



Politecnico  
di Torino

ScuDo  
Scuola di Dottorato ~ Doctoral School  
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation  
Doctoral Program in Control and Computer Engineering (38<sup>th</sup> cycle)

# Experiencing the Metaverse: Designing Immersive and Interactive Virtual Worlds

Alessandro Visconti

\* \* \* \* \*

## Supervisors

Prof. Fabrizio Lamberti (Supervisor)

Dr. Alberto Cannavò (Co-supervisor)

## Doctoral Examination Committee:

Prof. Lucio Tommaso De Paolis, University of Salento, Italy (Referee)

Prof. Sofia Seinfeld, Universitat Oberta de Catalunya, Spain (Referee)

Prof. Minjuan Wang, The Education University of Hong Kong, Hong Kong, China

Prof. Abdulmotaleb El Saddik, University of Ottawa, Canada

Prof. Valentina Gatteschi, Politecnico di Torino, Italy

Politecnico di Torino  
September 18, 2026

This thesis is licensed under a Creative Commons License, Attribution - Noncommercial-NoDerivative Works 4.0 International: see [www.creativecommons.org](http://www.creativecommons.org). The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

I hereby declare that, the contents and organisation of this thesis constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

.....  
Alessandro Visconti  
Turin, September 18, 2026



# Summary

The Metaverse refers to a persistent and interconnected network of digital, often virtual worlds in which users, typically embodied as avatars, can socialize, collaborate, interact, and participate in shared activities. As immersive technologies become increasingly accessible, these virtual worlds are evolving into platforms capable of hosting a wide range of social experiences. However, realizing this potential requires more than technological advancements: it demands a principled understanding of how virtual worlds should be structured and how avatar-mediated communication can be made expressive, natural, and socially meaningful. Current social virtual worlds still struggle on both fronts, often producing interactions that feel artificial or disengaging due to limitations in spatial design, tracking fidelity, and interaction paradigms.

This doctoral thesis contributes to addressing these shortcomings through two complementary lines of investigation. The first examines how the structural configuration of social virtual worlds, encompassing spatial layout, architectural choices, visual composition, and scene organization, influences user perception, engagement, and social dynamics. A virtual library application serves as a case study, illustrating how a socially oriented immersive world can be designed to support community-centered activities around reading and cultural participation. Building on this work, the thesis derives a structured set of design guidelines for immersive social virtual worlds through literature review, expert consultation, and empirical evaluation.

The second line of investigation addresses the fidelity and augmentation of social interaction in avatar-mediated environments, exploring how users perceive and respond to different technological solutions for reproducing nonverbal behavior, including facial expression conveyance through hardware tracking and software approximation, gesture-based design of social contact interactions such as handshakes, machine learning-based refinement of close-contact avatar animations, and the artificial amplification of facial mimicry to enrich emotional communication between users.

The results emerging from these investigations provide practical insights into the factors that shape the quality of social experience in immersive virtual worlds, and advance a human-centered framework for designing Metaverse worlds that are not only technically functional, but also able to support meaningful social interaction.



# Acknowledgements

I would like to sincerely express my heartfelt gratitude to all those who have supported and contributed to the completion of my doctoral thesis.

First and foremost, I would like to extend my deepest appreciation to my supervisor, Prof. Fabrizio Lamberti, for his exceptional guidance, continuous support, and invaluable insights throughout this research journey. His expertise, enthusiasm, and constant encouragement have played a fundamental role in shaping the development of this work, helping me refine my ideas and pursue my research goals with confidence. I am truly grateful for his dedication, trust, and unwavering support.

I would also like to express my sincere gratitude to my co-supervisor, Dr. Alberto Cannavò, for his valuable advice and continuous support throughout my doctoral studies. His guidance and suggestions have been greatly appreciated and have contributed to the development of this work.

I would also like to thank Dr. Davide Calandra for his support during the darkest moments and for his “acknowledgements”.

I am deeply grateful to the faculty members, administrative staff, researchers, and colleagues of the Department of Control and Computer Engineering at Politecnico di Torino for providing such a stimulating and supportive academic environment. The exchange of ideas, collaboration, and shared enthusiasm for research have greatly enriched my doctoral experience and contributed to my personal and professional growth.

I would like to express my heartfelt thanks to my family and friends for their constant affection, encouragement, and support throughout this challenging yet rewarding journey. Their trust in me, their patience, and their continuous motivation have been an essential source of strength and inspiration.

I would also like to acknowledge all the research participants who generously dedicated their time and shared their experiences and insights. Their valuable contributions have been crucial to the development of this research and have helped provide greater depth and meaning to the results presented in this thesis.

Finally, I would like to extend my gratitude to everyone who has been part of this journey and supported me along the way. Their encouragement, trust, and kindness have made this experience truly meaningful, and I am deeply grateful for all the contributions that have helped me reach this important milestone.

# Contents

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| <b>List of Tables</b>                                                                  | 8  |
| <b>List of Figures</b>                                                                 | 9  |
| <b>1 Introduction</b>                                                                  | 13 |
| 1.1 Background and motivation . . . . .                                                | 13 |
| 1.2 Thesis structure . . . . .                                                         | 17 |
| <b>2 Structural design of immersive social virtual worlds</b>                          | 19 |
| 2.1 MetaLibrary: bringing libraries into the metaverse . . . . .                       | 22 |
| 2.1.1 Background . . . . .                                                             | 23 |
| 2.1.2 Materials and methods . . . . .                                                  | 24 |
| 2.2 Finding best practices in immersive social virtual worlds design: In-              |    |
| sights from the metaLibrary platform . . . . .                                         | 27 |
| 2.2.1 Background . . . . .                                                             | 28 |
| 2.2.2 Materials and methods . . . . .                                                  | 30 |
| 2.2.3 Experiment . . . . .                                                             | 38 |
| 2.2.4 Results and discussion . . . . .                                                 | 40 |
| 2.3 Considerations and remarks . . . . .                                               | 45 |
| <b>3 Design and Augmentation of Social Interaction in Avatar-Mediated Environments</b> | 49 |
| 3.1 Comparing technologies for emotion facial expressions in social vir-               |    |
| tual worlds . . . . .                                                                  | 52 |
| 3.1.1 Background . . . . .                                                             | 53 |
| 3.1.2 Materials and methods . . . . .                                                  | 54 |
| 3.1.3 Experiment . . . . .                                                             | 56 |
| 3.1.4 Results and discussion . . . . .                                                 | 58 |
| 3.2 Controller-based non-verbal interaction design for social virtual worlds           | 60 |
| 3.2.1 Background . . . . .                                                             | 61 |
| 3.2.2 Materials and methods . . . . .                                                  | 62 |
| 3.2.3 Experiment . . . . .                                                             | 65 |

|          |                                                                                               |            |
|----------|-----------------------------------------------------------------------------------------------|------------|
| 3.2.4    | Results and discussion . . . . .                                                              | 67         |
| 3.3      | Refining close social interaction animations in social virtual worlds .                       | 70         |
| 3.3.1    | Background . . . . .                                                                          | 71         |
| 3.3.2    | Materials and methods . . . . .                                                               | 72         |
| 3.3.3    | Experiment . . . . .                                                                          | 73         |
| 3.3.4    | Results and discussion . . . . .                                                              | 79         |
| 3.4      | Augmenting virtual interpersonal communication through artificial<br>facial mimicry . . . . . | 81         |
| 3.4.1    | Background . . . . .                                                                          | 82         |
| 3.4.2    | Materials and methods . . . . .                                                               | 86         |
| 3.4.3    | Experiment . . . . .                                                                          | 91         |
| 3.4.4    | Results and discussion . . . . .                                                              | 95         |
| 3.5      | Considerations and remarks . . . . .                                                          | 103        |
| <b>4</b> | <b>Conclusions</b> . . . . .                                                                  | <b>107</b> |
| 4.1      | Summary of contributions . . . . .                                                            | 107        |
| 4.2      | Limitations . . . . .                                                                         | 110        |
| 4.3      | Future works . . . . .                                                                        | 111        |
|          | <b>Glossary</b> . . . . .                                                                     | <b>114</b> |
|          | <b>Bibliography</b> . . . . .                                                                 | <b>117</b> |

# List of Tables

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |     |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 2.1 | Best practices extracted from the analysis of the literature. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 32  |
| 2.2 | Best practices extracted from the analysis of social VR platforms and immersive applications. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         | 34  |
| 2.3 | P-values of the follow-up analysis on the PQ subscale of the AttrakDiff tool for the interaction effect between task and layout conditions. Bold values indicate statistically significant differences. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                               | 43  |
| 2.4 | P-values of the follow-up analysis on the HQ-S subscale of the AttrakDiff tool for the interaction effect between task and layout conditions. Bold values indicate statistically significant differences. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                             | 44  |
| 2.5 | P-values of the pairwise comparison on configuration rankings for T1 (overall Friedman test: $p < .001$ ). Bold values indicate statistically significant differences. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | 44  |
| 3.1 | Average percentages of AFM module usage across all participants during their speaking turns ( $\sim 10$ minutes), categorized by detected emotion. Standard deviations are reported in parentheses. Column (a) reports the proportion of speaking time during which each emotion was detected. The module operates in three modes: Augmentation (b), Camouflage (c), and AFM Off (e), where $b + c + e = 100\%$ . Column (d) reports the combined active AFM time ( $b + c$ ). Column (f) reports the proportion of time during which the peer's emotion naturally aligned with the speaker's without AFM intervention (Natural Facial Mimicry, NFM). . . . . | 101 |

# List of Figures

|     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |    |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Overview of the research context addressed in this thesis, illustrating the Metaverse as a collection of virtual worlds and highlighting the interplay between virtual world design and avatar-mediated interaction within a virtual world. . . . .                                                                                                                                                                                                                                          | 14 |
| 2.1 | First-person point of view of the user within the MetaLibrary virtual world. . . . .                                                                                                                                                                                                                                                                                                                                                                                                         | 25 |
| 2.2 | MetaLibrary virtual world: (a) overview of the world; (b) view from the central hub. . . . .                                                                                                                                                                                                                                                                                                                                                                                                 | 26 |
| 2.3 | The redesigned virtual world of MetaLibrary: (a) overview, (b) high ceilings and interconnected rooms, (c) furniture, (d) central hub, (e) room of the day, (f) most-read room, (g) conference room, and (h) social activities area. . . . .                                                                                                                                                                                                                                                 | 36 |
| 2.4 | Overview of the environment configurations and tasks included in the user study. . . . .                                                                                                                                                                                                                                                                                                                                                                                                     | 41 |
| 2.5 | Objective results: average fixation time percentages across tasks and environment configurations. . . . .                                                                                                                                                                                                                                                                                                                                                                                    | 42 |
| 2.6 | Subjective results: average AttrakDiff scores across environment configurations and tasks [84]. . . . .                                                                                                                                                                                                                                                                                                                                                                                      | 43 |
| 3.1 | First-person view of the user's avatar. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                              | 56 |
| 3.2 | (a) Actor performing and recording the happiness scene using the VR setup. (b) First-person view of the recorded emotional scene inside the virtual environment. . . . .                                                                                                                                                                                                                                                                                                                     | 57 |
| 3.3 | Subjective results for each scene and emotion (a–f). Statistically significant differences ( $p < .05$ ) are marked with brackets; error bars represent standard deviations. Questions #1 to #8 asked participants to indicate which of the two modalities: #1 made them feel more comfortable; #2 was more expressive; #3 was more realistic; #4 was more natural; #5 was more plausible; #6 was more pleasant; #7 better conveyed the emotional content; #8 was preferred overall. . . . . | 58 |

|      |                                                                                                                                                                                                                                                                                                                                                                                                                             |    |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.4  | The four handshake interaction variants, ordered by increasing gestural fidelity. (a) HS1: GUI-triggered. (b) HS2: controller-button-triggered. (c) HS3: grab-based, with predefined completion animation. (d) HS4: grab-based, with free shaking motion and no predefined animation. . . . .                                                                                                                               | 63 |
| 3.5  | Usability measures across the four handshake variants based on the VR-USE questionnaire. Brackets indicate statistically significant differences ( $p < .05$ ). . . . .                                                                                                                                                                                                                                                     | 68 |
| 3.6  | Scores from the ad-hoc handshake experience questionnaire across the four variants. Brackets indicate statistically significant differences ( $p < .05$ ). . . . .                                                                                                                                                                                                                                                          | 69 |
| 3.7  | Overview of the proposed refinement framework. User 1 and User 2 are engaged in a social interaction, while User 3 acts as an external observer. Each user runs a local neural network module that processes avatar pose data and refines the animation output of the base IK system. . . . .                                                                                                                               | 74 |
| 3.8  | The two hug types considered in the study: (a) criss-cross, and (b) neck-waist. . . . .                                                                                                                                                                                                                                                                                                                                     | 76 |
| 3.9  | The two animation conditions across the four types of hugs performed in a virtual environment. . . . .                                                                                                                                                                                                                                                                                                                      | 77 |
| 3.10 | Participant positions and viewpoints considered in the user study: (a) position of the participant as an external observer (red marker), (b) the corresponding external observer viewpoint, and (c) the front-facing viewpoint of one of the users engaged in the hug. . . . .                                                                                                                                              | 78 |
| 3.11 | Distribution of scores for the base IK system (B-IK) and the proposed refinement framework (PRF) across custom questionnaire items. Statistically significant results are marked with an asterisk. Q1: no overlaps between avatars; Q2: hug beginning and end well coordinated; Q3: visually harmonious hug; Q4: well-aligned postures and movements; Q5: arm and hand movements adapted to the other body's shape. . . . . | 80 |
| 3.12 | Distribution of scores for the base IK system (B-IK) and the proposed refinement framework (PRF) across items of the Godspeed anthropomorphism subscale. Statistically significant results are marked with an asterisk. Q1: Fake / Natural; Q2: Machinelike / Humanlike; Q3: Unconscious / Conscious; Q4: Artificial / Lifelike; Q5: Moving rigidly / Moving elegantly. . . . .                                             | 80 |



|      |                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |     |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.13 | AFM methods in immersive social VR. User 1 (blue, left) observes User 2 (orange, right) reaching the same emotional state under four different conditions: (a) the user’s own expression is projected onto the peer; (b) a lower-intensity emotion on the peer is amplified to match the user’s level; (c) a conflicting emotion on the peer is replaced with a matching one; (d) the peer’s expression is remapped into a form more easily understood by the user. . . . . | 83  |
| 3.14 | Architecture of the AM system, illustrated assuming User 1 is speaking. Speech input detected on the microphone activates AFM, making its effects visible on User 2’s avatar. The same logic applies symmetrically for User 2; AFM is disabled on the listening user’s end.                                                                                                                                                                                                 | 87  |
| 3.15 | Operational logic of the AFM Management module for User 1 (speaker) and User 2 (listener) across three emotional states: happiness, sadness, and neutrality. User 1’s blendshapes are directly tracked via facial tracking; User 2’s blendshapes are synchronized over the network. Labels A to G correspond to the states described in Section 3.4.2.                                                                                                                      | 90  |
| 3.16 | Overview of the experimental procedure phases. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                      | 94  |
| 3.17 | Distribution of IOS scores assessing interpersonal closeness. The bracket indicates a statistically significant difference between groups.                                                                                                                                                                                                                                                                                                                                  | 96  |
| 3.18 | Distribution of NMMS scores (#1 Co-presence, #2 Perceived attentional engagement, #3 Perceived emotional contagion, #4 Perceived comprehension, #5 Perceived behavioral interdependence). Brackets indicate statistically significant differences. . . . .                                                                                                                                                                                                                  | 97  |
| 3.19 | Distribution of social attraction scores (#1 I like my partner; #2 I would get along well with my partner; #3 I would enjoy a casual conversation with my partner; #4 My partner is friendly). Brackets indicate statistically significant differences. . . . .                                                                                                                                                                                                             | 98  |
| 3.20 | Distribution of Godspeed anthropomorphism scores (#1 Fake / Natural; #2 Machinelike / Humanlike; #3 Unconscious / Conscious; #4 Artificial / Lifelike; #5 Moving rigidly / Moving elegantly). Brackets indicate statistically significant differences. . . . .                                                                                                                                                                                                              | 99  |
| 3.21 | Distribution of scores for custom interpersonal trust items (#1 The other person’s reactions made me feel comfortable; #2 When I was happy I felt understood; #3 When I was sad I felt understood; #4 I think the other person deserves my trust; #5 I think the other person was empathic with me). Brackets indicate statistically significant differences. . . . .                                                                                                       | 100 |



# Chapter 1

## Introduction

### 1.1 Background and motivation

The Metaverse can be described as a persistent network of interconnected virtual worlds in which people can interact, collaborate, and share experiences through embodied avatar representations [1]. Within these worlds, often leveraging immersive technologies, users are not just passive observers, but active participants who inhabit spaces, communicate with others, and engage in collective activities [2]. As enabling technologies such as [Head-Mounted Displays \(HMDs\)](#) for [Virtual Reality \(VR\)](#) experiences become increasingly accessible, virtual worlds are becoming platforms capable of hosting a wide range of social, collaborative experiences.

At this stage of evolution, one of the key challenges for the Metaverse is understanding how users actually experience and inhabit virtual worlds, and which design solutions can be adopted to improve such experiences [3], [4]. Beyond technological capabilities, the success of these virtual worlds depends on whether they can support interactions that users perceive as engaging, coherent, and socially meaningful [5]. This shifts the focus to the user at the center of the design process, highlighting the need to identify the conditions and design choices that effectively enhance immersive experiences.

From this standpoint, two complementary aspects become particularly relevant (Figure 1.1). The first concerns the design of virtual worlds themselves: understanding how spatial layout, architectural structure, and environmental features influence user engagement, navigation, and interaction with activities within the virtual world [6], [7]. The second aspect concerns how users communicate and interact within these environments. Because social interaction is mediated by avatars, it is important to examine how communicative cues, such as facial expressions and gestures, are conveyed, how close-range interactions unfold under these constraints, and which technological and interaction design solutions can improve or even augment the expressiveness, clarity, and naturalness of avatar-mediated communication [8], [9].

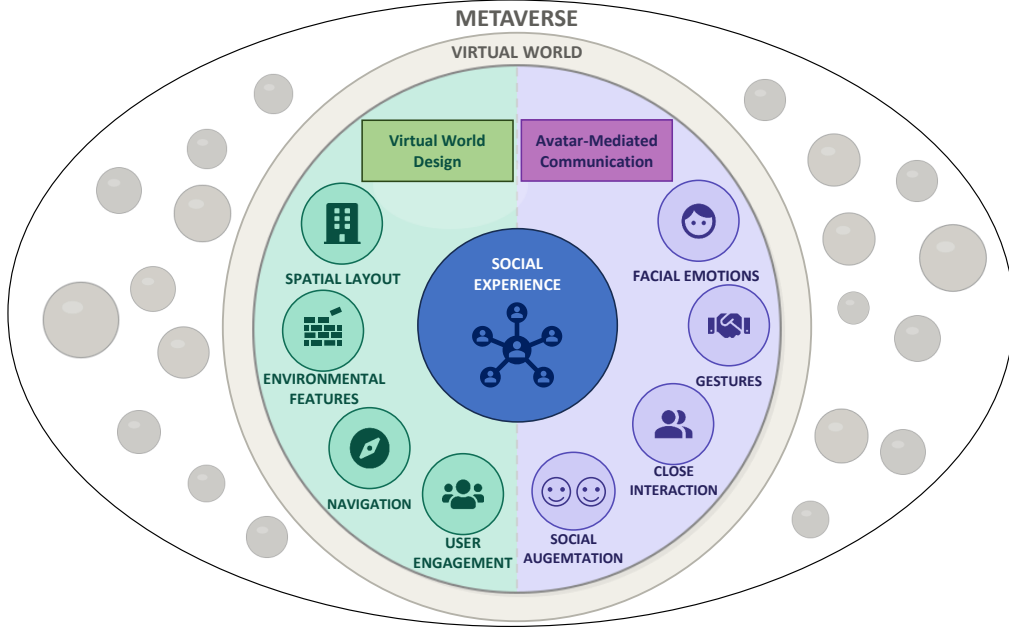


Figure 1.1: Overview of the research context addressed in this thesis, illustrating the Metaverse as a collection of virtual worlds and highlighting the interplay between virtual world design and avatar-mediated interaction within a virtual world.

Focusing first on the design of virtual worlds, the spatial structure of the environment plays a fundamental role in shaping how users experience and interact within it. Elements such as spatial layout, architectural organization, environmental scale, navigational structure and spatial lighting influence how users explore virtual environments, how they move through space, and how they encounter other participants [10], [11], [12].

In immersive virtual worlds, spatial architecture does not only define the visual appearance of the environment but also directly affects the user experience [12]. The arrangement of spaces, the presence of open or enclosed areas, and the organization of pathways and meeting points can facilitate or hinder the emergence of spontaneous encounters between users [13]. Environments that are visually rich but spatially confusing may reduce orientation, limit exploration, and ultimately decrease opportunities for meaningful interaction [14]. Conversely, carefully structured environments can guide user movement, support discovery, and create conditions that naturally encourage users to gather, interact, and participate in shared activities [11].

For this reason, the design of socially oriented virtual worlds requires particular attention to spatial configuration and environmental structure [13]. Understanding how architectural design choices influence user behavior and interaction dynamics

represents an important step toward creating immersive environments that effectively support social experiences [15].

At the same time, social interaction in virtual worlds relies heavily on avatar-mediated communication. Avatars act as the primary interface through which users express identity, emotions, and intentions while interacting with others [16], [17]. In face-to-face communication, a substantial portion of interpersonal meaning is conveyed through non-verbal signals such as gestures, gaze behavior, facial expressions, posture, and interpersonal distance [18], [19]. These cues play a critical role in regulating conversational flow, expressing emotions, and establishing social presence. In immersive virtual environments, however, many of these signals must be reconstructed through technological mediation [20], [21]. As a result, the fidelity with which avatars can reproduce non-verbal behavior becomes a key factor in determining how natural and believable social interactions appear to users [22], [23].

Despite recent progress in immersive technologies, reproducing credible social behavior in virtual worlds remains a complex challenge [18], [24]. Current consumer VR systems typically track only a limited set of body movements, such as head and hand motion, leaving many aspects of human behavior only partially represented. Limitations in tracking technologies, animation systems, and interaction techniques can therefore produce discrepancies between the visual realism of avatars and the behaviors they are able to exhibit [25], [26]. When the visual appearance of an avatar suggests a high level of realism that is not matched by its behavioral expressiveness, interactions may appear artificial or uncomfortable, a phenomenon often associated with breaks in social presence and behavioral plausibility [27], [28]. Improving the fidelity and expressiveness of avatars is essential for fostering trust, empathy, and engagement in immersive social environments [29].

Many aspects of everyday social interaction that occur naturally in physical environments are surprisingly difficult to reproduce in virtual worlds. Gestures such as greeting another person through a handshake [30], [31], expressing emotions through facial expressions [32], [33], or subconsciously mimicking the emotional expressions of others play an important role in interpersonal communication [34], [35]. These behaviors contribute to establishing rapport, conveying intentions, and reinforcing social bonds. Understanding how such non-verbal behaviors can be effectively represented through avatars, and how technological solutions can enhance their fidelity, represents an important step toward more natural and engaging social interaction in immersive systems.

These challenges require addressing both the design of the virtual world itself and the mechanisms that enable expressive avatar behavior. On one hand, it is necessary to explore how virtual worlds can be designed to support socially oriented activities and meaningful shared experiences [9], [12], [36]. On the other hand, it is essential to investigate how interaction techniques, animation systems, and emerging technologies can improve the realism and expressiveness of avatar-mediated

communication [37], [38].

This doctoral thesis investigates these aspects through a series of studies that examine how socially oriented virtual worlds can be designed and how avatar-mediated interactions can be enhanced to support richer social experiences. The research combines environment design, interaction design, and technological experimentation in order to explore different dimensions of social interaction in immersive virtual worlds.

A first part of investigation focuses on the design of virtual worlds intended to support cultural and community-oriented activities. Within this context, the thesis introduces the concept of MetaLibrary, a social virtual world designed to foster shared experiences related to reading and knowledge exploration [39]. Through the design and evaluation of this environment, the research examines how spatial organization, environmental storytelling, and interaction affordances can facilitate engagement, discovery, and community participation within immersive spaces [40].

A second line of investigation explores the fidelity of avatar-mediated communication in social virtual worlds. Several studies have been reported, whose aim was to examine how different technological approaches influence the perception of emotions, realism, and naturalness in interactions between users. This includes the comparison of techniques for conveying emotions through realistic avatars [41], the design of interaction methods that allow users to perform greeting gestures such as handshakes using handheld controllers [30], and the development of a machine learning based framework aimed at improving the fidelity of close social interaction animations between avatars [42]. In addition, the thesis investigates mechanisms for enhancing emotional communication through [Artificial Facial Mimicry \(AFM\)](#). By enabling avatars to reproduce subtle facial responses during interaction, this approach explores how automatic behavioral cues can contribute to more expressive and engaging interpersonal communication in immersive environments [34].

The research presented in this thesis contributes to a broader understanding of how immersive technologies can support social user experience in virtual worlds. The findings emerging from the different studies highlight several aspects that are particularly relevant for the design of socially oriented immersive systems.

From the perspective of virtual world design, the research suggests that the spatial architecture of immersive worlds plays a significant role in shaping user experience. Elements such as spatial organization, environmental richness, and scene configuration influence how users explore virtual spaces, orient themselves within the environment, and engage with shared activities. These aspects are particularly relevant when designing immersive environments intended to support cultural and community-oriented experiences.

From the perspective of avatar-mediated communication, the results indicate that the fidelity and expressiveness of non-verbal behavior are key factors for effective social interaction in immersive virtual worlds. Different technological solutions

for conveying emotions may perform differently depending on the specific emotional expression involved. In addition, interaction techniques that allow users to perform gestures that closely resemble real-world social behaviors tend to be perceived as more natural and engaging. The findings also suggest that computational approaches can play an important role in improving the realism of avatar interactions: machine learning techniques can enhance the fidelity of complex social animations, while mechanisms based on augmented behavioral responses, such as [AFM](#), can enrich interpersonal communication and positively influence perceived social presence and interpersonal connection.

Finally, all these insights highlight the importance of addressing both environmental and behavioral aspects when designing immersive systems for social interaction. The spatial configuration of virtual environments, the fidelity of avatar behavior, and the integration of computational techniques for enhancing non-verbal communication all contribute to shaping the quality of social experiences within immersive virtual worlds.

## 1.2 Thesis structure

The thesis is structured around the two main themes introduced above: the role of virtual world design in shaping user experience and the mechanisms that support effective avatar-mediated social interaction in immersive virtual worlds.

The present chapter provides a general introduction to the research context and motivations of the thesis. It presents the evolution of immersive technologies and the emergence of the Metaverse paradigm, highlighting the growing importance of virtual worlds as environments for social interaction, collaboration, and shared experiences. The chapter also introduces the main research directions addressed.

Chapter [2](#) focuses on the design of immersive virtual worlds and its influence on user experience. In particular, the chapter investigates how aspects related to spatial architecture, environmental organization, and scene configuration can affect how users explore virtual worlds' environments and engage in shared activities. Within this context, MetaLibrary is presented as a case study of a socially oriented immersive environment designed to support cultural experiences centered on reading, knowledge exploration, and community interaction.

Chapter [3](#) addresses the topic of social interaction in immersive virtual worlds, with particular attention to avatar-mediated communication and non-verbal behavior. This chapter presents a series of studies investigating different aspects of interpersonal interaction in immersive environments, including techniques for conveying emotions through realistic avatars, interaction methods for performing social gestures such as handshakes using handheld controllers, approaches for improving the behavioral fidelity of close avatar-to-avatar interactions, and mechanisms for

enhancing communication through artificial facial mimicry. Together, these investigations explore how different technological solutions influence the realism, expressiveness, and perceived naturalness of social interactions in immersive virtual worlds.

This thesis, as a whole, progresses from the design of immersive environments to the study of avatar-mediated interaction mechanisms, providing a comprehensive perspective on how immersive technologies can support richer and more natural social experiences in virtual worlds.



## Chapter 2

# Structural design of immersive social virtual worlds

*The work described in this chapter has been formerly published in [39], [40].*

The design of virtual worlds strongly shapes how users experience and interact within immersive social environments. Unlike physical spaces, virtual worlds are not limited by material or structural limitations. This offers a high degree of freedom in spatial organization: the layout, the density of visual elements, lighting, and overall aesthetic coherence, all contribute to whether users feel comfortable, oriented, and engaged [6], [43]. However, it also introduces challenges. In the absence of principled design guidelines, virtual spaces risk becoming disorienting, visually inconsistent, or ineffective in supporting meaningful social interaction [6], [9], [40]. Many design principles applied to virtual environments are informed by studies conducted in physical settings. This is especially evident in work on spatial cognition and wayfinding, where authors have argued that studying physical environments can make virtual environments easier to understand and navigate, and that knowledge from wayfinding in the physical world can be applied to virtual worlds [44]. For instance, in real environments, lighting conditions have been shown to influence the depth of interpersonal self-disclosure and the impressions formed during social exchanges [45]; ceiling height affects how abstractly or concretely people process information, with documented downstream effects on cooperative behavior [46]; and environment brightness has been linked to ethical and affiliative conduct in social tasks [47]. Whether and to what extent these findings transfer to immersive virtual environments is a question that has only recently begun to receive systematic empirical attention [46], [48]. The answer is not straightforward. A virtual world introduces additional layers of mediation, such as [HMDs](#), avatar-based

navigation, and altered depth perception. These factors can modulate, attenuate, or in some cases even reverse effects that have been well documented in real-world settings [49]. Recent findings suggest that this transfer is real but partial and task-dependent. Spatial dimensions appear particularly influential for social attention and affective responses, lighting quality shows measurable effects on visual comfort and mood, but the magnitude of these effects can differ from the physical baseline in ways that are not yet fully understood [46], [48]. This gap between knowledge from physical environments and the still limited evidence in immersive social VR is one of the field’s main open challenges.

Research in environmental psychology has long established that the physical characteristics of a space influence human behavior, emotional states, and social dynamics [6]; these effects do not disappear in virtual contexts but rather take on new forms, shaped by the interplay between immersive technology and the perceptual and cognitive responses of users [46]. Empirical studies have confirmed that users respond to virtual environments in ways broadly consistent with their responses to analogous physical spaces [43], [46], [49], suggesting that the principles of environmental psychology retain explanatory power in VR settings. At the same time, virtual environments introduce design variables with no direct physical counterpart, such as the ability to transition instantaneously between spatially disconnected areas, to render physically impossible architectural configurations, or to dynamically alter the environment in response to user behaviors that require dedicated frameworks going beyond those developed for physical spaces [6], [50], [51]. In immersive virtual worlds, architectural and aesthetic choices can influence the sense of presence, the ease of navigation, the direction of attention, and the willingness to engage with other users, making structural design a critical determinant of the overall quality of the social experience [6], [43]. Despite this relevance, the question of how to structure immersive virtual worlds so that they genuinely support social engagement and cultural participation is still largely open in the current literature [12], [52] and empirically grounded design guidelines are still scarce [40].

In recent years, domains such as culture and education have increasingly explored new ways of connecting people through social VR. In particular, the emergence of social virtual worlds has opened new opportunities to collective experiences, enabling users to meet, interact, and participate in activities regardless of their physical location.

In this context, libraries are a particularly relevant application domain [53]. As institutions dedicated to the preservation and dissemination of culture, they can benefit from virtual worlds not only by extending access to their collections, but by reimagining themselves as interactive and socially engaging environments. Existing efforts have primarily focused on technological aspects, such as AI-based recommendation systems [54] and the digitization and online accessibility of archival materials [55]. However, relatively little attention has been given to how these virtual worlds are actually designed and experienced, particularly in ways that

reflect the specific needs and practices of cultural institutions.

This gap points to a broader issue: the effectiveness of social virtual worlds, particularly in cultural contexts such as virtual libraries, depends not only on their capacity to connect users or provide access to content, but also on how the environment itself is structured and perceived [7]. The spatial configuration, visual organization, and overall design of the environment play a crucial role in determining whether users can meaningfully explore content, orient themselves within the space, and engage in social activities conducted within a virtual setting intended as a cultural and social venue. Despite the growing interest in social virtual worlds and the broader vision of the Metaverse [56], the question of how to design these environments to effectively support user experience in domain-specific contexts, such as culture and education, remains largely underexplored.

This chapter addresses the open challenge outlined above through two complementary contributions. The first introduces MetaLibrary, an immersive social VR platform developed to bring the experience of public libraries into the Metaverse, allowing users to explore a shared virtual world, connect with other readers, attend author-led events, and receive personalized book recommendations [39]. The platform was designed to support three interrelated modes of engagement: individual browsing, simulated social interaction through avatar-mediated communication, and structured cultural events addressed to a live virtual audience. Each mode places different demands on the environment in terms of spatial layout, lighting, visual signposting, and avatar proximity management, making MetaLibrary a particularly relevant testbed for investigating how real-world design principles transfer to, and must be adapted for, a social virtual setting. Developed within the framework of the Reading (&) Machine project funded by Fondazione TIM<sup>1</sup>, the platform was deployed in a real institutional context in Turin, Italy, allowing the system to be evaluated in a real institutional context. From this experience, the second contribution presents a systematic design study aimed at deriving and partially validating a structured set of best practices for the development of immersive social virtual worlds in the Metaverse [40]. Drawing on expert interviews, a review of the relevant literature, an analysis of existing social VR platforms, and a dedicated user study, the study identifies design principles spanning spatial organization, visual style, lighting, layout density, and the use of biophilic elements, providing a foundation for designers and researchers working on next-generation immersive social worlds [12], [57].

---

<sup>1</sup>Reading (&) Machine project: <https://tinyurl.com/2pcfu82k>

## 2.1 MetaLibrary: bringing libraries into the metaverse

In recent years, domains such as culture and education have undergone transformations driven by the growing integration of virtual technologies into their core processes. A key factor behind these changes is the widespread adoption of tools designed to connect people regardless of their physical location, thereby broadening the accessibility and circulation of content of all kinds. Libraries represent one of the contexts that stand to benefit most from this digital shift [53], as embracing new technologies and tools could help them strengthen their role as institutions dedicated to the preservation and dissemination of culture.

Among the technologies considered most relevant for the modernization of libraries, AI and immersive media and technologies such as VR are widely regarded as particularly promising. While AI applications for book recommendation are already well established [54], the potential of immersive virtual technologies in this domain remains largely unexplored. At the time the work was developed, it was found that VR has so far been employed primarily as an alternative interface for exploring the contents of individual books, rather than as a means of enhancing the library experience more broadly.

As a medium, VR supports the creation of complex and realistic simulations [58], which can be extended with additional functionalities such as guidance systems or assessment modules, as demonstrated in training contexts [59]. A notable characteristic of the virtual technology is its flexibility in terms of access modalities: virtual scenarios can be experienced either through a HMD combined with hand controllers, or through conventional input devices such as a mouse and keyboard, making it possible to accommodate users with different levels of familiarity and different hardware setups. VR has consequently been adopted as a foundation for social platforms in which multiple users can connect from geographically dispersed locations and share a common virtual world where they can interact and engage in shared and collaborative activities. Such applications are closely associated with the concept of the Metaverse [56].

Building on these premises, the Reading (&) Machine project investigated how Metaverse technologies can be applied to enrich the library experience and attract a broader audience. Project activities encompassed the design and implementation of MetaLibrary, an immersive VR platform conceived to bring the library into the Metaverse. The platform allows users to access a shared virtual world where they can meet other readers, attend author presentations, and interact with a book recommendation system that suggests titles based on their reading history or current preferences, elicited through an engaging and playful interface.

### 2.1.1 Background

The ongoing advancement of digital and virtual technologies is opening new opportunities to reshape how users engage cultural institutions, including libraries. Studies examining the digital transformation of library spaces suggest that the adoption of new technologies can meaningfully improve the reader experience while reinforcing the social and intellectual role of these institutions. Research focused on the Oodi Library in Helsinki, for instance, concluded that digital transformation can strengthen libraries as centers of intellectual life while enabling more personalized and responsive services [60]. A subsequent analysis of criteria for integrating digital technologies into library design further confirmed that technologically enriched library environments tend to be strongly user-oriented and can contribute positively to urban cultural life [53].

Despite this growing body of evidence, concrete implementations of digitalized library environments remain scarce in the literature. Among the most notable examples is again the Oodi Library, where an AI-powered mobile application was developed to provide users with personalized book recommendations through a virtual assistant [54]. The system proved effective in making the recommendation process faster and more engaging compared to traditional approaches. However, no prior examples of library experiences enhanced through immersive VR technologies have been identified in the literature.

The potential of immersive virtual technologies in related domains nevertheless provides a relevant foundation for this research. VR has been shown to support the development of social platforms with considerable educational value [61], a domain closely related to the cultural mission of libraries. In educational settings, immersive VR has been used as a practical complement to conventional instruction, demonstrating measurable improvements in learners' preparedness [58], and VR-based virtual classrooms have been shown to introduce new and engaging learning modalities for both educators and students [62].

The quality of the social experience within these environments has also been linked to the richness of avatar representation. Research suggests that shared VR experiences featuring expressive and detailed avatars, including animated body movements and facial expressions, tend to support higher levels of collaboration and communication among participants [41], [63]. Beyond educational contexts, immersive technologies have been successfully employed to create social VR experiences in which users meet, exchange opinions, and engage in shared activities [64]. Relevant examples include virtual conference systems that replicate the experience of attending oral presentations and poster sessions [65], as well as hybrid applications combining VR and mobile technologies to enable asymmetric collaborative and competitive interactions [66].

From an implementation perspective, existing social VR platforms such as VR-Chat have been adopted in a number of studies in the fields of education and cultural

heritage [67], [68]. However, these platforms are subject to technical constraints that often limit their adaptability to specific use cases, making the development of custom dedicated solutions necessary [67], [69].

### 2.1.2 Materials and methods

The social VR platform was developed to enable readers to access a shared virtual world where they can meet other users, discover new books, and engage with a community of fellow readers (Figure 2.1). The platform integrates a book recommendation system capable of suggesting books either on the basis of a user's borrowing history or through preferences expressed directly within the virtual world. The underlying data are sourced from a real physical library, where users can subsequently locate and borrow the recommended books. MetaLibrary was developed primarily for the Meta Quest 2 headset<sup>2</sup>, exploiting its integrated handheld controllers for interaction within the virtual world. To broaden accessibility, support for mouse and keyboard input was also included, allowing users without a VR headset to participate in the experience. The social VR platform was built in Unity Engine, with VR functionalities implemented through the XR Toolkit framework<sup>3</sup> and the OpenXR library<sup>4</sup>.

#### Multi-user architecture

MetaLibrary was conceived as a multi-user social platform in which several participants can share the same virtual world simultaneously. The underlying networking architecture follows a client-server model, implemented using the Mirror networking library<sup>5</sup> in Unity. A centralized server handles all incoming connections, while individual clients run independently on separate personal devices, either as immersive VR sessions or as desktop applications.

During the login phase, users select their preferred access modality (Immersive VR or Mouse and Keyboard) and customize their appearance by choosing an avatar from a set of predefined characters. Users can then join the platform either anonymously or by authenticating with their library credentials, taking on the role of *Readers*. A dedicated *Librarian* role is also available for authorized staff, granting administrative capabilities such as server moderation, user management, and the assignment of presenter privileges for the Conference room. Regardless of role or access modality, all users can communicate via Voice Over Internet Protocol

---

<sup>2</sup>Meta Quest 2: <https://www.meta.com/it/quest/products/quest-2/>

<sup>3</sup>XR Toolkit framework: <https://tinyurl.com/46p2b8kv>

<sup>4</sup>Open XR Library: <https://www.khronos.org/openxr/>

<sup>5</sup>Mirror Networking for Unity: <http://tiny.cc/jndmzz>



Figure 2.1: First-person point of view of the user within the MetaLibrary virtual world.

(VOIP), adjust their audio settings, and choose between continuous movement and teleportation as their preferred locomotion method.

## Virtual world

The virtual world was designed to evoke the atmosphere of a physical library with surreal and playful features inspired by established social VR platforms (Figure 2.2). A low-poly visual style was deliberately chosen to reduce the computational load of the application, ensuring compatibility with a wider range of hardware configurations. Upon joining, users are placed in a central hub that provides an overview of the virtual world and serves as a navigation point for all available areas. Each room within the platform fulfills a distinct function, as outlined below.

- *Room of the day*: this room displays a curated selection of five books updated daily, chosen according to a topic set by the administrator. The topic, extracted from the library database, acts as a thematic filter, and the room’s visual appearance adapts accordingly, with walls decorated with quotes and imagery associated with relevant authors.
- *Most-read room*: this room presents the most frequently borrowed books, ranked according to lending records retrieved from the library database.
- *Conference room*: this area hosts virtual events organized by the administrator. A designated presenter can share a slideshow uploaded as a PDF file and address the audience. The room is laid out similarly to a lecture hall or theater, with a central stage for the presenter and tiered viewing positions for attendees.
- *Community rooms*: users can create private community spaces and invite selected participants, supporting smaller and more focused social interactions.



