Presence • Exploring Interaction Fundamentals for Spatial Computing

Lukas Moro
Umeå Institute of Design
MFA • Interaction Design
June 2024

Thanks to Alex, Kay, Jakub, Davide, Shengzhi, Ben, Felix, Cosima, Simon, Ylva and Stoffel for providing feedback and helping me to gain clarity throughout this project.

Abstract

This project aims to explore how integration between virtual and physical spaces in terms of physics, light, controlling, scene representation and scene semantics could lead towards a more meaningful and less obtrusive embodiment of software in the world.

As physical beings we experience bodily and mental presence. While physical objects acknowledge our existence fully, today's virtual objects are occupying us mostly mentally. This leads to a disconnection of spaces where presence in one of them diminishes presence in the other.

The goal of the conducted prototypes and the proposed interaction fundamentals is to achieve a state of presence where both body and mind are fully engaged in a blended space. This space serves as a common perceptual stage where virtual and physical objects combine their advantages, allowing users to be completely mentally and physically present.

This should serve as a stage for applications infused with artificial intelligence that have much higher grades of agency when compared with today's applications.

Table of Contents

Sharing a memory. No direct interaction with a computer, direct conversations with people instead. No matter where you are, you are present. Memories remain in space. The space responds, you respond to the space, it responds back. It understands your intention and shapes around you, fluid like water. Infinite options experienced in the world, facilitating presence of body & mind. Everything is connected, helping us to make sense. No distraction from what matters to us, never...

An operating system to the world.

Introduction

Virtuality becomes more prevalent in our perceivable reality. Many services and objects of daily use are becoming increasingly virtual. They accelerate our problem-solving capabilities and provide us with fast and targeted information access. Interface abstractions and metaphors like windows, closed applications and mouse cursors were set in a time when computers used to be less powerful then today. While they are practical, many of those paradigms lead to a separation of presence in virtual and physical space.

This separation creates a binary state of focus, where being engaged with one space creates an interference of presence in the other space.

Spatial interfaces, compared to flat software front-ends, are more directly coupled with the senses. Hardware like mixed-reality headsets, display outputs arranged in space and spatial audio, makes software that is experienceable un-bound to a physical medium. Coupling with the senses provides a perceptual stage, where virtual and physical can effect each other in terms of interaction, physics, light and controlling.

This work is an exploration of how this blend of realities could lead towards a world where engaging with virtual and physical spaces can happen on a common stage and does not enforce presence in one space.

The initial research question is:

How can spatial computing lead towards interaction with meaningful augmentation of our minds and uninterrupted presence?

Design Opportunity

Mixed-reality headsets provide the opportunity to blend realities and give an impression on how a future with software perceivable in the world without intermediary hardware for manipulation could feel like. Key-enablers for the proposed spatial interfaces are:

- Superimposition of virtual objects onto a representation of user's field of view through camera passthrough and user's audition through spatial audio speakers.
- Semantic understanding of surroundings and behaviours happening in the physical world in real-time as an input system.
- Interpretation of this semantics to assemble interfaces and derive information.
- Real-time simulation of physics and hyper-physics[1] in world-space to give virtual objects the cues our perception is used to.
- Scene model of physical context for realistic occlusion, physics and lighting interaction between virtual and physical matter.

Flexible Trajectory

During the process the trajectory on the research question changed constantly and lead to its fragmentation into several clearer and shorter questions within the question. Keeping this flexible trajectory, instead of implementing a holistic product idea was chosen in order to enable generation of tacit knowledge originating from material talk-back of the explorations. The following questions represent those fragments in chronological order.

- What makes virtual and physical realities interweave?
- What makes mixed reality a balanced blended space?
- What makes contextual interfaces unobtrusive?
- What makes a perception gap, between people with different or no hardware smaller?
- Do augmented virtual objects allow interactions assembling around the intention?
- How might implicit voice input through conversation feel?
- How might agency of technology be made more transparent?

Values for Prototyping

With altering a person's reality through technology comes responsibility for designing those interventions. The following values are formulated to guide the building process.

1. Mixed-reality should stream-line the experience of human-machine interaction, while creating utility, joy and curiosity.
2. Virtual objects should saturate and satisfy human perception as much as possible and respect blended space to not stick out when superimposed on physical reality.
3. Virtual objects should not add unnecessary cognitive load to an experience and respect human attention whenever desirable.
4. Experiences should prioritise fostering connection between people in any reality.

Approach

Throughout the process a research-through-design[2] approach is used as a method to gain clarity about the nature of new ideas. When designing things through making, Mousette distinguishes between the actions of sketching and prototyping: "Sketches contribute back to its creator, prototypes radiate outward to the others and the world".[3]

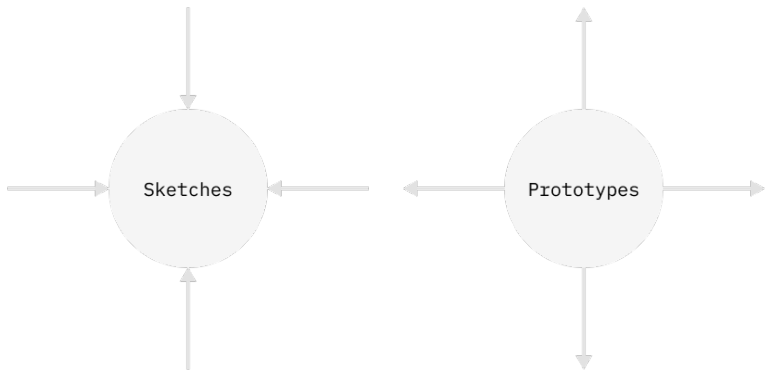


Figure 1: Radiation Pattern of Sketches & Prototypes

Mixed reality headsets, in combination with a game-engine, are used as a shapeless prototyping platform for different kinds of spatial computing archetypes during this project. This project is about the idea of direct interaction with software in perceivable reality. Available hard- and software is used as a prototyping platform during this project. It should not be seen as the final outlet allowing for adoption on scale of those kind of experiences. It should be seen more as an enabler to gain understanding of how those experiences could feel like further down the road once hardware is available that allows for extended or continuous use.

Following Redström's idea of technology as a material for design, early-on sketches help to gain understanding about this materiality.[4] They provoke a conversation with the material, due to their interactive nature. This conversation derives smaller sub-problems inhabiting the bigger over-arching initial idea. Those sub-problems often refine and concretise the initial idea. After sharing ideas and sketches with stakeholders, prototypes are built that fuse insights of the sketches into interactive artefacts for further communication. Those prototypes are attempts to answer the research question in higher fidelity.

Flexibility in trajectory is kept up to the latest point to exploit the maximum insight from the building process.

Stapper's idea that functional artefacts are where ideas touch-down on earth, stop being theory and start becoming reality is a guiding motif to this thesis.[5]

Definition of Success

The project is considered successful if the generated prototypes make my communication around the idea of embodied software in the world clearer. The prototypes should illustrate moments of this idea and provoke thoughts and discourse about how spatial operating systems could look like. It is my personal goal to further hone my skills in working with real-time rendering applications for experience prototyping and gain an understanding of the underlying technologies for creation of mixed reality.

Overview

Chapter 1: Reality & the Idea of Software in the World for a Higher Bandwidth between Humans & Computers

Chapter 1 introduces the topic through a brief definition of reality, revisits idea of tangible bits in the context of current mixed reality platforms and then extends on it through the formulation of embodied software in the world for a higher bi-directional information exchange rate between humans and computers.

Chapter 2: Materiality of Used Technologies & Spatial Interactions

Chapter 2 introduces sketches conducted to gain understanding about the underlying materiality of the technologies used to build contextual mixed reality experiences. Technologies for semantic understanding, scene interpretation, measuring intent, blending realities, persistence in positioning virtual objects, dynamic occlusion and integration of control systems are explored. Furthermore research insights from sketches and literature about spatial interactions through hands, eyes and voice are summarised.

Chapter 3: Interactive Prototypes

Chapter 3 revolves around the final design interventions that were created during the course of the project. The insights from three consecutive prototypes are summarised. Accompanying literature is reviewed to reason on tacit insights from exposure and usage.

Chapter 4: Articulated Outcome & Reflection

Chapter 4 summarises the insights synthesised from the interactive prototypes through the proposal of new interaction fundamentals for an operating system structure for spatial computing. It also revisits the process and outcome from the perspective of the brief.

Reality & the Idea of Embodied Software in the World

Virtual reality, mixed reality and augmented reality are common terms in the field of spatial computing. They are used to describe different forms of altering one's perception of reality through technology tightly coupled with the senses.

Cognitive scientist Joscha Bach defines reality as the following:

"It's similar to a game engine. It's a physics engine that you can use to describe and predict how things look in a particular way, feel when you touch them in a particular way, the law of proprioception, auditory perception and so on. So basically the geometry of all these things. And this is probably 80% of what our brain is doing — it's dealing with this real-time simulation."

(Bach, 2021)[5]

This idea implies that reality might not be perceived directly, but our experience of reality is a simulation constructed by our brains for guiding our actions from the inputs reaching our sensory organs. If technological actuation succeeds in saturating our senses to similar extends like the physical world it will become increasingly difficult to clearly draw a line between physical and virtual objects.

Tangible Bits Revisited

This project revisits Iishi's and Ullmer's paper "Tangible Bits: Towards Seamless Interfaces Between People, Bits and Atoms." The paper proposes tangible user interfaces (TUIs) as a close coupling of physical objects and environments to perceivable digital information. The intention of the interfaces is to take advantage of sensorial properties and affordances of the physical world to achieve heightened legibility and seamlessness of interaction between people and information. The absence of these couplings is critiqued for leaving a divide between the spaces of atoms and bits.[6]

This paper looks at the idea of coupling physical and virtual objects tightly, through the lens of mixed reality platforms available today and their capabilities integrating directly with our senses, "hi-jacking" physical objects for haptics, control and seamlessness (like TUIs) without neglecting advantages of modern virtual objects like scalability, transmission speed, accessibility and personalisation of experience.

Higher Bandwidth between Humans & Computers

The direct integration of virtual objects into our perceived realities suggests interactions between humans and computers with a higher information exchange rate than other interfaces. This information exchange rate is defined as the efficiency with which output is delivered and input is interpreted on both sides.

