CHI 2018
Engage with CHI

Pervasive Human Computer Interaction
Department of Computer Science and Technology
Tsinghua University

VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval

Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, Yuanchun Shi

清華大學
Tsinghua University

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

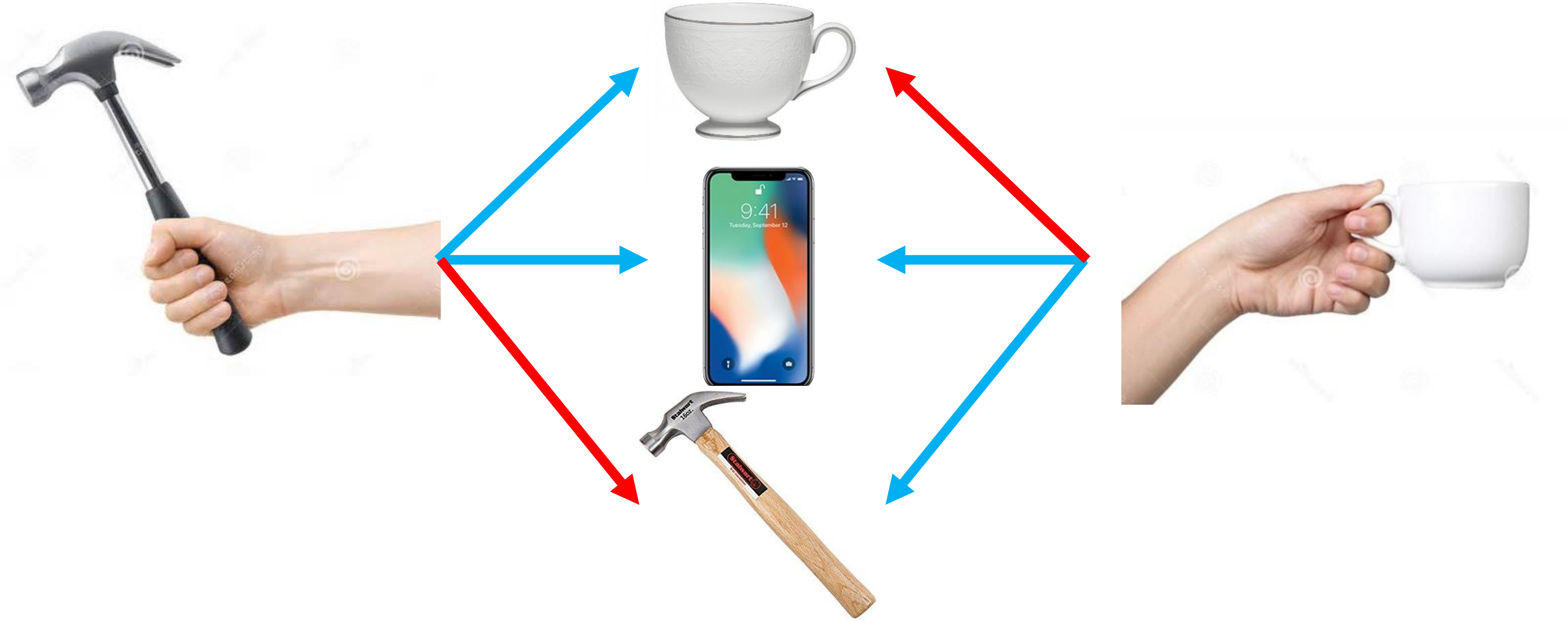
Thor's Hammer

To retrieve a **Virtual** object in VR, users perform the gesture of **Grasping** it in physical world.

Users perform **Swipe** gesture to select among the objects, and **Dwell** to confirm

MOTIVATION

To provide a set of **Self-Revealing Gestures** for object retrieval



MOTIVATION

To provide a set of **Self-Revealing Gestures** for object retrieval

1. Will users **consistently** perform the same grasping gesture for each object?

2. Can the grasping gestures of objects be **distinguished** by algorithms?

BACKGROUND

Gesture Interaction

Advantages

- Intuitive
- Direct Interaction
- Semantic Meaning
- Eyes-Free Interaction

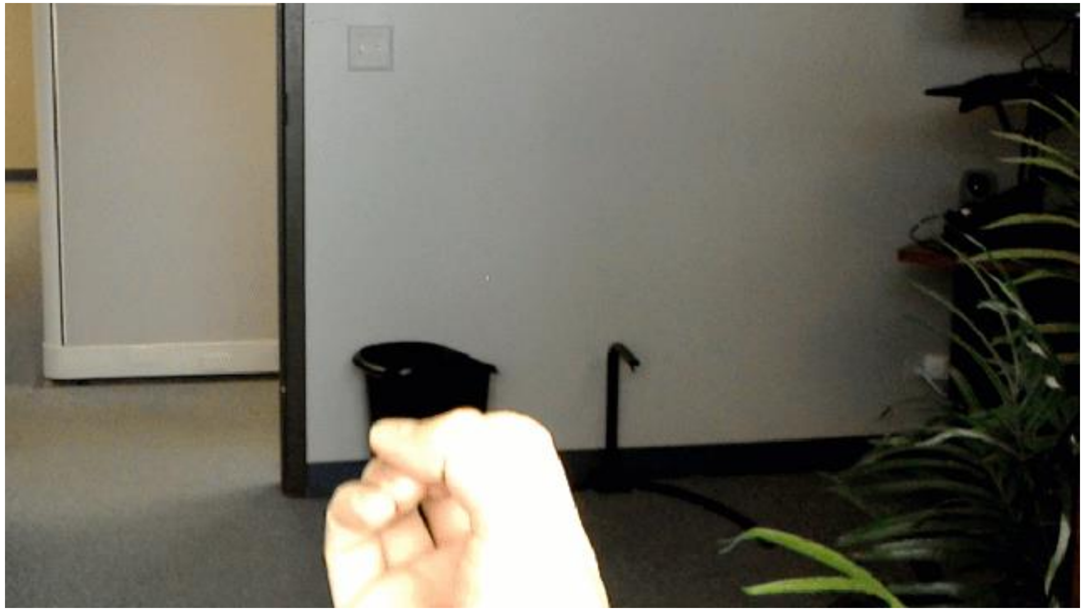
Disadvantages

- **Non Self-Revealing**
- Fatigue
- Fuzzy Input

BACKGROUND

Gesture Interaction

Hard to **Discover**



Bloom gesture to open the menu

Draw circle to open the camera

BACKGROUND

Gesture Interaction

Hard to Learn

BACKGROUND

Gesture Interaction

Hard to Remember

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Gesture Interaction

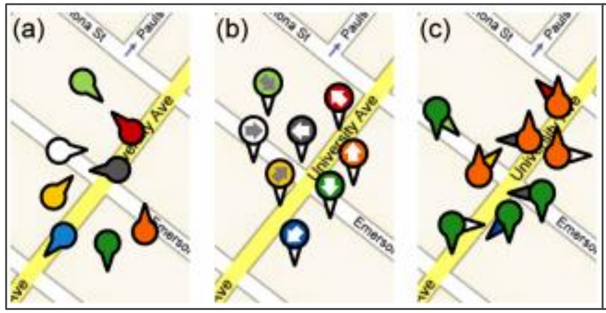
Mappings from Targets to Gestures

- Simple and easy to understand
- Consistent with acquired experience
- Consensus across different users

