

An Improved Avatar for Automatic Mouth Gesture Recognition

Ronan Johnson¹, Maren Brumm², Rosalee Wolfe¹

¹DePaul University School of Computing Chicago, USA

²University of Hamburg Institute of German Sign Language and Communication of the Deaf Hamburg, Germany

Motivation

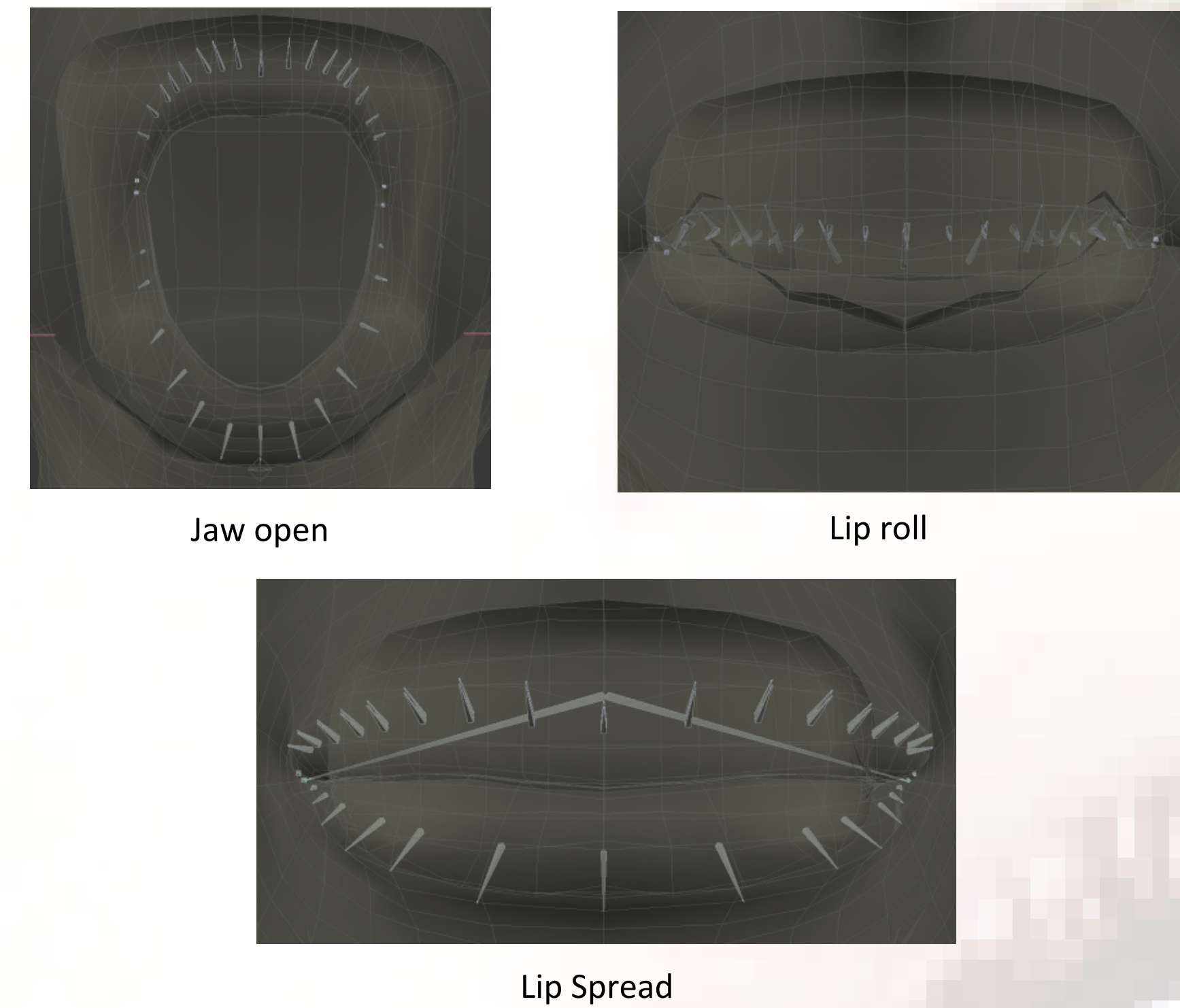
- Sign language avatars are used to synthesize signed languages as naturally and accurately as possible.
- Obtaining large, annotated corpora is essential to research efforts.
- Mouth gestures provide important linguistic information, but must be manually annotated.
- Neural networks can be trained to assist in annotation.
 - Require large amounts of training data to be effective.
 - Sensitive to unbalanced classes.
- Artificial training data generated with an avatar may improve predictive accuracy.
- This can alter or destroy the meaning of a signed utterance.

