reported that traditional video formats provided better clarity for facial expressions

(El Ghoul & Othman, 2022)

Prior Work: Interprete SignWriting Using Avatar

- The system translates SignWriting Markup Language (SWML), an XML-based format, into 3D animations of signed gestures using avatar technology.
- Converts gesture descriptions into Sign Modeling Language (SML), an XML-based format tailored for avatar animation.
- SML scripts are interpreted to manipulate avatar joints in real time.The avatar's rendering engine computes joint rotations in real time, updating the avatar's posture and movements frame-by-frame.
- The system defined specific facial articulators based on SignWriting notation, but there are limitations in capturing and rendering these features accurately (e.g., subtle facial expressions) were noted as areas for improvement.

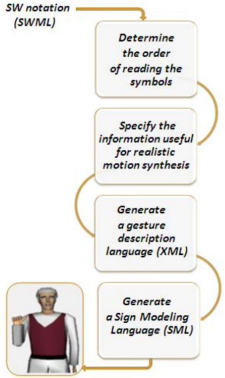
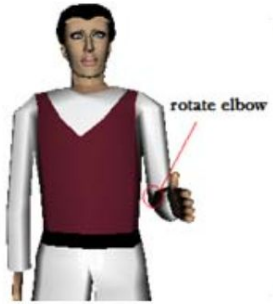


Figure 1. Proposed system process.



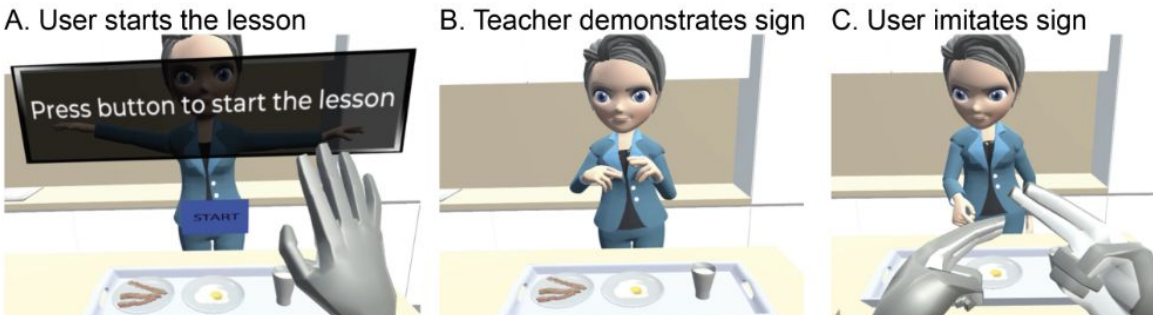

```
<Sentence>
   <word>
      <mouvement>
         <duration>0.5</duration>
         <join name="l_elbow">
            <rotation type="euler">
               <heading>80</heading>
               <attitude>50</attitude>
               <bank>50</bank>
            </rotation>
         </join>
      </mouvement>
   </word>
</Sentence>
```

Figure 3. Rotation of elbow joint int SML

(Bouzid & Jemni, 2013)

Prior Work: Signing Avatars and Immersive Learning (SAIL)

- A VR program for learning individual manual signs in which a teacher avatar demonstrated a sign and the learner saw their hands as they imitated it.
- Manual signs tracked using a LEAP motion sensor mounted to an Oculus Rift S headset.



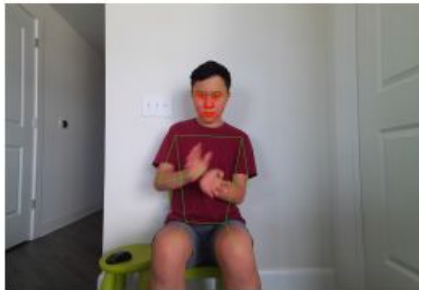
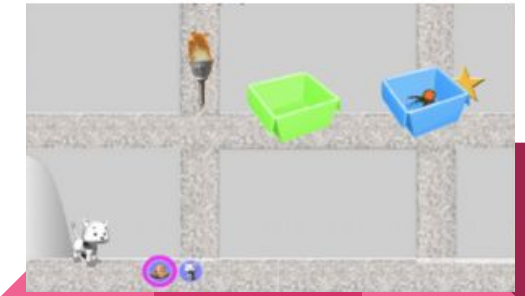
(Quandt, 2020)

Prior Work: CopyCat

- A user uses ASL to describe a scene to a mascot in VR. The system tracks their motions and if the description is correct, the user may advance.
- Previous models used custom gloves with a built in accelerometer
- More recent efforts have tracked movements with AlphaPose, Azure Kinect, and MediaPipe and showed improved accuracy using cameras on both the front and side. Hidden Markov Models increased accuracy of sign recognition

