

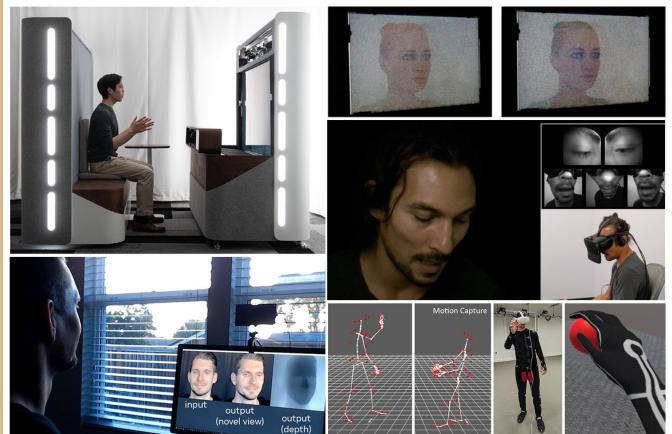


SIGGRAPH 2023
LOS ANGELES+ 6-10 AUG

State of the Art in Telepresence

Part 1

Jason Lawrence
Google Research



Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this header and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
SIGGRAPH '23 Courses, August 06-10, 2023, Los Angeles, CA, USA
ACM 979-4-007-0145-0/23/08.
10.1145/3587423.3595489

Speakers



Jason Lawrence
Google Research



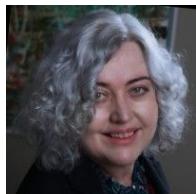
Ye Pan
Shanghai Jiao Tong
9:20a - 9:50a



Dan B Goldman
Google Research
9:50a - 10:20a



Rachel McDonnell
Trinity College
10:30a - 11:00a



Carol O'Sullivan
Trinity College
11:00a - 11:30a



Dave Luebke
NVIDIA Research
2:15p - 2:45p



Koki Nagano
NVIDIA Research
2:45p - 3:15p



Michael Zollhöfer
Reality Labs Research
3:30p - 4:00p



Jason Saragih
Reality Labs Research
4:00p - 4:30p

Part 1 (9a-12p)		
9:00a-9:20a (20m)	Welcome and Introduction	Jason Lawrence
9:20a-9:50a (30m)	Human Factors for Telepresence	Ye Pan
9:50a-10:20a (30m)	Google's Project Starline	Dan Goldman
10:20a-10:30a (10m)	BREAK	
10:30a-11:00a (30m)	Perception of Virtual Humans	Rachel McDonnell
11:00a-11:30a (30m)	Physical Interactions in Telepresence	Carol O'Sullivan
11:30a-12:00p (30m)	Discussion and Q&A	All
Part 2 (2p-5p)		
2:00p-2:15p (15m)	Welcome and Introduction	Jason Lawrence
2:15p-2:45p (30m)	Opportunities for AI-Mediated 3D Telepresence	Dave Luebke
2:45p-3:15p (30m)	AI-Driven Synthesis for 3D Telepresence	Koki Nagano
3:15p-3:30p (15m)	BREAK	
3:30p-4:00p (30m)	Towards Complete Codec Telepresence	Michael Zollhöfer
4:00p-4:30p (30m)	Neural Representations of Humans for Telepresence	Jason Saragih
4:30p-5:00p (30m)	Discussion and Q&A	All

Schedule of this two-part course with the first part in a morning session and the second part in the afternoon.

tel·e·pres·ence

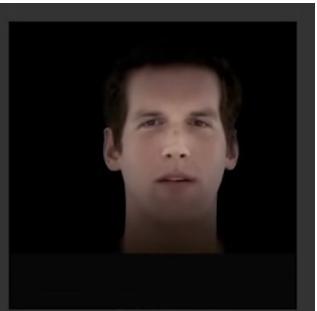
the use of virtual reality technology, especially for remote control of machinery or for apparent participation in distant events

The textbook definition of telepresence, which encapsulates “teleoperation” of distant machinery and apparent participation in distant events...

tel·e·pres·ence

the use of virtual reality technology, especially for remote control of machinery or for apparent participation in distant events

which is the aspect we will discuss in this course, that of participating in a conversation or interaction with other people who are, in fact, separated by distance.



And to help ground the kind of work we will review in this course, here are a few different prototype telepresence systems. Each of these combines some type of 3d display with some type of real-time 3d imaging and rendering technique in an attempt to bring people together in a convincing and immersive and co-present way despite the fact that they are in different locations.

These are also some of the specific systems that we will look at in detail in this course.

Top left image credit: Project Starline (Google 2023)

Bottom left image credit: NVIDIA AI-Mediated Telepresence eTech Demo (SIGGRAPH 2023)

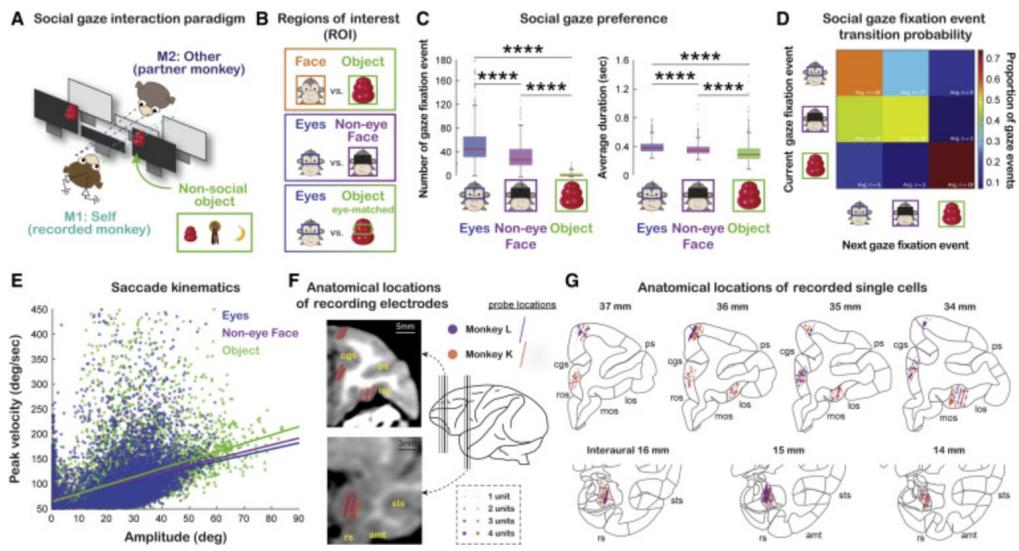
Right image source: <https://www.youtube.com/watch?v=86-tHA8F-zU>



But before we dig into the different research questions and latest developments in this space, it's worth asking the question: what is the potential impact of advancing this area?

Simply put, face-to-face communication and in-person interactions are fundamental to being human and, of course, we do this all of the time.

Image: Adobe Stock 187752803



Dal Monte, et al., Widespread implementations of interactive social gaze neurons in the primate prefrontal-amygda networks. *Neuron*. May 10, 2022.

One example of our highly evolved ability for in-person communication is the way we use non-verbal cues, like mutual eye gaze. (You'll be hearing plenty about eye contact and mutual gaze in today's course given its importance in human face-to-face conversation.)

In fact, eye gaze and the role it plays in communication a very active topic of research in the psychology and neuroscience fields, such as this study that was recently published in the journal Neuron that reveals new links between mutual eye gaze and measurable brain activities in primates.

[https://www.cell.com/neuron/fulltext/S0896-6273\(22\)00358-0?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627322003580%3Fshowall%3Dtrue](https://www.cell.com/neuron/fulltext/S0896-6273(22)00358-0?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627322003580%3Fshowall%3Dtrue)



And, because of this basic human need, we go to great lengths to be together in person, at significant financial and environmental costs.

Source of pie chart:

<https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions#transportation>

Image source:

<https://www.publicdomainpictures.net/pictures/260000/velka/airplane-sunset-travel.jpg>

Karen Arnold has released this “Airplane Sunset Travel” image under Public Domain license ([CC0 Public Domain](#)).



Advanced telepresence systems could also expand access to critical resources, such as bringing better medical care to remote areas, along with significant opportunities across many other industries.

Image: Adobe Stock 327070130



And the experience of endless video calls that are common in today's society, accelerated by the COVID-19 pandemic, highlights the significant gap that remains between prevailing video conferencing technology and truly being together in person. In other words, there is a big opportunity to close this gap.

Image: Adobe Stock 355042323



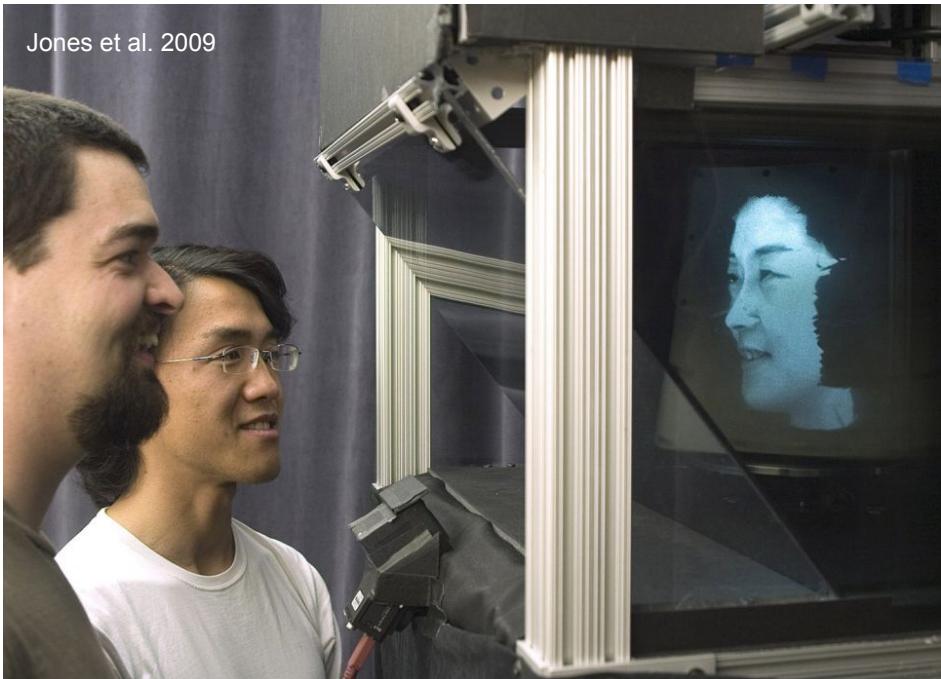
Mulligan et al. 2004

And because of these enormous opportunities, there is a considerable body of research that employs 3d displays and 3d imaging techniques in different ways in an attempt to reproduce the experience of being with a remote person.

One such research prototype is pictured here, developed by researchers at the University of Pennsylvania and the University of North Carolina around 2000. They combined a high-frame-rate head tracking system and a stereo projection display in order to convey both binocular stereo cues and motion parallax cues to an observer who is wearing glasses. The 3d representation of the remote scene is reconstructed at real-time video rates using binocular and trinocular stereo techniques.

Reference: J. Mulligan, X. Zabulis, N. Kelshikar and K. Daniilidis, "Stereo-based environment scanning for immersive telepresence," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 3, pp. 304-320, March 2004, doi: 10.1109/TCSVT.2004.823390. (Figure caption: Working prototypes from May and October 2000. Left: In May, the University of Pennsylvania and advanced network and services each transmitted five simultaneous trinocular depth streams to the University of North Carolina Chapel Hill. Right: In October, 3-D interaction and synthetic objects were added to the collaborative environment.)

Jones et al. 2009



Other work has explored automultiscopic or autostereoscopic displays – where the ‘auto-’ prefix indicates that the observer isn’t required to wear any type of special glasses or headwear, yet these displays are still able to provide stereo cues.

The work shown here by Jones and colleagues from 2009 combined a high-frequency projector with a very fast spinning mirror(!) to enable a multiscopic 3d display. They described using this display for telepresence scenarios, where it displays a live 3d image of a remote person being reconstructed using a structured light technique – and their image is able to be seen and heard by multiple people from different viewing directions, simultaneously.

Reference: Andrew Jones et al. 2009. Achieving eye contact in a one-to-many 3D video teleconferencing system. ACM ToG 28(3)



Maimone et al. 2012

Maimone and colleagues contributed important work to this area as well, showing how a lenticular-based autostereoscopic display could be combined with commodity real-time depth cameras to enable telepresence applications.

The system shown here combined 5 Kinect depth cameras to generate stereo images for their display.

Reference: Andrew Maimone et al. 2012. Enhanced personal autostereoscopic telepresence system using commodity depth cameras. C&G 36(7)



Zhang et al. 2013

Zhang and colleagues also experimented with combining multiple streaming depth cameras and a lenticular-based autostereo display to enable multiple endpoints to connect within a shared virtual space.

Their work is noteworthy in considering how more than two people could participate in this type of interaction.

Reference: Cha Zhang et al. 2013. Viewport: A distributed, immersive teleconferencing system with infrared dot pattern. IEEE MM 20(1)



Project Starline (Google, 2021)

This area also includes some of our work at Google as part of Project Starline, described in our 2021 Siggraph Asia paper. My colleague, Dan Goldman, will give a review of this work later in this course.

Reference: Jason Lawrence, Dan B Goldman, Supreeth Achar, Gregory Major Blascovich, Joseph G. Desloge, Tommy Fortes, Eric M. Gomez, Sascha Häberling, Hugues Hoppe, Andy Huibers, Claude Knaus, Brian Kuschak, Ricardo Martin-Brualla, Harris Nover, Andrew Ian Russell, Steven M. Seitz, and Kevin Tong. 2021. Project starline: a high-fidelity telepresence system. ACM Trans. Graph. 40, 6, Article 242 (December 2021).



Trevithik et al. (2023)

And there is exciting research that is being shared at this conference by our co-presenters from NVIDIA – a new AI technique for real-time view synthesis of people from a single camera viewpoint.

Reference: Alex Trevithick, Matthew Chan, Michael Stengel, Eric R. Chan, Chao Liu, Zhiding Yu, Sameh Khamis, Manmohan Chandraker, Ravi Ramamoorthi, and Koki Nagano. Real-Time Radiance Fields for Single-Image Portrait View Synthesis. ACM Transactions on Graphics (SIGGRAPH) 2023.
[\(https://research.nvidia.com/labs/nxp/lp3d/\)](https://research.nvidia.com/labs/nxp/lp3d/)



**** Live Demo at Emerging Technologies ****

And they demonstrate how this view synthesis technique can be used to drive a 3D display to enable communication. You can see a demo of this for yourself in the Emerging Technologies area of the conference.

Image credit: NVIDIA AI-Mediated Telepresence eTech Demo (SIGGRAPH 2023)



Oculus Quest 2 (2020)

Another important thread of work in this field explores using virtual- or augmented-reality head-mounted displays, which have seen immense progress in terms of their capabilities, comfort, and commercial availability.

Image credit: photo of an individual using an Oculus Quest 2 VR system. (Adobe Stock 443087061)

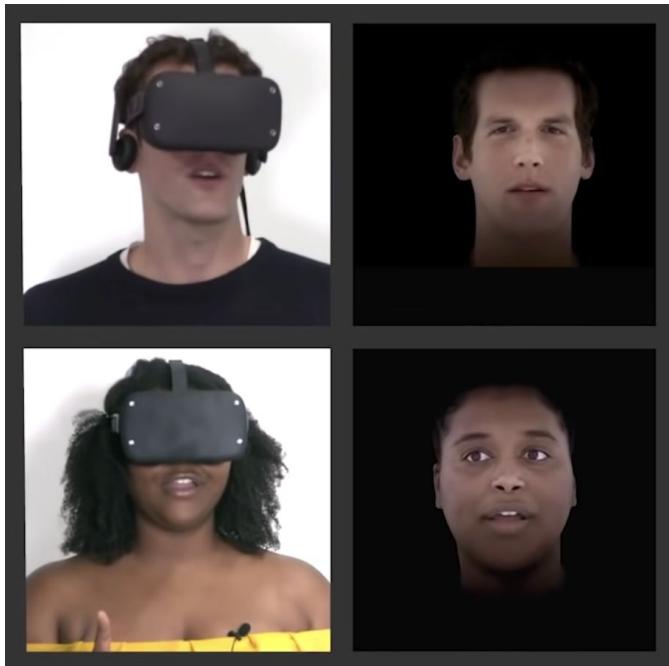


Orts-Escalano et al. 2016

A group at Microsoft demonstrated a complete symmetric prototype telepresence system based on the Microsoft Hololens AR display that they called “Holoportation.” The real-time 3D imaging was accomplished using a small array of streaming depth and color video cameras.

Reference: Sergio Orts-Escalano, Christoph Rhemann, Sean Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016.

Holoportation: Virtual 3D Teleportation in Real-time. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16). Association for Computing Machinery, New York, NY, USA, 741–754.
<https://doi.org/10.1145/2984511.2984517>



Researchers at Meta's Reality Labs Research group have also made many contributions to this area.

This image shows one of their earlier prototype systems capable of visualizing highly photorealistic virtual representations of two people having a conversation with one another in a virtual space viewed through VR HMDs.

This line of work will be covered in greater detail later in the course by our two presenters from this research group.

Source: <https://www.youtube.com/watch?v=86-tHA8F-zU>

Research Areas

- Enabling technologies (3d displays, real-time 3d imaging, spatial audio, etc.)
- Perception
- Human factors
- Physical interactions
- Application domains
- Ethical considerations

Indeed, there are a number of research topics in this broad area, and we will touch on most of these in this course.

One important area is the technologies that enable these types of systems, such as 3d displays and real-time view synthesis.

Another important research area is around how humans perceive virtual representations of other humans, along with human factors (and even form factors) associated with presenting virtual representations of people.

There is also important work around enabling physical interactions within telepresence systems, which we will touch on in this course.

And there are many application-specific questions, with so much potential to deploy these types of systems across so many domains.

And, finally, there are important ethical considerations surrounding this type of work.

Enabling Technologies

- Real-time 3d imaging and rendering
- Digital human representations (e.g. avatars)
- 3d displays (“fixed” or “free standing” and headsets)
- Streaming compression
- Spatial audio capture and display

Since we will spend a large part of this course talking about enabling technologies, it is helpful to break that category out a bit further.

One key component in these systems is a method for imaging and then rendering the 3d appearance of a remote person at real-time video rates, which typically means between 24 - 60Hz.

Another key question is how the 3d appearance of that remote person is represented, and this could range from stylized depictions to photorealistic ones.

Another is of course the display itself, where the ultimate goal is to reproduce the full set of visual cues that the human brain uses to perceive our three-dimensional world, such as stereopsis, motion parallax, and presenting people at their true scale. It's helpful to make the distinction between “fixed” or “free standing” displays where nothing is required to be worn by the observer, and head-mounted displays or simply headsets.

Another key component are techniques for compressing the streaming visual and audio data driving these systems.

And, finally, technologies for measuring and then reproducing at a remote location a sound field that gives a convincing impression of the sound emanating from the remote talker's mouth.

In this course we will spend time discussing the first three with an emphasis on the first two.

Progress Drivers

- GPU performance-cost trends
- Advancements in 3D ML, specifically real-time view synthesis of people
- Progress in VR/AR headsets and display performance-cost trends

It's useful to also understand some of the drivers of the significant progress in this area over the last ten years.

This includes the incredible gains in performance-cost trends in GPUs, which has increased the viability and scale of these compute-intensive systems.

Of course, the fields of computer graphics and computer vision have seen a paradigm shift away from classical methods in favor of AI and ML-based techniques, which has led to the development of entirely new real-time image synthesis techniques, and new methods for synthesizing media, including images, video, and audio of human forms. These techniques are rapidly advancing and finding applications in telepresence systems.

And, finally, there has been significant progress in commercially available high-fidelity AR/VR headsets from companies like Meta, Microsoft, and Apple, which has expanded the available performance of this critical technology component.