Avatar Video Requirements

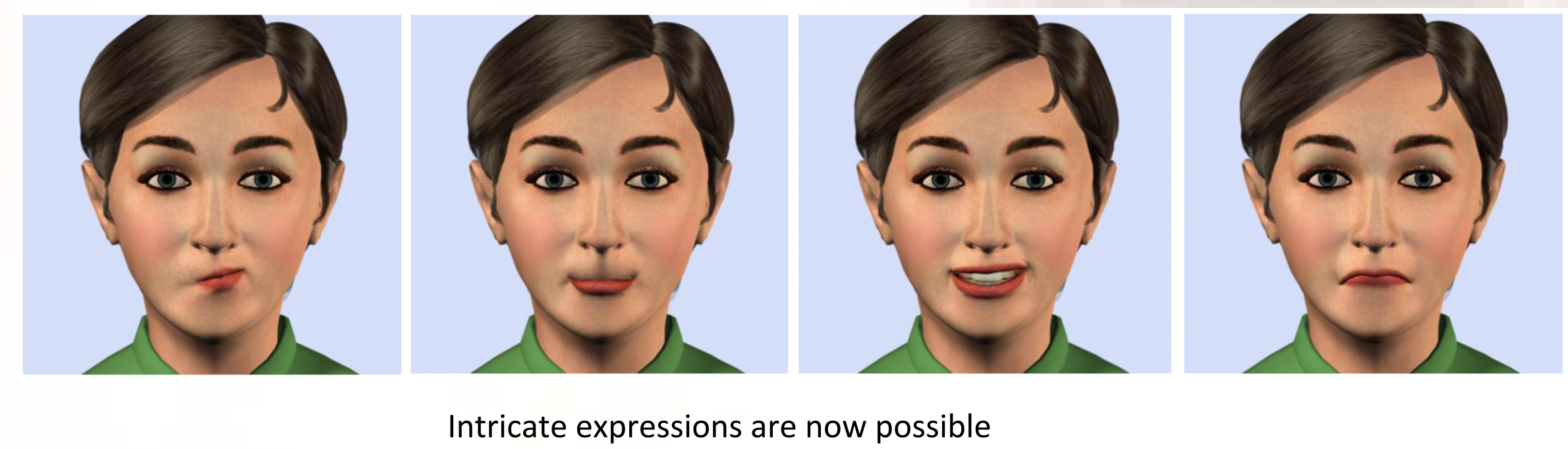
- Training data is made up of short videos of various mouth gestures in context.
- Artificial training videos must mimic real recordings as closely as possible.
- Considerations:
 - Beginning and ending head orientation
 - Range of head motion
 - Intensity and duration of the mouth gesture