(Bansal et al., 2021)

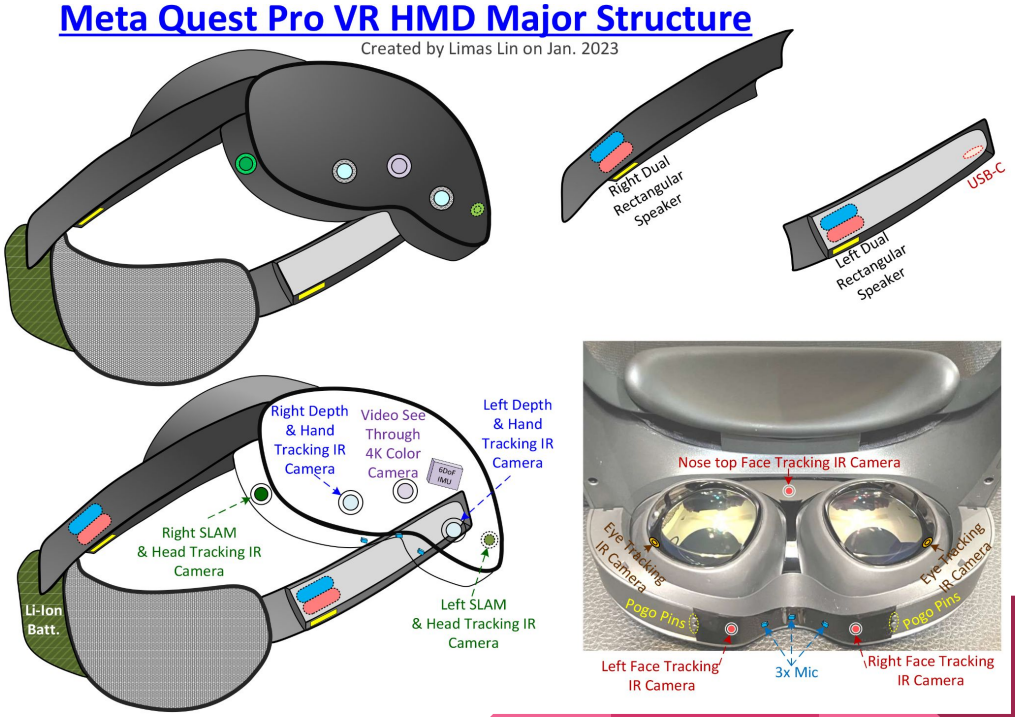


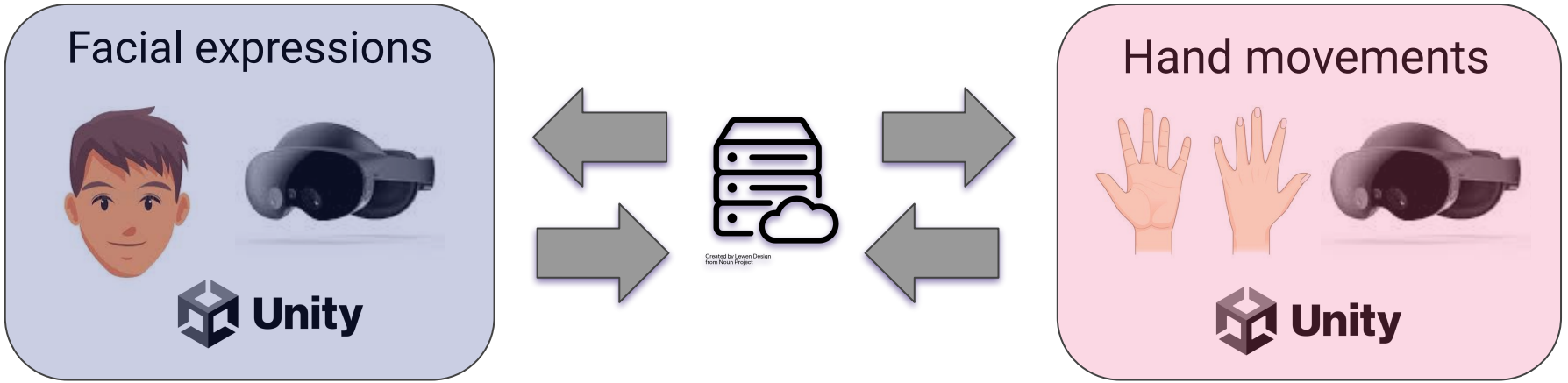
Meta Quest Pro

- Goal:
  - Capture Manual (hand motions for ASL) and Non-manual (head tilt, eyebrow movement, ...) expressions
  - Synchronize expressions across multiple devices
  - Show a virtual avatar of combined expressions
  - Detect specific ASL and gestures