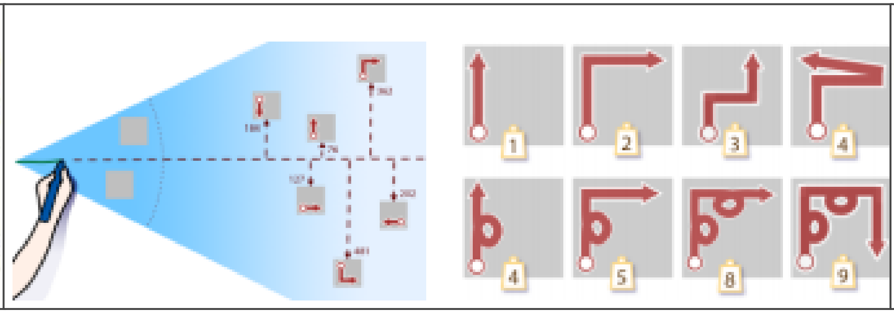
RELATED WORK

Approaches for Mapping Problems

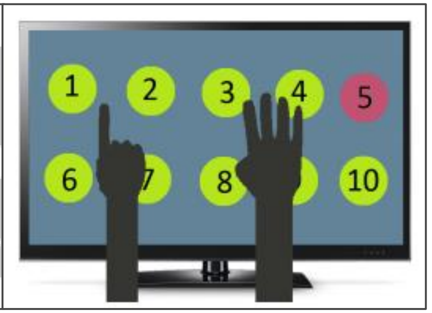
Look-and-Feel Design of the Targets



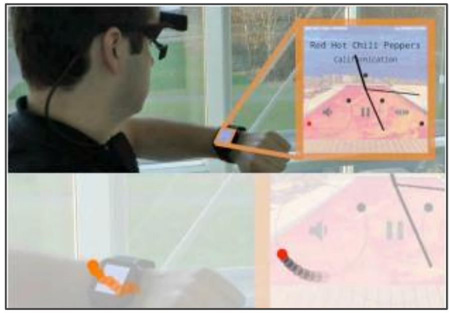
Yatani et al. CHI 08



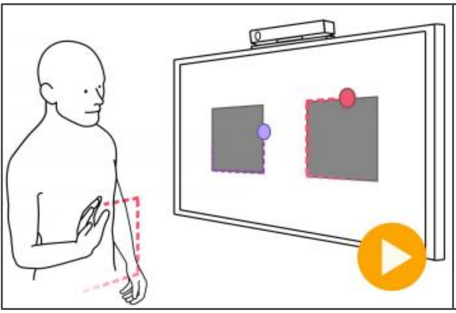
Bragdon et al. CHI 11



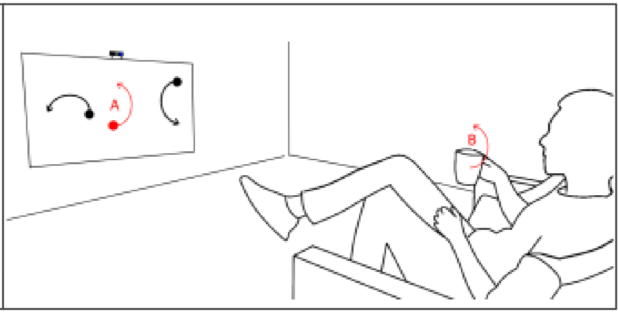
Wagner et al. CHI 14



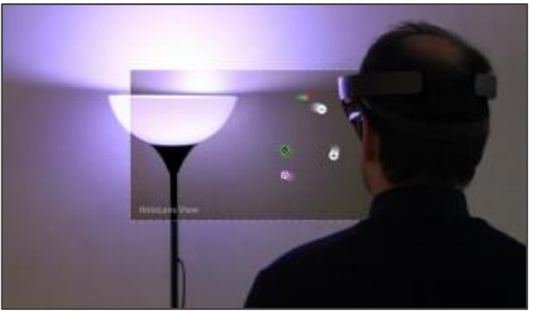
Esteves et al. UIST 15



Carter et al. CHI 16



Clarke et al. UIST 17



Esteves et al. UIST 17

BACKGROUND

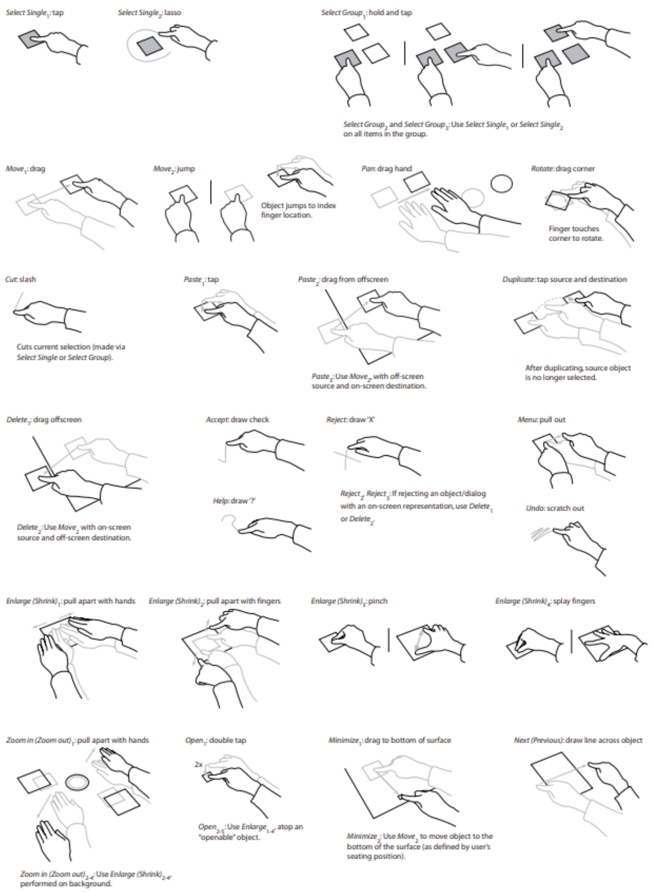
Gesture Interaction

Mapping from Targets to Gestures

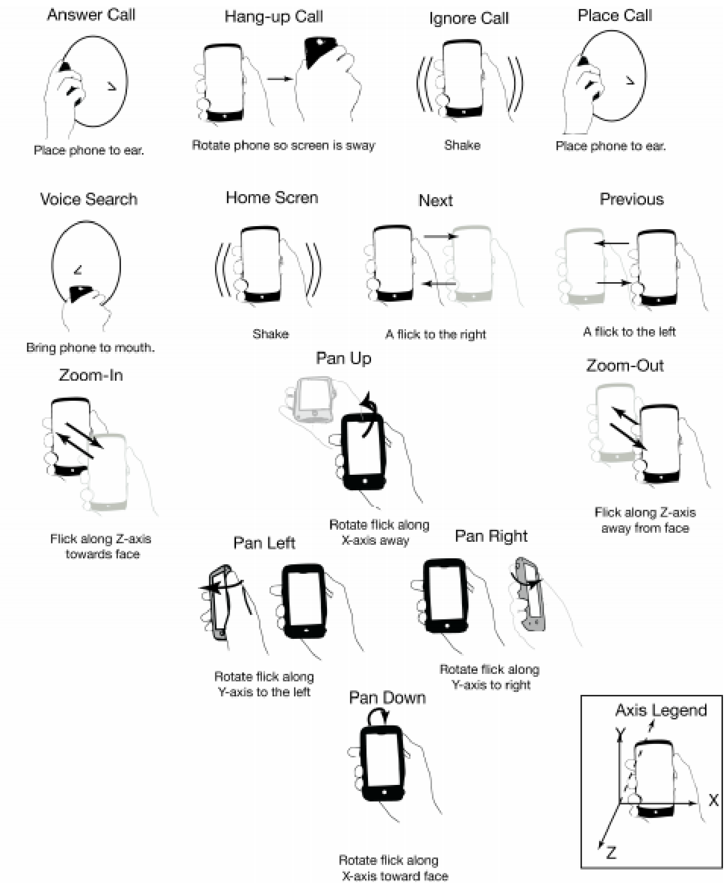
- Simple and easy to understand
- Consistent with acquired experience
- Consensus across different users

Approaches for Mapping Problems

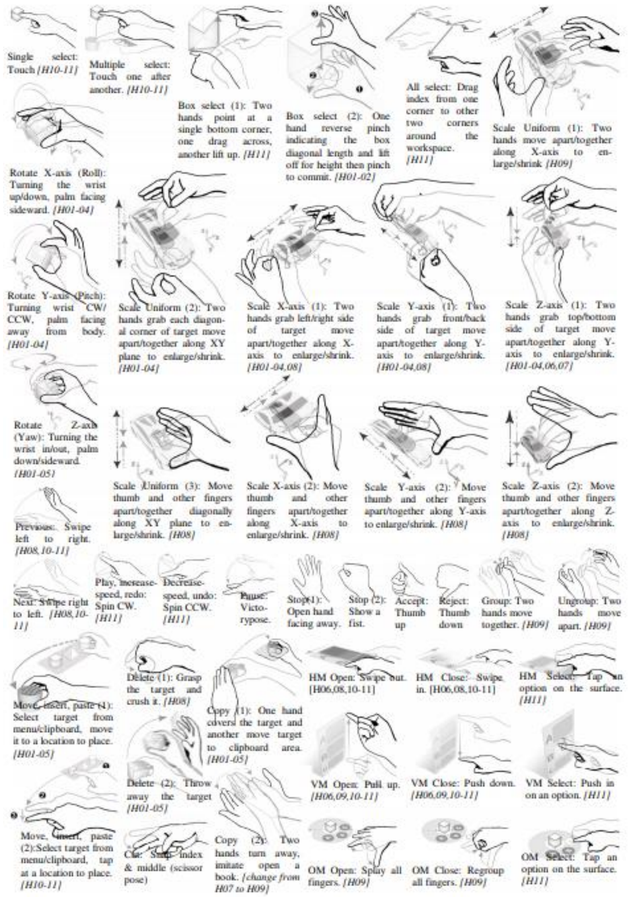
User Defined Gestures (Participatory Design)