Human factors for telepresence



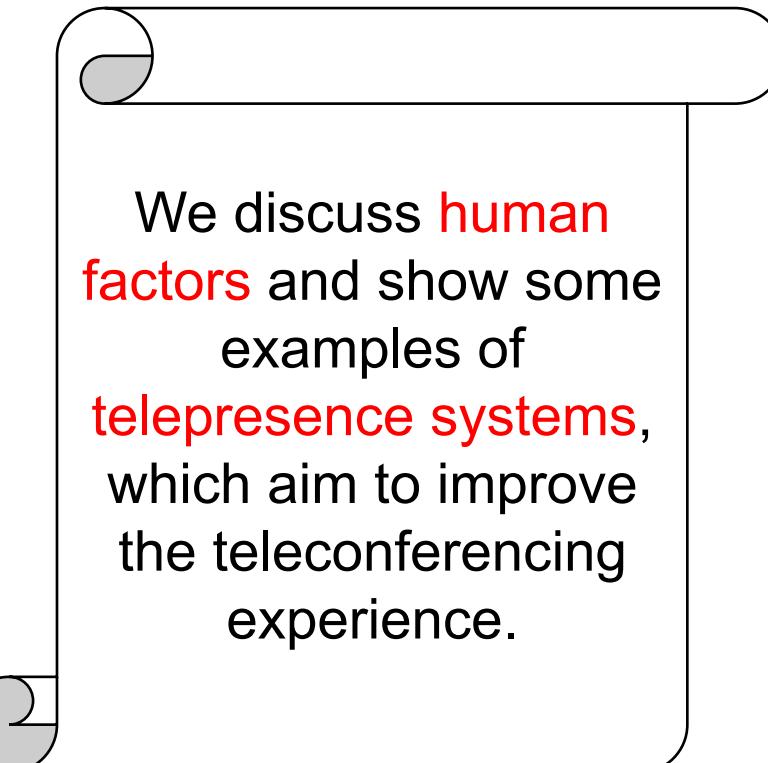
Ye Pan

2023.2.10

饮水思源 · 爱国荣校

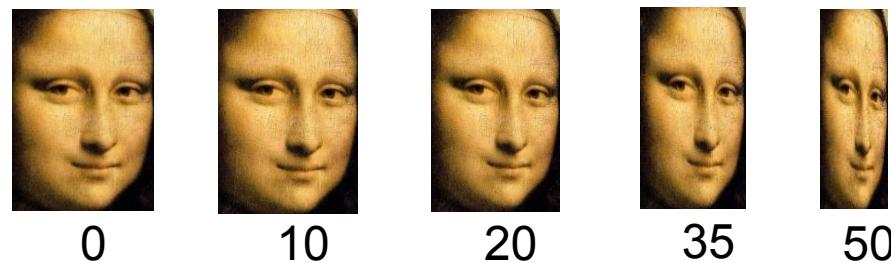
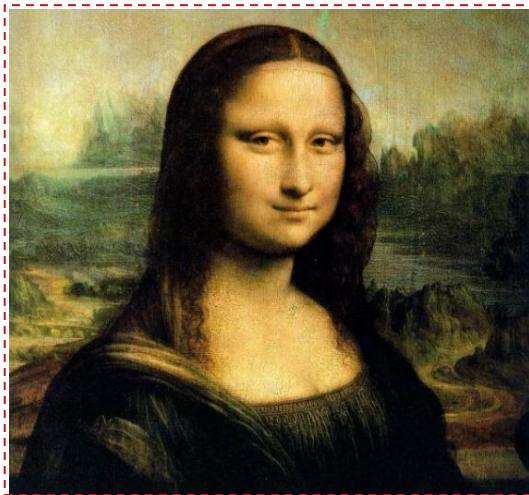
Outlines

- Eye gaze
 - Spherical video-mediated display
 - Cylindrical multiview display
 - Random hole multiview display
- Trust
 - Spherical avatar display
- Physical embodiment & Privacy
 - Humanoid robot
- Leadership effect
 - Collaborative mixed reality
- Engagement
 - Adaptive Dynamic Anamorphosis System
- Emotion
 - Facial Mocap with Live Mood Dynamics
 - Emotional Voice Puppetry



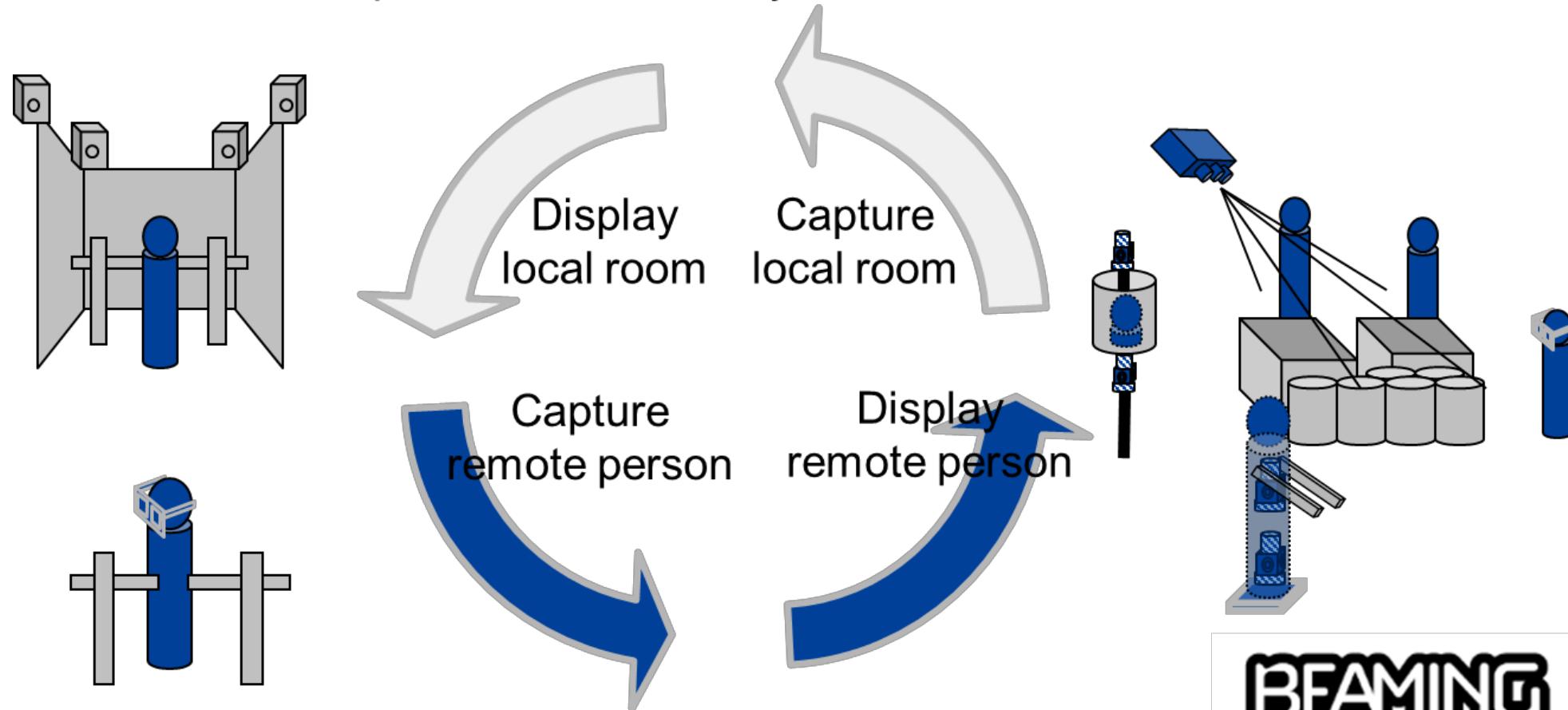
We discuss **human factors** and show some examples of **telepresence systems**, which aim to improve the teleconferencing experience.

Eye Gaze & Mona Lisa Effect



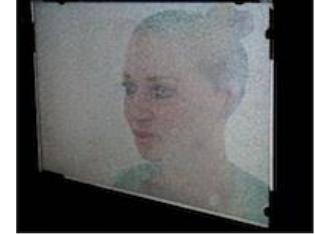
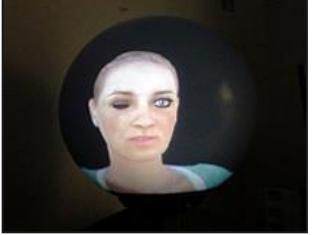
High Level Experience/ Technical Goals

Give the remote person “telepresence” in the local room
The remote person acts as if they were located in the local room

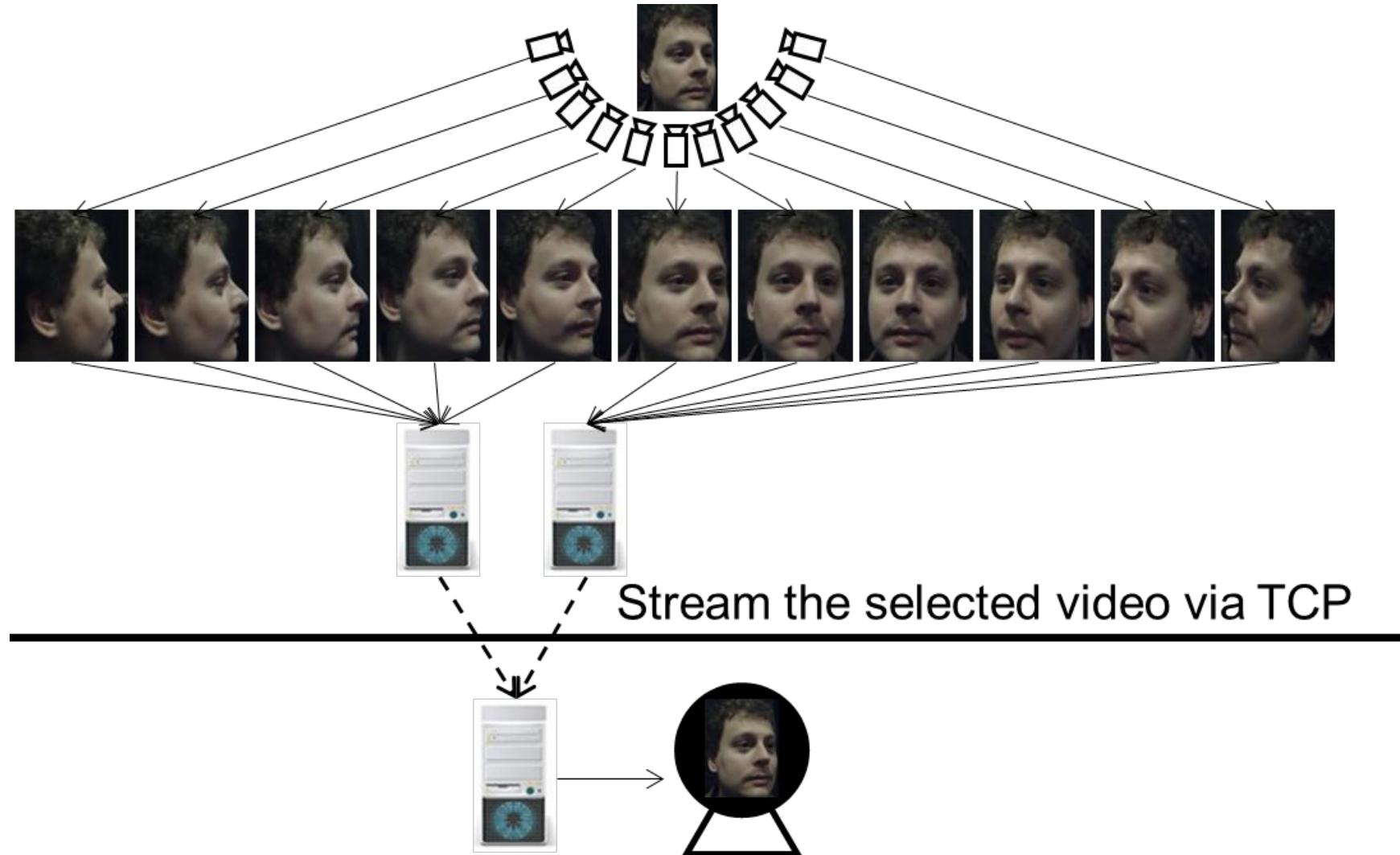


Give the remote person “virtual physical presence” in the local room
Locals act as if the remote person was co-located in the destination

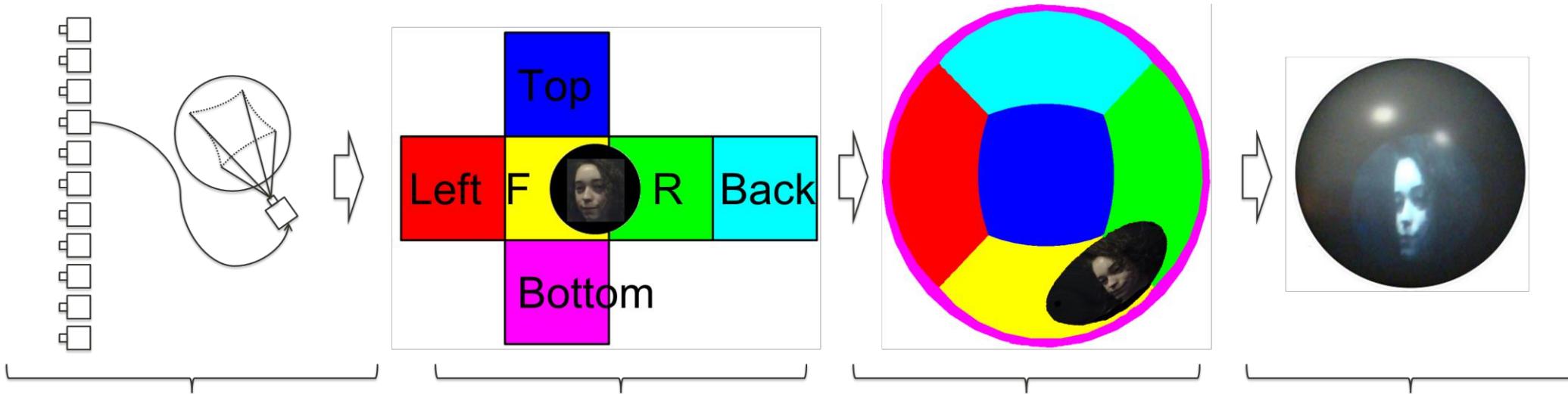
The remote person representation

Telepresence display	Spherical video display	Cylindrical multiview display	Random hole multiview display	Spherical avatar display
Photo				
Gaze	√	√	√	√
Multiple users		√	√	
Stereo views from arbitrary positions			√	
360° view	√	√		√

Spherical video-mediated display



Spherical video-mediated display



Stage 1:
Video from one camera texture-mapped on to a geometric sphere as a projective texture (using calibrated camera position)

Stage 2:
Sphere with texture is rendered into a cube map using six virtual cameras

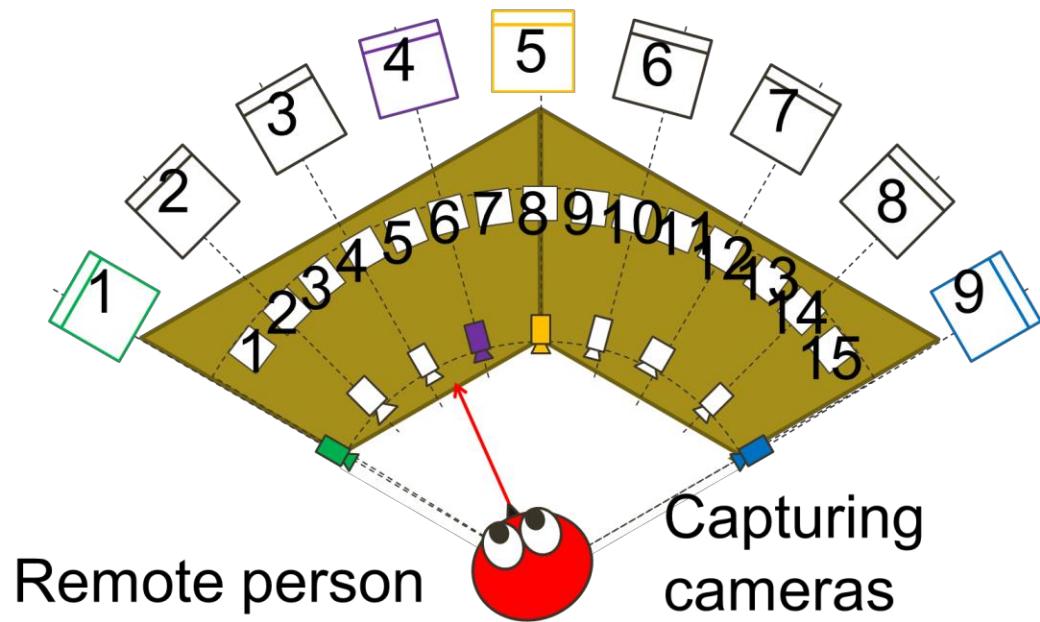
Stage 3:
Cube map is rendered as an environment map to creation distortion needed for fish-eye projection lens

Stage 4:
Once projected onto the display, the observer sees the head at approximately life-size

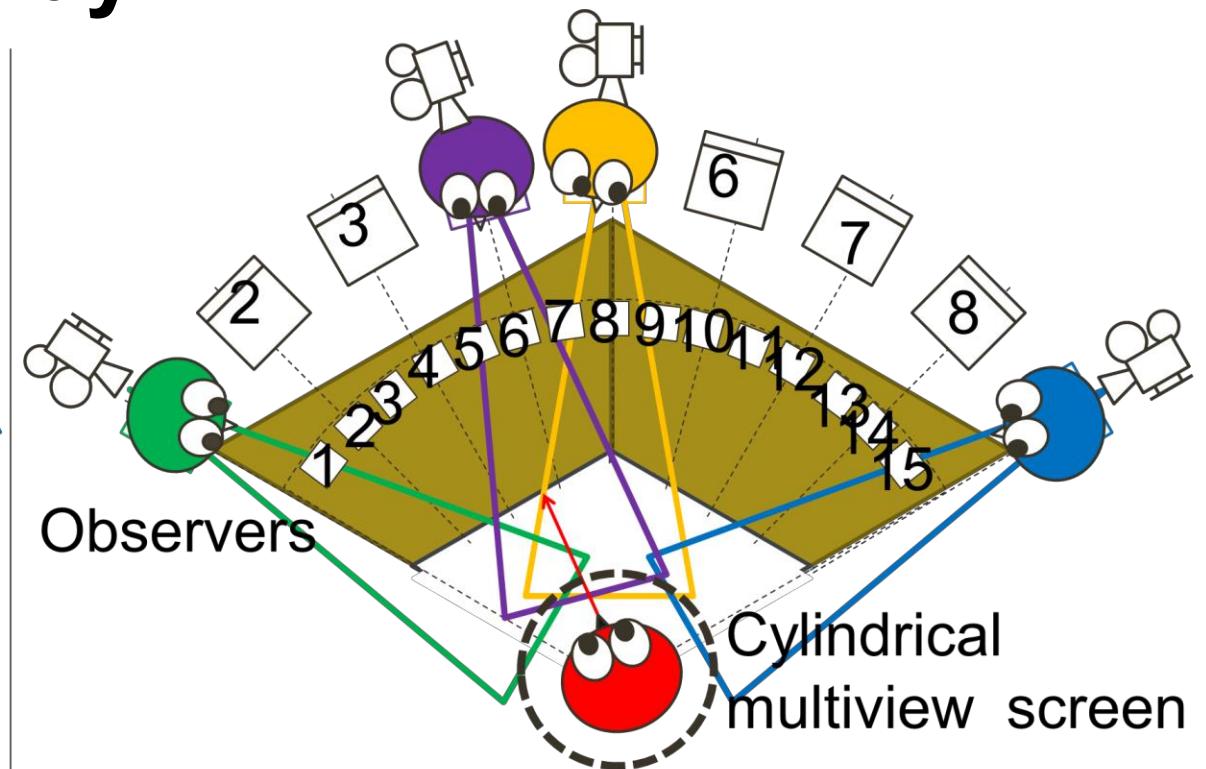
Spherical video-mediated display

- The spherical display offers a **360 degree view** whereas flat displays are only visible from the front.
- By using a surrounding camera array, we allow **principal observers** to **accurately** tell where the actor is **looking** from **multiple** observing positions.

Cylindrical multiview display



Video 1 Video 4 Video 5 Video 9



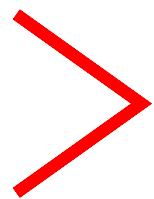
View 1 View 4 View 5 View 9

Cylindrical multiview display

Target error = | the observer's perceived target number - the actual target number|



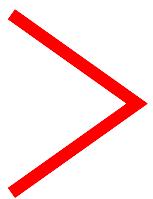
Cylinder
multiview
single-video



Flat
diffuse
single-video



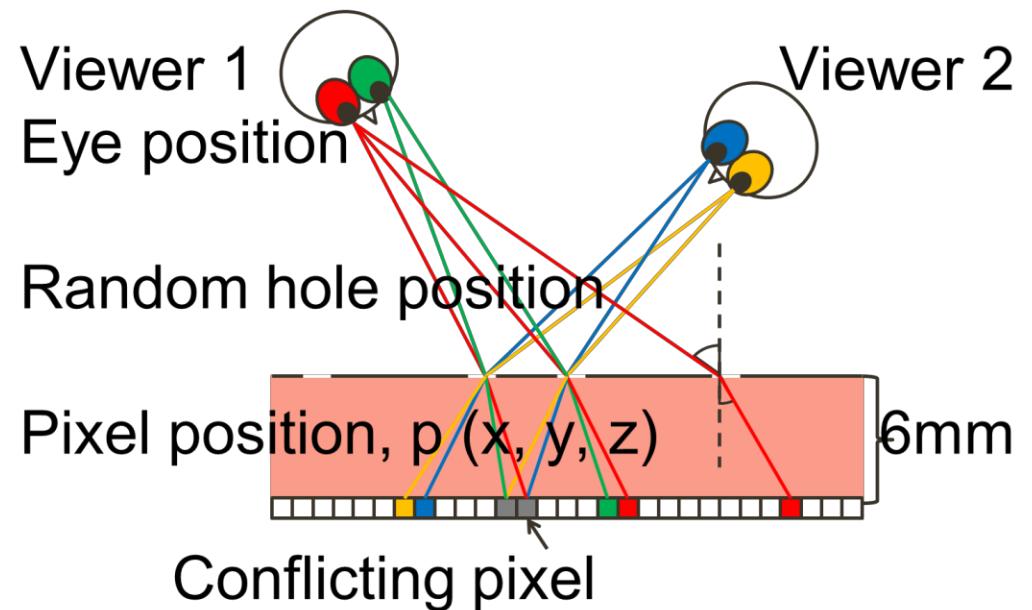
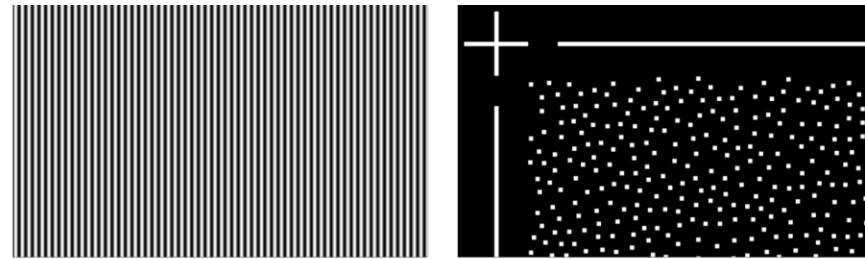
Cylinder
diffuse
single-video



Cylinder
multiview
multi-video

Random hole multiview display

- It allows **multiple users** from **multiple viewpoints** to **simultaneously** tell where the remote person is looking at
- Cheap and easy to configure
- Dense camera arrays that are further from the users



Random hole multiview display

- Supporting **multiple perspective-correct stereo** viewers in arbitrary locations



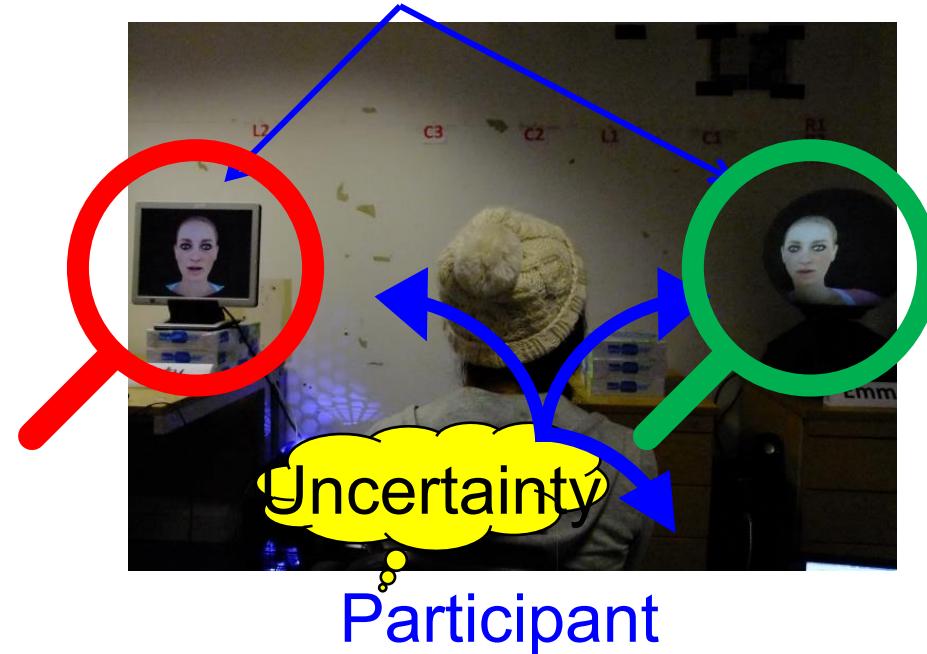
Random hole multiview display

- Motion parallax provides a dominant effect in improving the effectiveness with which users were able to estimate the gaze direction
- Additional effect for perspective-corrected augmented by stereoscopy.
- Simulating scenarios that require multiple simultaneous stereo views from arbitrary positions.

