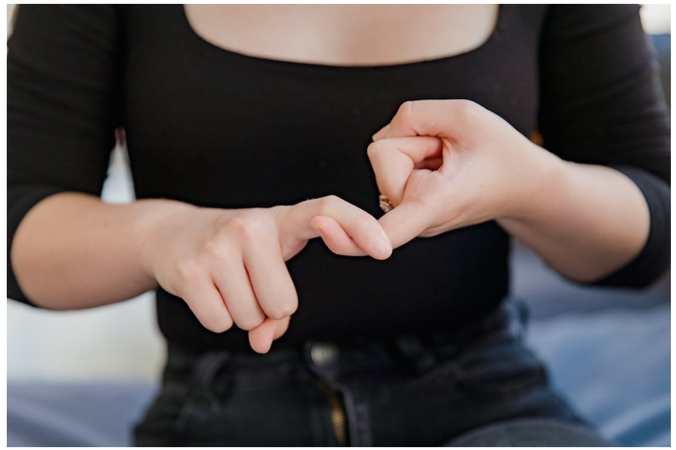
Sign-to-911: Emergency Call Service for Sign Language Users with Assistive AR

Yunqi Guo*, Jinghao Zhao*, Boyan Ding*, Congkai Tan*, Weichong Ling*, Zhaowei Tan†, Jennifer Miyaki‡, Hongzhe Du*, Songwu Lu*

*UCLA Computer Science, †UC Riverside, ‡UCLA Linguistics

ASL Interpreter: Zachary Goldstein

Background: Deaf and Sign Language



Global

- **> 70M** people are deaf [1]

In the US

- **15%** (~37.5 million) of adults have some trouble hearing [2]
- **0.5-2M** primarily communicate using sign language [3]
- **American Sign Language (ASL) is a natural language** with distinct vocabularies and grammars

[1] Deafness And Hearing Loss Statistics
[2] Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: National Health Interview Survey, 2012 (PDF). National Center for Health Statistics. Vital Health Stat 10(260). 2014.
[3] Mitchell, Ross E., et al. "How many people use ASL in the United States? Why estimates need updating." Sign Language Studies 6.3 (2006): 306-335.

Communication Barrier

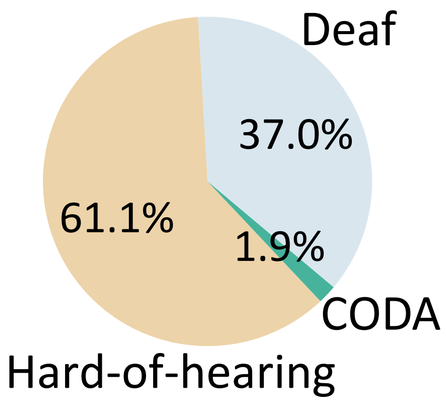
- **Lack of interpreter:** 1 ASL interpreter per 50 deaf individuals [1].
- **First step:** support emergency communication
  - This topic is suggested by my Deaf friend, Mark.

[1] https://deafservicesunlimited.com/asl-interpreter-shortage-and-accessibility-in-higher-education/

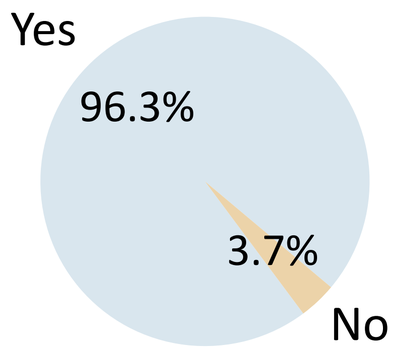
Communication Upon Emergencies

Our user survey (54 responses from Deaf community)

