MARCH, **2024**

Mixed Reality Multi-user Asymmetric Telecollaboration

PHD ORAL DEFENCE

Faisal Zaman

Supervisors

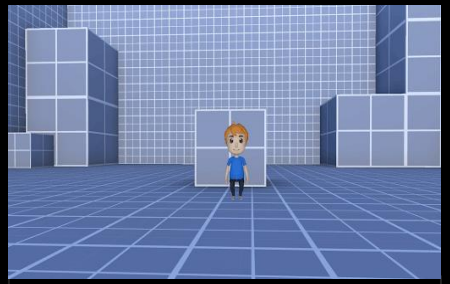
A/Prof Taehyun Rhee

Dr Craig Anslow

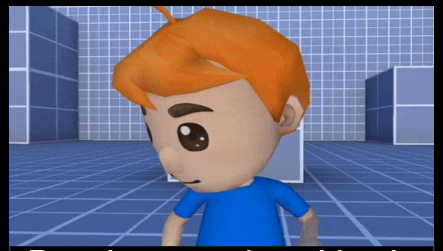
CMIC | Computational Media Innovation Centre

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

Collaborative Tools



Limited Field of view



Barely see others' body language & gaze

Fixed viewpoint

Collaborative XR Tools

Project starline, *Lawrence et al. 2021*

Microsoft Remote Assist

Spatial.io

Microsoft Mesh

Research Problem

totally disconnected from the real world's physical space.

face tracking only work for one person at a time and setup is quite expensive.

remote users have limited physical presence and interaction.

users collaborate in a shared virtual space not in physical space.

shared physical space with sense of presence and tools to assists their task

Collaborate in Asymmetric Physical Space with VR and AR Users

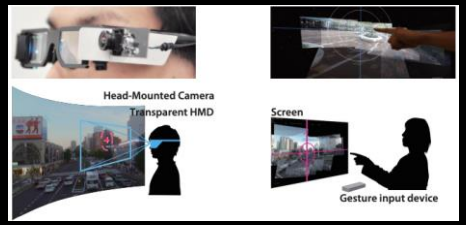
Remote Users

- VR systems are ideal for Strong storytelling
- Provide a fully immersive/3D experience
- Easily change the scale and representation of the virtual body

Local Users

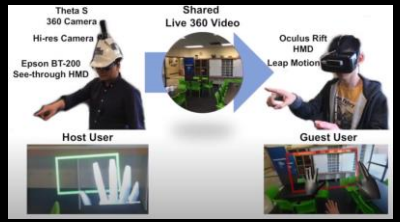
- AR systems are ideal for live sharing
- Allow real world overlay/can share viewpoints
- Remote annotation/link support
- Improve the real-world task

Related Work

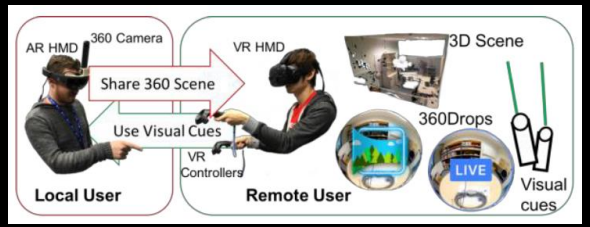


[S. Kasahara and J. Rekimoto 2014]

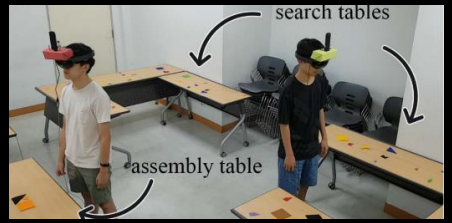
[S. Nagai et al. 2015]



[Lee et al. 2017]



[Teo et al. 2017]



[Lee et al. 2020]

[Rhee et al. 2020]

- Only One-One & One-Many
- No Audio-Visual Sync
- No Multi-Viewpoints Management

Research Questions

How can multiple remote (VR) users effectively connect and collaborate with local (AR) users in the Mixed Reality Collaboration space?

teleportation

RQ1 How can **multiple remote users have a high sense of presence** while maintaining spatial awareness and understanding in the Mixed Reality Collaboration space?

representation

RQ2 How can **multiple remote users be effectively represented** in the Mixed Reality Collaboration space to achieve **co-presence**?

interaction

RQ3 How can **multiple users interact effectively** in the Mixed Reality Collaboration space?

viewpoint perspective

RQ4 How can users' **viewpoints be managed effectively** to support various types of collaborative tasks in the Mixed Reality Collaboration space?

multi-perspective collaboration

RQ5 How can **multi-perspective collaboration be effectively** achieved for supporting collaborative tasks?

Contributions

1. Mixed Reality Multi-user Asymmetric Collaboration (**MRMAC**) system

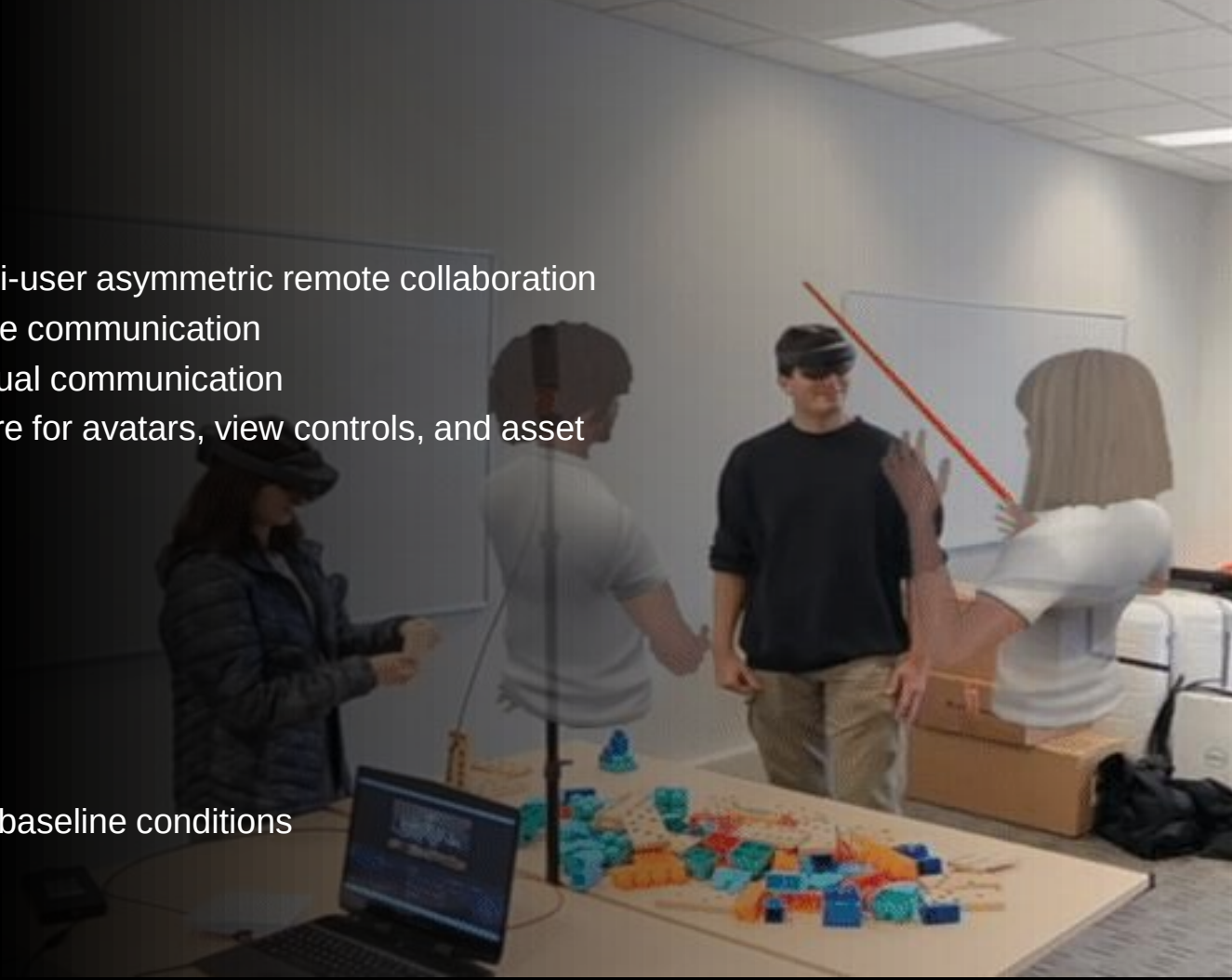
MRMAC

Design & Implementation

- Design concept for multi-user asymmetric remote collaboration
- Bidirectional face-to-face communication
- Synchronised audio-visual communication
- Client-server architecture for avatars, view controls, and asset streaming

Evaluation

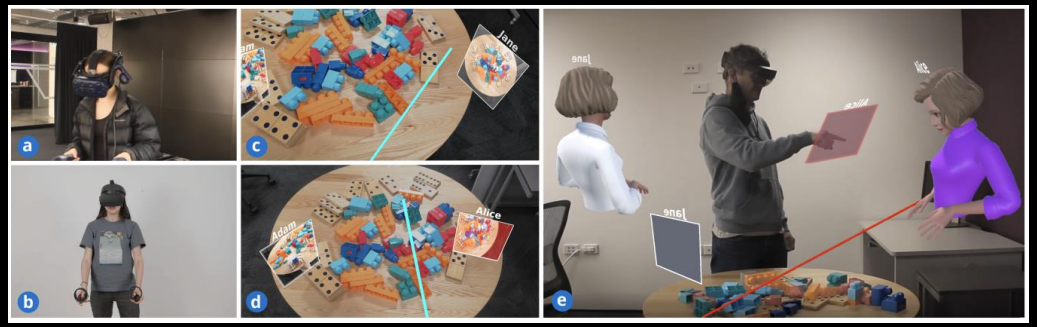
- System evaluation
  - scalability and latency
- User evaluation
  - impact of user roles
  - compared against two baseline conditions



Contributions

1. Mixed Reality Multi-user Asymmetric Collaboration (**MRMAC**) system

2. **Vicarious**, a context-aware viewpoint selection method



Vicarious

Design & Implementation

- Context-aware viewpoint selection method based on visual saliency, user actions, and speech patterns.

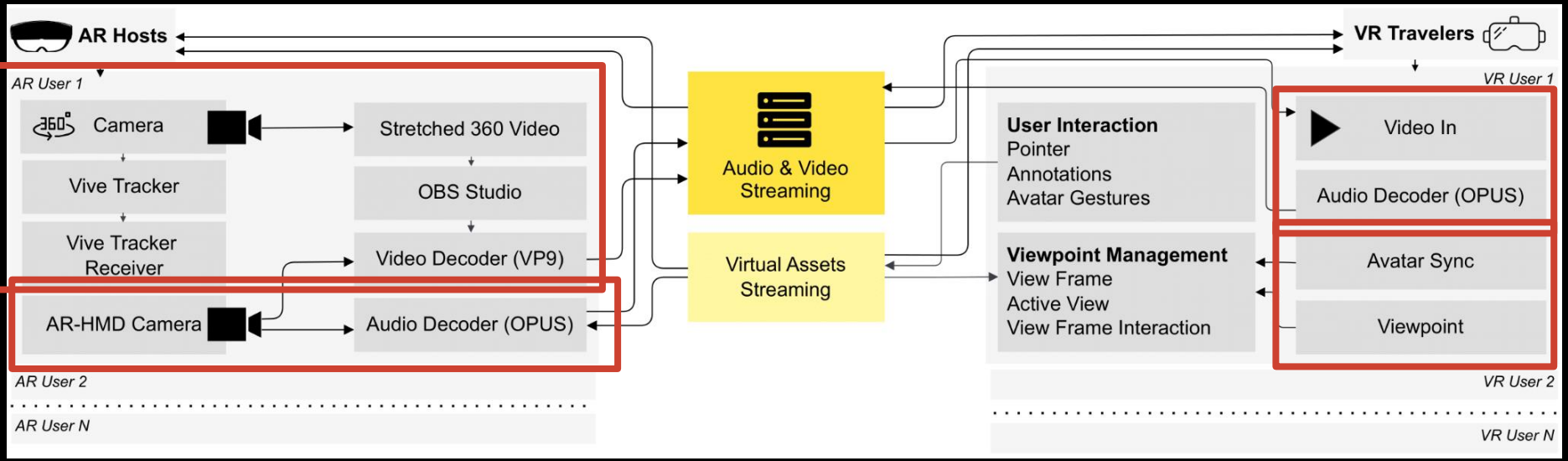
- Various visualization and interaction techniques

User Study

- Evaluated the impact of viewpoint selections method under four distinct conditions