- Meta Quest Pro
  - IR Camera for head tracking
  - IR Camera for eye tracking, nose tracking, & face tracking
  - Accelerometer for head movement
  - SLAM Cameras for hand tracking



Original Solution - Face and Hand Tracking in Meta Quest Pro

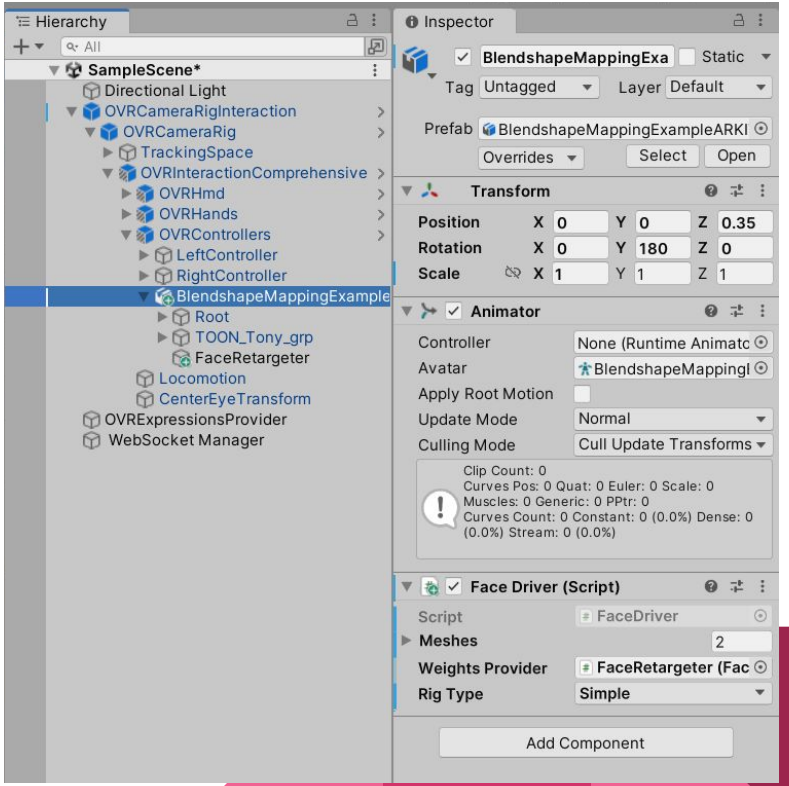


- XR Interaction for hand tracking
- Meta SDKs for face tracking
- Web socket server to coordinate users

Original Facial Expression Tracking Solution

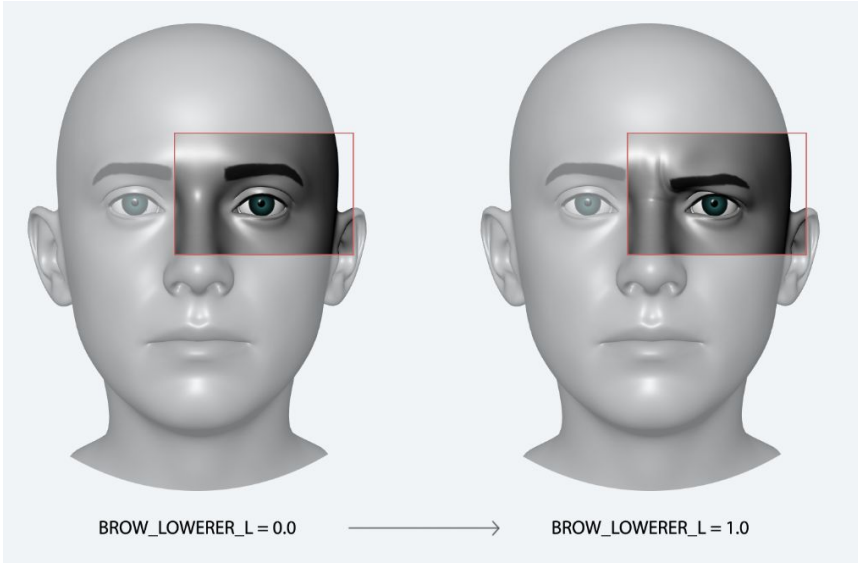
- Meta All-in-One SDK (includes Meta Core SDK & Meta Interaction SDK)
- Oculus XR Plug-in
- Meta Movement SDK (with Face Tracking API)
- Networking between multiple devices