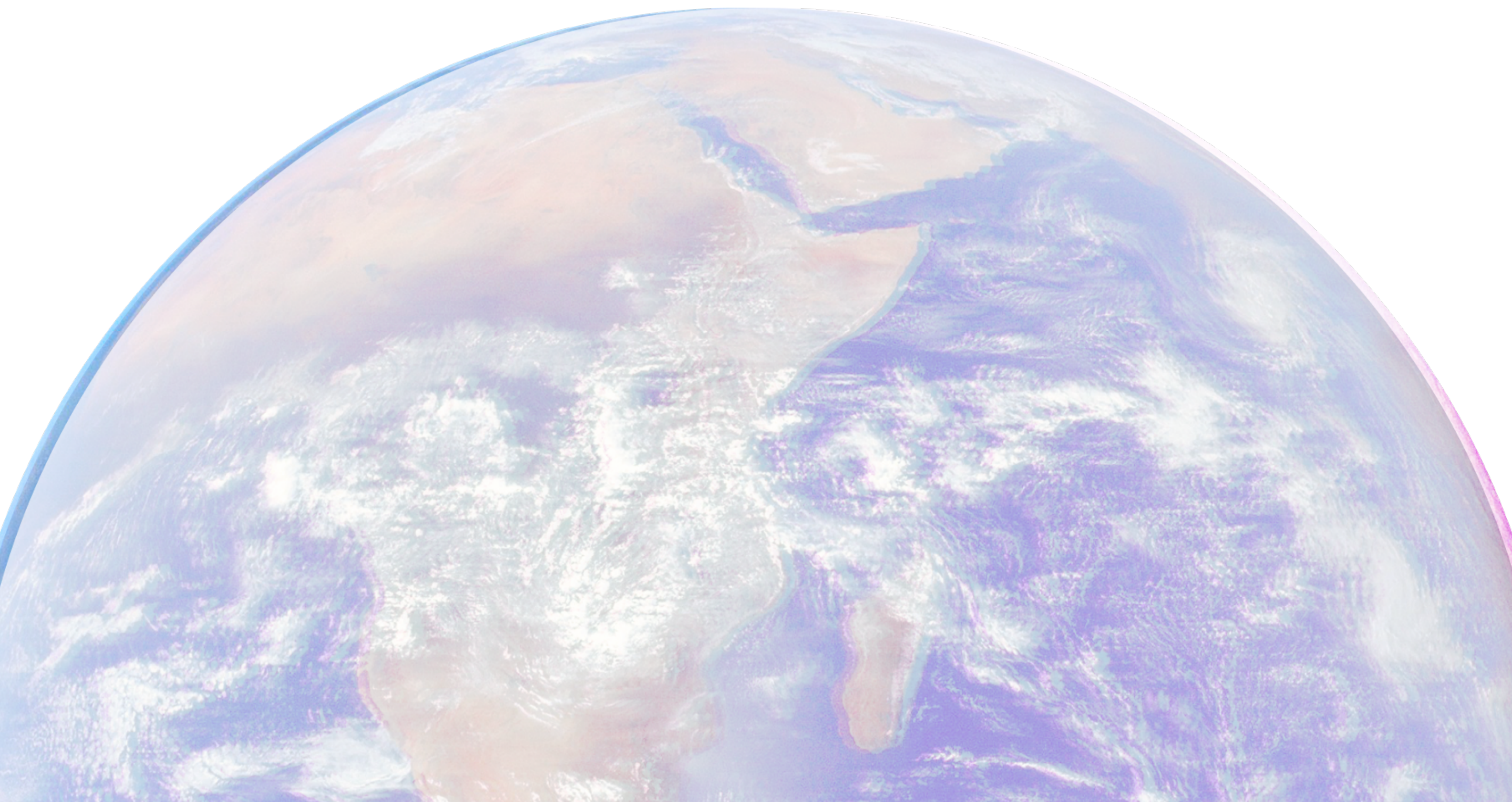
In 1962 Engelbart proposed computers as a way of augmenting human intellect through real-time interactive systems to expand on our own capabilities of problem-solving.[7] This proposition is a motivator for heightening the information exchange rate between humans and computers as higher bandwidths democratise intelligence, defined as ability to solve problems. Making this process more intuitive through accessible interfaces paints a picture were artificial intelligence is used for empowerment through augmenting intellect instead of automation for replacement of human labour.[8]

Embodied Software in the World

Intermediary hardware layers shape our interactions with machines. With spatial computing direct integration of the outputs with our senses and precise tracking of environments and context as input, opens up the possibility of having software embodied in the world as an interface.

This embodiment enables the creation of a space were physical, virtual and blended objects co-exist, to facilitate fluid experiences when interacting with technologies for different purposes. Arising interactions have inputs on a spectrum of ubiquitous to engaging and outputs on a spectrum of ambient to immersive.

Embodied software in the world leverages the advantages of both physicality and virtuality within a commonly perceived reality. It enhances physicality with traits we appreciate in virtuality, and vice versa.



Materiality of Used Technologies

The following part of this chapter puts words around the tacit insights synthesised in the process of sketching for material exploration. Recent technologies were used as a starting point to the process. The sketches were not targeted towards a specific product vision in the beginning. The goal was to provoke the maximum material talk-back to understand how the technology could be exploited for meaningful human-computer interaction and not force the material into a shape foreign to its nature.

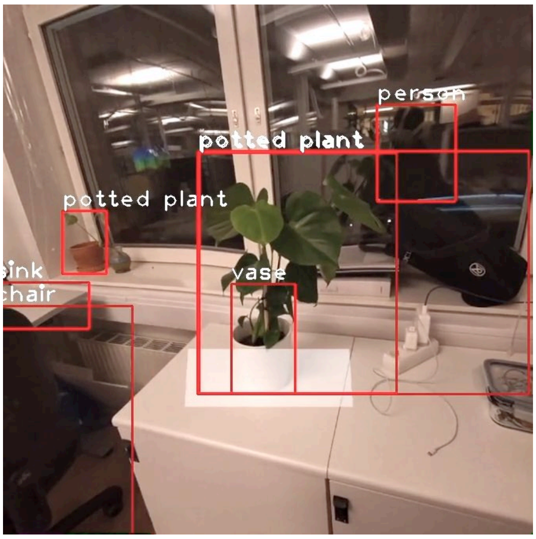


Figure 2: Object Detection Model running on FOV as Input

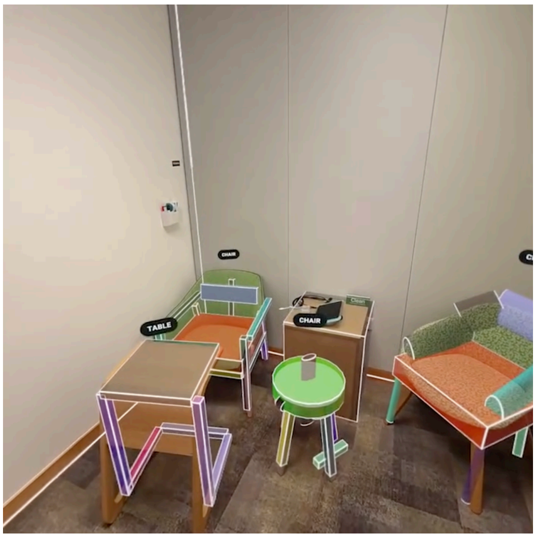


Figure 3: Meta's Scene Script Language Model

Scene Semantics

Scene semantics for mixed reality applications describe the idea of running object detection methods on the field-of-view, while simultaneously generating a 3D representation of the scene the user moves in to gain understanding of the context.

Methods for achieving this understanding include Meta's Scene Script, a method for representing and inferring scene geometry using an auto-regressively structured language model, and end-to-end learning. Scene Script is able to generate CAD-like 3D representations from the generated point cloud of the headset while simultaneously labelling the tracked objects with their names. This enables user experience with contextual user interfaces where interactions can become more implicit.[9]

To experience semantic scene understanding first-hand the object detection model Ultralytics YOLO V8 is executed via OpenCV on a laptop.[10] The video feed of the users field of view is streamed from a Quest 3 mixed-reality headset into that OpenCV application. The detected object classes and bounding boxes are then streamed into the Unity application via TCP/IP. In the Unity application running on the headset user interface elements are spawned based on combinations of object classifications. Meta currently does not allow for direct camera feed access on-device, which is the reason for the work-around. This is a bit of a shame, as it inhibits the process for developing contextual interfaces and other interesting possibilities.

This sketch validates the possibility of using scene semantics for understanding objects and context to spawn interfaces in real-time. For the sketch a user interface element is spawned in front of the user, as well as an audio source points out: "Give me some water!", whenever YOLO V8 detected an object with the class name plant. This feels alienating as the user interface and sound appears seemingly instantaneous and always when a plant enters the field-of-view, catering the attention towards that object. The sketch provokes the question, how spatial interfaces need to be designed in order to be unobtrusive when spawned contextually.

Context Interpretation

To assemble contextual interfaces an instance or layer interpreting the semantic scene understanding and translating it into machine- and then human readable output is required.

This is needed for generating information tailored to the context, as well as assembling the interfaces tailored to the generated information. Large language models offer potential for that use cases in a prototyping context.

For exploring the materiality of large language models for that specific purpose two sketches were conducted:

The first sketch is a Unity application with a terminal input and a particle system with a number of exposed parameters and a JSON parser that sets those parameters. The large language model has the system-prompt: “Interpret this metaphorical description of a particle system's behaviour and output your interpretation in only JSON with the properties: colour (RGBA), count (integer), speed (integer), drag (integer)!” This prompt results in consistent output of the metaphorical descriptions in machine-readable JSON format that can be used to dynamically assemble the visuals.

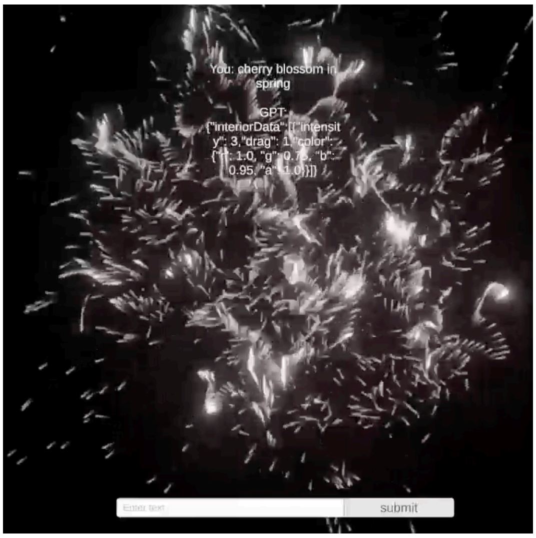


Figure 4: Prompt 1 → Cherry Blossom in Spring

The outcome is an interaction that enables to describe in human language how the visual output should behave and makes it adapt to this description. This sketch gives insight into the possibility of using the object classifications from a machine learning model instead of the terminal input to prompt a large learning model and through the system prompt adjust its output into the right format for dynamically assembling interfaces.

For the second sketch about context interpretation OpenAI's ChatGPT API was integrated into a Unity application that was deployable for the Quest 3. This sketch was not very insightful, however it was essential for prototypes further down the line.

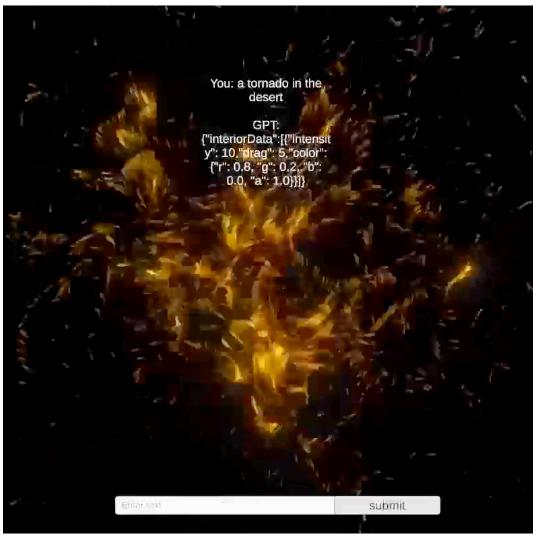


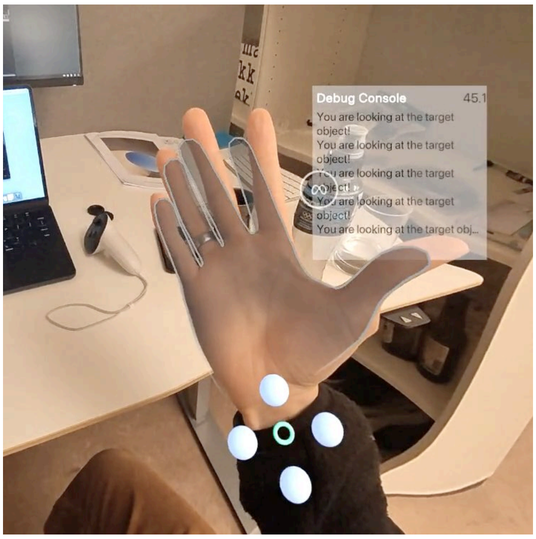
Figure 5: Prompt 2 → Tornado in the Desert

Using Gaze Direction for Intent Tracking

Unobtrusive contextual interfaces require the ability of the headset to determine the intention of the user. Eye-tracking can deliver reliable metrics for understanding intention.

Investigation into this data started from the insight that user interfaces solely spawned from the context a user moves in can feel obtrusive and distracting.

As the headsets available for this project do not have any eye-tracking sensors, an approximation of the user experience is built through a ray-cast projected from the centre of the field-



of-view that when certain parts of the scene model are hit triggers contextual interfaces and behaviours. Even this primitive form of tracking gaze leads to a significant improvement of perceived unobtrusiveness for contextual implicit interfaces.

The intent is treated as a form of world space hover state in the conducted sketches and prototypes. When the gaze hits an interactable target an unobtrusive small interface appears through a subtle micro-animation affording interactability. Depending on behaviour and position of the user further engagement with the interface can lead to expansion for balancing information density. When the gaze leaves the interactable the interface enters passive state again to minimise distraction when in the periphery of the user.

Scene Representation & Blended Space

To achieve meaningful virtual objects that adapt and influence to the perceived environments of the users a generic space is created that includes generated 3D geometry of the physical environment, semantic descriptions and world transforms of physical and virtual objects, a representation of the user in the scene and integration of control systems of physical outputs like home appliances. High accuracy of this generic model allows for deep integration of realities creating a perceivable blended space.

