ASSETS 2026

Design and Evaluation of an AI-Based American Sign Language Video Comprehension Tool: Exploring the Benefits and Use of Automatic Segmentation and Sign Lookup



Chaelin Kim[1]
ckim13@tulane.edu



Denise Crochet[1]
crochet@tulane.edu



Maty Bohacek[2]
maty@stanford.edu



Saad Hassan[1]
saadhassan@tulane.edu

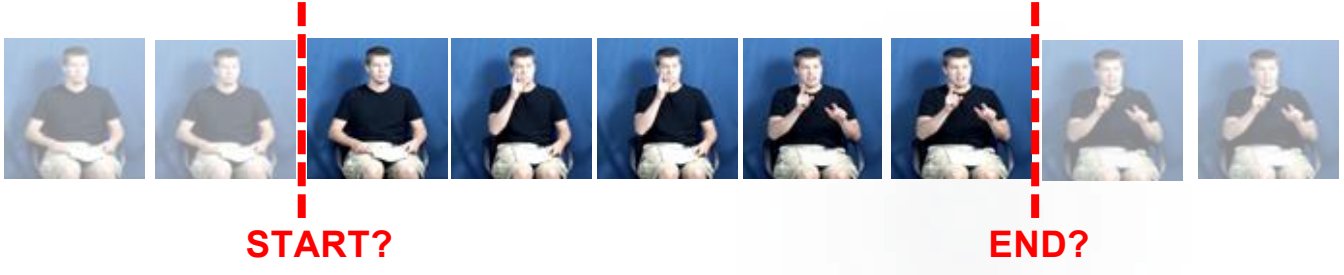
1 Tulane University

2 Stanford University

ASSETS '26 | October 25–28, 2026 | Vila Nova de Gaia, Portugal

Motivation

Many learners turn to **sign language videos** as a learning resource.



Learners struggle with determining sign boundaries

Signs change shape in context (coarticulation).

Existing dictionaries recognize isolated signs in their “citation” form.

Learners of spoken languages can highlight a word or read captions. But there is no comparable support for ASL videos.

There is no video-based learning support tool for segmenting and looking up an unfamiliar sign directly within a continuous ASL video.

Motivation

Prior work (e.g., Hassan et al., ASSETS 2022)

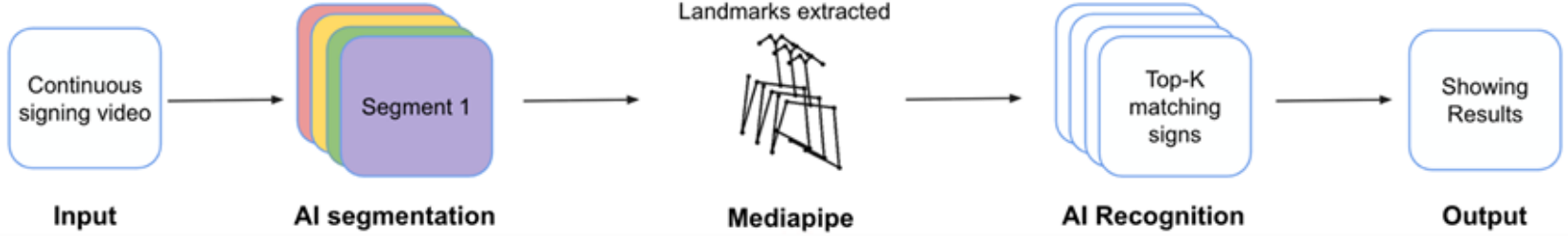
- Showed benefits in providing video-span selector to learners understanding continuous sign language videos and showed improvement over *feature-based* dictionaries.
- Mostly Wizard-of-Oz, with simulated recognition and hand-annotated segments

This work

- Compares video search with and without automatic segmentation to **study the benefits of automatic segmentation**
- Uses real AI models, so participants encounter real segmentation and recognition errors

Can AI-based automatic segmentation support comprehension of continuous signing videos for intermediate ASL learners?