Select Single1: tap
Select Single2: lasso
Select Group1: hold and tap
Select Group2 and Select Group3: Use Select Single1 or Select Single2 on all items in the group.
Move1: drag
Move2: jump
Object jumps to index finger location.
Pan: drag hand
Rotate: drag corner
Finger touches corner to rotate.
Cut: slash
Cuts current selection (made via Select Single or Select Group).
Paste1: tap
Paste2: drag from offscreen
Paste3: Use Move2, with off-screen source and on-screen destination.
Duplicate: tap source and destination
After duplicating, source object is no longer selected.
Delete1: drag offscreen
Delete2: Use Move2 with on-screen source and off-screen destination.
Accept: draw check
Help: draw '?'
Reject: draw 'X'
Reject2, Reject3: If rejecting an object/dialog with an on-screen representation, use Delete1 or Delete2.
Menu: pull out
Undo: scratch out
Enlarge (Shrink)1: pull apart with hands
Enlarge (Shrink)2: pull apart with fingers
Enlarge (Shrink)3: pinch
Enlarge (Shrink)4: splay fingers
Zoom in (Zoom out)1: pull apart with hands
Zoom in (Zoom out)2-4: Use Enlarge (Shrink)2-4, performed on background.
Open1: double tap
2x
Open2-5: Use Enlarge1-4, atop an "openable" object.
Minimize1: drag to bottom of surface
Minimize2: Use Move2 to move object to the bottom of the surface (as defined by user's seating position).
Next (Previous): draw line across object

Wobbrock et al. CHI 09



Ruiz et al. CHI 11

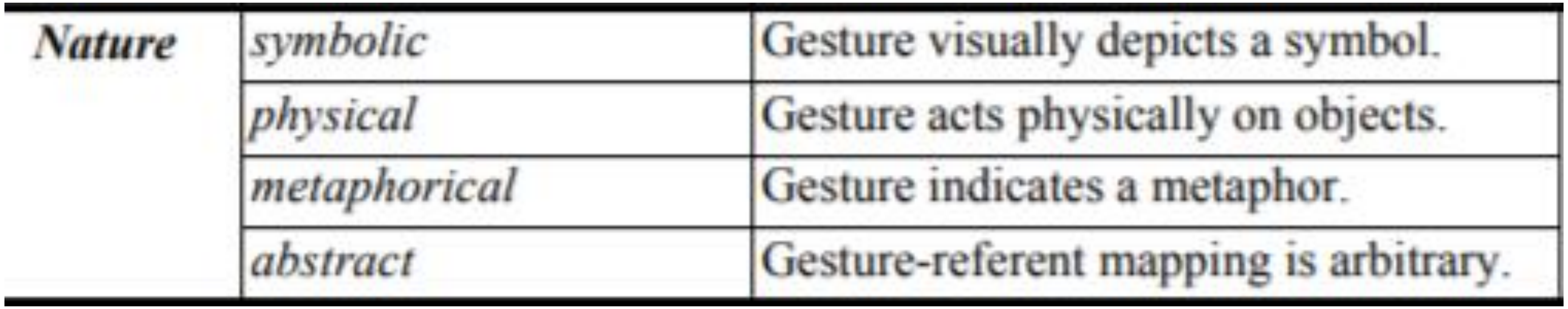

Single select: Touch [H10-11]
Multiple select: Touch one after another. [H10-11]
Box select (1): Two hands point at a single bottom corner, one drag across, another lift up. [H11]
Box select (2): One hand reverse pinch indicating the box diagonal length and lift off for height then pinch to commit. [H01-02]
All select: Drag index from one corner to other two corners around the workspace. [H11]
Scale Uniform (1): Two hands move apart/together along X-axis to enlarge/shrink [H09]
Rotate X-axis (Roll): Turning the wrist up/down, palm facing sideward. [H01-04]
Rotate Y-axis (Pitch): Turning wrist CW/CCW, palm facing away from body. [H01-04]
Scale Uniform (2): Two hands grab each diagonal corner of target move apart/together along XY plane to enlarge/shrink. [H01-04]
Scale X-axis (1): Two hands grab left/right side of target move apart/together along X-axis to enlarge/shrink. [H01-04,08]
Scale Y-axis (1): Two hands grab front/back side of target move apart/together along Y-axis to enlarge/shrink. [H01-04,08]
Scale Z-axis (1): Two hands grab top/bottom side of target move apart/together along Y-axis to enlarge/shrink. [H01-04,06,07]
Rotate Z-axis (Yaw): Turning the wrist in/out, palm down/sideward. [H01-05]
Scale Uniform (3): Move thumb and other fingers apart/together diagonally along XY plane to enlarge/shrink. [H08]
Scale X-axis (2): Move thumb and other fingers apart/together along X-axis to enlarge/shrink. [H08]
Scale Y-axis (2): Move thumb and other fingers apart/together along Y-axis to enlarge/shrink. [H08]
Scale Z-axis (2): Move thumb and other fingers apart/together along Z-axis to enlarge/shrink. [H08]
Previous: Swipe left to right. [H08,10-11]
Next: Swipe right to left. [H08,10-11]
Play, increase-speed, redo: Spin CW. [H11]
Decrease-speed, undo: Spin CCW. [H11]
Pause: Victorypose.
Stop (1): Open hand facing away.
Stop (2): Show a fist.
Accept: Thumb up
Reject: Thumb down
Group: Two hands move together. [H09]
Ungroup: Two hands move apart. [H09]
Move, insert, paste (1): Select target from menu/clipboard, move it to a location to place. [H01-05]
Delete (1): Grasp the target and crush it. [H08]
Copy (1): One hand covers the target and another move target to clipboard area. [H01-05]
HM Open: Swipe out. [H06,08,10-11]
HM Close: Swipe in. [H06,08,10-11]
HM Select: Tap an option on the surface. [H11]
Delete (2): Throw away the target [H01-05]
VM Open: Pull up. [H06,09,10-11]
VM Close: Push down. [H06,09,10-11]
VM Select: Push in on an option. [H11]
Move, insert, paste (2): Select target from menu/clipboard, tap at a location to place. [H10-11]
Cut: Snap index & middle (scissor pose)
Copy (2): Two hands turn away, imitate open a book. [change from H07 to H09]
OM Open: Splay all fingers. [H09]
OM Close: Regroup all fingers. [H09]
OM Select: Tap an option on the surface. [H11]

Piumsomboon et al. INTERACT 13

Gesture Interaction

Mapping from Targets to Gestures

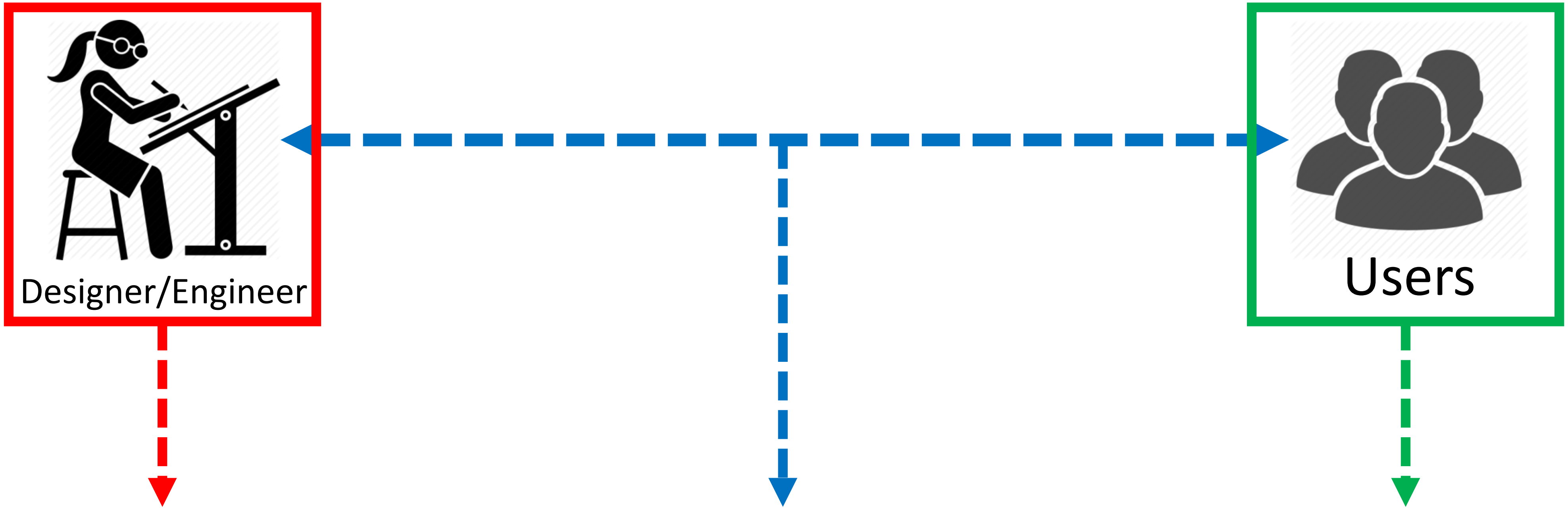
- Simple and easy to understand → **One Simple Metaphor**
- Consistent with acquired experience
- Consensus across different users



Nature	*symbolic*	Gesture visually depicts a symbol.
	physical	Gesture acts physically on objects.
	metaphorical	Gesture indicates a metaphor.
	abstract	Gesture-referent mapping is arbitrary.