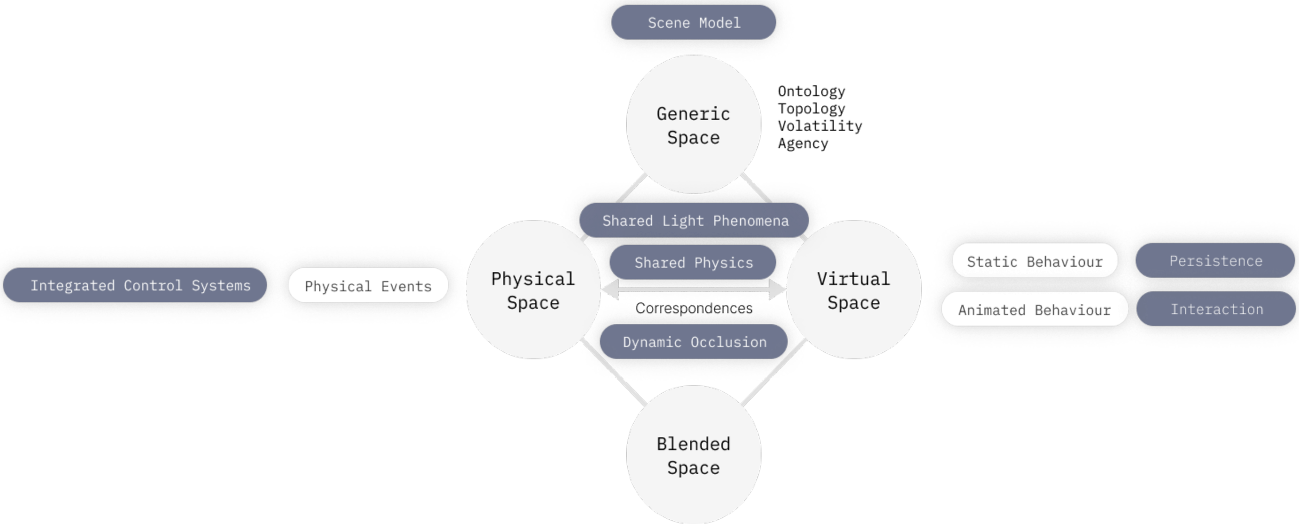


Figure 7: Expansion of Benyon's Blended Space Concept

Benyon defines the blended space as physical space with deliberate close-knit integration of digital space. Furthermore he maps out the importance of creating a blended space:

"The aim of blended space design is to enable people to feel present in a blended space, acting directly on the content of the blended space."

(Benyon, 2014)[11]

This idea is central to conducted prototypes later on, as it allows integration of virtual objects into perceived reality giving them similar weight as physical objects and reducing contrast between realities by a high magnitude of order.

Coupling of physical and virtual reality require some technical considerations. The following sketches were conducted to explore the technologies needed to achieve a blended space.

Dynamic Occlusion

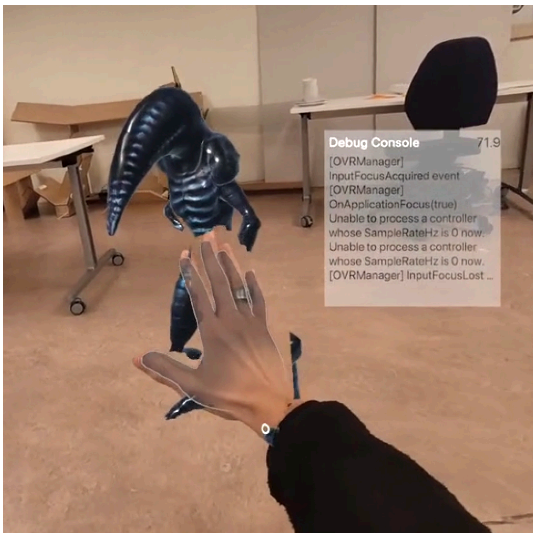


Figure 8: Using Meta's Depth API for Dynamic Occlusion

Occlusion is an unsolved problem for mixed reality interfaces. If a physical object does not have a geometry representation in the generic scene model it will be covered by virtual objects even if they are behind it in world space. Animated virtual and moving physical objects introduce more complexity to this problem. In most prototyping settings accurate hand models and consideration of the render queue is sufficient. To explore current possibilities a sketch with Meta's Depth API[12] was conducted. This approach to dynamic occlusion uses depth-maps to measure per-frame sensed environment depth to estimate the position of physical objects in world space to occlude in case a virtual object is covered.

The sketch showed that Depth API is not refined enough for experience prototyping yet as occluded objects have jagged silhouettes, however it is promising for future experiences as it might enable easy and efficient coupling of virtual objects with physical objects.

Simulated Physics

Game physics are laws of physics defining behaviour of virtual objects in mixed reality setups. They can be defined resembling physics of the physical world but can also be designed exhibiting hyper-physical phenomena as defined by Crawford.

This opens up new areas for designers that were unexplored till lately. The possibility to alter perceived physics in blended space gives virtual objects expressive properties that can be used as a

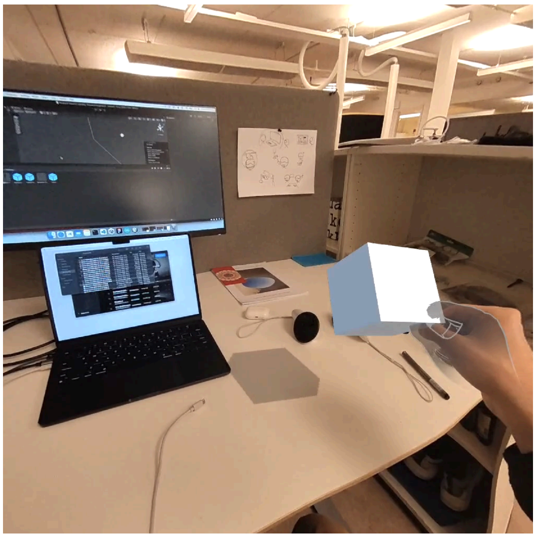


Figure 9: Using Scene Model & Shader for Shadow Interaction between Physical & Virtual Space

canvas for many aesthetic functionalities like affordance, indication, preview, effect, distortion of materiality and many more. Running physics simulations where virtual objects, through the generic scene model or the avatar, seemingly converse with the physical world and vice versa can lower reality contrast and create an integrated user experience.

Simulated Light & Shadows

Simulating light and shadow interactions between both physical and virtual realities influencing blended space is an important building block of integrating realities.

To illustrate the behaviour of shadow interactions a shader is used on the scene model, which grabs the shadow pass and renders it onto an otherwise transparent mesh. A grabbable cube is then added to the scene that interacted with the directional light source in Unity. When grabbed and dragged around the illusion is created that the cube throws a shadow onto a physical table.

The shadow helps reasoning about the spatial positioning of the object in physical space as it gives a second point of reference, even-though the light source came at a steep angle. The virtual directional light source in the setup does not match a physical light source, however in the environment this sketch is conducted in with many different lights this did not become apparent.

Accurate light reflections between realities are sketched through a similar approach where a shader is used that additionally to the real-time shadows also renders specular and diffuse lighting to the transparent mesh. This enables virtual lights to effect blended space. Meta's Quest 3 platform does not allow direct access to the passthrough layer which prevents accurate physical light representation on virtual objects through a real-time skybox and accurate rendering of translucent objects. As a work-around point-lights representing physical light-source and an HDRI for ambient lighting and a directional light source resembling sun-light are used. This setup enables the experience of lighting of the physical space influencing virtual objects.

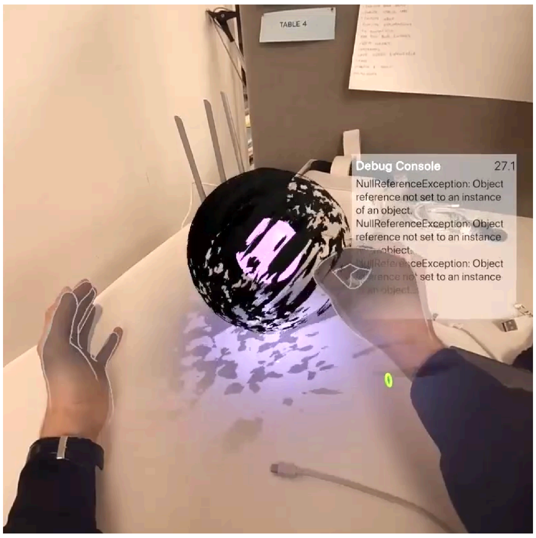


Figure 10: Expanding Setup for Light Interaction

Reality Contrast

Reality contrast is occurring between virtual and physical objects when experiences are not blended in terms of physics, light and interaction. In general it should be kept as low as possible because objects with high contrast guide the users attention towards them. A high contrast can be used stylistically to guide attention for non-persistent objects.

Integrated Control Systems

Integrating control systems of physical products into blended space adds accessibility, potential for user experiences and democratisation beyond goggles to spatial computing.

The first iteration of sketching integrates an Arduino Uno that is controlling an Adafruit LED ring with inputs from a Unity application via TCP/IP and allows to send different hard-coded RGBA colour values by pressing virtual buttons in the application running on the Quest 3. The resulting experience got positive resonance from people trying it. Being able to control the "real" world with a virtual object feels new.

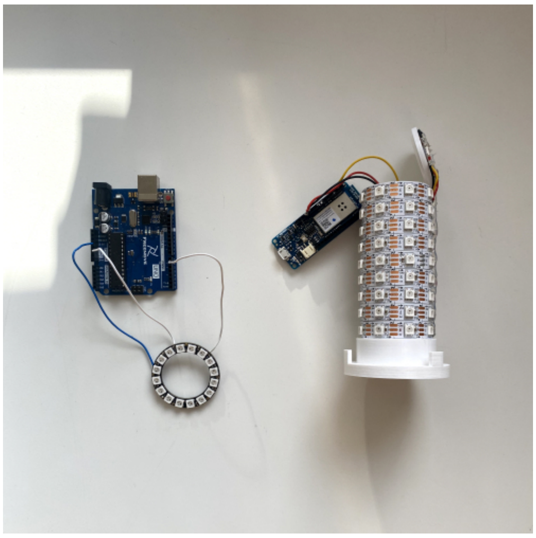


Figure 11: Iterations of Lamps with Integrated Control System

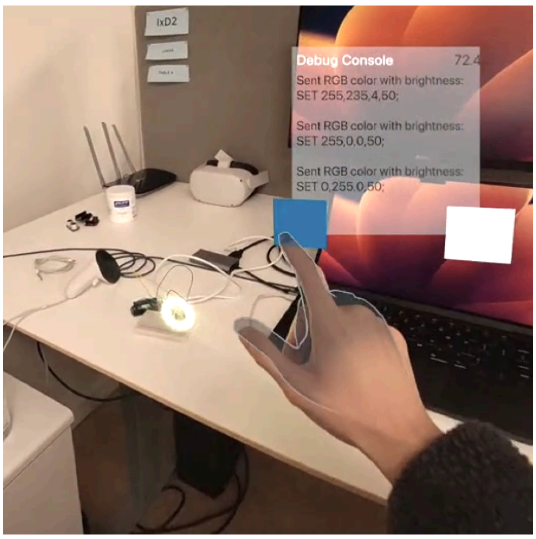


Figure 12: Changing Light Colour of Lamp with Virtual Buttons

Manipulating objects perceived from outside the headset creates a reality perception gap. The layout of the buttons in space provokes an expressive poke interaction. This leads to different perceptions of the interaction from headset perspective vs. from outside perspective. In first-person view the virtual interaction elements are clearly visible. The layout in front of the user is accessible and invites for interaction. The interaction itself is satisfying as the colour change is immediate and accompanied by auditive feedback of pressing the button. In third-person view the interaction looks irritating. The user is poking in the air with no visible interaction elements and no feedback, except the lamp changing its colour at the tipping-point of the poke gesture.

Reality Perception Gap

This disconnect between perspectives of people wearing compatible hardware and people wearing non-compatible hardware or no hardware is a challenge for extended mixed reality adoption. Resulting cleavage between people able to perceive virtual objects and people unable to do so often makes mixed reality an isolated first-person experience.