Trust: Advice seeking behavior

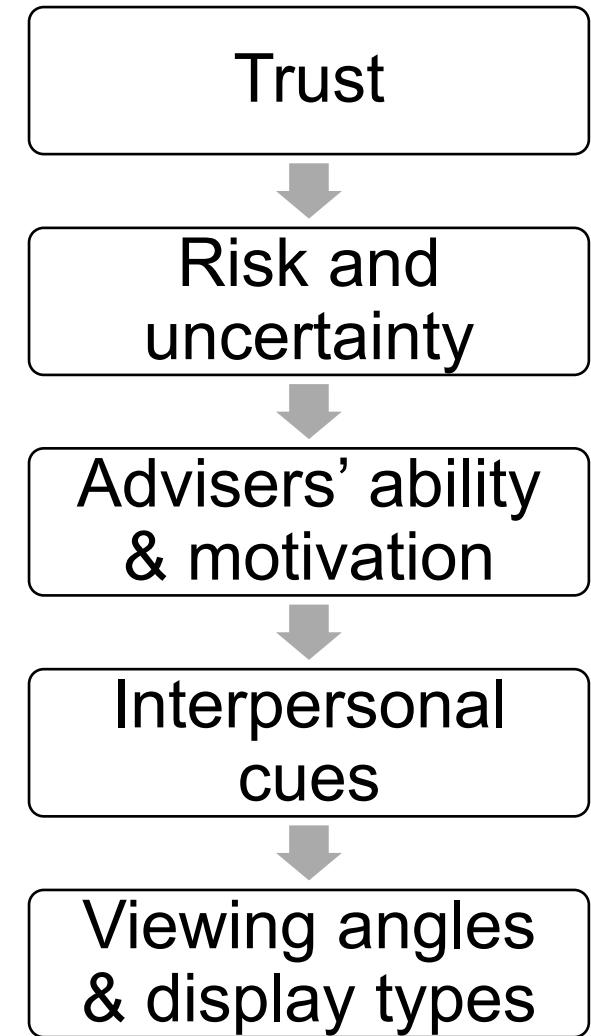


Adviser: expert vs. non-expert

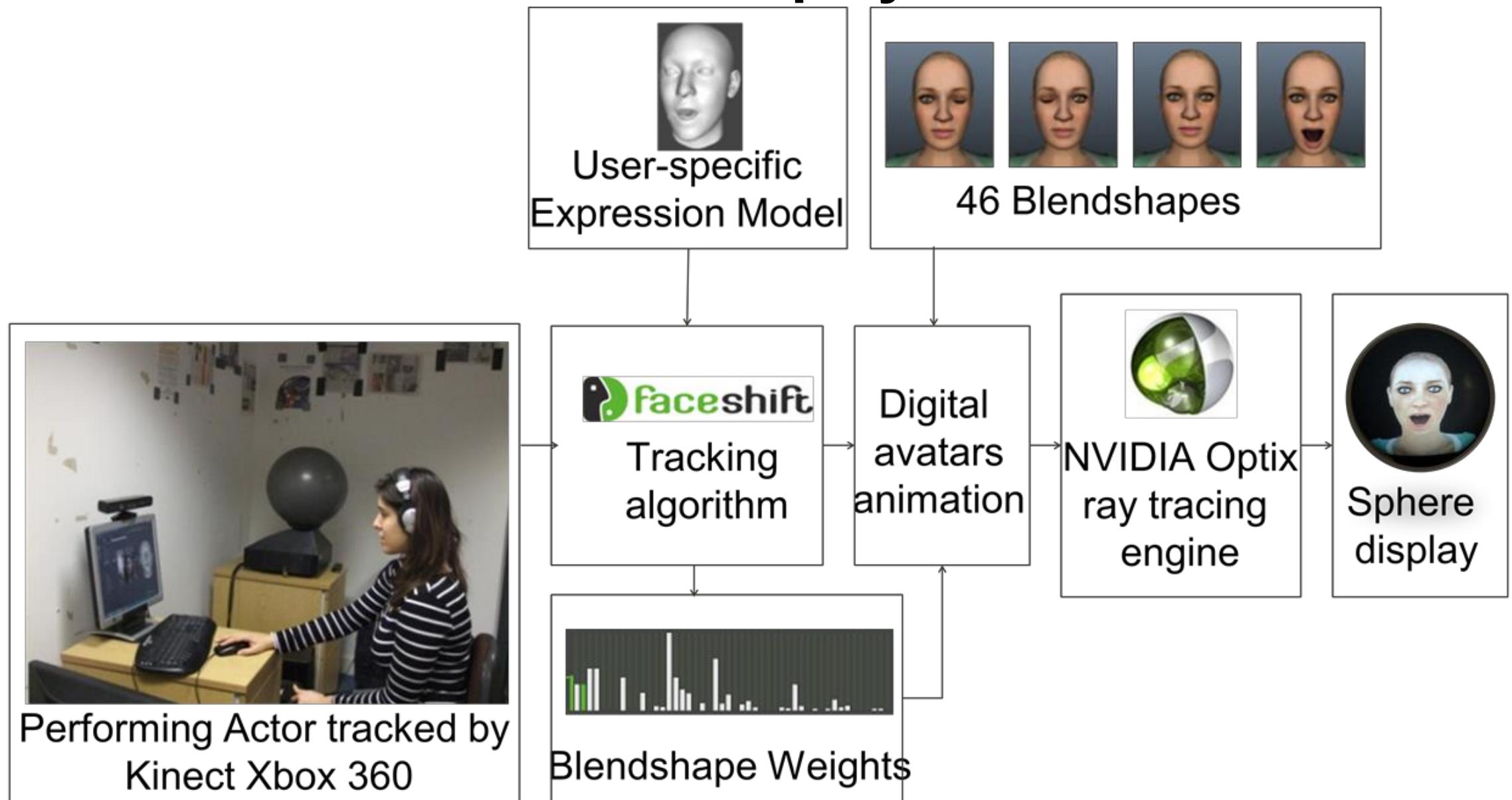


What subject did Bill Gates study?
Extremely difficult?

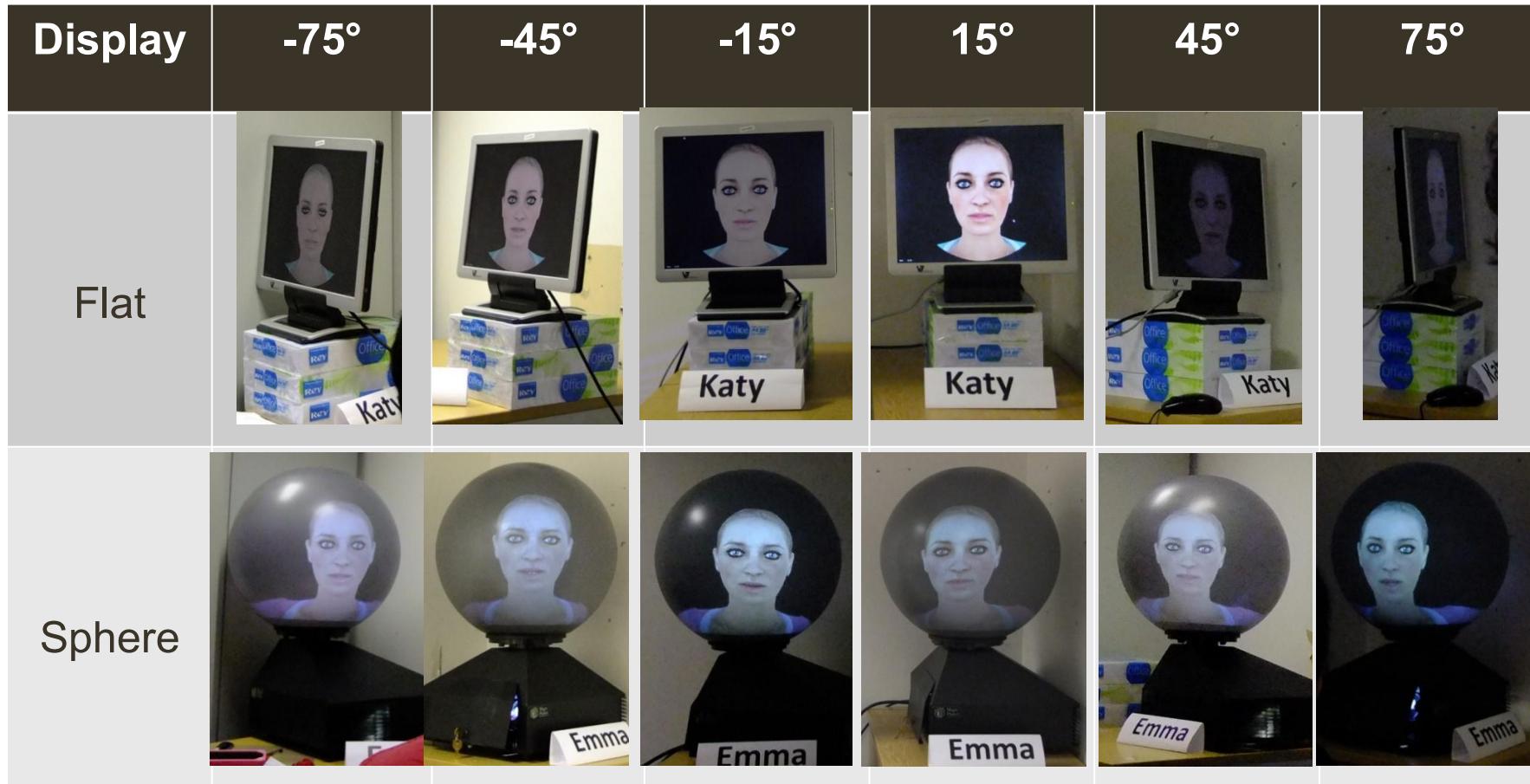
A. Law B. Medicine C. Computer Science D. Maths



Spherical avatar-mediated display



Spherical avatar-mediated display

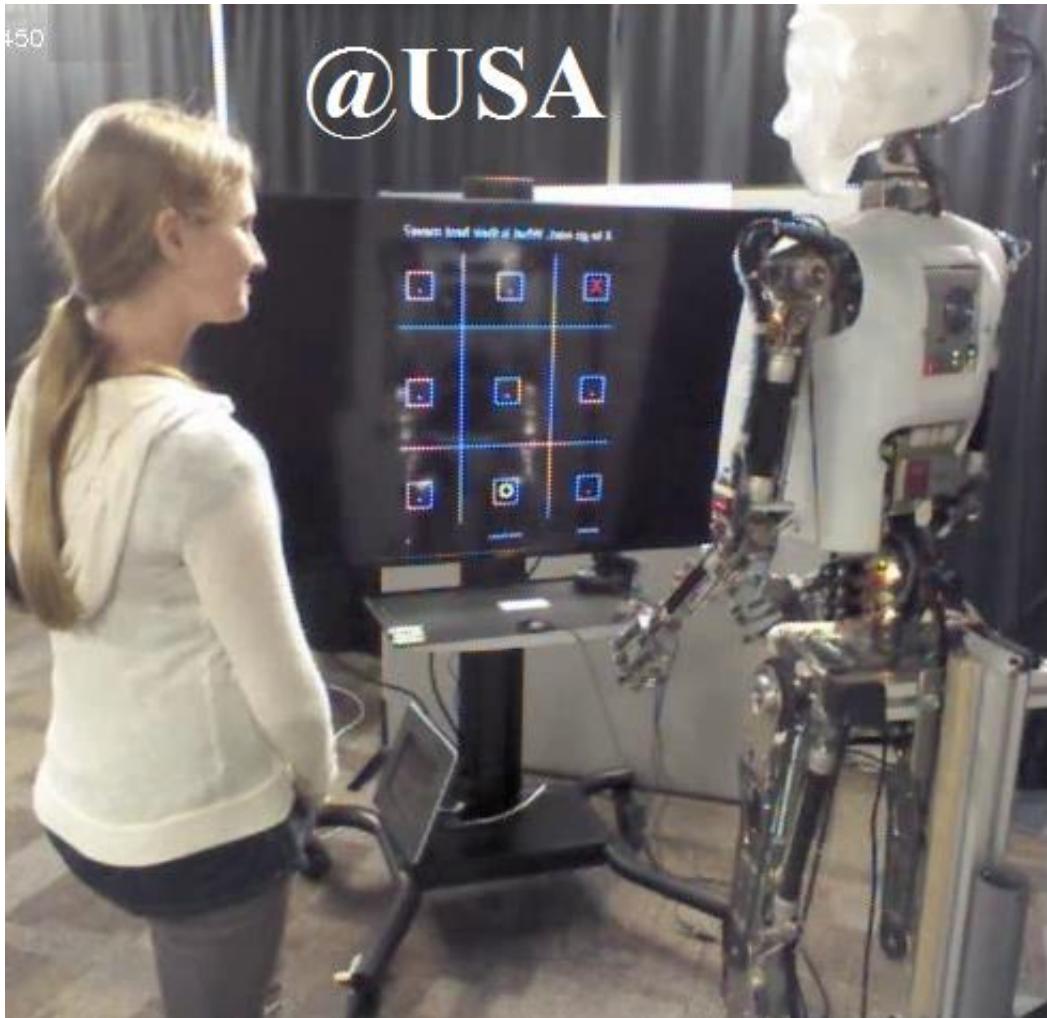


- If gaze is better preserved on a display (e.g. sphere display), the user may exhibit more trust towards avatars on that display, or show trust-related behaviours

Spherical avatar-mediated display

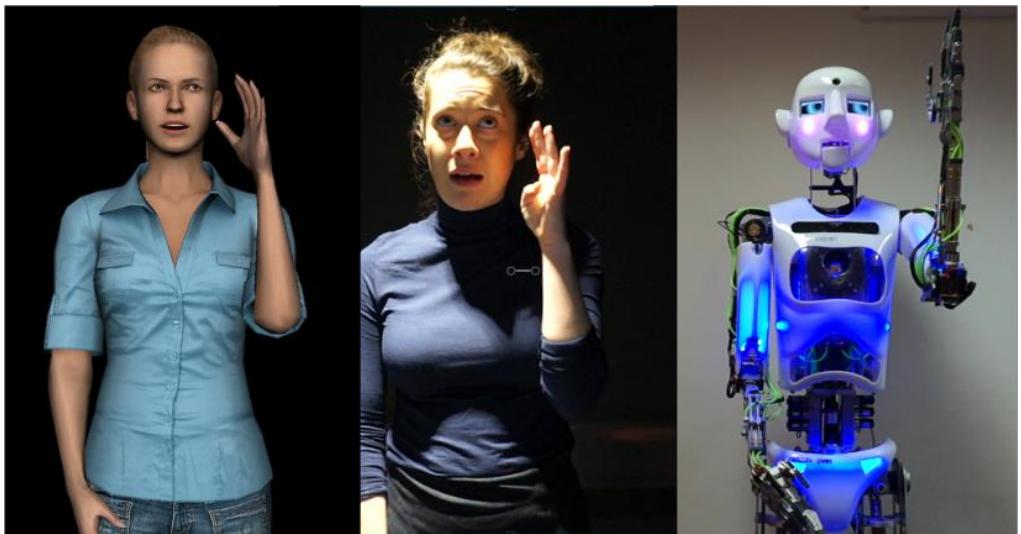
- We detail a method for enabling the displayed avatar to reproduce **the facial expression** captured from a person in real-time and with high-fidelity.
- Our display is small enough to situate almost anywhere in a room, and it is visible from **all directions**.
- Our view depending rendering method could be extended to other display systems that have a **three dimensional display surface**.
- Participants were able to discriminate trustworthy and less trustworthy advisers irrespective of display type
- A negative bias for flat display **can interfere** with users' ability to discriminate effectively
- Trust can be easily and significantly manipulated in mediated interaction by adjusting display viewing angle

Symmetric robotic system



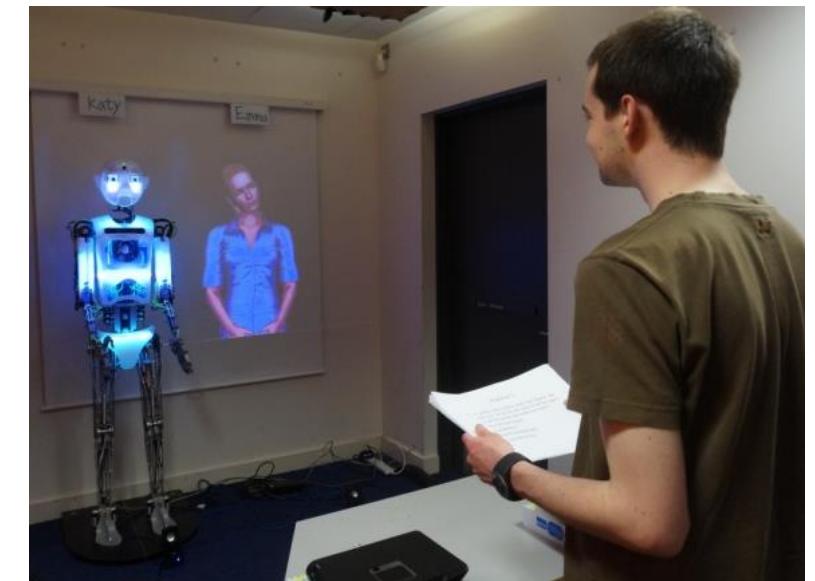
Humanoid robot

- Physical embodiment
 - Robot
- Physical contact
 - Robot
- Motion fluency
 - Avatar, video
- Masked identity
 - Robot, avatar
- 2D vs. 3D
 - Robot, avatar



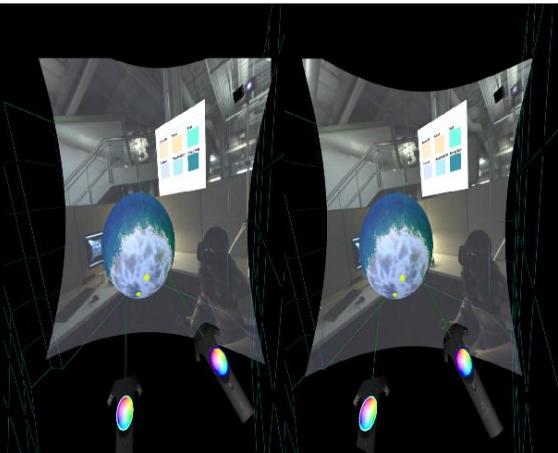
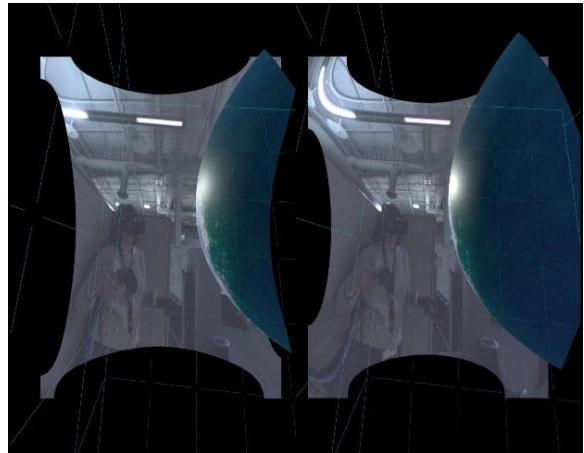
Symmetric robotic system

- Participants' advice seeking behaviour under risk as an indicator of their trust in the advisor.
- Users' trust assessments: robot = video > avatar
- The physical presence of the robot representation might compensate for the lack of identity cues.