- Recommendations for design implications and direction for future research

MRMAC Architecture



Communication Cues

Viewpoint Sharing

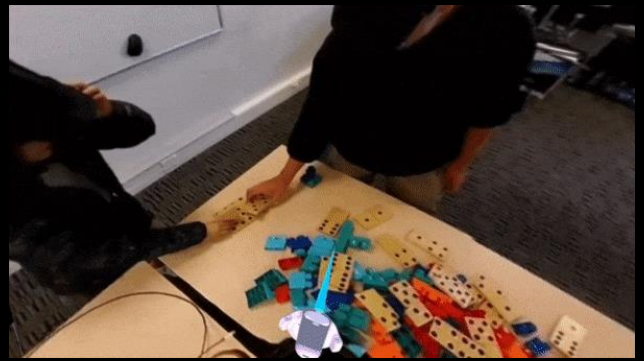
Multiple Avatar Control

Personalised Avatar

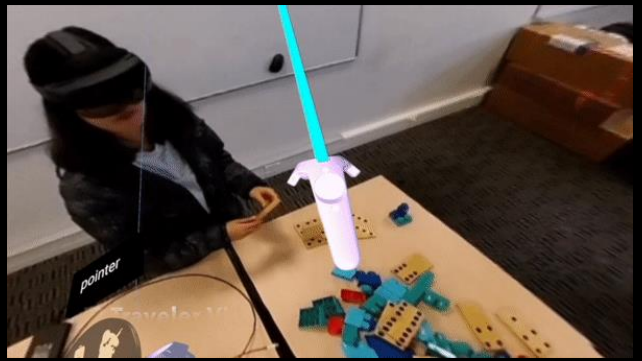
Spatial Audio

Interaction Cues

3D Pointer



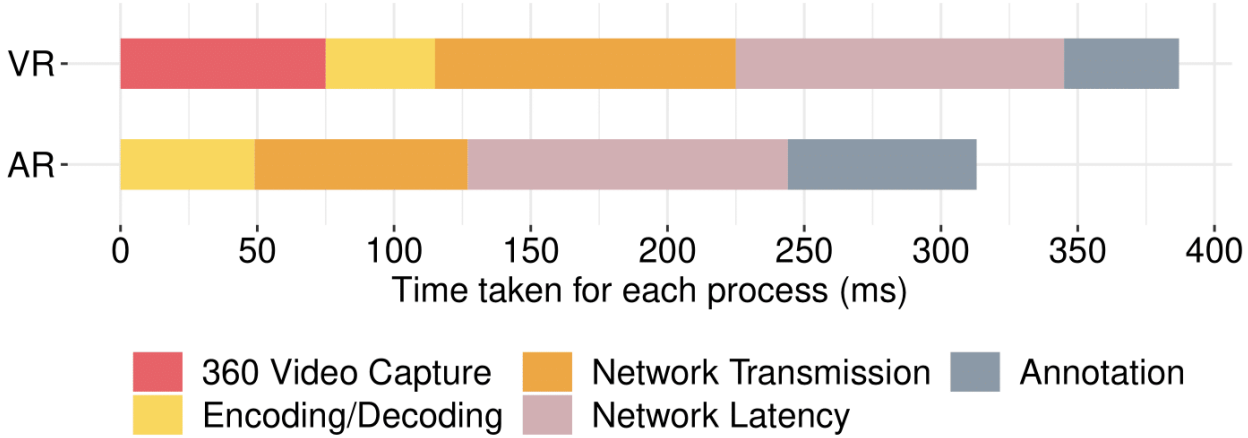
Annotation



System Evaluation: Latency

Average end-to-end latency

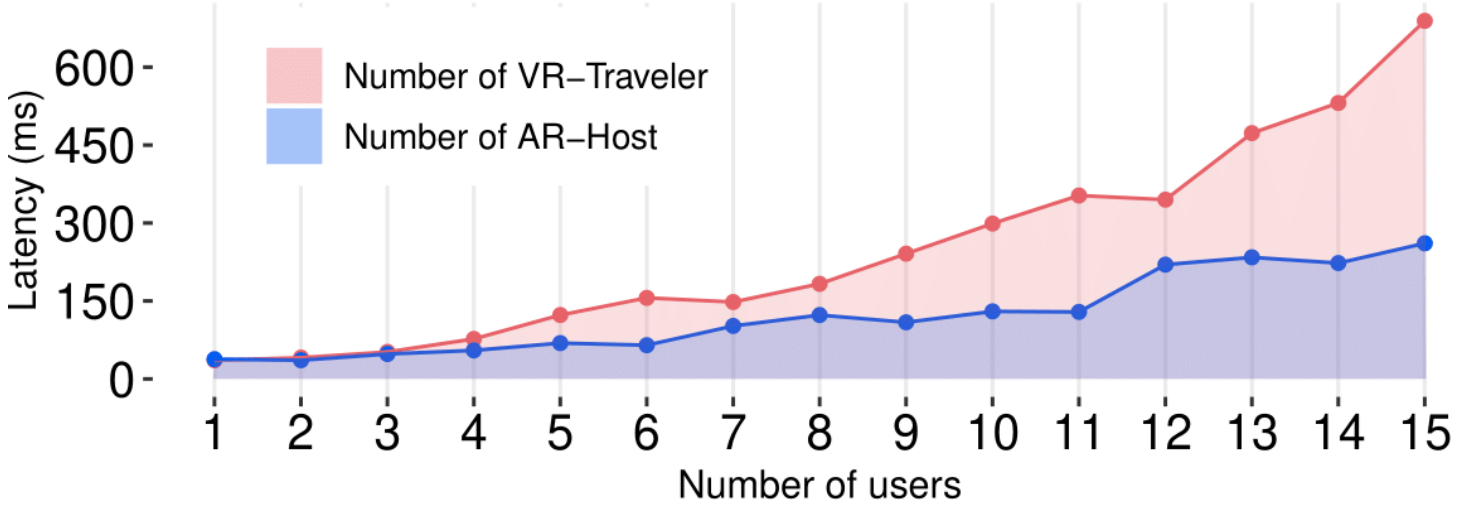
- Changes in the 360-degree video to show up on the local computer and in the VR/AR window.



System Evaluation: Scalability

Network latency as the number of users increases

- average frame rate = video streaming 30FPS
- audio streaming 44.1kHz
- rendering time of 60±10 FPS.
- wireless bandwidth capacity (AX3000 Wi-Fi 6 router) 3 Gbps



User Evaluation

36 participants

18—81 years
μ = 30.83, σ = 14.10

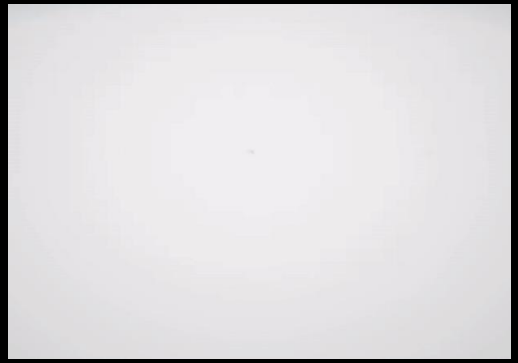
3×2 mixed factorial design

between-subjects factor:
local vs. remote

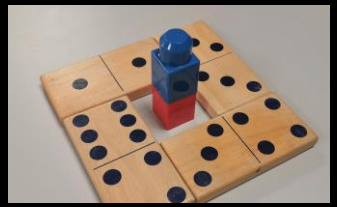
within-subjects factor:
C1 vs. C2 vs. C3

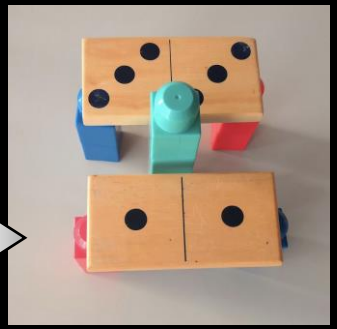
VR Users

AR Users



Virtual replica





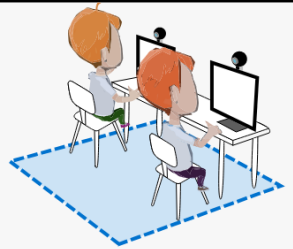
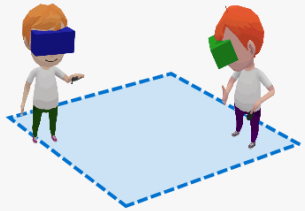
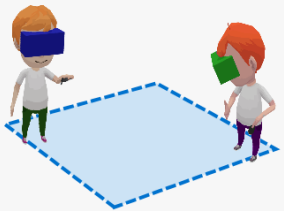
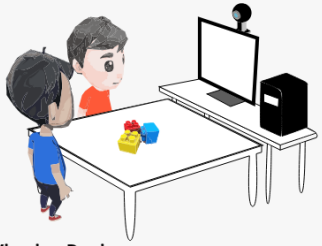
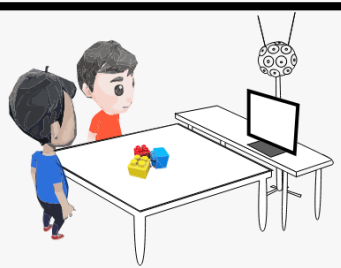
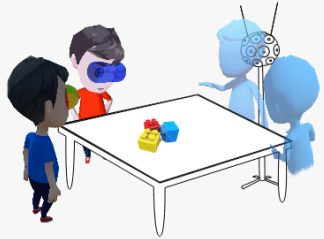
Physical model

User Evaluation

3×2 mixed factorial design

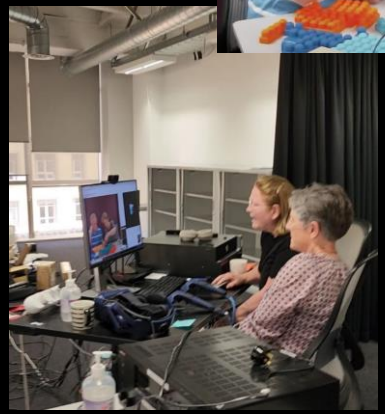
between-subjects factor:
local vs. remote

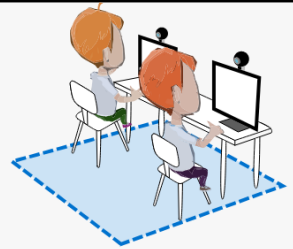
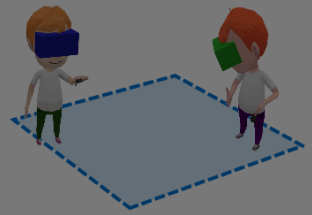
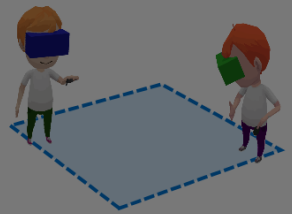
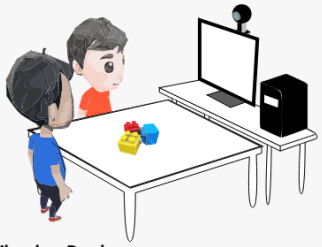
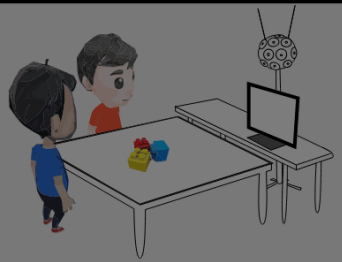
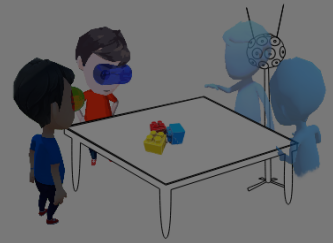
within-subjects factor:
C1 vs. C2 vs. C3

	1 Condition 1 Conventional video with 2D annotation	2 Condition 2 360 Video without augmented visual cues	3 Condition 3 360 video with augmented visual cues
VR-Travelers	 **Viewing Device:** Desktop Monitor **Camera control:** N/A **Interaction:** Mouse for 2D Annotation **Representation:** Appears in 2D video	 VR HMD (Viewing 360 Video) Camera Control by VR HMD N/A No Visual, Voice Only	 VR HMD (Viewing 360 Video) Camera Control by VR HMD VR Controller 3D Avatar
AR-Hosts	 **Viewing Device:** Desktop Monitor **Camera control:** N/A **Interaction:** N/A **Representation:** Appears in 2D video	 Desktop Monitor N/A N/A Appears in 360-degree Video	 AR HMD N/A N/A Appears in 360-degree Video

