Supporting Sign Language Video Comprehension through In-Situ Automatic Segmentation and Sign Lookup Tool

TULANE

Chaelin Kim (ckim13@tulane.edu) and Saad Hassan (saadhassan@tulane.edu)

NOLA A11Y

ASSETS 2026

Background

American Sign Language (ASL) is the third most studied language in the United States. However, deaf and Hard of Hearing (DHH) learners and ASL students face limitations in accessing effective video-based learning tools.

Existing ASL learning platforms primarily focus on isolated sign lookup rather than sentence-level comprehension. When learners encounter long signing videos, they must manually search for unfamiliar signs, which is time-consuming and cognitively demanding.

We propose AI-based platforms designed to automatically segment continuous ASL sentences to support efficient sentence-level learning.

System Overview

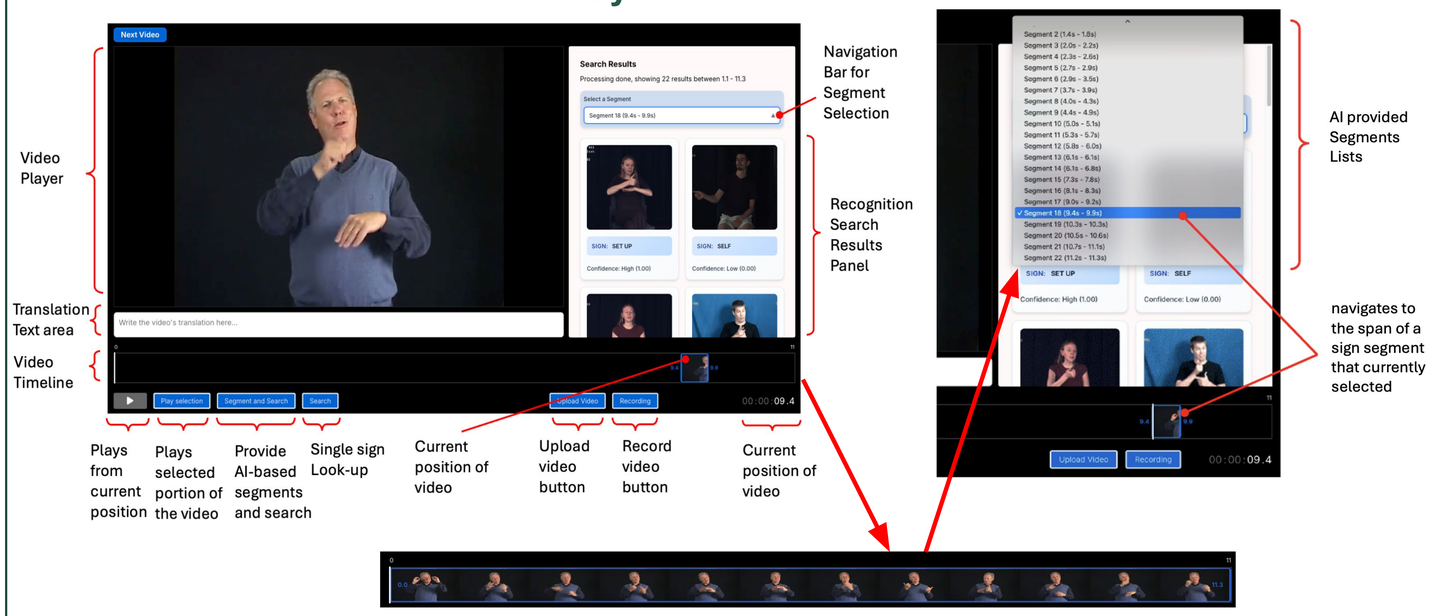


Figure 1. Users upload a signing video and activate **"Segment & Search".** The system automatically segments continuous signing into individual sign units and displays candidate recognition results for each segment in the results panel.

System Components

Segmentation

For automatic segmentation, we build upon [1] model. The model adopts a BIO-tagging scheme to identify the beginning and end boundaries of each sign in continuous signing sequences.

