Applying Hidden Markov Models (HMMs) to a Sign Language Translation System for the Deaf in Rwanda

My capstone project aims to build an assistive mobile application that will enable individuals with hearing (*deaf*) and speech disabilities to communicate more easily with their surrounding communities in different aspects of life. The application will recognize visual sign gestures performed by users (people with deaf disabilities), and automatically translate them into readable text and synthesized speech using AI. A Hidden Markov Model (HMM) can also be applied to model the sequential nature of sign gestures and improve gesture recognition.

1. Describe the observations

In this system, the observations are the measurable features extracted from sign gestures captured through the user device's camera. These may include:

- Hand coordinates (x, y position in each frame)
- Hand orientation and angle
- Motion direction and velocity
- Hand shape or keypoint vector (from MediaPipe or similar models)
- Temporal sequence (frame-by-frame movement)

Each gesture is represented as a time-series of these observations, which feed into the HMM.

2. Type of HMM problem

This is an unsupervised learning problem. The hidden states which represent the actual intended meaning of a sign, ex., "hello", "thank you", "water", and others; are not directly labeled or known in advance. The task is to:

- Learn these hidden patterns from sequences of observed visual features
- Decode the most probable meaning (state) given a sequence of signs

So, this is both a learning and decoding problem.

3. Training Algorithms

a. What values are known at the start?
   - The structure of the observed features (video-based gesture data)
   - Number of observable sequences (recorded gesture examples)
   - Possibly a rough estimate of the number of intended sign meanings

b. What values are unknown and need to be learned?
   - Transition probabilities between signs (e.g., how often one gesture follows another)
   - Emission probabilities (likelihood of a specific hand motion pattern given a sign)
   - Initial state probabilities (likelihood of a certain sign starting a sentence)

These unknowns can be estimated using the Baum-Welch algorithm, which iteratively improves estimates through Expectation-Maximization (EM).

4. Parameter Updates

The HMM training algorithm will update the following parameters:

- **A (Transition probabilities):** Probability of moving from one sign (state) to another.
- **B (Emission probabilities):** Probability of a specific visual feature given a sign.
- **π (Initial state probabilities):** Probability of a certain sign starting a sequence.

These updates allow the model to better recognize gesture patterns and translate them into meaningful, readable text and audio in real time.