User Evaluation

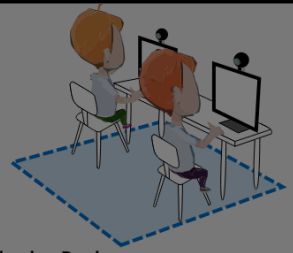
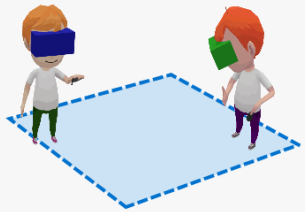
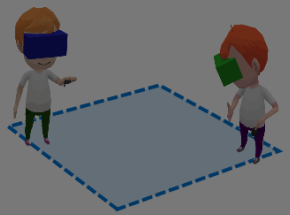
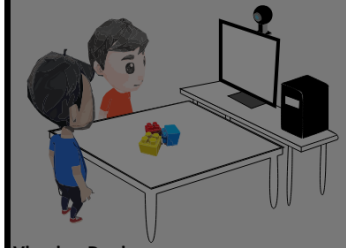
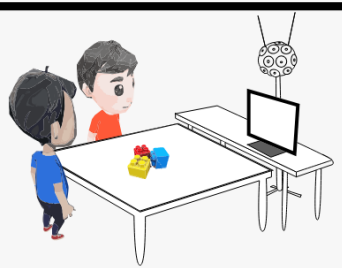
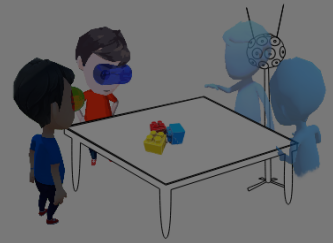
Condition 1



	1 Condition 1 **Conventional video with 2D annotation**	2 Condition 2 **360 Video without augmented visual cues**	3 Condition 3 **360 video with augmented visual cues**
VR-Travelers	 **Viewing Device:** Desktop Monitor **Camera control:** N/A **Interaction:** Mouse for 2D Annotation **Representation:** Appears in 2D video	 VR HMD (Viewing 360 Video) Camera Control by VR HMD N/A No Visual, Voice Only	 VR HMD (Viewing 360 Video) Camera Control by VR HMD VR Controller 3D Avatar
AR-Hosts	 **Viewing Device:** Desktop Monitor **Camera control:** N/A **Interaction:** N/A **Representation:** Appears in 2D video	 Desktop Monitor N/A N/A Appears in 360-degree Video	 AR HMD N/A N/A Appears in 360-degree Video

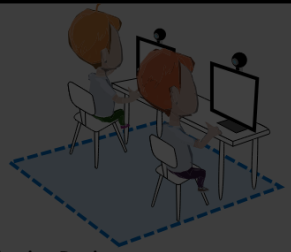
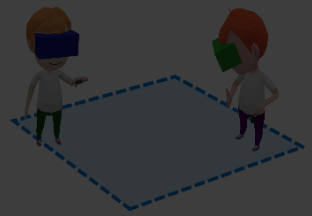
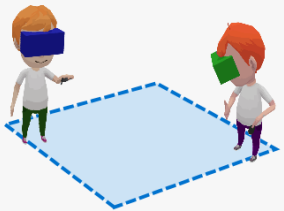
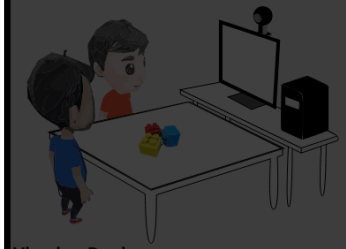
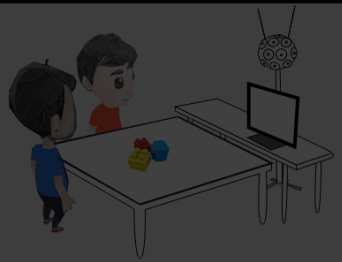
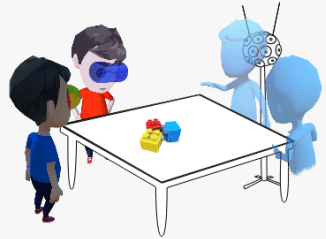
User Evaluation

Condition 2

	1 Condition 1 Conventional video with 2D annotation	2 Condition 2 360 Video without augmented visual cues	3 Condition 3 360 video with augmented visual cues
VR-Travelers			
	Viewing Device: Desktop Monitor **Camera control:** N/A **Interaction:** Mouse for 2D Annotation **Representation:** Appears in 2D video	VR HMD (Viewing 360 Video) Camera Control by VR HMD N/A No Visual, Voice Only	VR HMD (Viewing 360 Video) Camera Control by VR HMD VR Controller 3D Avatar
AR-Hosts			
	Viewing Device: Desktop Monitor **Camera control:** N/A **Interaction:** N/A **Representation:** Appears in 2D video	Desktop Monitor N/A N/A Appears in 360-degree Video	AR HMD N/A N/A Appears in 360-degree Video

User Evaluation

Condition 3

	1 Condition 1 Conventional video with 2D annotation	2 Condition 2 360 Video without augmented visual cues	3 Condition 3 360 video with augmented visual cues
VR-Travelers	 **Viewing Device:** Desktop Monitor **Camera control:** N/A **Interaction:** Mouse for 2D Annotation **Representation:** Appears in 2D video	 VR HMD (Viewing 360 Video) Camera Control by VR HMD N/A No Visual, Voice Only	 VR HMD (Viewing 360 Video) Camera Control by VR HMD VR Controller 3D Avatar
AR-Hosts	 **Viewing Device:** Desktop Monitor **Camera control:** N/A **Interaction:** N/A **Representation:** Appears in 2D video	 Desktop Monitor N/A N/A Appears in 360-degree Video	 AR HMD N/A N/A Appears in 360-degree Video

Results

Statistically Significant

- Spatial Presence
- Social Presence
- Task Workload
- System Usability
- Preference

Not Statistically Significant

- Task Completion Time
- Simulator Sickness

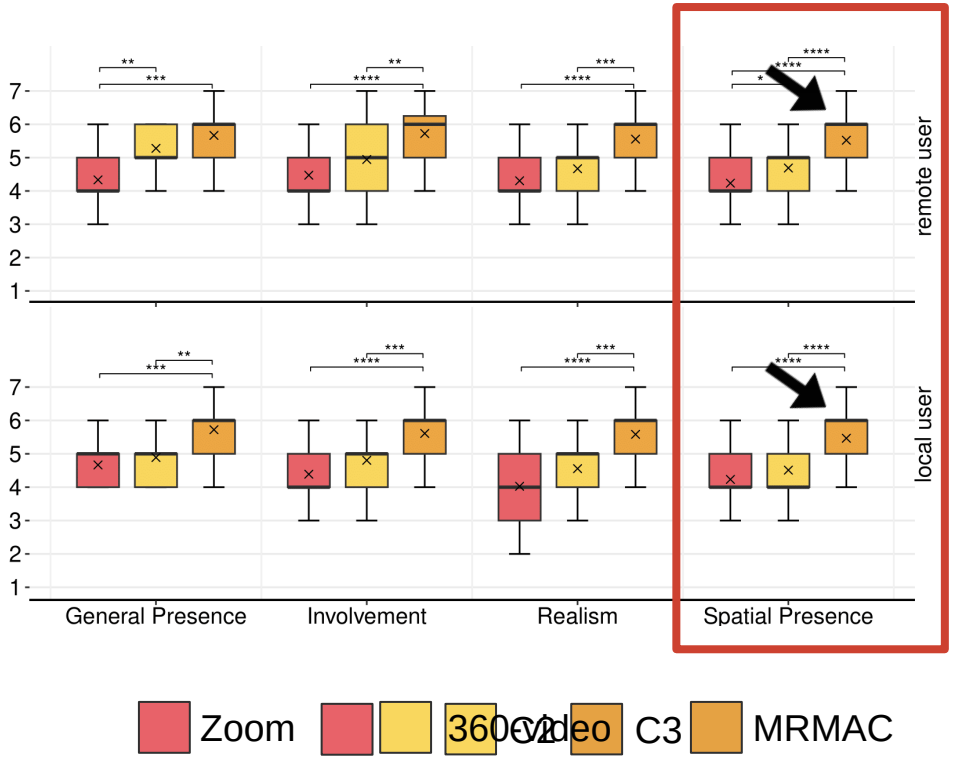
Results

Statistically Significant

- Spatial Presence
- Social Presence
- Task Workload
- System Usability
- Preference

Not Statistically Significant

- Task Completion Time
- Simulator Sickness



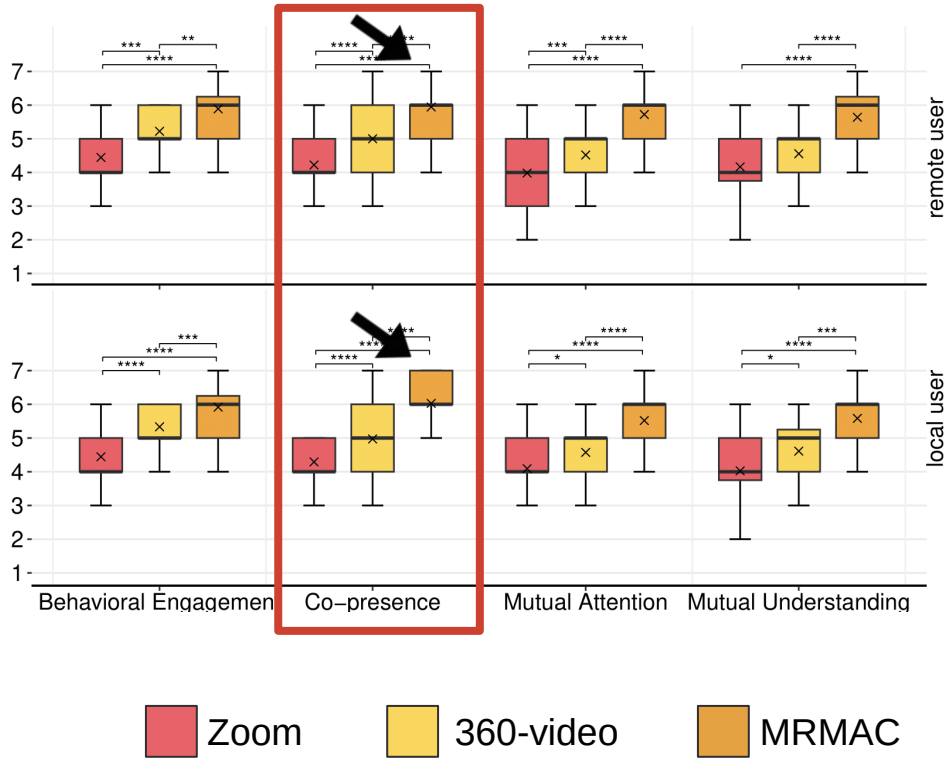
Results

Statistically Significant

- Spatial Presence
- Social Presence
- Task Workload
- System Usability
- Preference

Not Statistically Significant

- Task Completion Time
- Simulator Sickness



Results

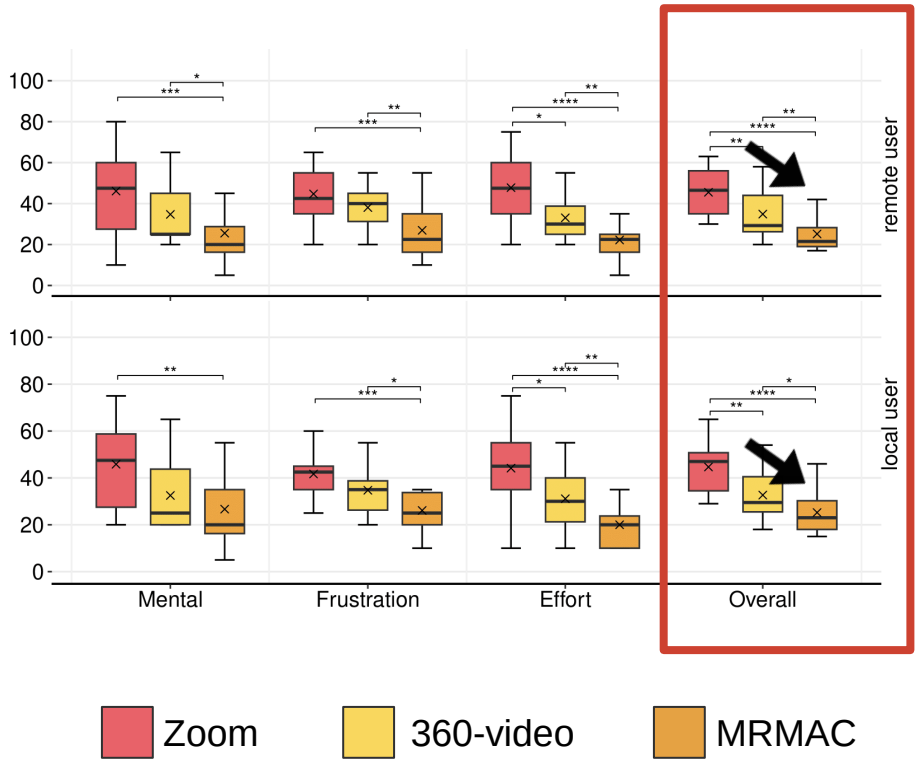
Statistically Significant

- Spatial Presence
- Social Presence
- Task Workload
- System Usability
- Preference