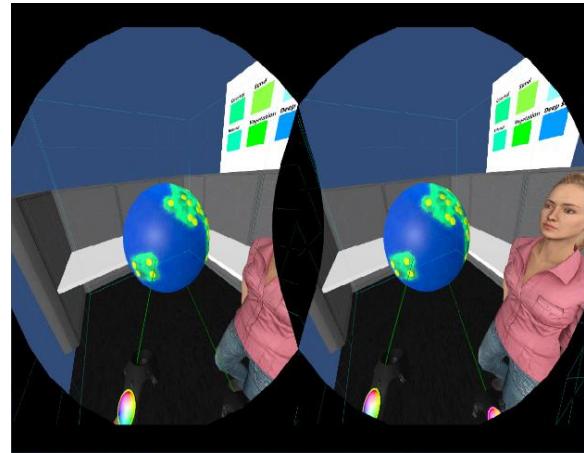
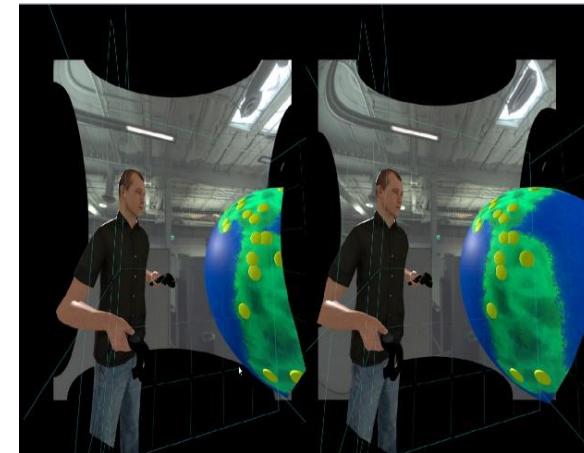


Leadership effect

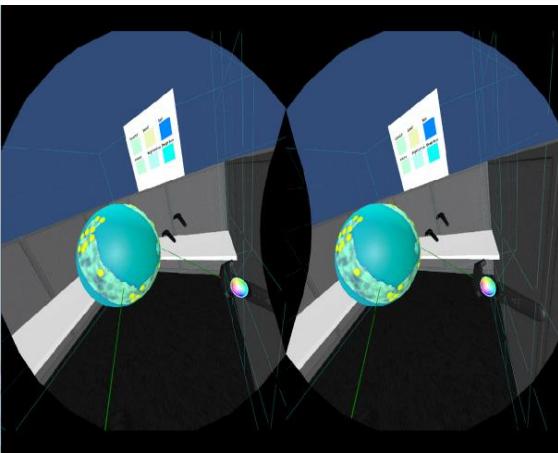
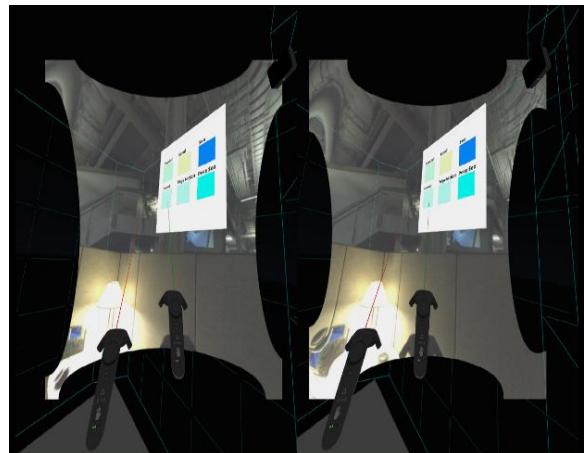
Each pair of screenshots was simultaneously captured from the first-person view



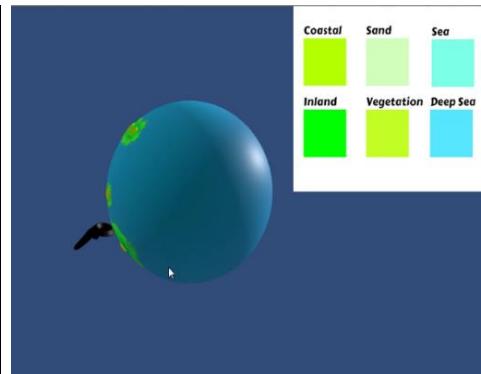
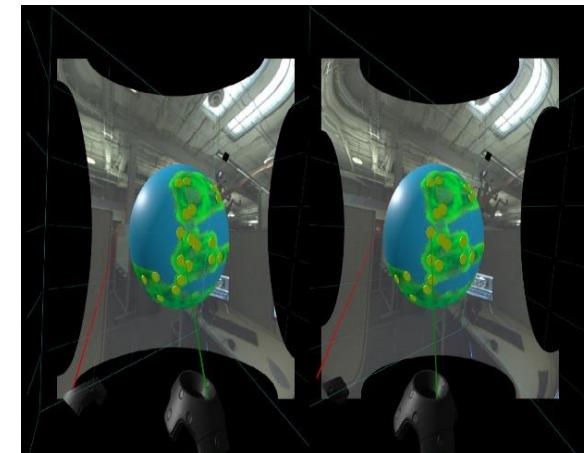
Augmented reality (AR)-to-AR



AR-to-VRBody



AR-to-virtual reality (VR)



AR-to-Desktop



Disney Research

Leadership effect

- The more immersed participant was singled out as the leader.
- Leadership effect: AR-to-desktop > AR-to-VR
- No Leadership effect: AR-to-VRBody & AR-to-AR
- Leadership effect only emerged in 3D interactions but not in 2D interactions.

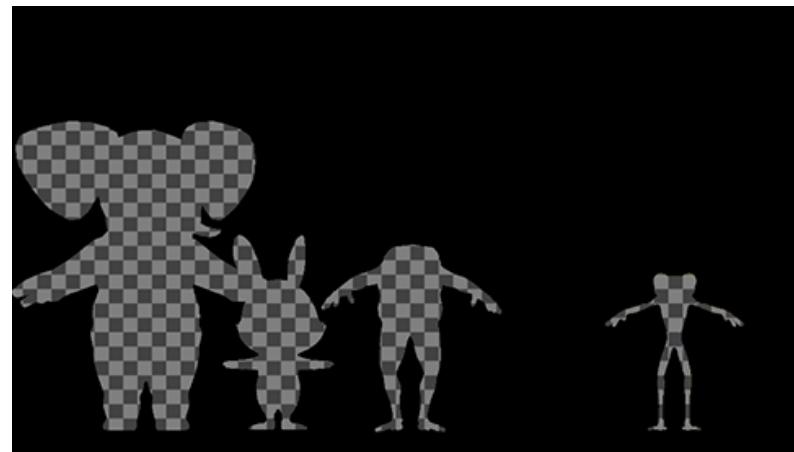
Adaptive Dynamic Anamorphosis System



(a)



(b)



(c)



(d)

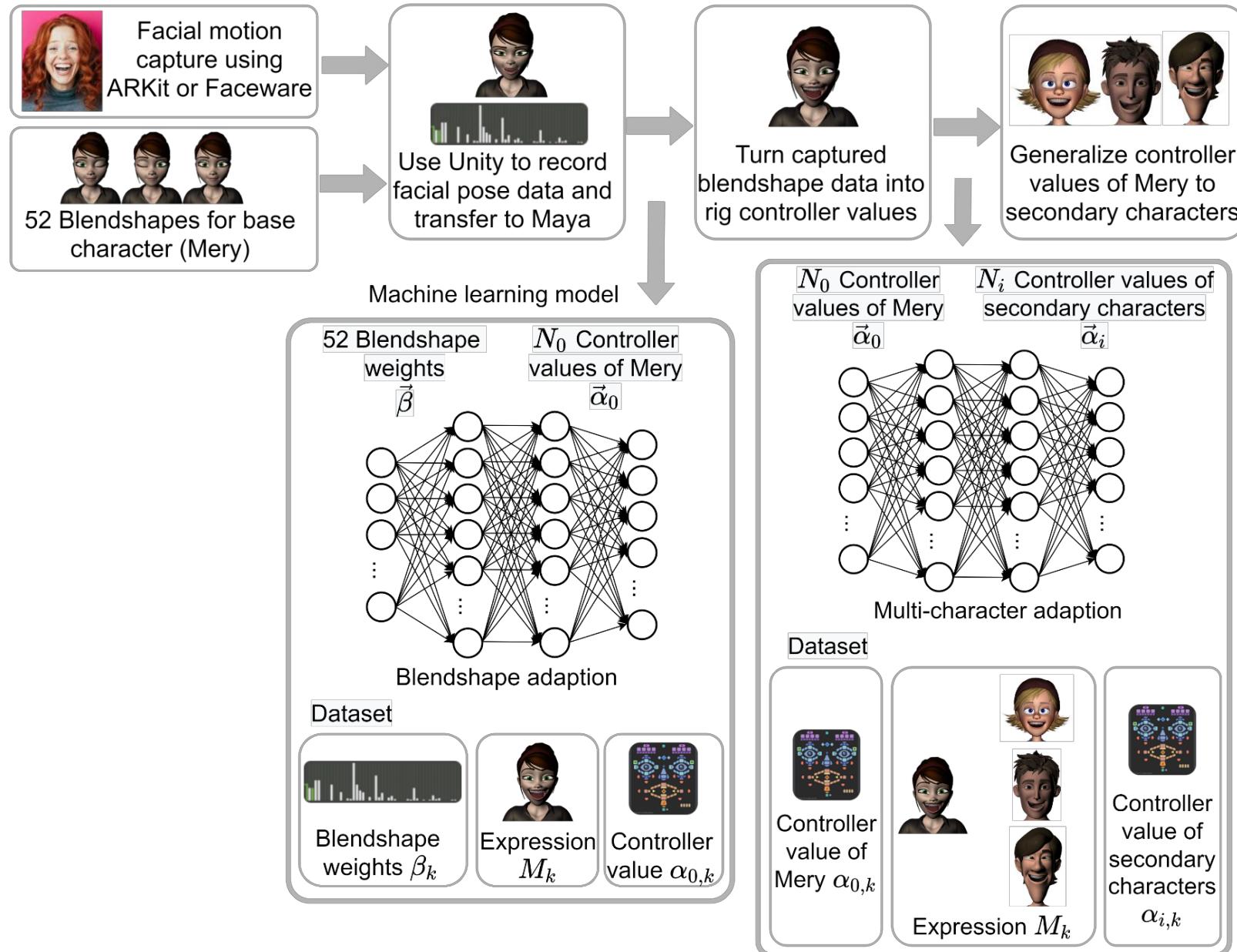
Composition of the adaptive dynamic anamorphosis, (a) Anamorphic projected image. (b) Normal perspective. (c) Mask used for composition. (d) Compositing (a) and (b) into a single resulting dynamic adaptive anamorphic perspective image.

Eye gaze & engagement

- We propose a dynamic adaptive anamorphosis method based on non-linear projections.
- Anamorphic rendering of a selective object with normal view rendering of the rest.
- The VIP guest results in an improved gaze and engagement estimation.
- This is performed without sacrificing the other guests' viewing experience.

Expression	Neutral	Joy	Surprise	Sadness	Disgust	Anger	Fear
Human							
Faceware							
MienCap primary character							
MienCap Secondary characters							

Facial Mocap with Live Mood Dynamics



MienCap:
Performance-Based
Facial Animation with
Live Mood Dynamics

網易 NETEASE

ROBLOX

Character Lab

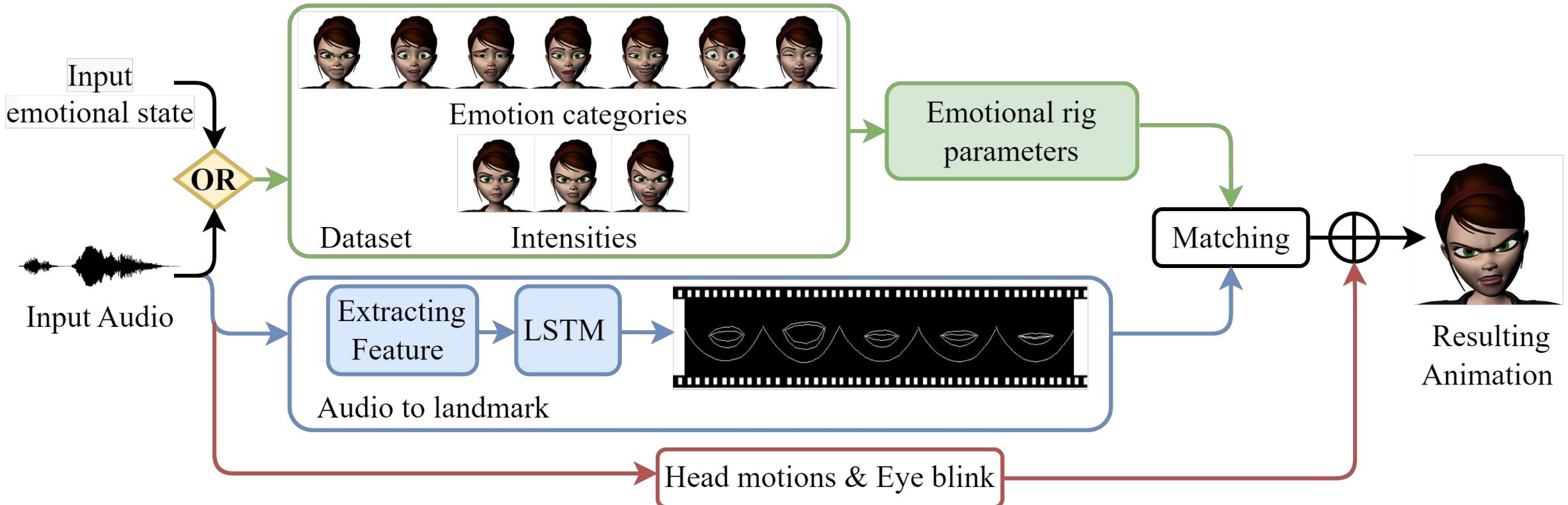
Emotion

- Blendshape animation techniques + machine learning models
- The first **real time** system transferring human facial expressions to multiple 3D stylized characters in a **geometrically consistent** and **perceptually correct** way.
- Creating a complete set of blendshapes for each new character needs a significant amount of manual effort, whereas MienCap can flexibly mapping the expressions from the existing character to a new one without the need of creating blendshapes.

Dataset



Emotional Voice Puppetry

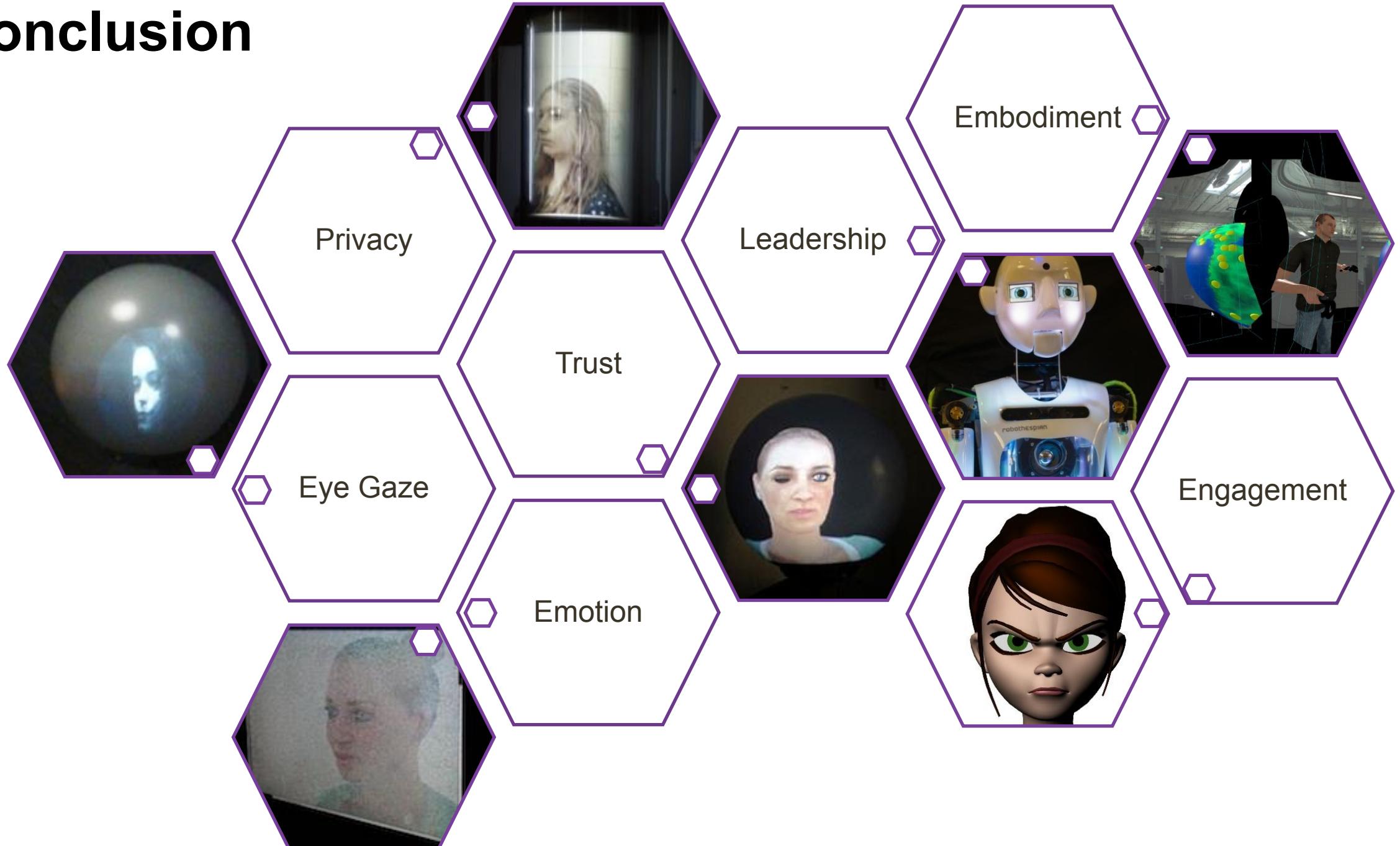


Conclusion



- We discussed a couple of **human factors** enhance telepresence, and insight into the understanding of how people behave and respond when engaged in these novel telepresence technologies.
- We built a series of **displays/systems** which could be used in future teleconferencing.
- Together these demonstrations motivate the further study of novel display configurations and suggest parameters for the design of future telepresence systems.

Conclusion



Selected Reference

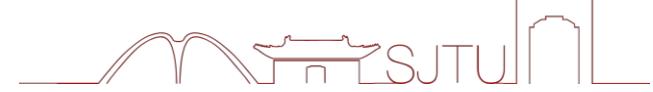


- Y. Pan, R. Zhang, S. Cheng, S. Tan, Y. Ding, K. Mitchell, and X. Yang. Emotional Voice Puppetry. *IEEE Conference on Virtual Reality and 3D User Interfaces*, 2023
- Y. Pan and K. Mitchell. Improving VIP viewer gaze estimation and engagement using adaptive dynamic anamorphosis. *International Journal of Human - Computer Studies*, 2021
- Y. Pan and A. Steed. Effects of 3D Perspective on Gaze Estimation with a Multiview Autostereoscopic Display. *International Journal of Human - Computer Studies*, 2016
- Y. Pan and A. Steed. A gaze-preserving cylindrical multiview telepresence system. *ACM CHI Human Factors in Computing Systems, Toronto, Canada, April 26-May 1, 2014*
- Y. Pan, W. Steptoe and A. Steed. Comparing flat and spherical displays in a trust scenario in avatar-mediated interaction. *ACM CHI Human Factors in Computing Systems, Toronto, Canada, April 26-May 1, 2014*



上海交通大学

SHANGHAI JIAO TONG UNIVERSITY



Thanks!

Ye Pan

whitneypanye@sjtu.edu.cn

饮水思源 · 爱国荣校



THE PREMIER CONFERENCE & EXHIBITION ON COMPUTER
GRAPHICS & INTERACTIVE TECHNIQUES



SIGGRAPH 2023
LOS ANGELES+ 6-10 AUG

Project Starline: A high-fidelity telepresence system

Dan B Goldman, Google Research

© 2023 SIGGRAPH. ALL RIGHTS RESERVED.

My name is Dan Goldman, and I'll be presenting Project Starline, a high-fidelity telepresence system developed by Jason, myself, and many others at Google.



[Maimone and Fuchs 2011]

There has been a considerable amount of work in this area over the last 30 years, and prior research systems have been described that also use free-standing displays and depth sensors. Our work builds upon these earlier approaches, offering a fully symmetric and networked system that achieves an unprecedented level of video and audio fidelity.

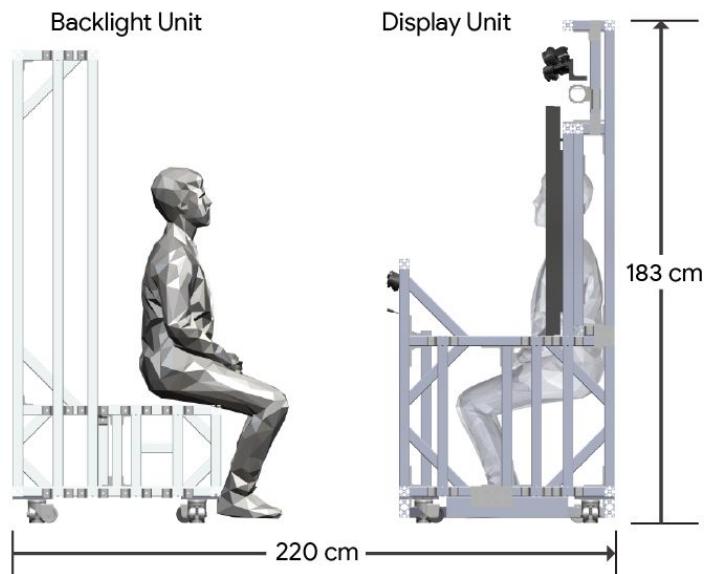


[Wei et al. 2019]

More recent work has investigated head-mounted displays for telepresence applications, including the Holoportation project from Microsoft Research that Jason mentioned, and the codec avatars work being pursued at Meta Reality Labs.

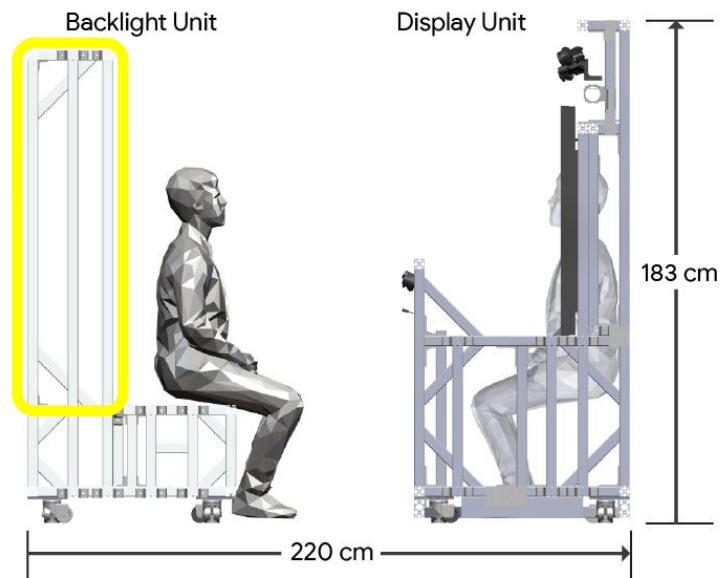
Although this direction shows a path to eventual low-cost and portable systems, we think a key benefit of our approach is that it enables a fully unencumbered user experience – the participant doesn't have to wear anything or perform any pre-meeting data captures.

