How should we take input?

Doing this research we wanted to decide between using a camera or some kind of motion capture system to take in a user's movement to decipher what they are signing. Before the research we expect the camera method will have more support and may be easier to implement, but the motion capture method may give us more accurate data while being a bit harder to implement.

Camera
Pros:
- Easier setup
- More resources detailing using a camera as input to a neural network
- Ability to consider facial expressions
- Can capture movements that are not motion captured.
- Less expensive

Cons:
- Must be facing the camera to capture data
- Could get bad readings if the user is at an angle to the camera

Motion Capture
Pros:
- No matter where the user is standing, the input will be accurate
- The input would not have additional, unnecessary information, like the background

Cons:
- Setup will be harder, putting up the sensors
- Have not been able to find many resources detailing using motion capture for ML
- We would not be able to capture facial expressions, which we recently were informed is an important part of sign language.
- We would need to use full motion capture sleeves and gloves to capture the full range of motion

It seems to us that using a camera would be the ideal solution for taking input during this project. We've found multiple examples of people successfully using a camera to identify individual signs, so we think with that information, as well as the pros and cons, using a camera makes the most sense.

We found a camera, the OAK-D, which is high resolution and built to work well with openCV, a library used for image recognition. This will help us get the best possible results.