Not Statistically Significant

- Task Completion Time
- Simulator Sickness

μ =1.43 (σ =.77, SEM=.14)



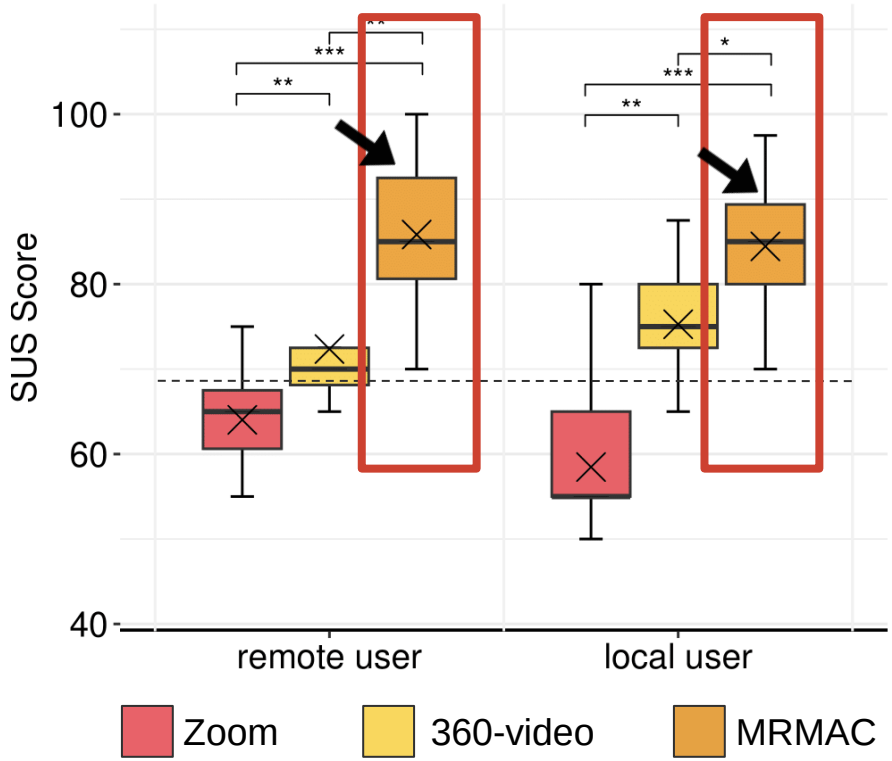
Results

Statistically Significant

- Spatial Presence
- Social Presence
- Task Workload
- System Usability
- Preference

Not Statistically Significant

- Task Completion Time
- Simulator Sickness



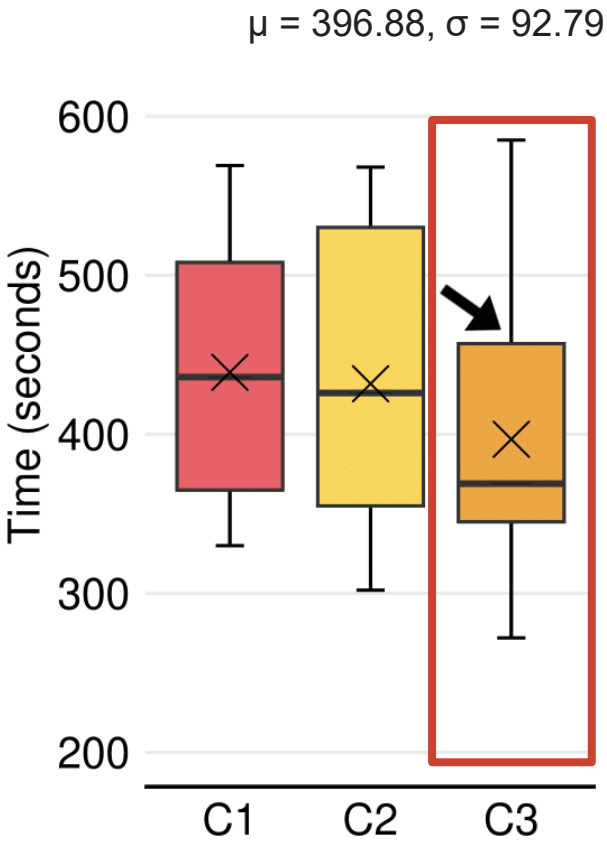
Results

Statistically Significant

- **Spatial Presence**
- **Social Presence**
- **Task Workload**
- **System Usability**
- **Preference**

Not Statistically Significant

- **Task Completion Time**
- **Simulator Sickness**



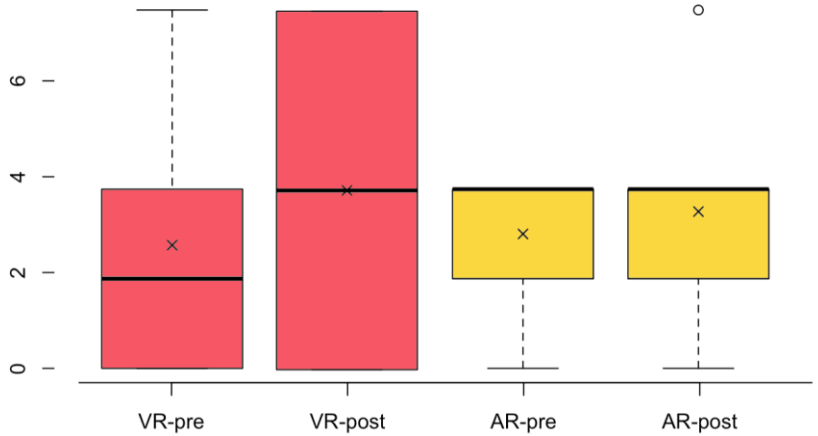
Results

Statistically Significant

- Spatial Presence
- Social Presence
- Task Workload
- System Usability
- Preference

Not Statistically Significant

- Task Completion Time
- Simulator Sickness



Summary of MRMAC

RQ1: teleportation | RQ2: representation | RQ3: interaction

Fully bidirectional asymmetric collaboration system

- Real-time collaboration
- Audio-visual synchronisation
- Virtual assets blending

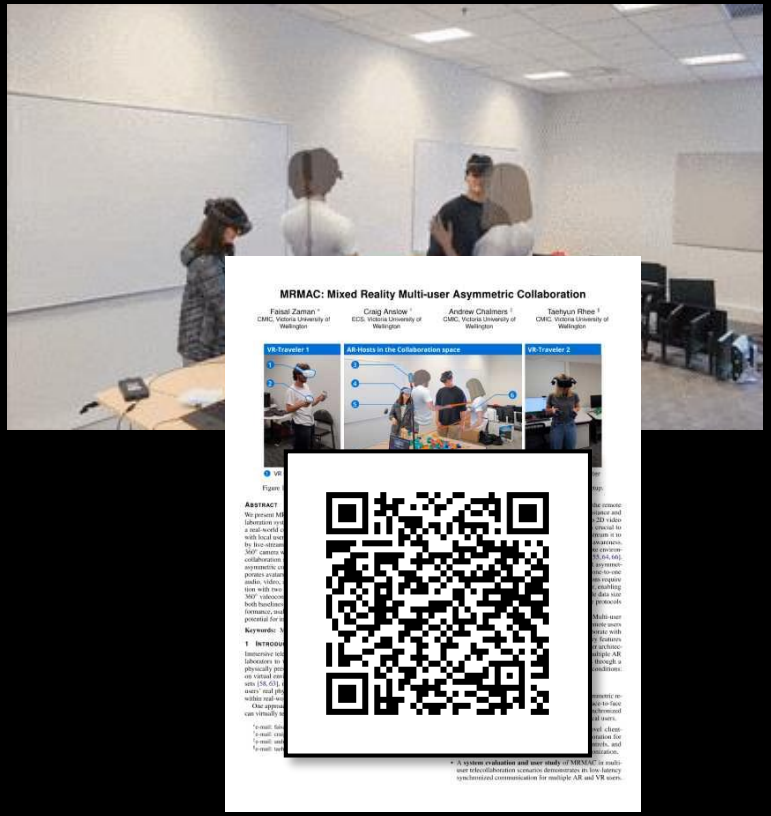
System evaluation

- Reasonably Scalable
- Low-latency

User evaluation

- MRMAC *performed* collaborative tasks faster
- MRMAC *preferred* over conventional 2D and traditional 360

Why travel with hefty backpack when you can teleport with *MRMAC*



F. Zaman, C. Anslow, A. Chalmers, T. Rhee. **MRMAC: Mixed Reality Multiuser Asymmetric Collaboration**. IEEE ISMAR, 2023.

Multiple Viewpoints

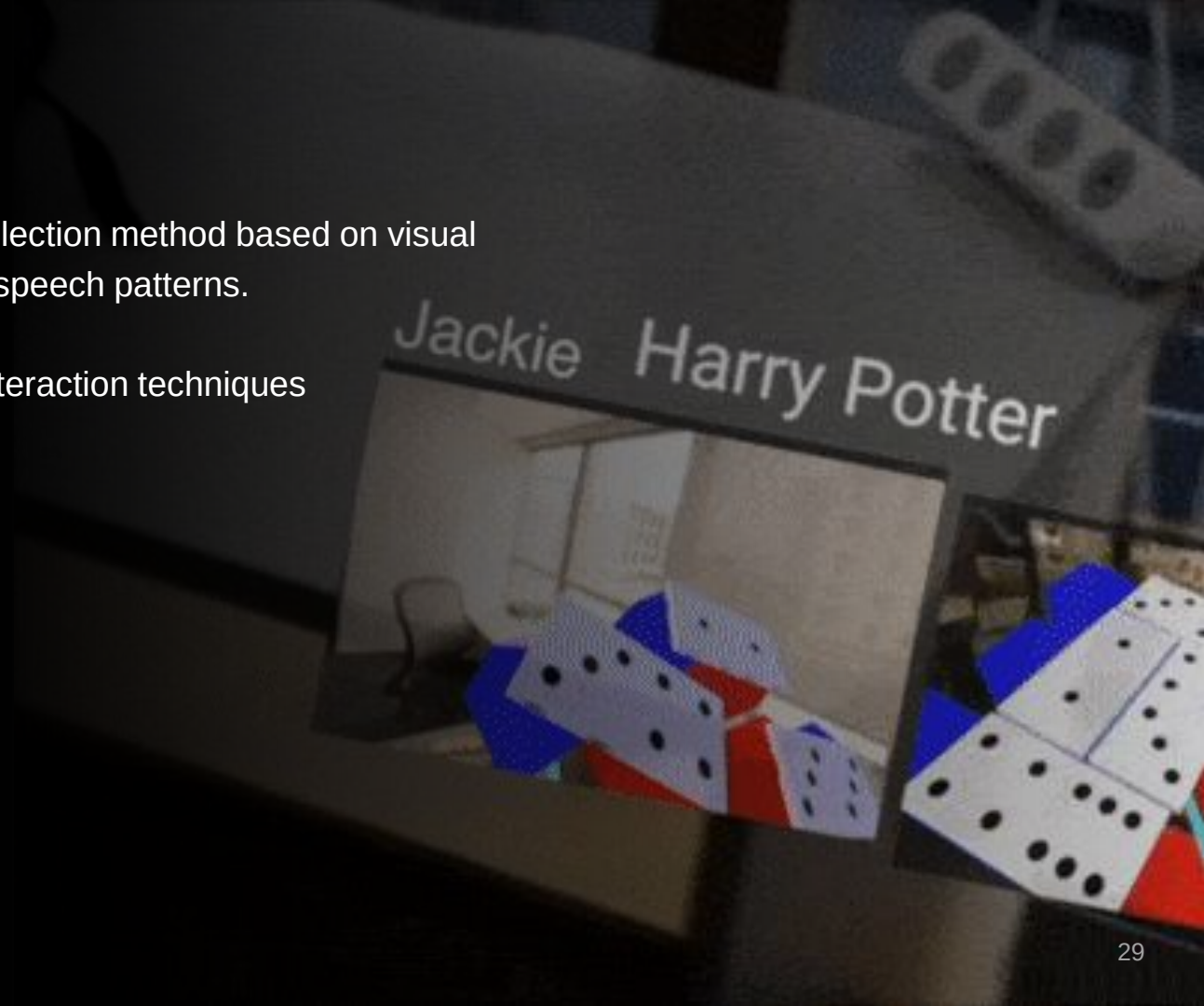
Mixed-perspective shown to improve the collaborative experience in remote settings

managing multiple viewpoints challenging for users to decide which views to focus on

Vicarious

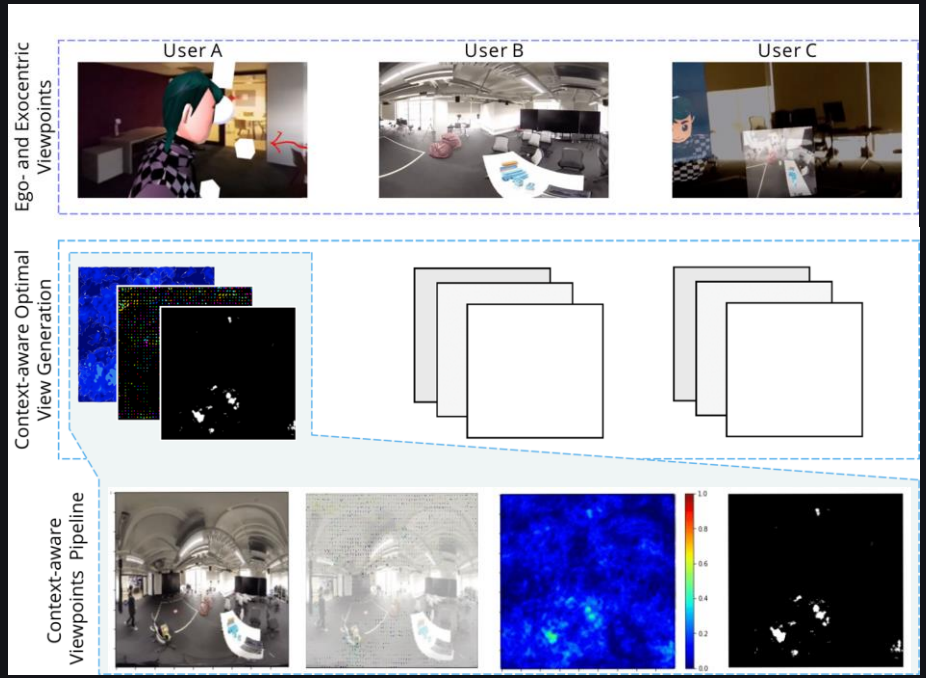
- Context-aware viewpoint selection method based on visual saliency, user actions, and speech patterns.