Also, our display and setup achieve a retina resolution of 45 ppd across the target field of view, showing the other person in greater detail than what is possible with today's state-of-the-art VR headsets and AR glasses.



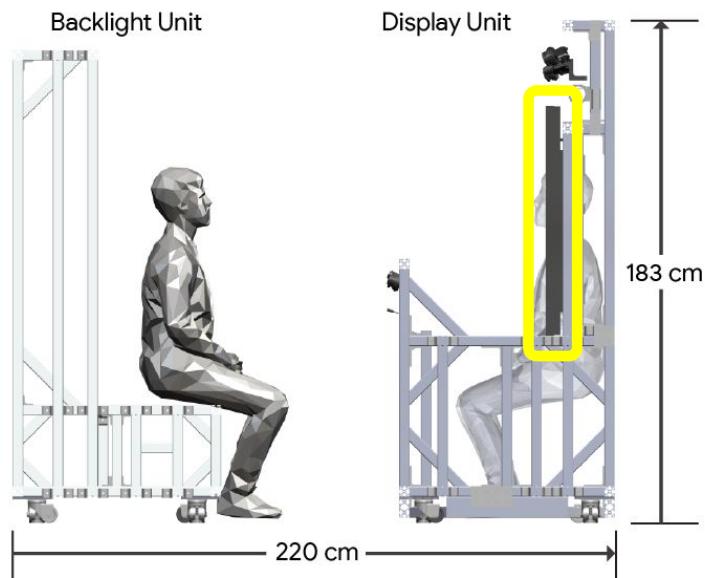
Let me start with a high-level overview of our design.

Our system is intended to be used by a single person at a time.



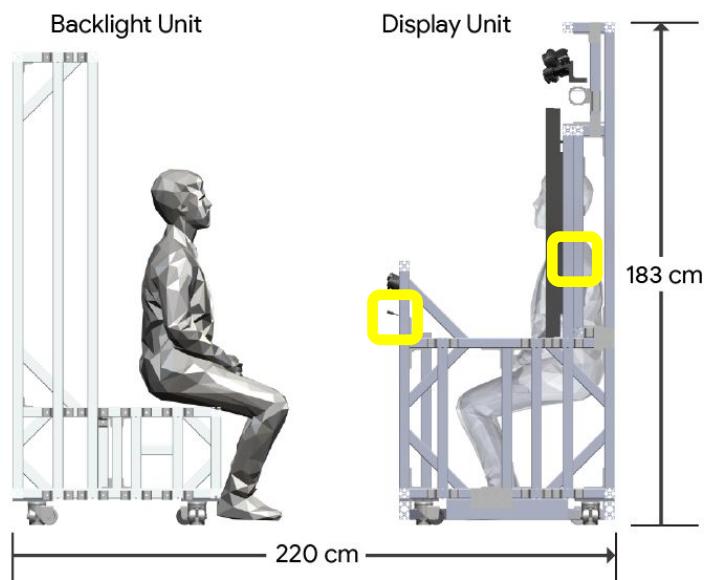
They sit on a bench that is connected to a large infrared backlight located directly behind them.

They see their remote conversation partner...

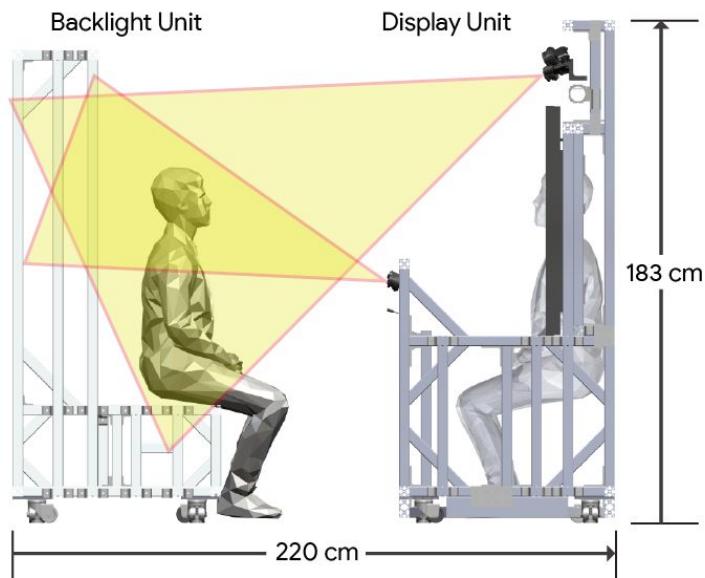


...through a 65" autostereoscopic display, located roughly 1.2 meters in front of them.

This display conveys both stereo and parallax cues to the seated participant and shows the remote person at their true physical size.



The system also produces 3d spatialized audio that appears to emanate from the remote person's mouth.



A 3d video of the participant is reconstructed at 60 frames per second from two camera viewpoints located above the display and one viewpoint located in a central "middle wall".

This middle wall also serves to hide the bottom edge of the display to avoid depth conflicts that would otherwise appear whenever the remote person's body at that edge extends beyond the plane of the display.



The combined effect is that each participant can see and hear the other person as they truly are, within a headbox of roughly 1m-cubed centered about 1.2m in front of the display.

Key non-verbal cues like eye contact and hand gestures are all intuitively conveyed.

Proprietary + Confidential



Next, I'll walk through the main technology components that power this experience in a bit more detail.

Autostereo display



Proprietary + Confidential

We designed our system around a 60Hz 65-inch 8K autostereoscopic display. For a typical observer sitting 1.25 m away, the lens array presents each eye a separate subset of the display pixels - about 5M pixels of each red, green, and blue primary - providing angular resolution of about 45 pixels per degree. This is more than twice the effective resolution of the Meta Quest Pro, with 22 pixels per degree. (Apple hasn't disclosed the PPD of their new Vision Pro yet.)

High frame-rate face tracking cameras



Proprietary + Confidential

To steer this display to the viewer's eyes, we use four face-tracking cameras running at 120 frames per second. These estimate the 3D location of the eyes, ears, and mouth within about 5 millimeters of precision.



We use a fast face tracker to find 2D facial features, and triangulate to find the 3D locations. These are used to render the appropriate viewpoints, to steer the 3D display, and to drive free-space spatialized 3D audio.

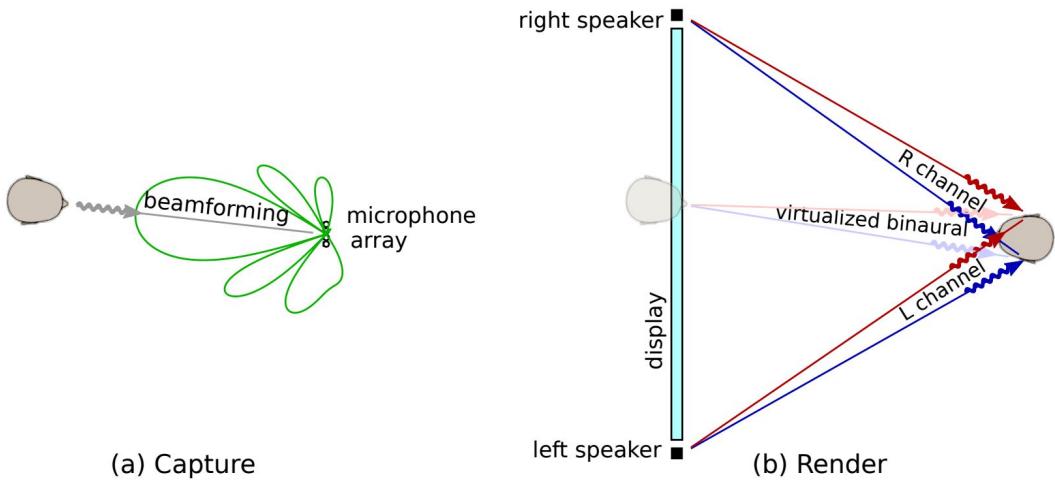
Spatialized audio: speakers and microphone array



Proprietary + Confidential

The spatialized audio system uses two speakers and an array of 4 microphones. Together these can capture and reproduce speech as if it came from the mouth of the other participant.

Freespace spatialized audio



Proprietary + Confidential

On the input side, face tracking data enables dynamic beamforming, sharpening the microphone's directionality to combat noise and reverberation.

On the output side, tracking enables the system to spatialize playback at the location of the speaker's mouth, and binaural crosstalk cancellation to target the correct waveforms at the listener's ears.

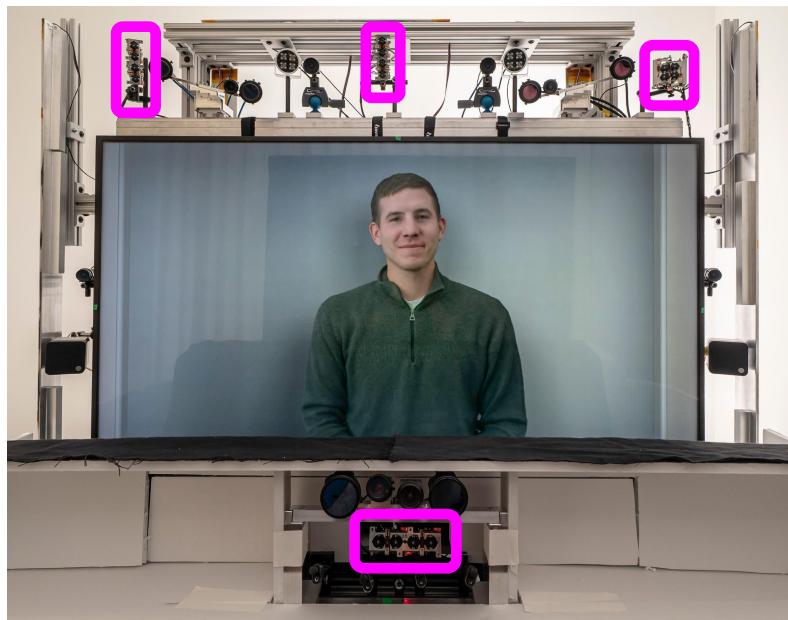
Thus, even though the speakers are spaced far apart, the sound appears to emanate from the remote user's mouth.

**ESPReSSo
RGBD “pods”**



To capture a 3D video of the subjects, we use 3 groups of cameras we call pods, each with two infrared cameras and one color camera. The bottom pod contains an extra color camera zoomed into the face for higher resolution there.

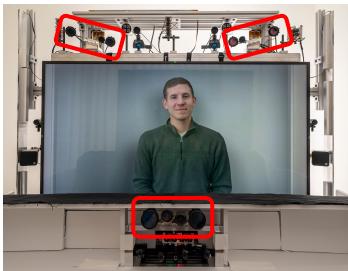
**IR projectors
(DOE)**



Proprietary + Confidential

For stereo reconstruction, we use time-varying infrared pattern generators, that create dot images only visible in infrared.

Stereo capture



Four patterned IR images and IR guide image

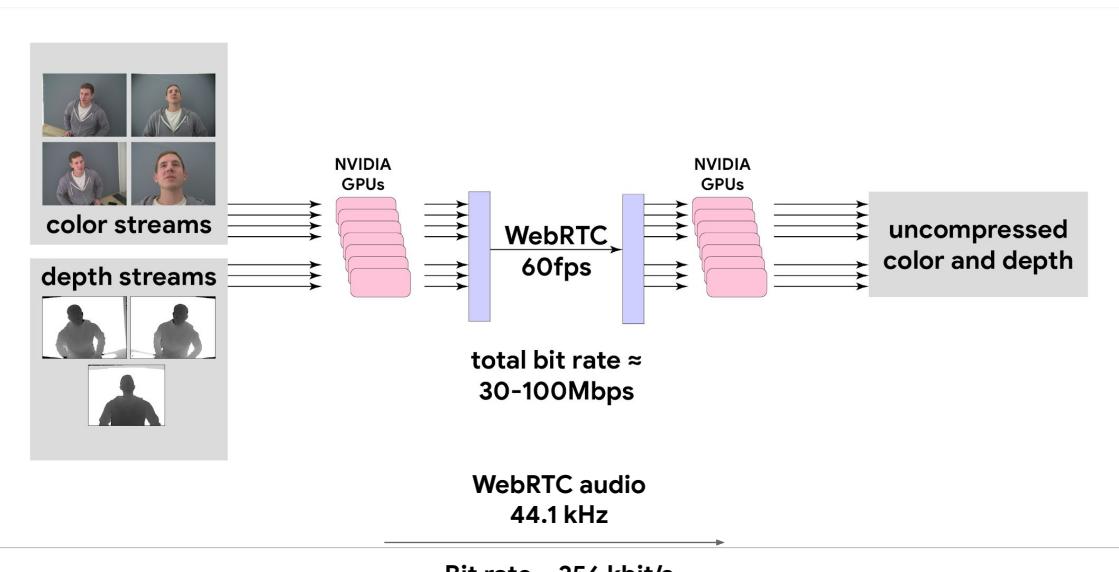
Background subtraction using IR backlight

We use windows of five infrared image pairs - four with dot patterns, and one with an infrared backlight - to compute depth from spacetime stereo, using the ESPReSSo algorithm we shared at 3DV 2018.

This algorithm takes as input infrared image pairs at 180 fps and computes synchronized output depth images at 60 fps.

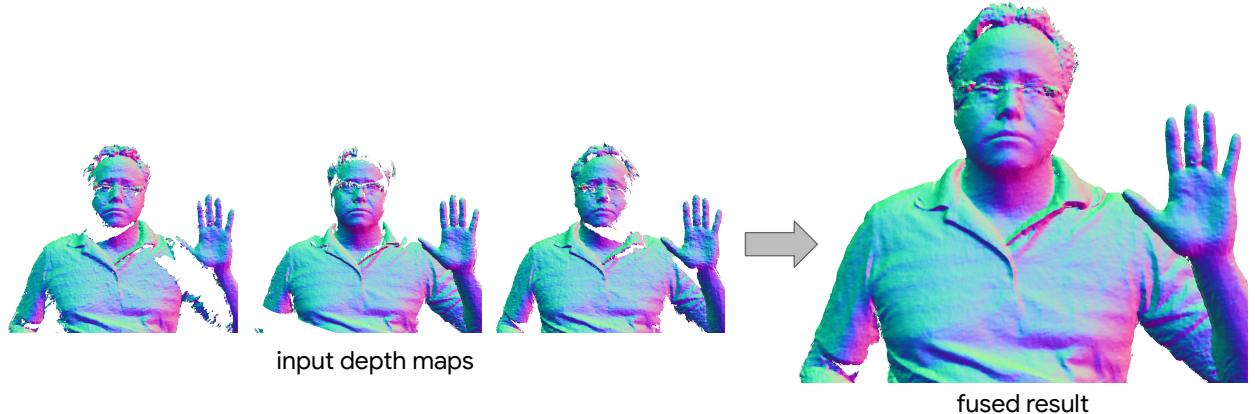
The infrared backlight is used to carve noisy background data from the stereo images and provide a reliable boundary for stereo estimation, improving accuracy at silhouette edges, as you can see in this result.

Streaming compression and transmission



In total, 3 depth, and 4 color streams are sent over WebRTC, using GPU video codec hardware.

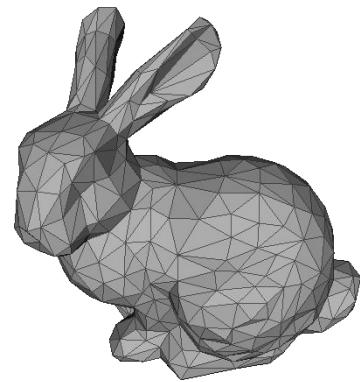
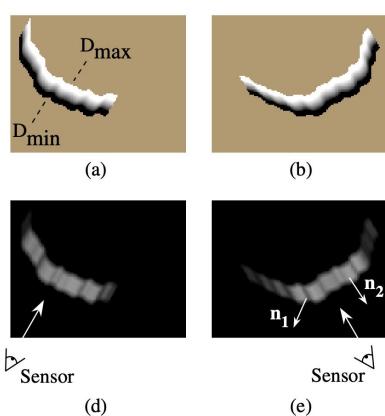
Image-based depth fusion



Proprietary + Confidential

On the receiving side, after decompression, the system reprojects 3 depth images to the local subject's eye positions.

Volumetric fusion and rendering [Curless and Levoy 1996]



1. Accumulate weighted SDFs

2. Extract isosurface

3. Render triangles

Proprietary + Confidential

A traditional volumetric fusion system would take these three depth images and fuse them in a voxel representation,

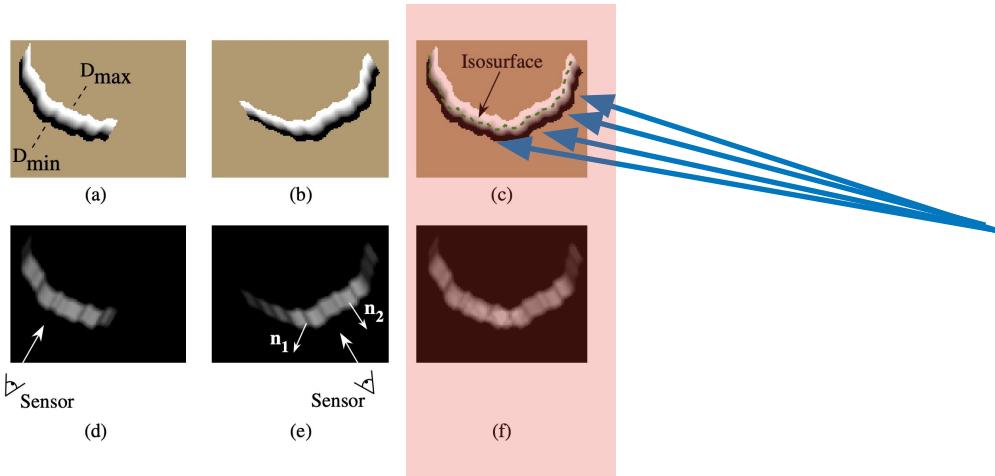
—

extract the isosurface using marching cubes, and

—

render the triangles.

Volumetric fusion and raycasting



1. Accumulate weighted SDFs

2. Ray-cast isosurface

Proprietary + Confidential

We can improve a little bit using modern GPU hardware, by eliminating the surface extraction step, and just raycasting the voxels directly. This eliminates the additional data structures and unpredictable memory usage of a triangle mesh.

—

However, it still requires a lot of GPU memory to store the voxel grid, and a lot of memory bandwidth for the raycasting kernel to retrieve it.

Volumetric fusion and raycasting

1. Voxelize weighted SDFs

```
Loop over every voxel
  Loop over every input depth image (3)
    Project voxel position into the image
    Accumulate SDF and weights
```

2. Ray-cast isosurface

```
Loop over output pixels
  Step through voxels along the eye ray
    Check SDF to determine step size
    Stop when the surface is reached
```

Proprietary + Confidential

By examining the pseudocode for this two-pass algorithm, we can make some observations.

—

First, it's making two passes over most voxels, just in a slightly different order.

—

And second, we only fuse three depth images for each voxel, so the inner loop here is very fast.

Image-based fusion and raycasting

Loop over output pixels

Step through voxels along the eye ray

Loop over every input depth image (3)

Project voxel position

Accumulate SDF and weights

Check SDF to determine step size

Stop when the surface is reached

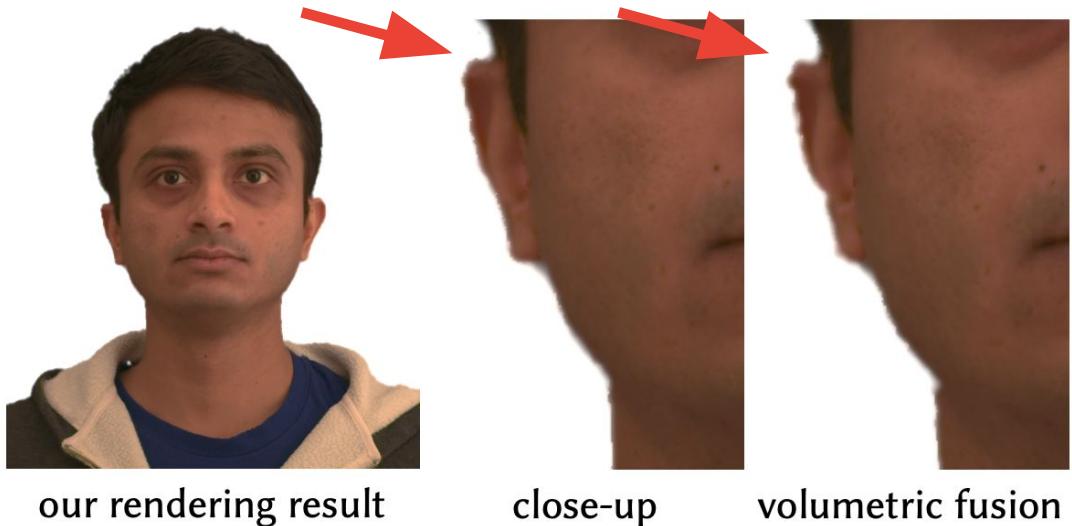
Proprietary + Confidential

Using these observations, we can interleave the fusion and raycasting passes into a single kernel,

-

fusing the depth images on the fly as we step through rays. This eliminates the need to store a voxel grid in GPU memory, dramatically reducing memory usage, and improving runtime by a factor of 6 over separate fusion and raycasting kernels.