Choreography of Interaction

Figure 13: Poke Hand Interaction can look irritating for Observers without Context

When designing hand interactions the choreography of gestures are perceived from inside and outside perspectives. People without hardware will perceive interactions out of context. This can be irritating for users and observers. Therefore the choreography of hand interactions should be carefully considered. Taking parts of the experience out of the headset can make interactions with the technology less irritating by providing context. Another solution can be micro-interaction with very little hand movement.

Spatial Interactions

The second part of this chapter summarises the insights that were synthesised from a secondary literature research on spatial interactions after the initial insight from sketching.

World Interaction Space & Hand Interaction Space

In mixed reality Wu draws a line between two different types of interaction spaces.[13] They are defined as world interaction space and hand interaction space. Without using tools physical reality allows for manipulation of objects that are within the bodies reach. With virtual objects this is different as interactions are not restricted by physics. Through sequential gesturing virtual objects are interactable even if they are not positioned within the bodies direct reach. If an interaction is done that is out of reach of the body, it is in world interaction space. Blended objects incorporate properties of both interaction spaces. Eventhough their physical structure and world transform is bound to the world's physics, their manifestation and their virtual output can be altered from a far.

Figure 14: Deictic Gesture

Tracking Gestures

Gestures can be defined as a movement of the hands, head, or body to express intention, ideas or emotions. Through advancements in hand- and eye-tracking, hand gestures and gesture-gaze combinations are becoming popular input mediums for spatial computing. A distinction is made by Karam and Schraefel[14], between deictic, manipulative and semaphoric gestures and extended by Pfeuffer with gaze-enhanced gestures.[15]

Deictic Gestures

According to Karam and Schraefel, deictic gestures use acts of pointing to identify or place objects in space. Deictic gestures are commonly used input in spatial computing. An example would be a ray pointer interaction where a specific pose of the hand makes an ray appear between index finger and thumb which can be used to precisely target a position in world interaction space.

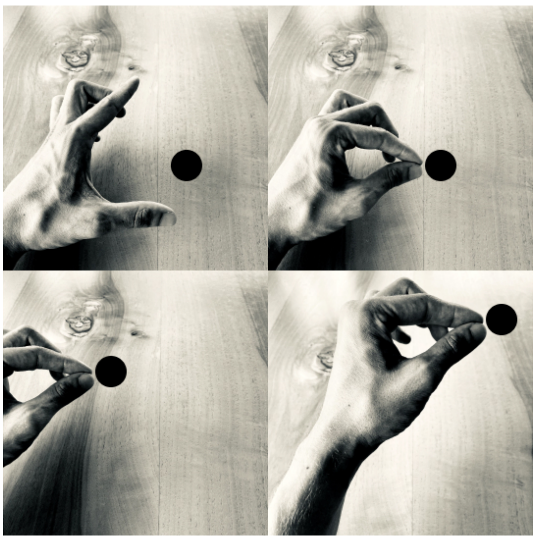


Figure 15: Manipulative Gesture & Sequence

Manipulative Gestures & Sequences

Manipulative gestures are performed with the goal of changing an object. They can be single gestures like pinch-to-zoom, but in spatial computing most often are sequential because of the added complexity of a third dimension. In world space interactions an example for a manipulative gesture sequence would be a point and pinch sequence. Here a grabbing position is specified through a ray and then the object is grabbed through a pinch gesture and transformed in world space through movement of the hand.

Figure 16: Semaphoric Gesture

Semaphoric Gestures

Semaphores are gestures that are pre-defined with meaning for interaction. Usually they are hand poses like for example a thumbs up communicating a like.

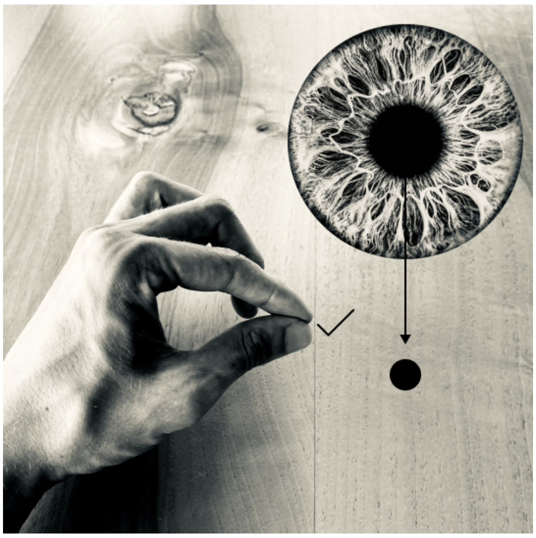


Figure 17: Gaze-Enhanced Gesture

Gaze-Enhanced Gestures

The term eye-hand symbiosis was coined by Ken Pfeuffer. Pfeuffer describes it as a form of human-computer interaction that combines data from hand- and eye-tracking to extend beyond the limitations of the individual modalities.[15]

A combined input modality is the gaze and pinch interaction. Here the precise gaze direction is used as a deictic gesture communicating attention and intention and the semaphoric pinch gesture is used as confirmation. Combinations like this are especially suited for world interaction space, as they enable quick interactions for example for hierarchical menu structures without using tiring arm movement deictic input requires.

Tracking Eye-Movements

Many of the current mixed-reality headsets include multiple inward facing infra-red cameras to track data giving clues about gaze direction and are potentially able to track pupil dilation and blink rate.

This data is a comprehensive window into the mental state of user’s and allows the computer to better understand intention, attention and health-related clues. It is a building block for creating implicitly triggered contextual interfaces that are not feeling intrusive. Unfortunately the hardware available for prototyping does not include dedicated infra-red cameras. This lack of data is bridged by sending a ray-cast from the centre of the field-of-view towards the objects in front of the user, which gives a rough approximation of the possible user experience.

Dedicated eye-tracking rounds up interaction patterns by giving deeper behavioural cues for adjusting interfaces by analysing the

Figure 18: Fixations last between 50 & 600ms

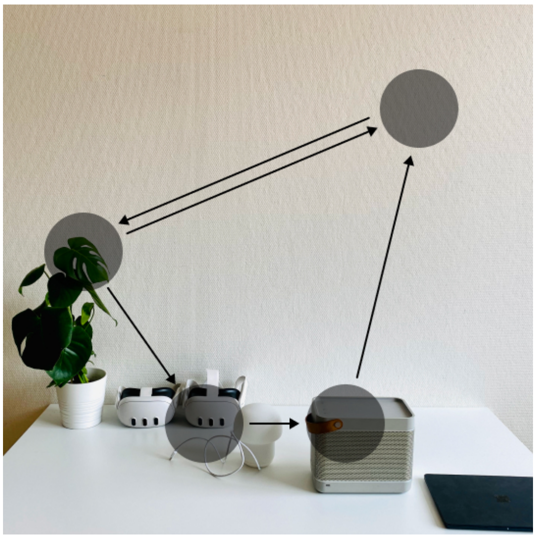


Figure 19: Saccades can be observed when a Person makes Sense of their Environment

Figure 20: During Smooth Pursuits moving Objects are tracked & kept in Focus

following eye-movement motifs as defined by the eye-tracker company Tobii.[16]

Fixations

Fixations are periods between 50 - 600 milliseconds of continuous focus of the gaze onto a static object. They can be observed when a person wants to maximise the visual information gain from a focussed object.[16]

Saccades

Saccades are classified as rapid eye-movements with switches in focal length from point-to-point of a scene. They are observable when a person needs to make sense of an environment combined with interruptions of fixations.[16]

Smooth Pursuits

Smooth pursuits incorporate continuous movement of the eyes combined with adjustment of focus in order to track a moving object.[16]

Privacy Considerations with Eye-Tracking

Eye-tracking used in a mixed reality context enables to collect data about intimate and unconscious responses to real-world cues. This allows for modelling individualised attention and behaviour models with high accuracy based on behaviours exhibited in both virtual and physical realities. These are valuable for novel ways of human computer interaction and ad-driven business models.[17]

Transcribing Voice

Real-time transcribed spoken language can be an effective and accessible input for spatial interfaces. While building, implicit inputs, where conversation between users spawns contextual interface elements are compared to explicit inputs, where task are directly instructed to the computer. Implicit voice input in combination with other contextual input offers potential to unobtrusively trigger virtual objects without diverting attention.

Interactive Prototypes

After primary and secondary literature review and building a total of twelve sketches for material dissection of spatial computing, three refined prototypes in one common space were crafted for further exploration of the potential of embodied software in the world.

A Lamp in Blended Space

This prototype combines Benyon's ideas of “Blended Space” and Iishi’s an Ullmer’s ideas of “Tangible Bits” to create a desk lamp that leverages traits of physicality, like haptics, perceivability without hardware and influence on physical reality as well as virtuality like accessibility, fluidity in appearance and fine-grain real-time reaction and exists as a blended object and example of embodied software in the world.

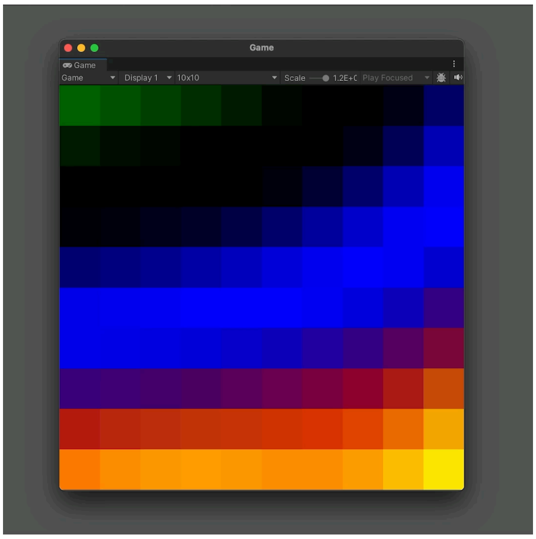


Figure 21: Unity Game View mapped to the LED Matrix

Figure 22: Lamp displaying Shader from Game View above

Figure 23: Different Shaders displayed on the LED Matrix

Technical Setup

The setup consists of a 3D-printed inlay, to replace the standard bulb of an IKEA Tokabu desk lamp. This inlay carries a total of 70 NeoPixel LEDs. This LED matrix is controlled by a Genuino MKR-1000 microcontroller with WiFi capabilities. The Genuino establishes a network server via TCP/IP and runs logic for handling mapped RGBA values to drive the matrix like a display from a Unity camera output.

A light-weight Unity application setup with the build-in render pipeline running on a M2 MacBook Air subscribes to this server as a client. This application contains logic and a custom renderer feature that translates a camera output displaying a shader to a byte array containing the RGBA data to send it over the network with low latency and control the matrix at an adjustable refresh-rate.

The displayed shader can be configured to react to data input coming from another Unity application running on the Quest 3 headset. The application on the headset handles the rendering of the virtual objects as well as generation of the generic scene model necessary for blending realities in terms of lighting and hyper-physical phenomena. It also measures intent of adjusting the lamp's settings based on gaze direction. It provides a contextual palm menu to adjust brightness and colour when in world interaction space and contextual buttons when the index finger moves towards it in hand interaction space.

This split setup with two separate applications was chosen because the universal render pipeline used to run the Quest application for fidelity reasons did not support a C# functionality for converting the pixel data into a byte array. A better setup would run entirely on the headset and just off-load the driving of LEDs and network configuration to the Genuino.