- Various visualization and interaction techniques



Viewpoint Selection

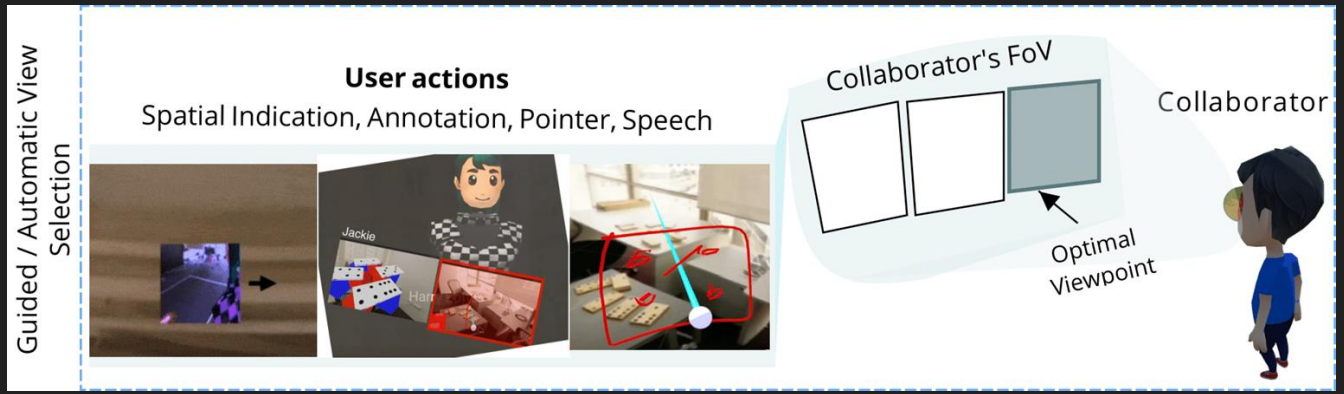
- Visual saliency



$$S(i) = w1 * K(i) + w2 * O(i) + w3 * G_\sigma(i)$$

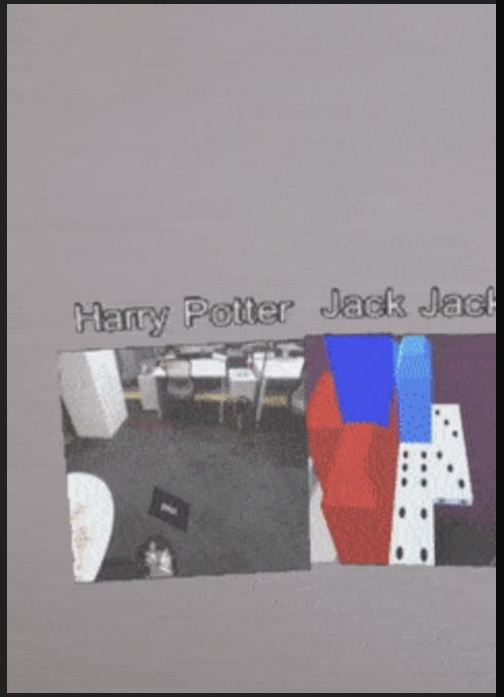
Viewpoint Selection

Predefined actions
- Verbal cues (*Voice activity + Intensity)*
- Visual cues (*Point, Annotate, Gestures)*