RATIONALE

Trade-Off between Mapping and Recognition



Current Gestures

✓ Robust to Recognize
✓ No Conflicts within the Set
✗ Non Self-Revealing to Users

Balanced Gestures

✓ Self-Revealing to Users
✓ Consistent across Users
✓ Robust to Recognize
✓ Large Vocabulary

User Defined Gestures

✓ Intuitive to Users
✓ Consistent across Users
✗ No Concerns of Recognition

RESEARCH QUESTION

Object Retrieval with **VirtualGrasp**

1. **Consistency**: Can users achieve high agreement on the mappings between the objects and their grasping gestures?

2. **Recognition:** Can grasping gestures of different objects be correctly distinguished by algorithms?

3. **Self-Revealing:** Can users discover the object-gesture mappings themselves? If not, can they learn and remember them easily?

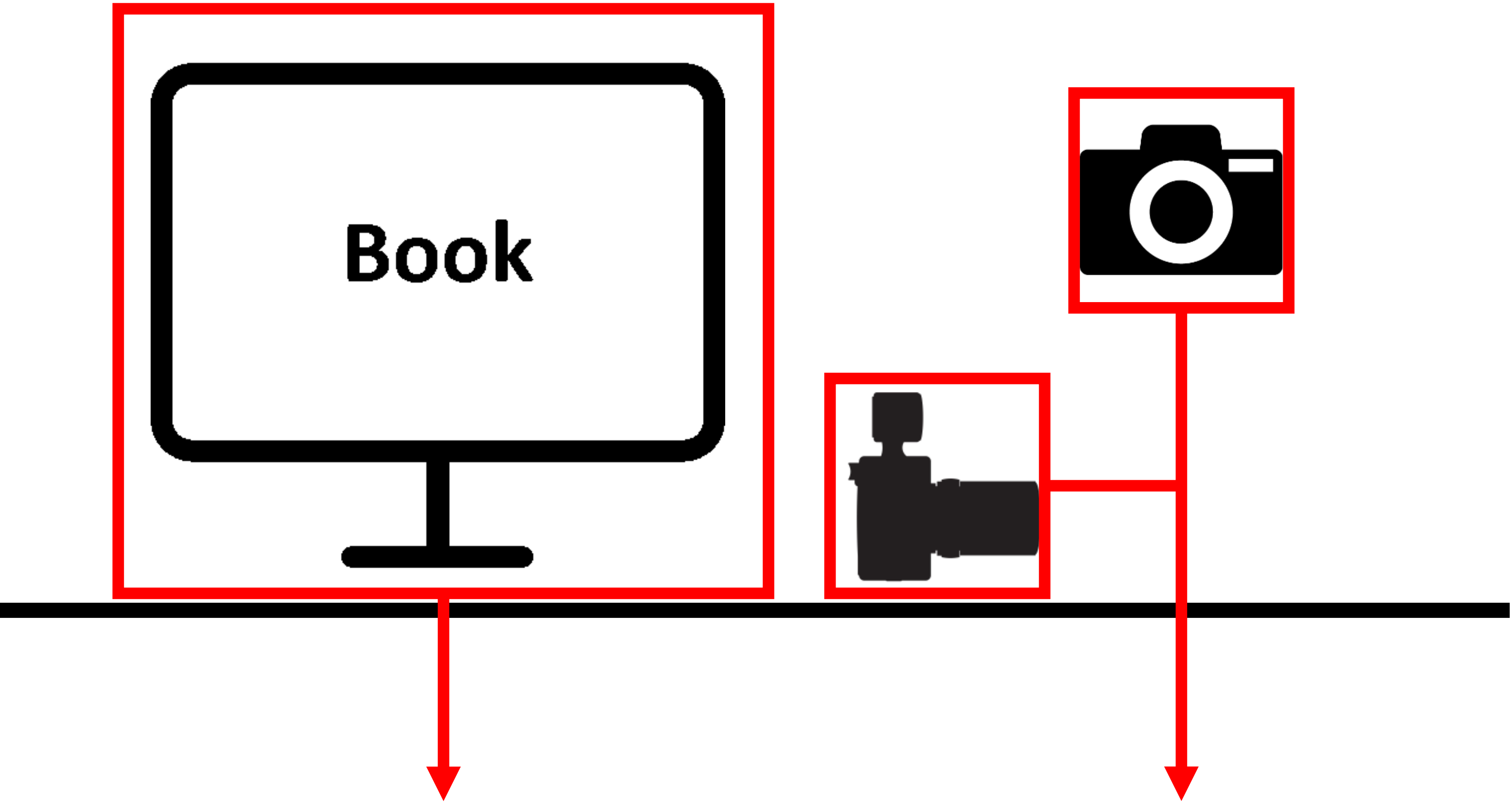
OUTLINE

- User Study: Gesture Elicitation -> **Consistency**
- Experiment: Gesture Recognition -> **Recognition**
- User Study: Object Retrieval -> **Self-Revealing**
- Summary
- Discussion

OUTLINE

- User Study: Gesture Elicitation -> **Consistency**
- Experiment: Gesture Recognition -> **Recognition**
- User Study: Object Retrieval -> **Self-Revealing**
- Summary
- Discussion

USER STUDY: Gesture Elicitation



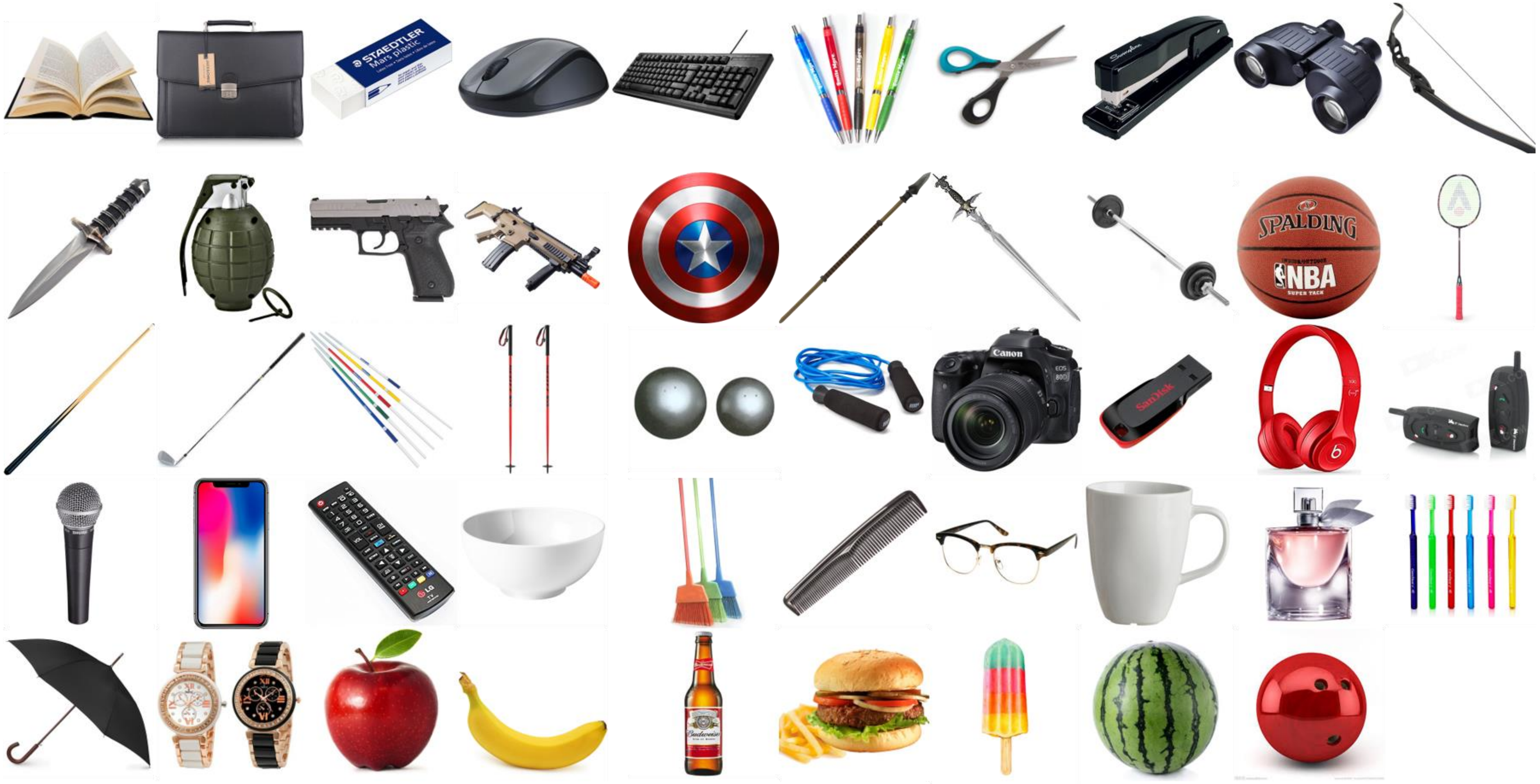
Names of **49** different objects were shown on the front screen.

Two cameras recorded the gestures from the front and the side view.

We recruited 20 participants (14M/6F) to perform the grasping gestures for each object.

USER STUDY: Gesture Elicitation

Object Set (49 Objects)

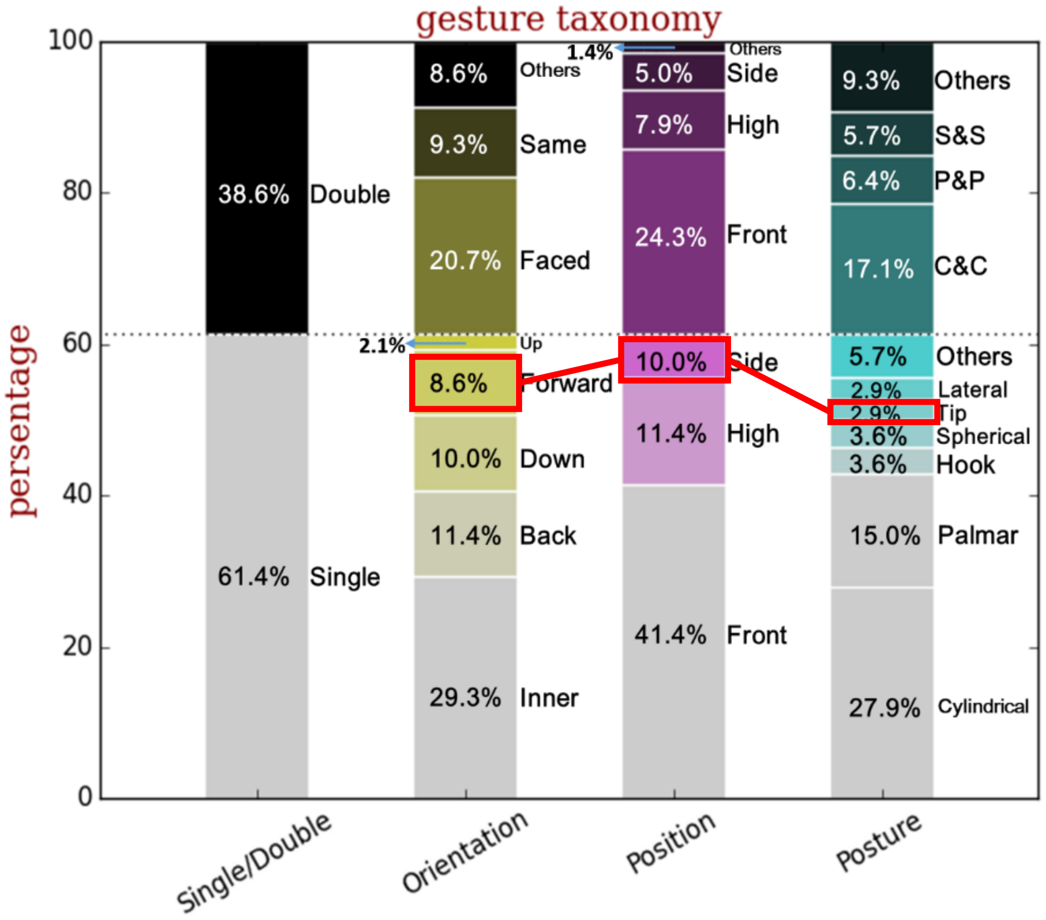


USER STUDY: Gesture Elicitation

- **Consensus across Users**
  - $20 \times 49 =$ 980 gestures, 140 gesture-object pairs.
  - 18/49 objects mapped to one unique gesture
  - 49/49 objects mapped to no more than five gestures
  - Agreement Score: AVG = 0.68, SD = 0.27
- **Key Properties of Objects**
  - Shapes:41.3/49 Usages: 40.8/49 Sizes: 29.8/49

USER STUDY: Gesture Elicitation

- **Breakdown and Distribution of the Gestures**
  - The taxonomy: "Single/Double Hands", "Hand Position", "Palm Orientation" and "Hand Shape“.
  - One-to-one V.S. N-to-one mapping.
  - Infrequent gestures to be leveraged.



USER STUDY: Gesture Elicitation

Discussion

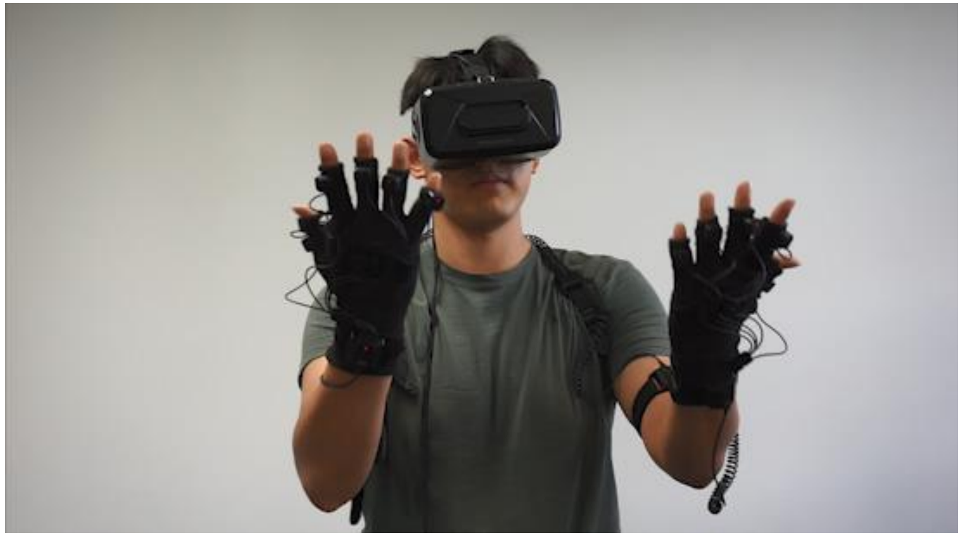
- Half Open-Ended Elicitation Study
  - The power of the metaphor: high consistency across users.
  - The using experience of the objects are required (39/980).

OUTLINE

- User Study: Gesture Elicitation -> **Consistency**
- Experiment: Gesture Recognition -> **Recognition**
- User Study: Object Retrieval -> **Self-Revealing**
- Summary
- Discussion

EXPERIMENT: Gesture Recognition

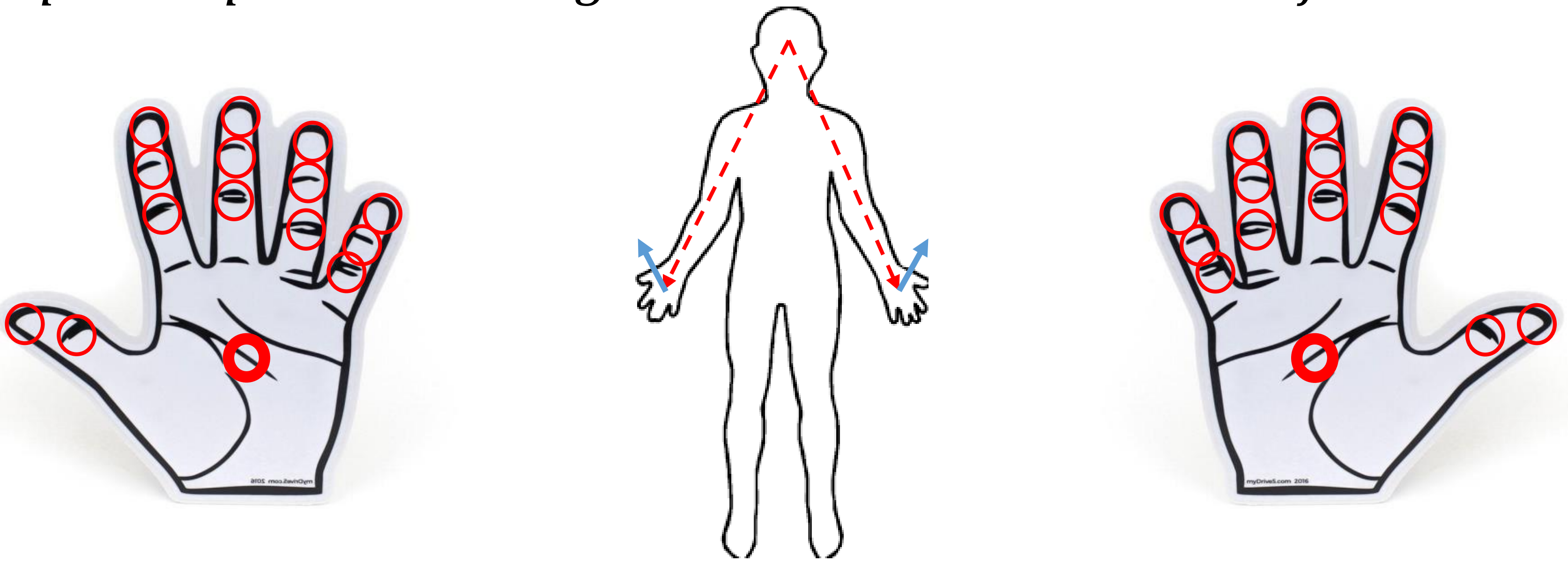
- **Participants**
  - **12** participants, with an average of 24.3 (SD = 1.5). Four of them had experience of mid air gesture interaction. All were familiar with touchscreen gesture interaction.
- **Apparatus**
  - Perception Neuron, which was a MEMS (Micro-Electro-Mechanical System) based tracking device, with a resolution of **0.02** degrees.