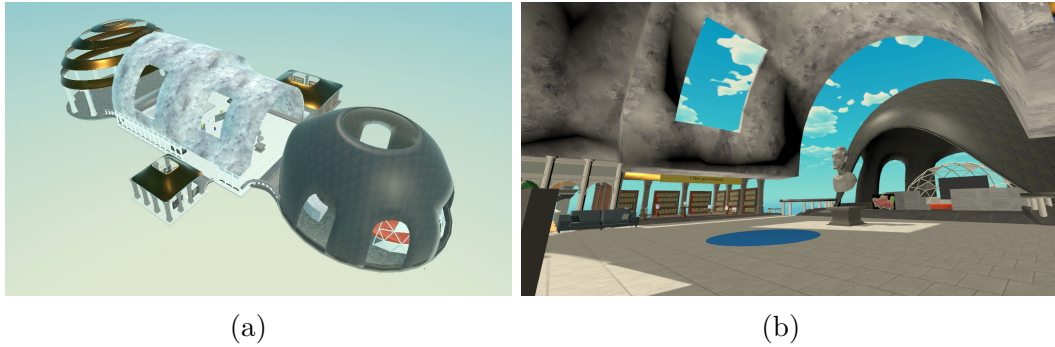


Figure 2.2: MetaLibrary virtual world: (a) overview of the world; (b) view from the central hub.

- *Assisted research rooms:* these private rooms provide access to the personalized recommendation system, which generates book suggestions based on the user's borrowing history and expressed preferences.

### Assisted Research

The assisted research feature allows users to receive a personalized list of up to 30 book recommendations by progressively refining their preferences through a three-step interaction procedure.

- *Step 1 – Genre selection:* users choose one or more genres from a set of 20 options derived from the library database, each represented by a 3D model placed within the scene. The recommendation list updates dynamically as genres are selected, and combinations of genres yield results spanning multiple categories.
- *Step 2 – Keyword filtering:* users can further narrow the recommendations by selecting keywords associated with the genres chosen in the previous step, allowing for a more targeted search.
- *Step 3 – Emotional filtering:* in the final step, users can filter recommendations by emotional tone, choosing from four categories (sadness, fear, anger, and happiness). These emotional labels were automatically assigned to each book by an [AI](#) system analyzing the corresponding descriptions in the library database.

At any point during the procedure, users may return to a previous step to revise their choices. Doing so automatically clears any selections made in subsequent steps, ensuring consistency in the recommendation output.



### Book recommendations

Books surfaced through the Room of the day, the Most-read room, or the Assisted research system are introduced as structured metadata cards retrieved from the library database and enriched with information from the Anobii platform<sup>6</sup>, an independent social network for readers. For each title, the following information is displayed: Cover image, Title, Author(s), Genre(s), Number of pages, Description, Average user rating.

## 2.2 Finding best practices in immersive social virtual worlds design: Insights from the metaLibrary platform

The previous section introduced MetaLibrary and showed, through its deployment in a real institutional context, that a socially oriented virtual world can support cultural and community activities beyond the controlled conditions of a laboratory. The design choices that made this possible, however, were taken case by case, driven by the requirements of a single platform rather than by a systematic understanding of why particular spatial, visual, and lighting decisions succeed. Translating that situated experience into principles that generalize to immersive social virtual worlds more broadly is the concern of the present section. As immersive social virtual worlds have grown in sophistication and adoption, the question of how to design their environments effectively has become increasingly central to the field. Research has demonstrated that architectural parameters such as ceiling height, spatial extent, lighting conditions, color, and visual complexity, significantly influence users' affective states, cognitive responses, and behavioral patterns within virtual environments [46], [70], [71]. These effects are not merely aesthetic, they condition how comfortably and effectively users can navigate a space, how much social attention they direct toward others, and how positively they appraise the overall experience.

This growing awareness of the role of virtual world design has led the research community to recognize the need for structured and actionable design guidelines for immersive virtual worlds. This need, despite the accumulating empirical evidence, remains unmet in the current literature [72]. The reasons for this gap are multiple. On one hand, the field has evolved rapidly, with hardware capabilities, rendering quality, and interaction paradigms shifting faster than research can consolidate findings [73]. On the other hand, most available evidence originates from studies conducted in physical settings or in tightly controlled single-user VR experiments,

---

<sup>6</sup>Anobii: <https://www.anobii.com/>

and its transferability to persistent, multi-user social virtual environments where the design must simultaneously support individual exploration, group interaction, and event attendance has not been systematically investigated [46], [48]. This leaves designers of social virtual worlds with limited empirical foundations on which to base their decisions.

This contribution addresses such gap through a focused case study: the redesign of the MetaLibrary virtual world. The original version of MetaLibrary was designed primarily as a proof of concept, with the main effort directed toward developing the platform’s core functionalities. As a consequence, the virtual world received comparatively less attention in terms of visual and experiential quality, relying heavily on cartoon-style graphics and lacking systematic grounding in user experience principles or perceptual considerations. Therefore, a major redesign was undertaken, guided by a systematic analysis of the literature and of existing social VR platforms, complemented by consultations with domain experts. While this represents only one of the many possible Metaverse applications, a virtual library constitutes a particularly fitting testbed for exploring environmental design criteria [60]: the multi-modal nature of the library experience spanning quiet individual browsing, spontaneous social encounters, and structured cultural events places simultaneous and sometimes competing demands on spatial layout, lighting, visual signposting, and avatar proximity management, making it a representative and challenging case for studying how design choices interact with user behavior in a social VR context.

A dedicated user study was additionally conducted to evaluate two specific design factors, i.e., lighting conditions and scene layout richness, for which the prior analysis did not yield sufficiently clear indications. The best practices emerging from this process are intended to serve not only as a foundation for the redesigned MetaLibrary, but also as a broader reference for designers and developers working on immersive social virtual worlds across a variety of application scenarios.

### **2.2.1 Background**

As noted above, the literature currently lacks comprehensive and structured guidelines for the design of virtual worlds in the Metaverse. Nevertheless, a number of studies have investigated how specific design choices affect users navigating and interacting within virtual spaces, offering a partial but informative view of the main factors in this context.

An influential contribution in this area is the work of [73], which proposed a taxonomy of space types relevant to the design of virtual worlds based on the results of a qualitative experiment. The taxonomy identifies five categories: personal spaces, supporting individual and focused work; collaboration spaces, designed for group interaction and discussion; presentation spaces, suited to the one-directional sharing or consumption of knowledge; making spaces, dedicated to practical experimentation and hands-on activities; and intermission spaces, serving as connective

areas for transition, recreation, and informal interaction.

Regarding the visual richness and level of detail of virtual worlds, a systematic review [7] indicated the importance of calibrating detail to the specific needs of the scenario, rather than pursuing a comprehensive replication of the physical world. High fidelity in terms of geometry, textures, and lighting is recommended primarily for interactive elements, as it helps users identify objects of functional relevance within the scene. The review also emphasized the value of including contextually meaningful elements, and noted that openings to the outside, such as windows, can support the construction of a coherent mental model of the environment and contribute to a stronger sense of presence.

Spatial perception in immersive virtual environments was the focus of [70], in which participants performed reading comprehension tasks across four virtual worlds varying in ceiling height and type. Significant differences emerged primarily in emotional response and room appreciation, with higher ceilings associated with more positive evaluations. Complementary evidence was provided by [74], a neurocognitive study examining how geometric properties including protrusion, curvature, scale, and proportion influence emotional responses. The findings indicated that in simple environments, larger and more symmetrical spaces tend to elicit positive emotional responses, whereas in more complex environments, positive reactions are more strongly associated with asymmetrical and protruding elements.

The visual attention and exploratory behaviors of users in virtual worlds' environment were examined in [75] through gaze tracking. Results showed that users tend to direct their attention toward objects presenting novel or unexpected information, as well as toward elements that correspond to the functional expectations associated with the specific application context.

A parametric framework for assessing how progressive modifications to architectural variables affect users' emotional states was developed in [76]. Across 54 virtual architectural configurations varying in sidewall distance, window and ceiling height, and color, the study found that negative emotional responses were consistently triggered by narrower wall spacing, elevated windowsill heights, and deviations from standard ceiling height in either direction.

The importance of natural elements in virtual worlds was suggested by a systematic review in [77], which gathered evidence showing that exposure to biophilic design patterns, an approach rooted in the built environment industry that seeks to reconnect users with natural features, consistently reduces stress and promotes positive emotional states comparable to those experienced in real natural settings.

Interior color schemes and their effects on human emotions, physiological responses, and task performance were investigated in [78]. Participants completed a proofreading task in four virtual world's virtual environments differentiated by color (red, white, blue, and green). Red was perceived as significantly warmer, more oppressive, and less pleasant than the alternatives, while the white environment was associated with the best task performance. Physiological measurements

showed a notable reduction in heart rate in the blue, white, and green conditions, suggesting a calming effect associated with these color schemes.

Finally, [79] compared users' emotional responses to a real urban environment and its digitally reconstructed counterpart, rendered with on-site recorded audio, high-fidelity visual details, and animated elements under both daylight and night-time conditions. The results revealed that emotional responses to the simulated environment did not always align with those elicited by the real one, and notable differences emerged between the two lighting conditions: while the night-time scenario was perceived as more realistic, it was also associated with a reduced sense of safety. These findings suggest that lighting conditions and their interaction with the overall scene design merit further investigation, particularly in indoor virtual environments.

### 2.2.2 Materials and methods

This section describes the overall process followed in the redesign of the MetaLibrary virtual world, detailing the analysis steps that informed the design decisions.

#### Design process

The redesign process began with a series of interviews conducted with staff members at three public libraries in the city of Turin: the Biblioteca Civica Centrale, the Bianca Guidetti Serra library, and the Primo Levi library. Prior to the interviews, the original version of MetaLibrary virtual world was demonstrated to library visitors, allowing the expert staff to gather direct feedback from members of the general public. The comments collected in this way were subsequently elaborated by the library professionals and reported during the interview sessions.

As a result, both usability issues specific to the case study and more general design limitations emerged from the interviews. These can be summarized as follows.

- *Visual style.* The cartoonish style adopted for both the virtual world and the avatars was not well received by users, who tended to associate the platform with a video game rather than a social cultural space. Library staff also noted that user satisfaction varied considerably depending on individual backgrounds and attitudes toward technology, a finding consistent with results reported in previous studies [80].
- *Visual inconsistency.* Marked differences in architectural structure and furniture layout across the various rooms created a sense of confusion and disorientation during navigation, undermining the coherence of the overall experience.
- *Absence of dedicated social spaces.* The original virtual world did not include areas specifically designed for informal interaction. Users wishing to engage in

casual conversations were limited to the central hub, which also served as the main navigation point, or to purpose-specific rooms whose primary function was thereby compromised.

It is worth noting that, while these limitations emerged from a specific institutional context, they are largely generalizable to the design of social virtual worlds more broadly, and are therefore relevant beyond the MetaLibrary case study.

Based on the limitations identified through the interviews, the following requirements were established to guide the redesign of the platform.

- *Visual realism.* The overall style should move away from cartoonish aesthetics in favor of more realistic assets, for both the virtual environment and the avatars. A more grounded visual language would better reflect the platform’s potential to host not only recreational activities, but also professional and culturally significant initiatives. The interviews further underlined the need to accommodate heterogeneous and multi-generational user groups, as reported also in related work [80].
- *Visual coherence.* While differentiation between areas can help users orient themselves within the environment, design choices should maintain a consistent visual language throughout the platform to prevent disorientation, in line with findings from the literature on virtual world design [72].
- *Dedicated social spaces.* The redesigned virtual world should include areas specifically designated for informal social interaction, consistent with evidence in the literature emphasizing the importance of purpose-defined spaces for fostering spontaneous social exchange [7].

The insights gathered through the interviews were complemented by a systematic review of the literature. The best practices extracted from this analysis are summarized in Table 2.1.

To further inform the design of the new virtual world, a selection of representative social VR platforms and Metaverse platforms were analyzed and, where possible, directly tested. The analysis aimed to identify recurring design patterns that have proven effective across different contexts and user bases.

- **VRChat<sup>7</sup>:** The platform hosts a wide variety of user-generated worlds spanning different visual styles, from photorealistic to cartoonish, making it difficult to identify a consistent design language. Among the most frequently visited virtual environments, recurring features included high ceilings, open spatial layouts, minimal room configurations, and a limited use of natural elements.

---

<sup>7</sup>VRChat: <https://hello.vrchat.com/>

Table 2.1: Best practices extracted from the analysis of the literature.

| Source                 | Best Practice                                                                                                                                                                                                                     |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Thoring et al. [73]    | Design environments around distinct space types to accommodate different kinds of activities. Flexible layouts that can adapt to varied user needs enhance both the experience and the functional versatility of the environment. |
| Neo et al. [7]         | Reserve high levels of visual realism for elements with which users directly interact, in order to support engagement while avoiding unnecessary resource consumption.                                                            |
| Cha et al. [70]        | Incorporate open layouts with high ceilings to promote positive emotional responses in users.                                                                                                                                     |
| Shemesh et al. [74]    | Favor larger, symmetrical, and curved spaces in simple environments; introduce protruding and asymmetrical elements in more complex settings to elicit positive emotional responses.                                              |
| Lee et al. [75]        | Populate environments with novel and contextually meaningful objects to capture users' attention and encourage exploratory behavior.                                                                                              |
| Presti et al. [76]     | Prioritize wide spaces and lower windowsills to support user comfort and prevent negative emotional responses associated with the perception of insufficient natural light.                                                       |
| Mollazadeh et al. [77] | Incorporate biophilic design patterns to promote positive emotional states and reduce perceived stress.                                                                                                                           |
| Cha et al. [78]        | Select color schemes that support task performance and minimize physiological stress responses.                                                                                                                                   |
| Bishop et al. [79]     | Account for the distinct emotional responses associated with day- and night-time lighting conditions when calibrating the perceived realism and sense of safety of the environment.                                               |

- **RecRoom**<sup>8</sup>: Similarly to VRChat, virtual environments are largely user-generated. The prevailing visual style is low-poly, with open and minimal spaces characterized by high ceilings and occasional natural elements.

---

<sup>8</sup>RecRoom: <https://recroom.com/>

- **Mozilla Hubs**<sup>9</sup>: The analyzed virtual environments reflected the heterogeneity typical of user-created content. Design choices tended toward minimalism, with a predominance of color palettes.
- **Spatial**<sup>10</sup>: Virtual environments are characterized by low-poly and cartoonish aesthetics, with polygon budgets constrained by platform performance requirements. Popular spaces, often used for art exhibitions set within natural contexts, typically feature high ceilings with overhead openings, open layouts, and minimal furnishing.
- **Zepeto**<sup>11</sup>: Virtual environments vary widely in style but tend to draw on real-world architectural references. Common features include high ceilings, large windows, sharp geometric forms, and minimal rooms. Lighting is predominantly natural, and color palettes are generally soft and desaturated.
- **Decentraland**<sup>12</sup>: User-generated virtual environments built using a dedicated game engine exhibit considerable variety. Architectural forms tend toward the complex, with curved geometries frequently preferred over simple rectangular layouts. The dominant visual style is low-poly with cartoonish elements.
- **Second Life**<sup>13</sup>: Virtual environments display a high degree of visual realism, supported by high-polygon models, detailed textures, and sophisticated lighting. The range of settings is broad, encompassing urban environments, natural landscapes, and architectural spaces of varying complexity, all shaped by the creative preferences of their creators.
- **Tangra**<sup>14</sup>: Virtual environments exhibit a comparatively higher degree of visual realism than other platforms analyzed. Despite variation across different functional contexts, recurring design features include high ceilings, natural elements, natural lighting, and cool color schemes.
- **Viverse**<sup>15</sup>: Virtual environments tend toward a low-poly and cartoonish aesthetic, with open spatial layouts and natural elements commonly incorporated.

---

<sup>9</sup>Mozilla Hubs: <https://hubs.mozilla.com/>

<sup>10</sup>Spatial: <https://www.spatial.io/>

<sup>11</sup>Zepeto: <https://web.zepeto.me/en>

<sup>12</sup>Decentraland: <https://decentraland.org/>

<sup>13</sup>Second Life: <https://secondlife.com/>

<sup>14</sup>Tangra: <https://tangra.link/>

<sup>15</sup>Viverse: <https://www.viverse.com/>



Table 2.2: Best practices extracted from the analysis of social VR platforms and immersive applications.

| Source                                                  | Best Practice                                                                                                         |
|---------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| VRChat, Spatial, Zepeto, Tangra, Meta Horizon Workrooms | Incorporate high ceilings, open spatial layouts, and large windows to foster a sense of openness and spatial comfort. |
| RecRoom, Spatial, Tangra, Viverse                       | Enrich the environment with biophilic design elements to create a more natural and pleasant atmosphere.               |
| Mozilla Hubs, Zepeto                                    | Adopt clean and visually uncluttered color schemes to support a simple and aesthetically coherent experience.         |
| Second Life, Tangra                                     | Leverage lighting effects to enhance the perceived realism of the environment.                                        |
| Second Life, Viverse                                    | Apply appropriate levels of visual fidelity to avatars and interactive objects to balance realism and performance.    |
| Decentraland                                            | Explore curved and non-rectangular room geometries to introduce visual interest and spatial variety.                  |

- **Meta Horizon Workrooms<sup>16</sup>**: Designed to support virtual collaboration and professional activities, the virtual environments in this platform are characterized by clean, modern aesthetics with customizable layouts. High-quality rendering supports a sense of presence conducive to focused work. Meeting spaces typically feature semicircular layouts, minimal furnishing, high ceilings, and large windows offering views of natural landscapes. Both warm and cool color schemes are employed, and natural lighting plays a prominent role in the overall visual atmosphere.

The design patterns most commonly shared across the analyzed platforms are summarized in Table 2.2. Notably, several of these recurring features, including high ceilings, open spatial layouts, biophilic elements, and natural lighting, align closely with the best practices identified through the literature review, further reinforcing their relevance for the redesign of MetaLibrary.

<sup>16</sup>Meta Horizon Workrooms: <https://forwork.meta.com/it/horizon-workrooms/>



## Implementation

The best practices identified through the analysis described in the previous section guided the development of the redesigned virtual world for the MetaLibrary platform, an overview of which is provided in Figure 2.3.

A first fundamental departure from the original version concerns the visual style. Cartoonish-style assets were replaced with more realistic representations for both the virtual environment and the avatars, in line with recommendations from the literature on virtual world design [7] and avatar representation [81].

Spatial openness was another central design priority. Consistent with evidence from both the scientific literature [70], [76] and the analysis of existing social VR platforms, the environment was designed around high ceilings and interconnected rooms without doors, as illustrated in Figure 2.2b, in order to promote a sense of openness and facilitate fluid navigation between areas.

Regarding spatial geometry, room walls were designed with curved profiles, and the overall layout incorporates several symmetrical elements, following the recommendations of [74]. These characteristics are clearly visible in the overview provided in Figure 2.3a.

The furnishing of the environment draws on contextual references appropriate to a library setting [75], with tables and chairs distributed across the various rooms, as depicted in Figure 2.3c. The decision was made not to include bookshelves populated with large numbers of volumes, as would be expected in a traditional library. This choice reflects the recommendation to avoid visual clutter that could distract users from the interactive elements, namely the recommended books displayed in each room [7].

All spaces are equipped with large windows, effectively implemented as transparent walls, which serve a dual purpose: ensuring adequate natural lighting within the virtual environment and providing views of outdoor natural scenery, in accordance with the guidelines derived from [76] and [77]. The color scheme, informed by the findings of [78], is based on a combination of white and blue as the dominant tones throughout the virtual shared environment.

**Central hub.** The central hub (Figure 2.3a) was designed taking the main hall of Shanghai Library East [82] as a primary architectural reference. This choice was motivated by the observation that, unlike the public libraries of Turin, Shanghai Library East embodies several of the design characteristics identified through the analysis, making it a particularly suitable model for reproduction in immersive VR. In the terminology proposed by [73], the hub functions as an *intermission space*: a large, open area where users spawn upon joining the platform and from which all other rooms are accessible. Like its real-world reference, the hub features high ceilings and curved walls. One notable departure from the original building concerns vertical circulation: all rooms in the redesigned environment are placed

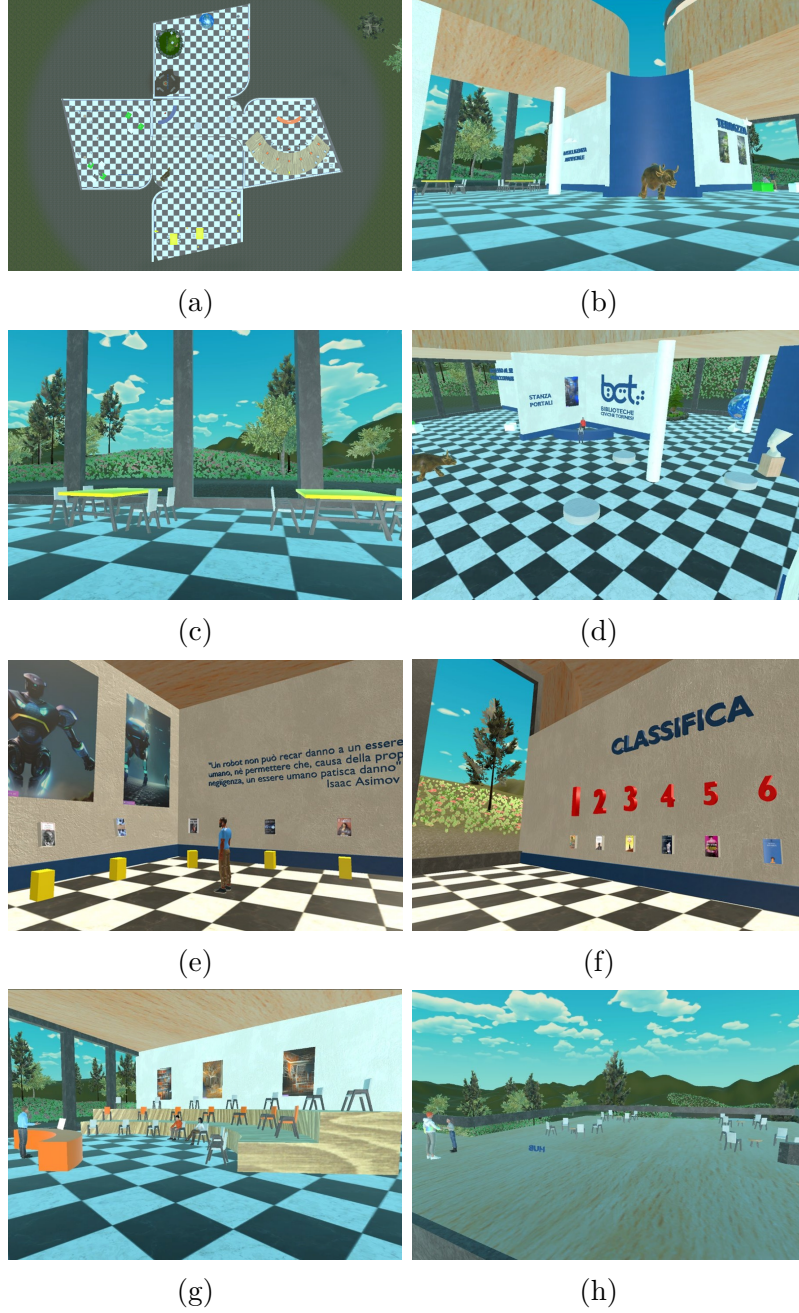


Figure 2.3: The redesigned virtual world of MetaLibrary: (a) overview, (b) high ceilings and interconnected rooms, (c) furniture, (d) central hub, (e) room of the day, (f) most-read room, (g) conference room, and (h) social activities area.

on the same floor, as navigating stairs or ramps in immersive [VR](#) is known to be a potential source of motion sickness [83]. Large openings toward the external virtual

environment are uniformly distributed across the hub’s perimeter, in line with the recommendations of [77].

**Room of the day and most-read room.** Both rooms reflect the characteristics of a *personal* and *making space* [73], offering focused virtual environments in which users can engage with curated book content. Their design was again inspired by the interior spaces of Shanghai Library East, and the two rooms are positioned opposite each other within the central hub to reinforce the overall spatial symmetry of the environment.

In the room of the day (Figure 2.3e), thematic elements related to the topic of the day, such as quotes and images associated with relevant authors and books, are displayed on the walls. This form of contextual customization has been shown to support engagement and a sense of purpose within cultural environments [4]. The five featured books are placed on individual pedestals to maximize their visibility and draw users’ attention, following the principle of highlighting interactive elements within the scene [7]. Tables and chairs are included to evoke a space conducive to focused reading and consultation, and to invite users to linger and interact with the content, thereby fostering knowledge exchange and collaborative exploration.

In the most-read room (Figure 2.3f), a dedicated wall displays the ranking of the seven most borrowed titles. Position numbers are rendered in red to create a clear focal point and attract users’ attention. Beyond its informational function, the ranking serves as an engagement mechanism, encouraging users to explore titles they may not have previously encountered.

**Conference room.** The conference room (Figure 2.3g) is designed as a *presentation space* [73], following a traditional semi-circular layout. A tiered seating arrangement ensures unobstructed sight lines to the presenter and the large display screen, supporting full audience engagement during events.

**Social activities area.** The social activities area (Figure 2.3h) functions as a *collaboration space* [73], dedicated to informal interaction and community building. Located on the building’s terrace, the area incorporates biophilic elements consistent with the outdoor natural scenery visible through the hub’s openings, reinforcing the environmental coherence of the platform and providing a welcoming setting for spontaneous social exchange.

**Technical implementation.** The virtual world was created using Blender for 3D modeling and Unity as the runtime engine. The platform targets the Meta Quest 2 headset, with VR functionalities handled through the OpenXR framework

and the XR Interaction Toolkit. The outdoor natural scenery was generated using the Unity Terrain Sample Asset Pack<sup>17</sup>.

Avatar representation follows a full-body humanoid approach consistent with prior work [41]. Avatars were generated using the Ready Player Me online tool and subsequently optimized within Unity through simplified rigging and inverse kinematics configurations, balancing animation quality with computational efficiency. Pre-recorded animations from Mixamo<sup>18</sup> were integrated and fine-tuned within Unity to ensure smooth transitions and naturalistic movement without requiring custom animation pipelines.

Several technical challenges were encountered during implementation. Achieving a balance between visual fidelity and rendering performance was a primary concern, given the sensitivity of immersive VR users to frame rate drops and the associated risk of motion sickness. Each 3D model was designed with a minimal polygon count, and textures and materials were optimized to reduce computational load. Asset bundling and dynamic loading strategies were adopted to prevent system overload during navigation. Collision detection was handled through simplified box colliders applied to non-interactive objects, avoiding the computational cost of mesh-accurate collision boundaries while still preventing undesired interpenetration with users' virtual bodies.

Maintaining visual coherence across the virtual world environment presented an additional challenge. This was addressed through a systematic revision workflow applied to each 3D asset, ensuring consistent color schemes, texture resolutions, and lighting effects prior to integration into the scene. Lighting optimization was handled through a combination of Unity's baked lighting and strategically placed real-time light sources, providing a visually convincing result while keeping the computational cost of dynamic shadows within acceptable bounds.

### **2.2.3 Experiment**

The majority of design decisions underlying the redesigned MetaLibrary virtual world were grounded in the best practices identified through the literature review and platform analysis. However, two factors, namely lighting conditions and scene layout richness, could not be resolved through this process alone, as no sufficiently actionable guidance was available in the existing literature. In particular, the reviewed studies on lighting focused predominantly on natural light, leaving the effects of artificial lighting in virtual environments largely unaddressed. Similarly, no clear guidelines were identified regarding the optimal density and arrangement

---

<sup>17</sup>Unity Terrain Sample Asset Pack: <https://tinyurl.com/y8nfm86>

<sup>18</sup>Mixamo: <https://www.mixamo.com/>

of furniture and objects within virtual spaces. While other factors might in principle have been included, the investigation was deliberately limited to these two variables to avoid the interpretive complications that arise when multiple factors vary simultaneously, which would have made it difficult to isolate the contribution of each.

A dedicated user study was therefore conducted to evaluate alternative configurations of the platform across these two factors, with the aim of identifying the combination best suited to each of the tasks performed within the virtual world environment.

### Procedure and tasks

Two levels were defined for each factor. For lighting, the conditions considered were: daytime with natural lighting (D) and night-time with artificial lighting (N). For layout, environments with a high density of objects and furniture (F) and largely empty environments (E) were compared. Combining these conditions yielded four experimental configurations: D+F, D+E, N+F, and N+E.

Participants were asked to perform three tasks within the redesigned MetaLibrary environment, each set in a different area of the platform and designed to reflect a distinct mode of interaction.

- *Task 1 (T1) – Conference attendance:* participants attended a short presentation delivered by a non-player character (NPC) in the conference room. The presentation lasted under two minutes and covered topics related to well-known books and authors.
- *Task 2 (T2) – Book exploration:* participants were placed in the room of the day and asked to interact with two of the books on display, retrieving information about the titles and the topic of the day.
- *Task 3 (T3) – Social interaction:* participants were invited to interact with three NPC avatars in the social activities area, each programmed to share opinions about a selection of books.

Figure 2.4 illustrates the three tasks and the corresponding environment configurations used in the study.

The experiment followed a  $3 \times 2 \times 2$  within-subjects design, with tasks, lighting conditions, and layout conditions as the three factors. After completing a demographic questionnaire and receiving general instructions, participants familiarized themselves with the MetaLibrary platform and its navigation controls using a Meta Quest 2 headset. Continuous movement was adopted as the locomotion method. Once sufficiently comfortable with the interface and informed about the tasks,



participants proceeded to complete each task across all four environment configurations. The order of both tasks and configurations was randomized for each participant to mitigate learning effects. A post-test questionnaire was administered upon completion of the experiment.

### Evaluation metrics

Both objective and subjective measures were collected during the experiment. For the objective dimension, fixation time was selected as the primary metric, defined as the proportion of task completion time during which participants directed their gaze toward a predefined target element. In T1, the relevant target was the large screen displaying the presentation delivered by the NPC on stage. In T2 and T3, the targets were a specific book in the room of the day and the NPC avatars in the social activities area, respectively. Fixation time was selected as an indicator of user engagement: sustained attention toward the assigned target suggests that the surrounding environment was not perceived as distracting or disorienting, and that participants remained focused on the task at hand.

For the subjective dimension, participants were asked to rate each of the four environment configurations (D+F, D+E, N+F, N+E) using the 28 adjective pairs of the AttrakDiff questionnaire [84], completed separately for each of the three tasks. The tool employs a seven-point rating scale for each pair and organizes responses into four subscales: pragmatic quality (PQ), hedonic quality split into identity (HQ-I) and stimulation (HQ-S), and overall attractiveness (ATT). Upon completion of all tasks, participants were also asked to rank the four configurations according to their overall preference and to provide open comments on their experience.

### Participants

Twenty participants took part in the study, of whom 13 identified as male and seven as female. Ages ranged from 24 to 34 years ( $M = 26.04$ ,  $SD = 2.39$ ). The sample was heterogeneous with respect to prior VR experience: 55% of participants reported regular use of VR technology, 35% had limited experience, and the remaining 10% had none. Familiarity with social and virtual platforms was generally low, with 95% of participants reporting little or no prior use. Conversely, 75% of participants indicated frequent engagement with video games.

#### 2.2.4 Results and discussion

As the primary test, a statistical analysis was conducted applying a three-way repeated-measures ANOVA. Where appropriate, two-factor or one-factor repeated-measures ANOVA with Bonferroni correction was used for follow-up comparisons. The normality assumption was verified using the D'Agostino-Pearson test. Rankings were analyzed using the Friedman test, with the Wilcoxon signed-rank test

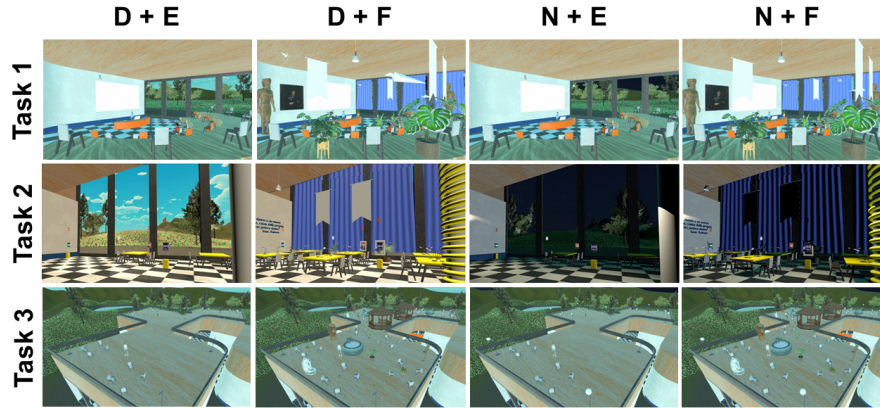


Figure 2.4: Overview of the environment configurations and tasks included in the user study.

applied for pairwise comparisons.

### Objective results

Figure 2.5 reports the average fixation time percentages across tasks and environment configurations.

The main effects of tasks and layout conditions were both statistically significant. Regarding tasks, fixation time differed significantly across the three conditions (T1: 58.80% vs T2: 49.94% vs T3: 45.67%,  $p < .001$ ), with follow-up analysis confirming significant differences between T1 and T2 ( $p = .011$ ) and between T1 and T3 ( $p = .001$ ). These results indicate that participants directed proportionally more attention toward the task-relevant target during the conference attendance task than during the book exploration and social interaction tasks, regardless of the environment configuration. Regarding layout, the full condition yielded significantly higher fixation time than the empty condition (F: 53.15% vs E: 47.78%,  $p = .008$ ), suggesting that a richer layout encouraged participants to focus more consistently on the designated target. A significant interaction was also found between task and layout conditions ( $p < .001$ ), indicating that the influence of layout on fixation time was not uniform across tasks.

### Subjective results

Figure 2.6 presents the average AttrakDiff scores across configurations and tasks.

**Pragmatic quality.** Significant main effects were found for both tasks (T1: 5.39 vs T2: 4.90 vs T3: 5.11,  $p = .009$ ) and layout conditions (E: 5.68 vs F: 4.58,  $p < .001$ ). Follow-up analysis confirmed significant differences between T1 and

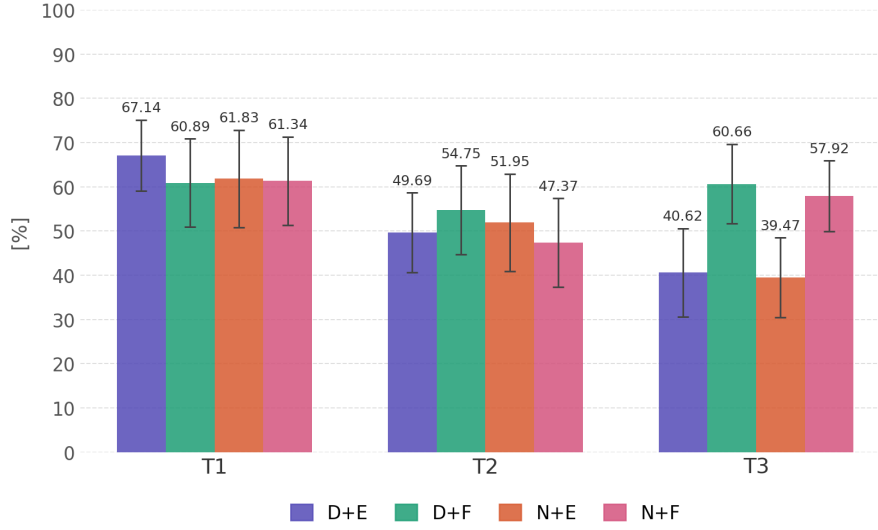


Figure 2.5: Objective results: average fixation time percentages across tasks and environment configurations.

T2 ( $p = .002$ ) and between T1 and T3 ( $p = .023$ ), indicating that the conference room environment was consistently rated as more usable than those associated with the other two tasks. Furthermore, empty layouts were rated as more usable than full ones across all tasks, a trend clearly visible in Figure 2.6, where the lines corresponding to the empty configurations (D+E in blue and N+E in gray) systematically exceed those of the full configurations (D+F in orange and N+F in yellow). A significant interaction between tasks and layout conditions was also observed; detailed comparisons are provided in Table 2.3.

**Hedonic quality - Identity.** No statistically significant differences were found for this subscale across any of the main effects or interaction terms in the three-way ANOVA.

**Hedonic quality - Stimulation.** Significant main effects were identified for both lighting (D: 4.54 vs N: 4.82,  $p = .046$ ) and layout conditions (E: 4.06 vs F: 5.30,  $p < .001$ ). The night-time configuration was perceived as more stimulating and engaging than the daytime one, and full layouts elicited higher stimulation ratings than empty ones. A significant interaction between tasks and layout conditions was also found ( $p = .018$ ); detailed comparisons are reported in Table 2.4.

**Overall attractiveness.** A significant main effect was found only for the layout factor (E: 4.88 vs F: 5.35,  $p < .001$ ), with full layouts rated as more appealing and attractive than empty ones across all tasks and lighting conditions.



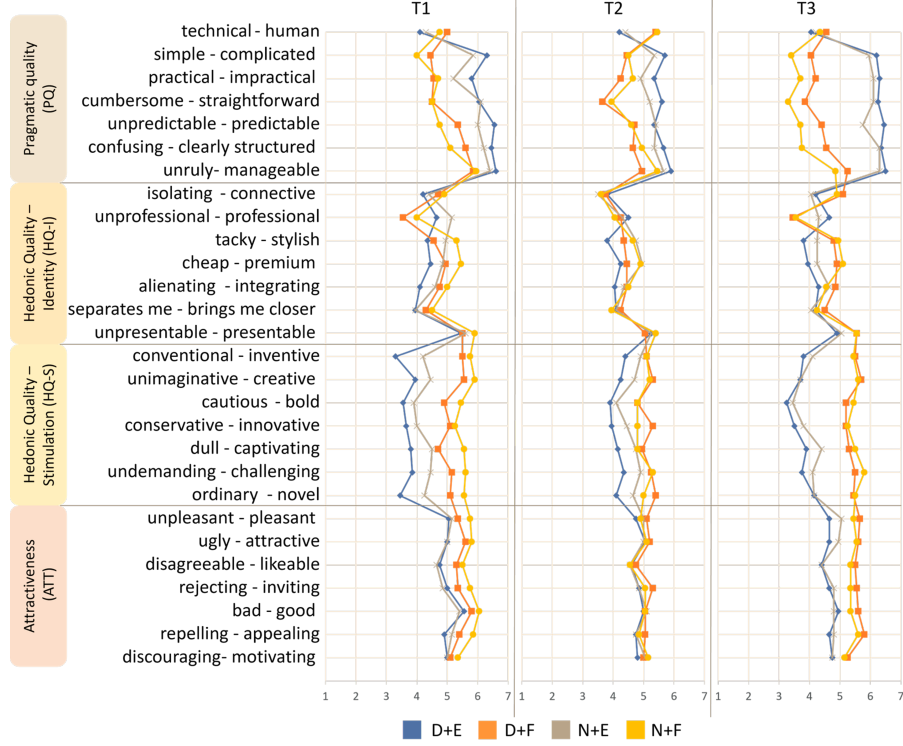


Figure 2.6: Subjective results: average AttrakDiff scores across environment configurations and tasks [84].