Extended Deformation Results

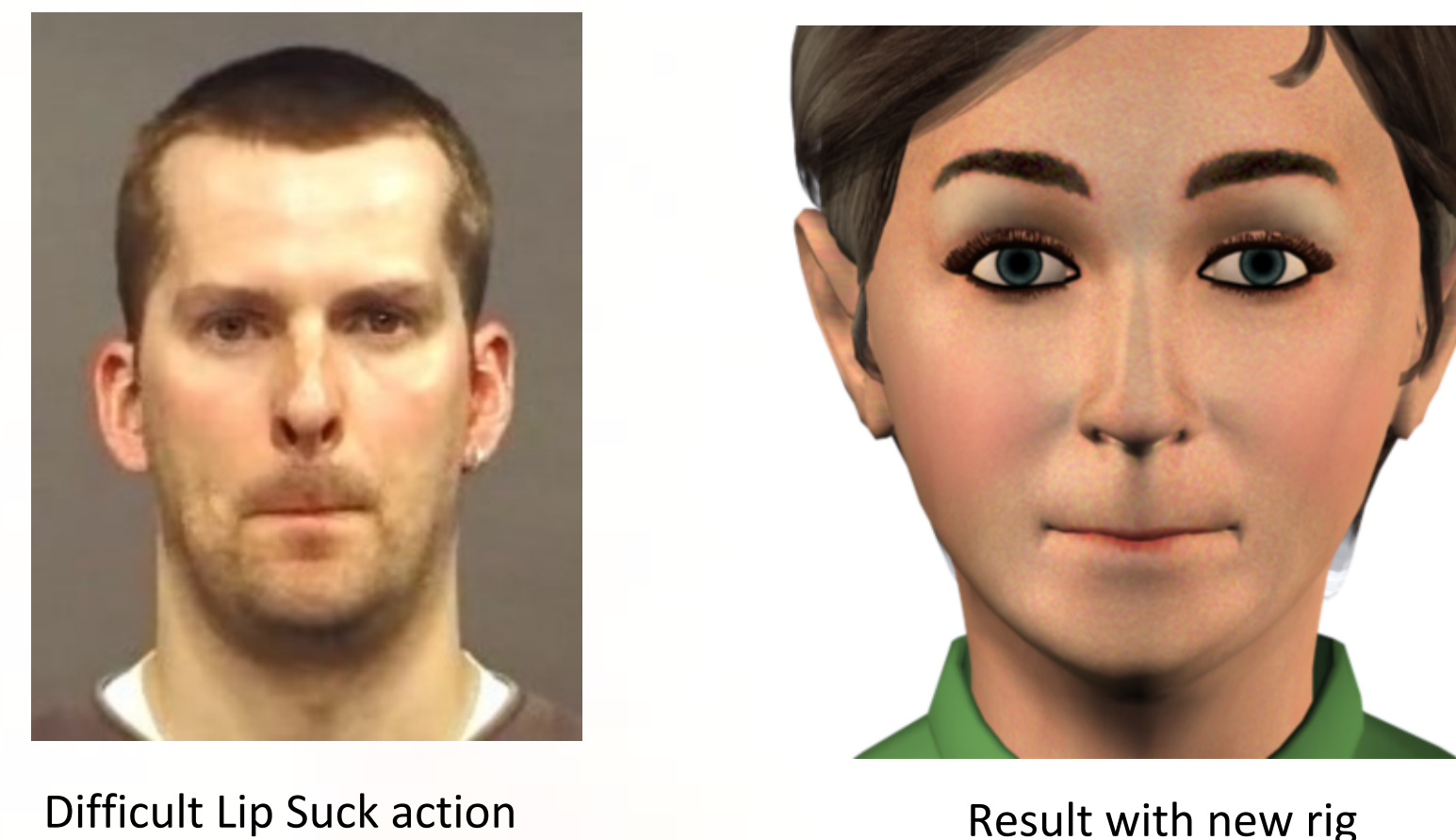
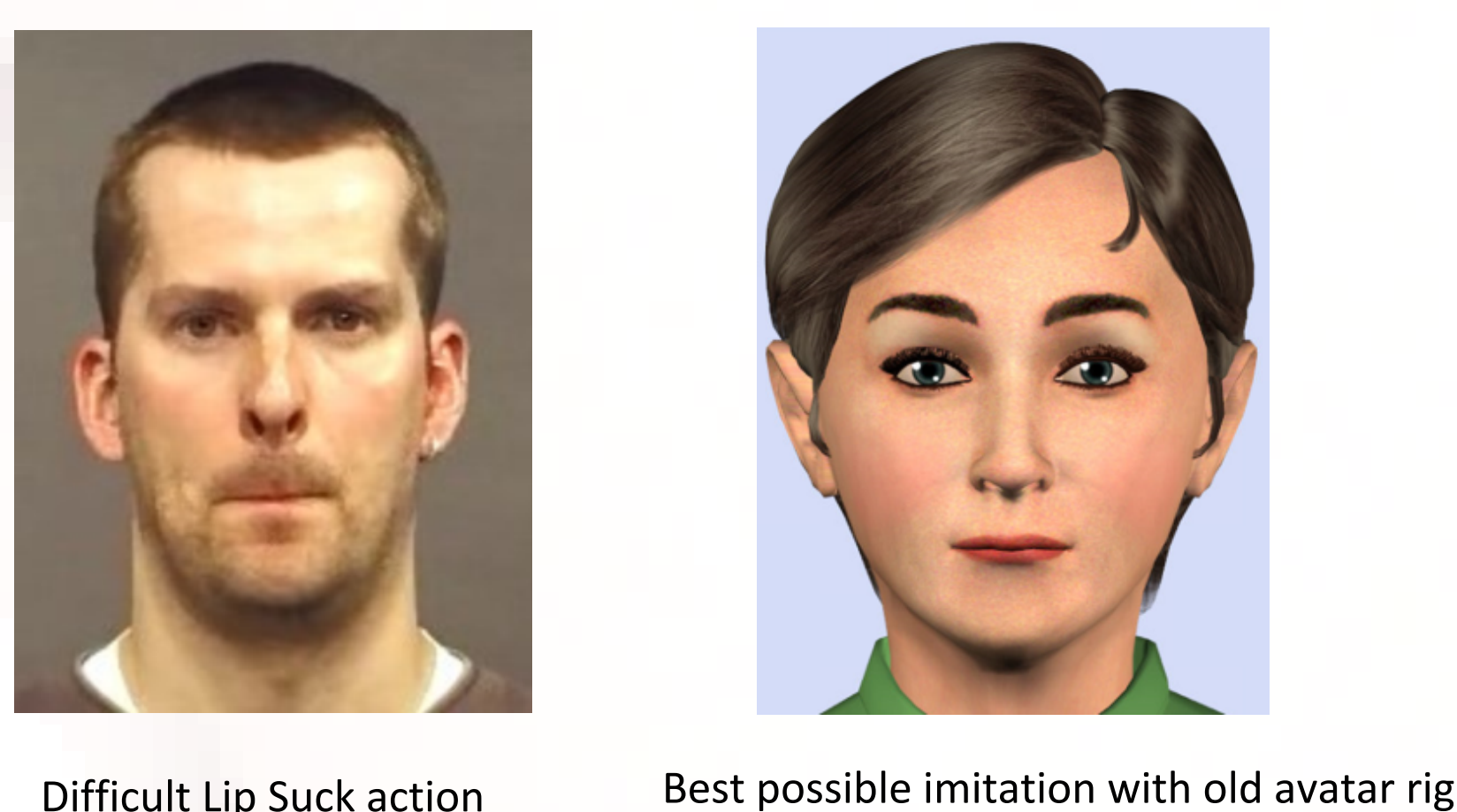