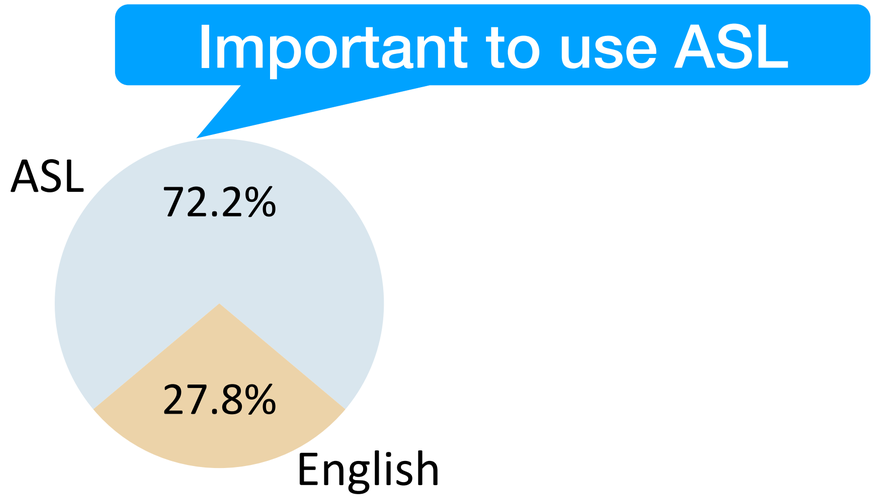
- **96%** face challenges to reach emergency services.



(a) User groups



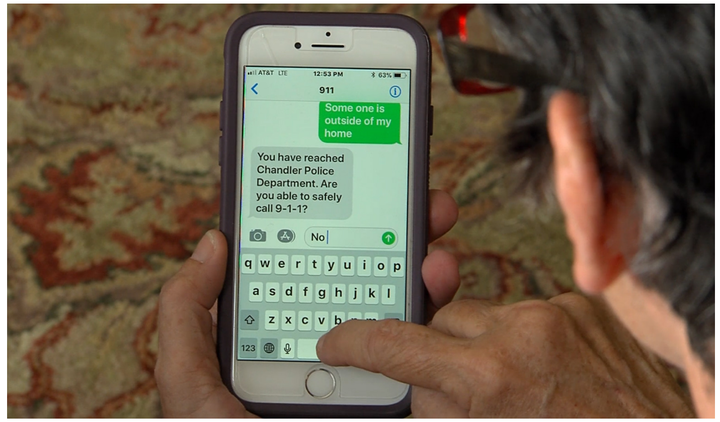
(b) Had difficulties in calling emergency?



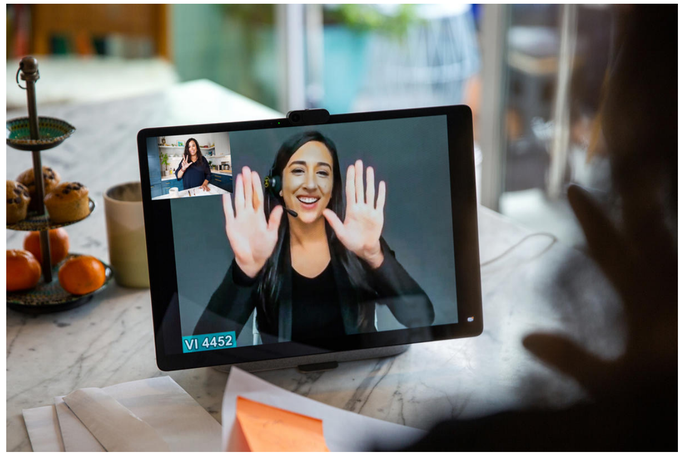
(c) Preference in emergency call?

Current Emergency Services for Deaf People

- **Text-based communication**
  - Not all deaf people are fluent in English
- **Video relay services (VRS)**
  - Requires network and human interpreters



Text-to-911



VRS

How to Provide Accessible Sign Emergency Services?