(Meta Horizon, 2024)



Face Tracking API

- “Blendshapes” represent possible facial expressions (70 available)
- Inward facing cameras on the headset share data regarding which expressions are made and the extent to which they are made (calculated as a “weight”)
- A FaceRetargeter component determines the expression to be made based on the blendshape triggered and its weight and the FaceDriver component mirrors the avatar’s expression accordingly



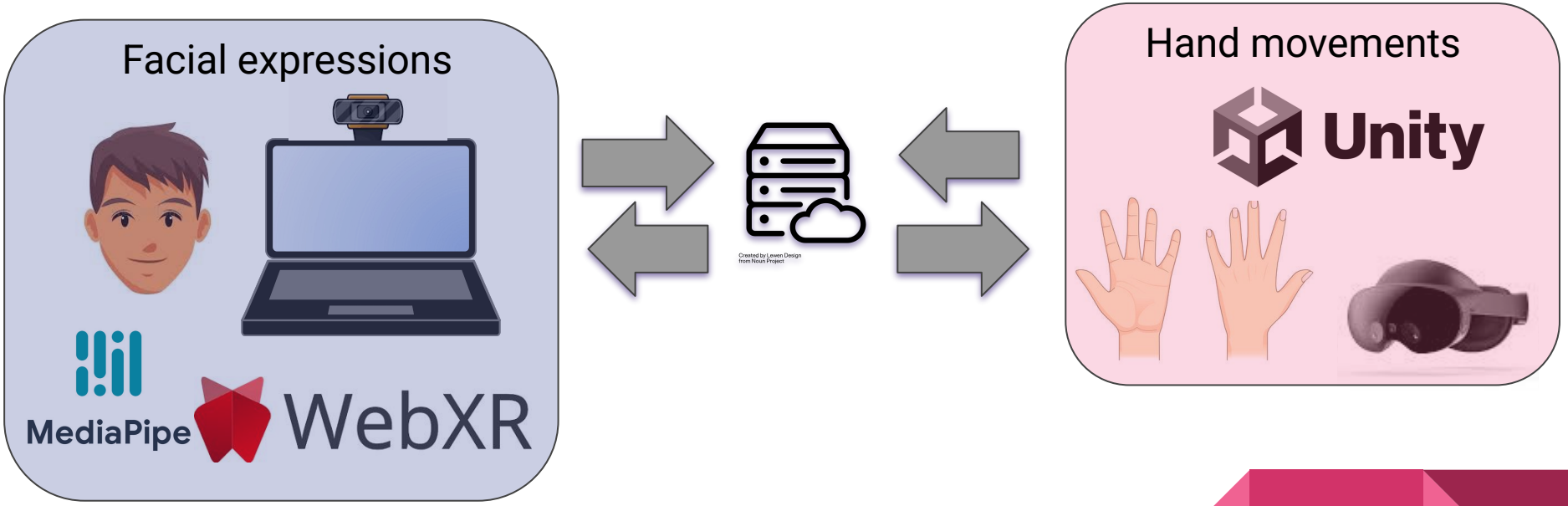
(Meta Horizon, 2024)

Original Solution Pitfalls

- Meta Quest SDKs were difficult to install on machines in the Studio X computer Lab

- The University of Rochester has a VERY strong firewall that made it difficult to network between multiple devices
  - We can send hand tracking data between devices across an AWS server outside of the University

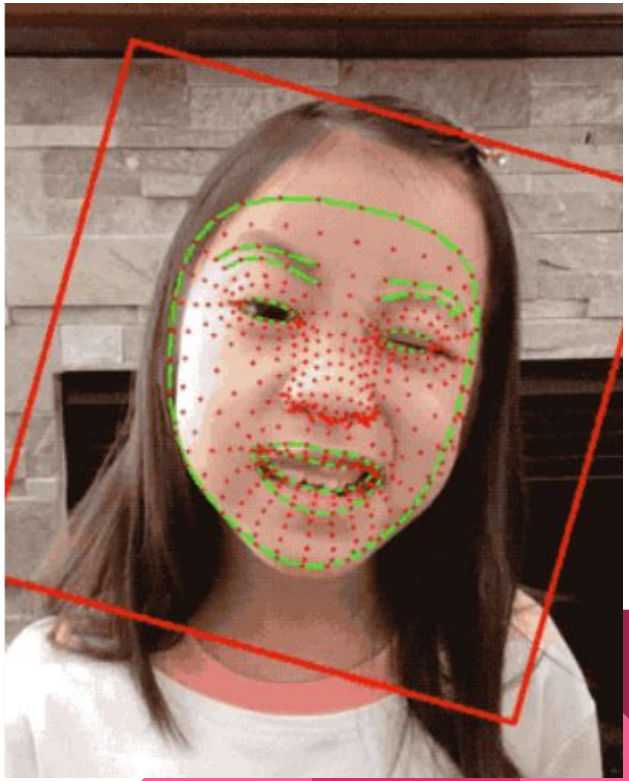
Solution 2 - WebXR for Face Tracking and Meta Quest Pro for Hand Tracking