The sensors that participants put on

EXPERIMENT: Gesture Recognition

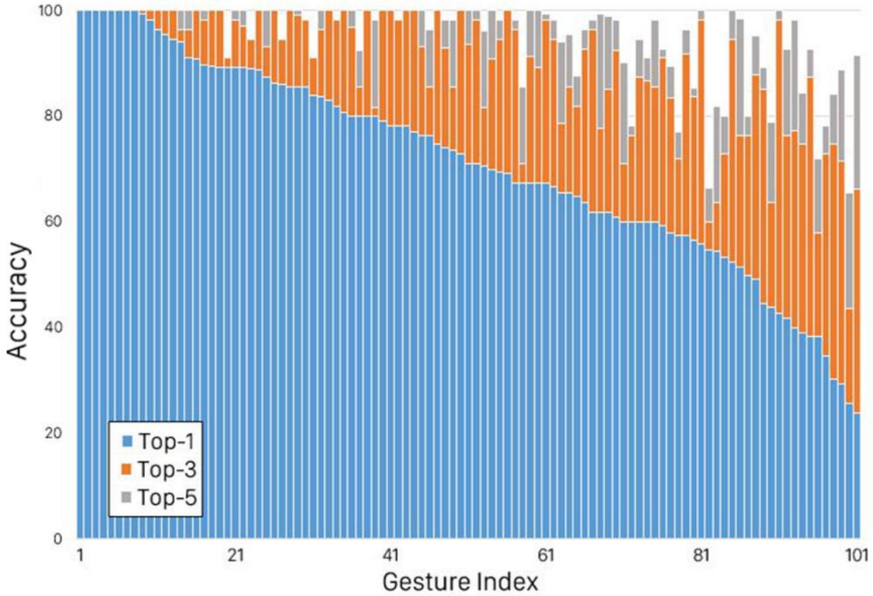
- **Data Collection**
  - The positions and orientations of the hand palms relative to the head.
  - The positions of the 14 joints relative to the hand palms.
  - 40 frames for each gesture that participants performed.
  - $2\ hands\ \times 16\ vectors\ \times 3\ values = 96\ values\ per\ frame$
  - $12\ participants\ \times 101\ gestures\ \times 2\ rounds\ \times 40\ frames = 96960\ frames$



EXPERIMENT: Gesture Recognition

- **Leave-Two-Out Validation**
  - Data of two participants as test set and the left as training set. ($C_{12}^{2} = 66\ rounds$)
  - Top-N accuracy: N most possible objects contain the target. (Top-1, Top-3, Top-5)
  - Average accuracy:

	Top-1	Top-3	Top-5
Mean	70.96%	89.65%	95.05%
SD	9.25%	6.39%	4.56%

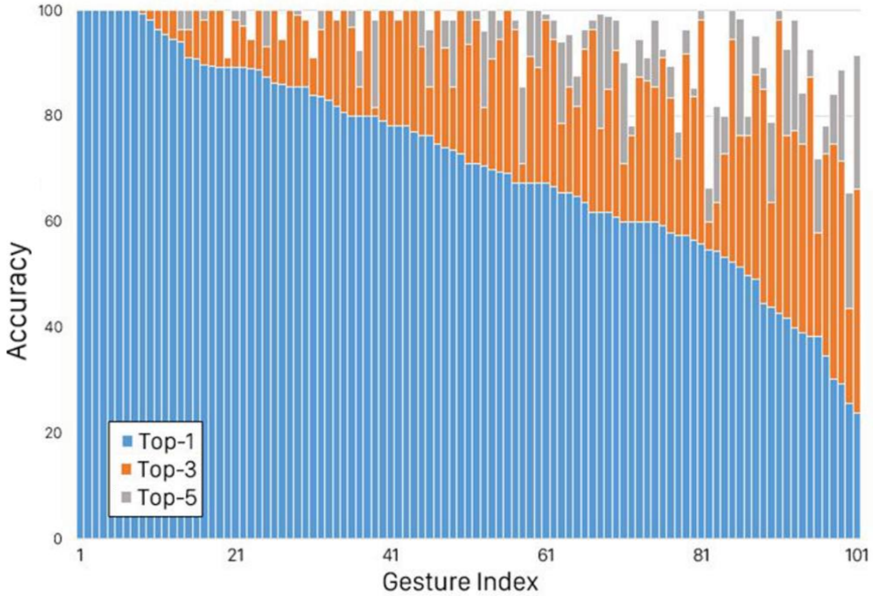


Too small objects

EXPERIMENT: Gesture Recognition

- **Leave-Two-Out Validation**
  - Data of two participants as test set and the left as training set. $(C_{12}^{2} = 66\, rounds)$
  - Top-N accuracy: N most possible objects contain the target. (Top-1, Top-3, Top-5)
  - Average accuracy:

	Top-1	Top-3	Top-5
Mean	70.96%	89.65%	95.05%
SD	9.25%	6.39%	4.56%



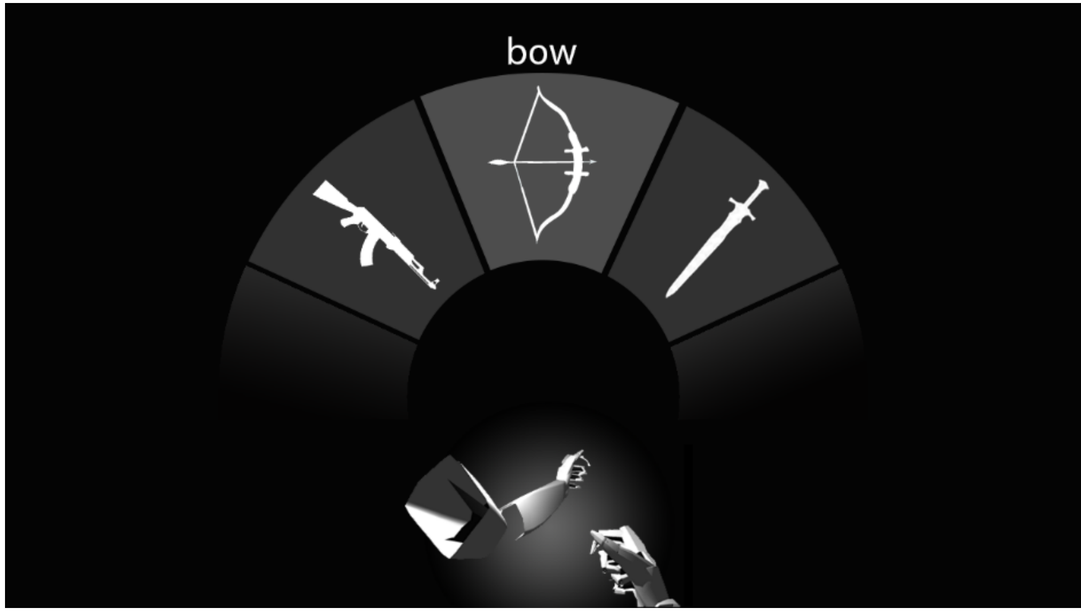
Strong connection to usages

OUTLINE

- User Study: Gesture Elicitation -> **Consistency**
- Experiment: Gesture Recognition -> **Recognition**
- User Study: Object Retrieval -> **Self-Revealing**
- Summary
- Discussion

USER STUDY: Object Retrieval

- **Participants**
  - 12 new participants, who never participated in STUDY1 or STUDY2.
- **Apparatus**
  - We showed the name of the target object on the top, visualized the current gesture of the participants, and showed the recognition result of top three possible objects in the center.