Goals

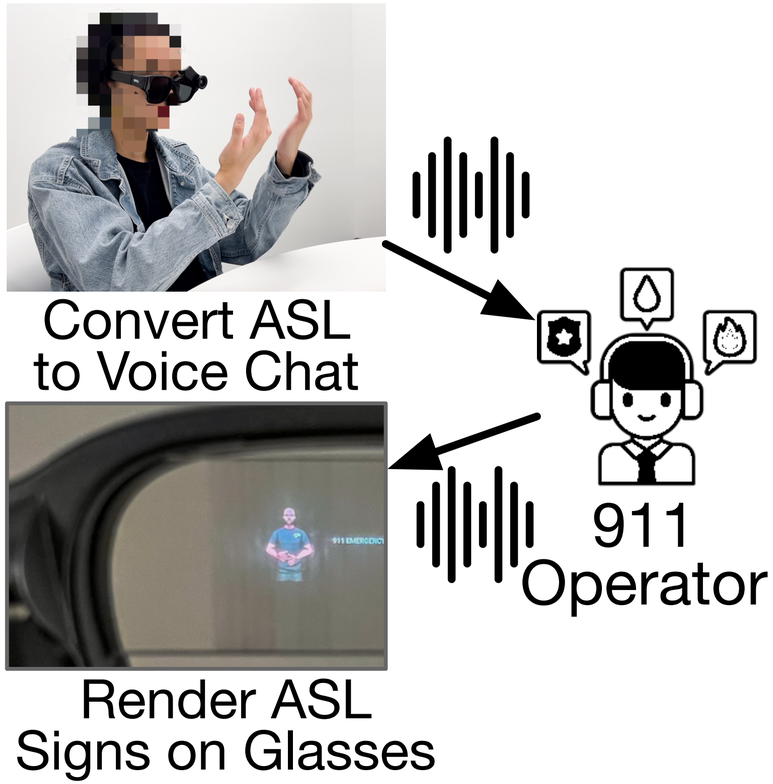
- **Accessible:** easy to use and carry
- **Bi-directional**: ASL ↔ Spoken English
- **Standalone:** No dependency on remote processing or connectivity

Core Ideas

1. **Leverage ASL linguistic features**
2. **Exploit mobile, wearable computing, and smaller models**
3. **Working with the Deaf community**

Sign-to-911 System Overview

- **AR Glasses:** capturing and displaying signs
- **Smartphone:** processing and voice call interface
- **Connection in between:** Bluetooth (150KB/s)

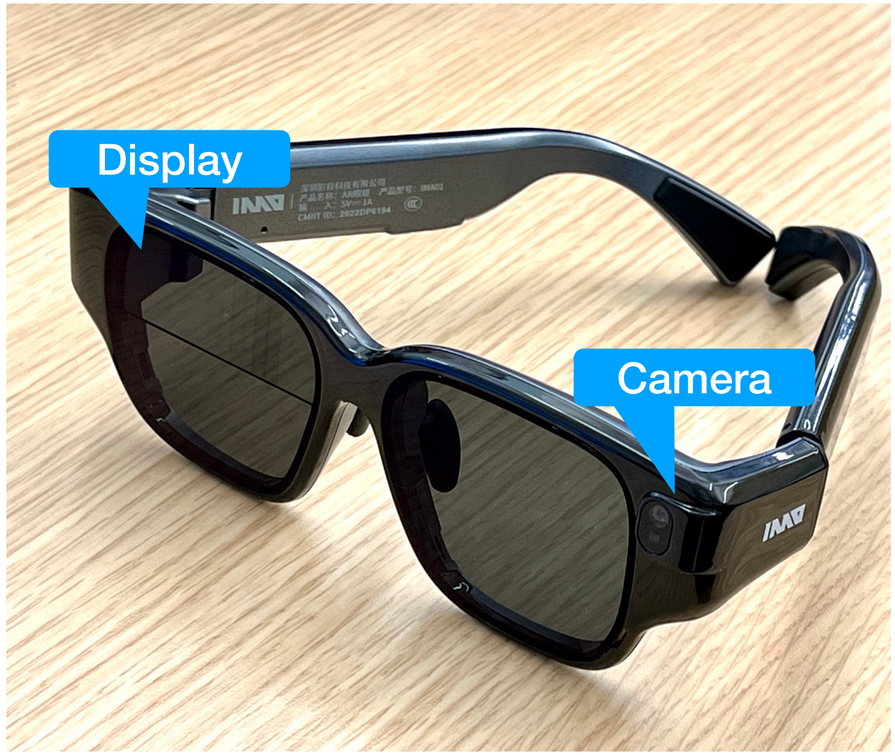


Why AR Glasses?

Bi-directional Capability

Accessibility:

- Compact and non-intrusive
- Keeps hands free for signing
- Already used by deaf people for live captions [1]



Assistive AR glasses (~$350)

[1] https://patient-innovation.com/post/7501?language=en

System Functions

1. **ASL-to-English:** Fast and Accurate Sign Translation
2. **English-to-ASL:** Efficient Production & Rendering from English

ASL-to-English: Domain-Oriented Translation

ASL-to-English translation steps

Step 1: Capture sign with sign parameters

From Ling. Parameters to ML Parameters

Sign parameters (Stokoe [1])

1. **Hand number**
2. **Handshape**
3. **Location**
4. **Movement**
5. **Palm orientation**
6. ~~**Non-manual**~~

Modified parameters for ML

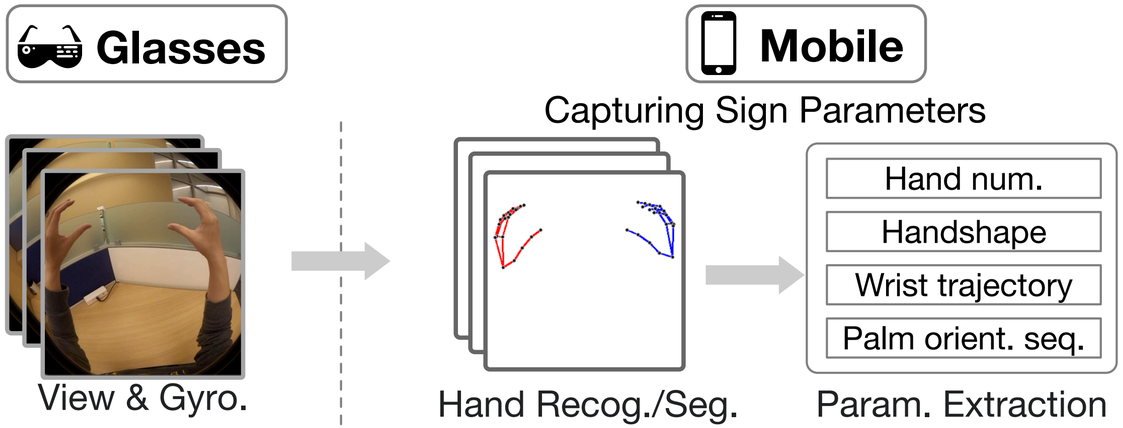
1. **Hand number:** probabilities, 1 or 2 handed
2. **Handshape:** probabilities of 40 basic handshape
3. **Wrist trajectory:** (x, y, z) over time
4. **Palm orientation:** (α, β, γ) over time

Captured through glasses

[1] Tennant, R. A., Gluszak, M., and Brown, M. G. The American sign language handshape dictionary. Gallaudet University Press, 1998.

How to Extract Sign Parameters?

1. **Hand recognition**
   - ➢ Offload processing to the phone
2. **Segmentation**
   - ➢ Pause time and hand neutral hand position
3. **Adaptive Extraction**
   - ➢ **Merge handshape:** 1) classify handshape with NN; 2) keep start/end handshapes



Parameter	Dimension #
Hand number	2
Handshape	40 x 2
Wrist trajectory	3 x 2 x t
Palm orientation	3 x 2 x t

Sign Parameter Dimensions

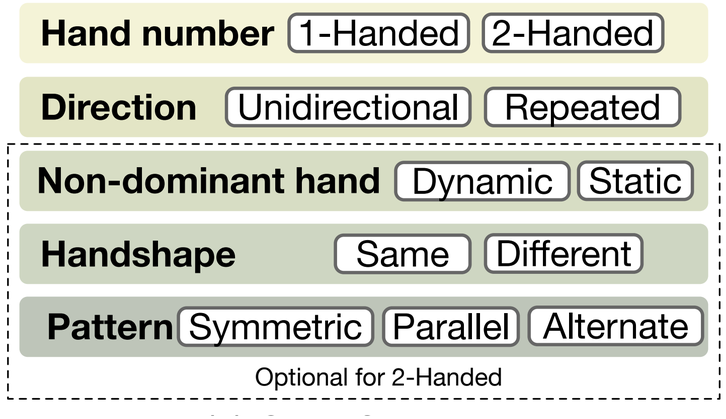
Step 2: Sign Recognition from Parameters

a) **Sign categorization:** Signs are classified along five hierarchical dimensions