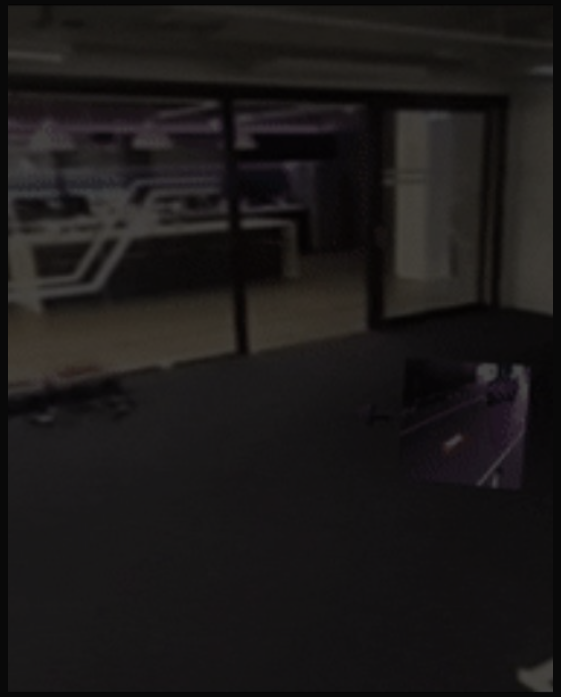
Visualization and Interaction

1. Smooth Follow



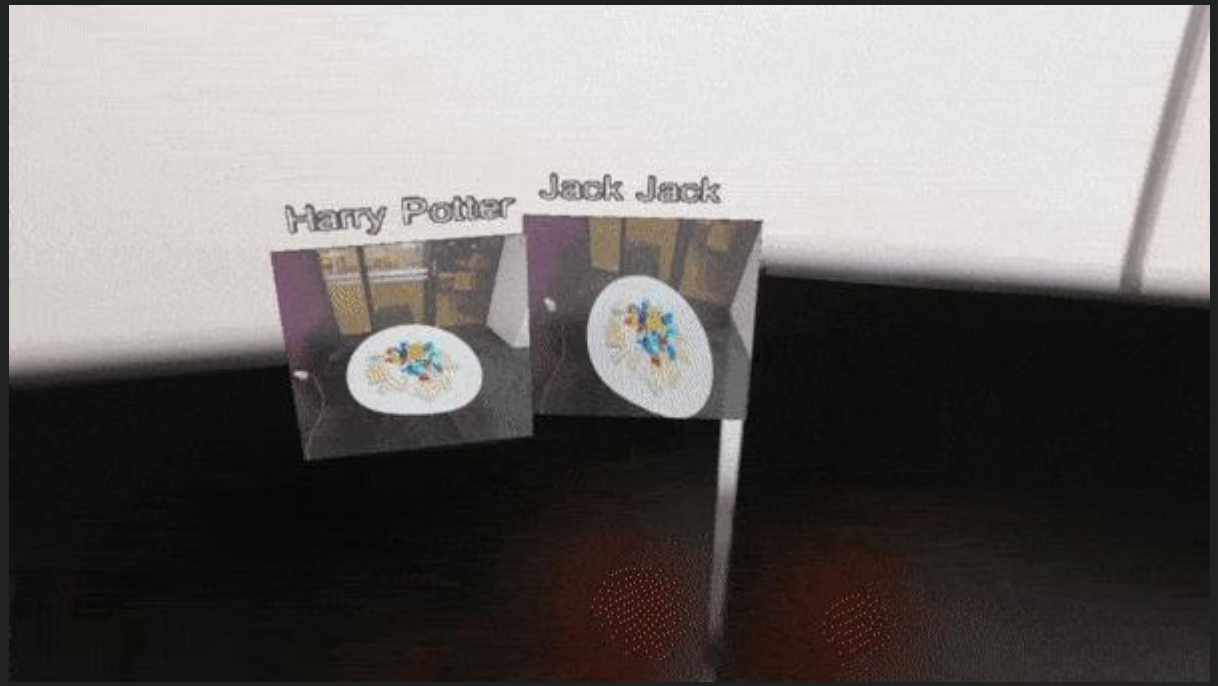
2. Pinned to Fixed Location

3. Spatial Indication



Visualization and Interaction

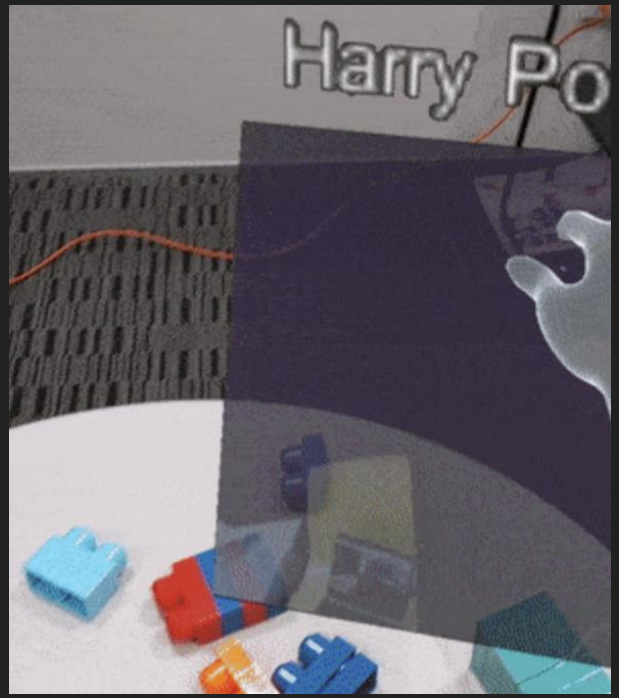
1. Smooth Follow



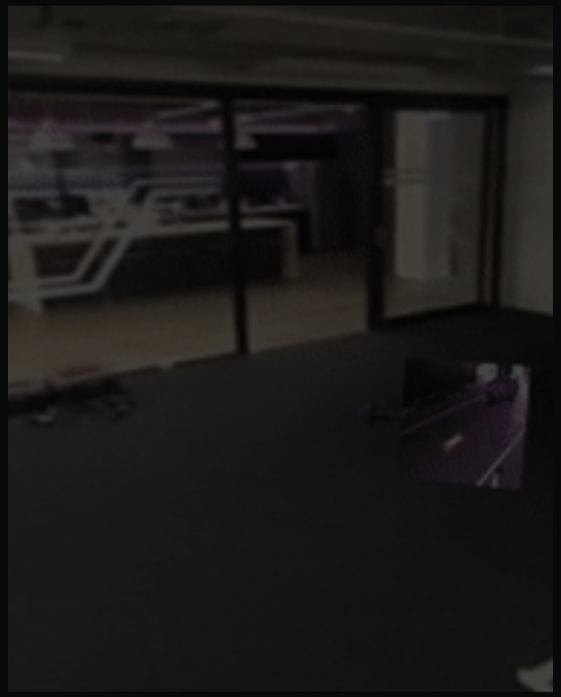
Visualization and Interaction

1. Smooth Follow

2. Pinned to Fixed Location

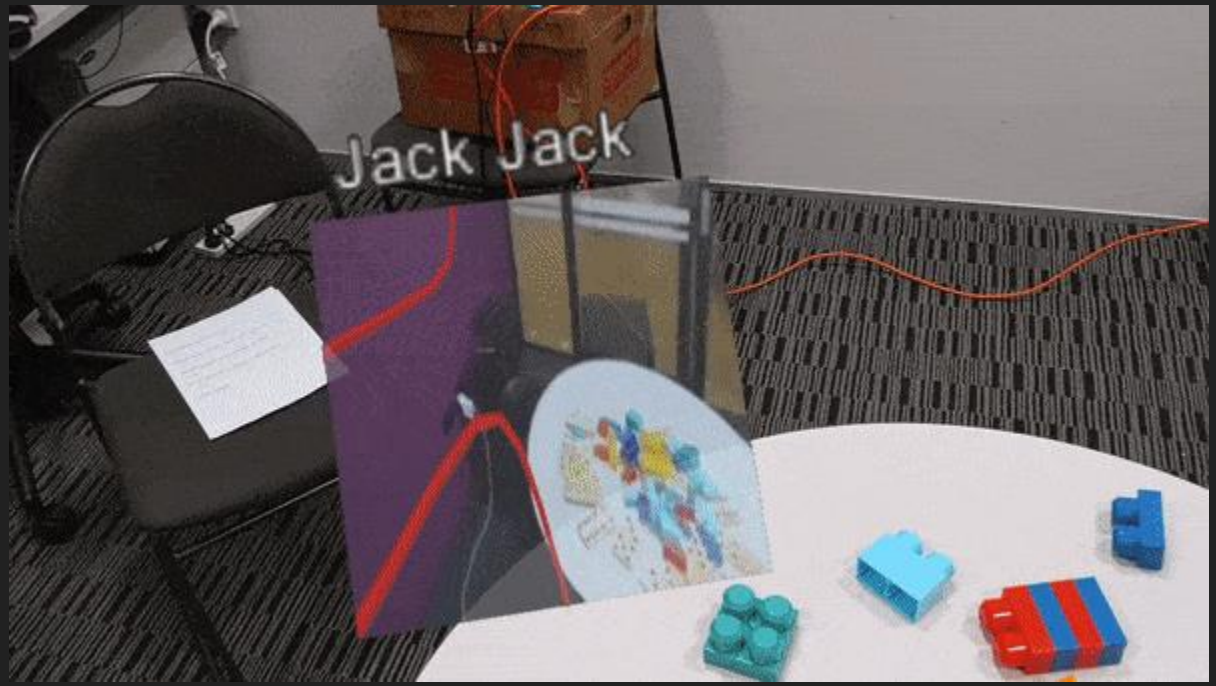


3. Spatial Indication



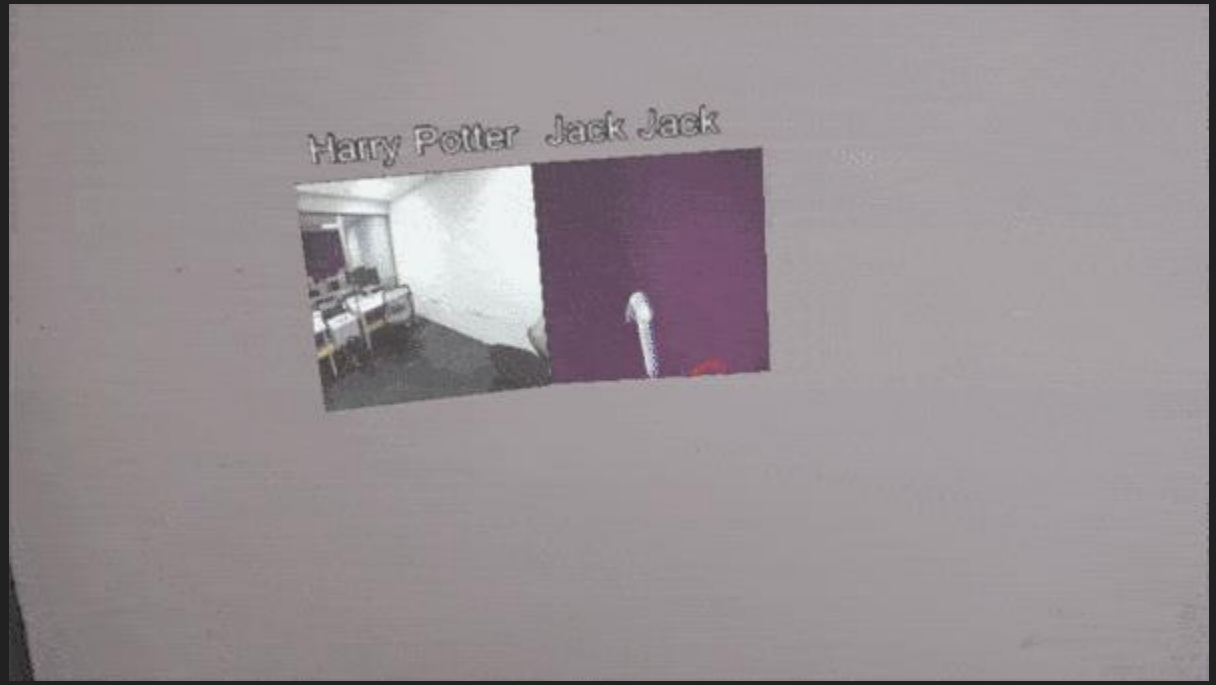
Visualization and Interaction

2. Pinned to Fixed Location



Visualization and Interaction

2. Pinned to Fixed Location (Zoom In/Out)

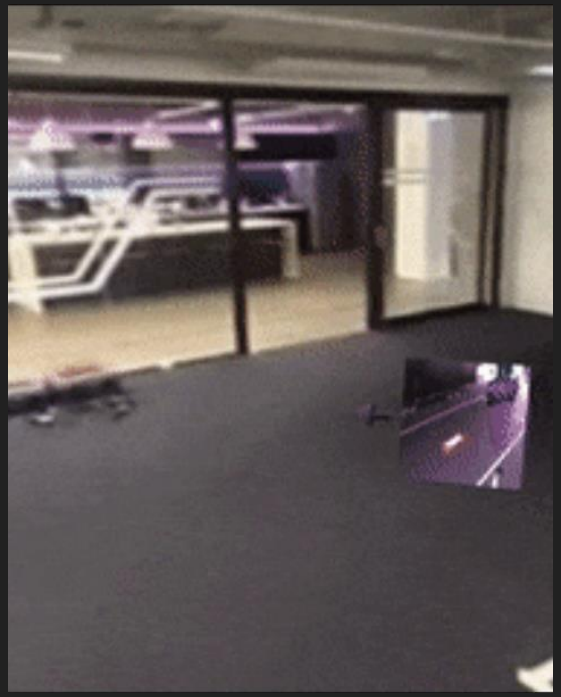


Visualization and Interaction

1. Smooth Follow

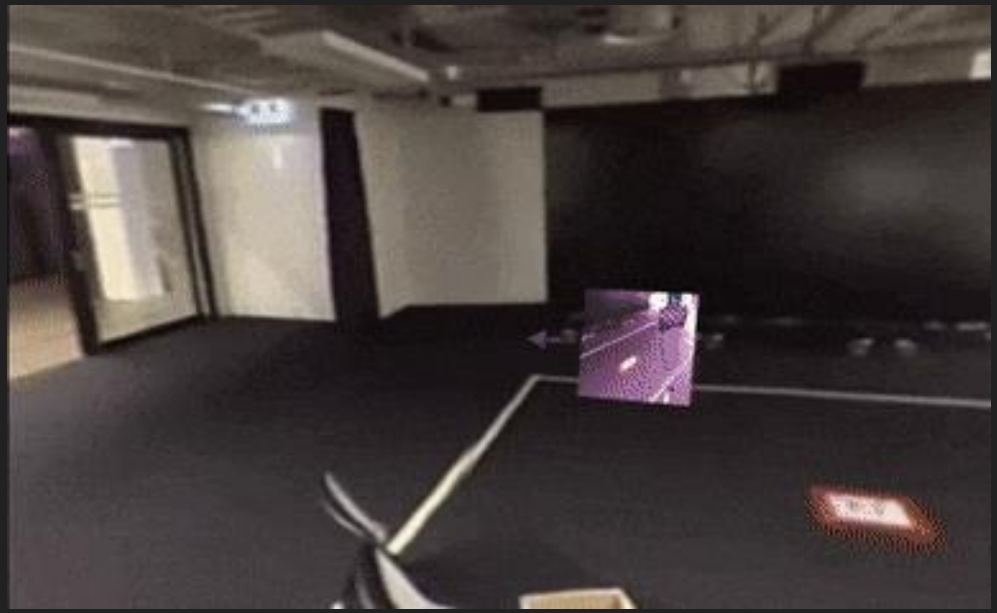
2. Pinned to Fixed Location

3. Spatial Indication



Visualization and Interaction

3. Spatial Indication

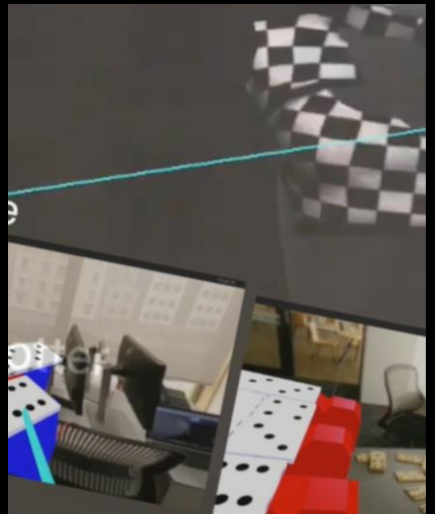


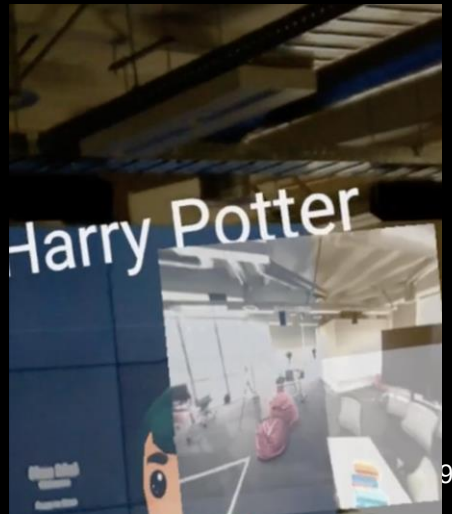
Vicarious: Research Questions

viewpoint perspective

RQ4: How can users' **viewpoints be managed effectively** to support various types of collaborative tasks in the Mixed Reality Collaboration space?

No Selection (NS)	Manual Selection (MS)	Guided Selection (GS)	Automatic Selection (AS)





Vicarious: Research Questions

viewpoint perspective

RQ4: How can users' **viewpoints be managed effectively** to support various types of collaborative tasks in the Mixed Reality Collaboration space?

multi-perspective collaboration

RQ5: How can **multi-perspective collaboration be effectively** achieved for supporting collaborative tasks?

RQ5.1 Would context-aware viewpoint selection increase the sense of presence in remote collaboration?

RQ5.2 Would context-aware viewpoint selection improve the task performance compared to using no-view selection?

Experimental Setup

4×2 mixed factorial design

between-subjects factor:
local vs. remote

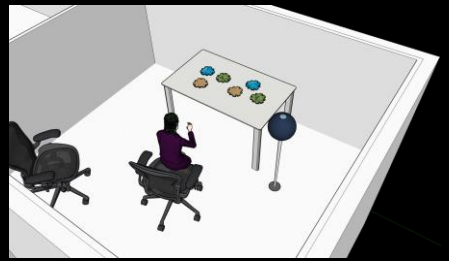
within-subjects factor:
NS vs. MS vs. GS vs. AS

