Gestures for Blind People to Specify Spatial Focus of Vision Language Models on Head-Worn Devices

Shoya Higuchi [1] Masaki Kuribayashi [1] Arata Ito [1] Shigeo Morishima [1]

1 Waseda University

VRST 2026

Introduction

Goal: Enable flexible framing on HWDs for blind.

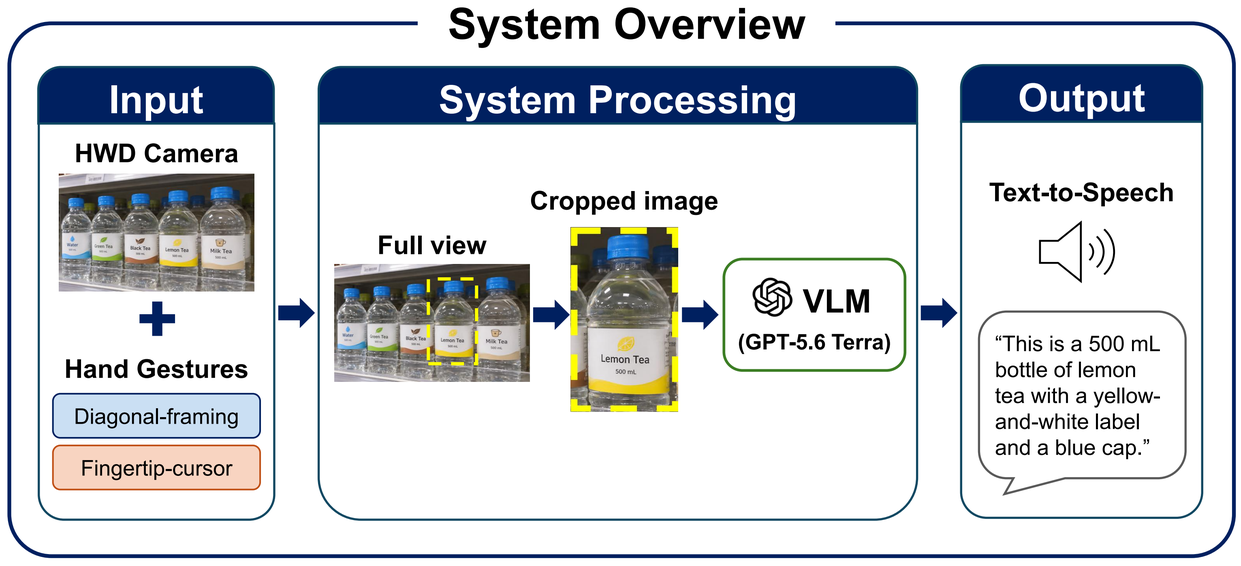
VLM-based head-worn devices (HWDs) enable hands-free visual assistance through egocentric cameras.
However, **blind users may struggle to capture only the content they want described** [2].



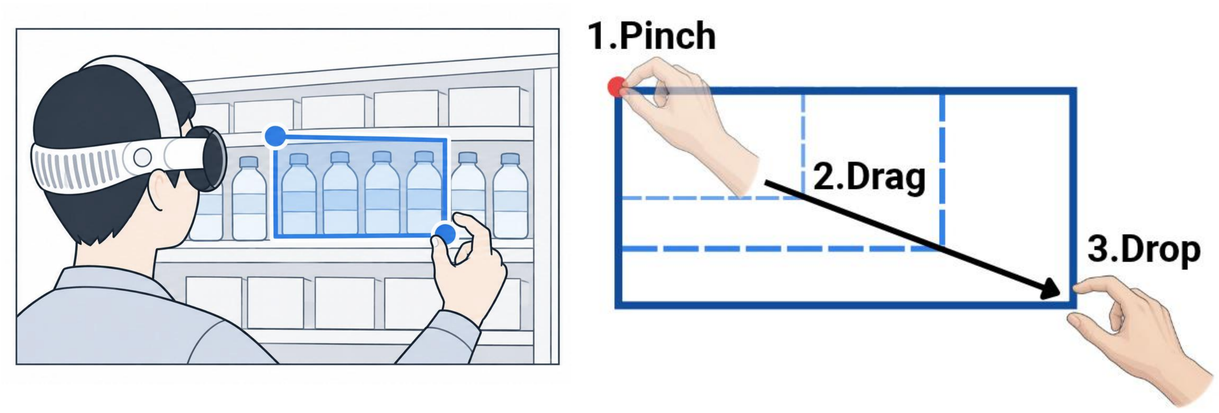
Automatic object identification may not reflect the user's intended focus [1,2].