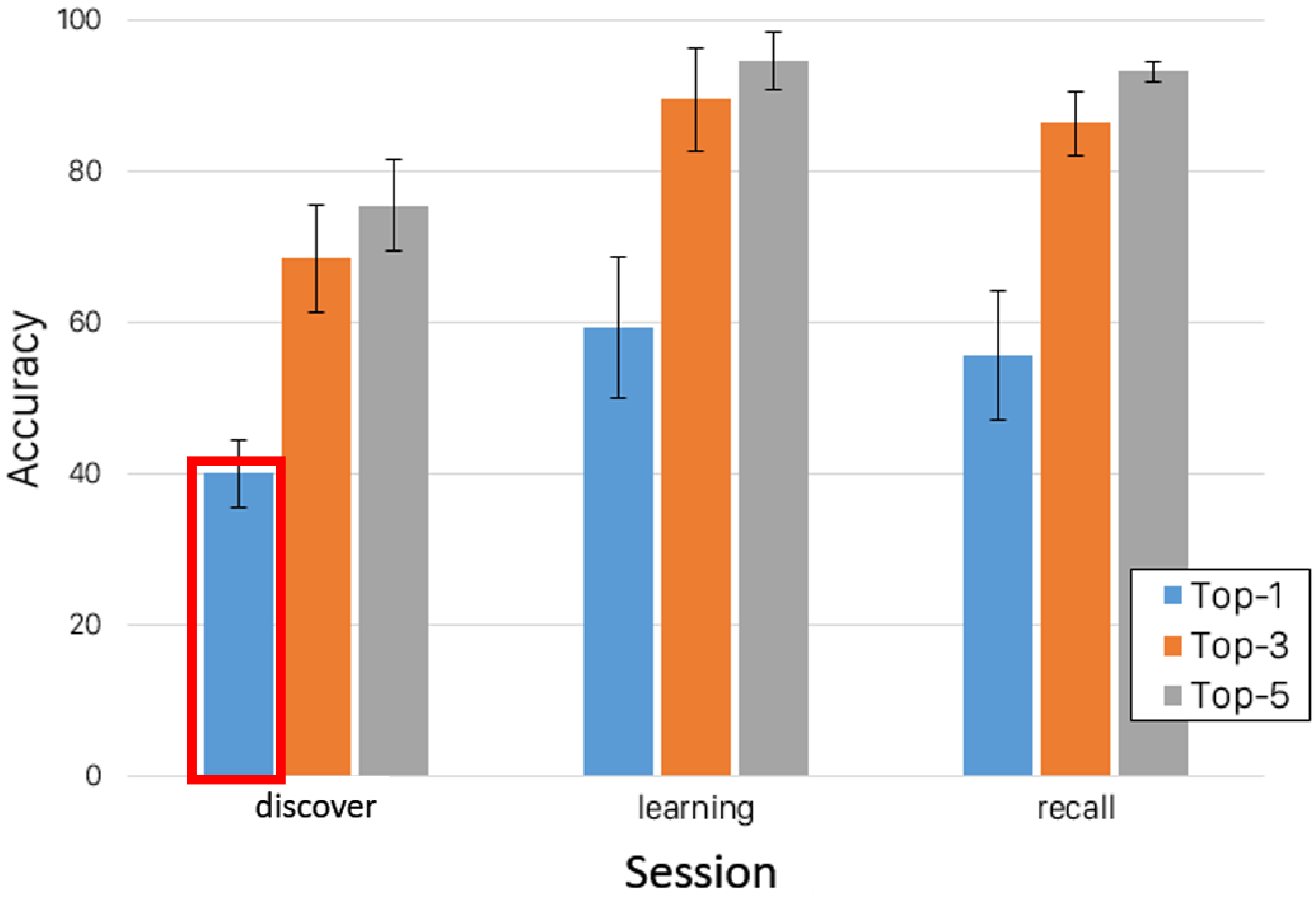
User Interface

USER STUDY: Object Retrieval

- **Discovery Session**
  - ***Without learning*** the gesture-object mappings in the system, we asked participants to perform their own grasping gestures.
- **Learning Session**
  - Before test, we let participants learn the standard gestures. They were free to ***practice*** the gestures until they confirm to be ready.
- **Recall Session**
  - ***A week later***, participants came back to lab and perform 49 object retrieval tasks again. During the week, they were not exposed to the standard gestures again.