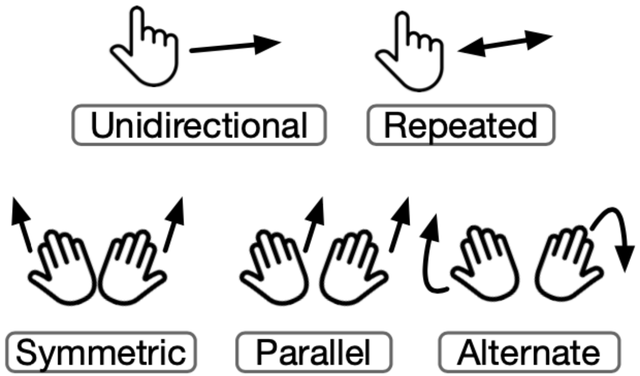
- E.g., Movement Trajectory → Pattern

b) **Recognize signs:**

- DTW + Regression model to learn sign parameter weights to classify signs



Sign Categories



Movement Patterns

Step 3: Sentence Translation

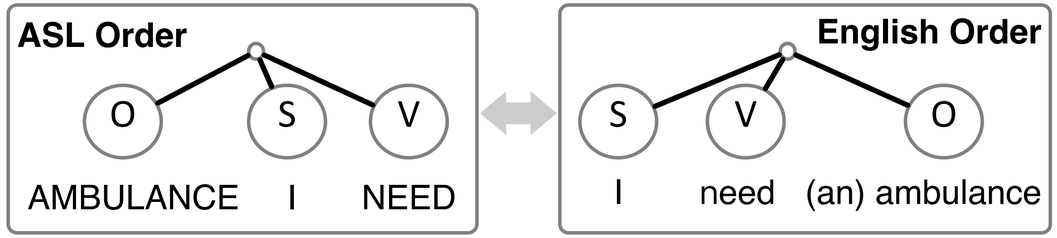
- **ASL grammar rules differ from that of English.**
- **Example:**
  - English: **I have a fire emergency.** (Subject-Verb-Object, SVO)

    ASL: **FIRE EMERGENCY I HAVE.** (Object-Subject-Verb, OSV)

I have fire emergency

Sentence Translation with Grammar Tree

- **Learn from samples**
  - 1233 pairs: [ASL Gloss Seq., English]
- **Sentence translation**
  - Parse and map with the tree structures
- **Leveraging 911 contexts to refine recognition**
  - Emergency type & questions



Grammar Mapping

English-to-ASL: Phoneme-based ASL Production

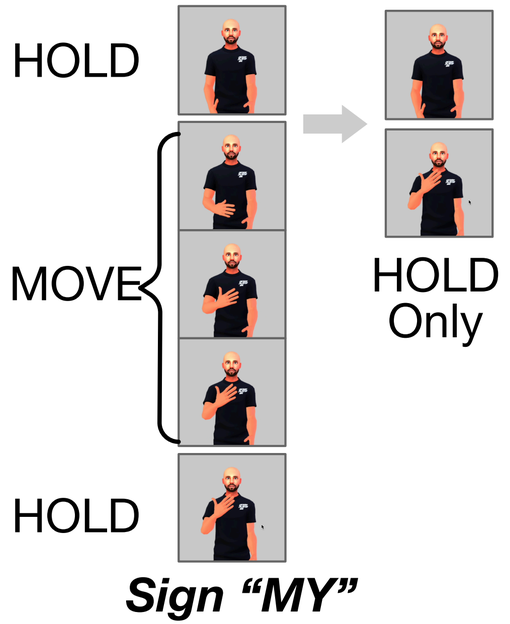
Phonemes are basic unit in ASL[1], contains

- **HOLD:** Static gestures; 1 to 5 per sign
- **MOVE:** Transition between two HOLD states

Phoneme extraction: Motion capture from sign videos

Up to 50x bandwidth reduction with kinematic compressions

→ Work with Bluetooth setup



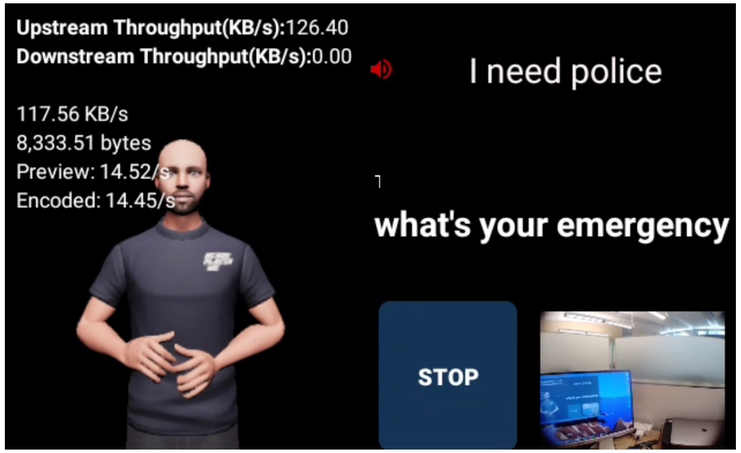
[1] Liddell, Scott K., and Robert E. Johnson. "American sign language: The phonological base." *Sign language studies* 64.1 (1989): 195-277.

Implementation

Commodity Devices:

- **Glasses:** INMO AIR priced at $350.
- **Smartphone:** Android OS, ranging from $200 to $550.

Voice call captured with **Android Accessibility APIs.**



Glass App

Datasets

Glass-view Sign Traces

- **911 conversations from multiple resources**

  D1 & D2: Derived from **real-world emergency templates**; 550 signs.

  D3: 911 recordings and ChatGPT synthesized, 30 conversations with 150 Q&As.

- **By authentic Signers*:** 2 native ASL user, 4 advanced ASL students and researchers

- **Data size:** 249 GB video traces; 11.5-hour signing

Glass-view collection

ASL production

- 3000+ signs from video samples

* UCLA IRB #: 23-000239

System Evaluation

Sign Capture and Sentence Translation:

- **Accrete:** Achieved +5.1% word acc. improvement

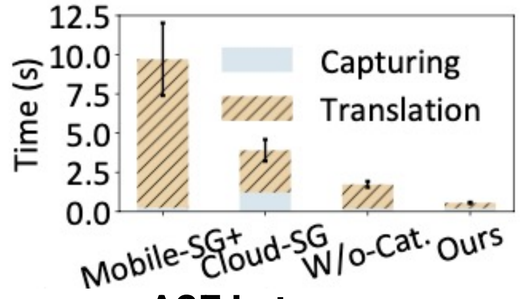
End-to-end Latency:

- **A2E:** 0.55s; 17x reduction from other models on mobile;7x reduction from cloud processing
- **E2A:** 0.21s; >15x reduction from skeleton/video streaming

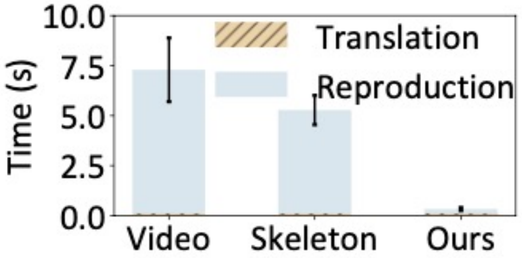
Overhead: at the level of video players

Model	D3
I3D	80.18
SL-GCN	86.31
Sign-to-911	**91.37**

Translation Word Acc.(%)



A2E Latency

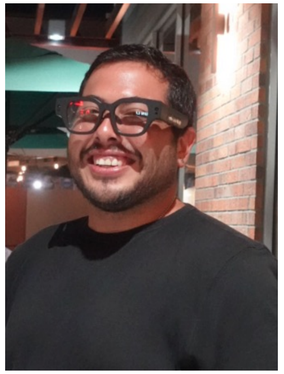


E2A Latency

User Study

Participants:

- 12 from CSUN, GLAD, and UCLA, and local communities
- Age Demographics: Spanning ages 20 to 80.



Approach	Accs.	Usab.	Oval.
Text-based	2.9	2.8	2.8
VRS	3.4	3.5	3.6
Sign-to-911	**4.2**	**4.3**	**4.6**

QoE from User Study

Summary

Sign-to-911

- Introducing bi-directional ASL interaction through Assistive AR
- Solutions oriented from linguistic domain: it’s lightweight, efficient, and accurate

Keep in Mind

- Must address the **needs, wants, and concerns** of the Deaf community
- **Linguistic diversity** of the Deaf community
  - > 200 sign languages in the world
  - The US has other sign language and varieties, e.g., SEE, Black ASL

Thank you!

Stay tuned: AnySign.net

Turning Accessibility into Action!

Contact:
linktr.ee/yqguo