Proposed Interaction

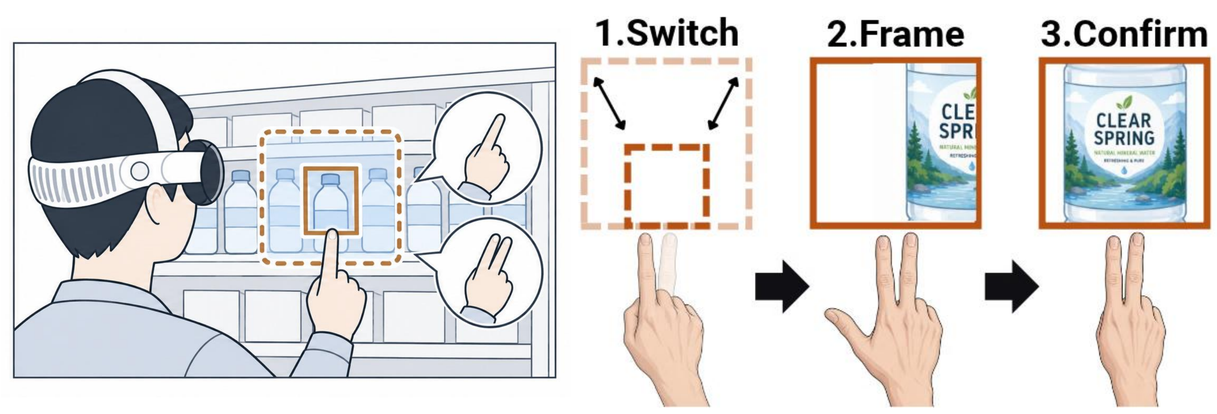
Developed **two gesture-based methods** on Apple Vision Pro.



Diagonal-framing Method: Define a rectangular frame by pinching and dragging diagonally.



Fingertip-cursor Method: Place a square at the fingertip, resize it, and confirm with the thumb.



User Study with Blind People (N=5)

Task: 3 targets × 3 conditions × 3 trials

Conditions: The two methods and Seeing AI [3].

Targets: Red outlines indicate the target areas.