AI Pipeline Overview

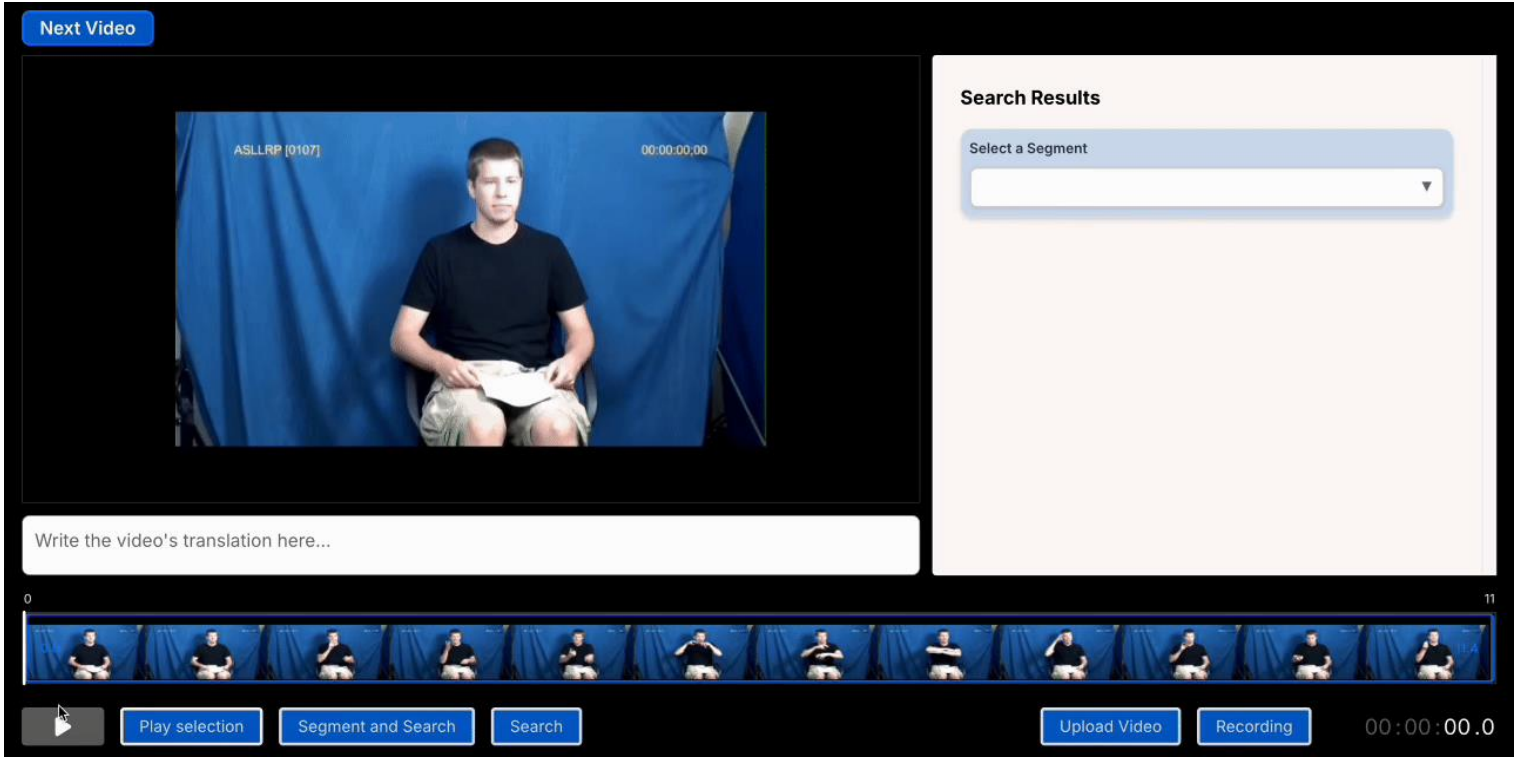


Segmentation: Linguistically motivated model (Moryossef et al., 2023) finds the start and end of each sign.

Recognition: ST-GCN pretrained on ASL Citizen and fine-tuned on 991 glosses. Returns top-20 matches with confidence level.

	Videos used in Study
Segmentation **IoU**	0.50 ± 0.19
Recognition **top-1**	0.61 ± 0.15
Recognition **top-20**	0.91 ± 0.07

Prototype Interface



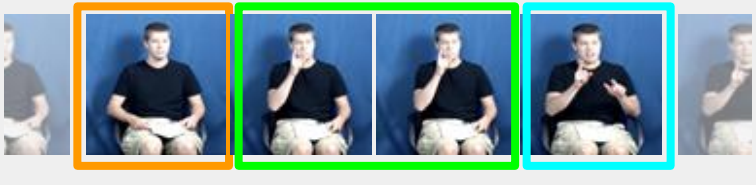
Two Conditions

Condition 1: Segment & Search (AI automatic segmentation)

1. Human selects a broad span



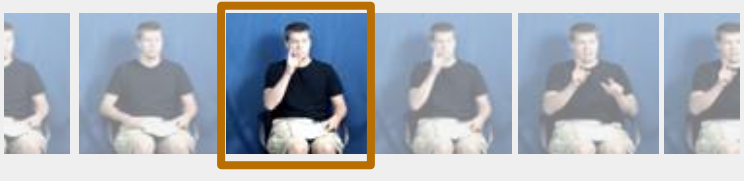
2. AI automatically segments the selected span into individual signs



3. Corresponding results for each individual segment within user selection

Condition 2: Search (Manual segmentation)

1. Human manually adjusts the start and end points to select a single sign



2. Get matching results for a single segment

Participants and Procedure

Participants

13 intermediate ASL learners with at least two ASL courses or two years of learning

Tasks

Four ASL videos, translated into English or gloss. Two videos per prototype, counterbalanced

Measures

NASA-TLX workload, search satisfaction, translation quality, interaction logs, and an interview



Condition 1: Segment & Search
(AI automatic segmentation)

Condition 2: Search
(Manual segmentation)

Questionnaire after each video



Interview

Research Questions

When ASL learners using an in-situ video-comprehension tool with and without AI-based automatic segmentation, are there differences in:

RQ1
perceived workload,

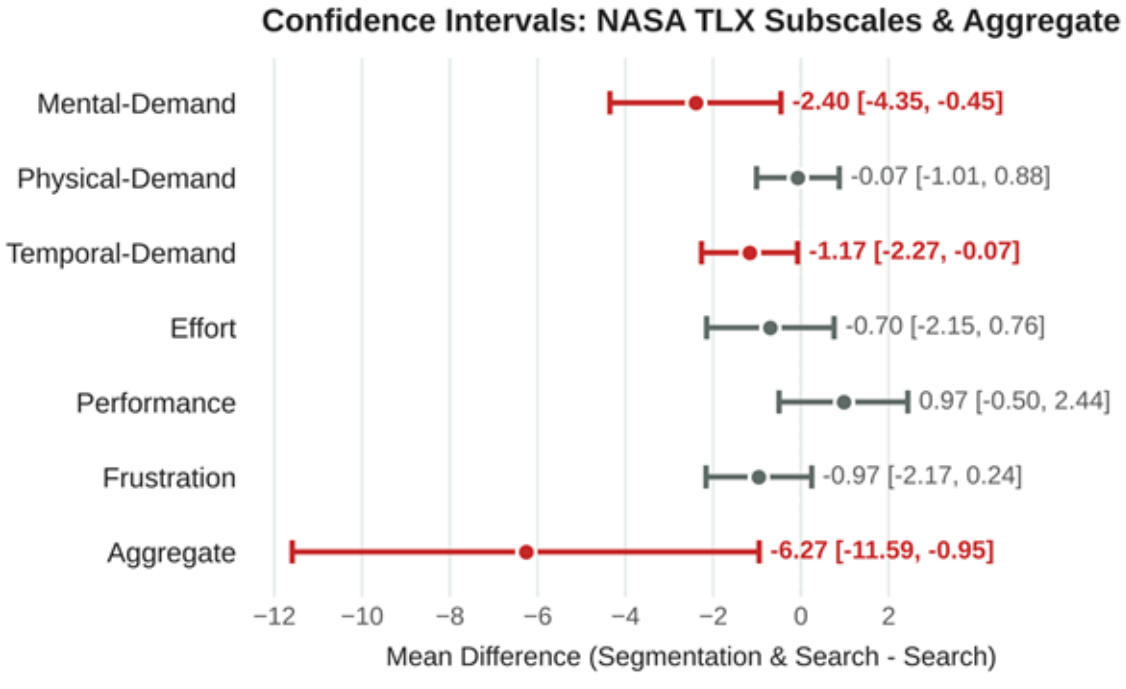
RQ2
reported satisfaction,

RQ3
translation quality,

RQ4
use of user-interface (UI) elements,
e.g., span selector,

in the prototype systems?

Findings 1. Automatic segmentation reduced perceived workload



Paired mean differences with 95% CIs.
Red marks $p < .05$

Findings 2. Higher satisfaction with the search experience



P3

"The first version (Segment & Search) was easier because it was pre-segmented for you, so you could easily look at the clips."

P11

"[I prefer the] AI one (Segment & Search)... It just made it quicker to find words. I also felt like it was more accurate. And finding the specific signs. "

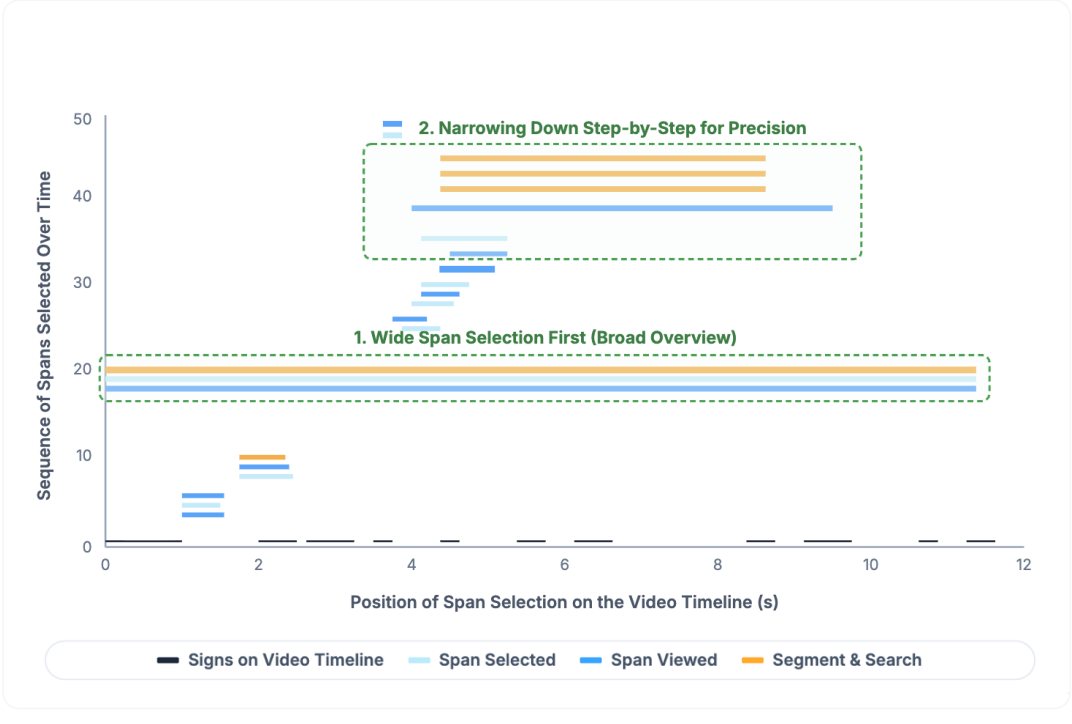
Findings 3. Segment & Search improved translation quality

58.9% → 66.8%	41.1% → 33.2%	5.49 → 7.04
More accurate signs	Fewer missing signs	Higer holistic score (/10)

- Eight participants translated at the sentence level; five used glosses or a combination of both.
- Sign-level measures are normalized by the number of signs per video