Table 2.3: P-values of the follow-up analysis on the PQ subscale of the AttrakDiff tool for the interaction effect between task and layout conditions. Bold values indicate statistically significant differences.

|             | <b>T1×E</b><br>$\mu = 5.85$<br>( $\sigma = 0.56$ ) | <b>T1×F</b><br>$\mu = 4.93$<br>( $\sigma = 1.03$ ) | <b>T2×E</b><br>$\mu = 5.26$<br>( $\sigma = 0.93$ ) | <b>T2×F</b><br>$\mu = 4.55$<br>( $\sigma = 1.06$ ) | <b>T3×E</b><br>$\mu = 5.94$<br>( $\sigma = 0.69$ ) | <b>T3×F</b><br>$\mu = 4.28$<br>( $\sigma = 1.09$ ) |
|-------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|
| <b>T1×E</b> | -                                                  | <b>&lt;.001</b>                                    | <b>.003</b>                                        | <b>&lt;.001</b>                                    | .484                                               | <b>&lt;.001</b>                                    |
| <b>T1×F</b> |                                                    | -                                                  | .109                                               | .084                                               | <b>.001</b>                                        | <b>.004</b>                                        |
| <b>T2×E</b> |                                                    |                                                    | -                                                  | <b>.001</b>                                        | <b>.005</b>                                        | <b>&lt;.001</b>                                    |
| <b>T2×F</b> |                                                    |                                                    |                                                    | -                                                  |                                                    | .200                                               |
| <b>T3×E</b> |                                                    |                                                    |                                                    |                                                    | -                                                  | <b>&lt;.001</b>                                    |
| <b>T3×F</b> |                                                    |                                                    |                                                    |                                                    |                                                    | -                                                  |

**Configuration ranking.** Statistically significant differences in overall ranking were observed only for T1 ( $p < .001$ ). Pairwise comparisons did not yield a complete ordering of the four configurations; however, D+E was consistently ranked lower than D+F ( $p = .001$ ), N+E ( $p = .018$ ), and N+F ( $p = .002$ ), indicating a clear dispreference for the daytime, empty configuration in the context of conference

Table 2.4: P-values of the follow-up analysis on the HQ-S subscale of the AttrakDiff tool for the interaction effect between task and layout conditions. Bold values indicate statistically significant differences.

|             | <b>T1×E</b><br>$\mu = 3.95$<br>( $\sigma = 1.30$ ) | <b>T1×F</b><br>$\mu = 5.36$<br>( $\sigma = 0.79$ ) | <b>T2×E</b><br>$\mu = 4.40$<br>( $\sigma = 1.18$ ) | <b>T2×F</b><br>$\mu = 5.08$<br>( $\sigma = 0.59$ ) | <b>T3×E</b><br>$\mu = 3.84$<br>( $\sigma = 1.26$ ) | <b>T3×F</b><br>$\mu = 5.46$<br>( $\sigma = 0.64$ ) |
|-------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|----------------------------------------------------|
| <b>T1×E</b> | -                                                  | <b>&lt;.001</b>                                    | .079                                               | <b>&lt;.001</b>                                    | .681                                               | <b>&lt;.001</b>                                    |
| <b>T1×F</b> |                                                    | -                                                  | <b>.001</b>                                        | .096                                               | <b>&lt;.001</b>                                    | .498                                               |
| <b>T2×E</b> |                                                    |                                                    | -                                                  | <b>.006</b>                                        | <b>&lt;.001</b>                                    | <b>&lt;.001</b>                                    |
| <b>T2×F</b> |                                                    |                                                    |                                                    | -                                                  | <b>&lt;.001</b>                                    | <b>.015</b>                                        |
| <b>T3×E</b> |                                                    |                                                    |                                                    |                                                    | -                                                  | <b>&lt;.001</b>                                    |
| <b>T3×F</b> |                                                    |                                                    |                                                    |                                                    |                                                    | -                                                  |

Table 2.5: P-values of the pairwise comparison on configuration rankings for T1 (overall Friedman test:  $p < .001$ ). Bold values indicate statistically significant differences.

|            | <b>D+E</b><br>med = 4 | <b>D+F</b><br>med = 2 | <b>N+E</b><br>med = 2.5 | <b>N+F</b><br>med = 2 |
|------------|-----------------------|-----------------------|-------------------------|-----------------------|
| <b>D+E</b> | -                     | <b>.001</b>           | <b>.018</b>             | <b>.001</b>           |
| <b>D+F</b> |                       | -                     | .194                    | .759                  |
| <b>N+E</b> |                       |                       | -                       | .399                  |
| <b>N+F</b> |                       |                       |                         | -                     |

attendance. Full results of the pairwise analysis are reported in Table 2.5.

## Discussion

The objective results highlight the significant influence of task type on users' visual attention within the virtual world. Participants directed more sustained attention toward the task-relevant target during the conference attendance task than during book exploration or social interaction. This outcome can be plausibly attributed to the architectural characteristics of the conference room, whose semi-circular layout naturally channels users' gaze toward the presentation stage. In contrast, the tasks carried out in more open spaces, namely the wide room of the day in T2 and the terrace in T3, offered fewer structural cues to guide visual attention, resulting in lower fixation rates. This observation reinforces the principle that the spatial design of a virtual world should be carefully tailored to the specific activities it is intended to support.

The results related to layout density provide a more nuanced picture. On one hand, the objective data indicate that a richer layout was associated with higher fixation time on the designated targets, a result that might appear counterintuitive

at first glance. On the other hand, the subjective PQ ratings consistently favored empty layouts in terms of perceived usability, suggesting that a higher density of objects introduced a degree of visual complexity that participants found less comfortable to navigate, even if it did not reduce their task focus. The HQ-S subscale, however, told a different story: full layouts were rated as significantly more stimulating and engaging than empty ones. These findings point to an inherent tension between usability and stimulation in virtual environment design. Reducing the density of objects improves perceived usability but risks creating an environment that feels sparse and less engaging, while richer layouts enhance stimulation at the cost of a somewhat more demanding user experience. Identifying the appropriate balance between these two dimensions therefore represents a key design challenge, and one that is likely to depend on the specific nature of the tasks users are expected to perform.

The significant interactions observed between task and layout conditions across both objective and subjective measures confirm that these two factors are not independent, and that the optimal layout configuration cannot be determined without reference to the intended use of the space.

Regarding lighting, the results were less conclusive. A marginal preference for the night-time condition over the daytime one was observed for the HQ-S subscale, suggesting a slightly higher perceived stimulation under artificial lighting. However, no significant effects were found for the other subscales or for the objective metric, and the overall influence of lighting on user experience in this study remains inconclusive. Further investigation involving a broader range of lighting configurations and a more diverse set of tasks will be necessary to draw more definitive conclusions on this aspect.

## 2.3 Considerations and remarks

This chapter investigates how the structural design of immersive virtual worlds can influence the overall user experience in social virtual worlds. In particular, the chapter explored how architectural choices, spatial organization, and environmental design shape the way users explore virtual spaces, focus their attention, and participate in shared activities.

The first study introduced MetaLibrary, an immersive social virtual world designed to support reading-based cultural experiences, knowledge exploration, and interaction among users. The platform was designed to recreate some of the social dynamics of physical libraries while using immersive technologies such as remote participation, virtual events with authors, and AI-based book recommendation mechanisms [39]. The deployment of the system within a real-world context, with a dedicated installation accessible to visitors at Turin’s central public library since April 2023, indicated the potential of immersive environments to extend the role of libraries beyond their physical boundaries, enabling new forms of community

engagement and shared cultural experiences [85].

The second study examined design aspects of immersive virtual worlds, investigating how different environmental configurations influence user attention and perceived quality of experience [57]. The study’s design process drew on expert interviews, a systematic literature review, and an analysis of existing social VR platforms, resulting in a set of design guidelines for immersive social virtual worlds. These practices offer actionable guidance for designers working in this space. Spatial layouts incorporating high ceilings, open configurations, and large windows promote a sense of openness, accommodate diverse activities, and contribute positively to users’ emotional responses. Flexible and adaptable layouts further enhance user experience while maximizing the functional versatility of the environment. High fidelity should be reserved for interactive elements such as avatars and objects, where it most directly supports engagement; for the broader environment, more stylized approaches can be equally effective, and the strategic use of lighting effects can enhance perceived realism. In simpler virtual environments, larger and curved spaces work better, symmetrical, and curved spaces are recommended to foster positive emotional responses, whereas in more complex environments, asymmetrical and protruding elements can serve the same purpose. Biophilic elements such as natural patterns and textures contribute to a more agreeable and emotionally positive atmosphere. Color schemes should be selected to support a clean and visually coherent aesthetic, with evidence suggesting benefits for both task performance and physiological stress reduction. Finally, the emotional responses associated with daytime and night-time lighting conditions should be taken into account to calibrate the perceived realism and sense of safety of the environment [57].

Beyond the derivation of these best practices, the second study also analyzed the effects of lighting conditions and scene layout complexity through both objective metrics and subjective user evaluations [86], targeting design factors for which no sufficiently actionable guidance could be derived from prior analysis. The results show that the spatial configuration of the environment plays a significant role in shaping user behavior. Tasks performed in environments with clearly structured spatial layouts, such as the semi-circular conference room, led participants to focus their attention more effectively on the intended targets [57]. Conversely, tasks performed in open spaces containing a higher number of visual elements encouraged exploration but could also reduce the ability to maintain attention on specific targets [87], [88].

Subjective evaluations also highlighted the presence of a trade-off between pragmatic and hedonic qualities of immersive environments. Simpler layouts with fewer objects were generally perceived as more usable and easier to navigate, leading to higher pragmatic quality scores. At the same time, richer and more complex environments were considered more stimulating and engaging from a hedonic perspective [89], [90]. These findings suggest a balance is needed between clarity and environmental richness in order to support both usability and engagement.

Another relevant observation concerns the relationship between spatial design and the specific activities that users are expected to perform within the virtual world. The results indicate that different tasks may benefit from different architectural configurations. Environments designed to support focused activities, such as attending presentations or lectures, may benefit from spatial structures that guide attention toward a specific focal point. Conversely, activities centered on exploration or informal social interaction may benefit from more open and visually diverse environments that encourage movement and discovery [12]. The nature of the task thus emerges as a key moderating factor in how users perceive spatial organization, underscoring the importance of aligning architectural decisions with intended use patterns.

Overall, the studies presented in this chapter highlight the importance of considering spatial architecture as a key design dimension in immersive social virtual worlds [6]. The organization of spaces, the density of environmental elements, and the visual characteristics of the environment can all influence how users perceive and interact with immersive systems. These aspects become particularly relevant in social virtual worlds, where spatial design not only affects individual user experience but also shapes the conditions under which social encounters and shared activities can emerge.

While the findings presented in this chapter provide useful indications for the design of immersive environments, several aspects deserve further investigation. The effects of lighting conditions in particular require deeper study, as the present results yielded only partial and inconclusive findings. Future work should examine interaction effects between design factors, potentially through factorial breakdown analyses, and explore additional dimensions not addressed in the current studies, including animated environmental elements, non-player characters, avatar interaction, and sound design. Dynamic lighting adjustments and the effects of varying degrees of user interactivity and environmental customization on presence and engagement also merit dedicated attention. From an environmental design perspective, it would also be valuable to compare the current low-poly visual approach with more photorealistic rendering styles, in order to assess how visual fidelity affects usability, perceived quality of experience, and social dynamics. Future studies should also include a broader range of tasks, which would strengthen the generalizability of the reported findings. In addition, the transferability of the identified best practices to application domains beyond the library context should be systematically assessed through comparative studies across diverse Metaverse scenarios.



## Chapter 3

# Design and Augmentation of Social Interaction in Avatar-Mediated Environments

*The work described in this chapter has been formerly published in [30], [34], [41], [42].*

A fundamental aspect of user experience in immersive social virtual worlds is the ability to communicate and interact with other users [9]. While the structural design of virtual environments defines how users inhabit and navigate these spaces, the quality of the overall experience is largely determined by how effectively users can engage with one another through avatar-mediated interaction. In these environments, communication is not a secondary feature but the primary mechanism through which users establish relationships, coordinate activities, and construct shared experiences [91], [92], [93]. Human communication is inherently multimodal: in face-to-face interaction, the exchange of meaning relies on a rich combination of verbal content, facial expressions, body posture, gaze direction, and physical contact, with non-verbal channels accounting for a substantial portion of the overall communicative bandwidth [94], [95]. This multimodal structure is not merely additive: research in embodied cognition suggests that the body itself is constitutively involved in the production and comprehension of meaning, such that the absence or degradation of bodily cues does not simply reduce communicative bandwidth but can qualitatively alter the nature of the interaction [96]. Replicating this richness in avatar-mediated environments is one of the central and most persistent challenges in social VR research [97], [98]. Unlike traditional video-conferencing, which preserves the visual appearance of participants at the cost of

spatial co-presence, social VR worlds place users within shared three-dimensional environments where they are represented by digital avatars. This shift opens new possibilities for naturalistic interaction, but also introduces a new layer of technical and perceptual complexity: the quality of social experience depends critically on how faithfully, and how expressively, avatars can convey the non-verbal signals that users produce. This dependency is further complicated by the Proteus effect [99]: users' behavior and self-perception are themselves shaped by the appearance and expressiveness of their own avatar, meaning that limitations in avatar fidelity do not merely impair signal transmission but may recursively affect how users engage with and present themselves within the interaction.

The challenges involved span multiple, interconnected dimensions of non-verbal communication. A first critical aspect concerns the representation of facial expressions and emotions. In immersive settings, the presence of a HMD and the limitations of current sensing technologies make it difficult to accurately capture and reproduce the full range of human facial behavior, often resulting in reduced emotional expressiveness or ambiguous signals [41], [100], [101], [102]. This is particularly consequential because facial expressions serve not only as signals of internal emotional states but also as regulators of turn-taking and social coordination [103]: their degradation therefore risks disrupting the temporal fabric of conversation, not only its affective content. From a measurement standpoint, the dominant paradigm of Action Unit-based coding [104] provides a useful decomposition of facial movement, but its direct applicability to real-time avatar animation is constrained by the high-dimensional and context-dependent nature of expressive behavior, which resists reduction to a fixed set of discrete units [105]. A second dimension relates to full-body motion and the overall quality of avatar movement. Many social interactions rely on subtle variations in posture, gesture, and coordination, which are difficult to reconstruct from the sparse tracking data typically available in consumer VR systems. Consumer VR setups typically provide tracking for the head and two hand controllers, leaving the configuration of the remaining degrees of freedom of the body, torso, elbows, hips, knees, to be inferred through inverse kinematics [106]. While modern IK solvers can produce plausible poses for a single user in isolation, they are not designed to enforce the interpersonal coordination constraints that arise in dyadic interaction, where the relative positioning and timing of two bodies must jointly satisfy biomechanical and social plausibility criteria [107]. As a result, avatar movements may appear less natural or less synchronized with user intentions, potentially affecting the perceived realism and effectiveness of the interaction [108], [109]. A particularly challenging aspect concerns physically grounded social interactions between users, such as greetings or other forms of close-contact behavior. These interactions require the real-time coordination of multiple avatars within a shared space, raising issues related to spatial alignment, motion coherence, and behavioral plausibility. Unlike isolated gestures, contact-based interactions such as handshakes or hugs are inherently relational motor acts: their execution



is shaped by continuous sensorimotor prediction and mutual adjustment between participants [110]. In avatar-mediated settings, where haptic feedback is absent and visual feedback of the partner’s motion is subject to network latency, this predictive loop is broken, placing greater demands on the animation system to compensate through plausible motion synthesis. Ensuring that such interactions are perceived as natural and meaningful remains an open problem in the design of social VR experiences [111], [112], [113].

Beyond the faithful reproduction of natural interaction, social VR also opens the possibility of deliberately augmenting non-verbal communication in ways that go beyond what unaided human behavior can achieve. **Artificial Mimicry (AM)**, defined as the automated replication and amplification of verbal and non-verbal cues, is one of the most studied augmentation strategies in this space [93], [114]. The underlying mechanism draws on well-established findings from social psychology: unconscious mimicry of a conversational partner’s expressions and gestures is a powerful facilitator of rapport, trust, and affective convergence [115], [116], [117]. The pro-social effect of mimicry is thought to operate through multiple pathways: at the perceptual level, being mimicked increases the sense that one is understood and validated by the interlocutor [115]; at the affective level, the motor simulation of observed expressions engages the observer’s own affective system, producing a degree of emotional contagion that supports interpersonal convergence [96], [117]. Critically, these effects are strongest when mimicry is perceived as natural and unintentional: deliberate or overly literal imitation tends to be detected and perceived negatively [118], which imposes a design constraint on AM systems, they must introduce sufficient temporal and parametric perturbation to preserve the appearance of spontaneity. When applied to facial expressions through Artificial Facial Mimicry, avatars can dynamically modulate their displayed emotions in response to the affective state of their interlocutor, compensating for the expressive limitations of current systems while simultaneously enriching the non-verbal communication layer [111], [119]. Despite its theoretical appeal, the application of such augmentation techniques to fully immersive, avatar-mediated interaction remains comparatively underexplored, particularly with respect to their impact on social presence, trust, and interpersonal connection [114].

Addressing these challenges in a principled and empirically grounded manner is essential to realizing the potential of social VR as a medium for meaningful human connection, and represents one of the most active frontiers in the field [97], [120], [121]. This chapter addresses this set of open problems through four complementary investigations, each targeting a distinct aspect of social interaction in avatar-mediated environments. The first study investigates the fidelity with which realistic full-body avatars can convey emotions, comparing a hardware-based facial tracking approach against a software-based lip synchronization solution across six basic emotional expressions [41]. The second study examines how non-verbal interaction (specifically, greeting gestures like handshakes) can be effectively designed

using handheld controllers, evaluating four interaction variants ranging from GUI-based to gesture-driven approaches and assessing their impact on embodiment, social presence, and interaction naturalness [30]. The third study addresses the challenge of close-contact avatar interactions, proposing a machine learning-based framework that refines inverse kinematics-driven animations in real time to produce more natural and synchronized movements during physical social gestures such as hugs [42]. Finally, the fourth study explores the intentional augmentation of social interaction through **Artificial Facial Mimicry (AFM)**, demonstrating that the controlled amplification of users' emotional facial expressions can significantly enhance interpersonal closeness, trust, and social presence without compromising the perceived naturalness of avatar behavior [34]. Together, these contributions advance the understanding of how avatar-mediated communication can be designed and augmented to support richer and more meaningful social experiences in immersive virtual worlds.

### **3.1 Comparing technologies for emotion facial expressions in social virtual worlds**

The rapid diffusion of VR technologies has stimulated a growing body of research examining their effects across a wide range of application domains [122]. Among these, social VR and remote collaboration have emerged as particularly active areas, a trend further accelerated by the disruptions to physical interaction brought about by the global pandemic [100].

Within the broader context of the Metaverse, researchers are increasingly investigating how users can be effectively represented through avatars in shared virtual environments. Avatars serve as the primary interface through which users' movements, expressions, and social presence are projected into the virtual world [123]. They can vary considerably in visual style, ranging from photorealistic [124] to cartoonish [125], and in structural complexity, from minimalist representations showing only the head and hands [126] to full-body configurations that also reproduce facial expressions [100], [124]. Research on avatar representation and its psychosociological and perceptual implications remains at a relatively early stage [127], yet existing evidence consistently suggests that users prefer avatars that are realistic, full-body, and expressive [100], [128]. Reproducing facial expressions in VR, however, presents a significant technical challenge, particularly when users wear a head-mounted display that partially occludes the face. Dedicated facial tracking devices are increasingly available on the consumer market to address this challenge [129], but an alternative software-based approach relies on lip synchronization, which generates synchronized mouth animations on a speaking avatar without requiring additional hardware [102].

Against this backdrop, the study presented in this section compares two modalities for conveying emotions through realistic full-body avatars in a multi-user VR scenario. The first modality leverages consumer-level facial tracking hardware, specifically the HTC VIVE Facial Tracker, while the second relies on a software-based lip synchronization solution, SALSA LipSync Suite v2<sup>1</sup>. A total of 28 participants wore a head-mounted display and observed an avatar performing six scenes, each designed to portray one of Ekman’s six basic emotions [130]. Each scene was presented twice, once per modality, in a pseudo-randomized order, with the aim of replicating conditions as close as possible to real social VR usage. Following each pair of scenes, participants blindly evaluated the two modalities across a set of relevant dimensions, including comfort, expressiveness, realism, naturalness, pleasantness, and emotion conveyance.

### 3.1.1 Background

The influence of avatar appearance on user experience in multi-user extended reality scenarios has been the subject of numerous investigations. Researchers have examined a wide range of avatar characteristics, including aesthetic style, behavioral repertoire, movement quality, and expressive capacity, and their effects on interpersonal communication, non-verbal interaction, and social presence [131], [132], [133], [134], [135], [136], [137], [138], [139].

Roth et al. [140] investigated the effects of reducing social information channels in immersive virtual environments with full-body humanoid avatars, finding that lower avatar realism hindered social interaction. However, the absence of facial expression analysis in their study limited the scope of its conclusions. Dobre et al. [128] compared realistic and cartoonish full-body avatars in a mixed reality telepresence context, demonstrating that the non-verbal behavior of the realistic avatar was perceived as more appropriate and informative for social interaction than its cartoonish counterpart.

The role of body completeness in avatar representation was examined by Yoon et al. [124] in an augmented reality setting, where three levels of body visibility, covering head and hands only, upper body, and whole body, were compared. Consistent with the findings reported in the previous section, a realistic whole-body avatar was found to be the most effective configuration for supporting remote collaboration.

Kasapakis and Dzardanova [141] investigated the integration of facial and eye tracking in a multi-user VR learning environment, using a high-fidelity educator avatar driven by a combination of body, eye, and facial tracking devices. Their results indicated that the combined presence of body motion, eye movement, and facial cues helped 90% of students maintain attention and improve their understanding during virtual lectures.

---

<sup>1</sup>SALSA LipSync Suite: <https://tinyurl.com/2p7pj2mt>

Hube et al. [102], extending their earlier work [142], examined the contribution of facial visual parameters to emotion perception in VR. Using the SALSA LipSync Suite to apply real-time lip synchronization to a half-body avatar, they compared the animated avatar against an audio-only condition in which the avatar remained static. The results confirmed that the addition of facial motion improved participants' ability to identify the emotions conveyed in the audio clip. The study did not, however, account for the contribution of body language, and the range of non-verbal cues evaluated was limited relative to the full complexity of real-world emotional expression.

Overall, the reviewed literature supports the use of realistic, full-body avatars as the most effective representation approach across a range of social VR scenarios, and highlights the added value of facial expressiveness for user experience. Both hardware-based facial tracking and software-based lip synchronization have been explored as means of implementing facial cues, yet a direct comparison of these two approaches applied to a full-body animated avatar in a multi-user social VR context had not been conducted prior to the present study.

On the basis of this gap, the following hypothesis was formulated. Facial tracking technology is expected to convey emotional content more effectively than lip synchronization, given that the latter captures only lower face motion while the former reproduces the full range of facial and eye dynamics. However, since facial expressions in real social VR settings are always accompanied by voice and body movement, and since these contextual cues may compensate for lower facial fidelity in certain situations, the difference between the two modalities is expected to be less pronounced for some emotions than for others.

### **3.1.2 Materials and methods**

The virtual environment was developed in Unity 2021.3 as an OpenXR application and was experienced using an HTC VIVE Pro Eye<sup>2</sup> VR kit. Both modalities under evaluation were implemented on top of the realistic full-body avatar. Full-body motion was generated using the VRIK module of Final IK by RootMotion<sup>3</sup> [143], which reconstructs plausible body poses from the position and orientation of the headset and hand controllers through inverse kinematics for the upper body and blended animations for locomotion. Voice input was captured through the built-in microphone of the headset.

---

<sup>2</sup>HTC VIVE Pro Eye: <https://www.vive.com/us/product/vive-pro-eye/overview/>

<sup>3</sup>Final IK: <https://tinyurl.com/rMotionFinalIK>

## Evaluated configurations

The first modality, referred to as **Facial+eye Tracking (FT)**, combined the eye tracking capabilities built into the HTC VIVE Pro Eye with an externally mounted HTC VIVE Facial Tracker<sup>4</sup>. This configuration was selected for several reasons. The HTC VIVE ecosystem is widely used within the social VR community and supports the integration of additional tracking peripherals for full-body reconstruction [144]. The combination of the built-in eye tracker and the facial tracker provides a convenient, consumer-available solution for real-time facial expression capture. Together, these devices enabled continuous mapping between the user's facial movements and eye behavior and the corresponding blendshapes of the avatar mesh, which was created using Autodesk Character Generator<sup>5</sup>. Character models produced by this tool come with a complete body and facial rig, making them straightforward to integrate with Final IK (Figure 3.1). Minor manual adjustments were required in Autodesk Maya to add a *frown* blendshape not included in the default rig.

The second modality, referred to as Lip-sync Estimation (**LE**), replaced hardware-based facial capture with a purely software-driven approach. Real-time lip synchronization was implemented using the SALSA LipSync Suite v2, following a methodology similar to that of Hube et al. [102]. The *One-Click* automatic configurator for Autodesk Character Generator was used to streamline integration with the humanoid mesh, and the *Eyes* module was enabled to generate procedural eye motion. SALSA was selected for this role on account of its widespread adoption in Unity-based VR applications, its favorable cost-to-performance ratio, and its prior use in the research literature, all of which contributed to its suitability as a representative software-based baseline for comparison.

## Test scenario

Emotion conveyance was assessed through a per-emotion analysis covering all six of Ekman's basic emotions: anger, happiness, sadness, fear, disgust, and surprise [130]. This represents an extension of the approach adopted in Hube et al. [102], which focused on three emotions only. For each emotion, participants observed a short scripted scene in which an actor portrayed the target emotion in a social context. Scripts were designed with reference to the valence and arousal dimensions commonly used in emotion research [145], and scenes were kept under one minute in duration while being made as socially naturalistic as possible. The original scripts in Italian, their English translations, and video recordings of all six

---

<sup>4</sup>HTC VIVE Facial Tracker: <https://www.vive.com/us/accessory/facial-tracker/>

<sup>5</sup>Autodesk Character Generator: <https://charactergenerator.autodesk.com/>



Figure 3.1: First-person view of the user's avatar.

scenes are publicly available<sup>6</sup>.

To avoid potential biases arising from live variability or repeated exposure to the same performance, scenes were pre-recorded rather than performed live in a real multi-user session (Figure 3.2). A single amateur actor (Figure 3.2a) performed all six scenes to ensure consistency across stimuli. The recorded inputs used to drive the avatar, comprising voice, head and hand movements, and facial blendshapes, were captured using the same hardware described above (Figure 3.2b). In the **FT** condition, all recorded streams were played back simultaneously, reproducing the effect of a real-time performance. In the **LE** condition, the facial blendshape and eye tracking data were discarded and replaced with SALSA-generated lip sync, while body and voice inputs remained identical across both conditions.

### 3.1.3 Experiment

The study was conducted as a within-subjects experiment involving 28 participants (16 male, 12 female) aged between 21 and 68 years, recruited from among students and staff at Politecnico di Torino. The same hardware configuration used for scene recording was employed throughout the experiment.

Upon arrival, participants completed a demographic questionnaire covering age, gender, and prior experience with **VR** and social **VR** applications. The sample was largely inexperienced with immersive technologies: approximately 61% rated their **VR** experience at 2 or below on a five-point scale, and around 78% reported almost no prior use of social **VR** or Metaverse platforms.

---

<sup>6</sup>Scene scripts and recordings: <https://tinyurl.com/3rjpu8et>





Figure 3.2: (a) Actor performing and recording the happiness scene using the VR setup. (b) First-person view of the recorded emotional scene inside the virtual environment.

Participants were informed that the study focused on the expressiveness and communicative capacity of avatars, with particular attention to the upper body, where the three principal channels of non-verbal communication, namely vocal tone, arm and hand gestures, and facial expressions, are most prominently expressed. They were then given the headset and guided into a virtual environment modeled after a theater<sup>7</sup>, where the pre-recorded avatar performed the scripted scenes on a virtual stage. Participants were free to move around the stage area using natural walking to observe the avatar from their preferred viewpoint.

Scene order was counterbalanced using a Latin square design to control for ordering effects across the six emotions. Each scene was presented once per modality in randomized order, with participants unaware of which modality was active during each viewing. After watching both versions of a scene representing a given emotion, participants completed a comparative questionnaire assessing comfort, expressiveness, realism, naturalness, likelihood, pleasantness, emotional conveyance, and overall preference. For negatively valenced emotions such as disgust, anger, and sadness, participants were instructed to evaluate the scenes from a neutral and detached perspective to avoid conflating emotional content with perceived quality. This procedure was repeated for all six emotions, and participants were invited to provide open-ended comments at the end of the session. The full questionnaire is publicly available<sup>8</sup>.

---

<sup>7</sup>Madame Walker Theatre: <https://tinyurl.com/yvj6h7ca>

<sup>8</sup>Experiment questionnaire: <https://tinyurl.com/yykdffz9>

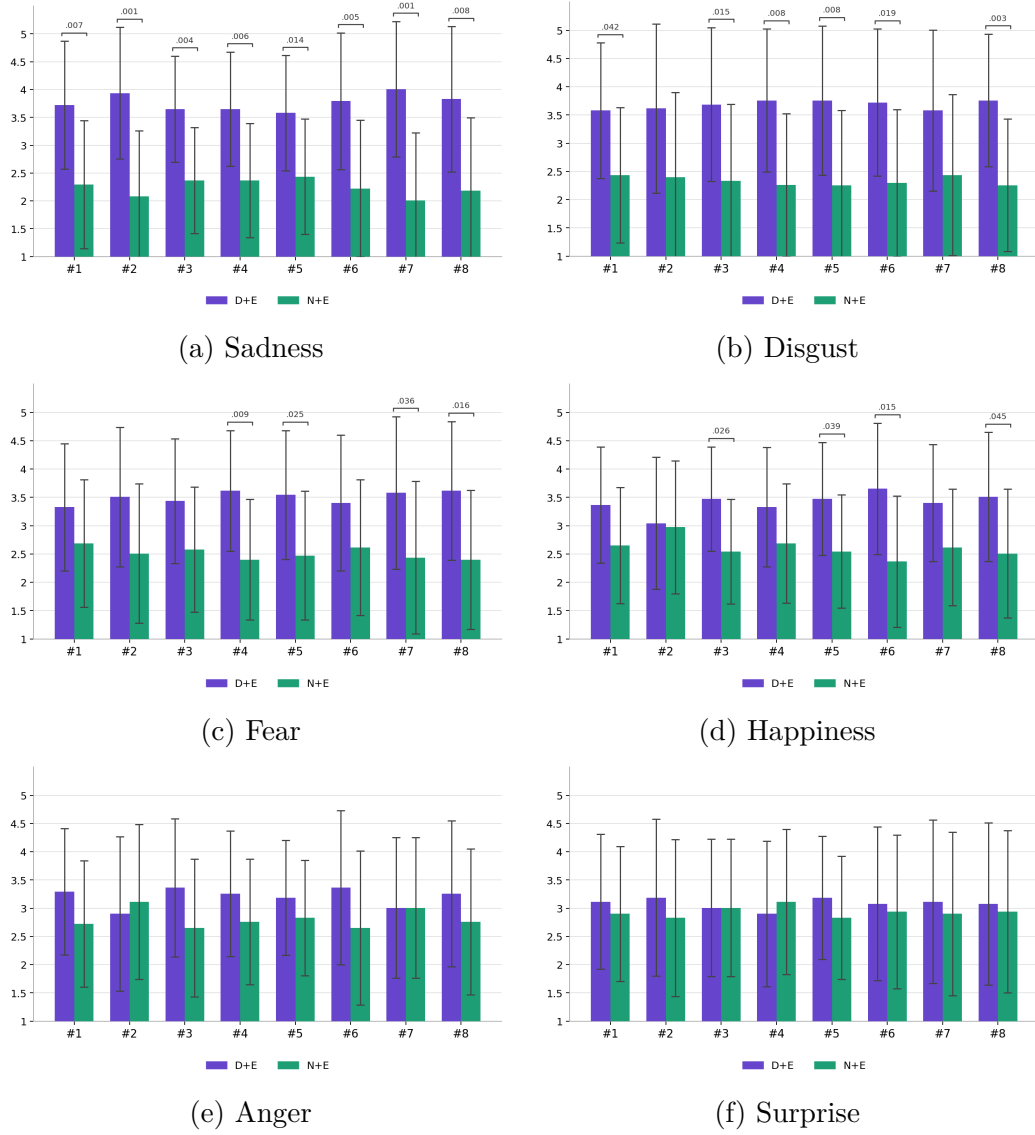


Figure 3.3: Subjective results for each scene and emotion (a-f). Statistically significant differences ( $p < .05$ ) are marked with brackets; error bars represent standard deviations. Questions #1 to #8 asked participants to indicate which of the two modalities: #1 made them feel more comfortable; #2 was more expressive; #3 was more realistic; #4 was more natural; #5 was more plausible; #6 was more pleasant; #7 better conveyed the emotional content; #8 was preferred overall.

### 3.1.4 Results and discussion

The results of the experiment are summarized in Figure 3.3.



Data normality was assessed using the Shapiro-Wilk test. Given that the distributions deviated significantly from normality, statistical comparisons were performed using the non-parametric Wilcoxon signed-rank test, with a significance threshold of  $p < .05$ .

**Sadness.** The **FT** modality was rated as significantly superior to **LE** across all eight evaluated dimensions (Figure 3.3a). Participants perceived **FT** as more expressive both in terms of general avatar expressiveness and emotional conveyance. Consistent advantages were also observed for pleasantness, comfort, realism, and overall preference. Differences were less pronounced but still statistically significant for naturalness and plausibility. These results align with theoretical expectations: according to Ekman’s framework [130], sadness is primarily communicated through coordinated eye and mouth movements, both of which are accurately reproduced by the **FT** modality but only partially approximated by the **LE** approach.

**Disgust.** In the disgust scene (Figure 3.3b), **FT** again outperformed **LE**, though the advantage was less consistent than for sadness. Participants rated **FT** as more realistic, more natural, more plausible, more pleasant, more comfortable, and preferable overall. However, no significant differences emerged for expressiveness or emotion conveyance. This likely reflects the transient nature of disgust, which tends to manifest as a brief and involuntary facial reaction rather than a sustained expression. When an emotion is displayed rapidly, participants may not be directing their attention toward the avatar’s face at the precise moment of expression, making subtle differences between the two modalities harder to detect. This interpretation is consistent with comments volunteered by several participants in the open feedback section.

**Fear.** The **FT** modality was perceived as superior to **LE** on four of the eight dimensions in the fear scene (Figure 3.3c): naturalness of expression, plausibility, emotion conveyance, and overall preference. No significant differences were found for the remaining dimensions. This pattern is consistent with the particular importance of upper face dynamics in the expression of fear, including eyebrow raises and cheekbone tension, which are faithfully reproduced by **FT** but absent in the **LE** approximation, which is restricted to the lower face.

**Happiness.** The happiness scene (Figure 3.3d) did not yield a clear overall winner, though **FT** was rated significantly higher than **LE** for realism, plausibility, pleasantness, and general preference. Notably, expressiveness scores were nearly identical across the two modalities. This unexpected result can be attributed to a hardware limitation: the HTC VIVE Facial Tracker experienced difficulty reliably detecting smiling gestures and triggering the corresponding blendshape. Several

participants explicitly noted in their feedback that the avatar did not appear to smile despite the context of the scene, effectively reducing the expressive advantage of the FT modality for this particular emotion.

**Anger.** No conclusive differences between the two modalities were found for the anger scene (Figure 3.3e). This outcome may partly reflect the impulsive and short-lived nature of anger as an expressed emotion, but participant feedback also pointed to a more specific factor: the scene script relied heavily on verbal delivery, which may have reduced the perceptual salience of facial differences between FT and LE. In a speech-dominant context, the lip synchronization provided by LE may be sufficient to meet participants' expectations, narrowing the gap with the higher-fidelity FT approach.

**Surprise.** Similarly, no significant differences were observed between the two modalities for the surprise scene (Figure 3.3f). Surprise shares with anger a fleeting quality: it is characterized by a brief, highly expressive peak that is difficult to sustain over time. In this context, the combination of body movement and vocal cues may have provided sufficient information for participants to perceive and evaluate the emotion, making differences in facial fidelity between FT and LE less perceptually relevant than they might otherwise be.

## 3.2 Controller-based non-verbal interaction design for social virtual worlds

The previous study examined how emotions are conveyed through the avatar's face, isolating a single channel of non-verbal communication. Facial expression accounts for only part of the signals exchanged during social interaction: a substantial share of interpersonal meaning is carried by the body, and in particular by gestures involving physical contact [146]. Such gestures are difficult to reproduce when users are represented by avatars and act through handheld controllers [95], and among them the handshake stands out as one of the most culturally widespread forms of social contact [94], [146]. The focus therefore shifts from facial to bodily signals. In this context, research efforts have extended beyond verbal communication to address the non-verbal dimensions of social interaction, which play a fundamental role in supporting meaningful interpersonal exchange [20], [94], [95], [98], [147]. Non-verbal communication and social touch, however, are particularly difficult to replicate in virtual environments [95]. Some social VR platforms have begun to address this challenge by supporting symbolic gestures such as the handshake, which has been shown to establish interpersonal bonds more effectively than verbal exchange alone [94], [146]. Despite its social significance, handshaking in social VR settings remains understudied.

Against this backdrop, the investigation reported in this section proposes and evaluates four handshake interaction designs based on the use of handheld controllers: a first design triggered through a GUI, a second leveraging a dedicated button on the controller, and two designs based on physically recreating the handshaking gesture by grabbing the other party’s virtual hand and shaking it. A user study was conducted to assess the impact of these different interaction modalities on the quality of the social VR experience.

### 3.2.1 Background

Mediated handshaking [148] has been investigated across several interaction contexts, including human-robot communication [31], video-based telepresence platforms [149], and non-immersive virtual environments [94]. However, the findings from these settings are not straightforwardly applicable to social VR. Handshaking robots and video-telepresence systems involved a strong kinesthetic haptic feedback component that is absent in controller-based VR. In the robot case, users also exhibited heterogeneous levels of anthropomorphism, while in the video-telepresence case interaction was not avatar-mediated. The avatar-mediated study by Tanenbaum et al. [94] is more directly relevant, but was conducted in a non-immersive environment, where the lower levels of embodiment perceived by users may limit the generalizability of its findings to immersive social VR.

More recent work has begun to address this gap more directly. Kasapakis et al. [20] proposed a system supporting handshaking and other non-verbal cues in virtual environments with a high degree of physical fidelity, obtaining preliminary favorable results in terms of usability and user experience. However, the system relied on hand-tracking gloves, which remain uncommon among everyday social VR users, and the study involved a limited sample size. Sykownik and Masuch [95] adopted a more accessible setup based on standard handheld controllers to investigate how different non-verbal gestures affect the development of intimacy in social VR. While their work demonstrated the relevance of touch in virtual social contexts, handshaking was not included among the gestures evaluated.

The handshaking gesture specifically was the focus of the study by Imaizumi et al. [147], who demonstrated that modifying the color of the user-controlled avatar’s hand could serve as a form of pseudo-haptic feedback, conveying the illusion of temperature and enhancing the perceived quality of the interaction. Their work used the UltraLeap hand-tracking system, which enables finger and hand tracking without controllers but provides no direct means of delivering haptic stimulation to the user’s hands. This limitation is particularly relevant given that handheld controllers remain the dominant input interface for social VR users, even though they involve a trade-off between tracking robustness and the sense of body ownership and naturalness [150]. In current commercial social VR platforms, the most common implementation of handshaking relies on gestures triggered through a GUI, a

solution that prioritizes technical simplicity over interaction naturalness.

The research reported in the following addresses the limitations of these prior contributions by proposing four controller-based handshake interaction designs that span a range of abstraction levels, from GUI-triggered to gesture-based, and evaluating their effects on embodiment, presence, social presence, usability, and the subjective quality of the handshake experience.

### **3.2.2 Materials and methods**

This section describes the four handshake interaction variants proposed and evaluated in the study. To ensure controllable and repeatable experimental conditions, a dedicated social VR testbed was developed, allowing participants to experience all four variants through interactions with both NPCs and a user-controlled avatar.

#### **Technology**

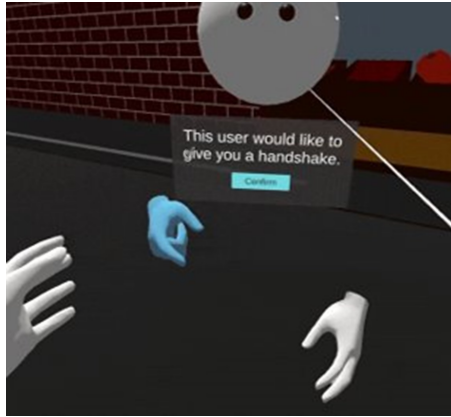
The testbed was implemented using the Unity game engine (v2021.3 LTS) with the XR Interaction Toolkit plugin. The virtual environment was populated with a combination of freely available 3D assets and custom models created in Blender (v3.3 LTS). Locomotion was implemented through joystick-based continuous movement, following the approach described in [83]. Multi-user connectivity and in-experience voice communication were handled through the Photon PUN 2 framework and its integrated voice module.

The application was deployed using the OpenXR framework on a Meta Quest 2 headset connected to a desktop PC via Air Link, with its bundled controllers as the primary input device. The headset features a per-eye display resolution of  $1832 \times 1900$  pixels, a  $90^\circ$  horizontal field of view, and a 120 Hz refresh rate.

#### **Testbed scenario**

The visual design of the testbed adopted a low-poly, cartoonish aesthetic inspired by popular social VR platforms such as Mozilla Hubs, Roblox, and RecRoom. The virtual environment was laid out as a small village square surrounded by commercial buildings, providing a contextually appropriate and socially recognizable setting for handshake interactions.

Avatar design followed a deliberately minimalist approach, with sphere-shaped heads, no body trunk, and stylized hands. This choice was made to direct participants' attention toward the gestural aspects of the handshake without visual distractions from more complex body representations.



(a) Handshake 1 (HS1)



(b) Handshake 2 (HS2)



(c) Handshake 3 (HS3)



(d) Handshake 4 (HS4)

Figure 3.4: The four handshake interaction variants, ordered by increasing gestural fidelity. (a) HS1: GUI-triggered. (b) HS2: controller-button-triggered. (c) HS3: grab-based, with predefined completion animation. (d) HS4: grab-based, with free shaking motion and no predefined animation.

### Variant design

The four interaction variants (Figure 3.4) were designed around three phases identified as constitutive of the handshake gesture:

- *Handshake Start (S)*: one party approaches the other within interaction range and initiates the handshake. The initiating party can cancel before the other responds; the receiving party can accept or decline.
- *Handshake Grant (G)*: the receiving party accepts the invitation and completes the gesture.
- *Handshake Feedback (F)*: both parties receive confirmation that the handshake

has been successfully completed.

A distance threshold of 1.5 m was set as the minimum proximity required to initiate a handshake. If either party moves beyond this range before the interaction is completed, the handshake is automatically cancelled. The S and G phases were kept consistent across variants, yielding the four designs described below.