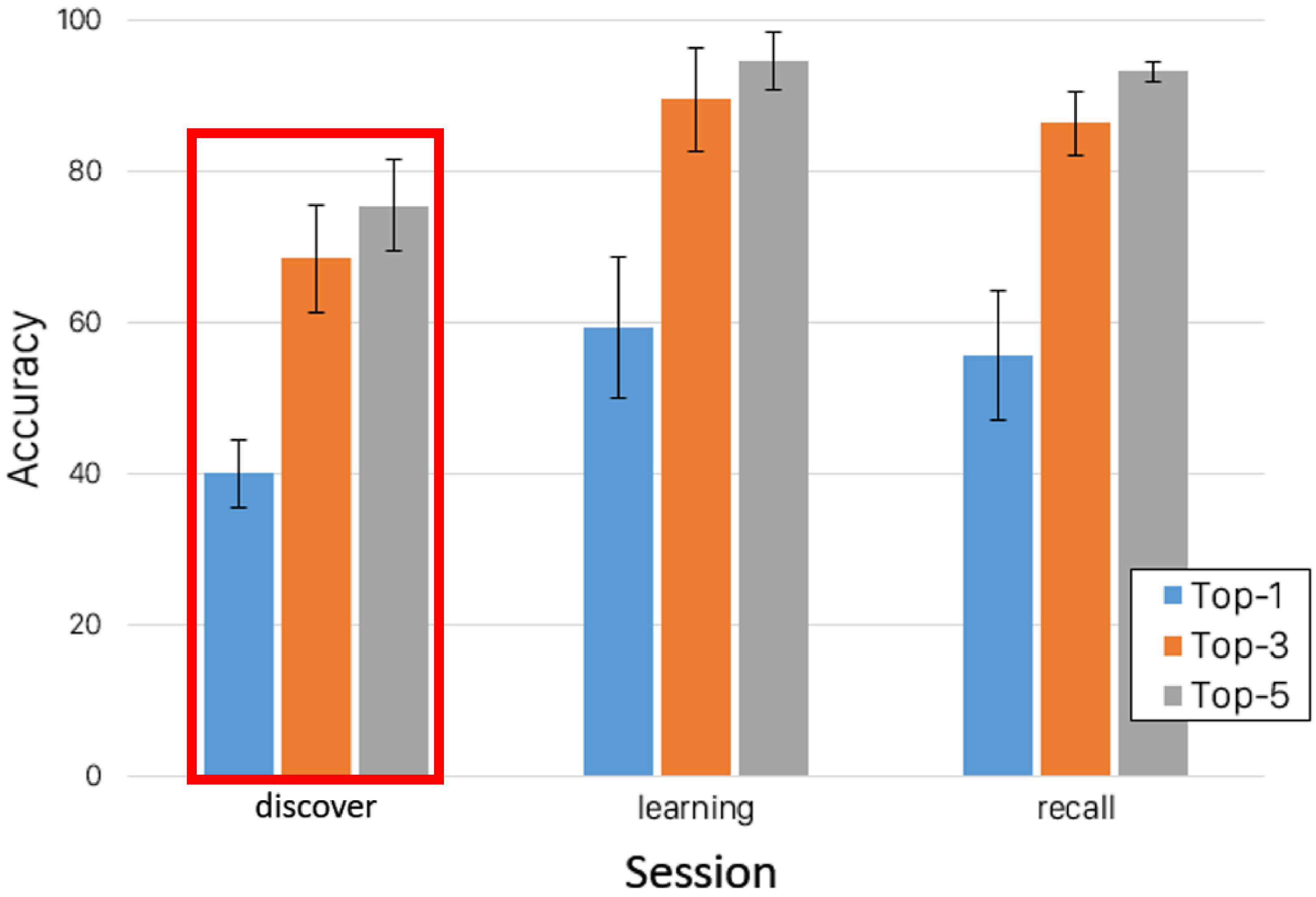
STUDY3: Object Retrieval

40% of the gestures were triggered without training

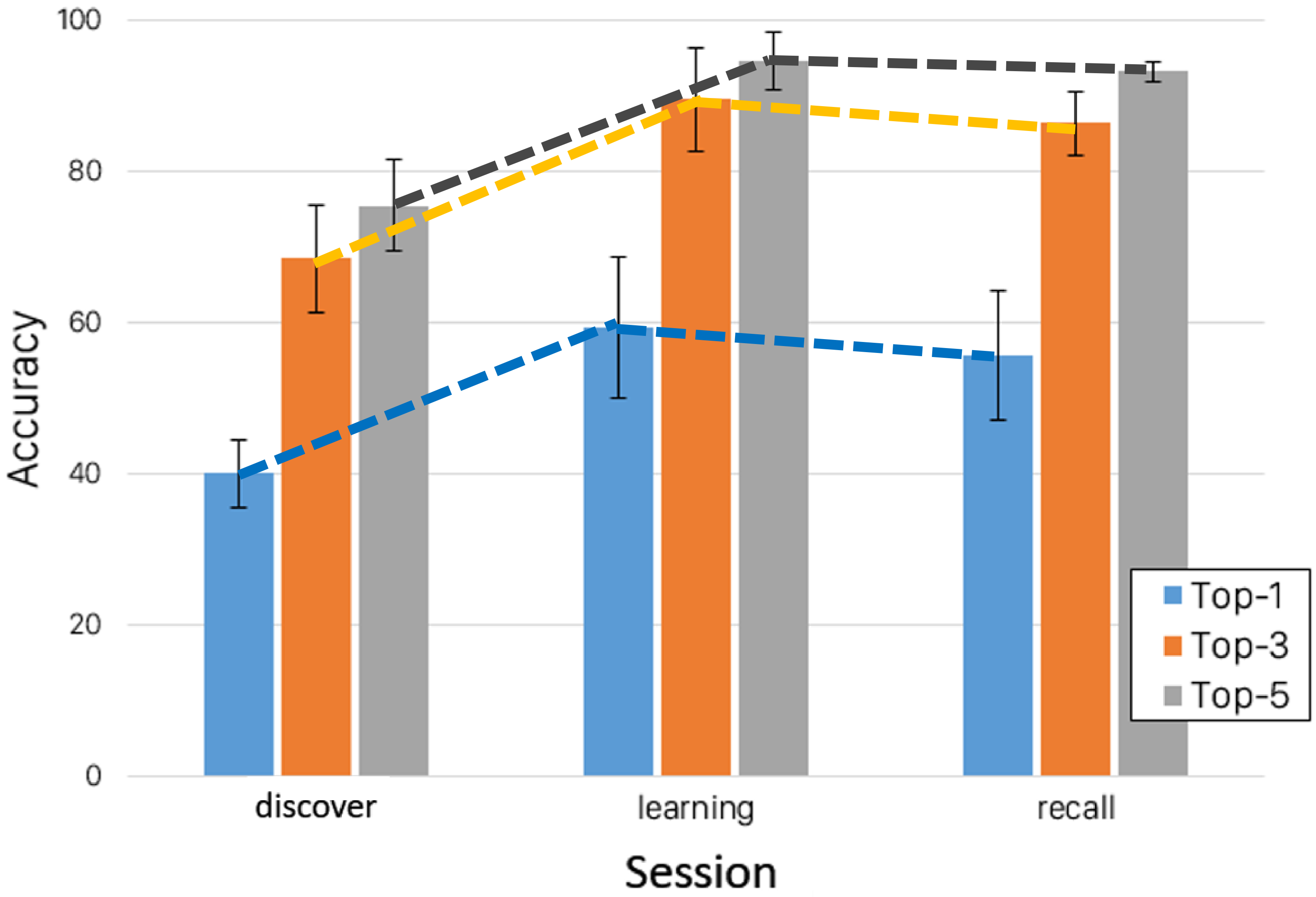


STUDY3: Object Retrieval

76% of the objects were successfully retrieved



STUDY3: Object Retrieval



STUDY3: Object Retrieval

- **Discoverability**: Without any training, participants could discover **40%** of the exact mappings by themselves, and could directly use VirtualGrasp to retrieve **76%** of the objects with top five candidates.

- **Memorability**: A week after the learning session, participants could still recall the mappings well, and could successfully retrieve **93%** of the objects with top five candidates.

STUDY3: Object Retrieval

Subjective Feedback

- **The system is intelligent**
- *"Two different gestures came to me for grasping the camera and it was intelligent that the system correctly recognized the one I performed." [P4]*
- **The gestures make sense**
- *"I never used a grenade before, but I agreed with Gesture 3 which was grasping it over the shoulder to throw it.“ [P6]*
- **New tricks under the concept**
- *“For 'Stapler', I chose to perform the gesture of pressing it instead of holding it, because few other objects require pressing." [P8]*

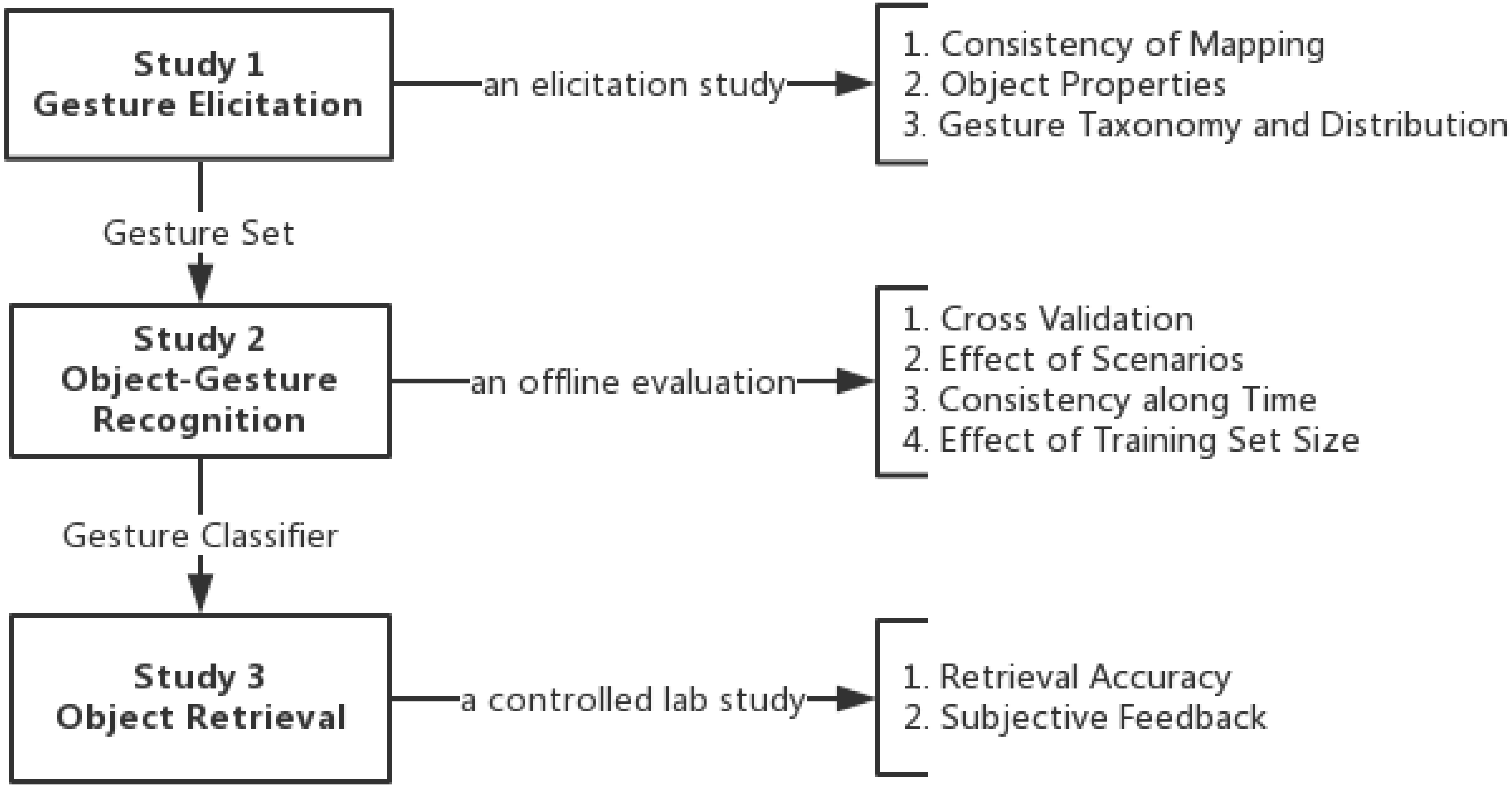
5-Point Scale	Discoverability	Fatigue	Memorability	Fun
Mean	4.2	4.4	4.5	4.4
SD	0.78	0.70	0.53	0.52

SUMMARY

High Consistency

Good Accuracy

Little Effort



Study 1
Gesture Elicitation

an elicitation study

1. Consistency of Mapping
2. Object Properties
3. Gesture Taxonomy and Distribution

Gesture Set

Study 2
Object-Gesture Recognition

an offline evaluation

1. Cross Validation
2. Effect of Scenarios
3. Consistency along Time
4. Effect of Training Set Size

Gesture Classifier

Study 3
Object Retrieval

a controlled lab study

1. Retrieval Accuracy
2. Subjective Feedback

DISCUSSION

- **Object-Gesture Mappings**
  - Objects with different property values.
    - Not from objects of the same type.

DISCUSSION

- **Object-Gesture Mappings**
  - Objects with different property values.
    - Not from objects of the same type.
  - Grasping gestures reflect different properties of objects.
    - Difficult to distinguish grasping gestures of too small objects.

DISCUSSION

- **Object-Gesture Mappings**
  - Objects with different property values.
    - Not from objects of the same type.
  - Grasping gestures reflect different properties of objects.
    - Difficult to distinguish grasping gestures of too small objects.
- **Sensing Technique**
  - Hand gesture, hand position and hand orientation.
    - Vision-based sensing techniques.
  - Hand gesture.
    - Data gloves, EMG sensors, Vision-based.
  - Hand position and hand orientation.
    - VR controllers.

Thanks