Findings 4. Different ways of using the span selector

P10 Working on Translating Video 2

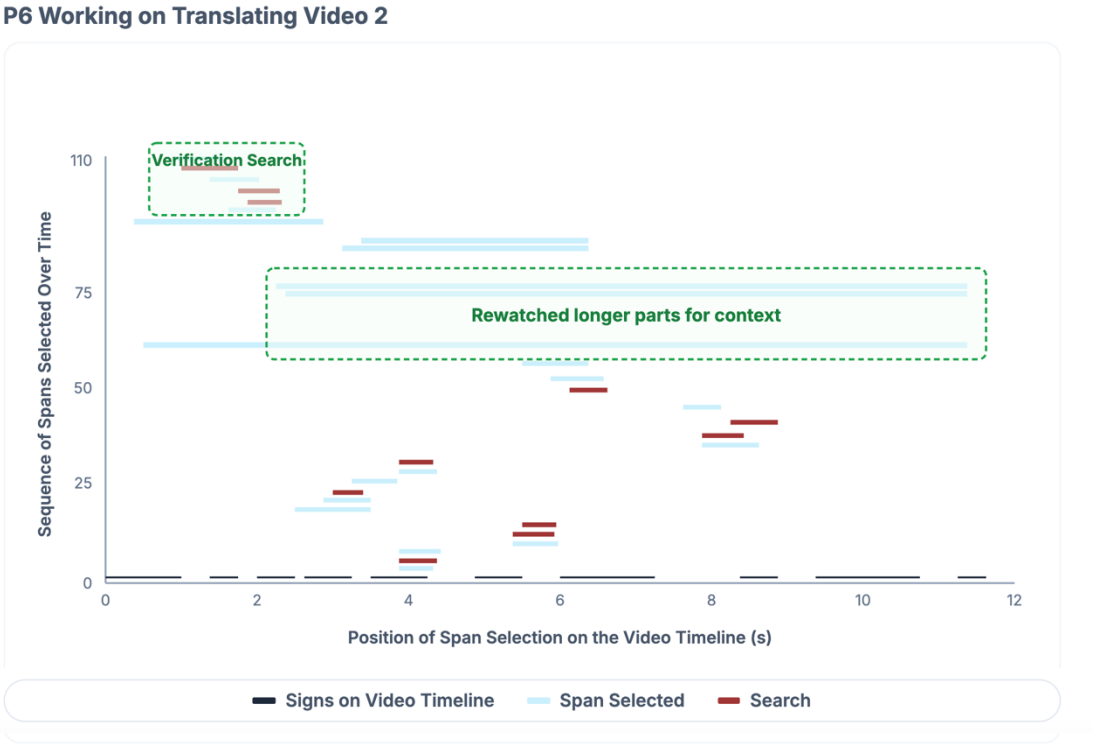


Segment & Search

Participant P10

Wide first, then narrow

Findings 4. Different ways of using the span selector



Search-only

Participant P6

Back and forth for context

Conclusion and Design Implications

Takeaways

- Automatic segmentation can be beneficial. It lowered workload and raised search satisfaction
- Manual selection encouraged closer, more deliberate analysis of signs

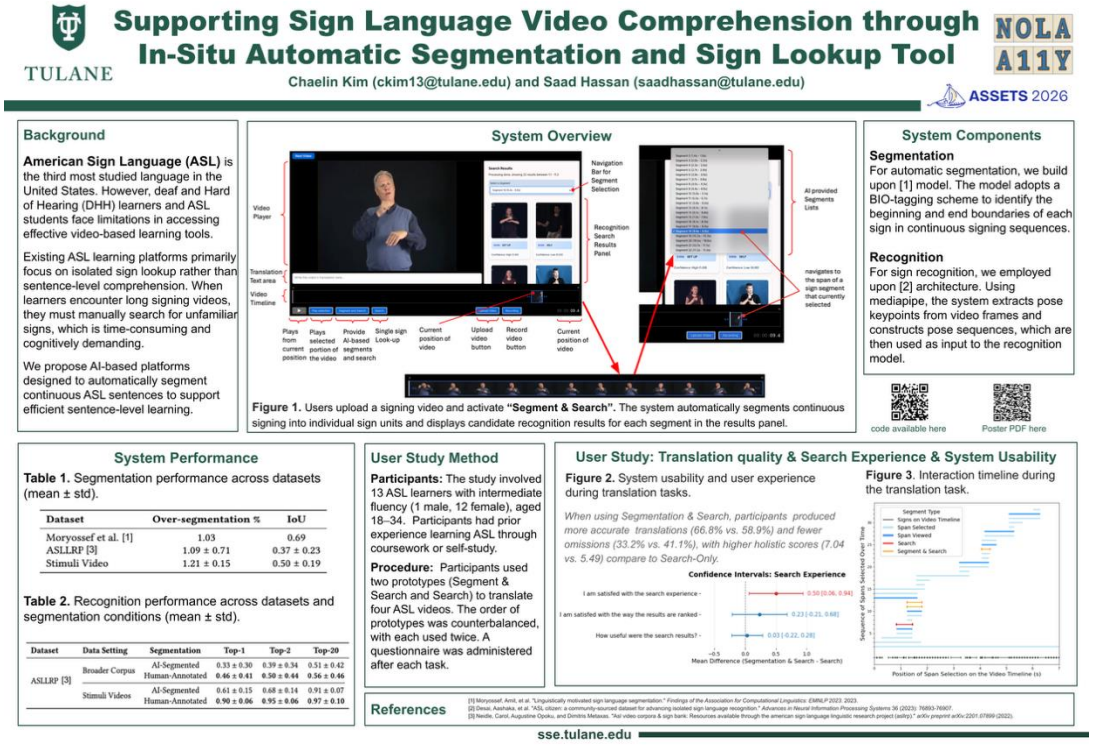
Design implications

- Feature-based filtering to refine imperfect results
- Slow motion, replay, and preset span lengths
- Feedback on span quality and confidence
- An in-situ dictionary as a fallback
- "Assist" and "practice" modes matched to learner goals

TAKE-HOME MESSAGE

Even with imperfect state-of-the-art AI, automatic segmentation made in-video sign lookup less demanding and more satisfying for ASL learners.

Demo Tomorrow!



Come try both prototypes!

DAY

Tuesday, Oct. 27

SESSION

Poster and Demo Session B

TIME

10:20–11:00 AM and 3:20–4:00 PM

LOCATION

Foyer Auditorium

Thank you


Chaelin Kim


Denise Crochet


Maty Bohacek


Saad Hassan

Contact

Chaelin Kim

Email: ckim13@tulane.edu

Website: chaelin0722.github.io

Lab

Human-Centered Computing Group, Tulane University

Lab website: saadh.info/a11y

NOLA A11Y

Paper available here!

Code available here!