Tacit Insights

This prototype expresses how embodied software could be infused into physical reality to create experiences that allow for fluidity of blended object's appearance and meaning but also saturate and satisfy human perception to make virtual objects not stick out as artificial overlays. The following points were observed while interacting with it:

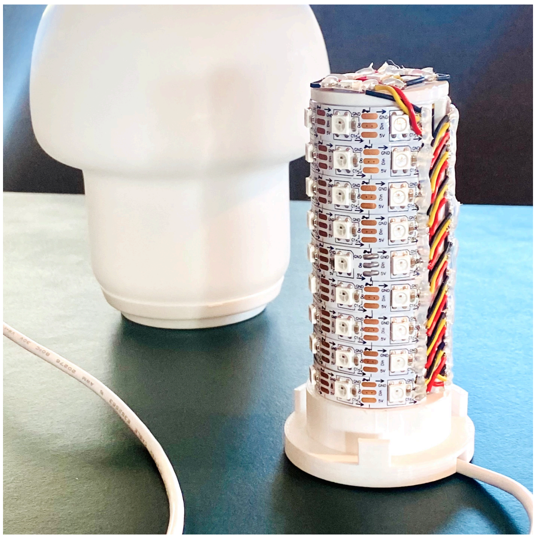


Figure 24: 70 NeoPixel for enough Resolution

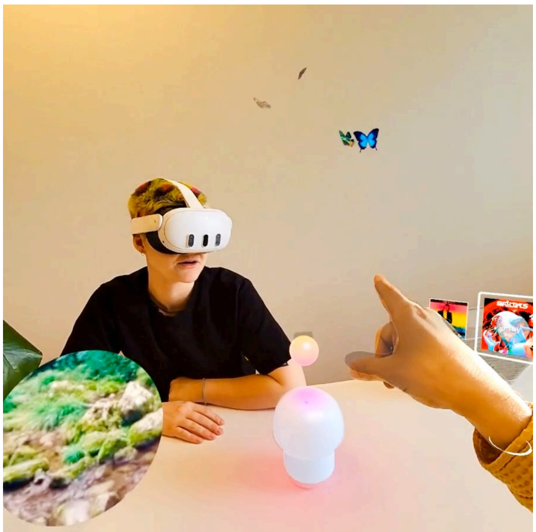


Figure 25: Butterflies interact with Shader mapped to Lamp

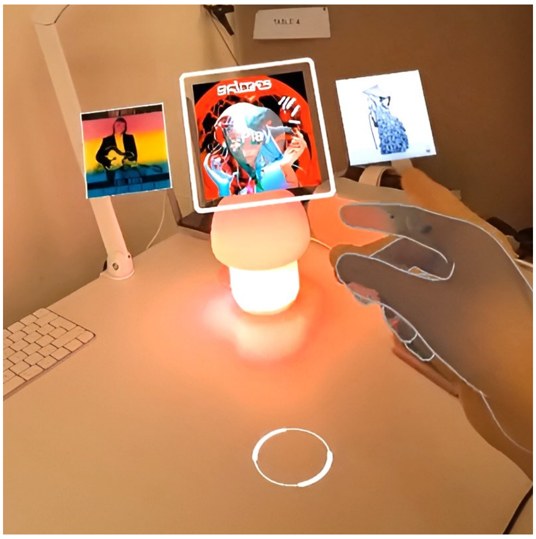


Figure 26: Light Colours change according to Colours of Cover

Output Resolution & Bridging Realities

The amount of LEDs was chosen consciously to have enough resolution available for changing the lamp's appearance in fine enough detail to communicate meaning and display basic applications on it. This is important for taking parts of the experiences happening in virtual reality beyond being perceivable only for users wearing mixed-reality headsets.

Limited communication of virtual content outside the headset can act as a bridge for people without goggles, however the input bandwidth of mixed reality headsets is hard to bridge fully.

Continuity of Experience

Being able to drive outlets beyond the head-mounted display from a real-time application enables creating continuity between experiences that lets different applications/virtual objects interact with each other. Prototyped examples include the butterflies spawned in another virtual object that then subtly guide the focus towards the lamp by flying there and interacting with emitted light of the lamp through reflections on their wings and throwing shadows to the floor.

Another example integrates a virtual music player discussed in the next chapter. When in proximity with the lamp switching through the album covers changes the colours of the lamp to match the artworks.

Scenarios of continuity between applications, virtual and mixed objects turned out to be promising for creating experiences that do not fragment reality into constraint spaces, but instead approach it as one space where those objects co-exist and cross pollinate each other.

Interaction Choreography

One observation from early sketches of controlling a lamp with virtual objects was that arrangement, positioning, sizing and type of the interface influences the choreography of interaction when engaging with virtual objects.

For the first sketch buttons, for switching the colour of the light source, were positioned right in front of the user. This interface required an expressive hand motion poking into the air for interaction. This poke motion felt unsatisfying because of lacking haptic feedback. From a third perspective the motion looked like an un-rational action.

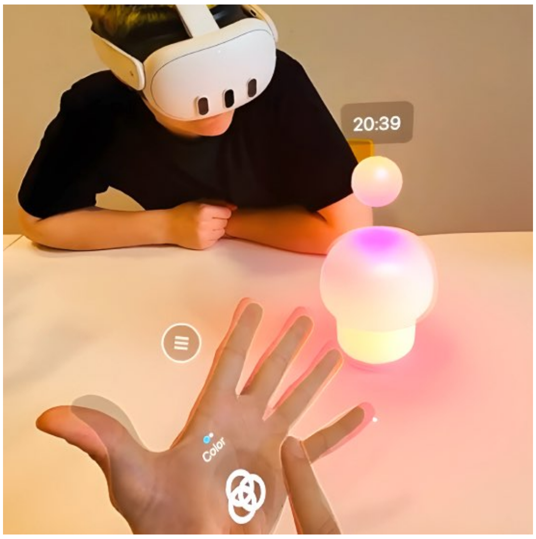


Figure 27: Contextual Palm Menu Triggered by Gaze Direction

By combination of gaze and proximity for determining intent and two contextual menus in shape of little buttons spawning right in the position of the lamp for close range interaction or on a palm menu for far range interaction a more subtle choreography of interaction is defined. The interactions picked were chosen for minimal irritation to outside observers. When switching colours with a finger in close proximity to the lamp, the movement is visually bound to the blended object. When switching colours with the palm menu the movement is minimal so the gesture becomes unobtrusive to outside observers, with the added benefit of hijacking the palms haptics.

Accessibility

Virtuality of objects enables for interaction in hand and world interaction space. This is also true for blended objects and has potential in terms of accessibility for people with limited mobility. Another enabler is the ability to update interfaces of blended objects based on usage data over time. This might enable future interfaces tailoring around specific needs of the individual.

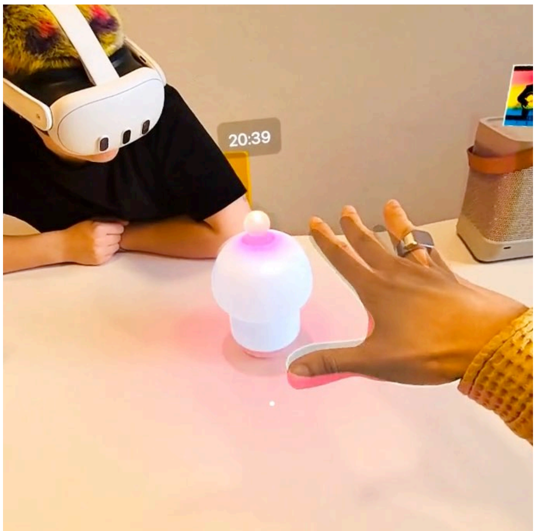


Figure 28: Transitioning Interaction States between Realities

Transitioning Interaction States across Realities

Physics in blended space can be used for transitioning between interaction states where one state is based in virtual and the other state is based in physical reality. This is useful for coupling realities, to indicate that an action influenced a physical object through a virtual object and vice versa.

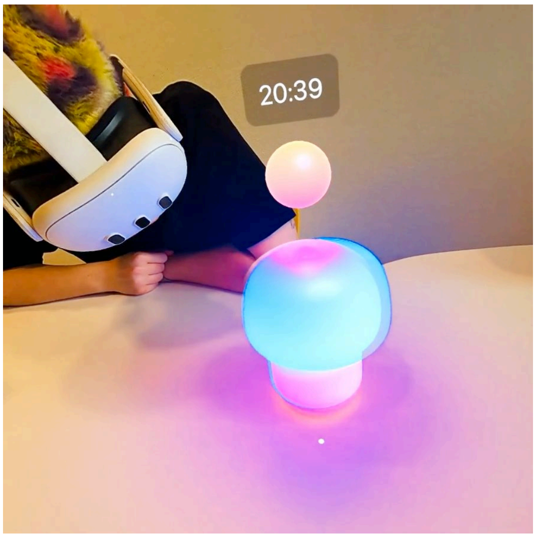


Figure 29: After the Droplet UI falls into the Lamp the Light Colour changes non-uniformly & Ripples appear on its Surface

Here a spherical interface element resembling a colour droplet is floating on top of the desk lamp as an indicator for a colour change. Once a fitting colour is selected and confirmed the virtual droplet falls into the physical lamp. On collision with the lamp a vertex transformation is applied to the scene model of the physical lamp through a shader. This deformer shader suggests that the droplet falls into the lamp, that itself has the material properties of a liquid, and creates some ripples on the lamps surface before mixing with the "lamp liquid" changing its colour.

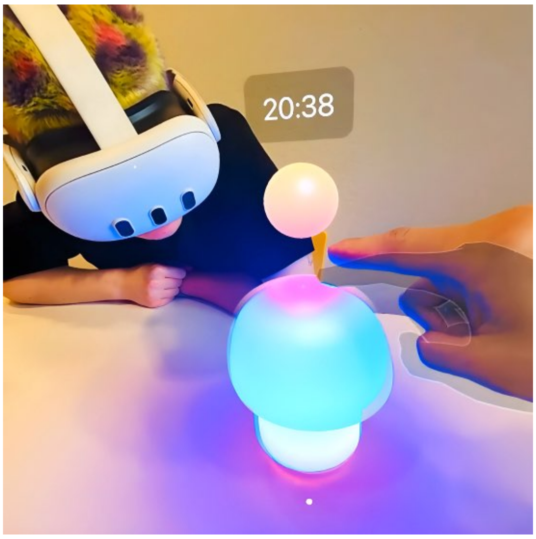


Figure 30: Vertex Deformation Shader animates on Scene Model when the Physical Lamp is touched

Altering Materiality in Hand Interaction Space

In hand interaction space simulated physics can be used to suggest causal relations with a mixed object. This can bridge a lack in detail of haptic feedback when hijacking haptics of a physical object or alter perceived materiality. An example would be a virtual button projected onto the scene model of a physical table. The lack of physical travel to an extent can be compensated through a visual deformation of the tables surface in combination with audio feedback.

In the lamp prototype this effect is used when the surface of the physical lamp is touched directly. This does not add any functional value but it creates consistency in altering the materiality of the lamp.

Physics in Combination with Integrated Control Systems

If a mixed object has high enough display fidelity and can be driven in real-time, physics can be used to bridge parts of the experience beyond the headset. This can create continuity of experiences as the real-time application adds fluidity to possible control, appearance and meaning.

In case of the lamp once the virtual colour droplet fell into it it does not uniformly change its colour but emphasises on the liquid materiality by mixing the colours.

Reality Blending with Light & Shadows

The prototype makes extended use of a precise scene model for sharing light and physics phenomena across realities. This approach is chosen as it seems necessary for creating virtual objects satisfying perception and not sticking out of perceived reality compared to physical objects. Virtual butterflies spawned around the lamp illustrate the dynamic shadowing on the desk and lamp itself. A virtual light source is created mirroring the physical light source to throw light onto the butterflies and desk. Those measurements enable a low reality contrast leading to the desirable effect of blurring the boundaries between physical and virtual objects. Exposing footage of the prototype to people showed that they could not precisely tell which parts are real and which parts are simulated.

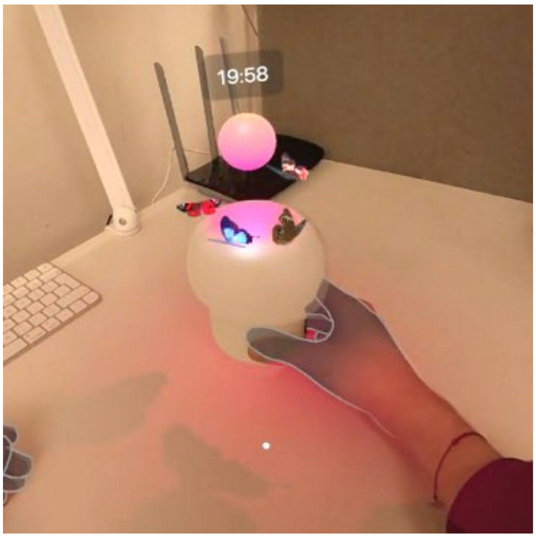


Figure 31: Light, Shadow & Occlusion between Physicality & Virtuality enabled by the Scene Model

A Music Player with Liquid Interactions

This prototype is inspired by material talk-back from previous interventions. When transforming the lamp to a blended object, the fluidity in appearance, meaning and application and the possibility of real-time reaction to inputs became apparent. This fluidity has potential for accessibility and affordance. The goal is to create a blended object that does not dictate its interactions but assembles affordance around the user's needs.

