

Perceptual Validation of 3D Pose, Guided Sign Language Synthesis

Ezekiel Maina, Lilian Wanzare, James Obuhuma
Maseno University

Background

Introduction (educational background)

Sign language corpora face a conflict between open-access requirements and the biometric risks of visual data. While raw video exposes facial morphology, standard anonymization obscures Non-Manual Markers (NMMs), severely degrading comprehension for native signers.

Problem statement

Privacy-Fidelity Balance: A 3D approach is essential to mask biometric identity while preserving critical linguistic meaning.

Integrated Validation: Effectiveness requires both technical metrics and human evaluation to ensure structural precision translates into subjective naturalness.

Current Work

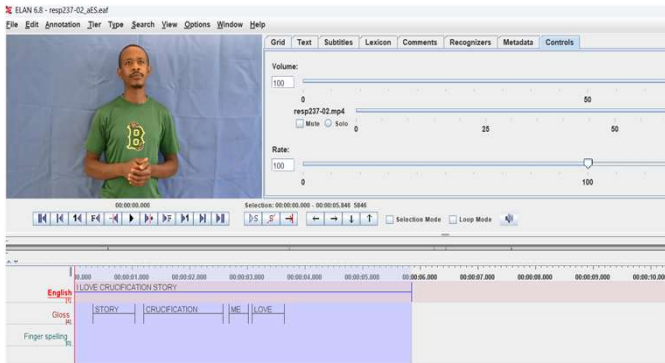


Figure 1: Video Segmentation using ELAN tool (Maina et al., 2025)

Methodology

We utilized the AI4KSL Corpus, consisting of 11,000 raw Kenya Sign Language (KSL) videos.

- Annotation using the ELAN tool; 30370 segments
- Segment Integrity Rate (SIR): 99.97%. $[SIR = \frac{1}{n} \sum_{i=1}^n S(V_i) * 100]$
- Validation Constraints: Non-zero file size, minimum 2 frames, H.264 codec, and 1280 * 720 resolution.
- Hierarchical Topology: Mediapipe (facial, hands and upper body)
- Kinematic Modeling: Employs signer-agnostic normalization relative to the torso centroid to remove stature biases.
- Avatar Synthesis: Deterministic retargeting to Ready Player Me skeletal hierarchies with temporal smoothing.
- NMM Activation: Translates 468 facial vertices into ARKit-compatible blendshapes to render subtle grammar like brow furrowing
- Technical and Human evaluation

Results

High-precision tracking achieved 92.7% PCP and 91.5% PCP, ensuring accurate handshape skeletal fidelity. Robustness against self-occlusion was confirmed by 93.4% PDJ and a 0.88 OKS, indicating superior spatial overlap. Model stability across complex signing depths and rotations was validated by an AUC of 0.91.



Figure 2: Architectural pipeline

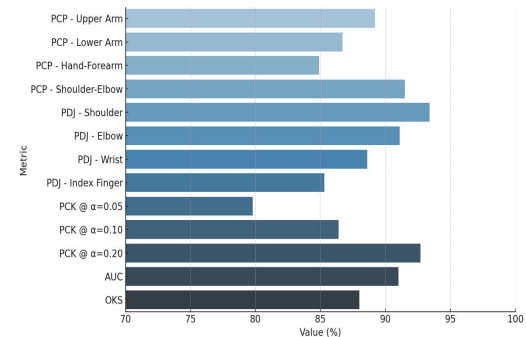
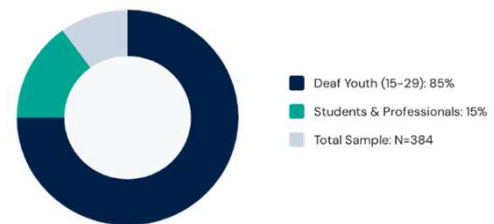


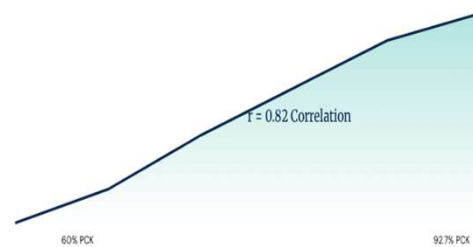
Figure 3: Technical Results

Perceptual naturalness results



A multi-stage design integrated stratified, purposive, and simple random sampling to ensure representativeness. Participants were stratified across four groups: primary, secondary and tertiary students, and Deaf professionals.

Dimension	Mean (M)	Std. Dev (SD)	Median
Gestural Authenticity	4.05	0.94	4.0
Environment	4.09	1.00	4.0
Body Language Clarity	4.15	0.94	4.0
Smoothness & Lifelike Motion	4.10	0.89	4.0
Facial Realism (NMMs)	3.95	0.87	4.0
Composite Score	4.07	0.65	High Acceptance (4.2)



There is a strong positive correlation ($r = 0.82$)

Conclusion & Recommendations

Privacy vs. Fidelity: The real-time pipeline (150ms latency) successfully masks biometric identity while achieving >90% accuracy.

Community Trust: High-fidelity avatars function as accepted "linguistic actors" rather than alienating technical artifacts.

Future Work: We aim to conduct adversarial biometric testing to ensure robustness against AI-driven re-identification attacks.