Image-based fusion and raycasting



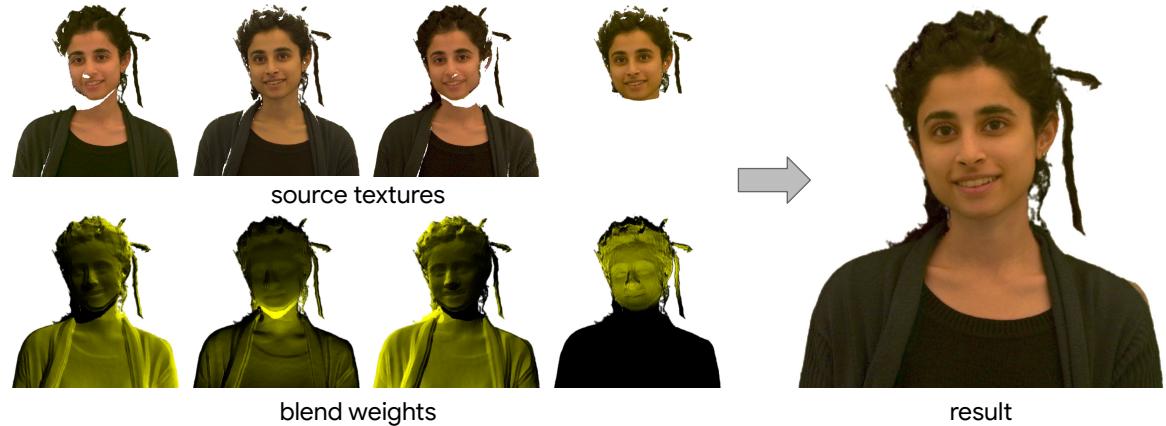
Proprietary + Confidential

By eliminating the need to sample on an arbitrarily aligned voxel grid, it can also

—

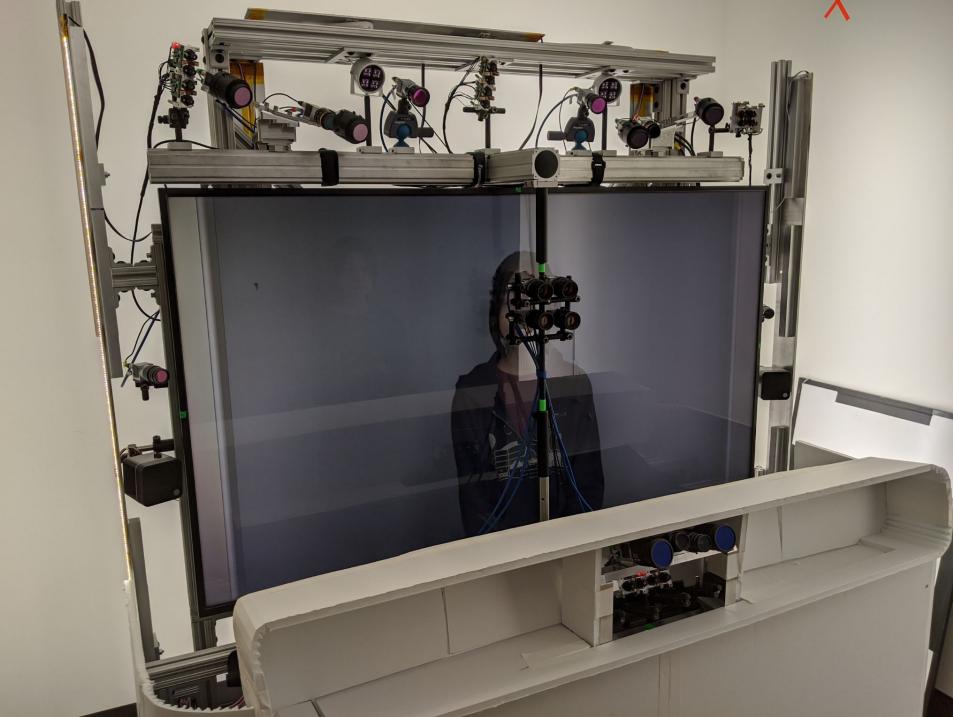
reduce aliasing artifacts, as seen along the silhouette here.

Texture mapping



Proprietary + Confidential

Next, we project the color camera images onto the fused geometry, and combine the colors using blend weights calculated from surface normals.



We can evaluate the quality of the image reconstruction by placing cameras in front of the display along the remote user's line of sight, and then render a reconstruction from the same viewpoints.

Real image



Proprietary + Confidential

Here's what the real image looks like...

Rendering

Proprietary + Confidential



...and here's our reconstruction, matted on a neutral gray.

Although you can see some differences in the background shadows, and some holes around the waist and hands, the majority of the torso and face are reconstructed with extremely high fidelity.

Here again is the real image....

And here is the reconstruction.

Real image



Proprietary + Confidential

Here are more comparisons illustrating the high reconstruction fidelity of this system.

Rendering



Proprietary + Confidential

Real image



Proprietary + Confidential

Here are more side by side examples illustrating the high reconstruction fidelity of this system.

Rendering

Proprietary + Confidential



Real image

Proprietary + Confidential



Our system doesn't work equally well for all scenes.

Rendering

Proprietary + Confidential



Notice that thin or frizzy hair is not well-reconstructed, as it falls below the resolution of our stereo system.

Real image



Proprietary + Confidential

Similarly, fast motion can break up the reconstructed geometry...

Rendering



Proprietary + Confidential

...resulting in holes and incorrect texture projections.



Likewise, eyeglasses also have thin geometric features and transparent surfaces that are missed by our 3D capture, causing incorrect texture projections.

Despite these limitations, Project Starline conveys a strong sense of remote co-presence, which we studied extensively.

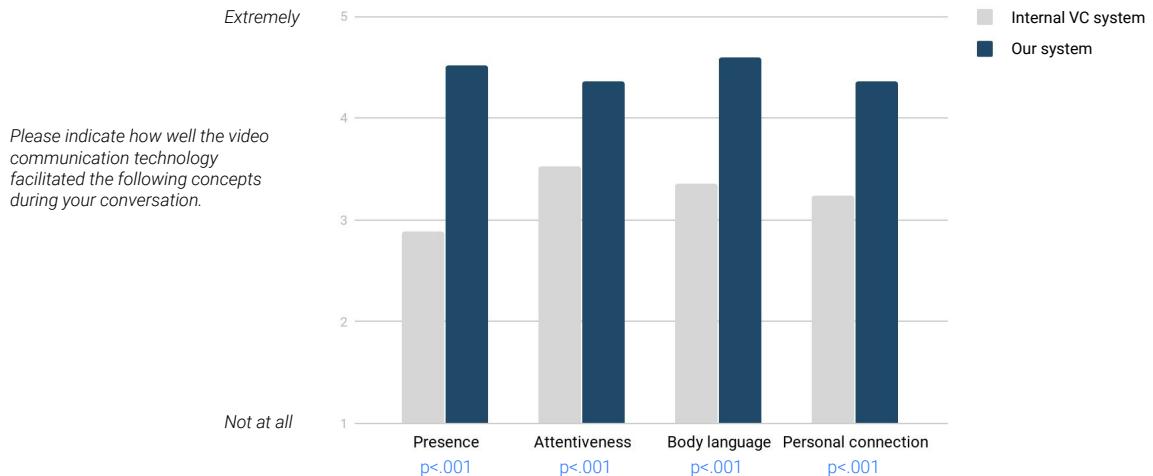


We evaluated our system in several different ways.

First, we used them for our own team 1:1 meetings over the period of about one year, connecting our team members in Seattle, Mountain View, and New York.

People reported experiencing a strong sense of being together with the other person, of occupying a shared physical space, and many commented that they recalled the exchange later as if it had happened in person.

We also conducted a pilot study with several other teams inside Google, who had access to these systems for their team 1:1 meetings, and were asked to complete a survey at the conclusion of each meeting comparing their experience to the normal 2D videoconferencing system deployed throughout the company.



117 participants, across 3 sites (some separated by >1,000 km), over period of 9 months

308 total meetings with average duration of 35.2 minutes ($SD = 16.7$ minutes)

296 post-meeting survey responses

All analyses between conditions use the Wilcoxon-Pratt Signed Rank Test, which accounts for skewness and ties (i.e., people who rate two conditions the same) in the data.

We collected a total of roughly 300 survey responses from 117 unique participants across 3 sites over a period of 9 months.

Our system was rated significantly higher in key value indicators for communication including: the sense of feeling present with your meeting partner, the degree of attentiveness between the two participants, the ability to effectively use body language, and producing an overall stronger personal connection.

Outperforming 2D videoconferencing is more challenging than it sounds for several reasons.

First, 2D video is highly realistic, whereas existing real-time 3D capture technologies are all known to suffer visual artifacts, putting them at an inherent disadvantage.

Second, compared to 2D displays, autostereoscopic displays introduce quality trade-offs such as lower resolution, tracking latency, or accommodation-vergence conflict, which degrade the experience for many viewers.

The fact that our system shows statistically significant user preference despite these challenges is noteworthy.

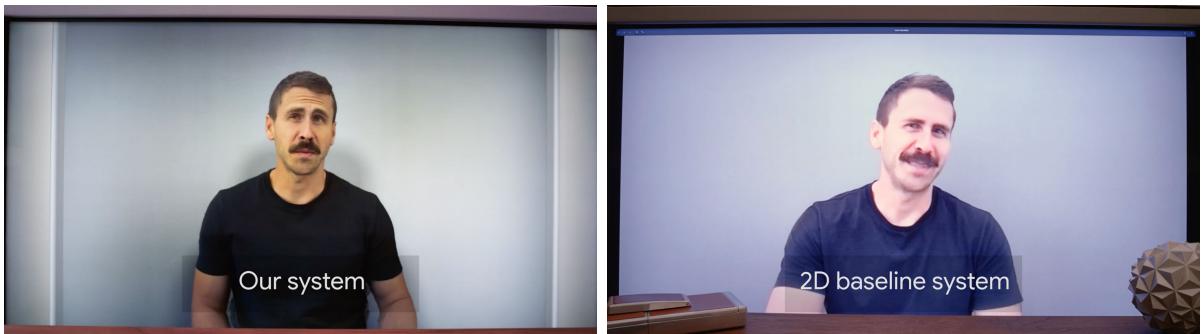


We also evaluated our system by measuring the frequency of specific non-verbal behaviors that are known to indicate effective meeting dynamics. We compared our system to a 2D baseline system with comparable image resolution and display size.



We observed statistically significant higher rates of important non-verbal cues like hand gestures, head nods, and eyebrow movements.

This finding suggests a greater level of engagement and sense of co-presence between the two participants.



Participants wrote more words after their conversation within our system ($M = 57.3$, $SD = 30.8$) compared to the 2d baseline system condition ($M = 44.8$, $SD = 24.4$), **suggesting that they recalled roughly 28% more meeting content** ($W-P$ statistic = 1.93, $p = 0.053$).

Last, we also observed a significant difference in the number of words people would use to recall what they had discussed in these two different conditions.

Our findings suggest that people recall roughly 28% more of their meeting content with our system.

Please refer to our paper for the details of these different studies, along with an audio realism study that we also conducted.



Project Starline

We have been working on this technology now for several years. Two years ago, Google announced this effort publicly at our I/O conference as Project Starline.



This year, we've started sharing our efforts to reduce the size, cost and complexity of the system, to make it available to even more trusted testers in the coming months. Our latest prototypes are powered by AI, enabling the same stunning experience using only standard color cameras. Here's a sneak peek at what these look like.

—

Looking ahead, we are excited about expanding access to Project Starline with these new systems!

→ THANKS!



SIGGRAPH Asia 2021 paper co-authors...

Jason Lawrence	Andy Huibers
Supreeth Achar	Claude Knaus
Gregory Major Blascovich	Brian Kuschak
Joseph G. Desloge	Ricardo Martin-Brualla
Tommy Fortes	Harris Nover
Eric M. Gomez	Andrew Ian Russell
Sascha Häberling	Steven M. Seitz
Hugues Hoppe	Kevin Tong

...and more!

Google
The entire Project Starline team
Clay Bavor
Andrew Nartker
Matthew DuVall
TJ Hayes
Jeff Prouty
Melba Tellez
Chad Lancaster



46

I'd like to thank the many people at Google who made this work possible.

State of the Art in Telepresence: Perception of Virtual Humans

Rachel McDonnell
Trinity College Dublin
rachel.mcdonnell@tcd.ie



Figure 1: Video call between participant (*left*) and experimenter (*right*), both embodied in virtual avatars and viewed side-by-side on-screen [Higgins et al. 2021b].

ABSTRACT

In this section, the focus will be how the human is represented within the shared virtual space, and what effect that has on the experience. In particular, we will look at recent results on the perception of virtual humans when various geometry and material properties are altered. We will first discuss the properties of virtual humans that affect how realistic and appealing they appear, such as shape, lighting, and materials. Then, we will discuss how the choice of avatar appearance and how motion tracking errors can impact social interactions in immersive virtual environments both for individuals and groups.

1 MODELING AND RENDERING

Virtual characters have few constraints in terms of design, and can be programmed to take on a multitude of different appearances, ranging from highly stylized to photorealistic, using a variety of modeling and rendering techniques. This section focuses on the perception of virtual humans utilizing different modeling and rendering techniques.

Peoples' attitudes towards virtual characters can be measured in various different ways. The most commonly used measures are subjective responses, where people are asked to give answers to a questionnaire, such as rate their experience or make a decision about what they had witnessed. Subjective responses are usually obtained by questionnaires, where Likert scales and semantic differential scales are used in the attempt to quantify data. Likert scale prompts the person to give a rating of an agreement with a particular statement (e.g., "On a scale from 1 to 7, how realistic is the character?"), while the semantic differential scale has two different descriptors on each end of the scale, for which it was previously established that they belong to the same dimension (e.g., an

emotional response scale can range from happy to sad). The most common approach, however, is to use standardized tests, which are created from a set of scales, measuring a specific construct, and have been tested on a large sample and controlled for validity (that the test is measuring the intended construct) and reliability (the test measures the construct consistently across time, individuals and situations). An example of a standardized measure which measures attitudes towards artificial humans is the Godspeed Questionnaire, introduced by Bartneck et al. [2009] and revised by Ho et al. [2010]. In this first section, we will present results from experiments conducted using these types of measures, with the aim of determining subjective impressions of virtual humans with different shape and material properties.

Firstly, we will review the literature on perception of virtual humans with alterations to their geometry, such as changing the facial proportions, stylization, and changing the level of detail.

1.1 Shape Stylization

Let's consider the shape of the character and how it affects perception of realism and appeal. Facial geometry can range from realistic to stylized with exaggerated/cartoon-like proportions.

In our work [Zell et al. 2015], we investigated a range of combinations of facial shape and material stylization, and found that the shape of the character is the main predictor of realism (Figure 4, bottom). In other words, if you require a character to appear realistic, you should aim for a facial geometry that is close to human-like (ideally by scanning real faces) rather than increasing the realism of the texture/material. Additionally, if you want a character to appear stylized, changing the geometry is more effective than stylizing the texture or material properties.

1.2 Level of Detail

Creating models with different levels of detail (LOD) is commonplace in real-time applications. The overall goal is to maintain rendering speed and a small memory footprint without losing visual accuracy. LOD or the number of polygons on the surface also affects the shape and subsequently the perception. As expected, higher polygon counts have been found in many studies to be associated with higher realism ratings. However, for appeal ratings the results are less consistent. Some older studies have shown that low resolution characters are perceived as less eerie and more appealing than their higher resolution counterparts especially if facial proportions are unnatural [Burleigh et al. 2013; MacDorman et al. 2009]. These findings are in line with the Uncanny Valley theory [Mori et al. 2012]. However, in contrast, our work [Higgins et al. 2021b] using state-of-the-art virtual humans showed that lower LODs were rated as more eerie and less appealing than their higher quality counterparts (Figure 2). This indicates that the modern state-of-the-art virtual humans have passed the Uncanny Valley, being both realistic in appearance but also rated as appealing. In our most recent study [Higgins et al. 2023] we further showed that empathic responses to virtual characters were the same, regardless of the level of realism shown.



Figure 2: High LOD (top) and low LOD (bottom) virtual humans used in the study by Higgins et al. [2021a].

1.3 Facial Proportions

Some studies have investigated altering facial proportions. In one study, it was found that realistic characters were perceived as less appealing if facial parts had strong deviations in terms of size (e.g., eyes have been locally increased) [Seyama and Nagayama 2007]. Results also showed that a mismatch of realism between facial parts negatively affects appeal [Burleigh et al. 2013; MacDorman et al. 2009].

Additionally, facial proportions (in particular the width of the face and size of the eyes) can have consequences on how we perceive personality traits of humans and subsequently virtual humans. Studies from human psychology on real faces have shown

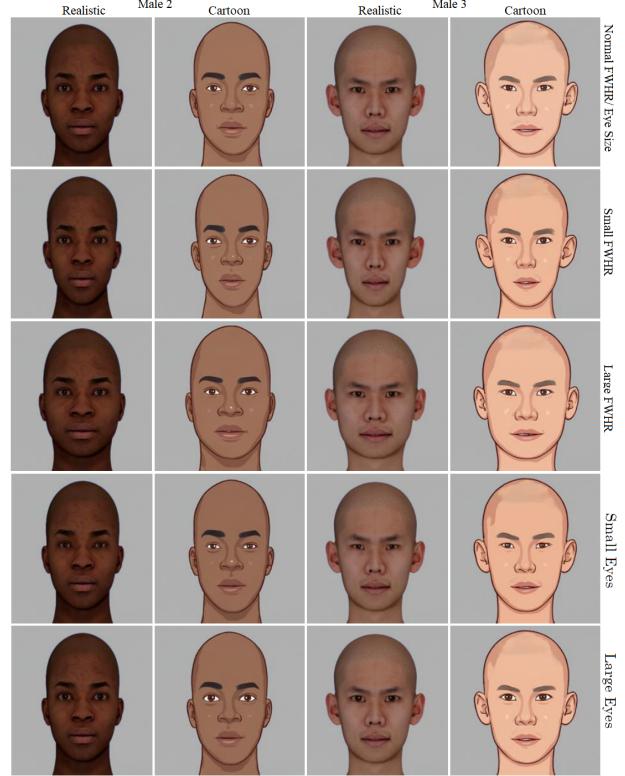


Figure 3: Examples of stimuli from [Ferstl et al. 2021] where the facial proportions of realistic and cartoon faces were altered. FWHR indicates the Facial Width to Height Ratio, with a small value indicating a narrower face.

that individuals with wider faces were judged by observers as more threatening, more dominant and less attractive, especially for male faces [Geniole et al. 2015]. In addition, larger eyes increase trustworthiness [Zebrowitz et al. 1996], while narrow eyes appear aggressive. Recent studies on virtual faces have shown that when facial feature manipulations are carefully kept within natural ranges (i.e., perceived as non-eerie), perceptions of both highly realistic virtual characters and cartoon characters follow the trends observed for human faces [Ferstl and McDonnell 2018; Ferstl et al. 2021] (Figure 3). For non-human virtual characters including robots, and human faces with manipulations outside the natural ranges, the results deviate from real human studies. In particular, we see the opposite effect where narrow faces are perceived as more dominant and aggressive [Ferstl et al. 2017].

In our recent study [Fribourg et al. 2021], we tested this effect further with self-perception when viewing one's own face in Augmented Reality. Interestingly, we found that people perceive their own faces in the same manner - with filters that widen the face appearing more dominant and aggressive, and filters that enlarge the eyes being perceived as more honest. This finding could have interesting implications for using facial filters for video conferencing and telepresence.

2 MATERIALS AND LIGHTING

Next, we look more closely at the importance of materials, textures and lighting on perception of virtual humans. Lighting has been used for centuries to enhance a character's realism, emotion and appeal, from historical paintings, such as Sfumato (light shadow, such as the Mona Lisa by daVinci) and tenebrism (dark shadow, such as John the Baptist by Carvagio) techniques, to the modern-day TV broadcasting. Artists experimented with different types of lighting designs and how they could influence the emotional response of the audience. While many artistic rules of lighting exist, there are few perceptual studies.

In our work, we have found that material is the main predictor of perceived appeal [Zell et al. 2015], specifically the albedo texture. In general, appeal, attractiveness, and eeriness are highly dependent on the material stylization. In particular, blurring realistic textures, while preserving feature contours (e.g., lip contours) made all types of characters (realistic and stylized) more appealing. Empirical observations that smoother skin is considered more appealing can also be found in many photograph retouching books and photo-retouching software for faces.

Keeping the stylization level of the material consistent with the stylization level of the geometry is another way to increase appeal, while strong mismatches (e.g., very realistic material on a stylized shape) result in unappealing and eerie characters. These types of characters with strong mismatches could be considered unfamiliar which is consistent with other research where we also found a drop in appeal came from characters that were rated as 'unfamiliar' or those that were in the middle of the 'abstract to realistic' scale [McDonnell et al. 2012] (Figure 4, top). This could be attributed to the fact that these characters were difficult for the brain to categorize due to their uncommon appearance, as suggested in the study of Saygin et al. [2012].

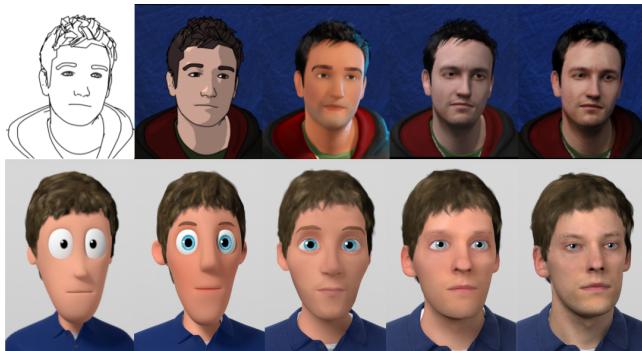


Figure 4: Examples of levels of material stylization [McDonnell et al. 2012] (top) and shape stylization [Zell et al. 2015] (bottom).

In other work, we investigated the effect of simply altering the brightness and key-to-fill ratio (darkness of shadow) [Wisessing et al. 2016, 2020]. We found strong evidence that the brighter the light (up to acceptable saturation levels) the more appealing the

characters appeared, regardless of stylization level. However, brightness did not improve the eeriness ratings. In order to reduce eeriness, lightening the shadows was effective for cartoons and mid-level stylized characters. However, lightening shadows did not improve eeriness or appeal ratings for realistic characters.

3 SOCIAL INTERACTIONS

Avatar use on video-conference platforms has found dual purpose in recent times as a potential method for ensuring privacy and improving subjective engagement with remote meeting, provided one can also ensure a minimal loss in the quality of social interaction and sense of personal presence.