Second Facial Expression Tracking Solution & Pitfalls

- Mediapipe works but networking still a challenge
- WebXR conversions from Unity
- Facial tracking with Mediapipe FaceMesh

(MediaPipe, n.d.)



Demonstration - Hand tracking (XR Interaction toolkit)

Demonstration - Facial Tracking (Meta Quest SDK)

Future Directions

- Fine-tune connection between devices across the server to integrate non-manual and manual signs in a single avatar
- Record signs to demonstrate in the game
- Use machine learning algorithms (such as hidden markov models) to train the program to detect correct combinations of manual and non-manual signs
- Test and finetune the game based on feedback from deaf individuals and new ASL Learners

References

Aristizábal, L. F., Cano, S., Collazos, C. A., Solano, A., & Slegers, K. (2017). Collaborative learning as educational strategy for deaf children: A systematic literature review. *Proceedings of the XVIII International Conference on Human Computer Interaction*, 1–8. https://doi.org/10.1145/3123818.3123830

Bansal, D., Ravi, P., So, M., Agrawal, P., Chadha, I., Murugappan, G., & Duke, C. (2021, May). Copycat: Using sign language recognition to help deaf children acquire language skills. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-10).

Beal, J. S. (2021). American Sign Language (ASL) development: Deaf students' ASL skills across age and time and implications for ASL instruction. *Deafness & Education International*, *23*(4), 335–352. https://doi.org/10.1080/14643154.2020.1737764

Bouzid, Y., Khenissi, M. A., & Jemni, M. (2016). The Effect of Avatar Technology on Sign Writing Vocabularies Acquisition for Deaf Learners. *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 441–445. https://doi.org/10.1109/ICALT.2016.127

El Ghoul, O., & Othman, A. (2022). Virtual reality for educating Sign Language using signing avatar: The future of creative learning for deaf students. In *2022 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1269-1274). IEEE. doi: 10.1109/EDUCON52537.2022.9766692.

Ghoul, O. E., & Othman, A. (2022). Virtual reality for educating Sign Language using signing avatar: The future of creative learning for deaf students. *2022 IEEE Global Engineering Education Conference (EDUCON)*, 1269–1274. https://doi.org/10.1109/EDUCON52537.2022.9766692

Kamnardsiri, T., Hongsit, L., Khuwuthyakorn, P., & Wongta, N. (2017). *The Effectiveness of the Game-Based Learning System for the Improvement of American Sign Language using Kinect*. *15*(4).

MediaPipe. (n.d.). *MediaPipe Face Mesh.* MediaPipe. https://mediapipe.readthedocs.io/en/latest/solutions/face_mesh.html

Meta Horizon (2024). *Face Tracking for Movement SDK for Unity.* Meta Developers. https://developers.meta.com/horizon/documentation/unity/move-face-tracking/

Non-manual markers in ASL / NMM's, (n.d). ASL American Sign Language. Retrieved from December 10, 2024, from https://www.lifeprint.com/asl101/pages-layout/nonmanualmarkers.htm

Quandt, L. (2020). Teaching ASL Signs using Signing Avatars and Immersive Learning in Virtual Reality. *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, 1–4. https://doi.org/10.1145/3373625.3418042

Reilly, J., & Anderson, D. (2002). The acquisition of non-manual morphology in ASL. *Directions in sign language acquisition*, 159-182.

Y. Bouzid, & M. Jemni (2013). An Avatar Based Approach for Automatically Interpreting a Sign Language Notation,P*roceedings of the 13th IEEE International Conference on Advanced Learning Technologies*, pp. 92-94.

Wolfe, R., McDonald, J. C., Hanke, T., Ebling, S., Van Landuyt, D., Picron, F., Krausneker, V., Efthimiou, E., Fotinea, E., & Braffort, A. (2022). Sign Language Avatars: A Question of Representation. *Information*, *13*(4), 206. https://doi.org/10.3390/info13040206

Zirzow, N. K. (2015). Signing Avatars: Using Virtual Reality to Support Students with Hearing Loss. *Rural Special Education Quarterly*, *34*(3), 33–36. https://doi.org/10.1177/875687051503400307

THANKS