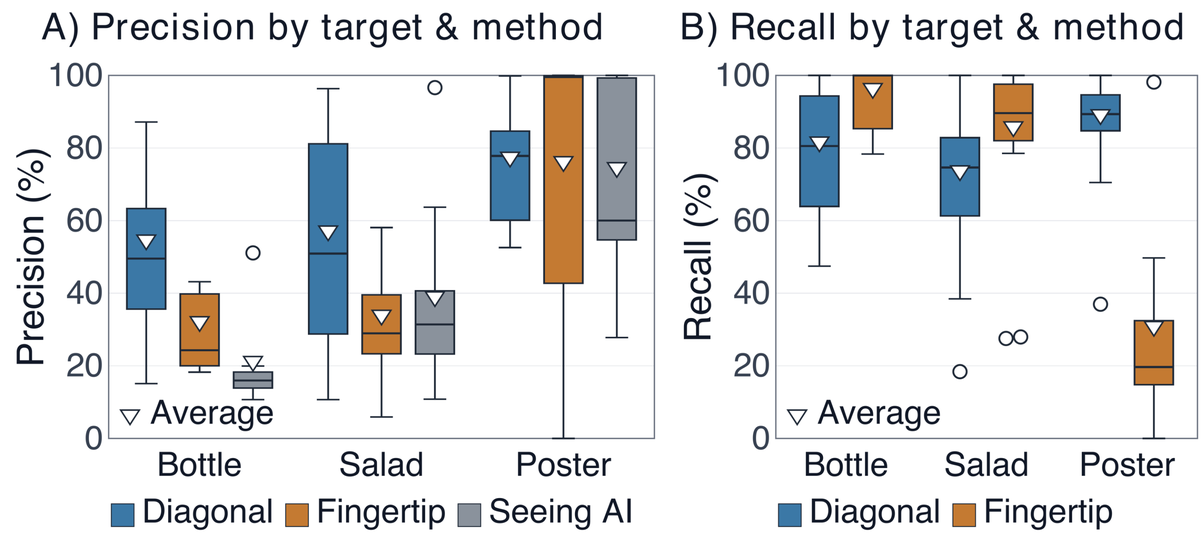
Metrics: We measured ***Precision*** and ***Recall***.

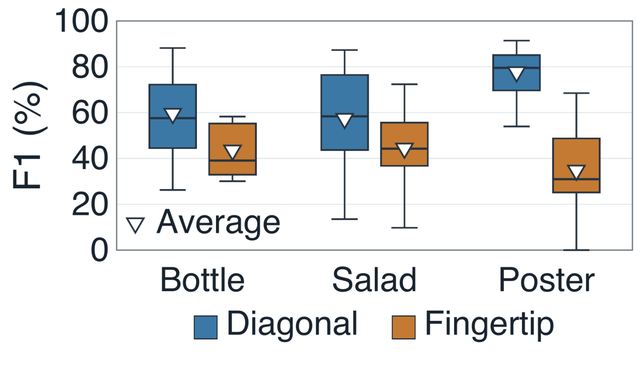
$$Precision = \frac{\text{Overlap}}{\text{Framed Area}} \qquad Recall = \frac{\text{Overlap}}{\text{Target Area}}$$

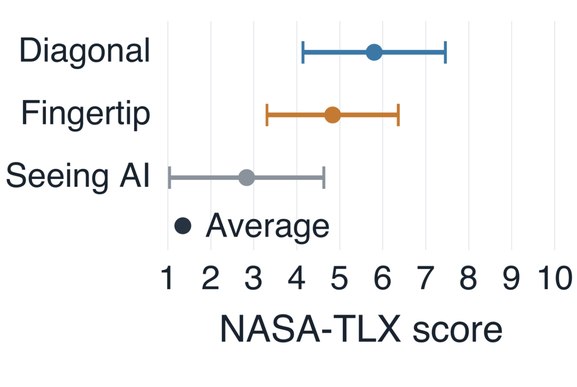


Result

- Higher framing accuracy tended to come with greater workload.
- High Precision did not always guarantee high Recall.
- Familiarity with the interface may influence subjective workload.







Comments:

P1: *"It's often hard to specify 'from here to here' just with words, but it might be convenient to specify the exact area spatially."*

P5: *"I think it's better to be able to know whether things are being captured properly in real time."*

Conclusion

- Our study showed a trade-off between performance and workload.
- Future work will focus on reducing this workload by incorporating the user feedback from this study.

References:
[1] Chang et al. (2026). "TouchScribe: Augmenting Non-Visual Hand-Object Interactions with Automated Live Visual Descriptions." CHI 2026.
[2] Lee, K., Shrivastava, A., & Kacorri, H. (2020). "Hand-Priming in Object Localization for Assistive Egocentric Vision." WACV 2020, pp. 3411–3421.
[3] Microsoft Corporation. 2017. Seeing AI. https://www.seeingai.com/ (accessed November 2, 2026).