In this section, we aim to answer the question - What effect does character appearance have on social interactions for telepresence? We will discuss the effect of character appearance on the feeling that a character is alive and present in a virtual space (Social Presence), and on the illusion that a virtual body or face is your own face (Ownership/embodiment). Furthermore, we will discuss how motion tracking errors could affect how the personality of the human driving the avatar is perceived, and the effect of long-term exposure.

While there are standardised questionnaires for the perceptual illusions, indirect measures where the participant is not aware of the purpose of the testing can be highly effective. For example, rather than asking the participant how threatening the character appears to him or her, we can measure participant's increase in heart rate or pupil dilation [Sterna et al. 2023]. Other indirect measures can be used such as the measure of proximity to virtual humans in the studies of Bailenson et al. [2005]. Indirect measures are extremely valuable as they bypass any conscious interpretation from the person which could affect the measured data. However, indirect measures may pose a question to validity - do these measures really reflect the nature of the studied construct? To increase validity, a combination of direct and indirect measures is usually the best choice for a rigorous perceptual study.

3.1 Social Presence

A vast body of research in VR has shown that even simple environments and representations of a virtual character can evoke strong sensations of being present in a real environment (place illusion) and that a virtual character is alive and present in the virtual space with them (co-presence or social presence). Place illusion is created when the user interacts with the environment and receives an appropriate response [Slater et al. 2009]. Social presence is elicited when the virtual character exhibits interaction cues, even as minimal as eye-contact [Bailenson et al. 2003]. Social presence is usually high for avatar interactions in virtual reality and teleconferencing, since the avatar is driven by a human in real-time, and therefore displays appropriate reactions in conversation and realistic movements. Visual fidelity (realistic appearance and animation) of the character was not found to play a significant role in increasing social presence [Garau et al. 2003; Nowak 2001]. However, the user's social presence can be reduced if the character's behaviour and appearance realism do not match [Bailenson et al. 2005; Garau et al. 2003; Zibrek et al. 2018]. Appearance can therefore affect the experience in the environment and with the character in VR.

In our large-scale study [Zibrek et al. 2019] where over 700 participants were immersed in VR, we investigated if photorealism could increase social presence, place illusion and concern for a virtual character in VR. We did so by measuring people's subjective and behavioural responses to a photo-realistic virtual character in a corresponding environment, and a simplified version of both. We also embodied the participant in the room by using real-time full-body optical motion capture [Zibrek and McDonnell 2019]. We found confirmation for the effect of photorealism on self-reported social presence and place illusion. However, proximity was unaffected by photorealism in our studies, meaning that participants used similar social distance to place themselves beside characters, regardless of how they were rendered or if they appeared particularly realistic.

3.2 Body and Face Ownership

Studies suggest that the level of realism of a virtual body does not affect the illusion of body-ownership when congruent visual-tactile or visual-motor cues are provided [Maselli and Slater 2013]. However, if the realistic avatar appears similar to the human (self-avatar), studies revealed higher ownership for the participants controlling their self-representation than with abstract representations [Gorisse et al. 2019]. Additionally, there is evidence that self-representation alterations can lead to self-concept and behaviour modifications, called the Proteus effect [Yee et al. 2007]. However, it is not yet clear if the level of photorealism of the avatar could cause the Proteus effect, or in what way.

For facial ownership, we have shown that high levels of face ownership and agency towards a virtual face can be induced through real-time mapping of the facial movements in a screen-based task with a virtual mirror [Kokkinara and McDonnell 2015]. In terms of the appearance of the virtual face, we found no differences on the reported ownership levels between the cartoon and realistic appearance styles, with both our styles also receiving high ratings on appeal.

In our recent study, we examined how study participants experienced enface toward high fidelity state-of-the-art avatars of different ages - ranging from young, middle-aged, to elderly [Jordan et al. 2023]. Participants rated the all avatars low on eeriness and high on realism, but rated the older avatars as less attractive. Interestingly, participants felt a similar sense of enface regardless of the given self-avatar's age or difference to their own age. This finding is consistent with Banakou et al.'s [2013] work, in which participants experienced similar levels of embodiment toward self-avatars regardless of age. However, more recent work by Fang et al. [2022] in VR showed lower levels of ownership for a cartoon than a realistic avatar, with their cartoon avatar rated as more attractive, human-like, and less eerie than the realistic one.

In another work, we focused on conversations through real-time motion captured 3D personalized virtual avatars in a 2D video-conferencing context [Higgins et al. 2021b]. Participants were embodied in cartoon avatars and had positive and negative valenced emotions induced (Figure 1). They then had to converse with the experimenter about their feelings towards the positive and negative content that they had viewed or listened to. Our results showed some evidence for face and body ownership, while participants also reported high levels of social presence with the other avatar,

indicating that avatar cameras could be a favorable alternative to non-camera feeds in video conferencing.

Results also highlighted subjective reports of social presence to be higher in avatar conversation compared to non-video conditions. There is also evidence that cartoon avatars were appropriate for use in negative conversations as participants felt comfortable to discuss negative topics even while embodied in a character with a happy appearance. In fact, the avatar conditions were favourable over voice-only conditions for negative valence discussions.

Other work by Oh et al. [2016] showed an ability to improve social presence and positive affect in avatar-to-avatar conversations by artificially enhancing the smile of the avatar. This ability to manipulate the appearance of the avatar to improve conversation could prove very important in future systems. Though the ethical implications of these sorts of manipulations would need to be considered.

The evidence from these studies suggests that avatars can be effective for remote collaboration and that users can potentially enface a diverse range of screen-based avatars. Furthermore, scores on fatigue were generally low, suggesting that the conversational aspect with the experimenter did not appear overly taxing and is likely to be a good choice in video-conference scenarios, which is also promising for the use of such avatars in this context.

3.3 Motion Tracking

Non-verbal cues are highly important in face-to-face communication and can be used as reliable indicators of personality. For example, Thomas et al. [2022] found that motion is the dominant modality for portraying an extraverted personality (Figure 5). However, motion tracking can frequently cause errors in social VR when parts of the body are obscured and the effects on our perception are not yet fully known.

However, some recent studies have begun to investigate the impact of tracking errors on social interactions with avatars. Ferstl et al. [2021] showed that motion errors are interpreted, at least in part, as a shift in interlocutor personality. Adkins et al. [2023] investigated the consequences of motions errors in hand and finger motions on comprehension, character perception, social presence, and user comfort. They showed that errors in hand motions affect the viewers' impression of a character and that jittering hand motions should be avoided as they significantly decrease user comfort.

3.4 Group behaviours

While much of the previous work focuses on one-to-one interactions with virtual humans on-screen and in VR, there are also many applications where humans will be interacting socially as groups of avatars. Some recent studies have investigated how avatar selection can affect group behaviour.

A recent large-scale study by Han et al. [2023] examined group behaviours using different avatar identities and environments over time. Interestingly, they found that entitativity, presence, enjoyment, and realism increased over 8 weeks, which is promising news for long-term use of social VR.

They also found that avatar appearance in a group affected group behaviours. For example, when group members used cloned avatars, they rated the experience as more fun, while personalised avatars

improved the realism of the experience and participants elicited more typical social behaviours such as synchronisation.

In our recent study on group behaviour in VR [Lauren Buck 2023] we showed that elements of work group inclusion are different between the physical world and VR, and that customization choices and user perceptions of avatars may shape the perception of inclusion. In particular, we found evidence that users who design avatars more closely linked to an idealized version of themselves – one that stands out and encapsulates their personality rather than physical appearance – may be likely to feel a stronger sense of work group inclusion.



Figure 5: Motion is the dominant modality for conveying an extraverted personality, when compared to voice and appearance. Example stimuli from study by Thomas et al. [2022].

4 DISCUSSION

Despite the number of publications on the topic, we are still far from fully understanding the perception of virtual characters. Virtual characters have been present in various domains for many years. Today's quick developments of real-time rendering technologies enable communication in virtual environments (VR, AR), and social media means that the need for understanding the implications of different forms of self-representations of the user either as a doppelgänger or an artificial avatar has never been greater. It is relevant to note, that the perception of virtual characters is not a static but quite a dynamic research area which needs continuous reassessment.

One important aspect that needs constant attention is diversity in the character models used for experiments, to ensure our results can be generalized. Much of the work in the literature on virtual human perception used just a single character for evaluation, due to lack of diversity in available character models. However, with the availability of new tools for avatar generation (such as Metahumans¹), it should be easier in the future to conduct experiments with diversity in aspects such as race, age, gender, etc. Already, many recent studies are starting to use more diverse sets of character models in their studies (e.g., [Ferstl et al. 2021; Higgins et al. 2021a], etc.), but it is important to continue this trend, in particular due to the historic bias in Computer Graphics research towards perfecting the rendering of skin and hair for white characters only [Theodore Kim 2020].

Finally, ethical considerations around avatar choice should also be addressed. For example, studies have shown that changing the avatar facial proportions can affect moral decisions in video games, with more aggressive-looking characters triggering less utilitarianistic choices in direct harm scenarios [Ferstl et al. 2017]. If this effect transferred to teleconferencing by simply altering the facial features

or adding a facial filter [Fribourg et al. 2021], it could have ethical implications. Additionally, alterations could be used for nefarious purposes, where a simple change in render style or facial features could alter the perceived personality [Ferstl et al. 2021; Zibrek et al. 2018] and even trust [McDonnell and Breidt 2010] towards the avatar. For a more detailed discussion on the potential threats posed by virtual humans used for conversing in virtual environments, please see [Buck and McDonnell 2022].

REFERENCES

- Alex Adkins, Aline Normoyle, Lorraine Lin, Yu Sun, Yuting Ye, Massimiliano Di Luca, and Sophie Jörg. 2023. How Important Are Detailed Hand Motions for Communication for a Virtual Character Through the Lens of Charades? *ACM Trans. Graph.* 42, 3, Article 27 (may 2023), 16 pages. <https://doi.org/10.1145/3578575>
- Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin* 29, 7 (2003), 819–833.
- Jeremy N Bailenson, Kimberly R Swinh, Crystal L Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 4 (2005), 379–393.
- Domna Banakou, Raphaela Groten, and Mel Slater. 2013. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proc. of the National Academy of Sciences* 110, 31 (2013), 12846–12851.
- Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.
- Lauren Buck and Rachel McDonnell. 2022. Security and Privacy in the Metaverse: The Threat of the Digital Human. In *1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality*.
- Tyler J. Burleigh, Jordan R. Schoenherr, and Guy L. Lacroix. 2013. Does the Uncanny Valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. *Computers in Human Behavior* 29, 3 (2013), 759–771.
- Ylva Ferstl, Elena Kokkinara, and Rachel McDonnell. 2017. Facial Features of Non-player Creatures Can Influence Moral Decisions in Video Games. *ACM Transaction on Applied Perception* 15, 1, Article 4 (2017), 12 pages. <https://doi.org/10.1145/3129561>
- Ylva Ferstl and Rachel McDonnell. 2018. A Perceptual Study on the Manipulation of Facial Features for Trait Portrayal in Virtual Agents. In *Proc. of Int. Conf. on Intelligent Virtual Agents (IVA)*. 281–288. <https://doi.org/10.1145/3267851.3267891>
- Ylva Ferstl, Rachel McDonnell, and Michael Neff. 2021. Evaluating Study Design and Strategies for Mitigating the Impact of Hand Tracking Loss. In *ACM Symposium on Applied Perception 2021 (Virtual Event, France) (SAP '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 12 pages. <https://doi.org/10.1145/3474451.3476235>
- Rebecca Fribourg, Etienne Peillard, and Rachel McDonnell. 2021. Mirror, Mirror on My Phone: Investigating Dimensions of Self-Face Perception Induced by Augmented Reality Filters. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 470–478. <https://doi.org/10.1109/ISMAR52148.2021.00064>
- Maia Carau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M Angela Sasse. 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 529–536.
- Shawn N. Geniole, Thomas F. Denson, Barnaby J. Dixson, Justin M. Carré, and Cheryl M. McCormick. 2015. Evidence from Meta-Analyses of the Facial Width-to-Height Ratio as an Evolved Cue of Threat. *PloS one* 10, 7 (2015), e0132726.
- Geoffrey Gorisse, Olivier Christmann, Samory Houzangbe, and Simon Richir. 2019. From Robot to Virtual Doppelganger: Impact of Visual Fidelity of Avatars Controlled in Third-Person Perspective on Embodiment and Behavior in Immersive Virtual Environments. *Frontiers in Robotics and AI* 6 (2019), 8. <https://doi.org/10.3389/frobt.2019.00008>
- Eugy Han, Mark R Miller, Cyan DeVeaux, Hanseul Jun, Kristine L Nowak, Jeffrey T Hancock, Nilam Ram, and Jeremy N Bailenson. 2023. People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse. *Journal of Computer-Mediated Communication* 28, 2 (01 2023), zmac031. <https://doi.org/10.1093/jcmc/zmac031>; arXiv:<https://academic.oup.com/jcmc/article-pdf/28/2/zmac031/48520441/zmac031.pdf>
- Darragh Higgins, Donal Egan, Rebecca Fribourg, Benjamin Cowan, and Rachel McDonnell. 2021a. Ascending from the Valley: Can State-of-the-Art Photorealism Avoid the Uncanny?. In *ACM Symposium on Applied Perception 2021 (Virtual Event, France) (SAP '21)*. Association for Computing Machinery, New York, NY, USA, Article 4, 12 pages. <https://doi.org/10.1145/3474451.3476236>

¹<https://www.unrealengine.com/en-US/metahuman-creator>

- France) (SAP '21). Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. <https://doi.org/10.1145/3474451.3476242>
- Darragh Higgins, Rebecca Fribourg, and Rachel McDonnell. 2021b. Remotely Perceived: Investigating the Influence of Valence on Self-Perception and Social Experience for Dyadic Video-Conferencing With Personalized Avatars. *Frontiers in Virtual Reality* 2 (2021). <https://doi.org/10.3389/frvr.2021.668499>
- Darragh Higgins, Yilin Zhan, Benjamin R. Cowan, and Rachel McDonnell. 2023. Investigating the Effect of Visual Realism on Empathic Responses to Emotionally Expressive Virtual Humans. In *ACM Symposium on Applied Perception 2023* (Los Angeles, CA, USA) (SAP '23). Association for Computing Machinery, New York, NY, USA, Article 5, 7 pages. <https://doi.org/10.1145/3605495.3605799>
- Chin-Chang Ho and Karl F. MacDorman. 2010. Revisiting the Uncanny Valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior* 26, 6 (2010), 1508–1518.
- Hugh Jordan, Lauren Buck, Pradyna Shinde, and Rachel McDonnell. 2023. What's My Age Again? Exploring the Impact of Age on the Enfacement of Current State-of-the-Art Avatars. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 865–866. <https://doi.org/10.1109/VRW58643.2023.000275>
- Elena Kokkinara and Rachel McDonnell. 2015. Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. Acm, 221–226.
- Rachel McDonnell Lauren Buck, Gareth W. Young, to appear 2023. Avatar Customization, Personality, and the Perception of Work Group Inclusion in Immersive Virtual Reality. In *ACM Conference On Computer-Supported Cooperative Work And Social Computing*.
- Fang Ma and Xueni Pan. 2022. Visual Fidelity Effects on Expressive Self-avatar in Virtual Reality: First Impressions Matter. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 57–65. <https://doi.org/10.1109/VR51125.2022.00023>
- Karl F. MacDorman, Robert D. Green, Chin-Chang Ho, and Clinton T. Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior* 25, 3 (2009), 695–710.
- Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Frontiers in human neuroscience* 7, 83 (2013).
- Rachel McDonnell and Martin Breidt. 2010. Face Reality: Investigating the Uncanny Valley for Virtual Faces. In *ACM SIGGRAPH ASIA 2010 Sketches* (Seoul, Republic of Korea) (SA '10). Association for Computing Machinery, New York, NY, USA, Article 41, 2 pages. <https://doi.org/10.1145/1899950.1899991>
- Rachel McDonnell, Martin Breidt, and Heinrich H. Bülthoff. 2012. Render me real? Investigating the effect of render style on the perception of animated virtual humans. *ACM Transaction on Graphics* 31, 4 (2012), 91:1–91:11.
- Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the field]. *IEEE Robotics and Automation Magazine* 19, 2 (2012), 98–100.
- Kristin Nowak. 2001. The influence of anthropomorphism on social judgment in social virtual environments. In *Annual Convention of the International Communication Association, Washington, DC*.
- Soo Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLOS ONE* 11 (09 2016), e0161794. <https://doi.org/10.1371/journal.pone.0161794>
- Ayse P. Saygin, Thierry Chaminade, Hiroshi Ishiguro, Jon Driver, and Chris Frith. 2012. The thing that should not be: Predictive coding and the Uncanny Valley in perceiving human and humanoid robot actions. *Social Cognitive Affective Neuroscience* 7, 4 (2012), 413–422.
- Jun'ichiro Seyama and Ruth S. Nagayama. 2007. The Uncanny Valley: Effect of Realism on the Impression of Artificial Human Faces. *Presence: Teleoperators and Virtual Environments* 16, 4 (2007), 337–351.
- Mel Slater, Daniel Pérez Marcos, Henrik Ehrsson, and Maria V Sanchez-Vives. 2009. Inducing illusory ownership of a virtual body. *Frontiers in neuroscience* 3 (2009), 29.
- Radosław Sterna, Artur Cybulska, Małgorzata Igras-Cybulska, Joanna Pilarczyk, and Michał Kuniecki. 2023. Does realism of virtual character influence arousal? Exploratory study with pupil size measurement. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 655–656. <https://doi.org/10.1109/VRW58643.2023.00170>
- booktitle=Scientific American Theodore Kim. 2020. The Racist Legacy of Computer-Generated Humans. <https://doi.org/3thBeye>
- Sean Thomas, Ylva Ferstl, Rachel McDonnell, and Cathy Ennis. 2022. Investigating how speech and animation realism influence the perceived personality of virtual characters and agents. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 11–20. <https://doi.org/10.1109/VR51125.2022.00018>
- Pisut Wisessing, John Dingiana, and Rachel McDonnell. 2016. Perception of Lighting and Shading for Animated Virtual Characters. In *Proc. of ACM Symp. of Applied Perception (SAP)*. 25–29.
- Pisut Wisessing, Katja Zibrek, Douglas W. Cunningham, John Dingiana, and Rachel McDonnell. 2020. Enlighten Me: Importance of Brightness and Shadow for Character Emotion and Appeal. *ACM Trans. Graph.* 39, 3, Article 19 (April 2020), 12 pages. <https://doi.org/10.1145/3383195>
- Nick Yee, Jeremy N Bailenson, and Kathryn Rickertsen. 2007. A Meta-analysis of the Impact of the Inclusion and Realism of Human-like Faces on User Experiences in Interfaces. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. 1–10. <https://doi.org/10.1145/1240624.1240626>
- Lesslie A. Zebowitz, Luminita Voinescu, and Mary Ann Collins. 1996. "wideeyed" and "crooked-faced": Determinants of perceived and real honesty across the life span. *Personality and Social Psychology Bulletin* 12, 12 (1996), 1258–1269.
- Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To Stylize or Not to Stylize?: The Effect of Shape and Material Stylization on the Perception of Computer-generated Faces. *ACM Transactions on Graphics* 34, 6, Article 184 (2015), 184:1–184:12 pages.
- Katja Zibrek, Elena Kokkinara, and Rachel McDonnell. 2018. The Effect of Realistic Appearance of Virtual Characters in Immersive Environments—Does the Character's Personality Play a Role? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1681–1690.
- Katja Zibrek, Sean Martin, and Rachel McDonnell. 2019. Is Photorealism Important for Perception of Expressive Virtual Humans in Virtual Reality? *ACM Transactions on Applied Perception* 16, 3, Article 14 (sep 2019), 19 pages. <https://doi.org/10.1145/3349609>
- Katja Zibrek and Rachel McDonnell. 2019. Social Presence and Place Illusion Are Affected by Photorealism in Embodied VR. In *Motion, Interaction and Games* (Newcastle upon Tyne, United Kingdom) (MIG '19). Association for Computing Machinery, New York, NY, USA, Article 13, 7 pages. <https://doi.org/10.1145/3359566.3360064>



Physical Interactions in Telepresence

Carol O'Sullivan

Professor of Visual Computing, Trinity College Dublin

Email: Carol.OSullivan@tcd.ie

In this section, we present some considerations around allowing remote participants to physically interact with each other and the environment in a shared space via telepresence.

We first present a simple scenario as a case study to explore a subset of interactions involved with sharing a virtual object, such as a ball, that can be thrown from one participant to another and that also interacts with the environment.

The perceptual factors that affect this scenario are then discussed, followed by some relevant previous work.

Physical Interactions in Reality

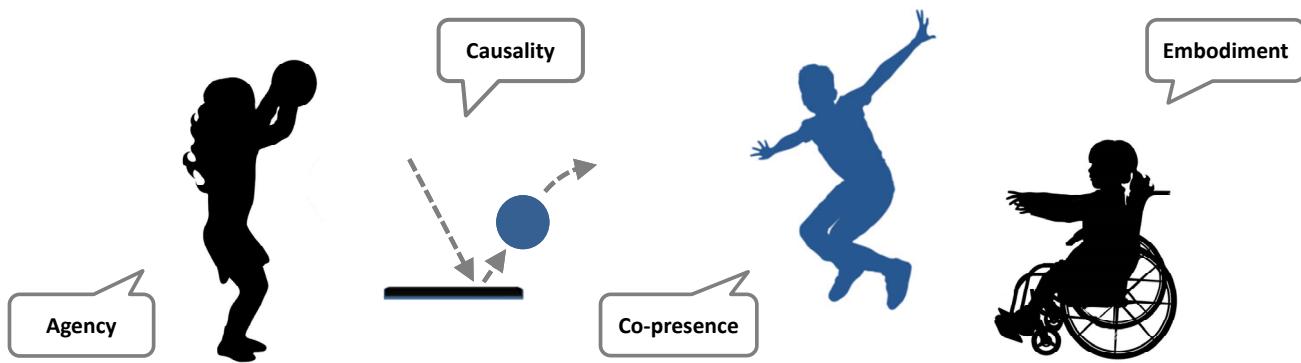


Consider the scenario above:

- Three children are playing dodgeball/catch in a room, surrounded by various rigid and deformable objects (e.g., walls, floor, hard and soft furnishings, ornaments).
- Jill picks up the ball in two hands, feels the weight, lifts it over her head and throws it to Jack.
- Jack dodges and the ball misses him, bounces off a wall before landing on and indenting a soft cushion, where it comes to rest.
- Jack picks up the ball and throws it back at Jill, catching her on the arm; she feels the impact and deflects the ball towards a vase, which falls on the floor.
- Mary is in a wheelchair with minimal mobility, so can only watch the others play.

How can we emulate this real-world experience if all three children are in different locations, and enhance it for Mary by enabling her to fully participate in the game?

Physical Interactions in Mixed Reality



Now imagine this situation in Mixed Reality, when the children are not co-located and perhaps the ball and one or more of the children are virtual.

- How will Jill pick up and throw that virtual ball?
- How will virtual Jack anticipate Jill's intentions and dodge the ball?
- How will the virtual ball appear to deform the real cushion or knock over the real vase?
- How will Jill experience being hit by the virtual ball?
- How can Mary participate fully in the game?
- How will we know when we have succeeded in bridging the perceptual gap between the real physical interactions and the MR ones?

Before answering these questions, we will first explore some perceptual factors that affect this experience, namely Causality, Agency, Co-presence and Embodiment

Perception of Physical Interactions

– Perception of causality

— perceiving that an event causes a particular response to occur

– Sense of personal agency

— when a user feels that they are both controlling their own body and affecting the external environment

– Sense of co-presence

— the feeling of “being able to perceive others while being actively perceived by them”, i.e., that one is sharing an environment with another person .

– Sense of embodiment

— “the ensemble of sensations that arise in conjunction with being inside, having, and controlling a body”

- The *perception of causality* occurs when it can be seen that an event causes a particular physical response to occur, e.g., a ball hitting a table causes it to bounce off; it hits a cushion and causes a deformation in the surface; it hits a person and knocks them backwards. The ability to perceive events as causal or not is developed early in infancy [1] and a delay of as low as 120 milliseconds between collision and response can cause an event to be perceived as non-causal.
- A sense of *personal agency* [2] in the world is achieved when a user feels that they are both controlling their own body and affecting the external environment, e.g., picking up a ball and throwing it.
- A sense of *co-presence* is the feeling of ‘being able to perceive others while being actively perceived by them’[3] and the feeling that one is sharing an environment with another person.
- A *sense of embodiment* is “the ensemble of sensations that arise in conjunction with being inside, having, and controlling a body” [4].

[1] Scholl, B., and Tremoulet, P., “Perceptual causality and animacy.” *Trends in Cognitive Sciences*, 4, 8, 299-309.

[2] Hannah, L., Coyle, D., and Moore J.W.(2014) “.” The experience of agency in human-computer interactions: a review *Frontiers in Human Neuroscience* 8.

[3] Zhao, S. (2003). “Toward a Taxonomy of Copresence.” *Presence: Teleoperators and Virtual Environments*. 12, 5, 445-455

[4] Kilteni, K., Grotens, R., and Slater, M. (2012). “The Sense of Embodiment in Virtual Reality.” *Presence Teleoperators & Virtual Environments*. 21, 4, 373-387.

Perception of Physical Interactions in Telepresence



Causality: can it be perceived that a telepresence event causes a physical response?



Agency: can a remote interaction give the sense of effecting a physical change?



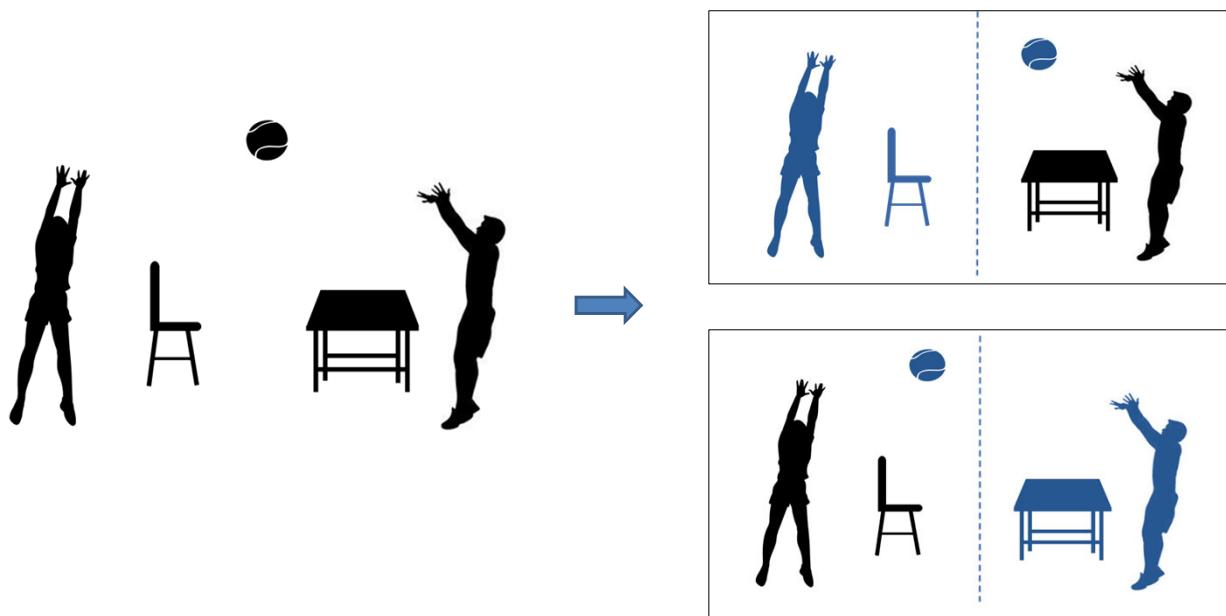
Copresence: can a remote interaction with a partner seem like he's sharing the same space?



Embodiment: can agency be felt when controlling an avatar for remote interactions?

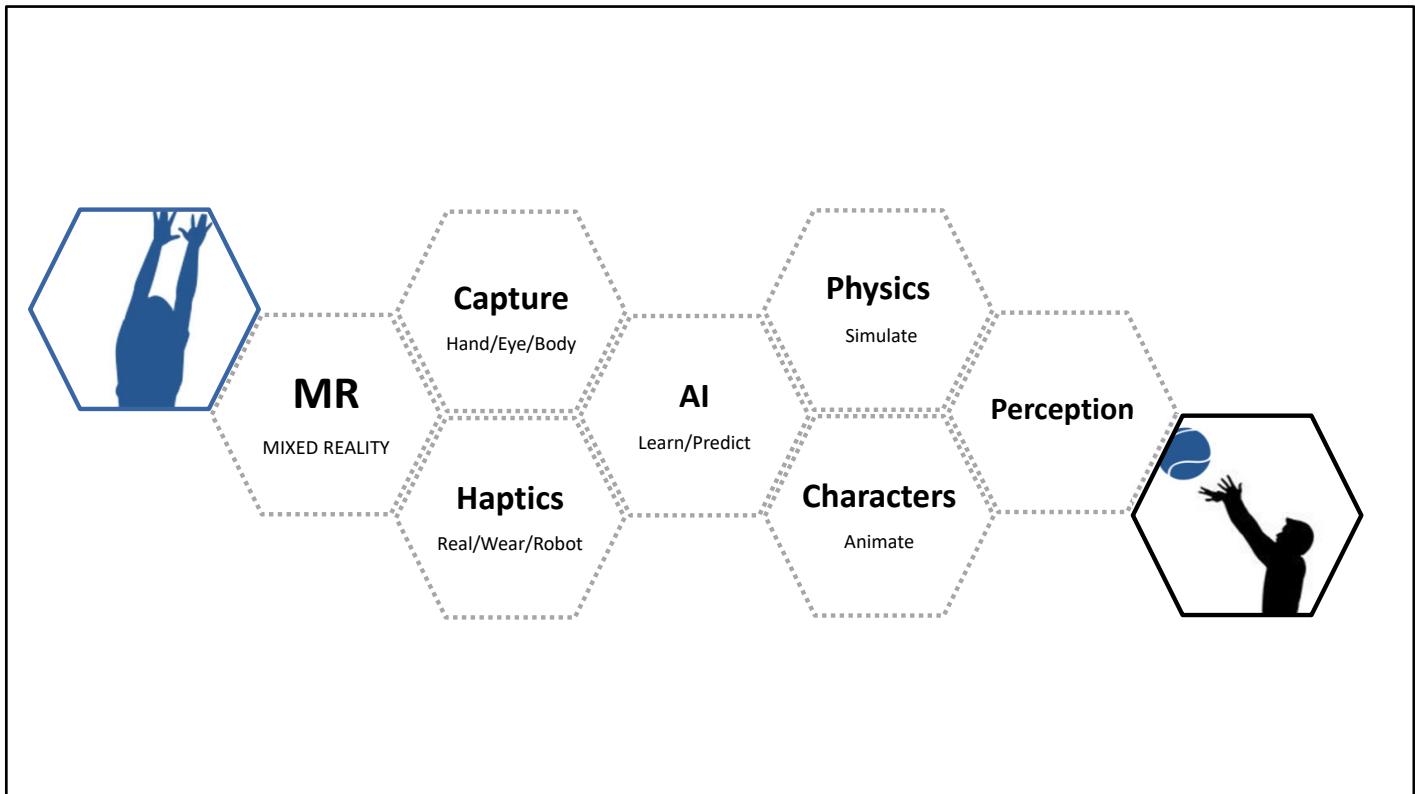
We now put these perceptual questions in the context of telepresence.

Physical Interactions in Telepresence



In a typical telepresence scenario, each participant occupies their own space and environment, and sees the others in some digital or mixed reality form.

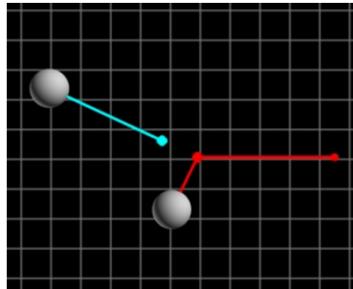
How can an object be shared between these remote spaces and seamlessly interact with the participants and their environments, to deliver the experience of a single, shared space?



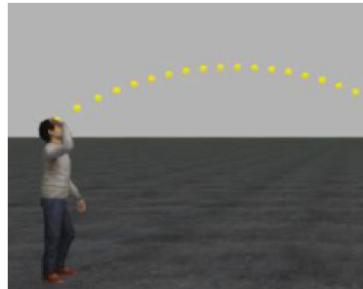
To achieve the goal of such a shared experience requires the convergence of multiple fields, including perception, physical simulation (e.g., collision detection and response), character animation (to simulate responsive avatars) and artificial intelligence (e.g., machine learning and intention prediction).

Next we will present some previous research works that are relevant for some of these problems.

Perception of Physical Interactions



Collisions:



Throwing:



Characters:

"Evaluating the visual fidelity of physically based animations"
O'Sullivan et al. *ACM TOG/SIGGRAPH*, 2003

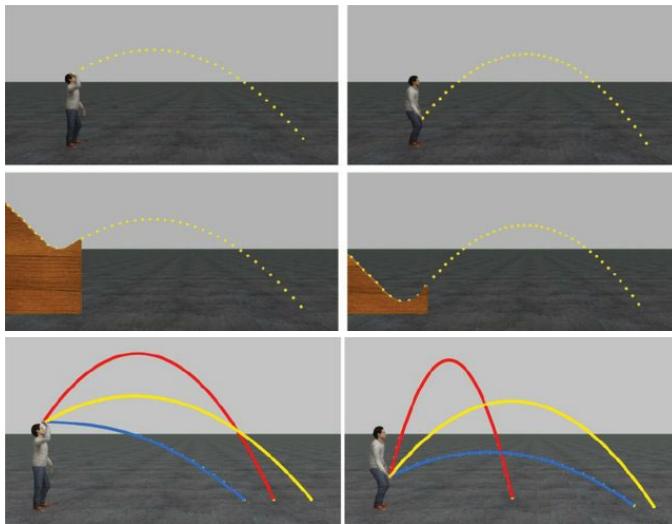


Let's look at some previous work on the perception of physical interactions such as collisions, throwing a ball, and characters pushing each other.

Firstly, we have explored the visual fidelity of simulated collisions and the perception of causality. The goal of this work was to determine the factors that affect a user's perception of a collision in a real-time simulation. In many such systems, a visually plausible result is more acceptable than a delay due to calculating a physically accurate response.

We conducted some perception experiments that established to determine how sensitive viewers are to visual errors in physically based simulations of rigid objects. Such errors can include delayed collision responses (which affect the perception of causality), and angular and momentum distortions. Please see the video for demonstrations and further details.

Perception of Throwing



"Perceptual Evaluation of Motion Editing for Realistic Throwing Animations"
Vicovaro et al. *ACM Transactions on Applied Perception*, 2014



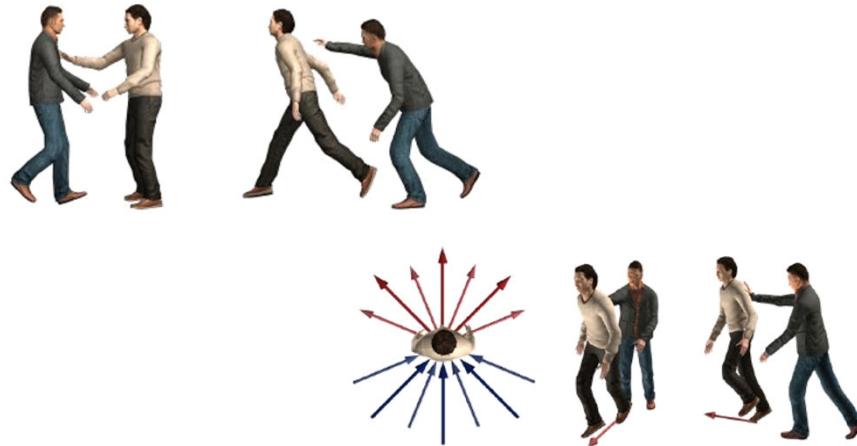
We have also studied the factors that can affect a viewer's perception of throwing animations. Often, manipulations of such animations are needed to adapt motion capture data to a particular real-time event (such as in games and VR). The results may be relevant for such interactions in a telepresence experience.

- First, the release velocity of the ball was manipulated, while leaving the original character motion and angle of release unchanged;
- Then, both the motion of the thrower and the release velocity of the ball were simultaneously modified, using dynamic time warping.

In both of these cases, shortened underarm throws were found to be unnatural, and editing the character's motion along with the ball's release velocity did not help to increase the plausibility of the throwing animation.

- Next, only the angle of release of the ball was changed, while the release velocity and character's motion were not. This simple modification significantly improved the perceived plausibility, especially for the shortened underarm throws.
- Finally, the virtual human thrower with a mechanical throwing device was replaced by a static ramp object, and opposite results were found, which suggests that biological and physical throwing events are perceived differently.

Perception of Character Interactions



Push it Real: Perceiving Causality in Virtual Interactions
Hoyet, McDonnell & O'Sullivan, *ACM TOG/SIGGRAPH Asia*, 2012



Another potentially relevant study explored collisions between human characters, specifically pushing interactions, where the real-time physical simulation of the collisions can lead to errors in timing, forces and response direction.

The motions of two actors pushing each other from different directions with varying forces were captured and animations were generated where errors were introduced. The participants were asked to state whether these animations were modified or not.

Participants were able to accurately identify timing errors of 150ms and over, with no difference between early or late responses. They could also detect force mismatches, especially when the force exerted by the pusher was matched with the reaction of the target character to a weaker push. Errors in the angle of the target's direction after being pushed were also quite perceptible, especially when the angle was contracted towards the pusher's body.

Throwing in VR



"Visual Perspective and Feedback Guidance for VR Free-Throw Training"
Covaci, Olivier & Multon, *IEEE CG&A*, 2015



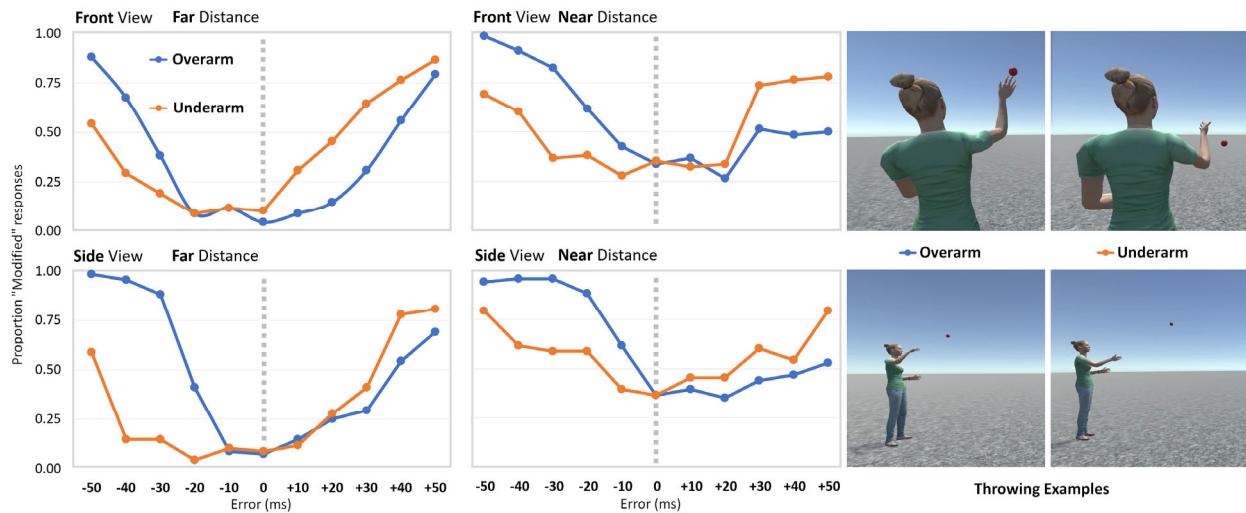
The aim of this research was to evaluate how effective VR would be for training beginner players to throw a basketball. In this study, a real ball was tracked, and its continued trajectory was then simulated within a Virtual Environment.

Three different visual conditions were tested:

- First-person view, where the ball's origin is at the location of the thrower;
- Third-person view, where the ball's origin is on the screen in front of the thrower; and
- Third-person view with guidance, where further visual cues were provided to the thrower, to indicate the ideal trajectory of the ball.

It was found that participants estimated the distance to the basket more accurately in the third-person view, while the added visual cues in the guidance condition resulted in more natural throwing behaviour. Experts and beginners performances were then compared and it was found that the motions of the beginners became more similar to those of the experts after training in the Virtual Environment.

Perception of Point of Release Timing Errors

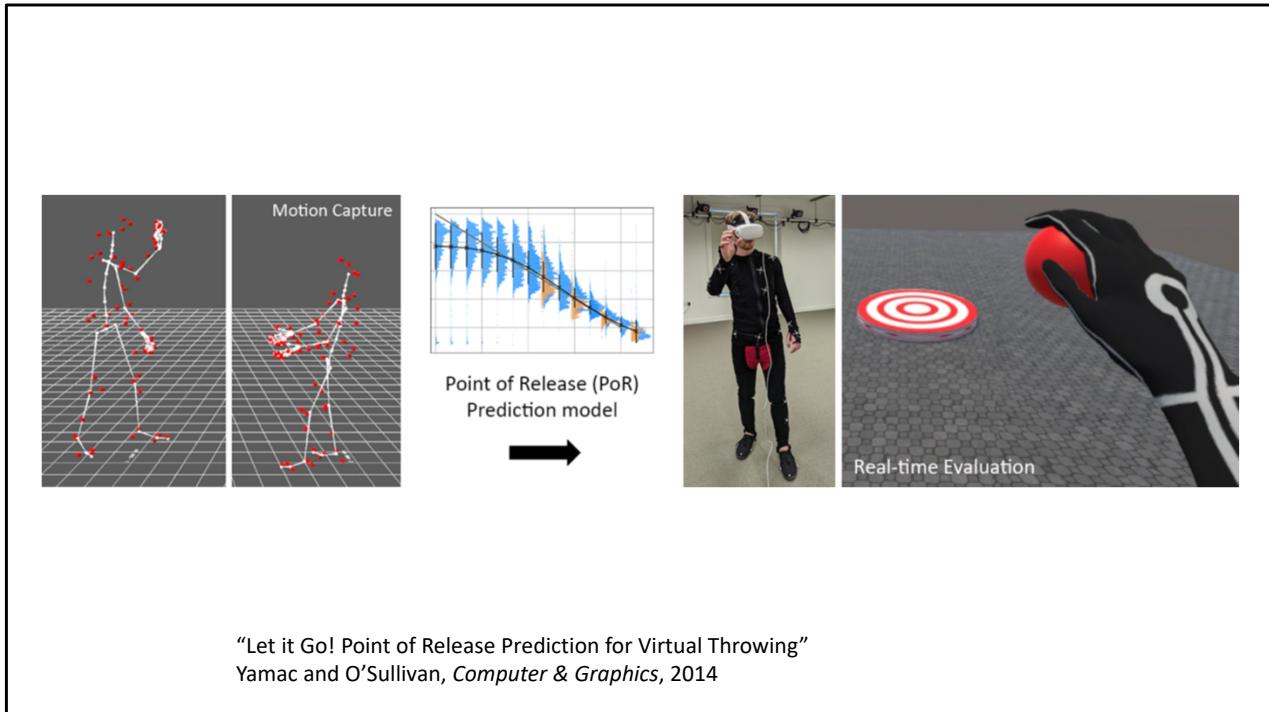


Yamac & O'Sullivan, 2022



More recently, we ran a study to explore how errors in the timing of the point of release can affect how animated throwing motions are perceived. The point of release was modified to be early or late and participants were asked to indicate whether the animation had been modified or not.

There was a difference found for overarm and underarm throws. A delay in the release of the ball was found to be more acceptable than an early release, while the opposite was found for underarm throws. The view (Side or Over the Shoulder) also affected the perceived accuracy of the animation.



These results helped to guide the training of a model to detect the point of release of a ball in an interactive application such as VR or a telepresence experience, where a limited number of simple sensors can be used to execute throwing motions.



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Thank you