Technical Setup

The setup consists of a virtual music collection of three albums and a speaker represented in blended space. To reduce development time and complexity the functionality of the music player for the collection was reduced to transitioning between passive and active state, browsing through albums, playing and pausing them and switching audio output from the headsets speakers to a speaker in the room.

Initially the music collection is in its passive state where the album covers are stacked on top of each other. On pressing on top of the stack the player expands in an animated fashion and the albums are arranged next to each other. Also a play/pause button appears in active state. If the player is not used and not engaged with for 30 seconds it cycles back into passive state.

Positioned around the album covers there are three spherical handles, that are invisible in their base state. After a proximity threshold between the user and the player is reached they monitor the distance between the tip segments of the index fingers and the handles. If the index fingers are getting into a range of 20 cm to the handles their visibility increases smoothly.

For gestures and poses the prototype makes use of Meta's Interaction SDK[18] to derive and process data about interactions from the hand tracking. Used gestures are poke for pressing buttons like activation, play and pause, pinch for the handles, 2 handed pinch for scaling the player, grab for the whole player and swipe for close- and far-range swiping through the album covers. A custom poke and rotate interaction was created to change where the album covers are facing towards. Used poses were a pointing pose as a play trigger and a stop pose as a hold trigger.

Another possible form of interactions is voice transcription either through direct commands e.g. saying “play” and “pause” or indirect implicit command e.g. talking with someone about the music at a specific event. This was prototyped using OpenAI's Whisper API[19].

If the player is positioned atop a surface with the semantic description of table or shelf within a distance range of 0 to 40cm, monitored through a sphere-cast, it projects a circular interface element onto this surface. This element can be used for rotating the covers to face into a certain position.

The rotation functionality is triggered by a poke gesture, which retrieves the touch position as a Vector3. The system computes the rotation direction between the touch position and the circular interface position. Next, it calculates the angle between the touch direction and the forward direction of the world. This angle is then adjusted to ensure it falls within the range of 0 to 360°. The current rotation of the circular interface is determined, and finally, the rotation of the interface and the virtual objects is updated with the calculated angle.

To integrate the music player with blended space a spatial audio interaction for switching between music playing on the speakers and headphones is prototyped. When music is played from the virtual object the sound emission's centre is its transformation in world space. If the player is grabbed and moved towards a speaker after a proximity threshold the audio volume from the headphones decreases in sound and increases on the speaker till it is playing on the speaker and silent on the headphones.

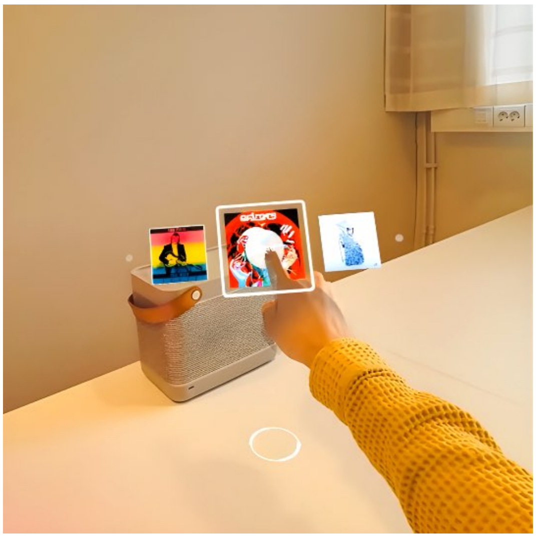


Figure 32: Poke Gesture for Playing a Song

This is achieved by having a second Unity application that communicates with the headset's application via TCP. Both Unity applications have audio sources that mirror each other in terms of track, play/pause state and timeline. The Unity application running on the headset has logic for determining the distance between the physical speaker and the virtual object. When the virtual object is moved in proximity to the speaker the volume of the audio source running on the headphones is turned down and the volume of the application for the speaker is turned up.

Tacit Insights

This prototype explores how virtual and blended object's fluidity in their experiential properties could be leveraged for accessibility by creating liquid interactions that form around the user's intention and needs. It illustrates how through multi-sensory spatial interactions, control of blended objects can be executed and how spatial responsiveness can increase ergonomics and precision. The following points were observed while interacting with it:

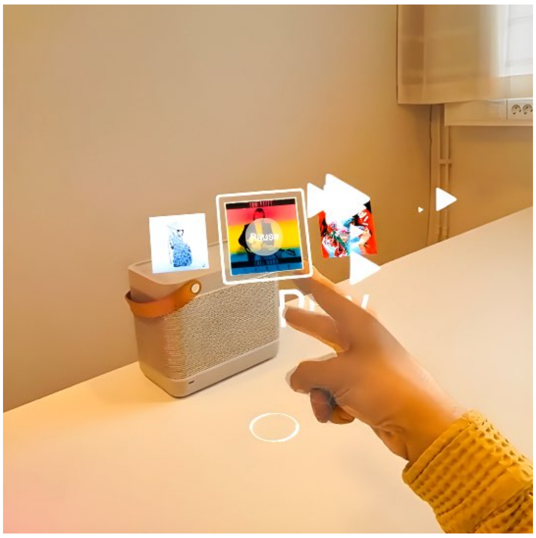


Figure 33: Pointing Hand Pose at Player for Playing a Song

Liquid Interactions for Accessibility

Embodied software in the world has the potential to make experiences of both virtual and blended nature more accessible at their core. While the accessibility potential of today's

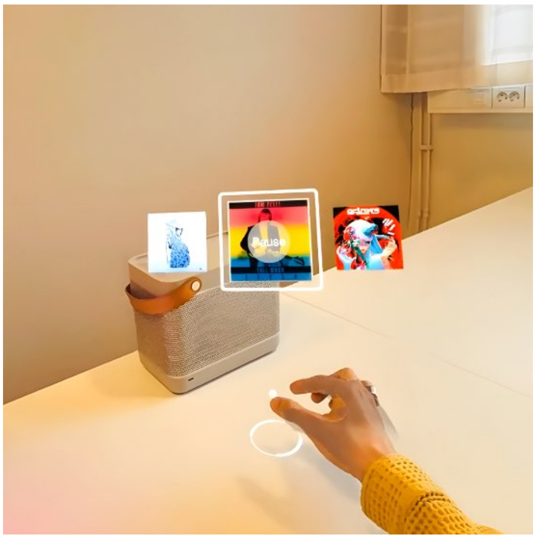


Figure 34: Handle is only afforded when Proximity of Finger signals Intent

virtuality is not exploited because of the intermediary hardware layer's dictating predefined interaction patterns and metaphors, fluidity of embodied software could create interfaces to the world that are tailored towards every individuals needs. The ability to fit a multitude of different types of interaction into embodied software creates a rich palette for assembling those interfaces.

With embodied software the goal of a set interaction can be constantly evaluated and the affordance can be adjusted in real-time to guide towards this intended goal. Imagine a hammer that is not consisting of a head and a handle, but distributes its weight based on how it is used to drive a nail into wood, while being able to dynamically shape a multitude of handles for all different hand sizes and shapes as well as needs.

Intersection Errors & Interaction Volume

A problem with virtual objects with a high saturation of interactions is that some of the incorporated interactions can be similar but inhabit different functionality. This can lead to intersection errors where interactions are misinterpreted and trigger the wrong functionality. Hand gestures and poses are prone to those errors as often the same or similar local hand poses can carry different meaning based on global hand movement. An example in the prototype is the similarity between the swipe with extended pointer finger and the point to play interaction, that can lead to triggering play while swiping.

To prevent intersection errors when designing virtual objects interactions should be carefully compared with each other. If there are similarities between them more contextual data must be used for distinction.

Hijacking Haptics

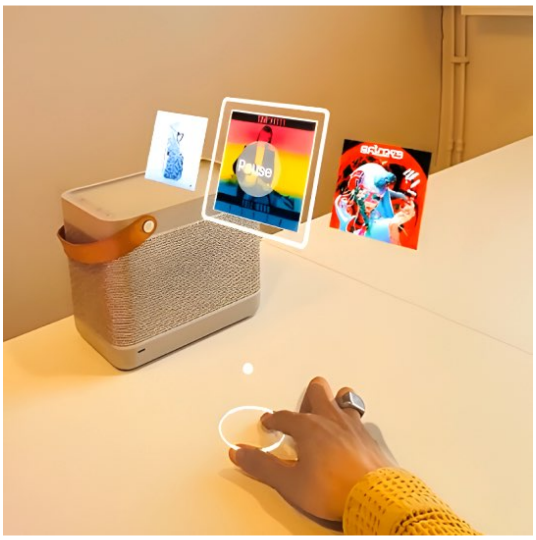


Figure 35: Rotation UI is projected on Table below to hijack its Haptics for precise Interaction

When changing the music player position in world space, sometimes its rotation on the global upwards axis needs to be adjusted. In an initial attempt to fix this problem an approach was picked where the rotation adjusts automatically to the position of the field-of-view camera. This turned out to be rather distracting as the result is an object moving in the periphery with every head movement. Also an object constantly "staring" at the user can feel rather discomforting for them.

Spatial rotation manipulations can involve two hands for locking the axis. The music player incorporates a two grab manipulation which turned out to be unnecessarily engaging for most use cases as it requires both hands and enables complete freedom of rotation.

When using the spatially responsive circular interface for rotation the haptics of the table are hijacked and only the upwards axis is unlocked. This enables precise rotation that just requires one hand. It is useable when in the periphery as the haptics are catering for missing visual cues.

Spatial Gestures for Controlling Blended Objects

Embodied software enables more seamless interfaces with blended objects, as virtual and the physical objects are inhabiting the same perceived space and can be put into direct causal relation with each other in terms of control and embodiment.

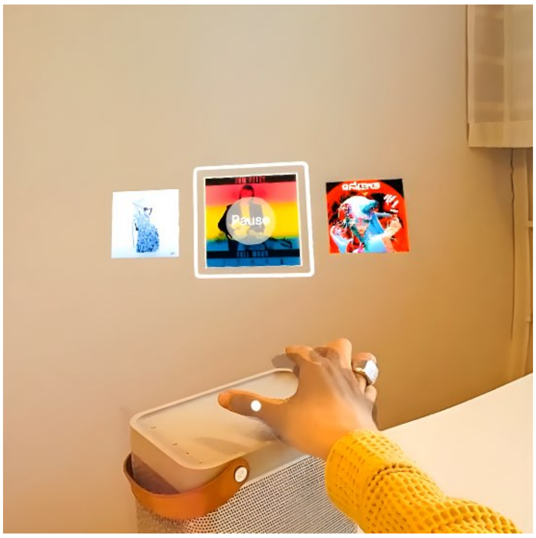


Figure 36: Moving the Player on the Speaker gradually changes its Audio Output from the Headphones to the Speaker

The spatial audio hand-off illustrates this idea by using sound and proximity of the interface to the blended object for control. The result feels satisfying as it adds a ritualistic/ceremonial dimension to the interaction, comparable to putting a LP onto a record player which is missing in other screen-based interactions with music.

Spatial Causal Relations

When using a smartphone for interacting with a speaker all the interaction happens as an abstraction in virtual space. The resulting large supply of music, efficiency and convenience, is traded in for the ceremonial ritual a vinyl player would offer. Through embodied software in the world, intricacy in interaction with directly perceivable causal relations for interactions set by the user can be revived in a virtual context.

The arising enhanced perceived meaning can hopefully introduce more transparency, visible agency and satisfaction into interaction with virtual objects. Appreciation for virtual objects should reach similar levels like for physical ones.

Spatial Responsiveness

Moving the player into different spatial contexts is used as a matter of changing its state and function. To create an experience where the virtual object uses its position and context in the world it needs to respond to it in terms of functionality and shape. This process is comparable to responsiveness in web design, where a website is viewed differently depending on the device used to display it.

One example for spatial responsiveness incorporated in the prototype is that the covers are changing in size if put on top of the player. Another one is the circular interfaces projected onto a table below it.

A Portal into Memories Triggered by Conversation

This prototype was created as an immanent loneliness was explored in the fact that all previous explorations are only fully experienceable in first perspective. Another motivation was to explore if interfaces spawned from implicit inputs can be perceived as unobtrusive. For this prototype is a portal into a past memory. It is spawned contextually if a conversation indicates an intention to share a memory with another person. It can be engaged with on a spectrum of immersion. In the portal a particle system is encountered that remains in space once the portal is closed again and interacts with physical reality.

Technical Setup

To build the memory portal a sphere geometry is animated to scale up when certain words are mentioned in a conversation. The transcription of the spoken language is done via OpenAI's Whisper API. Whisper is an automatic speech recognition system that allows transcribing multiple languages in chunks of 30 seconds.[19] Whisper cannot transcribe in real-time so for the prototype setting a work-around is used that pre-records a 10 second chunk triggered by a hand pose and then sends an API request. Once the API request is answered (usually in 1-2 seconds) it is parsed and a script checks for keywords triggering the animation. This setup turned out not to be optimal as it hindered the experience to feel fully implicit.

A part of this experience was the ability to share it with a second user wearing a headset. For enabling virtual objects with consistent location, the ability to interact with them from both perspectives and synchronising the hand-tracking across applications for occlusion culling from both perspectives a multiplayer setup was build through Unity's Netcode infrastructure that synchronises data for all those aspects.

The portal itself is a sphere geometry that clips a hole into the passthrough video feed used to superimpose the field of view into the application. This happens through a stencil buffer shader setup and setting up the render queue accordingly. Behind the passthrough layer another sphere has a 360° video mapped to it which becomes visible on the mesh of the portal when it is active.

The portal has a proximity triggered handle similar to the previous prototype. This handle can be used to move the portal around in space to focus it on specific sections of the 360° video.

The portal can be grabbed directly with one hand for adjusting its position or scaled up and down with a two handed grab interaction. If the portal exceeds a diameter of 28cm a script deactivates the sphere with the stencil shader, which makes the full 360° video representing the memory visible. It also spawns another sphere that has the passthrough layer rendered on it so users are able to still view parts of physical reality, when immersed in the memory.

After some time in the memory a particle system representing butterflies is spawned, that once the portal is closed again persists in space as a continuation of the experience.

The particle system is then animated to fly towards the desk lamp prototype. Once the physical lamp is reached by the butterflies it turns on and the butterflies interact with it in terms of light and shadow.

Tacit Insights

This prototype shows how experiencing embodied software in the world together can change the meaning of virtual objects. Furthermore it emphasises on the idea of togetherness by cutting out direct interaction with the technology unless it can benefit the situation by using implicit voice inputs to spawn unobtrusive interfaces. It explores how spatial videos in combination with animated content can be used as a medium for sharing memories and how this experience can be continuated in blended space. The following points were observed while interacting with it:

Figure 37: Butterflies from the Portal viewed in Co-Location

Sharing Perspectives on Virtual Objects & the World

Physical objects and spaces are experienceable from multiple perspectives. Physical co-location of people results in some inevitable data exchange between them through gestures, facial expressions and more. This quality of physical objects is not considered central to the quality of an artefact in the first explorations. After conversations about earlier prototypes and questioning the motivation behind bridging a perceived reality gap between people with and without hardware, the insight was gained that through embodiment of software in the world virtual objects can be truly experienced together for the first time.

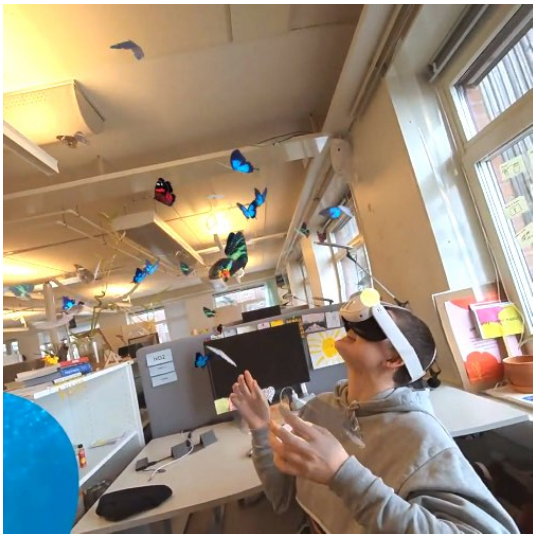


Figure 38: Second Perspective on the Scene from above

Although software can help us connect with people, whether they are nearby or far away, using it without physical spatial context isolates us from our immediate environments. Multi-player games, communication- and collaborative platforms are connecting people. When engaging with them, even if physically present in the same space, the interaction happens in a virtual abstracted space, where people cannot collaborate in true proximity to each other.

Co-located mixed reality changes this. It allows for direct experience and manipulation of virtual objects from multiple perspectives in the same physical space. Therefore it does not inhibit the data-exchange in form of gestures and facial expression between co-located users. The ability to experience virtual objects together with others enhances their meaning through perceiving the true response of others to those objects.

When building the prototype the importance of accurate virtual representation of the opposite user's hands and upper body was observed. This allows for accurate occlusion and physics interaction, which turned out important with higher immersion levels and indicates that the opposite has entered the same mixed reality.

Another observation was that the co-located multiplayer enables entering the perspectives of other users. This is of interest for user testing and can educate users about their usage of virtual objects through seeing themselves from first perspective.

Togetherness in mixed reality experiences is emphasised here as a key-stone for transitioning spatial computing from an isolating experience to a platform fostering exchange.

Figure 39: The Portal viewed non-immersive

Exploring a Spectrum of Immersion

One of the investigated user experiences with the memory portal prototype is how moving on a spectrum of immersion could feel like.

The choice of a sphere for the geometry of the portal is made because it creates the illusion of a hole that is poked into physical reality as it appears circular from every viewing angle and no shading or mapping is applied to it. This effect enables the user to grab the portal similarly to a magnifier glass and glimpse into the memory from different perspectives. Compared with the other prototypes where reality contrast was kept as low as possible to blend in, here the contrast is leveraged for sparking curiosity around the content behind the portal. This shows that high reality contrast can be used as a stylistic choice if used consciously, however generally most virtual objects should strive for low contrasts as discussed above.

Figure 40: The Portal viewed immersive

The two-hand grab interaction used to scale the portal, changes the experience to be immersive after a certain size threshold is reached. Then the portal reverses its direction to display

Figure 41: Having a Conversation about a past Memory

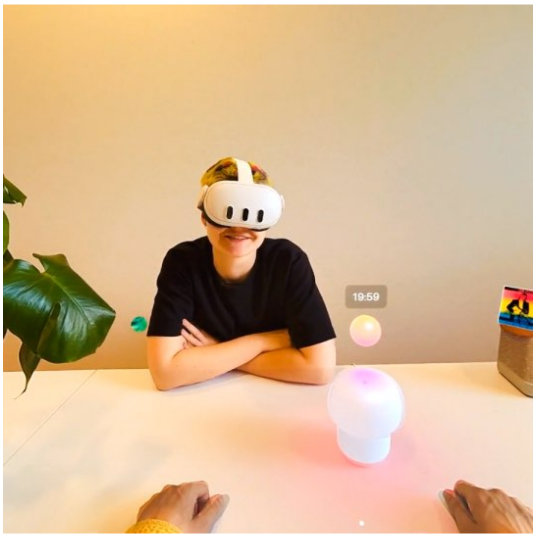


Figure 42: Parts of the Conversation are transcribed with Whisper & interpreted as Intention to share a Memory

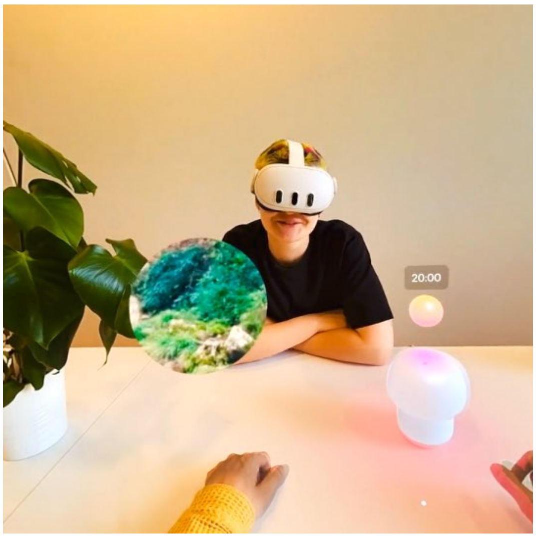


Figure 43: The Portal is augmented without any direct Interaction

physical reality instead. The abrupt switch from mixed reality into immersion was used to add a surprise momentum to the interaction, when exposed to users it turned out to create a feeling of discomfort instead though. In a further iteration a more gradual change of immersion should be prototyped instead.

When interacting with the portal together with another person the immersion change is applied to both perspectives. This creates a dynamic between the users, where one person alters another persons perceived reality and takes them into a virtual space partially disconnected from physical reality. Interactions like this should be handled with care as there is a fine line between adding value to experience and disturbing intrusion. In one case playing with the prototype a person having little experience with mixed reality was disturbed as they did not anticipate the sudden change in immersion while standing not grasping the other person was interacting with a portal.

Implicit Inputs & Unobtrusive Interfaces

In this prototype insights from the scene semantics and contextual interpretation sketches are expanded on to better understand the potential of implicit inputs spawning unobtrusive interfaces.

Compared to the first sketch where an object detection model is used to trigger interfaces the memory portal prototype emphasises on the idea of indirect voice input. Voice input is a interaction paradigm that usually requires users to give direct instructions and therefore shift focus to the interaction. This prototype tests the possibility of an experience where conversations with people are analysed and contextual interfaces are provided based on a need for information not interrupting the flow of of the conversation.

To make the interaction feel more unobtrusive, the spawned interface is positioned in the left periphery of the user triggering it. Compared to the initial sketch, where the interface element was placed in the centre of the outer periphery, placing it on the outer side periphery proved to be more effective. This is because the horizontal visual angle is wider than the vertical one.

If no direct engagement is shown the interface disappears again after a few seconds. As the periphery of one user might be the foreground of a conversation partner it would make sense to

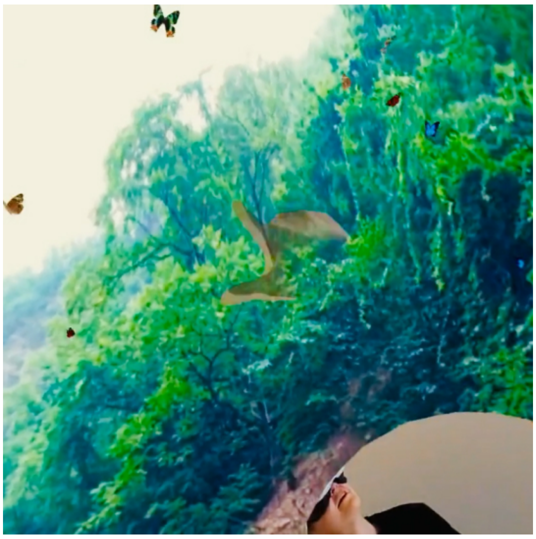


Figure 44: When the Portal is entered a Particle System spawns

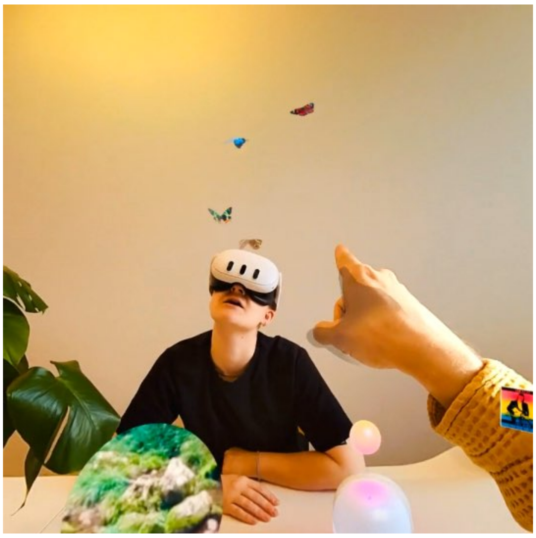


Figure 45: When the Portal is closed again the Particles remain in Space and start interacting with Physicality

introduce logic that makes the virtual object visible globally only when there is engagement with it.

Furthermore easy deflation and expansion of the experience with engagement is at the core of the design for the portal. This keeps information density down initially but gives room for more information once the object turns out to be useful.

Diving into Memories

One intention behind the memory portal prototype is to recall and share memories in more vivid ways than photos can do. 360° video was chosen for ease of technical implementation, alternatively 3D gaussian splatting or NeRFs could be used for scenes with volume rendering that could be entered fully.

To make the experience more engaging once the portal is opened a particle system spawns, which in case of the prototype resembles butterflies flying towards the users. The idea here is to make this particle system assemble dynamically based on the contents of the 360° video through an LLM, like sketched in one of the early explorations.

Those particles stay in space when the portal is minimised again and facilitate continuity of experience like adjusting lights to resemble light in a specific moment or adjusting the music to play a song that was listened to during a specific memory.

Articulated Outcome

This chapter summarises the insights from the interactive prototypes through the proposal of new interaction fundamentals for an operating system structure for spatial computing. It also revisits the process and outcome from the perspective of the brief.

Design Consolidation

Prototyping as the primary design and research tool was applied as a tool to generate knowledge insights during the process. The choice of tools for the prototyping and the ambition of investigating how the material of spatial computing can create presence in the moment lead to the creation of an eco-system of artefacts inhabiting interactions that suggest certain approaches to spatial operating systems. The interactions in this eco-system are diverse in shape and in their entireties exhibit some product-like properties. The final outcome of the thesis are not only the prototypes themselves, but the clues they provide about approaching design of an eco-system for spatial computing experiences with the following goals.

Enhanced Fluidity of Physical Objects

The prototypes exhibit how physical objects can be transformed into having properties of both physicality and virtuality. The arising artefacts are more fluid in their appearance and meaning. This continuation of virtual content into new products can bridge realities and make physical objects more accessible and interactive.

This can give physical reality some of the aspects we appreciate about virtuality.

Enhanced Purpose of Virtual Objects

The prototypes strive to be virtual objects that are purposeful and not perceived as something “not real” glued on top of reality. Enhanced meaning is achieved by careful integration of physical reality and the possibility of multiple perspective on the same virtual object.

This can give virtual reality some of the aspects we appreciate about physicality.

Coupling Realities for Facilitation of Presence

Through a coupling of realities the boundary between virtual and physical domains becomes blurry. This leads towards virtual and physical objects inhabiting one space that unites advantages of both of them.

Conclusion

The outcome of this thesis consists of the prototypes, the design consolidation and the formulation of new interaction fundamentals for spatial computing. Those fundamentals are malleable and should not be seen as absolutes but rather the raw experiential properties the technology enables at the current point of time for facilitating presence in the moment.

Technological Level

From a technological perspective spatial real-time applications for mixed reality headsets and other outlets were built through a designerly lens. Those applications illustrate user interactions and object interactions possible with available technologies.

Conceptual Level

From a conceptual perspective experiential properties were carved out of the material of spatial computing, through building, experiencing and exposing blended and virtual objects. Those properties can be considered when creating mixed reality experiences to make realities move closer together.

Interaction Fundamentals for Spatial Computing

Embracing Togetherness.

Data exchange happening between physical co-located people combined with exposure to the same virtual objects, changes the perceived meaning of these objects. Considering togetherness sparks discourse, collaboration and sharing, while preventing isolation and disconnection.

Integrating Realities.

Representation of the physical world in the generic scene model enables integration of physical and virtual realities. This integration takes shape in shared light, shadow and physics phenomena, integrated control system and world space. Considering integration leads to experiences satisfying the senses, interactions not neglecting our physical being and virtual objects carrying enhanced meaning.

Liquid Interactions.

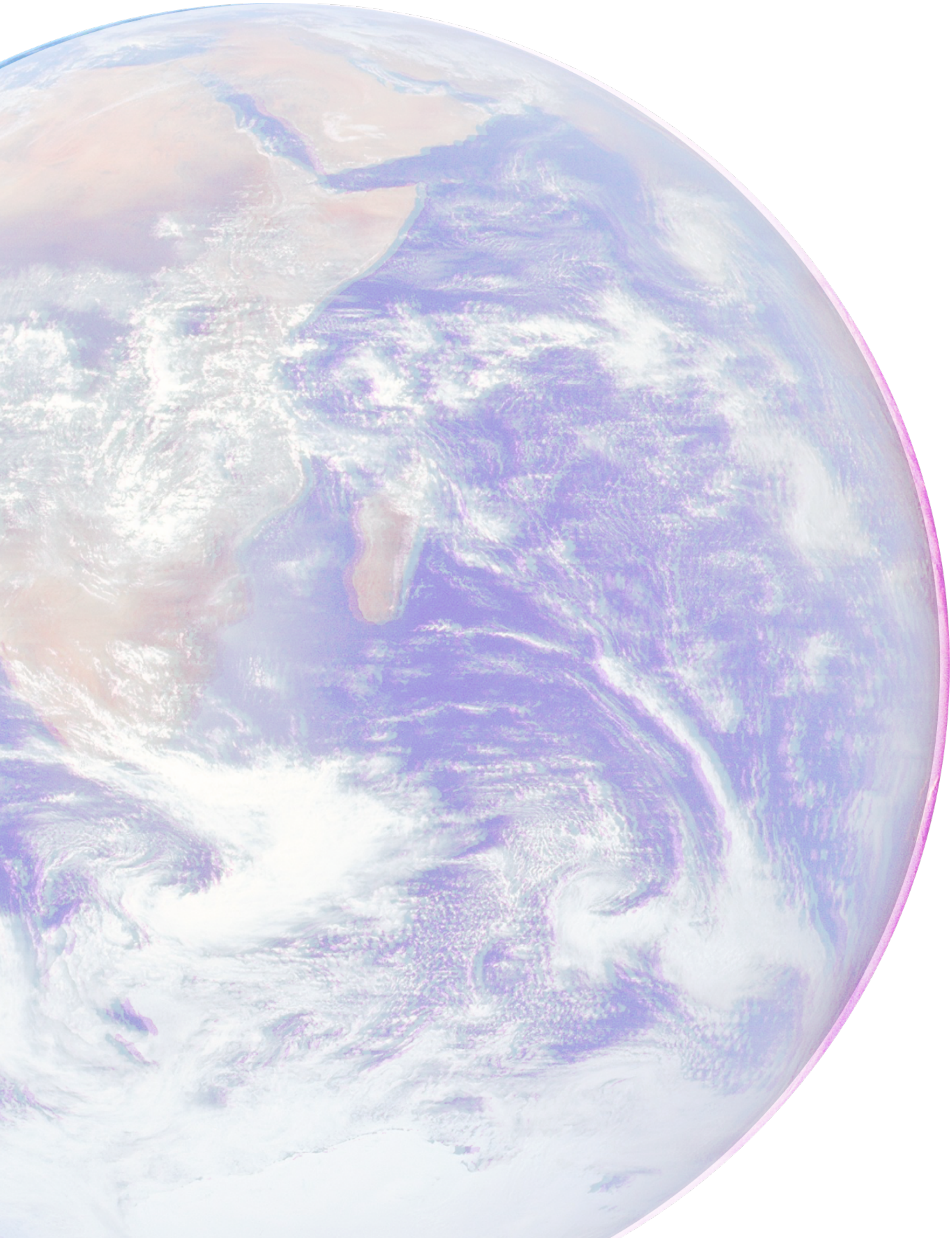
Embodied software in the world is directly coupled with our perception. Combined with the many possible input modalities of spatial computing, this enables to design virtual and blended objects with liquid interactions. Liquid interactions are forming around the intention, context, space and state of the object. Considering liquid interactions offers novel personalised user experiences and accessibility.

Opening Applications.

Spatial operating systems enable limited co-existence of virtual objects next to each other. This co-existence suggests a pivot from closed structures towards opening up and approaching software as virtual objects that exist in the same space. Considering this opening through designing virtual objects that elevate their experiential and functional value when influencing each other is a step towards closer coupling of realities.

Conscious Implicitness.

When designing interfaces triggering implicitly from context, the scenario should be understood thoroughly before. The interface itself should be expandable and positioned and shaped for minimal distraction. Considering conscious implicitness involving trigger scenarios and interfaces makes input triggered contextually less obtrusive.



Reflection

On the Process

The research-through-design approach throughout the project helped to become more precise in my formulation of the idea of spatial computing for living in the moment.

Ideas on how it could be possible to create more integrated realities turned out to have enabling technologies in place that could be dissected and used as a foundation to build on. The building process constantly confronted problems and inspired for new ideas through material-talkback. Keeping up a flexible trajectory helped in not inhibiting this process.

The prototypes helped explain ideas formulated for the brief to other people and immediately triggered insightful conversations that would not have been possible without interactive interventions.

The strong focus on technical implementation went hand in hand with increased time of acquiring background knowledge on underlying technologies. This focus was picked consciously as another approach would not have been sufficiently providing first hand insights on the experience and material.

On the Outcome

The brief was used as a starting point and guideline for the desired outcome. Through sketching and prototyping a lot of new topics of interaction design relevant for the initial idea of spatial software for presence were exposed.

The final outcome of the prototypes in combination with the design consolidation and the formulation of new interaction fundamentals for spatial computing provide a clear picture of derived insights.

Bibliography

1: Crawford, G. (2019). Developing Embodied Familiarity with Hyperphysical Phenomena. Carnegie Mellon University.

2: Stappers, P.J. (2007). Doing design as a part of doing research.

3: Moussette, C. (2012). Simple Haptics: Sketching Perspectives for the Design of Haptic Interactions. Umeå Institute of Design Research Publications

4: Redström, J. (2005). On Technology as Material in Design. In Willis, A-M. (ed.) Design Philosophy Papers: Collection Two, pp. 31-42. Team D/E/S Publications.

5: Fridman, L. (2021, August 22). Joscha Bach: Nature of Reality, Dreams, and Consciousness | Lex Fridman Podcast #212. YouTube.

6: Ishii, H., & Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97) (pp. 234–241). Association for Computing Machinery. https://doi.org/10.1145/258549.258715

7: Engelbart, D. (1962). Augmenting Human Intellect: A Conceptual Framework. Stanford Research Inst. Menlo Park CA.

8: Carter & Nielsen, "Using Artificial Intelligence to Augment Human Intelligence", Distill, 2017.

9: Avetisyan, A., Xie, C., Howard-Jenkins, H., Yang, T.-Y., Aroudj, S., Patra, S., Zhang, F., Frost, D., Holland, L., Orme, C., Engel, J. J., Miller, E., Newcombe, R. A., & Balntas, V. (2024). SceneScript: Reconstructing scenes with an autoregressive structured language model. arXiv. https://doi.org/10.48550/arXiv.2403.13064

10: Ultralytics. (2024). Ultralytics YOLO Documentation. Ultralytics. https://docs.ultralytics.com/

11: Benyon, D. (2014). Spaces of Interaction, Places for Experience. Synthesis Lectures on Human-Centered

Informatics. Morgan & Claypool Publishers. DOI: 10.2200/S00595ED1V01Y201409HCI022.

12: Meta. (2024). Depth API for Unity. Oculus Developers. https://developer.oculus.com/documentation/unity/unity-depthapi/

13: Wu, Shengzhi (2019). Affordances for Multi-device Gestural Interactions in Augmented Reality. Carnegie Mellon University. Thesis. https://doi.org/10.1184/R1/8434403.v1

14: Karam, M., & Schraefel, M. C. (2005). A taxonomy of gestures in human computer interactions. Technical report ECSTR-IAM05-009, Electronics and Computer Science, University of Southampton. https://eprints.soton.ac.uk/261149/

15: Pfeuffer, K. (2024). Eye-Hand Symbiosis. Ken Pfeuffer. https://kenpfeuffer.com/eye-hand-symbiosis-what-guide/

16: Tobii. (2024). Eye Movements. Tobii XR Devzone. https://developer.tobii.com/xr/learn/eye-behavior/eye-movements/

17: Marsden, C. (2023). The eyes are the prize: Eye-tracking technology is advertising's holy grail. Vice. https://www.vice.com/en/article/bj9ygv/the-eyes-are-the-prize-eye-tracking-technology-is-advertisings-holy-grail

18: Meta. (2024). Interaction SDK Overview. Oculus Developers. https://developer.oculus.com/documentation/unity/unity-isdk-interaction-sdk-overview/

19: OpenAI. (2024). Speech to Text. OpenAI. https://platform.openai.com/docs/guides/speech-to-text