New Rig Range of Motion



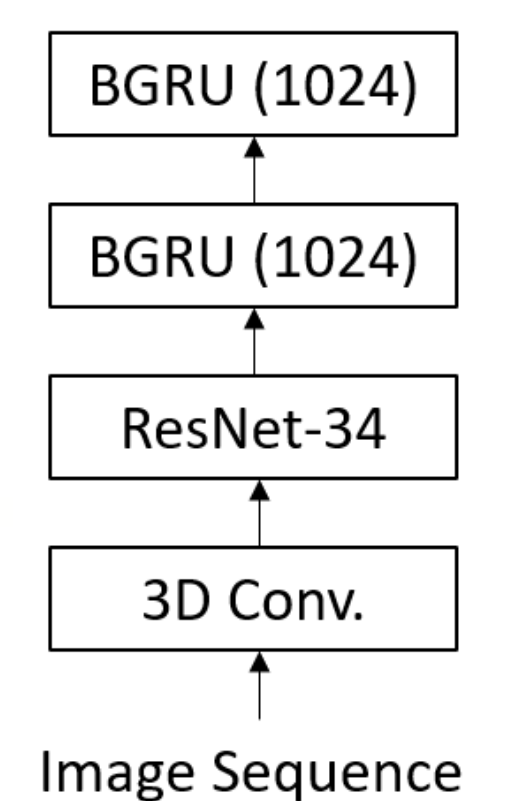
Motivation for Avatar Improvement

- University of Hamburg continues one of the most major corpora annotation efforts.
 - The corpora consists entirely of German Sign Language (DGS).
- DePaul's avatar, Paula, was designed for American Sign Language (ASL).
- Range of expression in DGS is more varied and exaggerated than in ASL.
- Paula was incapable of generating the necessary range of motion