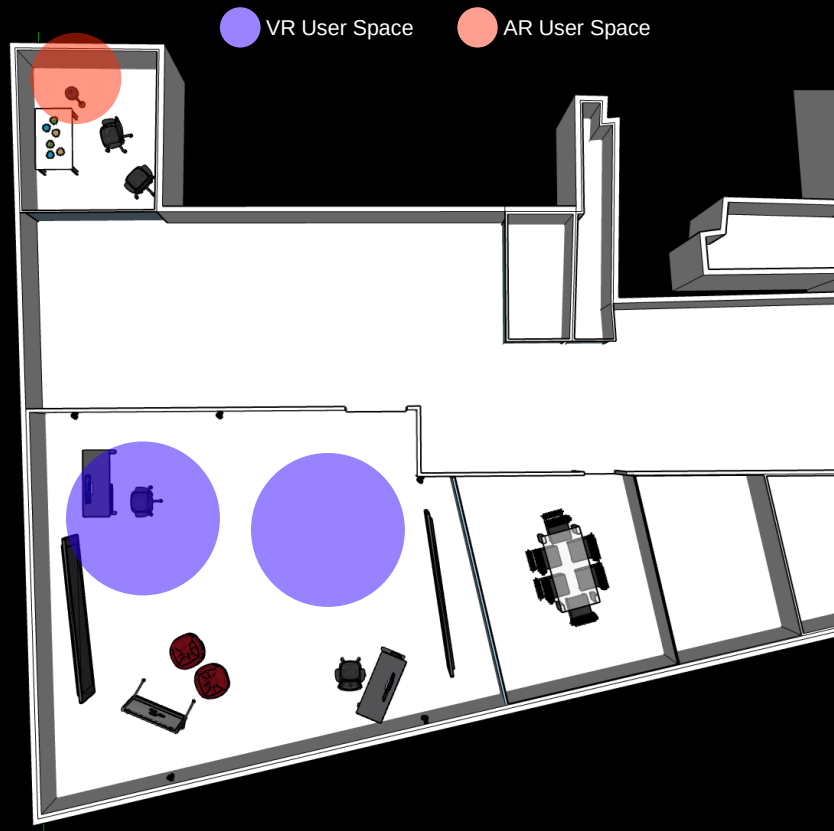
27 participants

18—55 years

μ = 28.13

σ = 7.32





Results

Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

Not Statistically Significant

- Social Presence
- System Usability
- Simulator Sickness

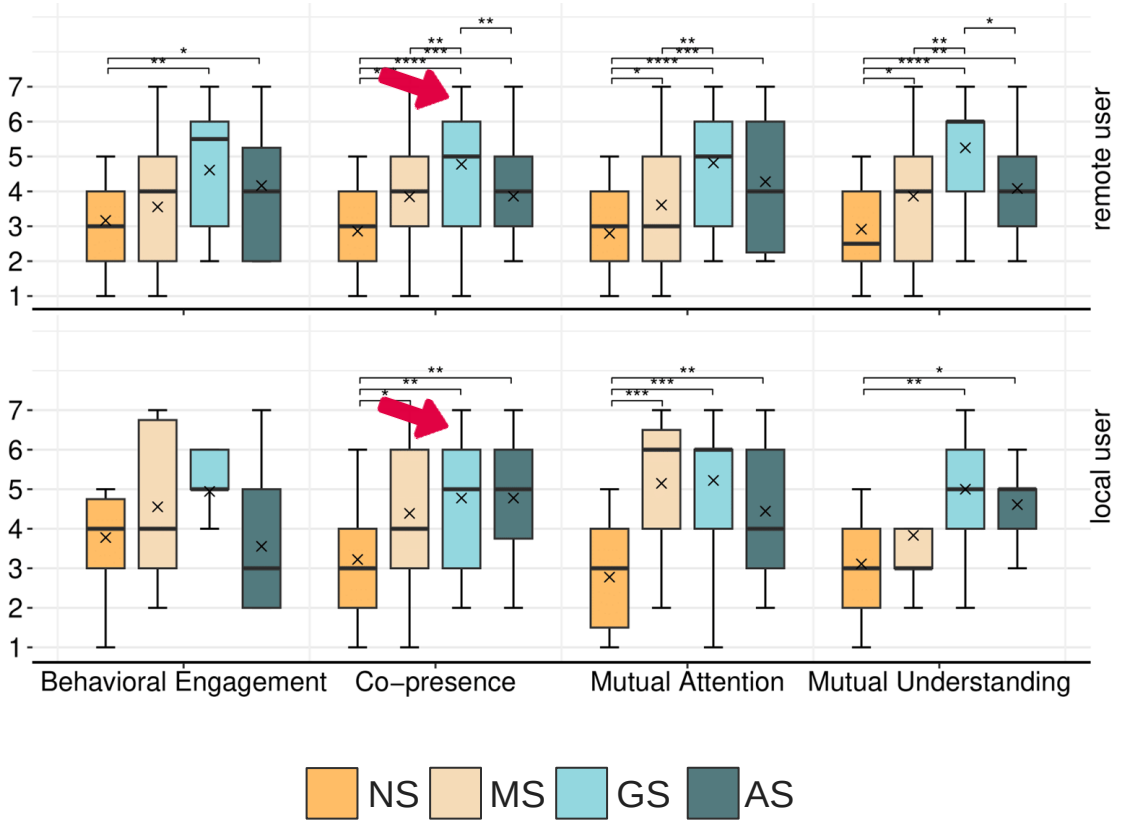
Results

Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

Not Statistically Significant

- Social Presence
- System Usability
- Simulator Sickness



Results

Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

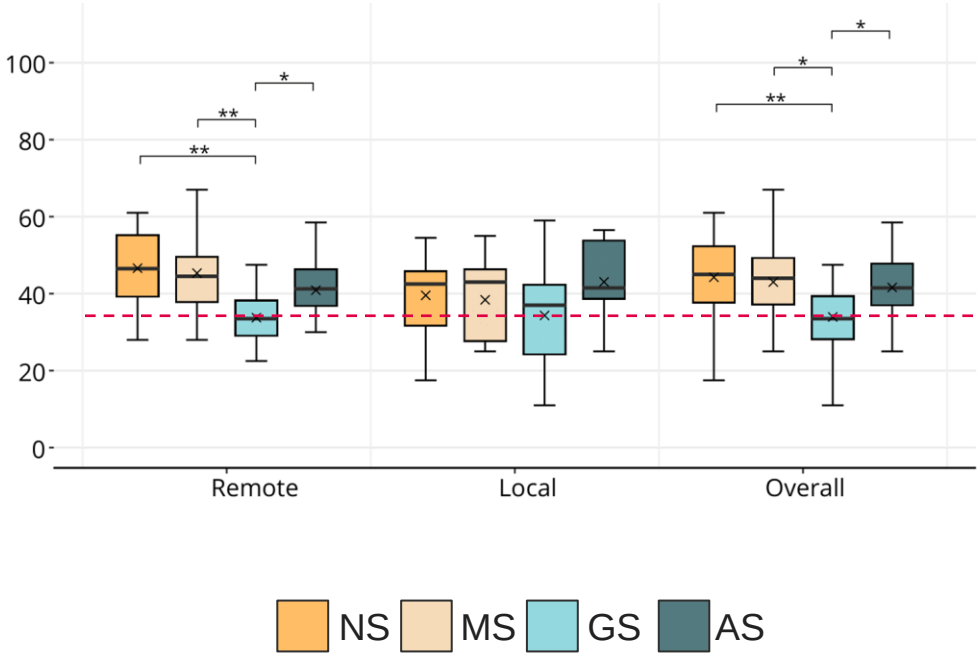
Not Statistically Significant

- Social Presence
- System Usability
- Simulator Sickness

GS

Local: M = 32.31, SD = 23.96

Remote: M = 34.76, SD = 25.41



Results

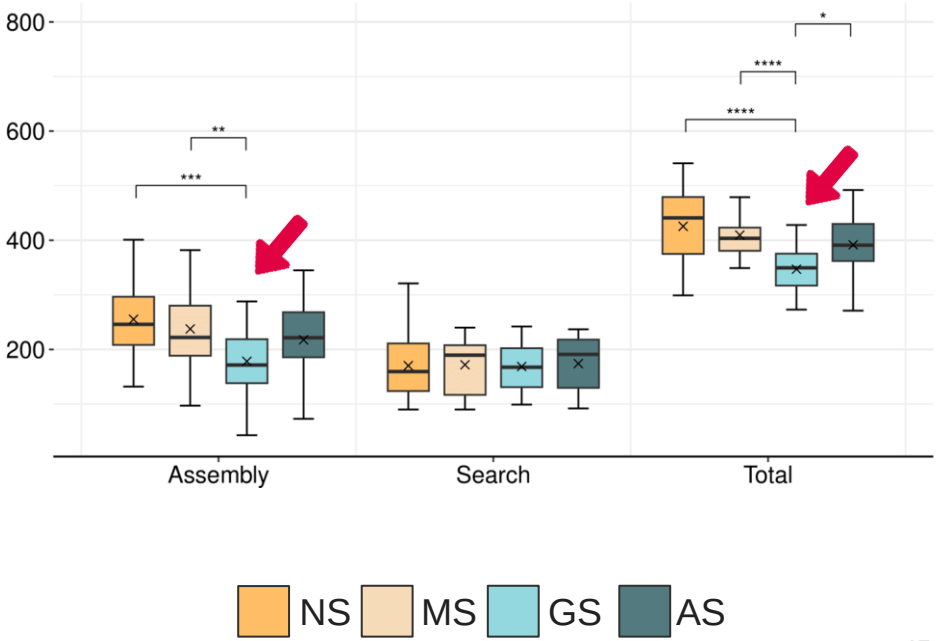
Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

Not Statistically Significant

- Social Presence
- System Usability
- Simulator Sickness

GS: M = 393.46, SD = 60.32



Results

Statistically Significant

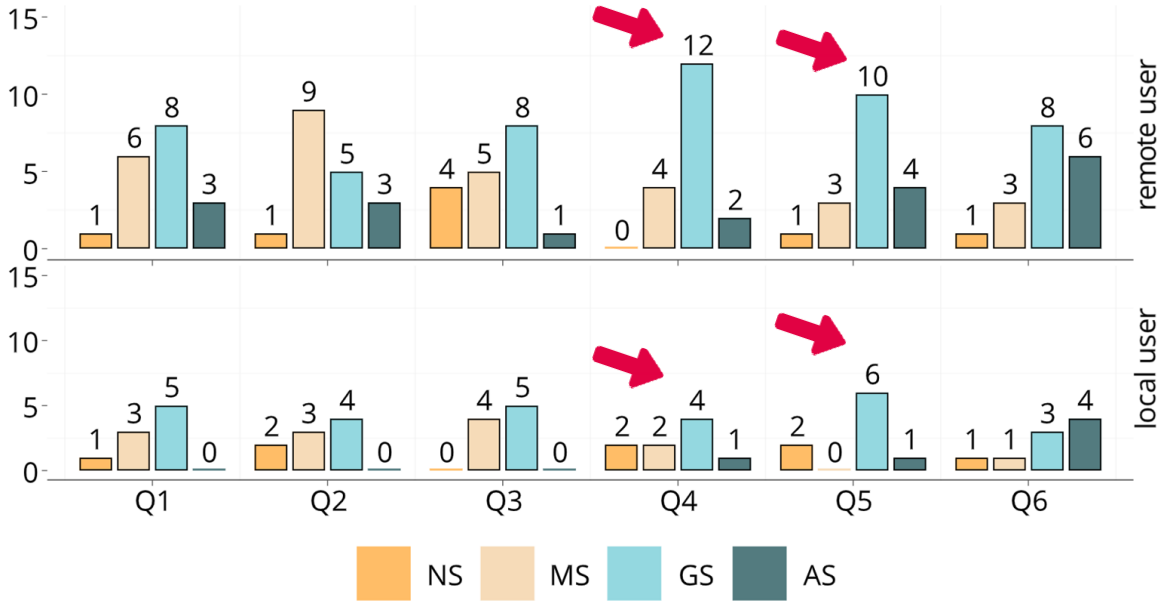
- **Spatial Presence**
- **Task Workload**
- **Task Completion Time**
- **Preference**

Not Statistically Significant

- **Social Presence**
- **System Usability**
- **Simulator Sickness**

Q4 Which condition did you find most effective for communicating with your partner?

Q5 Which condition did you find most helpful for completing the task?