- *Handshake 1 (HS1)*: this variant replicates the most common implementation found in commercial social VR platforms, in which a graphical user interface mediates the interaction [148]. When the two parties are within range, a floating menu appears above the dominant hand of each user. S is performed by pressing a button via ray-cast interaction; the button then transforms into a cancel option for the initiating party. Upon initiation, the first party's right hand extends and turns light blue, and a grant button appears near the receiving party's avatar. F is delivered through a predefined animation in which both right hands meet and shake three times, accompanied by a sound effect. This variant is illustrated in Figure 3.4a.
- *Handshake 2 (HS2)*: this variant retains the overall structure of HS1 while replacing GUI-based ray-cast interactions with physical button presses on the controller, with the aim of improving comfort and interaction efficiency. The menus appear as in HS1, but S is performed by pressing and holding one of the top controller buttons; releasing the button cancels the handshake. G is performed by the receiving party using the same button. F is unchanged. This variant is illustrated in Figure 3.4b.
- *Handshake 3 (HS3)*: this variant prioritizes interaction naturalness and gestural fidelity, at the potential cost of learnability and perceived control. S is initiated by physically reaching toward the other party's hand and pressing the grip button on the controller; releasing cancels the action. To manage the asymmetry of the grab interaction, the two parties are presented with complementary views: the initiating party sees a placeholder hand attached to their own, maintaining the visual impression of having grabbed the other user's hand, while the receiving party retains full control of their own hand and is notified through a GUI prompt and the light-blue hand animation that a handshake request is pending. G requires the receiving party to reach and grab the extended hand of the initiating user. This design was chosen to preserve a sense of agency and embodiment for both parties. F is delivered through the same predefined animation as HS1 and HS2, since the hands are already in the correct position upon completion of G. This variant is illustrated in Figure 3.4c.
- *Handshake 4 (HS4)*: this variant extends HS3 by removing the predefined animation from F, allowing the handshake to be performed freely by both



parties. G is slightly modified: the handshake is not considered granted as soon as mutual grabbing occurs, but only when both parties actively perform the shaking motion (moving the hand up and down while maintaining the grip button pressed). During F, both parties control their own hands freely while seeing the other avatar's hand attached to their own, with the sound effect retained as the only automated feedback element. This variant is illustrated in Figure 3.4d.

### 3.2.3 Experiment

This section describes the experimental setup designed to evaluate the four handshaking variants under controlled conditions. The study was structured around a dedicated handshaking activity in which participants systematically experienced all possible interaction outcomes for each variant. The following subsections detail the handshaking activity protocol, the overall experimental procedure, the participant sample, and the measures collected.

#### Handshaking activity

A structured handshaking activity was designed to ensure that each participant experienced the full range of possible interaction outcomes for each variant. Before starting, participants viewed a short tutorial video explaining the mechanics of the variant they were about to experience. Hands-on practice with another avatar was intentionally withheld at this stage, as learnability and intuitiveness were among the dimensions to be measured.

Following the tutorial, participants performed the handshaking activity with two NPCs and a user-controlled avatar operated by a confederate. The first NPC, playing the role of the village mayor, never initiated a handshake but could accept or decline one (with a 25% probability of declining). The second NPC, playing the role of a local retailer, always initiated the handshake and left the participant to respond. The user-controlled avatar was operated to cover the remaining interaction combinations. After each successful handshake, the user-controlled avatar moved away and reappeared after 30 seconds; to maintain the illusion of interacting with a different person each time, the avatar's color changed at each reappearance in a predefined order consistent across all participants.

The complete set of possible interaction outcomes comprised six distinct action types: the participant initiates and cancels before a response; the participant initiates and the other party declines; the participant initiates and the other party grants (successful handshake); the other party initiates and cancels; the other party initiates and the participant declines; the other party initiates and the participant grants (successful handshake). Each action type was experienced at least twice with both the NPCs and the user-controlled avatar, for a total of 36 interactions

per variant. The average time to complete the activity was approximately 12 minutes.

## Procedure

The study followed a within-subjects design. Before the handshaking activity, participants familiarized themselves with locomotion in the virtual environment through a sandbox session using the same layout but without other avatars present. They then completed the handshaking activity with each of the four variants. Exposure order was counterbalanced using a Latin square design to minimize ordering and learning effects.

Sample size was determined through an a-priori power analysis conducted with the G\*Power tool [151]. With  $\alpha = 0.05$  and a target minimum effect size of Cohen's  $f \geq 0.25$ , a sample of 24 participants was found to be sufficient to achieve a power of  $(1 - \beta) = 0.81$  for the planned within-subjects design [152]. Accordingly, 24 volunteers were recruited from the authors' professional network and from among staff and students at the authors' institution.

## Participants

A before-experience questionnaire was administered prior to the experiment, collecting information on demographics, familiarity with relevant technologies, and real-life handshaking habits. Participants ranged in age from 21 to 35 years ( $\bar{x} = 25.70$ ,  $s.d. = 2.95$ ); 75% identified as male and 25% as female. Regarding VR experience, 54% reported moderate to high familiarity with immersive VR, while the remaining 46% had little or no prior experience. Thirty percent of the sample had used a social VR platform at least once in the six months preceding the study. Only 8% reported a moderate or strong aversion to handshaking in real life, and 70% indicated that handshaking was a habitual part of their social introductions.

## Measures

Subjective data were collected through a multi-section questionnaire administered after each variant, comprising 81 items in total and organized as follows.

*Embodiment.* The influence of each variant on the level of embodiment with respect to the user-controlled avatar was assessed using the scale proposed in [153], reduced to the overall, location, ownership, and agency factors.

*Presence and Social Presence.* Presence was measured using the i-Group Presence Questionnaire (IPQ) [154]. Social presence was assessed through the Networked Minds Measure Scale (NMMS) [155].

*Usability.* Usability was evaluated from multiple perspectives. The System Usability Scale (SUS) [156] served as an overall indicator. Sections 1 and 8 of the VR-USE questionnaire [157] were additionally employed to assess functionality,



error correction, and robustness, along with the related subscales covering appropriateness, ease of use, intuitiveness, learnability, and system performance.

*Handshake experience.* A custom set of questions (available as supplemental material) was used to gather information on the perceived similarity of each variant to a real-life handshake, including whether participants could identify who was leading the interaction and how the personality of the other avatar was perceived across variants.

At the end of the experiment, participants were asked to explicitly rank the four variants across the following dimensions: ease of starting, ease of completing, ease of cancelling when self-initiated, ease of cancelling when other-initiated, ease of declining, naturalness, task control, distraction level, hand embodiment, and overall preference.

### 3.2.4 Results and discussion

Statistical significance was assessed using the Friedman test with Conover correction, followed by the Wilcoxon signed-ranks test as a post-hoc procedure.

**Embodiment, presence, and social presence.** No significant differences were found across the four variants for overall embodiment ( $\bar{x}_{HS1-4} = 4.95 \pm 0.24$ ,  $p = .113$ ) or any of its subscales, including body ownership ( $\bar{x}_{HS1-4} = 5.29 \pm 0.24$ ,  $p = .112$ ), body location ( $\bar{x}_{HS1-4} = 3.89 \pm 0.43$ ,  $p = .392$ ), and agency ( $\bar{x}_{HS1-4} = 5.69 \pm 0.23$ ,  $p = .058$ ). Similarly, no significant differences were identified for overall presence (IPQ:  $\bar{x}_{HS1-4} = 4.33 \pm 0.15$ ,  $p = .392$ ) or for most social presence subscales, including mutual awareness ( $\bar{x}_{HS1-4} = 6.79 \pm 0.09$ ,  $p = .392$ ), empathy ( $\bar{x}_{HS1-4} = 5.16 \pm 0.21$ ,  $p = .194$ ), behavioral interdependence ( $\bar{x}_{HS1-4} = 5.33 \pm 0.27$ ,  $p = .598$ ), and mutual assistance ( $\bar{x}_{HS1-4} = 6.75 \pm 0.13$ ,  $p = .999$ ). A small but statistically significant difference emerged only for the dependent action subscale ( $p = .037$ ), with HS4 perceived as allowing slightly more freedom of action compared to HS1 ( $p = .041$ ,  $\eta_p^2 = 0.28$ ) and HS2 ( $p = .041$ ,  $\eta_p^2 = 0.28$ ), suggesting that the gesture-based completion mechanism reduced the perceived reliance on the other avatar for completing the interaction.

**Usability.** Usability results were statistically significant across most evaluated dimensions, consistently reflecting the ordering  $HS4 > HS3 > HS2 > HS1$ , corroborated by the aggregate VR-USE measure shown in Figure 3.5. SUS scores confirmed this trend ( $HS1 = 68.44$ ,  $HS2 = 81.88$ ,  $HS3 = 89.17$ ,  $HS4 = 93.85$ ; Friedman  $p < .001$ ), with all pairwise differences reaching significance: HS1 vs. HS2 ( $p < .001$ ,  $d = 0.81$ ), HS1 vs. HS3 ( $p < .001$ ,  $d = 1.32$ ), HS1 vs. HS4 ( $p < .001$ ,  $d = 1.14$ ), HS2 vs. HS3 ( $p = .001$ ,  $d = 0.62$ ), HS2 vs. HS4 ( $p < .001$ ,  $d = 1.14$ ), and HS3 vs. HS4 ( $p = .009$ ,  $d = 0.52$ ).

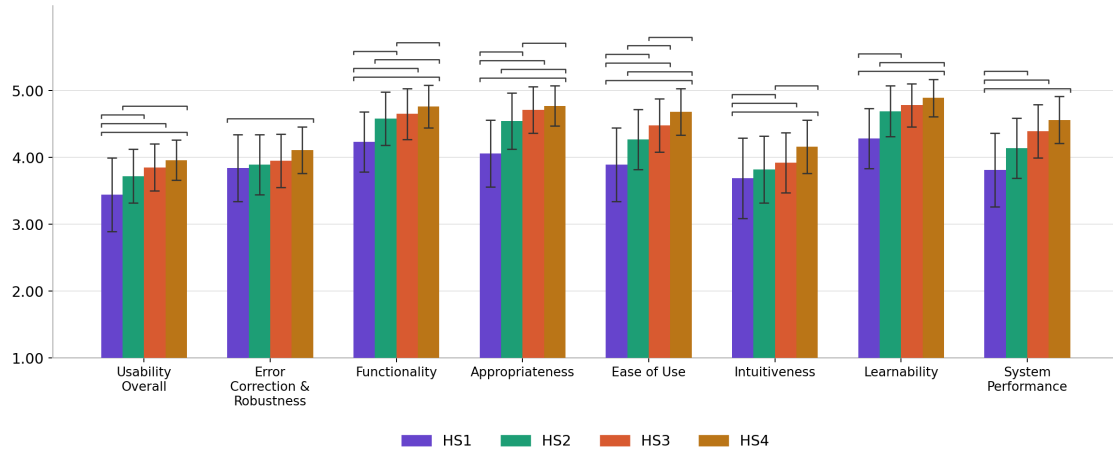


Figure 3.5: Usability measures across the four handshake variants based on the VR-USE questionnaire. Brackets indicate statistically significant differences ( $p < .05$ ).

Functionality scores followed the same overall ordering. Significant differences in error correction and robustness were found only between HS4 and the remaining three variants, suggesting that the absence of a predefined completion animation in HS4 may have affected participants' perception of interaction robustness. No significant effect of the animation was observed on the appropriateness dimension. Ease of use and system performance followed the general trend, as did intuitiveness and learnability, though no significant difference was found between HS2 and HS3 for these latter subscales.

**Handshake experience.** A more complete picture emerges from the handshake experience scores (Figure 3.6). The general trend was confirmed for overall score, fidelity, efficiency, and perceived bonding with the other avatar. Notably, no significant differences were found in the perception of who was leading the handshake across variants. With respect to personality assessment through the handshake, HS1 was rated as significantly worse than HS3 and HS4, suggesting that GUI-mediated interaction may act as a communication barrier by reducing the social expressiveness of the gesture. Regarding comfort, no significant differences were observed between HS3 and HS4, both of which were rated as the most comfortable variants, a result that was somewhat unexpected given the greater complexity of the gesture-based interaction.

**Subjective rankings.** Explicit preference rankings (median ranks, lower = better) were collected across ten dimensions; the Friedman test revealed significant differences for all of them except ease of cancelling when other-initiated, for which

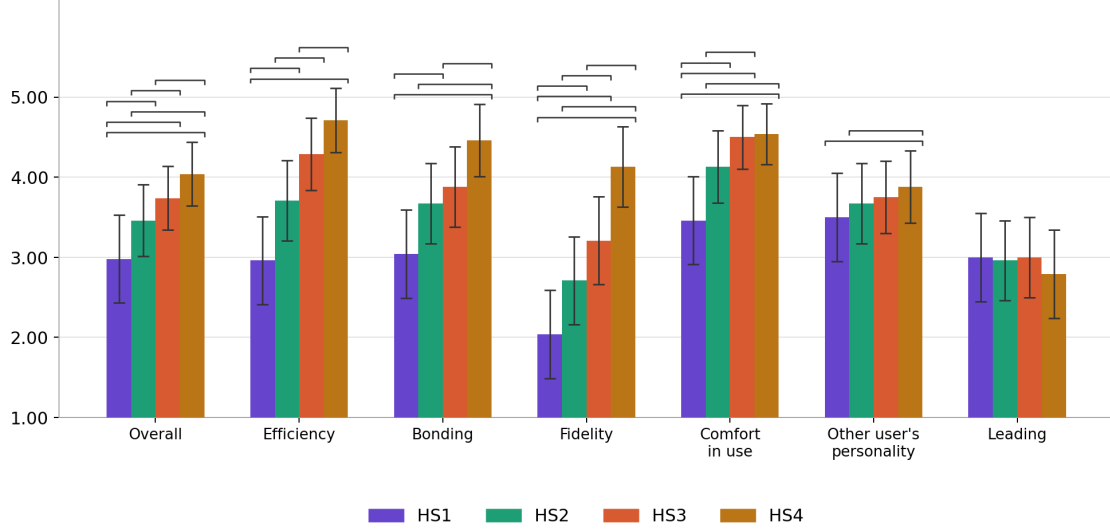


Figure 3.6: Scores from the ad-hoc handshake experience questionnaire across the four variants. Brackets indicate statistically significant differences ( $p < .05$ ).

all variants tied at rank 1 ( $p = .233$ ).

For dimensions related to task completion, HS4 ranked first overall but the difference with HS3 was not statistically significant. This held for ease of granting (HS3/HS4:  $p = .098$ ,  $d = 0.47$ ), ease of self-initiated cancellation, where HS3 and HS4 tied at median rank 1 (HS3/HS4:  $p = .065$ ,  $d = 0.56$ ), and ease of declining (HS3/HS4:  $p = .059$ ,  $d = 0.34$ ). Ease of starting the handshake was the only task-completion dimension where HS4 was rated significantly better than HS3 ( $p = .034$ ,  $d = 0.60$ ), though the effect size was small. These results suggest that the modified completion feedback introduced in HS4 did not meaningfully affect the perception of task completion relative to HS3. All button-based variants (HS2–HS4) were rated significantly better than HS1 across all task-completion dimensions ( $p \leq .024$  in all cases).

By contrast, the removal of the predefined animation in HS4 produced large and consistent advantages over HS3 on all remaining dimensions. HS4 was rated significantly more natural ( $p = .001$ ,  $d = 2.57$ ), less distracting ( $p = .002$ ,  $d = 1.29$ ), more controllable ( $p < .001$ ,  $d = 2.37$ ), and higher in hand embodiment ( $p < .001$ ,  $d = 1.83$ ) relative to HS3. Overall preference confirmed this pattern, with HS4 preferred over HS3 with a large effect size ( $p < .001$ ,  $d = 1.83$ ), and over all other variants ( $p < .001$  in all cases). The consistent ordering  $HS4 > HS3 > HS2 > HS1$  was significant across all pairwise comparisons for naturalness, distraction, control, embodiment, and overall preference.

### 3.3 Refining close social interaction animations in social virtual worlds

The handshake study addressed how a contact gesture can be initiated and performed through handheld controllers, treating the problem as one of interaction design. The handshake is a constrained gesture, brief and largely confined to the hands. Richer forms of close contact, such as hugs, involve the sustained and coordinated motion of two full bodies sharing the same space, where the challenge is no longer how to trigger the gesture but how to animate it convincingly. In these cases inverse kinematics alone often yields interpenetrations and unnatural poses [108], making behavioral fidelity, rather than interaction technique, the central problem. Avatar fidelity, encompassing both visual appearance and behavioral realism, plays a central role in supporting social presence and the overall quality of the experience [158].

From a visual standpoint, full-body avatar representations are consistently preferred by users in social VR settings [159]. Achieving behavioral fidelity for such representations, however, remains technically challenging. Hardware-based full-body tracking delivers the highest realism but requires expensive and complex setups [109]. *Inverse Kinematics* (IK) offers a more accessible alternative, though it tends to produce less natural movements, particularly during complex or physically demanding actions [108]. Machine learning-based approaches have recently emerged as a promising middle ground, capable of generating smoother and more realistic animations by exploiting large motion datasets [106].

Despite these advances, reproducing realistic close-contact interactions between avatars in virtual environments remains an open challenge [113]. While refinement methods for hand-object interactions have received some attention [160], [161], research on direct avatar-to-avatar interactions involving physical contact is comparatively scarce. The difficulty stems from the need to simultaneously model the movement adaptation of multiple avatars, each of which influences the other rather than simply responding to a fixed set of predefined actions [111]. This limitation is particularly visible in social VR, where close-contact gestures such as handshakes, high-fives, and hugs frequently appear unnatural, undermining immersion and social presence [111], [112].

To address this issue, a neural network-based framework was designed to model and refine avatar motion during social interactions in real time, producing more natural and synchronized movements between multiple avatars. The framework comprises two sequential components: a base IK module responsible for reconstructing avatar body poses from user input, followed by a *Multilayer Perceptron* (MLP) that processes the resulting motion data and applies learned refinements to improve synchronization and behavioral fidelity during social interactions. The network is trained on a dataset of avatar-to-avatar interactions, enabling it to learn and apply motion corrections dynamically at inference time.

Hugging interactions are used as the primary case study, as they represent one of the most physically complex and socially significant forms of close contact in social VR [162]. A dataset of authentic hugging interactions was collected for network training, with motion capture data used during training following the preprocessing pipeline described in [106], and IK-based animations used as input at inference time. The effectiveness of the framework was assessed through both objective metrics and a user study in which participants evaluated improvements in behavioral fidelity by observing two avatars hugging in VR.

### 3.3.1 Background

The need for avatar motion refinement arises primarily in situations involving physical-like interactions within virtual environments. When an avatar interacts with virtual objects [163], virtual agents [164], or remotely controlled avatars [111], its movements should adapt continuously to the ongoing physical action. In the absence of a physical counterpart in the real world, however, inconsistencies between the avatar’s pose and the virtual elements it interacts with are common, producing unnatural and visually unconvincing animations [112]. Addressing these artifacts through motion refinement is therefore essential for maintaining behavioral plausibility.

Early work in this direction by Schmidl and Lin [165] introduced a geometry-driven physics approach that computes penetration depth and contact points during avatar interactions with virtual elements, using this information to resolve overlaps and ensure realistic contact handling. Li et al. [160] subsequently proposed a grasp synthesis method based on shape matching to align virtual hand configurations with objects, though this approach was limited to single-hand interactions. Zhang et al. [166] extended this line of work by developing a bimanual interaction framework combining object geometry with motion data in a unified spatial representation; their neural network-based approach demonstrated that incorporating object movement significantly improves the accuracy of hand-object interactions. A further step was taken by Taheri et al. [167] with GRIP, a learning-based model for generating realistic full-body grasping motions that synchronizes hand and body movement before, during, and after interaction with a target object.

Beyond hand-object interactions, several studies have investigated full-body motion refinement more broadly. Choi et al. [168] proposed a motion optimization technique for mixed reality environments that refines both locomotion and object interactions through a penetration depth-based correction mechanism, though without incorporating learning-based components. Vogt et al. [164] developed an optimization framework for virtual agents that adapts and refines agent movements in real time using prerecorded motion capture data, enabling more natural human-agent interactions.

Research specifically addressing direct avatar-to-avatar interaction remains limited. Oh et al. [111] proposed an avatar animation technique for remote handshakes that refines hand contact to improve social telepresence, though without modeling fine-grained finger contact. Lee et al. [112] addressed this limitation by introducing a framework for multifinger contact interactions using a database of human hand gestures, improving realism for hand interactions but without extending to full-body animations.

Existing approaches have made meaningful progress in specific aspects of avatar motion refinement, yet close-contact interactions in social VR such as handshakes, high-fives, and hugs remain insufficiently addressed. A fundamental limitation of current methods is that they typically refine avatar motion in isolation, without accounting for the mutual influence that two interacting bodies exert on each other's movement.

The proposed framework addresses this gap by jointly modeling the movements of two interacting avatars, treating each as an active participant in the interaction rather than an independent object. By training a neural network on virtual social interaction data, the framework dynamically refines avatar motion to achieve behavioral fidelity, with the goal of enriching social presence and user experience in social VR.

### **3.3.2 Materials and methods**

This section presents the proposed refinement framework and the use case designed to evaluate its effectiveness.

#### **Framework**

The framework, illustrated in Figure 3.7, is designed to be integrated into a social VR platform accessible through consumer-grade VR kits comprising a headset and a pair of hand controllers. In the figure, User 1 and User 2 are engaged in a social interaction, while User 3 plays the role of an external observer, the role assumed by participants in the user study.

When users wear the VR headset, the position and rotation of their head and hands are tracked and processed by an IK system that reconstructs the avatar's full-body pose. A base IK module has been integrated as the first stage of the framework, serving as a functional foundation for the subsequent refinement stage. Although the base IK system introduces some artifacts and inaccuracies, its inclusion reduces the complexity of the learning problem by providing a plausible initial pose estimate, alleviating the need for the network to learn human motion from scratch. Running the IK system locally on each user's machine also limits the computational demand placed on the neural network module and ensures compatibility with a range of existing IK implementations.

The second stage of the framework is the neural network module, implemented as an [MLP](#). Neural networks are well suited to this task due to their ability to learn complex, non-linear motion patterns from data in ways that would be impractical to encode through hand-crafted rules. Running the network locally on each client, rather than on a centralized server, further reduces latency and improves the scalability of the system.

To limit unnecessary computational overhead, the neural network module is activated only when proximity between two or more avatars is detected. Specifically, the modules for User 1 and User 2 are activated when the distance between their hip joints falls below 1.2 meters (corresponding to the personal distance threshold) and both avatars fall within each other’s field of view. For User 3, the module activates when User 1 and User 2 are both detected as proximate and are within User 3’s field of view.

When active, the [IK](#)-based pose reconstruction is used to extract the position and rotation of each joint for User 1 and User 2. Per-joint velocities and angular velocities are then computed following the methodology of [106], yielding a four-parameter numerical pose representation for each frame. These representations are passed to the local neural network module, which has been trained on a dataset of close social interactions and applies learned motion refinements to improve behavioral fidelity. The refined poses are applied to the avatars of User 1 and User 2 and are rendered consistently for all observers within range, including User 3.

The framework operates independently of avatars’ global positions and orientations through a master-slave coordinate convention. The slave avatar’s root position and rotation are expressed in the local coordinate system of the master’s root, while all other joints are represented in local coordinates relative to their parent joints. During training, the master’s global root position is set to zero to ensure that the network learns motion patterns independently of absolute world positioning. At inference time, the master’s root position and rotation are retrieved from the base [IK](#) system output to maintain consistency between training and inference.

For User 3 acting as an observer, the master-slave assignment is determined dynamically: the avatar closest to the observer is designated as the master, and the other as the slave. For User 1 and User 2, the convention assigns the local user as the master and the remote user as the slave, ensuring that each participant’s module produces a refinement consistent with their own perspective.

### 3.3.3 Experiment

To evaluate the proposed refinement framework, hugging was selected as the primary use case. The following subsections describe the interaction type considered, the technologies employed, and the settings adopted for implementation and assessment.



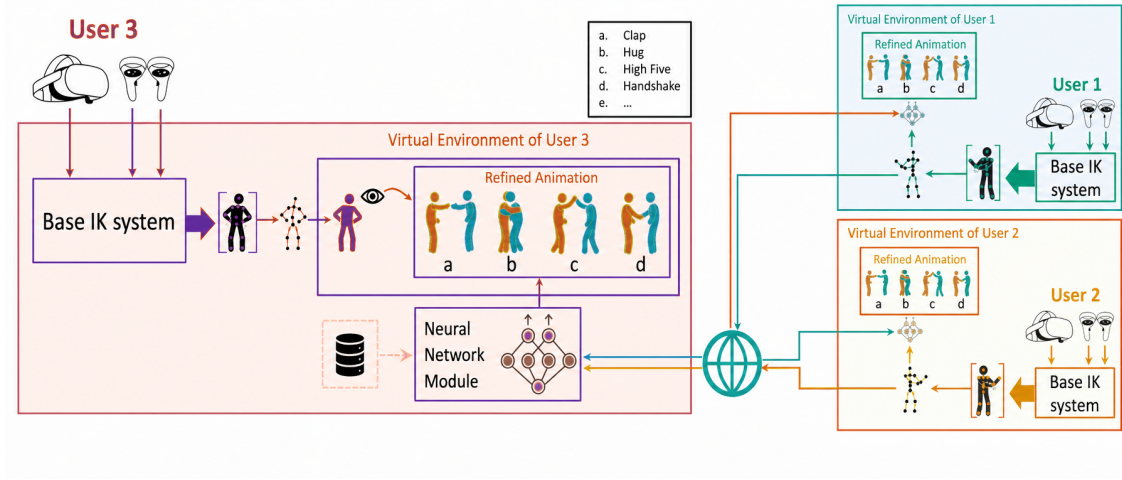


Figure 3.7: Overview of the proposed refinement framework. User 1 and User 2 are engaged in a social interaction, while User 3 acts as an external observer. Each user runs a local neural network module that processes avatar pose data and refines the animation output of the base IK system.

### Hugging interaction

Hugging represents one of the most physically complex forms of close social contact, involving a highly coordinated interplay between the limbs and bodies of two participants. Two hug types commonly observed in real-life interactions were considered: the criss-cross hug and the neck-waist hug [169], [170], illustrated in Figure 3.8. To introduce additional variability into the dataset, recordings were collected with different head crossing sides and arm positioning variations, distinguishing whether each arm passed over or under the other person’s body.

### Base IK system

Final IK was selected as the base IK system. Final IK is a full-body character animation solution featuring a VR-oriented IK solver that supports natural movement and real-time adaptation to head and hand tracking input, combining IK for the upper body with animation blending for locomotion. It has been widely adopted in studies involving virtual avatars [41], [171].

### Neural network module

The MLP was trained using stochastic gradient descent with a mean squared error (MSE) loss function. A systematic exploration of hyperparameter configurations was conducted, evaluating different hidden layer sizes alongside other training parameters including learning rate, batch size, and momentum, in order to identify



the optimal settings for the network.

## Dataset

Motion capture data were collected using the OptiTrack system<sup>9</sup>, following procedures established in prior work [109]. The built-in retargeting function from the OptiTrack streaming package for Unity was used to map the recorded movements onto the avatar skeletons, which were obtained from Mixamo.

The dataset comprises recordings of the two hug types described above, captured with different arm position variations, yielding 45 usable recordings with durations ranging from 10 to 30 seconds. To improve the robustness of the network, the dataset was augmented by reversing the master-slave assignment for each recording, doubling the number of viable samples to 90 and ensuring that the network can reconstruct the interaction regardless of which avatar assumes the master role.

Two simplifications were applied to reduce data collection complexity. First, finger tracking was omitted, as fine-grained finger motion contributes minimally to the overall perception of a hugging interaction. Second, only avatars of matching height were considered, as introducing height variability would have substantially increased the number of recordings required for each interaction type. Data collection involved two participants of approximately equal height, around 1.80 meters. Joint rotations and angular velocities were converted to a 6D representation, which has been shown to be effective for training motion-related neural networks [106], [172], [173]. The resulting dataset was partitioned into training (60%), validation (20%), and test (20%) subsets.

## Evaluation

This section presents the objective and subjective evaluations conducted to assess the proposed refinement framework. Specifically, pose error relative to the ground truth was computed for both the base IK system and the proposed framework, and a user study was conducted to determine whether the framework produces perceptible improvements in behavioral fidelity from a user perspective.

## User study

The user study followed a within-subjects design and involved 16 participants (12 male, 4 female) aged between 20 and 31 years, recruited from among the students and staff of the authors' institution as prospective social VR users. The

---

<sup>9</sup>OptiTrack: <https://www.optitrack.com/>



Figure 3.8: The two hug types considered in the study: (a) criss-cross, and (b) neck-waist.

VR experience was developed in Unity 2022.3 and deployed on a Meta Quest Pro headset<sup>10</sup>.

The study aimed to investigate the perceptual effects of the proposed refinement framework on the behavioral fidelity of hugging avatars, as assessed from the perspective of an external observer. Two animation conditions were compared: animations generated by the base IK system (Final IK) and animations produced by the proposed refinement framework built on top of Final IK. The two animation conditions in the four types of hugs performed are visible in Figure 3.9.

Upon arrival, participants completed a demographic questionnaire covering age, gender, VR experience, and familiarity with social VR platforms. The sample was largely inexperienced with immersive technologies: approximately 50% rated their VR experience at 3 or below on a seven-point scale, and around 94% reported little to no prior engagement with social VR environments.

Participants were then guided into a virtual environment designed to resemble a typical social VR setting<sup>11</sup>, where they observed four types of pre-recorded hug animations, each lasting a maximum of ten seconds: criss-cross right, criss-cross left, neck-waist high, and neck-waist low. Pre-recorded animations were used to ensure reproducibility and consistency across participants. Avatars with abstract visual representations were deliberately chosen to focus participants' attention on behavioral rather than visual fidelity [140].

Each animation was presented ten times: eight times from an external observer perspective with 45-degree incremental rotations to cover multiple viewpoints, and

<sup>10</sup>Meta Quest Pro: <https://www.meta.com/it/quest/quest-pro/>

<sup>11</sup>Sci-Fi Room: <https://tinyurl.com/shifiRoom>

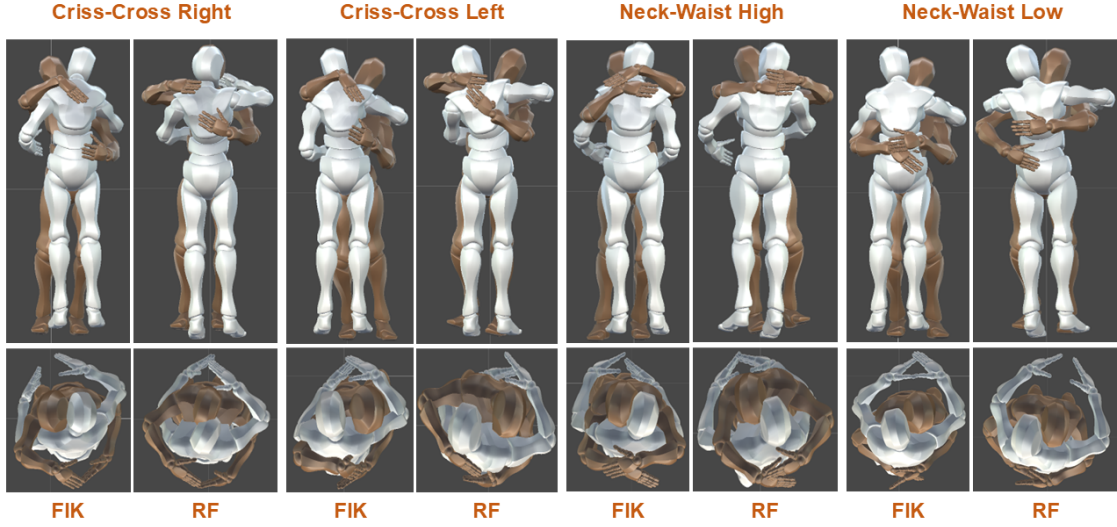


Figure 3.9: The two animation conditions across the four types of hugs performed in a virtual environment.

twice from the first-person perspective of one of the huggers, once from the master and once from the slave viewpoint. The participant positions and viewpoints are illustrated in Figure 3.10. A Latin square design was used to counterbalance both the order of hug types and the sequence of modality presentation.

After observing an animation in the first modality, participants completed a questionnaire assessing its naturalness and realism. They then viewed the same animation in the second modality and responded to the same set of questions, with the option to revise their earlier responses, enabling a direct within-subjects comparison. A video illustrating all hug types in both modalities as displayed in the virtual environment is available as supplemental material<sup>12</sup>. Participants were invited to provide open-ended feedback at the end of the session.

## Metrics

### Objective measures

Following the methodology of prior work [106], the ability of the proposed framework to refine hug animations was assessed using the **Mean Per Joint Rotation Error (MPJRE)**, which quantifies the average rotation error across all joints in degrees, comparing the output of the framework against the ground truth motion capture data.

<sup>12</sup>Video of hug animations: <https://tinyurl.com/2ja4zp7z>

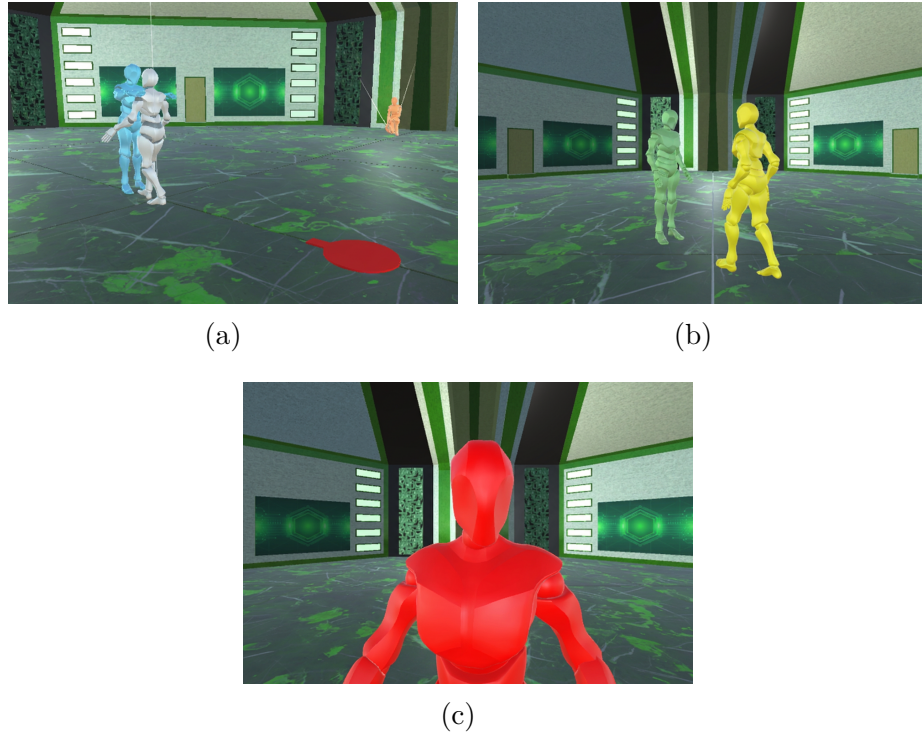


Figure 3.10: Participant positions and viewpoints considered in the user study: (a) position of the participant as an external observer (red marker), (b) the corresponding external observer viewpoint, and (c) the front-facing viewpoint of one of the users engaged in the hug.

### **Subjective measures**

Subjective evaluation was conducted through a custom questionnaire assessing the behavioral fidelity of the animations. Items were designed specifically for the hugging use case and addressed the following aspects: whether the beginning and end of the hug were well coordinated, whether the interaction appeared visually harmonious, whether the postures and movements of the avatars were well aligned, and whether the arm and hand movements adapted appropriately to the shape of the other avatar's body. All items were rated on a seven-point scale. A question from [174] on perceived avatar mesh interpenetration was also included. Overall realism was assessed using the anthropomorphism subscale of the Godspeed Questionnaire Series [175]. The full questionnaire is available as supplemental material<sup>13</sup>.

---

<sup>13</sup>Questionnaire: <https://tinyurl.com/54nbsaxu>

### 3.3.4 Results and discussion

The optimal neural network configuration identified through the hyperparameter search comprised a hidden layer of 1024 units, a momentum of 0.9, a learning rate of 0.1, a batch size of 32, 2000 training epochs, and a downsampling factor of 10, which was used to sample a subset of frames from each recording to increase variability and reduce overfitting.

**Objective results.** The proposed refinement framework achieved an [MPJRE](#) of 4.86 degrees, compared to 17.83 degrees for the base [IK](#) system alone, representing an improvement of approximately 13 degrees in the accuracy of joint rotation reconstruction relative to the ground truth poses.

**Subjective results.** Data normality was confirmed using the Shapiro-Wilk test, and statistical significance was assessed using a paired t-test with a significance threshold of  $p < .05$ . Regarding the custom questionnaire items (Figure [3.11](#)), no significant differences were found between the two conditions for hug coordination, visual harmony, or avatar posture alignment. This outcome suggests that the proposed refinement framework does not compromise the fluidity and coherence of the animations already produced by the base [IK](#) system. A significant difference in favor of the refinement framework was found for the item assessing how well arm and hand movements adapted to the shape of the other avatar’s body (3.84 vs 4.22,  $p = .036$ ), a result attributable to the network’s training on high-fidelity hugging data, which enables it to produce body-conforming motion that the base [IK](#) system cannot achieve. A further significant difference was observed for perceived interpenetration, with participants reporting fewer instances of unnatural visual collisions in the refinement framework condition (3.27 vs 3.56,  $p = .028$ ), consistent with the network’s ability to jointly adapt the movements of both avatars to avoid mesh overlap.

The Godspeed anthropomorphism subscale results (Figure [3.12](#)) showed no significant differences for most items, indicating that the refinement framework does not adversely affect the overall naturalness of the motion produced by the base [IK](#) system. A notable exception was the machinelike-humanlike item, for which animations from the base [IK](#) system were rated as more humanlike than those from the refinement framework (4.52 vs 4.03,  $p = .032$ ). This result suggests that while the refinement framework improves body adaptation, the refined movements may lack sufficient smoothness, producing a slightly more mechanical appearance that partially offsets the gains in behavioral fidelity.

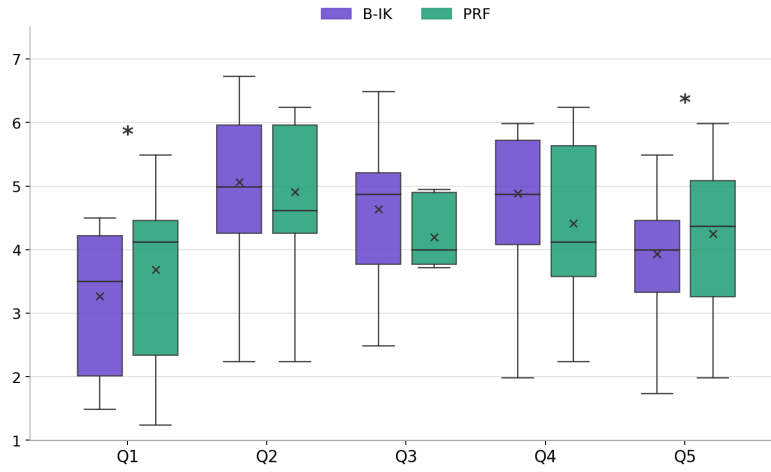


Figure 3.11: Distribution of scores for the base IK system (B-IK) and the proposed refinement framework (PRF) across custom questionnaire items. Statistically significant results are marked with an asterisk. Q1: no overlaps between avatars; Q2: hug beginning and end well coordinated; Q3: visually harmonious hug; Q4: well-aligned postures and movements; Q5: arm and hand movements adapted to the other body's shape.

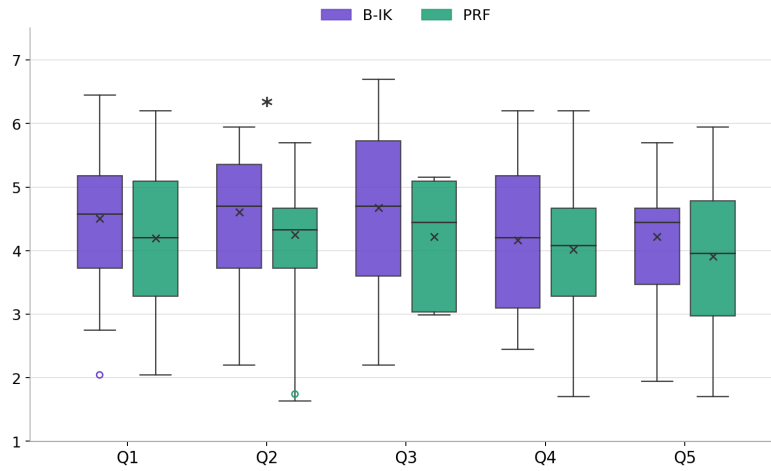


Figure 3.12: Distribution of scores for the base IK system (B-IK) and the proposed refinement framework (PRF) across items of the Godspeed anthropomorphism sub-scale. Statistically significant results are marked with an asterisk. Q1: Fake / Natural; Q2: Machinelike / Humanlike; Q3: Unconscious / Conscious; Q4: Artificial / Lifelike; Q5: Moving rigidly / Moving elegantly.



### 3.4 Augmenting virtual interpersonal communication through artificial facial mimicry

The studies presented so far shared a single objective: reproducing natural social behavior as faithfully as the technology allows, whether in facial expression, gesture, or close-contact motion. Social VR, however, is not confined to replicating face-to-face interaction. The same computational mediation that imposes fidelity constraints also makes it possible to modify and amplify non-verbal signals in ways unaided human behavior cannot achieve [114], [176]. The question therefore shifts from how closely an avatar can match natural behavior to whether the deliberate augmentation of that behavior can enrich social interaction. Such augmentations can compensate for sensory gaps, enrich non-verbal communication, and support users with social communication difficulties [114], [176]. Among these augmentations, **Artificial Mimicry (AM)** occupies a particularly prominent role. Mimicry, defined as the unconscious imitation of verbal and non-verbal cues, is a well-established mechanism for strengthening social bonds, building rapport, and facilitating communication [115], [116], [117], [177]. In social VR, AM has been explored across a range of behavioral channels, including body movements, eye contact, nodding, and facial expressions [93], [114], [178], [179].

When applied specifically to facial expressions, AM is referred to as **AFM**. Drawing on the same mechanisms observed in face-to-face interaction [180], AFM enables digital avatars to dynamically adjust their facial expressions in response to the emotional state of their conversational partner. Prior work has demonstrated its potential in human-agent interactions, where it has been shown to enhance social presence and user engagement [119], [181], [182]. Its application to avatar-mediated human-to-human interaction, however, remains comparatively underexplored [114].

AFM can be implemented at different levels of sophistication. The simplest approach, *Emotion Mirroring*, directly projects one user’s expression onto the peer’s avatar regardless of emotional context (Figure 3.13a), which can produce forced or incoherent results and may interfere with speech synchronization. More advanced methods based on automated emotion detection allow mimicry to be applied adaptively and dynamically, decoupling avatar animation from the mimicry augmentation. These include *Emotion Augmentation*, which amplifies the intensity of a detected emotion on the peer’s avatar (Figure 3.13b); *Emotion Camouflage*, which replaces a conflicting expression with one that aligns with the current emotional context (Figure 3.13c); and *Emotion Translation*, which remaps expressions into forms more readily interpretable by the conversational partner (Figure 3.13d) [93], [111].

Despite its potential, AFM has to date been applied almost exclusively in non-immersive VR settings, where avatar control relies on indirect input methods such as mouse and keyboard [179], [183]. The growing availability of immersive VR

headsets with integrated FT capabilities [101], [184], [185] makes AFM increasingly viable for direct and natural avatar control. Nevertheless, current tracking systems continue to struggle with the full range of facial expressions, particularly those involving the upper face [41], and AFM can play a complementary role in compensating for these limitations by enhancing expressiveness where tracking falls short.

This last investigation explores the potential of AFM to improve social interaction quality in immersive social VR. A machine learning-based AM system was developed to manage AFM alongside the mimicry of nodding and eye contact. The system dynamically adjusts the peer’s avatar facial expressions by combining FT data with the current user’s emotional state, detected through a classifier trained to recognize happiness, sadness, and neutrality. When emotional discrepancies arise between the two users, the system applies emotion camouflage to align their expressions at matching intensity. During natural mimicry, if the mimicked emotion is detected at a lower intensity on the peer, the system applies emotion augmentation to bring it in line with the user’s level; otherwise, the tracking data is relayed without modification.

### **3.4.1 Background**

Mimicry has been studied extensively across real-life interactions, human-agent communication, and avatar-mediated exchanges in virtual environments. This section reviews the relevant literature, covering the role of mimicry in face-to-face social dynamics, its use by virtual agents to enhance engagement, and its application in avatar-mediated human-to-human interaction, identifying the gaps in social VR that motivate this investigation.

#### **Mimicry in real-world interactions**

A substantial body of research has established mimicry as a powerful mechanism for fostering empathy, building rapport, and cultivating interpersonal trust [186], [187], [188], [189], [190]. Chartrand and Bargh [186] introduced the concept of the chameleon effect, demonstrating that being mimicked leads to smoother social interactions, increased affiliation, and higher likability ratings. Maddux et al. [191] extended these findings to negotiation contexts, showing that mimicry enhances trust and improves outcomes, while Gueguen [192] found that it promotes disclosure and honesty in interpersonal exchanges.

Methodologically, mimicry research has relied on controlled manipulation through photographs, video stimuli, or live interactions. The confederate paradigm, in which trained individuals systematically mirror participants, has been particularly instrumental in isolating its effects [193], [194], though results with respect to liking have not always been consistent, despite robust evidence for other positive social outcomes [195], [196]. Dyadic interaction paradigms enable more ecologically valid



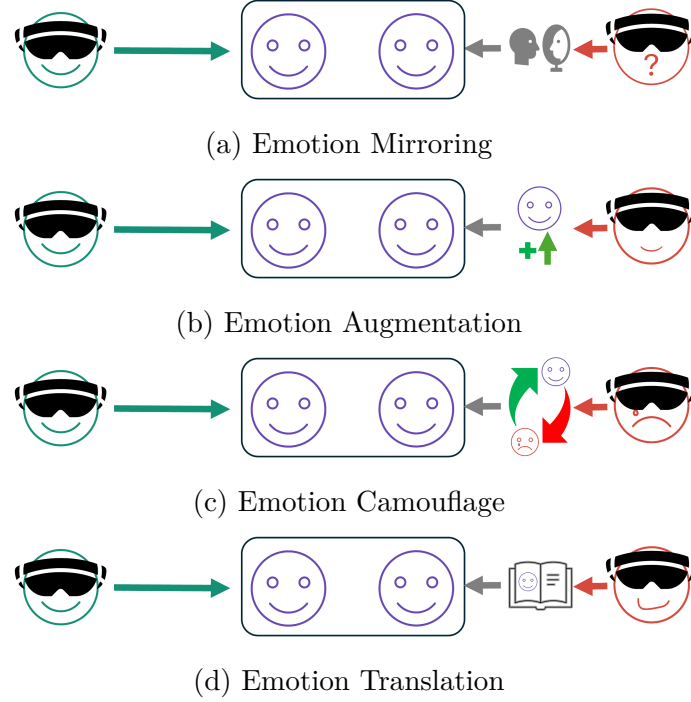


Figure 3.13: AFM methods in immersive social VR. User 1 (blue, left) observes User 2 (orange, right) reaching the same emotional state under four different conditions: (a) the user’s own expression is projected onto the peer; (b) a lower-intensity emotion on the peer is amplified to match the user’s level; (c) a conflicting emotion on the peer is replaced with a matching one; (d) the peer’s expression is remapped into a form more easily understood by the user.

observations of natural affiliation processes, though at the cost of reduced experimental control [197].

These findings position mimicry as a promising mechanism for enhancing persuasion, trust, and social affiliation, with clear implications for its application in shared virtual environments [198].

### Mimicry in human-agent interactions

The use of mimicry in human-agent interactions has emerged as an effective strategy for improving social engagement and perceived agent quality [181], [182], [183], [199]. Bailenson and Yee [198] provided an early and influential demonstration of this effect in immersive VR, showing that virtual agents that mirrored participants’ head movements with a four-second delay were perceived as more persuasive and likable, even when participants were unaware of the mimicry, and even in the absence of verbal turn-taking.

Verberne et al. [200] investigated the effects of head movement mimicry on interpersonal trust, likability, and self-other overlap. Their findings revealed that mimicry enhanced calculative trust in a context where perceived competence was central, but had no significant effect in a setting where relational trust based on perceived intentions was more relevant. The authors suggested that more naturalistic mimicry algorithms and the incorporation of facial expression mimicry could address the limitations of rigid, delayed approaches.

Hoegen et al. [201] examined the role of facial mimicry and counter-mimicry in human-agent interaction, finding that agents that mirrored participants' smiles and frowns promoted greater rapport and cooperation, while counter-mimicry produced no measurable effects. Their results also showed that mimicry encouraged participants to smile more frequently, underscoring the importance of emotionally aligned responses in social dynamics.

Doo Sung et al. [202] explored the combined effects of eye contact, facial mimicry, and head nodding on virtual counseling effectiveness in a non-immersive VR setting. Participants interacting with a compassionate virtual counselor employing these non-verbal cues reported reduced anger and greater perceived empathy compared to those assigned to a neutral counselor. The study highlighted the potential of face tracking and emotion recognition technologies to enhance non-verbal communication in human-agent contexts.

### **Mimicry in avatar-mediated human-to-human interactions**

The application of AM to avatar-mediated human-to-human interaction has been explored in a number of studies aiming to replicate the benefits observed in human-agent settings. Roth et al. [93] introduced the Mimicry Injector, a system that integrates artificial non-verbal mimicry of body movements into VR interactions. A negotiation study found that, while participants largely failed to detect the manipulation, AM had no significant effect on communication or negotiation outcomes, a finding the authors attributed to the comparatively limited social impact of body mimicry relative to facial mimicry. They also identified the need for more sophisticated AM approaches capable of capturing subtle non-verbal cues and for well-defined activation triggers, while raising ethical concerns regarding covert behavioral modification in virtual environments.

Earlier studies on AFM in human-to-human contexts were conducted in non-immersive settings. Suda et al. [203] examined partial AFM during avatar-mediated negotiations, mimicking eye and eyelid movements. Results showed improved cooperative outcomes, though the partial nature of the implementation underscored the need for systems capable of replicating a broader range of facial expressions while maintaining contextual appropriateness. Suzuki et al. [204] proposed FaceShare, a videophone system that automatically mirrors smiles using image processing. While the pseudo-smile mirroring enhanced closeness and conversation quality, the

system was limited to positive expressions, highlighting the need for more adaptive and context-aware mechanisms covering the full emotional spectrum.

Nodding mimicry has been studied by Wakabashaky et al. [179], who showed that an avatar mimicking customer nodding in a non-immersive virtual environment enhanced rapport, particularly among female participants. The study also identified the influence of contextual factors, such as the simulated retail environment, and called for more controlled settings to enable cleaner evaluations.

Roth et al. [205] proposed the Hybrid Avatar-Agent Technology, a framework that analyzes AM at the dyad level by accounting for bidirectional adaptation and synchronization processes. Hybrid Avatar-Agent Technology was capable of interpreting social signals and mapping them to the peer’s reactions in real time. Subsequent work using Hybrid Avatar-Agent Technology investigated hybrid and augmented gaze in multi-user VR environments [8], demonstrating its capacity to increase social presence [92]. Building on these foundations, Roth et al. [114] introduced injectX, a system integrating the mimicry of body movements, gaze, and facial expressions within immersive Virtual Environment through a context-aware, dynamic decision-making process. While a preliminary study focused on body movement blending techniques, facial mimicry was identified as a primary direction for future investigation.

## Considerations

The reviewed literature establishes mimicry, and facial mimicry in particular, as a promising mechanism for enhancing social interaction quality in shared virtual environments. While the benefits of AM in human-agent settings are well-documented, evidence for its effectiveness in avatar-mediated human-to-human interaction, especially in immersive contexts, remains limited and methodologically incomplete. Prior work in this direction, including [114], has relied on simplified mimicry implementations or lacked rigorous experimental validation. This investigation addresses the above gaps by introducing a complete AM system supporting the mimicry of nodding, eye contact, and facial expressions in immersive social VR, and by evaluating its impact on interpersonal communication through a controlled user study.

Two hypotheses are formulated and tested:

- **H1:** Experiencing social interaction with an avatar featuring AFM enhances the quality of avatar-mediated interpersonal communication in immersive social VR.
- **H2:** Using AFM in immersive social VR does not negatively affect the perception of the behavioral naturalness of avatars.

To test these hypotheses, an immersive social VR application was developed as a test-bench alongside the AM system, enabling two users to converse through their

avatars. A study with 40 participants organized into 20 dyads compared an experimental group in which **AFM** was active against a control group relying on natural mimicry through facial tracking, with both groups experiencing **AM** of nodding and eye contact. Results showed that **AFM** significantly enhanced avatar-mediated interpersonal communication, improving closeness, social attraction, social presence, and trust, while preserving avatar behavioral naturalness and conversational clarity.

### 3.4.2 Materials and methods

This section presents the architecture of the proposed **AM** system and the immersive social **VR** application developed to serve as an experimental test-bench.

#### Artificial mimicry system

The proposed **AM** system extends prior work [93], [114] and is designed for seamless integration into Unity-based multi-user **VR** applications. It relies on **VR** headsets with facial tracking capabilities to capture users' expressions and apply context-aware mimicry augmentations dynamically.

The system architecture, illustrated in Figure 3.14, comprises four components: an *Emotion Classifier*, an *AFM Management* module, a *Nodding Management* module, and an *Eye Contact Management* module. A demonstration video is available as supplemental material<sup>14</sup>.

**Emotion classifier** Rather than directly mirroring facial expressions, the **AM** system analyzes facial data to infer the user's emotional state, which then guides the application of **AFM**.

Emotion classification relies on the avatar's facial blendshapes, which are driven in real time by the facial tracking **SDK** of the **VR** headset. Several machine learning models available in the open-source *scikit-learn* library<sup>15</sup> were evaluated for this multi-class classification task, including Random Forest, Gradient Boosting, Support Vector Machine, and Logistic Regression. An initial training phase on a trial blendshape dataset was conducted for each model, followed by stratified 5-fold cross-validation and a grid search to compare performance. Random Forest consistently achieved the best results, with accuracy ranging from approximately 78% to 98%. To minimize computational overhead on a system running a **VR** application simultaneously, the model was configured with a maximum tree depth of 10 and

---

<sup>14</sup>Sample video of **AM** system operation: <https://tinyurl.com/37mzxv9m>

<sup>15</sup>Scikit-learn: <https://scikit-learn.org/stable/>

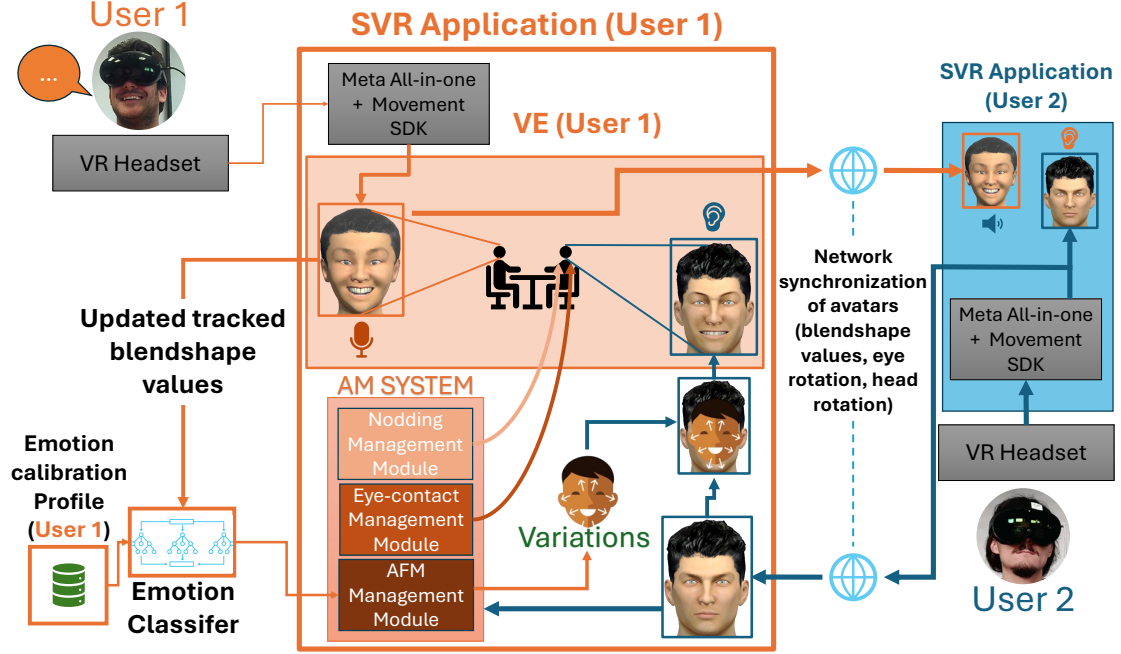


Figure 3.14: Architecture of the [AM](#) system, illustrated assuming User 1 is speaking. Speech input detected on the microphone activates [AFM](#), making its effects visible on User 2’s avatar. The same logic applies symmetrically for User 2; [AFM](#) is disabled on the listening user’s end.

50 estimators. Despite these constraints, all configurations exceeded 90% accuracy, making Random Forest a suitable choice for real-time classification without impacting frame rate.

Before the [AM](#) system can be used, the Emotion Classifier requires a calibration procedure conducted in a dedicated social [VR](#) application. This procedure trains the classifier to recognize the individual user’s facial expressions for specific emotions and comprises three sequential phases carried out in separate [VR](#) scenes.

In the *Acclimatization* phase, the user observes their avatar in a virtual mirror to promote embodiment and familiarize themselves with the virtual environment. This phase is conducted once and lasts two minutes [182].

In the *Emotion Stimulation* phase, a combination of video clips, environmental colors, and musical elements is used to evoke each of Ekman’s six basic emotions alongside a neutral state [130], [206], [207]. Media playback and customizable environment colors [208] are used to align the scene with the target emotion. Visual stimuli are sourced from established databases, including IAPS [209] for images and the Standardized Database of Chinese Emotional Film Clips [210] for videos.

In the *Emotion Recording* phase, the user is placed in a virtual environment

configured with colors and sounds associated with the previously stimulated emotion, and is asked to verbally reflect on the emotional impact of the viewed clip for two minutes. Throughout this period, the avatar’s facial blendshape values are recorded frame by frame via the headset’s facial tracking system, producing a Calibration Recording file. The user then returns to the Emotion Stimulation phase to repeat the process for the same or a different emotion. The collection of Calibration Recording files produced across a session, potentially covering multiple emotions, constitutes the user’s Emotion Calibration Profile.

Once calibration is complete, the [AFM](#) Management module can operate for that user by loading their Emotion Calibration Profile, enabling real-time emotion recognition and adaptive mimicry behavior.

**Artificial facial mimicry management** The [AFM](#) Management module controls the avatar’s facial expressions by applying targeted modifications based on the emotional states identified by the classifier. Although the module is designed to support any emotion the classifier has been trained on, the current implementation is limited to happiness, sadness, and neutrality, a scope chosen to simplify system management for the purposes of the study.

Emotion recognition is governed by the concept of reaction time [211]. Happiness is detected when a corresponding facial expression is sustained for at least 0.5 seconds [212]. For sadness and neutrality, as well as for transitions from emotional to neutral states, a reaction time of 0.6 seconds was adopted, based on average values reported in prior studies [211], [213].

Since facial mimicry is more perceptible in listeners than in speakers [214], [AFM](#) is tied to speech input to promote emotional connection during empathetic listening [215]. When speech is detected via the microphone, [AFM](#) is activated for the speaker’s peer. After 0.5 seconds of silence, the user is classified as a listener and [AFM](#) is deactivated. Transitions between activation states are handled through the same interpolation mechanism used for all other mode transitions, described below.

Emotion detection via the classifier operates exclusively when [AFM](#) is active, applied only to the local user’s avatar blendshapes. This design avoids the need to transmit detection results over the network or share Emotion Calibration Profiles between users, keeping the system self-contained and enabling [AFM](#) to function as a personal setting comparable to, for example, a colorblind display mode. Its operation depends only on the peer’s avatar having compatible blendshapes driven by facial tracking.

Figure 3.15 illustrates the module’s operational logic for User 1 as speaker and User 2 as listener, covering the three supported emotional states.

The module operates in three modes: [AFM Off](#), [Augmentation](#), and [Camouflage](#), with the latter two collectively referred to as [AFM On](#). Listener avatar modifications are applied continuously through per-frame blendshape updates, using Unity’s

linear interpolation (Lerp<sup>16</sup>) with a 50% adjustment factor to produce an exponential smoothing effect. At 90 fps, this configuration produces perceptually natural transitions, covering the full range from 0 to 100 in approximately 0.148 seconds.

At initialization, the module starts in *AFM Off* mode with User 1 in a neutral state. When the classifier processes the speaker’s blendshape data and detects an emotion for at least the required reaction time (Figure 3.15a), that emotion is recorded as User 1’s current state. For non-neutral emotions, the module compares User 1’s and User 2’s blendshapes (Figure 3.15b) using the *Action Units* (AUs) defined by the *Facial Action Coding System* (FACS) [216]. This approach enables the module to move beyond simple one-to-one expression mapping, focusing on universally recognizable emotional signals while filtering out undesirable movements such as those associated with speech [203]. For example, when happiness is detected for User 1, emotional alignment for User 2 requires AUs 6 and 12 (associated with happiness) to be greater than zero, while AUs 1, 4, and 15 (associated with sadness) should approach zero [217].

If misalignment is detected, the module transitions to *Camouflage* mode (Figure 3.15c). If alignment is confirmed, the module compares the AU blendshape values associated with User 1’s emotion between the two users (Figure 3.15d). If User 2’s values are higher, natural facial mimicry at greater intensity than User 1’s is already occurring, and the module remains in *AFM Off* mode (Figure 3.15e) with target values set to zero, preventing excessive emotional mirroring that could undermine trust and rapport [218]. Otherwise, the module transitions to *Augmentation* mode (Figure 3.15f). In both *Augmentation* and *Camouflage* modes, target values for the listener’s AU blendshapes associated with the detected emotion are computed as the difference between the speaker’s and listener’s values. In *Camouflage* mode, blendshape target values for AUs unrelated to the detected emotion are additionally set to zero (Figure 3.15g), suppressing conflicting expressions and promoting smoother emotional alignment [219]. If User 1 is detected as neutral, the module returns to *AFM Off* mode.

It should be noted that the system does not implement emotion translation. While FACS AUs are used to interpret the peer’s expressions and regulate AFM application, true translation would require detecting the peer’s emotion via their Emotion Calibration Profile and remapping it to a target expression. This is not feasible in the current implementation, as the classifier operates only on the local user’s data and detection results are not exchanged between users.

**Nodding and eye contact management** Eye contact and nodding are among the most fundamental non-verbal cues in social interaction: the former signals attention and strengthens interpersonal connection, while the latter indicates active

---

<sup>16</sup>Unity Scripting API - Mathf.Lerp: <http://tiny.cc/eb34001>



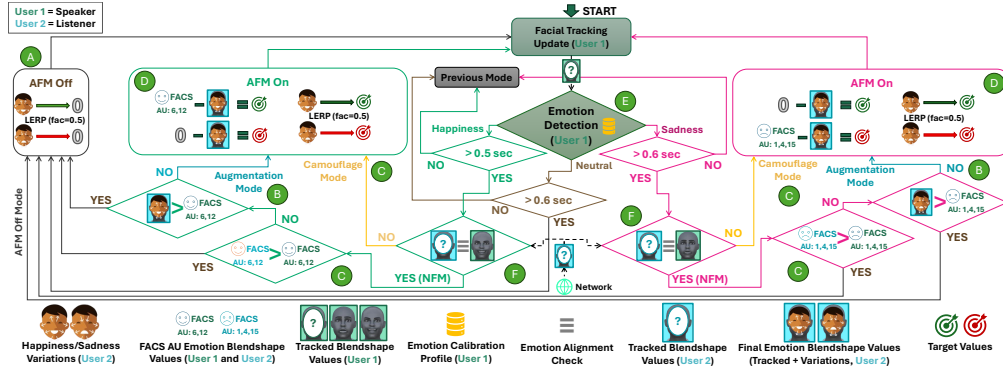


Figure 3.15: Operational logic of the AFM Management module for User 1 (speaker) and User 2 (listener) across three emotional states: happiness, sadness, and neutrality. User 1’s blendshapes are directly tracked via facial tracking; User 2’s blendshapes are synchronized over the network. Labels A to G correspond to the states described in Section 3.4.2.

engagement in the conversation [220]. Drawing on these principles and on prior work [179], [202], the AM system incorporates artificial mimicry for both behaviors.

Eye contact is managed through raycasts projected from the user’s avatar eyes; when these intersect with a sphere collider placed at the peer’s avatar head, mutual gaze is detected. If eye contact is not reciprocated, the peer’s avatar eye rotations are adjusted via interpolation, completing the transition within one second to simulate natural gaze redirection. The mimicry is sustained for up to nine seconds, consistent with the optimal duration for eye contact in conversational settings [202].

Nodding detection relies on the open-source *HeadNodYesSensor.cs* Unity script<sup>17</sup>. When a nod is identified, the peer’s avatar mirrors it with head rotation angles randomly varied between 15 and 20 degrees, introducing natural variability and avoiding the perception of mechanical imitation.

Together, the Eye Contact, Nodding, and AFM Management modules form the complete AM system, which can be integrated into any multi-user immersive social VR application supporting facial tracking and standard avatar blendshapes.

## Social virtual reality application

The AM system was designed to investigate the benefits of AFM in avatar-mediated human-to-human interaction within immersive social VR. To this end, a dedicated immersive VR application was developed as an experimental test-bench for deploying and evaluating the system.

<sup>17</sup>HeadNodYesSensor: <https://gist.github.com/kurtdekker/0f5f353911d7a064e20868e4e5f4dba4>

The application was implemented in Unity 2021.3, using the Meta XR All-in-One SDK<sup>18</sup> for immersive VR functionalities including eye and facial tracking, the Mirror networking library for multi-user synchronization, and the Dissonance Voice Chat asset<sup>19</sup> for real-time voice communication. Particular attention was paid to avatar design, environment simplicity, and the seamless integration of the AM system.

Full-body humanoid avatars were created using Ready Player Me, balancing visual fidelity with a low polygon count and providing facial blendshapes compatible with the Meta XR SDK. Two avatars, one male and one female, were designed in a matching visual style with black and white textures, minimizing distinguishing features to reduce biases related to ingroup favoritism and outgroup discrimination, documented both in real-life settings [221] and in social VR platforms [222], [223]. Users are placed in a virtual room seated across from each other, with avatar control restricted to head, face, and eye movements managed through the Meta Movement SDK<sup>20</sup>. Facial expressions are driven through 52 blendshapes automatically updated by the SDK from facial tracking input.

### Virtual environment

The virtual environment was designed to be minimal yet visually coherent, creating a relaxed social atmosphere conducive to natural conversation [7], [78]. A small set of environmental elements, including a door, a window, and a table, was included to provide contextual grounding while minimizing potential distractions [179]. Users were seated in armchairs facing each other, each accompanied by a virtual notebook used to display the conversation-guiding questions described in Section 3.4.3. To prevent visual clipping artifacts, the local user's own avatar mesh was disabled, while the peer's avatar was rendered in an idle seated animation holding the notebook. An acclimatization scene was also included to help users adjust to the experience prior to the main session, using the same environmental configuration adopted during calibration.

### 3.4.3 Experiment

The user study was designed to evaluate the effects of facial mimicry in multi-user social VR, employing a between-dyads design in which participant pairs were randomly assigned to one of two equal groups: an experimental group in which AFM was active, and a control group relying on natural facial mimicry through facial tracking alone. Both groups included eye contact and nodding mimicry as

---

<sup>18</sup>Meta XR All-in-One: <https://tinyurl.com/bpaejask>

<sup>19</sup>Dissonance Voice Chat for Unity: <http://tiny.cc/cndmzz>

<sup>20</sup>Meta Movement SDK for Unity: <http://tiny.cc/aynmzz>

consistent non-verbal cues, isolating the contribution of facial mimicry as the independent variable.

Each dyad comprised two participants engaged in a natural conversation. The use of a confederate in place of one participant was deliberately avoided, as repeated participation by a confederate could introduce variability in sociability and affect the ecological validity of the results [195], [196]. In the experimental group, AFM was applied to the listener’s avatar during the speaker’s turn and automatically switched when speaking roles reversed. The Fast Friends Protocol (FFP) [224], [225] was adopted to structure the conversation, a well-established method in digital sociality research for promoting closeness between strangers through three sets of 12 progressively personal questions, designed for a 45-minute session. A reduced version of the protocol was implemented to better suit the experimental context.

The hardware setup comprised two Meta Quest Pro VR headsets with integrated facial and eye tracking, connected via Meta Quest Link to two Intel Core i7-12700 PCs (16 GB RAM, GeForce RTX 3070 Ti) on the same wired local network. The two workstations were placed in separate rooms to ensure acoustic privacy throughout the session.

**Emotion calibration profile creation** Target emotions for the calibration procedure were selected based on two criteria: affiliative relevance and ease of elicitation. Happiness and sadness were identified as the most suitable candidates, as both are readily stimulated and tend to produce sustained facial responses.

For the Emotion Stimulation phase, the emotion stimuli were drawn from the AVDOS dataset [226], which comprises 60 video clips of 60 seconds each categorized as positive, negative, or neutral. The ten clips with the highest positive valence and the ten with the lowest negative valence were shortlisted, and six were selected for use in the study. Each participant viewed two clips associated with happiness, two with sadness, and two with neutrality.

For the Emotion Recording phase, two color profiles were applied to the virtual environment to reinforce the targeted emotional state. A happiness profile increased ambient brightness and introduced warmer color tones, while a sadness profile reduced brightness and shifted the palette toward cooler tones. Audio stimuli were incorporated following established practice [227]: a major chord composition (Vivaldi’s Allegro molto in C major) was played in loop during happiness recording, and a minor chord composition (Sainte Colombe’s Les pleurs) during sadness recording. Each participant’s Emotion Calibration Profile comprised six Calibration Recording files corresponding to the emotions elicited during the session.

**Participants** The study involved 40 participants (24 male, 16 female) aged between 19 and 33 years ( $M = 25.56$ ,  $SD = 2.75$ ), recruited from among students and staff at the authors’ institution as prospective users of social VR applications. On a five-point scale ranging from never used to daily use, participants reported

moderate prior VR experience ( $M = 2.44$ ,  $SD = 1.27$ ) and limited familiarity with social VR platforms ( $M = 1.28$ ,  $SD = 0.57$ ).

Participants were divided into two groups with no statistically significant differences in age or prior VR and social VR experience. The *AFM* group experienced the social VR application with the full AM system active, including facial mimicry. The *Control* group used the same application with only nodding and eye contact mimicry enabled. None of the participants were aware that the AM system was operating during the experiment. To minimize potential biases, participants were paired with strangers and grouped by gender, following the approach adopted in similar studies [224], [225], [228].

## Procedure

The experimental procedure (Figure 3.16) comprised five sequential phases: (1) pre-experience questionnaire, (2) calibration, (3) acclimatization in the virtual environment, (4) conversation, and (5) post-experience questionnaire.

**Pre-experience questionnaire** Following a brief introduction to the study procedure, participants completed a pre-experience questionnaire collecting demographic information, prior experience with relevant technologies, and personality traits.

**Eye-tracking calibration and emotion calibration profile setup** Each participant was equipped with a VR headset and first completed the built-in eye-tracking calibration for the Meta Quest Pro. They were then guided through the classifier setup procedure described in Section 3.4.2, using the stimuli outlined in Section 3.4.3. Upon completion, the classifier was activated and loaded with the participant’s Emotion Calibration Profile in preparation for the main experiment. Dyads were formed based on personality compatibility assessed through the pre-experience questionnaire; participants were occasionally asked to return at a later time if a suitable partner was not immediately available. Throughout the experiment, the two members of each dyad were in separate rooms under the supervision of different experimenters.

**Acclimatization** Before beginning the main experiment, each participant engaged with the single-user acclimatization scene described in Section 3.4.2. Following the approach adopted in prior work [93], [182], participants were given two to three minutes to explore their avatar in the virtual mirror, promoting a sense of embodiment and familiarization with their digital representation.

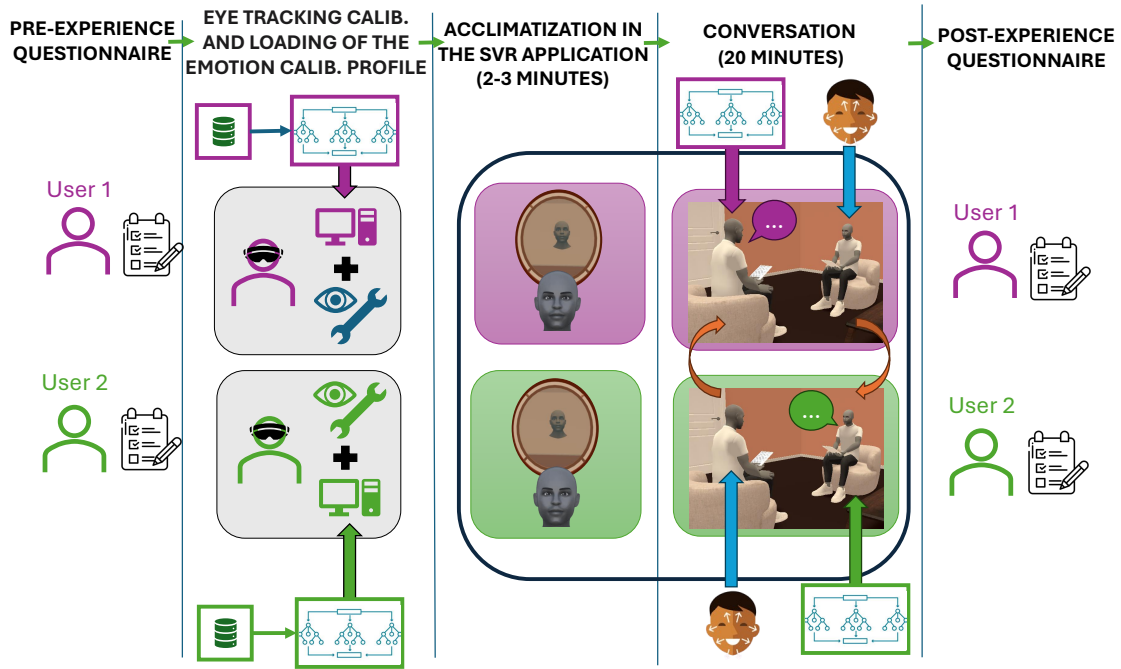


Figure 3.16: Overview of the experimental procedure phases.

**Conversation** Following acclimatization, participants entered the multi-user social VR scene and engaged in a structured conversation lasting 20 minutes. Conversation topics were drawn from a subset of FFP questions, consistent with the methodology used in prior studies [224], [225], [229], [230]. To maintain immersion while allowing participants to read the prompts naturally, each avatar held a virtual notebook displaying the questions. Participants alternated turns: one would read the first question and answer it, followed by the other participant answering the same question before reading and responding to the next prompt, proceeding in this manner throughout the session.

**Post-experience questionnaire** Upon completion of the 20-minute conversation, participants completed a post-experience questionnaire assessing various dimensions of interpersonal communication and avatar perception.

## Measures

The evaluation combined subjective and objective measures. Subjective data were collected through pre- and post-experience questionnaires; objective data were obtained from AM system logs recording detected emotions and the operational states of the AFM Management module.

The pre-experience questionnaire gathered demographics, VR familiarity, and

social VR experience, assessed on a five-point Likert scale. It also included the abbreviated Big Five Inventory-Short (BFI-S) [231], used both to characterize participants' social tendencies and to support compatible pairing within dyads.

The post-experience questionnaire incorporated several validated instruments. Interpersonal closeness was measured using the Inclusion of Others in the Self (IOS) scale [229], which asks participants to select one of seven images depicting progressively overlapping circles representing levels of perceived closeness on a one-to-seven scale. Social presence was assessed using the NMMS [155], covering co-presence, attentional involvement, perceived emotional contagion, comprehension, and behavioral interdependence on a seven-point scale.

Interpersonal attraction was evaluated following the methodology of McCroskey and McCain [232], distinguishing between social attraction, measured using four statements from Davis and Perkowski [233], and task attraction, assessed through four statements from Burgoon [234]. All items were rated on a seven-point Likert scale.

Avatar behavioral naturalness was evaluated using the corresponding section of the questionnaire by Kullmann et al. [235], designed to identify discrepancies between expected and observed behavior in immersive social VR contexts, rated on a one-to-seven scale. The realism of avatar behavior was additionally assessed using the Godspeed questionnaire [203], widely employed in virtual agent and robotics research, on a one-to-five scale.

Interpersonal trust was assessed through a set of custom questions rated on a seven-point Likert scale, covering perceived comfort, sense of being emotionally understood for both happiness and sadness, trust, and empathy. No standardized questionnaire was available for this purpose, consistent with the approach adopted in similar studies [17]. Dimensions such as communication satisfaction and attention to behavioral cues were excluded due to redundancy with other measures or misalignment with the study protocol. All questionnaires and the adapted FFP questions used during the experiment, provided in both English and Italian, are available as supplemental material<sup>21</sup>.

Objective data from the AM system logs included the number of nods detected, the proportion of time each participant spent looking at the peer's face, and the distribution of operating time across detected emotions and AFM module states.

### 3.4.4 Results and discussion

#### Statistical analysis

Data normality was assessed using the Shapiro-Wilk test. Given that distributions deviated significantly from normality, the non-parametric Mann-Whitney

---

<sup>21</sup>Questionnaire and FFP questions: <https://tinyurl.com/4expjpyf>

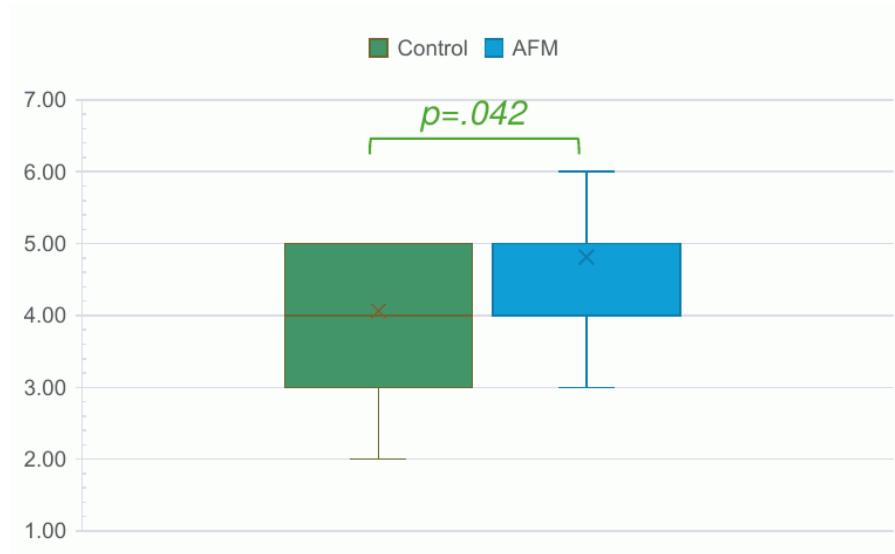


Figure 3.17: Distribution of IOS scores assessing interpersonal closeness. The bracket indicates a statistically significant difference between groups.

test was used to assess statistical significance across groups, with a threshold of  $p < .050$ . Internal consistency of each measure was evaluated using Cronbach's alpha, and effect sizes were computed as Cohen's  $d$ .

### Interpersonal closeness

Participants in the AFM group reported significantly higher IOS scores than those in the Control group ( $M = 4.81$ ,  $SD = 0.81$  vs.  $M = 4.06$ ,  $SD = 0.97$ ;  $d = 0.84$ ), indicating a stronger sense of interpersonal closeness following the conversation. Results are illustrated in Figure 3.17.

### Social presence

Internal consistency of the NMMS was acceptable for both groups (AFM:  $\alpha = .95$ ; Control:  $\alpha = .91$ ). Significant differences in favor of the AFM group were found for three of the five subscales: co-presence ( $M = 6.05$ ,  $SD = 0.56$  vs.  $M = 5.16$ ,  $SD = 0.95$ ;  $d = 1.13$ ), perceived emotional contagion ( $M = 5.09$ ,  $SD = 1.00$  vs.  $M = 4.13$ ,  $SD = 1.06$ ;  $d = 0.93$ ), and behavioral interdependence ( $M = 4.96$ ,  $SD = 0.94$  vs.  $M = 3.90$ ,  $SD = 1.22$ ;  $d = 0.98$ ), as shown in Figure 3.18.

At the item level, the AFM group reported a significantly stronger sense of physical co-presence: participants felt less strongly that the other was not actually in the same room ( $M = 1.94$ ,  $SD = 0.97$  vs.  $M = 3.19$ ,  $SD = 1.59$ ;  $p = .011$ ;  $d = 0.95$ ), and similarly perceived that the other felt less distant from them ( $M = 1.81$ ,  $SD = 0.63$  vs.  $M = 3.25$ ,  $SD = 1.68$ ;  $p = .009$ ;  $d = 1.13$ ). Mutual awareness was



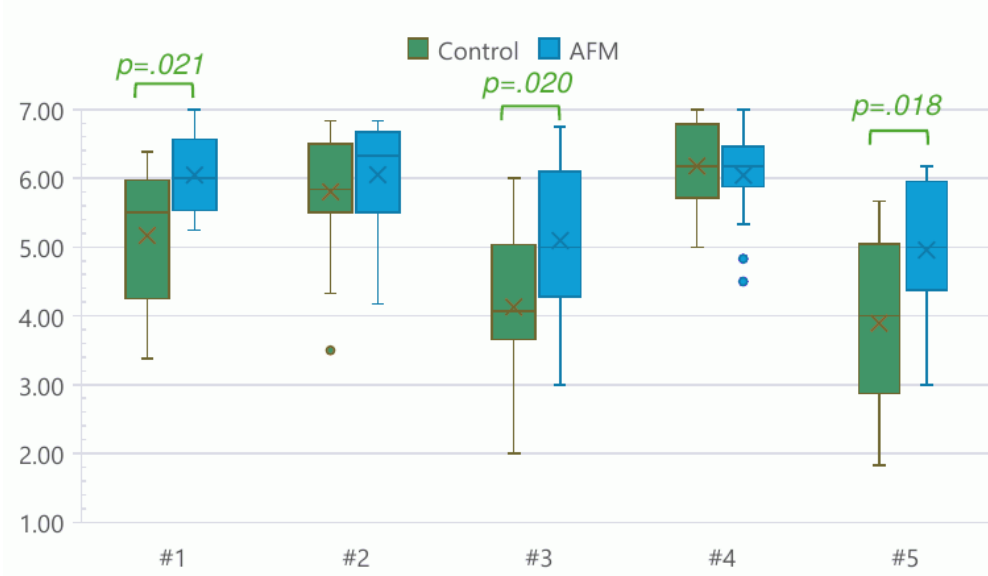


Figure 3.18: Distribution of NMMS scores (#1 Co-presence, #2 Perceived attentional engagement, #3 Perceived emotional contagion, #4 Perceived comprehension, #5 Perceived behavioral interdependence). Brackets indicate statistically significant differences.

also higher in the AFM group, both in terms of perceiving that the other was aware of them ( $M = 6.06$ ,  $SD = 0.83$  vs.  $M = 5.12$ ,  $SD = 1.05$ ;  $p = .016$ ;  $d = 0.99$ ) and in terms of believing that the other perceived them as aware ( $M = 5.63$ ,  $SD = 0.93$  vs.  $M = 4.75$ ,  $SD = 1.09$ ;  $p = .040$ ;  $d = 0.87$ ). For emotional contagion, the AFM group rated both happiness ( $M = 5.75$ ,  $SD = 0.90$  vs.  $M = 4.63$ ,  $SD = 1.36$ ;  $p = .024$ ;  $d = 0.97$ ) and sadness contagion ( $M = 5.06$ ,  $SD = 1.43$  vs.  $M = 3.56$ ,  $SD = 1.27$ ;  $p = .008$ ;  $d = 1.11$ ) significantly higher. Regarding behavioral interdependence, AFM group participants reported a stronger sense of influencing the other's actions ( $M = 5.00$ ,  $SD = 1.37$  vs.  $M = 3.75$ ,  $SD = 1.15$ ;  $p = .005$ ;  $d = 0.99$ ), greater perception that the other's behavior depended on their own ( $M = 4.88$ ,  $SD = 1.36$  vs.  $M = 3.56$ ,  $SD = 1.32$ ;  $p = .005$ ;  $d = 0.98$ ), and a stronger sense that the other was responding directly to them ( $M = 5.06$ ,  $SD = 0.97$  vs.  $M = 3.88$ ,  $SD = 1.41$ ;  $p = .020$ ;  $d = 0.98$ ). No significant differences were observed for perceived attentional engagement or comprehension.

### Interpersonal attraction

Social attraction was significantly higher in the AFM group, both for the individual items of likability ( $M = 6.13$ ,  $SD = 1.05$  vs.  $M = 5.00$ ,  $SD = 1.50$ ;  $d = 0.87$ ) and friendliness ( $M = 6.56$ ,  $SD = 0.61$  vs.  $M = 5.69$ ,  $SD = 1.04$ ;  $p = .014$ ;

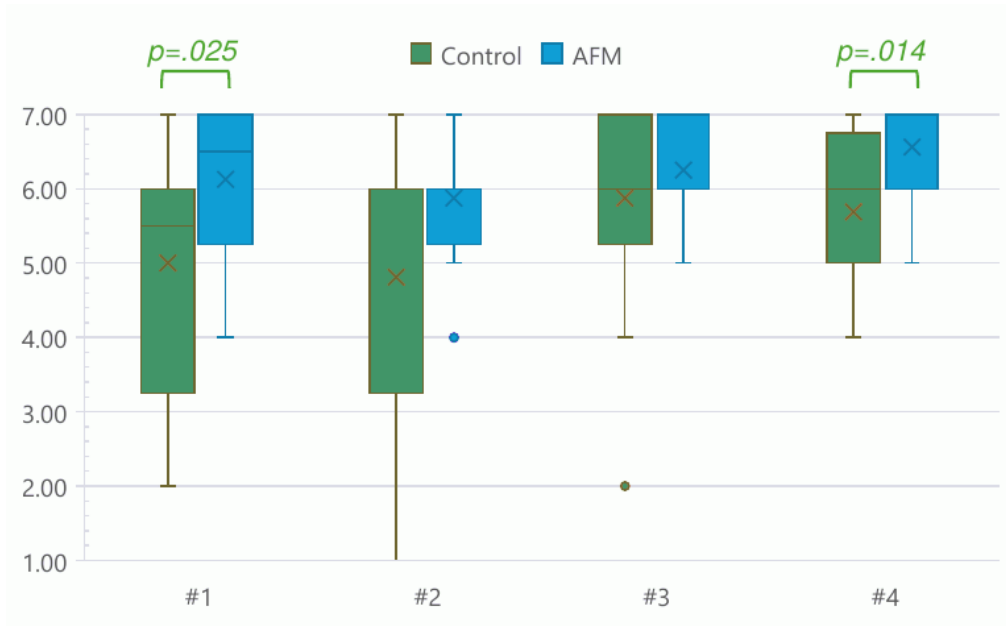


Figure 3.19: Distribution of social attraction scores (#1 I like my partner; #2 I would get along well with my partner; #3 I would enjoy a casual conversation with my partner; #4 My partner is friendly). Brackets indicate statistically significant differences.

$d = 1.02$ ), and for the aggregate score ( $M = 6.20$ ,  $SD = 0.71$  vs.  $M = 5.34$ ,  $SD = 1.30$ ;  $p = .041$ ;  $d = 0.82$ ), as shown in Figure 3.19. No significant differences were found for task attraction, either at the item level or for the aggregate ( $M = 6.18$ ,  $SD = 0.90$  vs.  $M = 5.76$ ,  $SD = 0.90$ ;  $p = .150$ ;  $d = 0.47$ ). Internal consistency was acceptable for both groups (AFM:  $\alpha = .94$ ; Control:  $\alpha = .93$ ).

### Behavioral naturalness and avatar realism

Overall behavioral naturalness was rated significantly higher by the AFM group ( $M = 5.64$ ,  $SD = 0.68$  vs.  $M = 5.16$ ,  $SD = 0.66$ ;  $p = .050$ ;  $d = 0.71$ ), though no significant differences emerged at the individual item level. Internal consistency was good for the AFM group ( $\alpha = .83$ ) but lower for the Control group ( $\alpha = .66$ ).

The Godspeed questionnaire revealed significantly higher avatar realism ratings in the AFM group ( $M = 4.05$ ,  $SD = 0.72$  vs.  $M = 3.33$ ,  $SD = 1.00$ ;  $p = .040$ ;  $d = 0.83$ ), driven primarily by the perception of less artificial behavior ( $M = 4.13$ ,  $SD = 1.11$  vs.  $M = 3.00$ ,  $SD = 1.17$ ;  $p = .001$ ;  $d = 0.99$ ), as illustrated in Figure 3.20. Internal consistency was acceptable for both groups (AFM:  $\alpha = .73$ ; Control:  $\alpha = .85$ ).

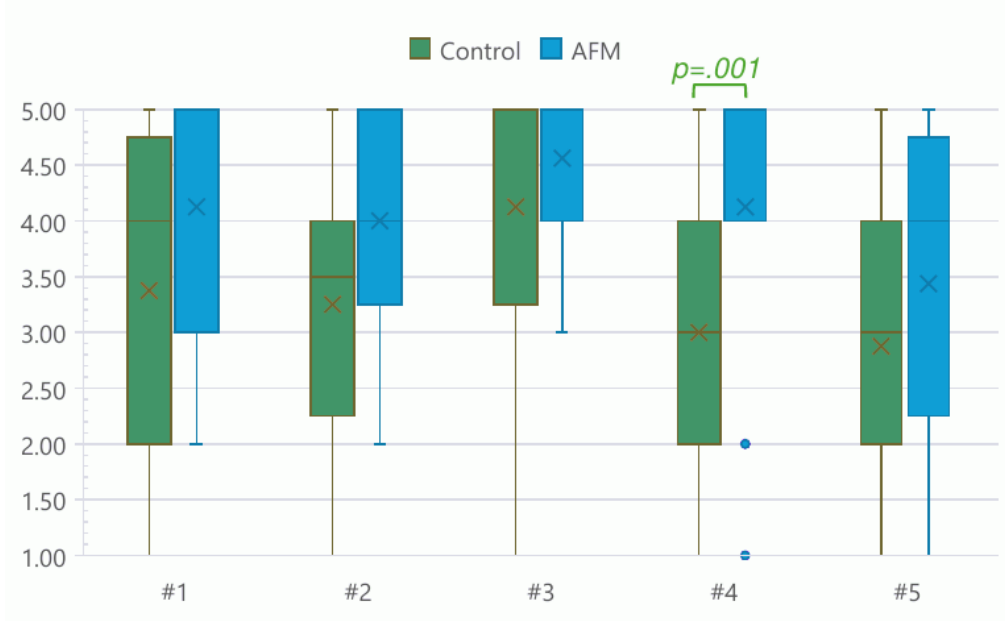


Figure 3.20: Distribution of Godspeed anthropomorphism scores (#1 Fake / Natural; #2 Machinelike / Humanlike; #3 Unconscious / Conscious; #4 Artificial / Lifelike; #5 Moving rigidly / Moving elegantly). Brackets indicate statistically significant differences.

### Interpersonal trust

Significant differences in favor of the AFM group were found for all custom trust items, as shown in Figure 3.21. Participants in the AFM group reported feeling more comfortable with the other person ( $M = 6.31$ ,  $SD = 0.46$  vs.  $M = 5.56$ ,  $SD = 0.99$ ;  $d = 0.96$ ), feeling better understood when happy ( $M = 6.19$ ,  $SD = 0.73$  vs.  $M = 5.19$ ,  $SD = 1.18$ ;  $d = 1.02$ ) and when sad ( $M = 5.81$ ,  $SD = 1.01$  vs.  $M = 4.75$ ,  $SD = 0.75$ ;  $d = 1.19$ ), experiencing higher trust in the other ( $M = 5.69$ ,  $SD = 0.92$  vs.  $M = 4.88$ ,  $SD = 0.86$ ;  $d = 0.92$ ), and perceiving the other as more empathetic ( $M = 6.06$ ,  $SD = 0.75$  vs.  $M = 5.13$ ,  $SD = 0.86$ ;  $d = 1.17$ ). Internal consistency was acceptable for both groups (AFM:  $\alpha = .85$ ; Control:  $\alpha = .83$ ).

### Nested analysis and objective measures

To verify the absence of significant within-group variability attributable to dyad-level differences, a Nested Ranks Test<sup>22</sup> was conducted, a Mann-Whitney-Wilcoxon variant suited for nested data with non-normal distributions. Significant differences were found for only a small number of items, primarily within the Control group. A

<sup>22</sup>Nested Ranks Test: <https://github.com/douglasgscfield/nestedRanksTest>

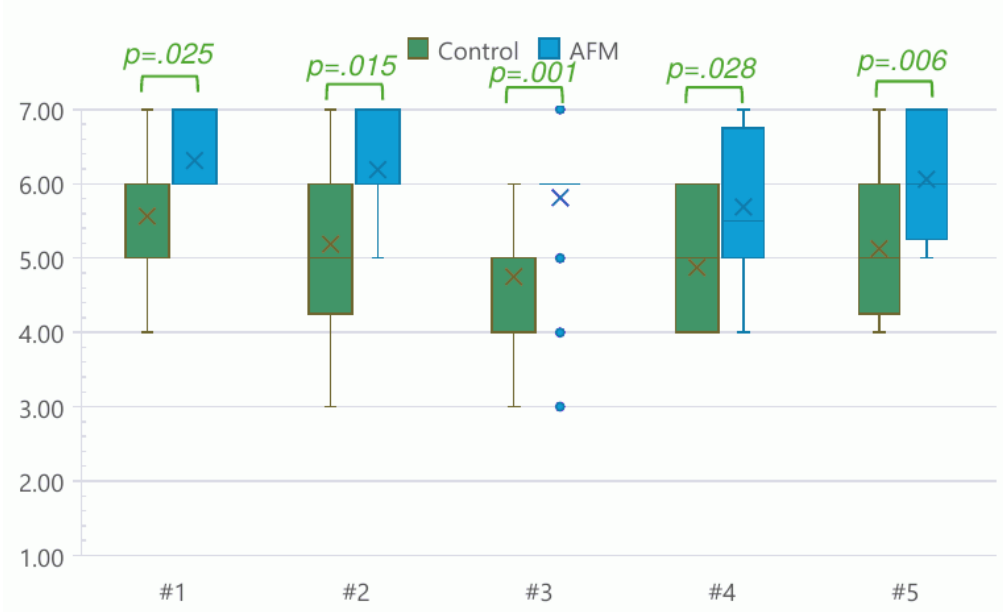


Figure 3.21: Distribution of scores for custom interpersonal trust items (#1 The other person’s reactions made me feel comfortable; #2 When I was happy I felt understood; #3 When I was sad I felt understood; #4 I think the other person deserves my trust; #5 I think the other person was empathic with me). Brackets indicate statistically significant differences.

subsequent variability assessment using the Coefficient of Variation (CV), with 20% as the threshold for moderate to high variability [236], revealed that the affected items exhibited higher variability in the Control group than in the AFM group. These items were: perception of being in the same room ( $p = .031$ ; CV: 33.32% vs. 20.35%), sadness emotional contagion ( $p = .010$ ; CV: 20.35% vs. 19.58%), personal liking of the partner ( $p = .038$ ; CV: 26.16% vs. 14.58%), and perception of the partner’s behavior as artificial versus lifelike ( $p = .027$ ; CV: 37.44% vs. 17.68%). These findings suggest that the AFM condition provided a more consistent user experience across dyads compared to the control condition.

Objective data from the AM system logs showed no significant differences between groups in terms of nodding frequency ( $M = 8.58$ ,  $SD = 5.36$  vs.  $M = 11.00$ ,  $SD = 7.77$ ;  $p = .620$ ;  $d = 0.36$ ) or proportion of time spent looking at the peer’s face ( $M = 87.00$ ,  $SD = 3.22$  vs.  $M = 85.20$ ,  $SD = 3.49$ ;  $p = .150$ ;  $d = 0.54$ ), confirming that differences in subjective outcomes cannot be attributed to differences in baseline non-verbal behavior between groups. A detailed breakdown of AFM module usage across emotional states is provided in Table 3.1.

Table 3.1: Average percentages of **AFM** module usage across all participants during their speaking turns ( $\sim 10$  minutes), categorized by detected emotion. Standard deviations are reported in parentheses. Column (a) reports the proportion of speaking time during which each emotion was detected. The module operates in three modes: Augmentation (b), Camouflage (c), and **AFM** Off (e), where  $b + c + e = 100\%$ . Column (d) reports the combined active **AFM** time ( $b + c$ ). Column (f) reports the proportion of time during which the peer’s emotion naturally aligned with the speaker’s without **AFM** intervention (**Natural Facial Mimicry**, **NFM**).

| Detected Emotion | (a) Time Detected | (b) Augm.     | (c) Camo.   | (d) <b>AFM</b> On (b+c) | (e) <b>AFM</b> Off | (f) <b>NFM</b> |
|------------------|-------------------|---------------|-------------|-------------------------|--------------------|----------------|
| Happy            | 32.68 (20.99)     | 22.96 (15.05) | 1.07 (1.15) | 24.04 (15.24)           | 8.65 (5.87)        | 30.48 (19.81)  |
| Sad              | 55.06 (18.74)     | 29.34 (10.77) | 3.19 (3.54) | 32.53 (11.57)           | 22.54 (17.36)      | 48.96 (15.85)  |
| Neutral          | 12.25 (2.38)      | –             | –           | –                       | 12.25 (2.38)       | –              |
| Total            | 100.00 (–)        | 52.30 (15.26) | 4.27 (3.03) | 56.57 (14.59)           | 43.43 (14.59)      | 79.45 (4.71)   |

## Discussion

The study investigated two hypotheses: that **AFM** can enhance the quality of avatar-mediated interpersonal communication in immersive social **VR** (H1), and that it does not negatively affect the perceived behavioral naturalness of avatars (H2). The experimental results provide substantial support for both, confirming **AFM**’s positive contribution to social interaction quality and avatar authenticity in immersive virtual environments.

**Support for H1.** The **AFM** group reported significantly higher levels of interpersonal closeness, a finding consistent with prior research demonstrating mimicry’s role in strengthening social bonds across face-to-face interactions [197], remote communication [204], and human-agent settings [183], [202]. This result was further corroborated by open comments from two participants in the Control group, one of whom noted that the peer seemed disengaged, while the other reported a lack of genuine connection during the conversation.

Social attraction, specifically in terms of likability and friendliness, was also rated significantly higher in the **AFM** group, reinforcing **AFM**’s capacity to enhance the perception of the peer’s responsiveness and presence [198], [200]. Consistent with prior work on interpersonal communication in social **VR** [17], no significant differences were found for communication satisfaction, with both groups achieving high scores. Similarly, task attraction did not differ significantly between groups, an outcome that may reflect the inherently engaging structure of the **FFP** conversation rather than a limitation of the **AFM** system itself.

**AFM** also contributed to stronger perceptions of interpersonal trust, mirroring findings from psychology and virtual agent research linking mimicry to increased

trust and cooperative behavior [191], [200]. Empathy, particularly for happiness and sadness, was perceived as higher in the AFM group, in line with evidence from both physical [237] and virtual settings [202]. Social presence benefits were most pronounced for co-presence, perceived emotional contagion, and behavioral interdependence. Higher co-presence ratings suggest that AFM led participants to perceive the peer as more physically present, aware, and attuned to them [238]. The emotional contagion results validate the system’s effectiveness in conveying emotional states through avatar facial dynamics, while the behavioral interdependence findings highlight AFM’s ability to amplify the perception of mutual responsiveness. No significant differences were found for perceived attentional engagement or comprehension, a result that is nonetheless informative: it indicates that AFM did not introduce distractions or interfere with participants’ ability to follow the conversation, despite the known role of non-verbal cues in supporting communication [228], [239].

**Support for H2.** The absence of attentional disruption is consistent with the finding that AFM did not negatively affect the perceived behavioral naturalness of avatars, supporting H2. Participants in the AFM group consistently rated avatars as more natural and lifelike, without reporting any discomfort or perception of artificial manipulation. These results align with prior work on mimicry in virtual environments [203], suggesting that AFM can enhance perceived realism without triggering adverse reactions such as the uncanny valley effect [240]. The subtle nature of the facial augmentation, combined with the use of avatar models that balanced visual realism without appearing overly artificial, contributed to this positive reception. Participants in the AFM group expressed surprise when informed at the end of the experiment that their peer’s expressions had been artificially modified, confirming that the system operated below the threshold of conscious detection throughout the interaction.

**Objective measures.** AM system logs revealed that AFM was active for just over 50% of the total speaking time, indicating that the system functioned as a supplementary tool rather than a wholesale replacement of participants’ natural expressive behavior. During the majority of its active time, the system operated in emotion augmentation mode, amplifying the listener’s mimicked emotional response and extending periods of emotional alignment, which has been associated with enhanced communication quality [219]. The prevalence of this mode is consistent with a substantial presence of natural facial mimicry, reflecting the spontaneous emotional alignment that characterizes real-life interaction [241]. The limited activation of emotion camouflage further suggests that participants tended to align emotionally without requiring corrective intervention. The high proportion of detected sadness is likely attributable to the predominantly reflective and emotionally serious tone of the FFP questions used in the experiment.

**Nested analysis.** The nested analysis reinforced the robustness of the between-groups findings by showing that significant within-group variability was concentrated almost exclusively in the Control group, suggesting that AFM produced a more consistent and reliable user experience across dyads. The high variability observed in both groups for same-room perception indicates that this difference was not driven by greater consistency in either condition specifically. For sadness emotional contagion, the Control group showed marginally higher variability, while for personal liking of the partner, the greater spread in the Control group likely reflects the subjective nature of this measure in the absence of preference-based pairing. The higher variability in the Control group for partner appearance ratings may be attributed to individual differences in facial tracking performance and varying degrees of familiarity with 3D avatar representations [242].

## 3.5 Considerations and remarks

The four studies in this chapter explore how social interaction in avatar-mediated environments can be designed and augmented to more closely approximate the richness of face-to-face human communication [97], [243]. Together, they address complementary dimensions of social VR experience, including facial expression fidelity, handshake interaction, close-contact animation, and facial mimicry, and their findings converge on several overarching considerations for researchers and designers in this space.

A key theme is the fidelity–naturalness trade-off. Across several studies, higher fidelity does not automatically translate into better perceived quality [120], [244], [245]. In the emotion conveyance study, full facial tracking yielded significant perceptual advantages over lip-sync estimation for emotions that rely heavily on eye and mouth dynamics, such as sadness and disgust [246]. However, for transient or speech-dominant emotions like anger and surprise, the two modalities were largely indistinguishable, suggesting that perception depends more on context than on tracking accuracy alone [247]. For anger, speech dominated the scene, favoring lip-sync, while for surprise the brief nature of the expression likely reduced its perceptibility regardless of the method. Similarly, the animation refinement framework achieved objective improvements in joint rotation error during hugging interactions, and participants perceived fewer collisions and greater harmony, yet a slight reduction in humanlikeness was noted, likely due to insufficient motion smoothness [29]. These observations suggest that future developments should pursue not only accuracy but also perceptual coherence, ensuring that improvements match users’ embodied expectations [248], [249].

A second theme is the importance of natural gesture-based interaction. The handshake study demonstrated that gesture-based interaction was consistently preferred over GUI-triggered or button-mapped alternatives [41], [250], showing the importance of natural motor behaviors rather than abstracting them into discrete



inputs [251]. This finding resonates with the animation refinement contribution, which highlights the challenge of preserving humanlike fluidity in contact-rich interactions [252], and with the facial mimicry study, where the subtlety of artificially augmented expressions went largely unnoticed by participants, who nonetheless rated interactions as more natural and socially engaging [34]. Across the three studies, these results suggest that the closer a system matches users' expectations of social behavior, the better it is received [97], [162].

A third theme is social signal augmentation. The AFM contribution shows that the controlled amplification of emotional facial cues enhanced interpersonal closeness, social attraction, trust, and social presence, without disrupting the perceived naturalness of the interaction [34]. This points to a promising design direction in which social VR platforms do not merely replicate the real interactions but actively support and enrich the non-verbal communication layer, compensating for the inherent limitations of current tracking and rendering pipelines [111], [114]. Previous work shows that injecting mimicry-based cues can positively affect rapport and cooperation in avatar-mediated contexts [115], [183], and that even non-realistic social augmentations such as visualised gaze or joint attention can measurably increase social presence [253]. At the same time, the modest but consistent gains achieved through animation refinement for close-contact interactions suggest that even partial improvements to physical interaction fidelity can meaningfully contribute to the sense of co-presence and shared embodiment [254]. The AFM study also raises important ethical considerations that deserve sustained attention as the technology matures. The subtle influence that AFM exerts on user perception and behavior raises questions about transparency and the potential for covert manipulation, particularly in commercial deployments [114]. Regulatory frameworks should ensure users are informed and retain control, while the long-term psychological implications of repeated exposure to artificially augmented social interactions, including the risks of dependency or emotional desensitization, merit dedicated investigation.

Finally, the studies highlight several open challenges. Hardware limitations, such as the HTC VIVE Facial Tracker's difficulty in detecting smiling, continue to constrain the realization of expressive avatars [245], and future work should consider both alternative hardware solutions and software-based approaches combining emotion detection algorithms with lip-sync approximation to better reconstruct the full range of facial dynamics. The relatively small sample sizes adopted across all studies limit the generalizability of findings and call for replication with larger and more diverse participant pools [162]. From a methodological standpoint, the emotion conveyance study relied on scripted scenes, which limits generalization. Future evaluations should consider multiple scene variants per emotion to account for the situational variability of emotional expression, as well as refined scene designs that reduce the reliance on speech for emotions such as anger. The handshake study, while establishing the superiority of gesture-based interaction, leaves open the question of how this design compares against hand-tracking solutions and whether the

addition of verbal or non-verbal cues to initiate the interaction, such as gaze-based triggers, would further improve the experience. For the animation refinement framework, the introduction of motion smoothing filters to address the humanlikeness limitation, the extension to a broader range of close-contact interactions, and the integration of haptic feedback systems represent the most immediate priorities. Regarding [AFM](#), extending the emotional scope beyond happiness and sadness to cover a wider range of affective states, integrating complementary non-verbal channels such as body posture and vocal intonation, and conducting longitudinal studies to assess the sustained effects of augmented social interaction on user behavior and emotional well-being are all directions that warrant future attention. More broadly, the focus on individual non-verbal channels in isolation across the four studies leaves open the question of how these modalities interact and compound when combined, motivating future multimodal investigations [\[162\]](#).

In summary, the studies show that improvements in avatar-mediated interaction can come from multiple directions and that technical fidelity, perceptual naturalness, and social augmentation are complementary pillars of a coherent design space rather than mutually exclusive goals. The findings lay a foundation for future research aimed at developing immersive social [VR](#) systems that are not only technically robust but also able to support essential interpersonal interaction and dynamics [\[97\]](#), [\[114\]](#).



## Chapter 4

# Conclusions

In this chapter, a summary of the contributions and their outcomes is presented. Then, the main limitations of the research are discussed, along with potential directions for future work.

### 4.1 Summary of contributions

This doctoral thesis investigates how immersive virtual worlds can be designed and augmented to support richer and more natural social experiences, addressing two complementary research directions: the structural design of socially oriented virtual worlds and the enhancement of avatar-mediated interaction. Across a series of contributions spanning these two areas, the thesis provides a clearer view of the current state of the field and the most promising directions for future advancements.

A key finding is that the quality of a virtual social experience cannot be attributed to any single component in isolation. Rather, it depends on the interaction of multiple design factors, each contributing in different but interconnected ways to shaping how users perceive, inhabit, and interact within immersive environments. The spatial organization of the virtual world, the fidelity of avatar behavior, and the computational mechanisms that support expressive non-verbal communication all proved to be relevant factors, and their combined influence on social user experience was consistently more important than any individual contribution considered alone. This observation reinforces the importance of adopting a holistic and user-centered design perspective when approaching the development of social [VR](#) applications and Metaverse platforms.

Chapter [2](#) addressed the first research direction, focusing on the structural design of immersive social virtual worlds through the lens of a concrete and practically grounded case study, i.e., virtual libraries. The MetaLibrary platform demonstrated that a purposefully designed social virtual world can successfully support a broad

range of community-oriented activities, including socialization, conference attendance, and personalized book recommendations, while remaining accessible to a general audience. The deployment of the system in Turin, Italy, with an installation at the city's central public library, allowed the system to be evaluated in a real institutional context, beyond the controlled conditions of a laboratory study. The lessons learned from this deployment informed the subsequent investigation into design best practices, which drew on expert interviews, a review of the relevant literature, and a dedicated user study to derive and partially validate a structured set of guidelines for the design of immersive social environments in the Metaverse.

The best practices identified in Chapter 2 span a wide range of design dimensions. At the spatial level, the adoption of versatile layouts incorporating high ceilings, open configurations, and large windows was found to foster a sense of openness and support diverse activities, positively influencing users' emotional responses and contributing to a welcoming atmosphere. Flexible spatial arrangements were shown to enhance user experience while optimizing the functional organization of the environment. At the visual level, the incorporation of biophilic design elements, including natural patterns and textures, was found to promote a more realistic and agreeable ambience, while carefully selected color schemes contributed to improved user performance and reduced physiological stress. Regarding the degree of realism pursued, the studies consistently indicated that photorealism should not be treated as a universal objective, but rather applied selectively to interactive elements such as avatars and objects, which benefit most from perceptual fidelity. For the broader environment, stylized or simplified approaches were found to be equally effective in sustaining presence and engagement, while considerably reducing resource consumption. The validation study further revealed that design factors such as lighting and spatial layout interact in non-trivial ways, reinforcing the view that virtual world design is a matter of coherent compositional choices rather than the optimization of isolated variables.

Chapter 3 addressed the second research direction, investigating how social interaction between users can be enriched and augmented within avatar-mediated virtual environments. Four complementary studies explored distinct but related dimensions of this challenge, collectively contributing to a more comprehensive understanding of the factors that determine the perceived quality and naturalness of interpersonal interaction in immersive settings. The emotion conveyance study compared two common approaches for managing facial expressions in social VR: a hardware-based solution relying on facial and eye-tracking to directly map the user's facial movements onto the avatar, and a software-based solution that approximates lip movements on the basis of voice input. The results highlighted the importance of accurate facial tracking for emotions that rely heavily on eye and mouth dynamics. Full facial tracking yielded perceptual advantages over lip-sync estimation for sadness and disgust across most of the evaluated dimensions, and

to a lesser but still meaningful extent for happiness and fear. However, for emotions such as anger and surprise, no conclusive differences were observed between the two modalities. This outcome suggests that, in conditions where speech and body movement provide strong contextual cues, the lower fidelity of the lip-sync approximation may represent an acceptable trade-off between technical complexity and expressive quality, at least for certain emotional states. The handshake interaction study investigated four possible designs for performing greeting gestures using handheld controllers, comparing them across dimensions including embodiment, presence, social presence, usability, and handshake quality of experience. The results consistently indicated that gesture-based interaction was preferred over alternatives relying on graphical user interface elements or simple button presses, underscoring the importance of supporting users' natural motor behaviors rather than abstracting social actions into discrete inputs. This finding points to a broader design principle: the closer an interaction technique aligns with users' intuitive expectations of real-world social dynamics, the more positively it is perceived, even when the technical realization remains imperfect.

The animation refinement study proposed and evaluated a machine learning-based framework designed to improve the behavioral fidelity of close-contact avatar interactions in real time. The framework combines a base inverse kinematics system for reconstructing avatar body movements with a multilayer perceptron module trained on motion capture recordings of avatar-to-avatar interactions, enabling the network to apply realistic motion refinements during social exchanges. Using hugging as a representative use case, the study collected both objective and subjective evidence of the framework's effectiveness. Objective results demonstrated improvements in mean per-joint rotation error compared to the base inverse kinematics system, while subjective results indicated a reduction in perceived interpenetrations between avatars and an improvement in animation harmony. However, a slight decrease in perceived humanlikeness was also observed, attributed to insufficient smoothness in the refined movements. These findings confirm that machine learning-based approaches hold considerable promise for enhancing the realism of complex social animations, while also identifying motion smoothness as a key area requiring further development. The [AFM](#) study investigated whether the controlled amplification of users' emotional facial cues could enhance the quality of social interaction in immersive [VR](#). The system, referred to as [AFM](#), operated by detecting the emotional state of a listener and augmenting their facial expressions to increase periods of emotional alignment during conversations. The experimental results provided substantial support for the hypothesis that [AFM](#) can enhance avatar-mediated interpersonal communication without negatively affecting the perceived naturalness of avatar behavior. Participants in the [AFM](#) group reported higher levels of interpersonal closeness, social attraction, trust, and social presence, particularly in terms of co-presence, perceived emotional contagion, and behavioral interdependence. Notably, the subtle nature of the augmentation went largely

unnoticed by participants during the experiment, who expressed surprise when informed afterward that their peer’s expressions had been artificially modified. These findings position social augmentation as a viable and promising design strategy for immersive virtual environments, demonstrating that computational interventions on non-verbal behavior can meaningfully enrich interpersonal dynamics without compromising the perceived authenticity of the interaction.

Finally, the contributions of this thesis advance the understanding of how socially oriented immersive virtual worlds can be designed and how avatar-mediated interactions can be enhanced to support more natural and engaging social experiences. The findings demonstrate that meaningful improvements are achievable across multiple dimensions of the social VR experience, and that technical fidelity, perceptual naturalness, and social augmentation are not mutually exclusive objectives but rather complementary pillars of a coherent design space.

## 4.2 Limitations

Several limitations must be acknowledged across the studies presented in this thesis. A first and recurring constraint concerns sample size. A number of the conducted user studies involved a relatively limited number of participants, which may restrict the statistical power of the obtained results and their generalizability to broader and more diverse populations. This limitation was particularly relevant in the context of the avatar-mediated interaction studies, where individual variability in response to social cues and augmentation techniques can be considerable, and where larger samples would be necessary to draw definitive conclusions across different user profiles and cultural backgrounds.

A second limitation pertains to the scope of the evaluated conditions within each individual study. Several investigations focused on a restricted set of variables or interaction scenarios, leaving other potentially relevant dimensions unexplored. The emotion conveyance study, for instance, concentrated on six basic emotions within a fixed set of scripted scenes and did not account for the full diversity of emotional expressions encountered in naturalistic social interaction. The animation refinement framework was evaluated exclusively on hugging interactions, leaving open the question of how the proposed approach generalizes to other forms of close-contact social behavior such as handshakes or high-fives. The artificial facial mimicry study addressed only happiness and sadness as target emotions, and the extent to which the system performs effectively across a broader emotional repertoire remains to be established. The handshake interaction study, while thorough in its comparison of controller-based designs, did not include a condition relying on hand-tracking technology, which represents an increasingly available and relevant alternative in contemporary consumer VR systems.

A third limitation concerns the gap between the controlled conditions in which the studies were conducted and the variability of real-world social interaction. All



user studies were carried out in laboratory settings designed to constrain relevant variables, with sessions following a predefined protocol and participants aware of being observed. While this degree of control is methodologically necessary to isolate the effects under investigation, it comes at the cost of ecological validity: participant behavior in these conditions may not fully reflect how users act when unaware of being observed or when interacting without a predefined task structure. In real-world social VR, interactions typically unfold over longer timescales, with dynamics that develop gradually rather than within the bounded duration of a laboratory session. Whether the effects observed under these controlled conditions generalize to naturalistic social VR use, such as informal interactions with remote acquaintances, extended collaborative sessions, or participation in cultural events, remains an open question for future research.

A fourth limitation relates to the design of the studies themselves. Most of the investigations addressed individual non-verbal communication channels, such as facial expressions, hand gestures, and body contact animations, largely in isolation from one another. While this decomposed approach was methodologically necessary to isolate the contribution of each factor and draw interpretable conclusions, it leaves open the important question of how these channels interact when combined in a fully integrated system. Social interaction in real environments is inherently multimodal, and the perceptual and behavioral effects observed in controlled studies may differ substantially when multiple augmentation and representation techniques operate simultaneously.

Finally, the generalizability of the design best practices derived with the first set of studies is subject to the constraints inherent in case-study-based research. Although the MetaLibrary platform was selected for its representativeness of social VR platforms more broadly, the conclusions are tied to a specific cultural, institutional, and technical context. The extent to which the identified guidelines transfer to different application domains, user populations, interaction tasks, and platform architectures remains to be systematically assessed through comparative studies involving a wider range of scenarios.

## 4.3 Future works

The findings and limitations described above suggest several promising directions for future research, both within the specific areas addressed by this thesis and at the broader intersection of immersive technologies and social experience design.

In the area of virtual environment design, future work should extend the investigation to design dimensions that were not fully addressed in the current studies. The role of sound design, animated environmental elements, non-player character behavior, and dynamic lighting adjustments in shaping user experience and social engagement is a promising direction, given the known influence of auditory and temporal cues on presence and immersion. Comparative studies across different

application scenarios and user populations would also contribute to establishing the broader validity of the proposed best practices, moving beyond the virtual library context toward a more general framework applicable to diverse Metaverse platforms. The MetaLibrary deployment itself, with usage data continuing to be collected in a real institutional setting, offers a concrete opportunity for longitudinal investigations into the long-term impact of social VR environments on community participation and cultural engagement.

In the area of avatar-mediated social interaction, a natural and important direction for future work is the integration of the individual augmentation techniques that have been investigated in this thesis into a unified multimodal framework. Combining facial mimicry augmentation, gesture-based interaction design, refined close-contact animations, and accurate emotion conveyance within a single coherent system would allow researchers to study how these components interact and whether their benefits compound in ecologically valid social scenarios. It would also enable a more realistic evaluation of the overall quality of avatar-mediated communication, moving closer to the complexity of real interpersonal interaction.

Expanding the emotional repertoire addressed in the facial mimicry and emotion conveyance studies represents another important direction. Future investigations should examine a broader range of emotional states, including culturally dependent expressions that may not conform to the universalist framework of basic emotions adopted in the current thesis. This expansion would provide a more comprehensive understanding of the conditions under which high-fidelity facial tracking provides meaningful advantages over approximation-based approaches, and of the circumstances in which artificial mimicry augmentation is most effective. The design of calibration procedures capable of reliably evoking distinct emotional states in participants, and of classifiers capable of detecting a broader range of emotions with sufficient accuracy, will be key technical challenges in this regard.

For the animation refinement framework, future developments should prioritize the incorporation of motion smoothing techniques to address the reduction in perceived human likeness observed in the current study, as well as the exploration of more advanced neural network architectures that better balance prediction accuracy, computational latency, and visual quality. Evaluating the framework across a wider range of close-contact social interactions, and involving participants who are actively engaged in those interactions rather than observing them, would provide a more ecologically valid assessment of its effectiveness. The integration of haptic feedback systems to complement the visual refinement of contact interactions represents a further avenue with considerable potential for enriching the sense of physical co-presence in immersive environments. The handshake interaction study identified gesture recognition as a particularly promising direction for future development. A direct comparison between the gesture-based controller design evaluated in the current thesis and an equivalent design relying on hand-tracking technology would help clarify the relative merits of these two approaches, and could inform

platform-level decisions about the interaction methods best suited to different social VR contexts. More broadly, investigating how gesture-based social interactions can be designed to accommodate a wider variety of greeting and contact behaviors, and how they can be made robust to individual differences in movement style and physical ability, would contribute to the accessibility and inclusivity of social VR platforms.

From a longer-term perspective, several broader research directions appear particularly relevant to the evolution of the field. The integration of augmented reality and VR into mixed reality environments opens new possibilities for social interaction that bridges the physical and digital worlds, and raises new questions about how avatar-mediated communication must adapt to contexts in which virtual and physical participants coexist. The exploration of brain-computer interfaces as an alternative or complement to conventional input modalities could enable more direct and expressive forms of interaction in immersive environments, though technical and ethical challenges remain to be addressed before such approaches can be considered for deployment in social contexts.

Finally, as systems capable of augmenting users' social behavior in virtual environments become increasingly sophisticated, the ethical dimensions of this research warrant careful and ongoing attention. The artificial facial mimicry system evaluated in this thesis, for instance, operates by subtly modifying the expressions that other participants perceive, without necessarily making this modification transparent to all parties involved. Questions concerning informed consent, user autonomy, the boundaries of acceptable behavioral augmentation, and the potential long-term psychological effects of sustained exposure to artificially enriched social interactions must be addressed alongside technical developments. Ensuring that the potential of these technologies is realized in ways that respect and empower users, rather than manipulating them in opaque or commercially exploitative ways, will be a defining challenge for the field as social VR platforms and Metaverse applications continue to evolve toward broader deployment and adoption.



# Glossary

**AFM** Artificial Facial Mimicry. [8](#), [11](#), [16](#), [17](#), [52](#), [81–105](#), [109](#)

**AI** Artificial Intelligence. [20](#), [22](#), [23](#), [26](#), [45](#)

**AM** Artificial Mimicry. [11](#), [51](#), [81](#), [82](#), [84–87](#), [90](#), [91](#), [93–95](#), [100](#), [102](#)

**ATT** Attractiveness. [40](#)

**AU** Action Units. [89](#)

**BFI-S** Big Five Inventory-Short. [95](#)

**FACS** Facial Action Coding System. [89](#)

**FFP** Fast Friends Protocol. [92](#), [94](#), [95](#), [101](#), [102](#)

**FT** Facial+eye Tracking. [55](#), [56](#), [59](#), [60](#), [82](#)

**GUI** Graphical User Interface. [52](#), [61](#), [62](#), [64](#), [68](#), [103](#)

**HMD** Head-Mounted Display. [13](#), [19](#), [22](#), [50](#)

**HQ-I** Hedonic Quality – Identity. [40](#)

**HQ-S** Hedonic Quality – Stimulation. [40](#), [45](#)

**IK** Inverse Kinematics. [10](#), [50](#), [70–76](#), [79](#), [80](#)

**IOS** Inclusion of Others in the Self. [95](#)

**IPQ** i-Group Presence Questionnaire. [66](#)

**LE** Lip-sync Estimation. [55](#), [56](#), [59](#), [60](#)

**MLP** Multilayer Perceptron. [70](#), [73](#), [74](#)

**MPJRE** Mean Per Joint Rotation Error. [77](#), [79](#)

**NFM** Natural Facial Mimicry. [8](#), [101](#)

**NMMS** Networked Minds Measure Scale. [11](#), [66](#), [95–97](#)

**NPC** Non-Player Character. [39](#), [40](#), [62](#), [65](#)

**PQ** Pragmatic Quality. [40](#), [45](#)

**SDK** Software Development Kit. [86](#), [91](#)

**SUS** System Usability Scale. [66](#)

**VOIP** Voice Over Internet Protocol. [24](#), [25](#)

**VR** Virtual Reality. [8](#), [11](#), [13](#), [15](#), [20–25](#), [27](#), [28](#), [31](#), [34–38](#), [40](#), [46](#), [49–56](#), [60–62](#), [64](#), [66](#), [70–72](#), [74–76](#), [81–87](#), [90–95](#), [101](#), [103–105](#), [107–113](#)

# Bibliography

- [1] L.-H. Lee et al., “All one needs to know about Metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda”, *Foundations and Trends in Human-Computer Interaction*, vol. 18, no. 2-3, pp. 100–337, 2024. DOI: [10.1561/11000000095](https://doi.org/10.1561/11000000095).
- [2] J. Guan and A. Morris, “Design frameworks for hyper-connected social xri immersive Metaverse environments”, *IEEE Network*, vol. 37, no. 4, pp. 12–21, 2023. DOI: [10.1109/MNET.005.2300129](https://doi.org/10.1109/MNET.005.2300129).
- [3] A. Koohang et al., “Shaping the Metaverse into reality: A holistic multidisciplinary understanding of opportunities, challenges, and avenues for future investigation”, *J. of Computer Information Systems*, vol. 63, no. 3, pp. 735–765, 2023. DOI: [10.1080/08874417.2023.2165197](https://doi.org/10.1080/08874417.2023.2165197).
- [4] H. Park, D. Ahn, and J. Lee, “Towards a Metaverse workspace: Opportunities, challenges, and design implications”, in *Proc. of the CHI Conference on Human Factors in Computing Systems (CHI 2023)*, ACM, 2023, pp. 1–20. DOI: [10.1145/3544548.3581302](https://doi.org/10.1145/3544548.3581302).
- [5] D. Shin, “Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience?”, *Computers in Human Behavior*, vol. 78, pp. 64–73, 2018. DOI: [10.1016/j.chb.2017.09.012](https://doi.org/10.1016/j.chb.2017.09.012).
- [6] A. Yassien, P. ElAgroudy, E. Makled, and S. Abdennadher, “A design space for social presence in VR”, in *Proc. of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI 2020)*, ACM, 2020. DOI: [10.1145/3419249.3420112](https://doi.org/10.1145/3419249.3420112).
- [7] J. R. J. Neo, A. S. Won, and M. M. Shepley, “Designing immersive virtual environments for human behavior research”, *Frontiers in Virtual Reality*, vol. 2, 2021. DOI: [10.3389/frvir.2021.603750](https://doi.org/10.3389/frvir.2021.603750).
- [8] D. Roth, P. Kullmann, G. Bente, D. Gall, and M. E. Latoschik, “Effects of hybrid and synthetic social gaze in avatar-mediated interactions”, in *Proc. | ISMAR-Adjunct*, 2018, pp. 103–108. DOI: [10.1109/ISMAR-Adjunct.2018.00044](https://doi.org/10.1109/ISMAR-Adjunct.2018.00044).



- [9] X. Wei, X. Jin, and M. Fan, “Communication in immersive social virtual reality: A systematic review of 10 years’ studies”, in *Proc. of the Tenth International Symposium of Chinese CHI*, ACM, 2024, pp. 27–37. DOI: [10.1145/3565698.3565767](https://doi.org/10.1145/3565698.3565767).
- [10] H. Nakanishi, “Freewalk: A social interaction platform for group behaviour in a virtual space”, *Int. J. of Human-Computer Studies*, vol. 60, no. 4, pp. 421–454, 2004. DOI: [10.1016/j.ijhcs.2003.11.003](https://doi.org/10.1016/j.ijhcs.2003.11.003).
- [11] J. Williamson et al., “Social density and its impact on behaviour in virtual environments”, in *Proc. of the 3rd International Workshop on Interactive EXtended Reality (IXR 2025)*, ACM, 2025, pp. 37–46. DOI: [10.1145/3746269.3760421](https://doi.org/10.1145/3746269.3760421).
- [12] J. Han, Y. Sun, S. Ppali, A. Covaci, and A. Vande Moere, “To cuddle, mingle, venture, or guide: How architectural affordances influence the experience of social vr places”, in *Proc. of the 2025 ACM Designing Interactive Systems Conference (DIS 2025)*, ACM, 2025, pp. 3355–3374. DOI: [10.1145/3715336.3735772](https://doi.org/10.1145/3715336.3735772).
- [13] J. R. Williamson, J. O’Hagan, J. A. Guerra-Gomez, J. H. Williamson, P. Cesar, and D. A. Shamma, “Digital proxemics: Designing social and collaborative interaction in virtual environments”, in *Proc. of the 2022 CHI Conference on Human Factors in Computing Systems (CHI 2022)*, ACM, 2022. DOI: [10.1145/3491102.3517594](https://doi.org/10.1145/3491102.3517594).
- [14] R. Darken and J. Sibert, “Wayfinding strategies and behaviors in large virtual worlds”, in *Proc. of the CHI Conference on Human Factors in Computing Systems (CHI 1996)*, ACM, 1996, pp. 142–149. DOI: [10.1145/238386.238459](https://doi.org/10.1145/238386.238459).
- [15] M. P. McCreery, D. B. Vallett, and C. Clark, “Social interaction in a virtual environment: Examining socio-spatial interactivity and social presence using behavioral analytics”, *Computers in Human Behavior*, vol. 51, pp. 203–206, 2015. DOI: <https://doi.org/10.1016/j.chb.2015.04.044>.
- [16] E. Etienne, A.-L. Leclercq, A. Remacle, L. Dessart, and M. Schyns, “Perception of avatars nonverbal behaviors in virtual reality”, *Psychology & Marketing*, vol. 40, no. 11, pp. 2464–2481, 2023. DOI: [10.1002/mar.21871](https://doi.org/10.1002/mar.21871).
- [17] S. Aseeri and V. Interrante, “The influence of avatar representation on interpersonal communication in virtual social environments”, *IEEE Trans. on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2608–2617, 2021. DOI: [10.1109/TVCG.2021.3067783](https://doi.org/10.1109/TVCG.2021.3067783).
- [18] A. Giovannelli, J. Thomas, L. Lane, F. Rodrigues, and D. A. Bowman, “Gestures vs. emojis: Comparing non-verbal reaction visualizations for immersive collaboration”, *IEEE Trans. on Visualization and Computer Graphics*, vol. 29, no. 11, pp. 4772–4781, 2023. DOI: [10.1109/TVCG.2023.3320254](https://doi.org/10.1109/TVCG.2023.3320254).

- [19] H. L. Tarnec, E. Bevacqua, O. Augereau, and P. De Loor, “Effect of avatar facial expressiveness on team collaboration in virtual reality”, in *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents (IVA 2023)*, ACM, 2023. DOI: [10.1145/3570945.3607330](https://doi.org/10.1145/3570945.3607330).
- [20] V. Kasapakis, E. Dzardanova, V. Nikolakopoulou, S. Vosinakis, I. Xenakis, and D. Gavalas, “Social virtual reality: Implementing non-verbal cues in remote synchronous communication”, in *Virtual Reality and Mixed Reality: 18th EuroXR International Conference, EuroXR 2021*, vol. 13105, Springer, 2021. DOI: [10.1007/978-3-030-90739-6\\_10](https://doi.org/10.1007/978-3-030-90739-6_10).
- [21] I. Xenakis, D. Gavalas, V. Kasapakis, E. Dzardanova, and S. Vosinakis, “Nonverbal communication in immersive virtual reality through the lens of presence: A critical review”, *PRESENCE: Virtual and Augmented Reality*, vol. 31, pp. 147–187, 2022. DOI: [10.1162/pres\\_a\\_00387](https://doi.org/10.1162/pres_a_00387).
- [22] E. Dzardanova, V. Nikolakopoulou, V. Kasapakis, S. Vosinakis, I. Xenakis, and D. Gavalas, “Exploring the impact of non-verbal cues on user experience in immersive virtual reality”, *Computer Animation and Virtual Worlds*, vol. 35, no. 1, p. 15, 2024. DOI: [10.1002/cav.2224](https://doi.org/10.1002/cav.2224).
- [23] F. Herrera, S. Y. Oh, and J. N. Bailenson, “Effect of behavioral realism on social interactions inside collaborative virtual environments”, *Presence*, vol. 27, no. 2, pp. 163–182, 2020. DOI: [10.1162/pres\\_a\\_00324](https://doi.org/10.1162/pres_a_00324).
- [24] S. A. Aseeri and V. Interrante, “The influence of avatar representation and behavior on communication in social immersive virtual environments”, in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 823–824. DOI: [10.1109/VR.2018.8446261](https://doi.org/10.1109/VR.2018.8446261).
- [25] Y. Ferstl, S. Thomas, C. Guiard, C. Ennis, and R. McDonnell, “Human or robot? investigating voice, appearance and gesture motion realism of conversational social agents”, in *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (IVA 2021)*, ACM, 2021, pp. 76–83. DOI: [10.1145/3472306.3478338](https://doi.org/10.1145/3472306.3478338).
- [26] E. Makled, C. Gerhardt, T. Schwandt, et al., “Evaluating behavioral realism in AR and VR: A comparison of single-point IK and full-body motion capture virtual humans”, *The Visual Computer*, vol. 41, pp. 11 721–11 733, 2025. DOI: [10.1007/s00371-025-03934-5](https://doi.org/10.1007/s00371-025-03934-5).
- [27] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch, “The effect of avatar realism in immersive social virtual realities”, in *Proc. of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST 2017)*, ACM, 2017. DOI: [10.1145/3139131.3139156](https://doi.org/10.1145/3139131.3139156).

- [28] N. Amadou, K. I. Haque, and Z. Yumak, “Effect of appearance and animation realism on the perception of emotionally expressive virtual humans”, in *Proc. of the 23rd ACM International Conference on Intelligent Virtual Agents (IVA 2023)*, ACM, 2023. DOI: [10.1145/3570945.3607307](https://doi.org/10.1145/3570945.3607307).
- [29] M. Seymour, L. Yuan, A. Dennis, and K. Riemer, “Have we crossed the uncanny valley? understanding affinity, trustworthiness, and preference for realistic digital humans in immersive environments”, *J. of the Association for Information Systems*, vol. 22, pp. 591–617, 2021. DOI: [10.17705/1jais.00674](https://doi.org/10.17705/1jais.00674).
- [30] F. G. Praticcò, I. Checo, A. Visconti, A. Simeone, and F. Lamberti, “Designing Hand-held Controller-based Handshake Interaction in social VR and Metaverse”, in *Proc. of the 16th ACM SIGGRAPH Conference on Motion, Interaction and Games (MIG 2023)*, ACM, 2023. DOI: [10.1145/3623264.3624464](https://doi.org/10.1145/3623264.3624464).
- [31] Z. Wang, E. Giannopoulos, M. Slater, and A. Peer, “Handshake: Realistic Human-Robot Interaction in Haptic Enhanced Virtual Reality”, *Presence*, vol. 20, no. 4, pp. 371–392, 2011. DOI: [10.1162/PRES\\_a\\_00061](https://doi.org/10.1162/PRES_a_00061).
- [32] T. Suma, B. Sonia, K. Agyemang Baffour, and O. Oyekoya, “The effects of avatar voice and facial expression intensity on emotional recognition and user perception”, in *SIGGRAPH Asia 2023 Technical Communications (SA 2023)*, ACM, 2023. DOI: [10.1145/3610543.3626158](https://doi.org/10.1145/3610543.3626158).
- [33] X. Yi et al., “Catch my eyebrow, catch my mind: Examining the effect of upper facial expressions on emotional recognition for vr avatars”, in *Proc. of 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2023)*, 2023, pp. 761–762. DOI: [10.1109/VRW58643.2023.00223](https://doi.org/10.1109/VRW58643.2023.00223).
- [34] A. Visconti, D. Calandra, F. Giorgione, and F. Lamberti, “Enhancing social experiences in immersive virtual reality with artificial facial mimicry”, *IEEE Trans. on Visualization and Computer Graphics*, vol. 31, no. 5, pp. 3325–3335, 2025. DOI: [10.1109/TVCG.2025.3549163](https://doi.org/10.1109/TVCG.2025.3549163).
- [35] P. Weyers, A. Mühlberger, A. Kund, U. Hess, and P. Pauli, “Modulation of facial reactions to avatar emotional faces by nonconscious competition priming”, *Psychophysiology*, vol. 46, no. 2, pp. 328–335, 2009. DOI: [10.1111/j.1469-8986.2008.00771.x](https://doi.org/10.1111/j.1469-8986.2008.00771.x).
- [36] N. Sayadi, S. Co, and D. Gomez-Zara, “"Feeling that i was collaborating with them:" A 20-year scoping review of social virtual reality leveraging collaboration”, *arXiv*, 2024. DOI: [10.48550/arXiv.2412.20266](https://doi.org/10.48550/arXiv.2412.20266).

- [37] F. Weidner et al., “A systematic review on the visualization of avatars and agents in AR & VR displayed using head-mounted displays”, *IEEE Trans. on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2596–2606, 2023. DOI: [10.1109/TVCG.2023.3247072](https://doi.org/10.1109/TVCG.2023.3247072).
- [38] M. Baloup, T. Pietrzak, M. Hachet, and G. Casiez, “Non-isomorphic interaction techniques for controlling avatar facial expressions in vr”, in *Proc. of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST 2021)*, ACM, 2021. DOI: [10.1145/3489849.3489867](https://doi.org/10.1145/3489849.3489867).
- [39] F. De Lorenzis, A. Visconti, A. Cannavò, and F. Lamberti, “Metalibrary: Towards social immersive environments for readers”, in *Proc. of Extended Reality: International Conference, XR Salento 2023, Part II*, Springer-Verlag, 2023, pp. 79–87. DOI: [10.1007/978-3-031-43404-4\\_6](https://doi.org/10.1007/978-3-031-43404-4_6).
- [40] A. Cannavò, G. Arrigo, A. Visconti, F. De Lorenzis, and F. Lamberti, “Designing social immersive virtual environments for the Metaverse: The case study of MetaLibrary”, *Virtual Reality & Intelligent Hardware*, vol. 7, no. 3, pp. 279–298, 2025. DOI: [10.1016/j.vrih.2025.04.002](https://doi.org/10.1016/j.vrih.2025.04.002).
- [41] A. Visconti, D. Calandra, and F. Lamberti, “Comparing technologies for conveying emotions through realistic avatars in virtual reality-based Metaverse experiences”, *Computer Animation and Virtual Worlds*, vol. 34, no. 3-4, p. 11, 2023. DOI: <https://doi.org/10.1002/cav.2188>.
- [42] A. Visconti, R. Macaluso, G. Di Bartolomei, D. Calandra, and F. Lamberti, “Improving fidelity of close social interaction animations in social vr with a machine learning-based refinement framework”, in *Computer Animation and Social Agents (CASA 2025)*, Springer Nature Singapore, 2026, pp. 204–219. DOI: [10.1007/978-981-95-0100-7\\_13](https://doi.org/10.1007/978-981-95-0100-7_13).
- [43] C. S. Oh, J. N. Bailenson, and G. F. Welch, “A systematic review of social presence: Definition, antecedents, and implications”, *Frontiers in Robotics and AI*, vol. Volume 5 - 2018, 2018. DOI: [10.3389/frobt.2018.00114](https://doi.org/10.3389/frobt.2018.00114).
- [44] J. Arango, “Learning from the magic kingdom”, *J. of Information Architecture*, vol. 7, no. 2, pp. 83–106, 2022, ISSN: 1903-7260. DOI: [10.55135/1015060901/222.011/3.047](https://doi.org/10.55135/1015060901/222.011/3.047).
- [45] Y. Miwa and K. Hanyu, “The effects of interior design on communication and impressions of a counselor in a counseling room”, *Environment and Behavior*, vol. 38, no. 4, pp. 484–502, 2006. DOI: [10.1177/0013916505280084](https://doi.org/10.1177/0013916505280084).
- [46] E. Han, C. DeVeaux, J. T. Hancock, N. Ram, G. M. Harari, and J. N. Bailenson, “The influence of spatial dimensions of virtual environments on attitudes and nonverbal behaviors during social interactions”, *J. of Environmental Psychology*, vol. 95, p. 102 269, 2024. DOI: [10.1016/j.jenvp.2024.102269](https://doi.org/10.1016/j.jenvp.2024.102269).

- [47] T. Ishikawa and J. L. Nasar, “A review of effects of visual environmental factors on interpersonal cognition and behavior: Focusing on brightness, color, and depth”, *Japan Architectural Review*, vol. 6, no. 1, e12343, 2023. DOI: [10.1002/2475-8876.12343](https://doi.org/10.1002/2475-8876.12343).
- [48] A. Bellazzi et al., “Virtual reality for assessing visual quality and lighting perception: A systematic review”, *Building and Environment*, vol. 209, p. 108674, 2022. DOI: [10.1016/j.buildenv.2021.108674](https://doi.org/10.1016/j.buildenv.2021.108674).
- [49] M. V. Sanchez-Vives and M. Slater, “From presence to consciousness through virtual reality”, *Nature Reviews Neuroscience*, vol. 6, no. 4, pp. 332–339, 2005. DOI: [10.1038/nrn1651](https://doi.org/10.1038/nrn1651).
- [50] M. Yassin, A. El Antably, and M. A. Abou El-Ela, “The others know the way: A study of the impact of co-presence on wayfinding decisions in an interior virtual environment”, *Automation in Construction*, vol. 128, p. 103782, 2021. DOI: [10.1016/j.autcon.2021.103782](https://doi.org/10.1016/j.autcon.2021.103782).
- [51] S. Minocha and C. L. Hardy, “Designing navigation and wayfinding in 3d virtual learning spaces”, in *Proc. of the 23rd Australian Computer-Human Interaction Conference (OzCHI 2011)*, ACM, 2011, pp. 211–220. DOI: [10.1145/2071536.2071570](https://doi.org/10.1145/2071536.2071570).
- [52] Y. Lee, G. Kim, B. Lee, J. Park, and H. K. Kim, “Refining evaluation tools and scenarios for enhanced social virtual reality experiences across devices”, *Int. J. of Human-Computer Interaction*, vol. 42, no. 2, pp. 708–726, 2026. DOI: [10.1080/10447318.2025.2513027](https://doi.org/10.1080/10447318.2025.2513027).
- [53] S. Nur Şanlı and A. Sirel, “The effects of developing information technologies on 21st century library architecture”, *Proc. of the International Conference of Contemporary Affairs in Architecture and Urbanism-ICCAUA*, vol. 5, no. 1, pp. 12–26, 2022. DOI: [10.38027/ICCAUA2022EN0035](https://doi.org/10.38027/ICCAUA2022EN0035).
- [54] E. Hammais, H. Ketamo, and A. Koivisto, “Virtual information assistants on mobile app to serve visitors at Helsinki Central Library Oodi”, 2017. [Online]. Available: <https://repository.ifla.org/handle/20.500.14598/6587>.
- [55] P. F. Marty, “Museum websites and museum visitors: Before and after the museum visit”, *Museum Management and Curatorship*, vol. 22, no. 4, pp. 337–360, 2007. DOI: [10.1080/09647770701757708](https://doi.org/10.1080/09647770701757708).
- [56] D. Buhalis, D. Leung, and M. Lin, “Metaverse as a disruptive technology revolutionising tourism management and marketing”, *Tourism Management*, vol. 97, 2023. DOI: [10.1016/j.tourman.2023.104724](https://doi.org/10.1016/j.tourman.2023.104724).
- [57] S. Baldoni, M. S. Hadj Sassi, M. Carli, and F. Battisti, “Definition of guidelines for virtual reality application design based on visual attention”, *Multi-media Tools and Applications*, vol. 83, no. 16, pp. 49615–49640, 2024. DOI: [10.1007/s11042-023-17488-y](https://doi.org/10.1007/s11042-023-17488-y).

- [58] D. Calandra, F. De Lorenzis, A. Cannavò, and F. Lamberti, “Immersive virtual reality and passive haptic interfaces to improve procedural learning in a formal training course for first responders”, *Virtual Reality*, pp. 1–28, 2022. DOI: [10.1007/s10055-022-00704-9](https://doi.org/10.1007/s10055-022-00704-9).
- [59] F. De Lorenzis, F. G. Praticcò, M. Repetto, E. Pons, and F. Lamberti, “Immersive virtual reality for procedural training: Comparing traditional and learning by teaching approaches”, *Computers in Industry*, vol. 144, p. 103785, 2023. DOI: [10.1016/j.compind.2022.103785](https://doi.org/10.1016/j.compind.2022.103785).
- [60] A. Sirel, “Reflection of paradigm change in information technology to library architecture: The Helsinki Oodi library”, *Architecture and Urban Planning*, vol. 17, no. 1, pp. 123–135, 2021. DOI: [10.2478/aup-2021-0012](https://doi.org/10.2478/aup-2021-0012).
- [61] T. Uskali, M. Rautiainen, M. Juntunen, R. Tallavaara, and M. Hiljanen, *How to use social VR in higher education: Case study of the JYUXR campus in Finland*. 2022, pp. 99–111. DOI: [10.4324/9780367337032](https://doi.org/10.4324/9780367337032).
- [62] S. Williams, R. Enatsky, H. Gillcash, J. J. Murphy, and D. Gračanin, “Immersive technology in the public school classroom: When a class meets”, in *Proc. of 7th International Conference of the Immersive Learning Research Network (iLRN 2021)*, 2021, pp. 1–8. DOI: [10.23919/iLRN52045.2021.9459371](https://doi.org/10.23919/iLRN52045.2021.9459371).
- [63] D. Calandra, F. G. Praticcò, G. Lupini, and F. Lamberti, “Impact of avatar representation in a virtual reality-based multi-user tunnel fire simulator for training purposes”, in *Proc. of Computer Vision, Imaging and Computer Graphics Theory and Applications*, Springer International Publishing, 2023, pp. 3–20. DOI: [10.1007/978-3-031-25477-2\\_1](https://doi.org/10.1007/978-3-031-25477-2_1).
- [64] P. Manyuru, C. Dobbins, B. Matthews, O. Baumann, and A. Dey, “Focus group on social virtual reality in social virtual reality: Effects on emotion and self-awareness”, in *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct 2021)*, 2021, pp. 437–438. DOI: [10.1109/ISMAR-Adjunct54149.2021.00099](https://doi.org/10.1109/ISMAR-Adjunct54149.2021.00099).
- [65] D. A. Le, B. MacIntyre, and J. Outlaw, “Enhancing the experience of virtual conferences in social virtual environments”, in *Proc. of 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2020)*, 2020, pp. 485–494. DOI: [10.1109/VRW50115.2020.00101](https://doi.org/10.1109/VRW50115.2020.00101).
- [66] S. Nilsson et al., “Fruit golf: An asymmetrical shared space VR/mobile experience”, *Proc. of SIGGRAPH 2022 Immersive Pavilion*, 2022. DOI: [10.1145/3532834.3536214](https://doi.org/10.1145/3532834.3536214).



- [67] S. Iskalinov, A. Aliyev, N. Azimbay, E. Garifullin, and R. Omirgaliyev, “Alternative to the traditional learning: Usage of virtual reality platform for remote education”, in *Proc. of the 2025 IEEE European Technology and Engineering Management Summit (E-TEMS 2025)*, 2025, pp. 252–257. DOI: [10.1109/E-TEMS64751.2025.11239118](https://doi.org/10.1109/E-TEMS64751.2025.11239118).
- [68] H. Shim, K. T. Oh, and C.-K. Shi, “Exploring the social heritage Metaverse for virtual heritage communication from multi end-user centered experiences”, *J. Comput. Cult. Herit.*, vol. 18, no. 3, 2025, ISSN: 1556-4673. DOI: [10.1145/3736225](https://doi.org/10.1145/3736225). [Online]. Available: <https://doi.org/10.1145/3736225>.
- [69] Y. Zhao et al., “Virtual reality in heritage education for enhanced learning experience: A mini-review and design considerations”, *Frontiers in Virtual Reality*, vol. Volume 6 - 2025, 2025. DOI: [10.3389/frvir.2025.1560594](https://doi.org/10.3389/frvir.2025.1560594).
- [70] S. H. Cha, C. Koo, T. W. Kim, and T. Hong, “Spatial perception of ceiling height and type variation in immersive virtual environments”, *Building and Environment*, vol. 163, p. 106285, 2019. DOI: [10.1016/j.buildenv.2019.106285](https://doi.org/10.1016/j.buildenv.2019.106285).
- [71] Y. Li, M. Wang, D. Jin, F. Steinicke, and Q. Zhao, “Effects of virtual environment and self-representations on perception and physical performance in redirected jumping”, *Virtual Reality & Intelligent Hardware*, vol. 3, no. 6, pp. 451–469, 2021. DOI: [10.1016/j.vrih.2021.10.002](https://doi.org/10.1016/j.vrih.2021.10.002).
- [72] J. Han, A. Vande Moere, and A. L. Simeone, “Meaning follows purpose: Unravelling the architectural design conventions in the contemporary Metaverse”, in *Proc. of the CHI Conference on Human Factors in Computing Systems (CHI 2024)*, ACM, 2024, pp. 1–22. DOI: [10.1145/3613904.3642730](https://doi.org/10.1145/3613904.3642730).
- [73] K. Thoring, P. Desmet, and P. Badke-Schaub, “Creative environments for design education and practice: A typology of creative spaces”, *Design Studies*, vol. 56, pp. 54–83, 2018. DOI: [10.1016/j.destud.2018.02.001](https://doi.org/10.1016/j.destud.2018.02.001).
- [74] A. Shemesh, G. Leisman, M. Bar, and Y. J. Grobman, “The emotional influence of different geometries in virtual spaces: A neurocognitive examination”, *J. of Environmental Psychology*, vol. 81, p. 101802, 2022. DOI: [10.1016/j.jenvp.2022.101802](https://doi.org/10.1016/j.jenvp.2022.101802).
- [75] J. Lee, A. Eden, D. R. Ewoldsen, D. Beyea, and S. Lee, “Seeing possibilities for action: Orienting and exploratory behaviors in VR”, *Computers in Human Behavior*, vol. 98, pp. 158–165, 2019. DOI: [10.1016/j.chb.2019.04.022](https://doi.org/10.1016/j.chb.2019.04.022).
- [76] P. Presti, D. Ruzzon, P. Avanzini, F. Caruana, G. Rizzolatti, and G. Vecchiato, “Measuring arousal and valence generated by the dynamic experience of architectural forms in virtual environments”, *Scientific Reports*, vol. 12, no. 1, p. 13376, 2022. DOI: [10.1038/s41598-022-17689-1](https://doi.org/10.1038/s41598-022-17689-1).

- [77] M. Mollazadeh and Y. Zhu, “Application of virtual environments for biophilic design: A critical review”, *Buildings*, vol. 11, no. 4, p. 148, 2021. DOI: [10.3390/buildings11040148](https://doi.org/10.3390/buildings11040148).
- [78] S. H. Cha, S. Zhang, and T. W. Kim, “Effects of interior color schemes on emotion, task performance, and heart rate in immersive virtual environments”, *J. of Interior Design*, vol. 45, no. 4, pp. 51–65, 2020. DOI: [10.1111/joid.12179](https://doi.org/10.1111/joid.12179).
- [79] I. D. Bishop and B. Rohrmann, “Subjective responses to simulated and real environments: A comparison”, *Landscape and Urban Planning*, vol. 65, no. 4, pp. 261–277, 2003. DOI: [10.1016/S0169-2046\(03\)00070-7](https://doi.org/10.1016/S0169-2046(03)00070-7).
- [80] P. Siriaraya and C. Siang Ang, “Age differences in the perception of social presence in the use of 3D virtual world for social interaction”, *Interacting with Computers*, vol. 24, no. 4, pp. 280–291, 2012. DOI: [10.1016/j.intcom.2012.03.003](https://doi.org/10.1016/j.intcom.2012.03.003).
- [81] D.-Y. Kim, H.-K. Lee, and K. Chung, “Avatar-mediated experience in the Metaverse: The impact of avatar realism on user-avatar relationship”, *J. of Retailing and Consumer Services*, vol. 73, p. 103382, 2023. DOI: [10.1016/j.jretconser.2023.103382](https://doi.org/10.1016/j.jretconser.2023.103382).
- [82] F. Ng, *Picture postcard: Schmidt Hammer Lassen’s Shanghai Library East*, The RIBA Journal, [Online]. Available: <https://www.ribaj.com/buildings/picture-postcard-international-schmidt-hammer-lassen-shanghai-library-east-china-public-flora-ng>, Oct. 2023.
- [83] A. Cannavò, D. Calandra, F. G. Pratticò, V. Gatteschi, and F. Lamberti, “An evaluation testbed for locomotion in virtual reality”, *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 3, pp. 1871–1889, 2020. DOI: [10.1109/TVCG.2020.3032440](https://doi.org/10.1109/TVCG.2020.3032440).
- [84] M. Hassenzahl, “Hedonic, emotional, and experiential perspectives on product quality”, in *Encyclopedia of Human Computer Interaction*, Idea Group Reference, 2006, pp. 266–272. DOI: [10.4018/978-1-59140-562-7.ch039](https://doi.org/10.4018/978-1-59140-562-7.ch039).
- [85] M. Sra, A. Mottelson, and P. Maes, “Your place and mine: Designing a shared VR experience for remotely located users”, in *Proc. of the Designing Interactive Systems Conference (DIS ’18 2018)*, ACM, 2018, pp. 85–97. DOI: [10.1145/3196709.3196788](https://doi.org/10.1145/3196709.3196788).
- [86] S. Ronft, A.-K. Beck, and T. Lachmann, “Human-centric virtual lighting: Effects of color temperature and daylight simulation in virtual environments”, *Applied Ergonomics*, vol. 129, p. 104601, 2025. DOI: [10.1016/j.apergo.2025.104601](https://doi.org/10.1016/j.apergo.2025.104601).



- [87] M. Rettinger, X. Jiang, J. Yang, and G. Rigoll, “Visual complexity in vr: Implications for cognitive load”, in *Proc. of 2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2024)*, 2024, pp. 783–784. DOI: [10.1109/VRW62533.2024.00188](https://doi.org/10.1109/VRW62533.2024.00188).
- [88] M. Sanaei, S. B. Gilbert, N. Javadpour, H. Sabouni, M. C. Dorneich, and J. W. Kelly, “The correlations of scene complexity, workload, presence, and cybersickness in a task-based vr game”, in *Virtual, Augmented and Mixed Reality ((HCI 2024))*, Springer Nature Switzerland, 2024, pp. 277–289. DOI: [10.1007/978-3-031-61041-7\\_18](https://doi.org/10.1007/978-3-031-61041-7_18).
- [89] C. Sagnier, E. Loup-Escande, and G. Valléry, “Effects of gender and prior experience in immersive user experience with virtual reality”, in *Advances in Usability and User Experience ((AHFE 2019))*, Springer International Publishing, 2020, pp. 305–314. DOI: [10.1007/978-3-030-19135-1\\_30](https://doi.org/10.1007/978-3-030-19135-1_30).
- [90] L. Souza, A. Cunha, M. Lima, and D. Ricca, “Pragmatic and hedonic aspects of user experience in virtual reality: Analysis of novice users’ information mediation during their first interaction with a Metaverse platform”, *InfoDesign - Revista Brasileira de Design da Informação*, vol. 20, Dec. 2023. DOI: [10.51358/id.v20i3.1068](https://doi.org/10.51358/id.v20i3.1068).
- [91] N. Gueguen, C. Jacob, and A. Martin, “Mimicry in social interaction: Its effect on human judgment and behavior”, *European J. of Social Sciences*, vol. 8, no. 2, pp. 253–259, 2009. [Online]. Available: <https://www.scribd.com/document/414313762/Mimicry-in-Social-Interaction-Effect-on-Human-Judgement-and-Behav-BUENON>.
- [92] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, “Beyond replication: Augmenting social behaviors in multi-user virtual realities”, in *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2018)*, 2018, pp. 215–222. DOI: [10.1109/VR.2018.8447550](https://doi.org/10.1109/VR.2018.8447550).
- [93] D. Roth, D. Mal, C. F. Purps, P. Kullmann, and M. E. Latoschik, “Injecting nonverbal mimicry with hybrid avatar-agent technologies: A naïve approach”, in *Proc. of the ACM Symposium on Spatial User Interaction (SUI 2018)*, ACM, 2018, pp. 69–73. DOI: [10.1145/3267782.3267791](https://doi.org/10.1145/3267782.3267791).
- [94] J. Tanenbaum, M. Seif El-Nasr, and M. Nixon, “Nonverbal communication in virtual worlds”, in *Nonverbal Communication in Virtual Worlds*, ETC Press Pittsburgh, PA, 2014.
- [95] P. Sykownik and M. Masuch, “The experience of social touch in multi-user virtual reality”, in *Proc. of the 26th ACM Symposium on Virtual Reality Software and Technology (VRST 2020)*, 2020, pp. 1–11. DOI: [10.1145/3385956.3418944](https://doi.org/10.1145/3385956.3418944).

- [96] H. J. Smith and M. Neff, "Communication behavior in embodied virtual reality", in *Proc. of the CHI Conference on Human Factors in Computing Systems (CHI 2018)*, ACM, 2018, pp. 1–12. DOI: [10.1145/3173574.3173863](https://doi.org/10.1145/3173574.3173863).
- [97] D. Maloney, G. Freeman, and D. Y. Wohn, "'talking without a voice': Understanding non-verbal communication in social virtual reality", *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. CSCW2, 2020. DOI: [10.1145/3415246](https://doi.org/10.1145/3415246).
- [98] T. J. Tanenbaum, N. Hartoonian, and J. Bryan, "'how do i make this thing smile?': An inventory of expressive nonverbal communication in commercial social virtual reality platforms", in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI 2020)*, ACM, 2020, pp. 1–13. DOI: [10.1145/3313831.3376606](https://doi.org/10.1145/3313831.3376606).
- [99] A. Martin Coesel, B. Biancardi, and S. Buisine, "A theoretical review of the proteus effect: Understanding the underlying processes", *Frontiers in Psychology*, vol. Volume 15 - 2024, 2024. DOI: [10.3389/fpsyg.2024.1379599](https://doi.org/10.3389/fpsyg.2024.1379599).
- [100] J. D. Hart, T. Piumsomboon, G. A. Lee, R. T. Smith, and M. Billinghamurst, "Manipulating avatars for enhanced communication in extended reality", in *Proc. of 2021 IEEE International Conference on Intelligent Reality (ICIR 2021)*, 2021, pp. 9–16. DOI: [10.1109/ICIR51845.2021.00011](https://doi.org/10.1109/ICIR51845.2021.00011).
- [101] L. Wen, J. Zhou, W. Huang, and F. Chen, "A survey of facial capture for virtual reality", *IEEE Access*, vol. 10, pp. 6042–6052, 2022. DOI: [10.1109/ACCESS.2021.3138200](https://doi.org/10.1109/ACCESS.2021.3138200).
- [102] N. Hube, K. Vidackovic, and M. Sedlmair, "Using expressive avatars to increase emotion recognition: A pilot study", in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*, 2022, pp. 1–7. DOI: [10.1145/3491101.3519822](https://doi.org/10.1145/3491101.3519822).
- [103] J. P. Trujillo and J. Holler, "Conversational facial signals combine into compositional meanings that change the interpretation of speaker intentions", *Scientific Reports*, vol. 14, no. 1, p. 2286, 2024. DOI: [10.1038/s41598-024-52589-0](https://doi.org/10.1038/s41598-024-52589-0).
- [104] P. Ekman and W. V. Friesen, "Facial action coding system: A technique for the measurement of facial movement", *Consulting Psychologists Press*, 1978.
- [105] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Polak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements", *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019. DOI: [10.1177/1529100619832930](https://doi.org/10.1177/1529100619832930).
- [106] J. Jiang et al., "AvatarPoser: Articulated full-body pose tracking from sparse motion sensing", in *Proc. of the European Conference on Computer Vision (ECCV 2022)*, Springer, 2022, pp. 443–460. DOI: [10.1007/978-3-031-20065-6\\_26](https://doi.org/10.1007/978-3-031-20065-6_26).

- [107] J. L. Ponton, C. Andújar, and N. Pelechano, “SparsePoser: Real-time full-body motion reconstruction from sparse data”, *ACM Trans. on Graphics*, vol. 42, no. 6, 2023. DOI: [10.1145/3625264](https://doi.org/10.1145/3625264).
- [108] P. Caserman, P. Achenbach, and S. Göbel, “Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality”, in *Proc. of the 7th IEEE Int. Conference on Serious Games and Applications for Health*, 2019, pp. 1–8. DOI: [10.1109/SeGAH.2019.8882429](https://doi.org/10.1109/SeGAH.2019.8882429).
- [109] D. Roth, K. Waldow, M. E. Latoschik, A. Fuhrmann, and G. Bente, “Socially immersive avatar-based communication”, in *Proc. IEEE Virtual Reality*, 2017, pp. 259–260. DOI: [10.1109/VR.2017.7892275](https://doi.org/10.1109/VR.2017.7892275).
- [110] N. Sebanz and G. Knoblich, “Progress in joint-action research”, *Current Directions in Psychological Science*, vol. 30, no. 2, pp. 138–143, 2021. DOI: [10.1177/0963721420984425](https://doi.org/10.1177/0963721420984425).
- [111] C. S. Oh, J. N. Bailenson, N. Krämer, and B. Li, “Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments”, *PLOS ONE*, vol. 11, no. 9, 2016. DOI: [10.1371/journal.pone.0161794](https://doi.org/10.1371/journal.pone.0161794).
- [112] Y. Lee, S. Lee, and S.-H. Lee, “Multifinger interaction between remote users in avatar-mediated telepresence”, *Computer Animation and Virtual Worlds*, vol. 28, no. 3-4, e1778, 2017. DOI: <https://doi.org/10.1002/cav.1778>.
- [113] M. Mangalam, S. Oruganti, and G. Buckingham, “Enhancing hand-object interactions in virtual reality for precision manual tasks”, *Virtual Reality*, vol. 28, p. 166, 2024. DOI: [10.1007/s10055-024-01055-3](https://doi.org/10.1007/s10055-024-01055-3).
- [114] D. Roth et al., “Technologies for social augmentations in user-embodied virtual reality”, in *25th ACM Symposium on Virtual Reality Software and Technology (VRST 2019)*, ACM, 2019, pp. 1–12. DOI: [10.1145/3359996.3364269](https://doi.org/10.1145/3359996.3364269).
- [115] M. Stel and R. Vonk, “Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction”, *British Journal of Psychology*, vol. 101, no. 2, pp. 311–323, 2010. DOI: [10.1348/000712609X465424](https://doi.org/10.1348/000712609X465424).
- [116] T. L. Chartrand and R. van Baaren, “Chapter 5 human mimicry”, in *Adv. in Experimental Social Psychology*, vol. 41, Academic Press, 2009, pp. 219–274. DOI: [10.1016/S0065-2601\(08\)00405-X](https://doi.org/10.1016/S0065-2601(08)00405-X).
- [117] U. Hess and A. Fischer, “Emotional mimicry: Why and when we mimic emotions”, *Social and Personality Psychology Compass*, vol. 8, no. 2, pp. 45–57, 2014. DOI: [10.1111/spc3.12083](https://doi.org/10.1111/spc3.12083).
- [118] E. L. Fasya, E. van den Bos, D. K. Heylen, and M. E. Kret, “Smile mimicry smoothens human–virtual human interactions”, *Int. J. of Human-Computer Studies*, vol. 183, p. 103196, 2024. DOI: [10.1016/j.ijhcs.2023.103196](https://doi.org/10.1016/j.ijhcs.2023.103196).

- [119] I. Wang and J. Ruiz, “Examining the use of nonverbal communication in virtual agents”, *Int. J. of Human-Computer Interaction*, vol. 37, no. 17, pp. 1648–1673, 2021. DOI: [10.1080/10447318.2021.1898851](https://doi.org/10.1080/10447318.2021.1898851).
- [120] A. D. Fraser, I. Branson, R. C. Hollett, C. P. Speelman, and S. L. Rogers, “Expressiveness of real-time motion captured avatars influences perceived animation realism and perceived quality of social interaction in virtual reality”, *Frontiers in Virtual Reality*, vol. Volume 3 - 2022, 2022. DOI: [10.3389/frvir.2022.981400](https://doi.org/10.3389/frvir.2022.981400).
- [121] S. L. Rogers, R. Broadbent, J. Brown, A. Fraser, and C. P. Speelman, “Realistic motion avatars are the future for social interaction in virtual reality”, *Frontiers in Virtual Reality*, vol. Volume 2 - 2021, 2022. DOI: [10.3389/frvir.2021.750729](https://doi.org/10.3389/frvir.2021.750729).
- [122] N. Xi, J. Chen, F. Gama, M. Riar, and J. Hamari, “The challenges of entering the mMtaverse: An experiment on the effect of extended reality on workload”, *Information Systems Frontiers*, vol. 25, no. 2, pp. 659–680, 2023. DOI: [10.1007/s10796-022-10244-x](https://doi.org/10.1007/s10796-022-10244-x).
- [123] G. Freeman, S. Zamanifard, D. Maloney, and A. Adkins, “My body, my avatar: How people perceive their avatars in social virtual reality”, in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*, 2020, pp. 1–8. DOI: [10.1145/3334480.3382923](https://doi.org/10.1145/3334480.3382923).
- [124] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo, “The effect of avatar appearance on social presence in an augmented reality remote collaboration”, in *Proc. of 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2019)*, 2019, pp. 547–556. DOI: [10.1109/VR.2019.8797719](https://doi.org/10.1109/VR.2019.8797719).
- [125] C.-H. Rhee and C. H. Lee, “Cartoon-like avatar generation using facial component matching”, *Int. J. of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 69–78, 2013. [Online]. Available: [https://gvpress.com/journals/IJMUE/vol8\\_no4/7.pdf](https://gvpress.com/journals/IJMUE/vol8_no4/7.pdf).
- [126] F. De Lorenzis, F. G. Praticcò, and F. Lamberti, “HCP-VR: Training first responders through a virtual reality application for hydrogeological risk management.”, in *Proc. of 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2 (HUCAPP 2022)*, 2022, pp. 273–280. DOI: [10.5220/0011007800003124](https://doi.org/10.5220/0011007800003124).
- [127] K. L. Nowak and J. Fox, “Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations”, *Review of Communication Research*, vol. 6, pp. 30–53, 2018. DOI: [10.12840/issn.2255-4165.2018.06.01.015](https://doi.org/10.12840/issn.2255-4165.2018.06.01.015).