Recognition

For sign recognition, we employed upon [2] architecture. Using mediapipe, the system extracts pose keypoints from video frames and constructs pose sequences, which are then used as input to the recognition model.

code available here

Poster PDF here

System Performance

Table 1. Segmentation performance across datasets (mean ± std).

Dataset	Over-segmentation %	IoU
Moryossef et al. [1]	1.03	0.69
ASLLRP [3]	1.09 ± 0.71	0.37 ± 0.23
Stimuli Video	1.21 ± 0.15	0.50 ± 0.19

Table 2. Recognition performance across datasets and segmentation conditions (mean ± std).

Dataset	Data Setting	Segmentation	Top-1	Top-2	Top-20
ASLLRP [3]	Broader Corpus	AI-Segmented	0.33 ± 0.30	0.39 ± 0.34	0.51 ± 0.42
		Human-Annotated	**0.46 ± 0.41**	**0.50 ± 0.44**	**0.56 ± 0.46**
	Stimuli Videos	AI-Segmented	0.61 ± 0.15	0.68 ± 0.14	0.91 ± 0.07
		Human-Annotated	**0.90 ± 0.06**	**0.95 ± 0.06**	**0.97 ± 0.10**

User Study Method

Participants: The study involved 13 ASL learners with intermediate fluency (1 male, 12 female), aged 18–34. Participants had prior experience learning ASL through coursework or self-study.

Procedure: Participants used two prototypes (Segment & Search and Search) to translate four ASL videos. The order of prototypes was counterbalanced, with each used twice. A questionnaire was administered after each task.

User Study: Translation quality & Search Experience & System Usability

Figure 2. System usability and user experience during translation tasks.

When using Segmentation & Search, participants produced more accurate translations (66.8% vs. 58.9%) and fewer omissions (33.2% vs. 41.1%), with higher holistic scores (7.04 vs. 5.49) compare to Search-Only.

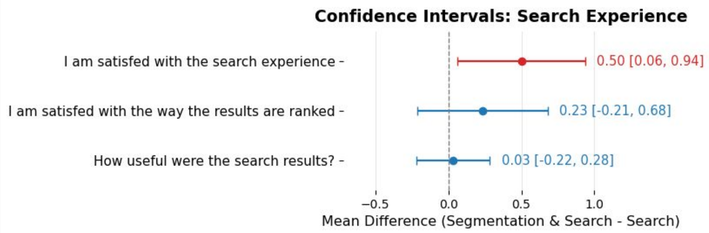
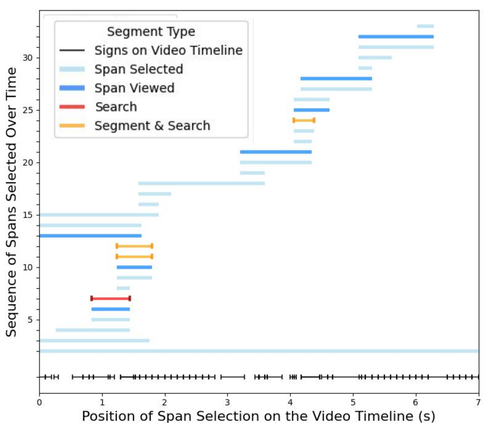


Figure 3. Interaction timeline during the translation task.



References

[1] Moryossef, Amit, et al. "Linguistically motivated sign language segmentation." *Findings of the Association for Computational Linguistics: EMNLP 2023*. 2023.
[2] Desai, Aashaka, et al. "ASL citizen: a community-sourced dataset for advancing isolated sign language recognition." *Advances in Neural Information Processing Systems* 36 (2023): 76893-76907.
[3] Neidle, Carol, Augustine Opoku, and Dimitris Metaxas. "Asl video corpora & sign bank: Resources available through the american sign language linguistic research project (asllrp)." *arXiv preprint arXiv:2201.07899* (2022).