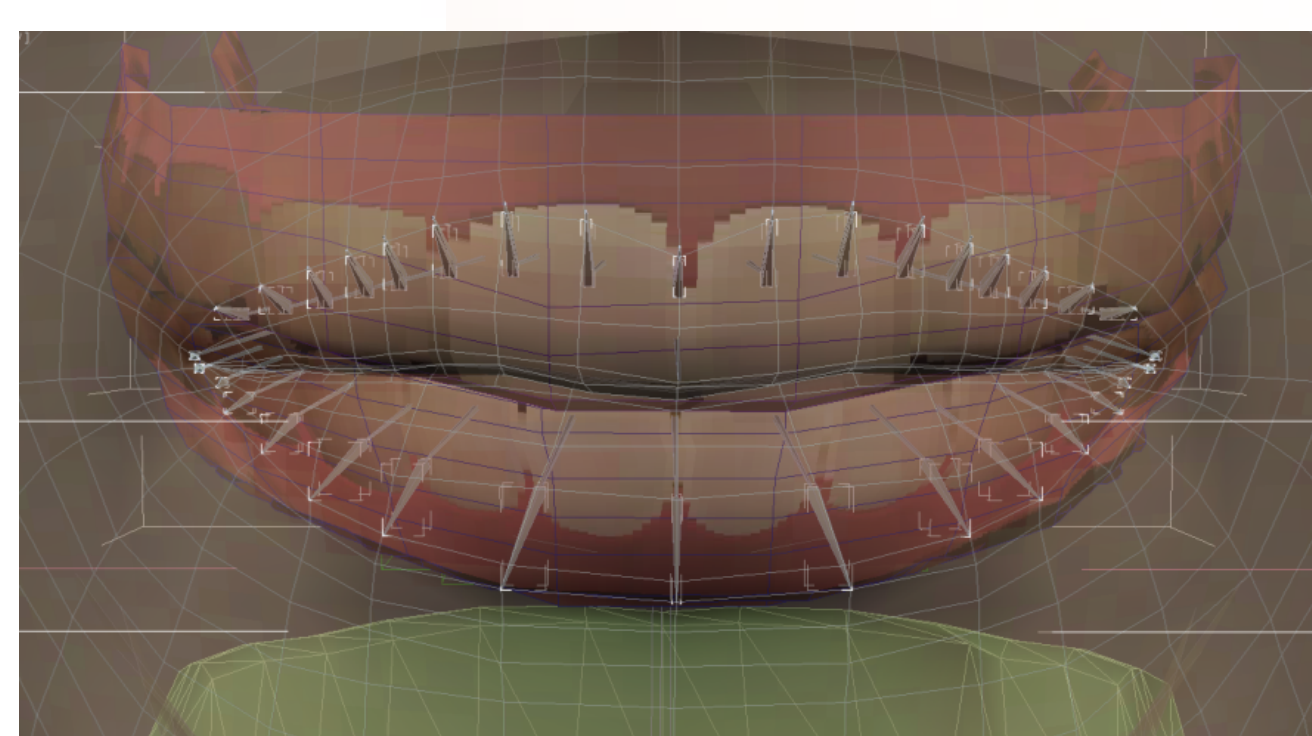
First Classification Results and Future Work

- For classification, we use a neural network pretrained on a lipreading dataset (LRW).
- Initially tested on five different mouth gesture classes.
 - Accuracy with natural data only: **83.59%**
 - Accuracy with included avatar data: **84.87%**
- Initial results show minor improvements.
- We expect more definitive results with more classes in future work.



Avatar Facial Rig Extensions

- Increased the number of mouth control points from 10 to 44
- Each "spike" represents a control point with endpoint along the surface of the lip
- We achieve 5 levels of control per spike
 - Roll, spread, jaw open, lip up/down, free control



References

- Please see the QR link.