- [128] G. C. Dobre, M. Wilczkowiak, M. Gillies, X. Pan, and S. Rintel, “Nice is different than good: Longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings”, in *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA ’22)*, 2022, pp. 1–7. DOI: [10.1145/3491101.3519628](https://doi.org/10.1145/3491101.3519628).
- [129] J. D. Hart, T. Piumsomboon, L. Lawrence, G. A. Lee, R. T. Smith, and M. Billinghamurst, “Demonstrating emotion sharing and augmentation in cooperative virtual reality games”, in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 2018, pp. 405–406. DOI: [10.1109/ISMAR-Adjunct.2018.00118](https://doi.org/10.1109/ISMAR-Adjunct.2018.00118).
- [130] P. Ekman, “Are there basic emotions?”, *Psychological Rev.*, vol. 99, no. 3, pp. 550–553, 1992. DOI: [10.1037/0033-295X.99.3.550](https://doi.org/10.1037/0033-295X.99.3.550).
- [131] P. Heidicker, E. Langbehn, and F. Steinicke, “Influence of avatar appearance on presence in social VR”, in *Proc. of 2017 IEEE Symposium on 3D User Interfaces (3DUI 2017)*, 2017, pp. 233–234. DOI: [10.1109/3DUI.2017.7893357](https://doi.org/10.1109/3DUI.2017.7893357).
- [132] O. Otto, D. Roberts, and R. Wolff, “A review on effective closely-coupled collaboration using immersive CVE’s”, in *Proc. of Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications (VRCIA ’06)*, 2006, pp. 145–154. DOI: [10.1145/1128923.1128947](https://doi.org/10.1145/1128923.1128947).
- [133] K. Kim, L. Boelling, S. Haesler, J. Bailenson, G. Bruder, and G. F. Welch, “Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR”, in *Proc. of 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2018)*, 2018, pp. 105–114. DOI: [10.1109/ISMAR.2018.00039](https://doi.org/10.1109/ISMAR.2018.00039).
- [134] A. Krekhov, S. Cmentowski, and J. Krüger, “The illusion of animal body ownership and its potential for virtual reality games”, in *Proc. of 2019 IEEE Conference on Games (CoG 2019)*, 2019, pp. 1–8. DOI: [10.1109/CIG.2019.8848005](https://doi.org/10.1109/CIG.2019.8848005).
- [135] B. Salem and N. Earle, “Designing a non-verbal language for expressive avatars”, in *Proc. of 3rd international conference on Collaborative virtual environments (CVE ’00)*, 2000, pp. 93–101. DOI: [10.1145/351006.351019](https://doi.org/10.1145/351006.351019).
- [136] J. N. Bailenson, N. Yee, D. Merget, and R. Schroeder, “The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction”, *Presence: Teleoperators and Virtual Environments*, vol. 15, no. 4, pp. 359–372, 2006. DOI: [10.1162/pres.15.4.359](https://doi.org/10.1162/pres.15.4.359).

- [137] I. Kotlyar and D. Ariely, “The effect of nonverbal cues on relationship formation”, *Computers in Human Behavior*, vol. 29, no. 3, pp. 544–551, 2013. DOI: [10.1016/j.chb.2012.11.020](https://doi.org/10.1016/j.chb.2012.11.020).
- [138] G. Bente and N. C. Krämer, “Virtual gestures: Embodiment and nonverbal behavior in computer-mediated communication”, *Face-to-face communication over the internet: Issues, research, challenges*, pp. 176–209, 2011. DOI: [10.1017/CB09780511977589.010](https://doi.org/10.1017/CB09780511977589.010).
- [139] S. J. Ahn and J. N. Bailenson, “Self-endorsing versus other-endorsing in virtual environments”, *J. of Advertising*, vol. 40, no. 2, pp. 93–106, 2011. DOI: [10.2753/JOA0091-3367400207](https://doi.org/10.2753/JOA0091-3367400207).
- [140] D. Roth et al., “Avatar realism and social interaction quality in virtual reality”, in *Proc. IEEE Virtual Reality*, 2016, pp. 277–278. DOI: [10.1109/VR.2016.7504761](https://doi.org/10.1109/VR.2016.7504761).
- [141] V. Kasapakis and E. Dzardanova, “Using high fidelity avatars to enhance learning experience in virtual learning environments”, in *Proc. of 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW 2021)*, 2021, pp. 645–646. DOI: [10.1109/VRW52623.2021.00205](https://doi.org/10.1109/VRW52623.2021.00205).
- [142] N. Hube, O. Lenz, L. Engeln, R. Groh, and M. Sedlmair, “Comparing methods for mapping facial expressions to enhance immersive collaboration with signs of emotion”, in *Proc. of 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct 2020)*, 2020, pp. 30–35. DOI: [10.1109/ISMAR-Adjunct51615.2020.00023](https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00023).
- [143] “Rootmotion”. [Online; accessed 10-July-2024]. [Online]. Available: <http://root-motion.com/>.
- [144] I. C. Campbell, *This VR doc shows how vibrant virtual life already is without Meta’s meddling*, 2022. [Online]. Available: <https://www.inverse.com/input/reviews/we-met-in-virtual-reality-documentary-shows-metaverse-vibrant-without-facebook>.
- [145] J. A. Russell, A. Weiss, and G. A. Mendelsohn, “Affect grid: A single-item scale of pleasure and arousal”, *J. of personality and social psychology*, vol. 57, no. 3, p. 493, 1989. DOI: [10.1037/0022-3514.57.3.493](https://doi.org/10.1037/0022-3514.57.3.493).
- [146] J. K. Burgoon and J. B. Walther, “Nonverbal expectancies and the evaluative consequences of violations”, *Human Communication Research*, vol. 17, no. 2, pp. 232–265, 1990. DOI: [10.1111/j.1468-2958.1990.tb00232.x](https://doi.org/10.1111/j.1468-2958.1990.tb00232.x).
- [147] S. Imaizumi, H. Saito, M. Kitajima, K. Kanari, and M. Sato, “Effects of hand color change on user’s psychological state during a pseudo handshake”, in *2022 International Conference on Cyberworlds (CW 2022)*, IEEE, 2022, pp. 126–129. DOI: [10.1109/CW55638.2022.00028](https://doi.org/10.1109/CW55638.2022.00028).



- [148] A. Haans and W. IJsselsteijn, “Mediated social touch: A review of current research and future directions”, *Virtual Reality*, vol. 9, pp. 149–159, 2006. DOI: [10.1007/s10055-005-0014-2](https://doi.org/10.1007/s10055-005-0014-2).
- [149] H. Nakanishi, K. Tanaka, and Y. Wada, “Remote handshaking: Touch enhances video-mediated social telepresence”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 2143–2152. DOI: [10.1145/2556288.2557169](https://doi.org/10.1145/2556288.2557169).
- [150] L. Lin et al., “The effect of hand size and interaction modality on the virtual hand illusion”, in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, IEEE, 2019, pp. 510–518. DOI: [10.1109/VR.2019.8797787](https://doi.org/10.1109/VR.2019.8797787).
- [151] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, “Statistical power analyses using g\*power 3.1: Tests for correlation and regression analyses”, *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–1160, 2009. DOI: [10.3758/BRM.41.4.1149](https://doi.org/10.3758/BRM.41.4.1149).
- [152] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Academic Press, 1977. DOI: [10.1016/B978-0-12-179060-8.50006-2](https://doi.org/10.1016/B978-0-12-179060-8.50006-2).
- [153] T. Peck and M. Gonzalez-Franco, “Avatar embodiment: A standardized questionnaire”, *Frontiers in Virtual Reality*, vol. 1, 2021. DOI: [10.3389/frvir.2020.575943](https://doi.org/10.3389/frvir.2020.575943).
- [154] T. W. Schubert, “The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness”, *Zeitschrift für Medienpsychologie*, vol. 15, no. 2, pp. 69–71, 2003. DOI: [10.1026/1617-6383.15.2.69](https://doi.org/10.1026/1617-6383.15.2.69).
- [155] C. Harms and F. A. Biocca, “Internal consistency and reliability of the networked minds social presence measure”, in *Seventh Annual International Workshop: Presence 2004*, 2004. [Online]. Available: [https://web.archive.org/web/2026/1/Harms\\_04\\_reliability\\_validity\\_social\\_presence\\_\(Biocca\).pdf](https://web.archive.org/web/2026/1/Harms_04_reliability_validity_social_presence_(Biocca).pdf).
- [156] J. Brooke, “Sus: A ‘quick and dirty’ usability scale”, in *Usability Evaluation in Industry*, 1996, p. 189. DOI: [10.1201/9781498710411-35](https://doi.org/10.1201/9781498710411-35).
- [157] R. S. Kalawsky, “VRUSE—a computerised diagnostic tool: For usability evaluation of virtual/synthetic environment systems”, *Applied Ergonomics*, vol. 30, no. 1, pp. 11–25, 1999. DOI: [10.1016/S0003-6870\(98\)00047-7](https://doi.org/10.1016/S0003-6870(98)00047-7).
- [158] S. Narayanan, N. Polys, and I. I. Bukvic, “Cinemacraft: Exploring fidelity cues in collaborative virtual world interactions”, *Virtual Reality*, vol. 24, no. 1, pp. 53–73, 2020. DOI: [10.1007/s10055-019-00382-0](https://doi.org/10.1007/s10055-019-00382-0).



- [159] G. Freeman, S. Zamanifard, D. Maloney, and A. Adkins, “My body, my avatar: How people perceive their avatars in social virtual reality”, in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20)*, ACM, 2020, pp. 1–8. DOI: [10.1145/3334480.3382923](https://doi.org/10.1145/3334480.3382923).
- [160] Y. Li, J. L. Fu, and N. S. Pollard, “Data-driven grasp synthesis using shape matching and task-based pruning”, *IEEE Trans. on Visualization and Computer Graphics*, vol. 13, no. 4, pp. 732–747, 2007. DOI: [10.1109/TVCG.2007.1033](https://doi.org/10.1109/TVCG.2007.1033).
- [161] S. Rezzonico, R. Boulic, Z. Huang, N. M. Thalmann, and D. Thalmann, “Consistent grasping in virtual environments based on the interactive grasping automata”, in *Virtual Environments '95*, M. Göbel, Ed., Springer Vienna, 1995, pp. 107–118. DOI: [10.1007/978-3-7091-9433-1\\_10](https://doi.org/10.1007/978-3-7091-9433-1_10).
- [162] C. Kyriltsias and D. Michael-Grigoriou, “Social interaction with agents and avatars in immersive virtual environments: A survey”, *Frontiers in Virtual Reality*, vol. 2, p. 786665, 2021. DOI: [10.3389/frvir.2021.786665](https://doi.org/10.3389/frvir.2021.786665).
- [163] W. Wan et al., “Learn to predict how humans manipulate large-sized objects from interactive motions”, *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4702–4709, 2022. DOI: [10.1109/LRA.2022.3151614](https://doi.org/10.1109/LRA.2022.3151614).
- [164] D. Vogt, S. Grehl, E. Berger, H. Ben Amor, and B. Jung, “A data-driven method for real-time character animation in human-agent interaction”, in *Intelligent Virtual Agents (IVA 2014)*, Springer International Publishing, 2014, pp. 463–476.
- [165] H. Schmidl and M. C. Lin, “Geometry-driven physical interaction between avatars and virtual environments”, *Computer Animation and Virtual Worlds*, vol. 15, no. 3-4, pp. 229–236, 2004. DOI: [10.1002/cav.25](https://doi.org/10.1002/cav.25).
- [166] H. Zhang, Y. Ye, T. Shiratori, and T. Komura, “ManipNet: Neural manipulation synthesis with a hand-object spatial representation”, *ACM Trans. on Graphics*, vol. 40, no. 4, 2021. DOI: [10.1145/3450626.3459830](https://doi.org/10.1145/3450626.3459830).
- [167] O. Taheri et al., “Grip: Generating interaction poses using spatial cues and latent consistency”, in *Proc. of Int. Conference on 3D Vision (3DV)*, 2024, pp. 933–943. DOI: [10.1109/3DV62453.2024.00064](https://doi.org/10.1109/3DV62453.2024.00064).
- [168] S. Choi, S. Hong, K. Cho, C. Kim, and J. Noh, “Online avatar motion adaptation to morphologically-similar spaces”, *Computer Graphics Forum*, vol. 42, no. 2, pp. 13–24, 2023. DOI: <https://doi.org/10.1111/cgf.14740>.
- [169] A. L. Dueren, A. Vafeiadou, C. Edgar, and M. J. Banissy, “The influence of duration, arm crossing style, gender, and emotional closeness on hugging behaviour”, *Acta Psychologica*, vol. 221, p. 103441, 2021. DOI: <https://doi.org/10.1016/j.actpsy.2021.103441>.

- [170] K. Floyd, “All touches are not created equal: Effects of form and duration on observers’ interpretations of an embrace”, *J. of Nonverbal Behavior*, vol. 23, no. 4, pp. 283–299, 1999. DOI: [10.1023/A:1021602926270](https://doi.org/10.1023/A:1021602926270).
- [171] L. Wagnerberger, D. Runde, M. T. Lafci, D. Przewozny, S. Bosse, and P. Chojewski, “Inverse kinematics for full-body self representation in VR-based cognitive rehabilitation”, in *Proc. IEEE Int. Symposium on Multimedia*, 2021, pp. 123–129. DOI: [10.1109/ISM52913.2021.00029](https://doi.org/10.1109/ISM52913.2021.00029).
- [172] B. N. Oreshkin, F. Bocquet, F. G. Harvey, B. Raitt, and D. Laflamme, “ProtoRes: Proto-residual network for pose authoring via learned IK”, *arXiv*, 2022. DOI: [10.48550/arXiv.2106.01981](https://doi.org/10.48550/arXiv.2106.01981).
- [173] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, “On the continuity of rotation representations in neural networks”, in *Proc. of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019)*, 2019, pp. 5738–5746. DOI: [10.1109/CVPR.2019.00589](https://doi.org/10.1109/CVPR.2019.00589).
- [174] L. Hoyet, A.-H. Olivier, R. Kulpa, and J. Pettr  , “Perceptual effect of shoulder motions on crowd animations”, *ACM Trans. on Graphics*, vol. 35, no. 4, 2016. DOI: [10.1145/2897824.2925931](https://doi.org/10.1145/2897824.2925931).
- [175] A. Weiss and C. Bartneck, “Meta analysis of the usage of the godspeed questionnaire series”, in *Proc. of 24th IEEE Int. Symposium on Robot and Human Interactive Communication*, 2015, pp. 381–388. DOI: [10.1109/ROMAN.2015.7333568](https://doi.org/10.1109/ROMAN.2015.7333568).
- [176] J. N. Bailenson, A. C. Beall, J. Loomis, J. Blascovich, and M. Turk, “Transformed social interaction, augmented gaze, and social influence in immersive virtual environments”, *Human Communication Research*, vol. 31, no. 4, pp. 511–537, 2005. DOI: [10.1111/j.1468-2958.2005.tb00881.x](https://doi.org/10.1111/j.1468-2958.2005.tb00881.x).
- [177] K. L. Poole and H. A. Henderson, “Shyness, self-focused attention, and behavioral mimicry during social interaction”, *J. of Research in Personality*, vol. 98, p. 104225, 2022. DOI: [10.1016/j.jrp.2022.104225](https://doi.org/10.1016/j.jrp.2022.104225).
- [178] A. Zheleva et al., “The impact of eye gaze on social interactions of females in virtual reality: The mediating role of the uncanniness of avatars and the moderating role of task type”, *Heliyon*, vol. 9, no. 10, 2023. DOI: [10.1016/j.heliyon.2023.e20165](https://doi.org/10.1016/j.heliyon.2023.e20165).
- [179] T. Wakabayashi, Y. Okada, and K. Zempo, “Effect of salesperson avatar automatically mimicking customer’s nodding on the enjoyment of conversation in virtual environments”, in *Proc. of the Augmented Humans International Conference 2023 (AHs 2023)*, ACM, 2023, pp. 334–337. DOI: [10.1145/3582700.3583711](https://doi.org/10.1145/3582700.3583711).

- [180] W. Sato and S. Yoshikawa, “Spontaneous facial mimicry in response to dynamic facial expressions”, *Cognition*, vol. 104, no. 1, pp. 1–18, 2007. DOI: [10.1016/j.cognition.2006.05.001](https://doi.org/10.1016/j.cognition.2006.05.001).
- [181] I. Wang, R. Calvo, H. Wang, and J. Ruiz, “Stop copying me: Evaluating non-verbal mimicry in embodied motivational agents”, in *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents (IVA 2023)*, ACM, 2023, pp. 1–4. DOI: [10.1145/3570945.3607322](https://doi.org/10.1145/3570945.3607322).
- [182] N. Aburumman, M. Gillies, J. A. Ward, and A. F. C. Hamilton, “Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions”, *Int. J. of Human-Computer Studies*, vol. 164, p. 102819, 2022. DOI: [10.1016/j.ijhcs.2022.102819](https://doi.org/10.1016/j.ijhcs.2022.102819).
- [183] J. Hale and A. F. d. C. Hamilton, “Testing the relationship between mimicry, trust and rapport in virtual reality conversations”, *Scientific Reports*, vol. 6, p. 35295, 2016. DOI: [10.1038/srep35295](https://doi.org/10.1038/srep35295).
- [184] P. Caserman, A. Garcia-Agundez, and S. Göbel, “A survey of full-body motion reconstruction in immersive virtual reality applications”, *IEEE Trans. on Visualization and Computer Graphics*, vol. 26, no. 10, pp. 3089–3108, 2020. DOI: [10.1109/TVCG.2019.2912607](https://doi.org/10.1109/TVCG.2019.2912607).
- [185] Meta, *Use Natural Facial Expressions on Meta Quest Pro*, [accessed 14 September 2024], 2022. [Online]. Available: <https://www.meta.com/en-us/help/quest/articles/getting-started/getting-started-with-quest-pro/facial-expressions/>.
- [186] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand, “The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry”, *J. of Nonverbal Behavior*, vol. 27, pp. 145–162, 2003. DOI: [10.1023/A:1025389814290](https://doi.org/10.1023/A:1025389814290).
- [187] T. Weronika and P. Brzóska, “How behavioral mimicry influences impression formation processes: A positive impression bias”, *Polskie Forum Psychologiczne*, vol. 27, no. 1, pp. 5–22, 2022. DOI: [10.34767/PFP.2022.01.01](https://doi.org/10.34767/PFP.2022.01.01).
- [188] T. L. Chartrand and J. L. Lakin, “The antecedents and consequences of human behavioral mimicry”, *Ann. Rev. of Psychology*, vol. 64, pp. 285–308, 2013. DOI: [10.1146/annurev-psych-113011-143754](https://doi.org/10.1146/annurev-psych-113011-143754).
- [189] J. L. Lakin and T. L. Chartrand, “Using nonconscious behavioral mimicry to create affiliation and rapport”, *Psychological Science*, vol. 14, no. 4, pp. 334–339, 2003. DOI: [10.1111/1467-9280.14481](https://doi.org/10.1111/1467-9280.14481).
- [190] R. B. Van Baaren, R. W. Holland, K. Kawakami, and A. Van Knippenberg, “Mimicry and prosocial behavior”, *Psychological Science*, vol. 15, no. 1, pp. 71–74, 2004. DOI: [10.1111/j.0963-7214.2004.01501012.x](https://doi.org/10.1111/j.0963-7214.2004.01501012.x).

- [191] W. W. Maddux, E. Mullen, and A. D. Galinsky, “Chameleons bake bigger pies and take bigger pieces: Strategic behavioral mimicry facilitates negotiation outcomes”, *J. of Experimental Social Psychology*, vol. 44, no. 2, pp. 461–468, 2008. DOI: [10.1016/j.jesp.2007.02.003](https://doi.org/10.1016/j.jesp.2007.02.003).
- [192] N. Guéguen, A. Martin, S. Meineri, and J. Simon, “Using mimicry to elicit answers to intimate questions in survey research”, *Field methods*, vol. 25, no. 1, pp. 47–57, 2013. DOI: [10.1177/1525822X12449710](https://doi.org/10.1177/1525822X12449710).
- [193] T. L. Chartrand and J. A. Bargh, “The chameleon effect: The perception–behavior link and social interaction.”, *J. of Personality and Social Psychology*, vol. 76, no. 6, p. 893, 1999. DOI: [10.1037/0022-3514.76.6.893](https://doi.org/10.1037/0022-3514.76.6.893).
- [194] M. Stel, S. Rispens, M. Leliveld, and A. M. Lokhorst, “The consequences of mimicry for prosocials and proselves: Effects of social value orientation on the mimicry–liking link”, *European J. of Social Psychology*, vol. 41, no. 3, pp. 269–274, 2011. DOI: [10.1002/ejsp.790](https://doi.org/10.1002/ejsp.790).
- [195] M. Drury and H. van Swol, “Are people who mimic others perceived as more friendly, likeable, persuasive, and knowledgeable?”, in *Proc. \ NCA*, 2005. [Online]. Available: <https://api.semanticscholar.org/CorpusID:146382743>.
- [196] R. B. Van Baaren, R. W. Holland, B. Steenaert, and A. Van Knippenberg, “Mimicry for money: Behavioral consequences of imitation”, *J. of Experimental Social Psychology*, vol. 39, no. 4, pp. 393–398, 2003. DOI: [10.1016/S0022-1031\(03\)00014-3](https://doi.org/10.1016/S0022-1031(03)00014-3).
- [197] U. Hess and A. Fischer, “Emotional mimicry as social regulation”, *Personality and Social Psychology Rev.*, vol. 17, no. 2, pp. 142–157, 2013. DOI: [10.1177/1088868312472607](https://doi.org/10.1177/1088868312472607).
- [198] J. N. Bailenson and N. Yee, “Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments”, *Psychological Science*, vol. 16, no. 10, pp. 814–819, 2005. DOI: [10.1111/j.1467-9280.2005.01619.x](https://doi.org/10.1111/j.1467-9280.2005.01619.x).
- [199] E. L. Fasya, E. van den Bos, D. K. Heylen, and M. E. Kret, “Smile mimicry smoothens human-virtual human interactions”, *Int. J. of Human-Computer Studies*, vol. 183, pp. 1–11, 2024. DOI: [10.1016/j.ijhcs.2023.103182](https://doi.org/10.1016/j.ijhcs.2023.103182).
- [200] F. M. F. Verberne, J. Ham, A. Ponnada, and C. J. H. Midden, “Trusting digital chameleons: The effect of mimicry by a virtual social agent on user trust”, in *booktitle="Persuasive Technology",E*, S. Berkovsky and J. Freyne, Eds., vol. 7822, Springer, Berlin, 2013, pp. 234–245. DOI: [10.1007/978-3-642-37157-8\\_28](https://doi.org/10.1007/978-3-642-37157-8_28).

- [201] R. Hoegen, J. van der Schalk, G. Lucas, and J. Gratch, “The impact of agent facial mimicry on social behavior in a prisoner’s dilemma”, in *Proc. of the 18th International Conference on Intelligent Virtual Agents (IVA 2018)*, ACM, 2018, pp. 275–280. DOI: [10.1145/3267851.3267911](https://doi.org/10.1145/3267851.3267911).
- [202] D. S. Choi, J. Park, M. Loeser, and K. Seo, “Improving counseling effectiveness with virtual counselors through nonverbal compassion involving eye contact, facial mimicry, and head-nodding”, *Scientific Reports*, vol. 14, no. 1, pp. 1–12, 2024. DOI: [10.1038/s41598-023-51115-y](https://doi.org/10.1038/s41598-023-51115-y).
- [203] *Evaluation of the effect of mimicry on facial expression in Avatar-Mediated Communication*, vol. ALIFE 2021: The 2021 Conference on Artificial Life, 2021, p. 48. DOI: [10.1162/isal\\_a\\_00394](https://doi.org/10.1162/isal_a_00394).
- [204] K. Suzuki et al., “Faceshare: Mirroring with pseudo-smile enriches video chat communications”, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI 2017)*, ACM, 2017, pp. 5313–5317, ISBN: 9781450346559. DOI: [10.1145/3025453.3025574](https://doi.org/10.1145/3025453.3025574).
- [205] D. Roth, M. E. Latoschik, K. Vogeley, and G. Bente, “Hybrid avatar-agent technology – a conceptual step towards mediated “social” virtual reality and its respective challenges”, *i-com*, vol. 14, no. 2, pp. 107–114, 2015. DOI: [doi: 10.1515/icom-2015-0030](https://doi.org/10.1515/icom-2015-0030).
- [206] R. Somarathna, T. Bednarz, and G. Mohammadi, “Virtual reality for emotion elicitation—a review”, *IEEE Tran. on Affective Computing*, vol. 14, no. 4, pp. 2626–2645, 2022. DOI: [10.1109/TAFFC.2022.3181053](https://doi.org/10.1109/TAFFC.2022.3181053).
- [207] F. Huo, R. Gao, C. Sun, and G. Hou, “Age differences in hazard perception of drivers: The roles of emotion”, *Front. in Psychology*, vol. 13, pp. 1–11, 2022. DOI: [10.3389/fpsyg.2022.867673](https://doi.org/10.3389/fpsyg.2022.867673).
- [208] R. Lipson-Smith, J. Bernhardt, E. Zamuner, L. Churilov, N. Busietta, and D. Moratti, “Exploring colour in context using Virtual Reality: Does a room change how you feel?”, *Virtual Reality*, vol. 25, no. 1, pp. 631–645, 2021. DOI: [10.1007/s10055-020-00491-x](https://doi.org/10.1007/s10055-020-00491-x).
- [209] M. M. Bradley and P. J. Lang, “International affective picture system”, in *Encyclopedia of Personality and Individual Differences*, V. Zeigler-Hill and T. K. Shackelford, Eds., Springer Int. Publishing, 2017, pp. 1–4. DOI: [10.1007/978-3-319-28099-8\\_42-1](https://doi.org/10.1007/978-3-319-28099-8_42-1).
- [210] Y. Ge, G. Zhao, Y. Zhang, R. J. Houston, and J. Song, “A standardised database of chinese emotional film clips”, *Cognition and Emotion*, vol. 33, no. 5, pp. 976–990, 2019. DOI: [10.1080/02699931.2018.1530197](https://doi.org/10.1080/02699931.2018.1530197).

- [211] L. O. H. KroczeK and A. Mühlberger, “Time to smile: How onset asynchronies between reciprocal facial expressions influence the experience of responsiveness of a virtual agent”, *J. of Nonverbal Behavior*, 2023. DOI: [10.1007/s10919-023-00430-z](https://doi.org/10.1007/s10919-023-00430-z).
- [212] T. L. Kraft-Feil, R. E. Ingram, C. Gorey, J. H. Luu, M. P. Cross, and S. D. Pressman, “The association of negative mood with automatic and effortful facial expression mimicry”, *Frontiers in Psychology*, vol. 14, pp. 1–14, 2023. DOI: [10.3389/fpsyg.2023.1056535](https://doi.org/10.3389/fpsyg.2023.1056535).
- [213] A. Orlowska, M. Rychlowska, and E. Krumhuber, “The interplay between mimicry and social context in facial expression perception”, in *Handbook on Facial Expression of Emotion - Vol. 3*. Leya, 2020, pp. 121–153. [Online]. Available: <https://pure.qub.ac.uk/en/publications/the-interplay-between-mimicry-and-social-context-in-facial-express>.
- [214] U. Dimberg, M. Thunberg, and K. Elmehed, “Unconscious facial reactions to emotional facial expressions”, *Psychological Science*, vol. 11, no. 1, pp. 86–89, 2000. DOI: [10.1111/1467-9280.00221](https://doi.org/10.1111/1467-9280.00221).
- [215] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, *Emotional Contagion*. Cambridge University Press, 1994. DOI: [10.1017/CB09781139174138](https://doi.org/10.1017/CB09781139174138).
- [216] P. Ekman and W. V. Friesen, “Facial action coding system”, *Environmental Psychology & Nonverbal Behavior*, pp. 1–6, 1978. DOI: [doi.org/10.1037/t27734-000](https://doi.org/10.1037/t27734-000).
- [217] B. Farnsworth, *Facial Action Coding System (FACS) – A Visual Guidebook - iMotions*, <https://imotions.com/blog/learning/research-fundamentals/facial-action-coding-system/>, [accessed 04 January 2025].
- [218] J. N. Bailenson, N. Yee, K. Patel, and A. C. Beall, “Detecting digital chameleons”, *Computers in Human Behavior*, vol. 24, no. 1, pp. 66–87, 2008. DOI: [10.1016/j.chb.2007.01.015](https://doi.org/10.1016/j.chb.2007.01.015).
- [219] B. Kidwell, V. Lopez-Kidwell, C. Blocker, and E. M. Mas, “Birds of a feather feel together: Emotional ability similarity in consumer interactions”, *J. of Consumer Research*, vol. 47, no. 2, pp. 215–236, 2020. DOI: [10.1093/jcr/ucaa011](https://doi.org/10.1093/jcr/ucaa011).
- [220] C. Oh, J. Bailenson, and G. Welch, “A systematic review of social presence: Definition, antecedents, and implications”, *Frontiers in Robotics and AI*, vol. 5, pp. 1–35, 2018. DOI: [10.3389/frobt.2018.00114](https://doi.org/10.3389/frobt.2018.00114).
- [221] U. Hess, M. Rychlowska, and T. e. a. StorZ, “Bridging the empathy gap: Or not? reactions to ingroup and outgroup facial expressions”, *J. of Cultural Cognitive Science*, 2022. DOI: [10.1007/s41809-022-00097-w](https://doi.org/10.1007/s41809-022-00097-w).



- [222] J. Dong, K. Ota, and M. Dong, “User experience of different groups in social vr applications: An empirical study based on user reviews”, *IEEE Trans. on Computational Social Systems*, vol. 11, no. 6, pp. 7133–7145, 2024. DOI: [10.1109/TCSS.2024.3416208](https://doi.org/10.1109/TCSS.2024.3416208).
- [223] H. R. Hatfield, S. J. (Ahn, M. Klein, and K. L. Nowak, “Confronting whiteness through virtual humans: A review of 20 years of research in prejudice and racial bias using virtual environments”, *J. of Computer-Mediated Communication*, vol. 27, no. 6, 2022. DOI: [10.1093/jcmc/zmac016](https://doi.org/10.1093/jcmc/zmac016).
- [224] L. Baccon, C. Elodie, and H. MacDougall, “Virtual reality for teletherapy: Avatars may combine the benefits of face-to-face communication with the anonymity of online text-based communication”, *Cyberpsychology, Behavior, and Social Networking*, vol. 22, 2018. DOI: [10.1089/cyber.2018.0247](https://doi.org/10.1089/cyber.2018.0247).
- [225] A. Bradbury, S. Schertz, and E. Wiebe, “How does virtual reality compare? the effects of avatar appearance and medium on self-disclosure”, *Proc. of the Human Factors and Ergonomics Society Annual Meeting (HFES 2022)*, vol. 66, no. 1, pp. 2036–2040, 2022. DOI: [10.1177/1071181322661123](https://doi.org/10.1177/1071181322661123).
- [226] M. Gnacek et al., “AVDOS-VR: Affective video database with physiological signals and continuous ratings collected remotely in VR”, *Scientific Data*, vol. 11, pp. 1–X, 2024. DOI: [10.1038/s41597-024-02953-6](https://doi.org/10.1038/s41597-024-02953-6).
- [227] M. Jeon and W. Zhang, “Sadder but wiser? effects of negative emotions on risk perception, driving performance, and perceived workload”, *Proc. of the Human Factors and Ergonomics Society Annual Meeting (HFES 2013)*, vol. 57, no. 1, pp. 1849–1853, 2013. DOI: [10.1177/1541931213571413](https://doi.org/10.1177/1541931213571413).
- [228] C. Kruzic, D. Kruzic, F. Herrera, and J. Bailenson, “Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments”, *Scientific Reports*, vol. 10, p. 20626, 2020. DOI: [10.1038/s41598-020-76672-4](https://doi.org/10.1038/s41598-020-76672-4).
- [229] C. R. Agnew, “Feeling close to a crab-thing in virtual reality: Does avatar appearance always matter in forming meaningful connections? A case study”, *Front. in Virtual Reality*, vol. 3, pp. 1–10, 2022, ISSN: 2673-4192. DOI: [10.3389/frvir.2022.889247](https://doi.org/10.3389/frvir.2022.889247).
- [230] A. Aron, E. Melinat, E. N. Aron, R. D. Vallone, and R. J. Bator, “The experimental generation of interpersonal closeness: A procedure and some preliminary findings”, *Personality and Social Psychology Bulletin*, 1997. DOI: [10.1177/0146167297234003](https://doi.org/10.1177/0146167297234003).
- [231] F. Lang, D. John, O. Lüdtke, J. Schupp, and G. Wagner, “Short assessment of the big five: Robust across survey methods except telephone interviewing”, *Behavior Research Methods*, vol. 43, pp. 548–67, 2011. DOI: [10.3758/s13428-011-0066-z](https://doi.org/10.3758/s13428-011-0066-z).



- [232] J. C. McCroskey and T. A. McCain, “The measurement of interpersonal attraction”, *Speech Monographs*, vol. 41, no. 6, pp. 261–266, 1974. DOI: [10.1080/03637757409375845](https://doi.org/10.1080/03637757409375845).
- [233] D. Davis and W. T. Perkowski, “Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction.”, *J. of Personality and Social Psychology*, vol. 37, no. 4, p. 534, 1979. DOI: [10.1037/0022-3514.37.4.534](https://doi.org/10.1037/0022-3514.37.4.534).
- [234] M. Burgoon, “Amount of conflicting information in a group discussion and tolerance for ambiguity as predictors of task attractiveness”, *Communications Monographs*, vol. 38, no. 2, pp. 121–124, 1971. DOI: [10.1080/03637757109375697](https://doi.org/10.1080/03637757109375697).
- [235] P. Kullmann, T. Menzel, M. Botsch, and M. E. Latoschik, “An evaluation of other-avatar facial animation methods for social VR”, in *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA 2023)*, ACM, 2023. DOI: [10.1145/3544549.3585617](https://doi.org/10.1145/3544549.3585617).
- [236] G. F. Reed, F. Lynn, and B. D. Meade, “Use of coefficient of variation in assessing variability of quantitative assays”, *Clinical and Vaccine Immunology*, vol. 9, no. 6, pp. 1235–1239, 2002. DOI: [10.1128/CDLI.9.6.1235-1239.2002](https://doi.org/10.1128/CDLI.9.6.1235-1239.2002).
- [237] R. E. Maurer and J. H. Tindall, “Effect of postural congruence on client’s perception of counselor empathy”, *J. of Counseling Psychology*, vol. 30, no. 2, pp. 158–163, 1983. DOI: [10.1037/0022-0167.30.2.158](https://doi.org/10.1037/0022-0167.30.2.158).
- [238] P. A. G. Forbes, X. Pan, and A. F. de C. Hamilton, “Reduced mimicry to virtual reality avatars in autism spectrum disorder”, *J. of Autism and Developmental Disorders*, vol. 46, no. 12, pp. 3788–3797, 2016. DOI: [10.1007/s10803-016-2930-2](https://doi.org/10.1007/s10803-016-2930-2).
- [239] S. A. M. Ali and S. Ali, “The use of non-verbal communication in the classroom”, in *Proc. \ ICFLTAL*, 2011, pp. 1096–1099. DOI: [10.1051/e3sconf/202453800001](https://doi.org/10.1051/e3sconf/202453800001).
- [240] D. Hepperle, C. F. Purps, J. Deuchler, and M. Wölfel, “Aspects of visual avatar appearance: Self-representation, display type, and uncanny valley”, *The Visual Computer*, vol. 38, no. 4, pp. 1227–1244, 2022. DOI: [10.1007/s00371-021-02151-0](https://doi.org/10.1007/s00371-021-02151-0).
- [241] K. Prehn et al., “The neural correlates of emotion alignment in social interaction”, *Social Cognitive and Affective Neuroscience*, vol. 10, no. 3, pp. 435–443, 2014. DOI: [10.1093/scan/nsu066](https://doi.org/10.1093/scan/nsu066).
- [242] S. Bai et al., “Universal facial encoding of codec avatars from VR headsets”, *ACM Trans. Graphics*, vol. 43, no. 4, pp. 1–22, 2024. DOI: [10.1145/3658234](https://doi.org/10.1145/3658234).

- [243] A. Kolesnichenko, J. McVeigh-Schultz, and K. Isbister, “Understanding emerging design practices for avatar systems in the commercial social VR ecology”, in *Proc. of the 2019 Designing Interactive Systems Conference (DIS 2019)*, ACM, 2019, pp. 241–252. DOI: [10.1145/3322276.3322352](https://doi.org/10.1145/3322276.3322352).
- [244] C. Dubosc, G. Gorisse, T. Lefrou, O. Christmann, and S. Richir, “Effect of avatar stylization and facial expression intensity in virtual interactions”, *Virtual Reality*, vol. 29, no. 163, pp. 1–17, 2025. DOI: [10.1007/s10055-025-01238-6](https://doi.org/10.1007/s10055-025-01238-6).
- [245] J. Hartbrich, F. Weidner, C. Kunert, A. Raake, W. Broll, and S. Arévalo Arboleda, “Eye and face tracking in VR: Avatar embodiment and enfacement with realistic and cartoon avatars”, in *Proc. of the Int. Conference on Mobile and Ubiquitous Multimedia (MUM 2023)*, ACM, 2023, pp. 1–9. DOI: [10.1145/3626705.3627793](https://doi.org/10.1145/3626705.3627793).
- [246] S. A. Nijman, W. Veling, B. L. Lestestuiver, and M. van der Gaag, “Virtual reality facial emotion recognition in social environments: An eye-tracking study”, *Internet Interventions*, vol. 25, p. 100390, 2021. DOI: [10.1016/j.invent.2021.100390](https://doi.org/10.1016/j.invent.2021.100390).
- [247] J. M. Fort et al., “Facial expression recognition in virtual reality environments: Challenges and opportunities”, *Frontiers in Psychology*, vol. 14, p. 1280136, 2023. DOI: [10.3389/fpsyg.2023.1280136](https://doi.org/10.3389/fpsyg.2023.1280136).
- [248] M. Mori, K. F. MacDorman, and N. Kageki, “The uncanny valley [from the field]”, *IEEE Robotics & Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012. DOI: [10.1109/MRA.2012.2192811](https://doi.org/10.1109/MRA.2012.2192811).
- [249] V. Schwind, K. Wolf, and N. Henze, “Avoiding the uncanny valley in virtual character design”, *Interactions*, vol. 25, no. 5, pp. 45–49, 2018. DOI: [10.1145/3236673](https://doi.org/10.1145/3236673).
- [250] Y. Li et al., “Gesture interaction in virtual reality”, *Virtual Reality & Intelligent Hardware*, vol. 1, no. 1, pp. 84–112, 2019. DOI: [10.3724/SP.J.2096-5796.2019.0006](https://doi.org/10.3724/SP.J.2096-5796.2019.0006).
- [251] G. Caggianese, L. Gallo, and P. Neroni, “Evaluation of spatial interaction techniques for virtual heritage applications”, in *Proc. of the 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2019)*, IEEE, 2019, pp. 399–405. DOI: [10.1109/SITIS.2019.00068](https://doi.org/10.1109/SITIS.2019.00068).
- [252] N. Haas et al., “Haptics in social interaction with agents and avatars in virtual reality: A systematic review”, *Virtual Reality*, 2024. DOI: [10.1007/s10055-024-01060-6](https://doi.org/10.1007/s10055-024-01060-6).

- [253] D. Roth, C. Kleinbeck, T. Feigl, C. Mutschler, and M. E. Latoschik, “Beyond replication: Augmenting social behaviors in multi-user virtual realities”, in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, IEEE, 2018, pp. 215–222. DOI: [10.1109/VR.2018.8448293](https://doi.org/10.1109/VR.2018.8448293).
- [254] K. Kilteni, R. Groten, and M. Slater, “The sense of embodiment in virtual reality”, *Presence: Teleoperators and Virtual Environments*, vol. 21, no. 4, pp. 373–387, 2012. DOI: [10.1162/PRES\\_a\\_00124](https://doi.org/10.1162/PRES_a_00124).

This Ph.D. thesis has been typeset by means of the T<sub>E</sub>X-system facilities. The typesetting engine was pdfL<sup>A</sup>T<sub>E</sub>X. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete T<sub>E</sub>X-system installation.