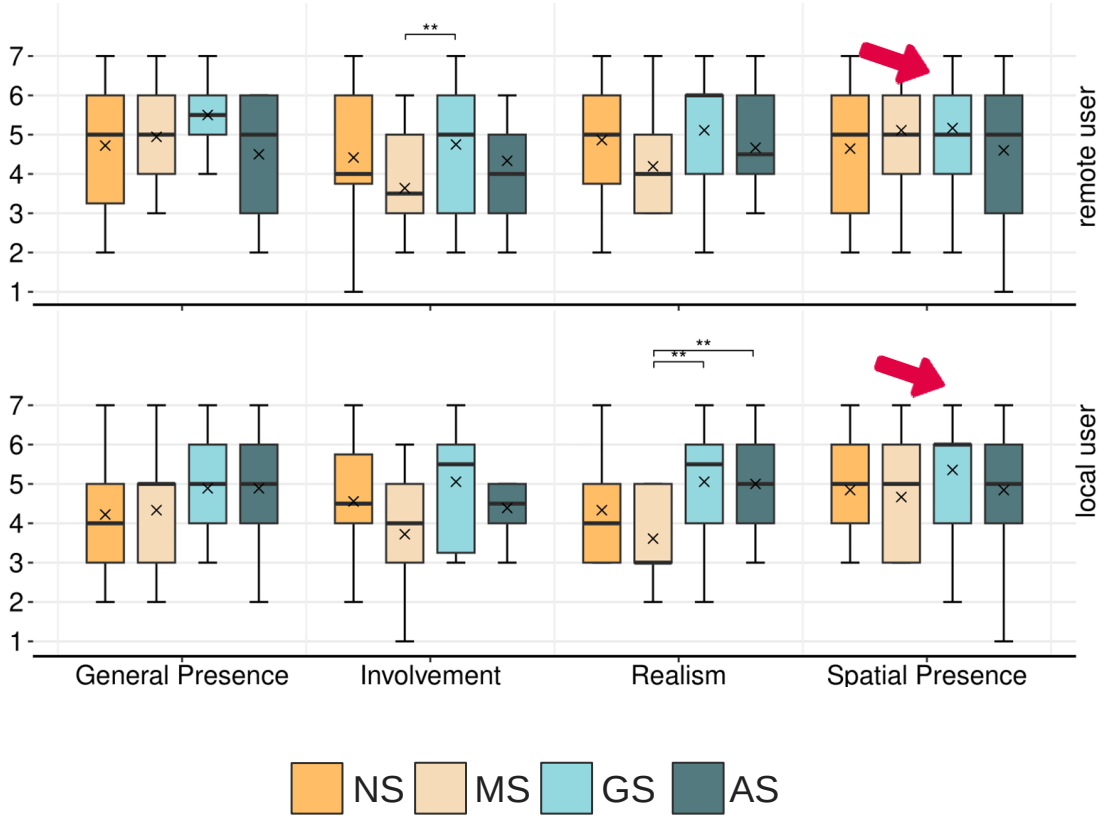
Results

Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

Not Statistically Significant

- Social Presence
- System Usability
- Simulator Sickness



Results

Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

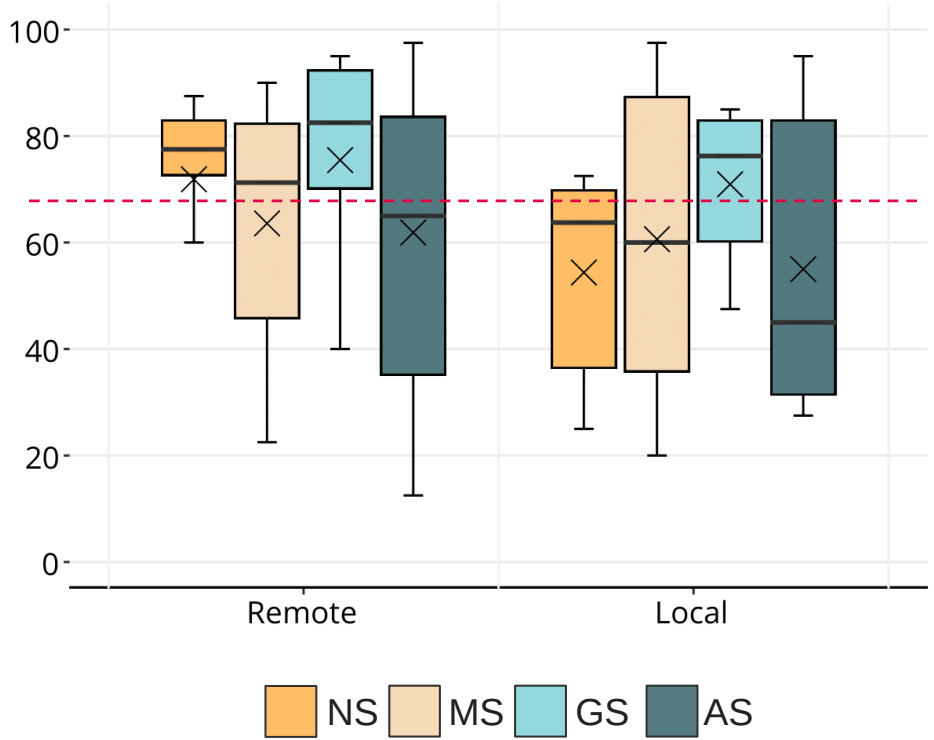
Not Statistically Significant

- Social Presence
- System Usability
- Simulator Sickness

GS

Local: M = 68.61, SD = 15.87

Remote: M = 74.58, SD = 19.82



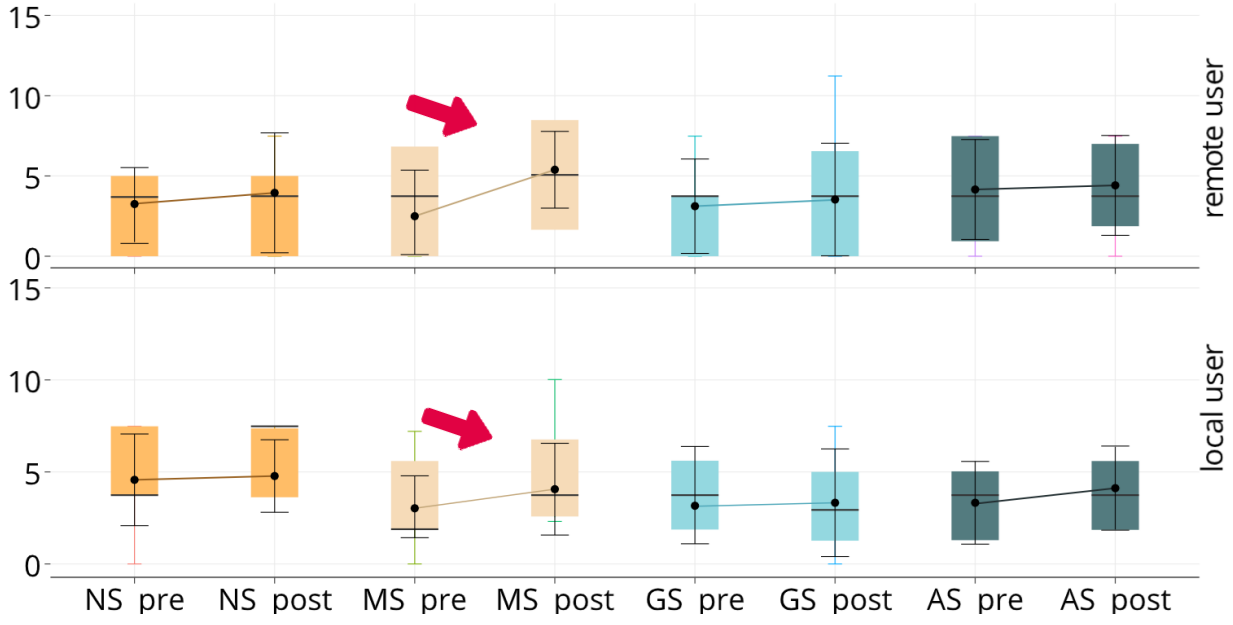
Results

Statistically Significant

- Spatial Presence
- Task Workload
- Task Completion Time
- Preference

Not Statistically Significant

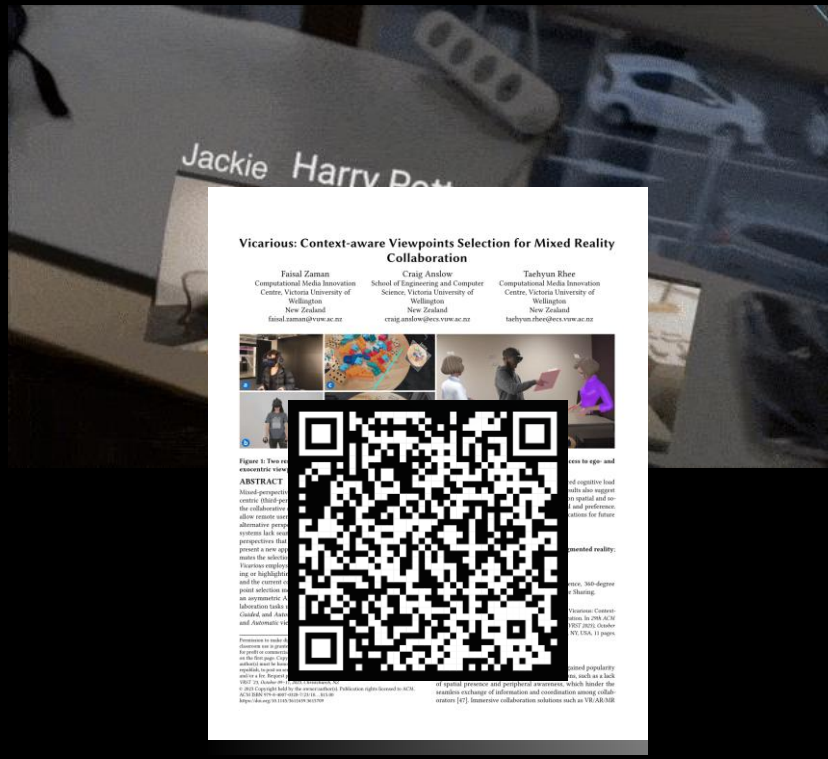
- Social Presence
- System Usability
- Simulator Sickness



Summary of Vicarious

- **Context-aware viewpoint selection method**
  - simplifies and automates multiple viewpoints.
  - finds optimal viewpoints based on audio-visual cues.
- **User evaluation**
  - 27 participants under four distinct conditions
  - asymmetric multiuser AR-VR setup
- **Findings and design implications**
  - improved task space understanding and performance reduced cognitive load
  - GS was the most preferred

Why work remotely when you can work *Vicariously?*



F. Zaman, C. Anslow, T. Rhee. **Vicarious: Context-aware Viewpoints Selection for Mixed Reality Collaboration**. ACM VRST, 2023.

Summary of Contributions

RQ1: teleportation

- Rich real-time telepresence experiences can feasibly be created
- MRMAC induces significantly more spatial presence than 2D and 360 video alone
- These experiences still pose a significant engineering challenge
- Opened a lot of future research opportunities

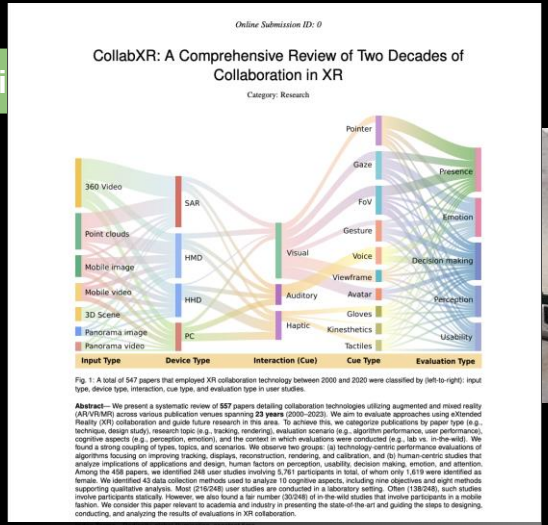
RQ2: representation

RQ3: interacti

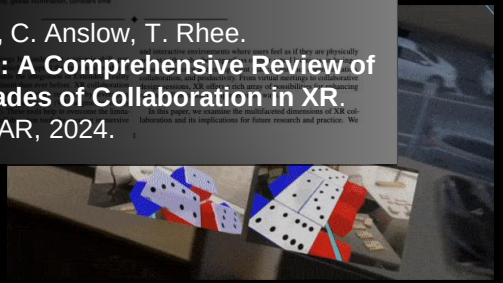
RQ4: viewpoint perspective

- Optimal viewpoint selection improved users' task space understanding
- Task performance and reduced cognitive load.
- Users preferred the intervention

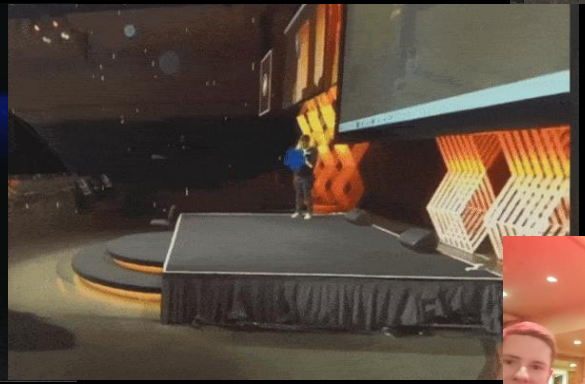
RQ5: multi-perspective collaboration



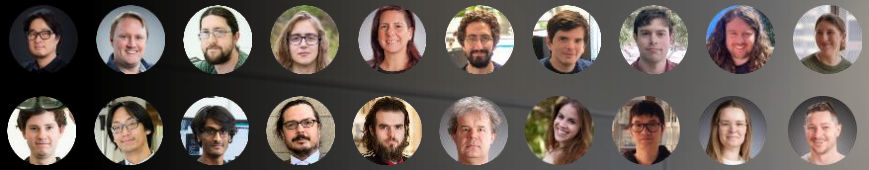
F. Zaman, C. Anslow, T. Rhee.
CollabXR: A Comprehensive Review of Two Decades of Collaboration in XR.
IEEE ISMAR, 2024.



Research Journey



Acknowledgements:



A/Prof Taehyun Rhee and Dr Craig Anslow for their guidance.
Dr Andrew Chalmers, Dr Jacob Young, and Dr Nadia Pantidi for their never-ending support.
Dr Ben Powley for being my PhD buddy.
Dr Rafael Dos Anjos and Dr Daniel Medeiros for their encouragement.
Stephen Thompson, Ian Loh and Nicolas Vergnaud for being great research buddies.
Everybody who sent me their sample slides.
My friends and coworkers at HCI group for their patience and understanding.
The people that attended this talk thank you for listening.
Thank you to CMIC for their support, the opportunity and the wonderful experiences.
This project was supported by the Entrepreneurial University Program funded by TEC, New Zealand.

CMIC Computational Media Innovation Centre

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA