Development of Human-Computer Interaction for Holographic AIs

Xinyu Huang

Thesis Submitted to the Open University for the Degree of Doctor of Philosophy

March 2023

Abstract

Virtual humans and embodied conversational agents play diverse roles in real life, including game characters, chatbots, and teachers. In Augmented Reality (AR), such agents are capable of interacting with the real world. To distinguish between both types of virtual agents, AR agents were conceptually redefined as "holographic Artificial Intelligences (AIs)". Holographic AIs are embodied virtual agents interacting with real objects in Augmented Reality (AR), and can respond to events both in virtual and real environments. This thesis provides a comprehensive investigation into holographic AIs, spanning from their design to their user experience.

The purpose of this thesis is to investigate the creation and use of holographic AIs, by creating specific holographic AIs, and then examining how users perceive such entities in order to contribute to the improvement of the user experience. As a result, this thesis explores the design space for and methods for creating holographic AIs, proposing the novel PICS model which include the dimensions of persona, intelligence, conviviality, and senses.

Following the PICS model, a set of holographic AIs are designed by using a method of semi-automatic reconstruction. An AI that resembles a human being in appearance and behaviour is endowed with multimodal interactions capable of creating the illusion of physicality. The initial proposed model is then refined based on the experience of creation.

Basic body language gestures, such as nodding and opening the arms, are insufficient to engage users, particularly when it comes to intelligent tutoring systems. Therefore, this thesis specifically focuses on an open problem, the generation of re-usable standard instructional gestures. In an experiment, key instructional movements that can be employed by holographic AIs were identified and extracted as animations. The hitherto known range of representational gestures is, epistemologically, further expanded by transformational and imitation gestures, which show how humans manipulate spatio-motor information and characterise posture using hand motion. Therefore, the model can be extended to describe the holographic AI's behaviour.

Moreover, in order to assess the empirical validity of holographic AIs, this research explores learners' trustworthiness towards this novel technology - as a key criterion for efficacy of this AI approach. Trust and trustworthiness, in terms of holographic AIs, refers to a mindset that aids users in achieving objectives based on good intentions. Young learners' perception of trust is largely influenced by affective aspects of trust, determined by how emotionally responsive a holographic AI is.

These findings contribute to the design of personal holographic AIs that can perform a series of meaningful gestures that engage the learner's attention for learning, which in turn fosters a reliable and trustworthy relationship. Both experiments are able to extend elements by adding gestures and holistic perception to this model.

To my parents, family, and friends

Acknowledgements

First, I am deeply grateful to my project supervisors, Prof. Fridolin Wild and Prof. Denise Whitelock, for their constant support, help and patience throughout my PhD study journey. I am grateful to Prof. Wild for his unwavering and invaluable guidance, encouragement and feedback throughout my research. His expertise and enthusiasm have inspired me, helped me to open my eyes, and taught me to be a positive person. Without his compassion and guidance, I doubt whether I would have completed my doctoral studies. I feel so fortunate to have been given the opportunity to work with him during this critical stage in my academic advancement.

I would also like to thank my second supervisor, Prof. Denise Whitlock, for her unwavering and insightful feedback and assistance. Her constructive suggestions have greatly contributed to the rigour of my research.

I am also grateful for the companionship of family and friends throughout my studies. I am indebted to my parents and even my cat for their affection and support during my studies in the UK.

I would also like to express my gratitude to my roommate, PianPian Zhao. She looked after me during the dispiriting times, and gave me invaluable suggestions relating to statistics.

I would like to thank Prof. John Mason and Prof. Anne Watson, who provided invaluable and professional suggestions in designing teaching content.

In addition, I am thankful to the participants who took part in my studies. Their willingness to share their experiences and perspectives has been instrumental to the success of this research.

Finally, I would like to thank the Institute of Educational Technology and the Faculty of Wellbeing, Education and Language Studies at the Open University for providing such a welcoming and academically stimulating environment.

Declaration of Authorship

I declare that this thesis has been composed solely by myself and that it has not been submitted, either in whole or in part, in any previous application for a degree. Except where otherwise acknowledged, the work presented is entirely my own.

References to Relevant Works:

Huang, X., Twycross, J. and Wild, F. (2019) ‘A process for the semi-automated generation of life-sized, interactive 3D character models for holographic projection’, 2019 International Conference on 3D Immersion (IC3D). 11 December. Brussels, Belgium, pp. 1–8.

Huang, X., Wild, F. and Whitelock, D. (2021) ‘Design dimensions for holographic intelligent agents: a comparative analysis’, Proceedings of the 1st International Workshop on Multimodal Artificial Intelligence in- Education (MAIED 2021). Utrecht, 10-14 June. Aachen: CEUR, pp. 17-26.

Huang, X. and Wild, F. (2021) ‘A new metric scale for measuring trust towards holographic intelligent agents’, Proceedings of the 34th British HCI Workshop and Doctoral Consortium 2021. Virtual Event, 20-21 July. Swindon: BCS Learning and Development Ltd, pp. 1-8.

Table of Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence
AR	Augmented Reality
ASD	Autism Spectrum Disorder
IoT	Internet of things
IT	Information technology
HCI	Human-computer interaction
Hypothesis	H
R&D	Research and development
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RNN	Recurrent Neutral Network
RQ	Research question
MR	Mixed Reality
UIs	User interfaces
VR	Virtual Reality
Woz	Wizard-of-Oz
3D	Three-dimensional

Chapter 1 Introduction

1.1 Introduction and Overview

The technology of Augmented Reality (AR) has developed rapidly over past decades, becoming increasingly integrated into people's lives. AR can be defined as an integration of physical and digital information that superimposes in real-time three-dimensional (3D) computer-generated objects into the user's interactive real world surrounding. Therefore, AR can be described as "a midground between synthetic and real environment" (Herpich, Guarese and Tarouco, 2017). The main difference between AR and Virtual Reality (VR) is that VR only observes the virtual environment, whereas AR technology yields a sense of coexistence, i.e. occupies a mixed environment. This technology relies on multimedia, visual elements, real-time tracking, and registration using optical cameras, inertia measurement units, and other sensors (Cheng et al., 2019).

However, the visual graphics that are generated by AR are different from that of hologram technology. The latter relies on light diffraction emitted from 3D objects using laser beams to record light patterns into thin film (Elmahal et al., 2020; Shimobaba et al., 2022) such that 3D graphics later to be observed by the naked eye. Whereas the holographic in AR refers to the display consisting of miniature holograms in the lens, creating computer-generated information using wearable smart glasses.

One weakness of hologram technology is that the 3D graphic lacks interactivity, and rely on other, additional equipment to receive information from the real world, such as voice recognition for speech interaction. The advantage of modern AR technology is the ability to recognize real-world spaces, sometimes referred to as 'marker-less' AR, which means it not only can be utilised for manipulation, but also serves the automatically and stable placement in suitable positions. For example, spatial mapping is one of the important features of AR, which can identify and map a 3D mesh surfaces by scanning the user-surrounding interaction space. Therefore, 3D objects can blend in with physical objects in order to provide an enhanced sense of physicality. This cannot be conducted by the classic hologram technology, that is used, for example, on the silvery credit card logos.

The application of 3D computer-generated graphics in AR is not limited to mere visual enjoyment, they also can be for user interfaces (UIs) capable of generating interactivity, and playing the role of embodied assistants that directly provide services. Such virtual embodied agents, including voice assistants and virtual agents/humans, have evolved in different domains, including education, entertainment, training, and for interacting with home devices.

Definitions of what such virtual agents are diverse. For example, intelligent virtual agents can be defined as sharing the virtual world with users when providing learning guidance and content (Rickel, 2001). Embodied conversational agents with life-sized and humanlike appearances can partner with a real human in performing a presentation (Trinh, Ring and Bickmore, 2015), or they could be a virtual agent or virtual interviewer, which can implement text-based conversation (Li et al., 2017). The common feature of these definitions of virtual agents is that these agents function in such a way that they can simulate to a certain degree a real human being.

AR agents differ from screen-displayed virtual agents. These VR agents offer immersive interaction also using 3D vision stimuli, sometimes combined with haptic devices, in a fully virtual space. In such interactive environment, users can observe the virtual body via a third person's point of view (can see fully body), or a first-person point of view (cannot see their heads).

However, people have a tendency to compare a current interlocutor with previous social experiences and feelings, which can lead to stereotyping and bias (Bourgeois and Hess, 2008). Neither screen-displayed nor VR agents can perform face-to-face interaction in the real world. Therefore, there exists a certain shortfall in credulity in that virtual agents cannot deliver a plausible illusion of physicality. Although screen-displayed virtual agents are able to recognize the human user's face, execute eye tracking, or perform gesture recognition, they cannot overcome the problem that the user and the agent are not in the same shared physical and virtual space.

On the other hand, virtual agents in AR are endowed with a certain degree of awareness of the real surroundings, based on spatial mapping and tracking of real objects. This produces a sense of co-presence for the users. Most research regarding AR agents has been focused on narrowing the gap between semi-virtual and real perception by reconstructing the real in an interactive space (Park et al., 2020).

Proposed by Holz et al. (2011), the Venn diagram depicted in Figure 1.1 helps narrow down the three research areas that need to come together in the investigation of embodied AR agents for this thesis. The term 'physical agents' refers to robots that exist in the real world, while MR agents, or embodied AR agents, perform at the interface of both physical reality and virtual reality.

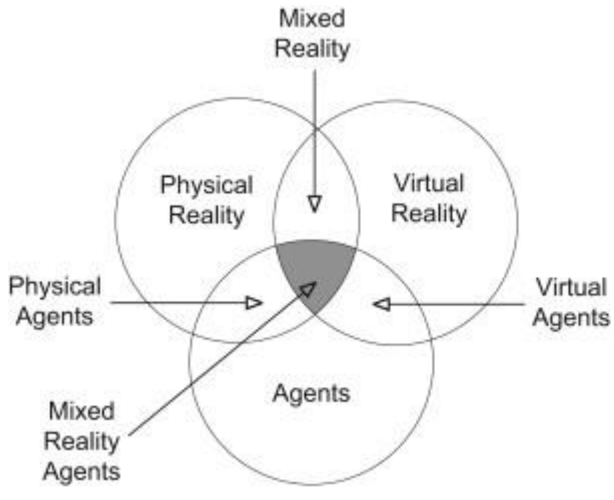


Figure 1.1. A Venn diagram (Holz et al., 2011)

However, due to the fact that screen-display, VR, and AR are different technologies, virtual agents in these environments have certain dissimilarities. Previous definitions of virtual agents, virtual humans, or embodied conversational agents cannot do justice to the AR characteristics. Holz et al. (2011), for example, envisioned "an agent embodied in a mixed reality agent", which illustrates that such agent has a visible body that can be projected into the real-world for interaction. This does, however, not provide a comprehensive and systematic investigation into the key characteristics of such AR agents.

To avoid confusion, this thesis prefers the term ‘holographic artificial intelligence’, first proposed in Huang, Wild and Whitelock (2021). The term holographic AIs equips agents with computer-graphics AR technology to respond to both events in the real and digital realm.

In order to create such agents with AR features, prior studies emphasize the study of how virtual and real-world information can be conveyed in real time, such as verbal/non-verbal interaction, physical-virtual recognition, or animations (Barakonyi and Schmalstieg, 2004; Hartholt et al., 2013; Chetty and White, 2019). There are different requirements for creating AR agents. For example, creation of a particular virtual agents may not require language generation, only verbal input interaction, and thus relies on different modules mixed together (Hartholt et al., 2013). For example, Ali et al. (2019) provided a basic architecture for creating the appearance of an anthropomorphic AR agent, from body design to body animation, and facial animations.

However, current models or frameworks for holographic AI are disparate, and there is a clear need for a cohesive structure to determine the requirements of holographic AI. Therefore, it is necessary to identify which elements of a potential harmonised model could feed into a specific holographic AI, and explore all possible permutations. Such a taxonomy, therefore, should be considered when customising a holographic AI’s features and requirements. A first step towards this goal is a systematic literature review.

Furthermore, it is crucial to distinguish between the definitions of a research model and a research framework. A research model can produce formal and visual representations that elucidate specific areas, enabling the description of influencing factors and their interrelationships (Fettke, 2009; Wand and Weber, 2002; Naffziger, Hornsby and Kuratko, 1994). Conversely, a research framework offers a foundational structure of ideas for exploring a phenomenon, incorporating conceptual and design elements derived from theories, concepts, and methodologies (Lester, 2005; Lithner, 2007). Therefore, this study is dedicated to developing a model for holographic AI—focusing on the components that constitute it—rather than a framework.

Holographic AIs, especially those with humanlike appearance, can produce a certain sense of realism and presence for the user (Reinhardt, Hillen and Wolf, 2020), which in turn facilitates the fidelity of social interaction and influence. Furthermore, static holographic AIs are unable to exhibit behavioural realism, which is particularly important for intelligent tutoring systems. Therefore, a holographic AI should have two basic elements that are similar to those of embodied virtual agents: appearance and animation. Holographic AI appearance can indicate a character’s level of friendliness (Dryer, 1999), which can influence the user’s perception of the holographic AI’s personality (Catrambone, Stasko and Xiao, 2002). Holographic AI behaviour can be defined as its ability of recognising and interacting with virtual and real objects. However, the Uncanny Valley hypothesis predicts that a character with humanlike appearance, but which is lacking certain human traits, can trigger a feeling of eeriness (Mori, MacDorman and Kageki, 1970). When a holographic AI with humanlike appearance and behaviours is projected into the real world, the incongruity between cartoon-like animation and realistic traits produces an unsettling mismatch of appearance and behaviour (McDonnell, Breidt and Bülthoff, 2012). Therefore, a holographic AI's appearance, behaviours and voice should be consistent with its functionality.

In creating a 3D representation of a holographic AI, the traditional approach is to build a basic human shape by sculpting a high-human likeness 3D model with high polygon counts to replicate good facial details, and then convert it to low-polygon counts. The process in question is complex, necessitating not only a thorough understanding of human anatomy and musculature but also sophisticated software capable of creating the avatar's clothing, textures, and 3D mesh arrangements for movement. As such, traditional methods may be unsuitable for designers with limited experience in avatar creation. Achenbach et al. (2017) proposed 3D scanning for generating 3D models. They used 8 cameras to generate 48 images of bodies and faces, and then generated a point cloud of an actor for model and texture reconstruction. The process of reconstruction is based on algorithms designed to align and match a template model, so that features of the actor, such as height and shoulder length, can be transformed to the template to produce the corresponding 3D meshes that are required to generate these types of 3D models.

In order to be effective in a teaching scenario, the behaviour of the holographic AI should be meaningful with pertinent gestures which match the intention of the instruction. It has been documented that virtual humans employing gestures can improve learning outcome (Twyford and Craig, 2013; Mierowsky, Marcus and Ayres, 2020), and it is known that children can track a teacher's gestures in order to discern which specific question or item the teacher is referring to (Wakefield et al., 2018). In the context of holographic AIs, Li et al. (2018) claimed that virtual agents with gestures can produce a higher sense of presence. Instructional gestures can maintain user attention, yet limited research identifies which specific gestural animations should be employed. Simple repetition of gestures, such as greeting, open-armed inviting, and relaxing movements, may be perceived negatively if performed in a mechanical fashion.

In investigating the influence of holographic AI features (understanding of physical objects and surroundings), studies have focused on the users' sense of social presence or co-presence (Li et al., 2018; Kim and Bruder, 2019; Schmidt, Ariza, and Steinicke, 2020). However, this aspect of user experience alone cannot be used to determine whether a holographic AI can cultivate a trusting relationship with users and maintain long-term positive interaction.

In social interaction, trust is fundamental, whether it is interpersonal or virtual. Golemiewski and McConkie (1975) claimed that "There is no single variable which so thoroughly influences interpersonal and group behaviour as does trust." The definition of interpersonal trust has multiple facets: it can be a willingness (Moonrman, Zaltman and Deshpandé, 1992), passion (Thomas, 1750), or expectation (Rempel, Holmes, and Zanna, 1989), "emotional security" (McAllister, 1995) or "reliance" (Rolin, 2021). These scholars also emphasised that by placing his/her trust in someone else, a person is taking a risk (Sheppard and Sherman, 1998; Bhattacharya, Devinney and Pillutla, 1998).

According to Borum (2010), trust involves cognitive perception that is contingent on judgements that inspire one's confidence in the knowledge and capabilities of others. The perception of a person being trustworthy is based on cognition, and such cognitive trust is dictated by the agent's competence and behaviours. On the emotional dimension, affective trust on the part of the person is determined by the concern and care shown by the other (ibid).

Although intelligent assistants endowed with 'human' capabilities can give rise to natural reactions (Krämer, Rosenthal-von der Pütten and Hoffmann, 2015), the definition of trust may be different in the context of human-computer interaction (HCI). Söllner et al. (2012) claimed that trust should be based on system attributes, such as functionality. Therefore, trust can be defined as the user's attitude stemming from the belief that he/she can achieve a goal under an agent's guidance in a vulnerable situation (Lee and See, 2004). Qiu and Benbasat (2009) investigated the sense of trust towards virtual human in shopping system. They found that a virtual human with voice output generally cultivates a stronger sense of trust, and that gestures can also stimulate a trusting perception (Parenti et al., 2022). Therefore, they claimed people prefer to interact with a virtual agent that has salient features of humanness. Holographic AIs should mimic a real person's appearance, behaviour, and cognitive ability a shared interactive space. The important questions being debated by scholars in the field are whether users' trust towards such agents is similar to human-human interpersonal trust, and what factors influence users' sense of trust in holographic AIs.

A child's perception is based on the specifications of the task, and empathy (Van Straten et al., 2018; Kory-Westlund, 2023), while adults' perceived trust depends on their objective feelings in relation to technology, such as technological functions, ability, and integrity (Sebo, Krishnamurthi and Scassellati, 2019). Adults can change their views on whether they consider a technology as a synthetic or real ('alive') product (Sweeney, 2020). By contrast, the cognitive foundation of trust in young children (for example, 3-year-old children, according to a study by Geiskkovitch et al. (2019)) is less stable (Calvo-Barajas and Castellano, 2022). Therefore, trust in the context of children is defined as dynamic process based on expectation and beliefs (Bernath and Feshbach, 1995; Calvo-Barajas and Castellano, 2022).

Prevailing research typically utilizes definitions of embodied virtual agents in VR or on screens, failing to consider the nuances of AR or to address the myriad considerations in creating a purpose-built holographic AI. While Norouzi et al. (2020) surveyed current trends in holographic AIs across different research fields, their review does not articulate the elements of a design model specifically, nor does it illustrate how to implement features like multimodal interaction within the holographic AI.

Additionally, the proposed model has not been empirically validated nor verified through the creation of a holographic AI, which limits the exploration of potential elements beyond this model. Ali et al. (2019) described a traditional approach to create a holographic AI based on their architecture of multimodal interaction, but failed to explicate the development of this structure, resulting in some elements being either redundant or overlooked. For instance, multimodal interaction encompasses object recognition, and virtual characters may incorporate chatbots.

The importance of body language is also often disregarded, a prevalent issue in the literature. Mechanistic and repetitive gestures do not effectively facilitate HCI.

Moreover, much of the existing research focuses on isolated factors affecting user experience rather than a comprehensive understanding of user perception. Studies have shown that physical-object interaction outperforms recognition in terms of co-presence (Schmidt, Ariza and Steinicke, 2020), yet they neglect other crucial aspects such as the holographic AI's usability or trustworthiness.

Thus, while the literature offers various recommendations, it often leaves gaps or presents contradictory information, leaving designers and developers uncertain about how to pragmatically construct holographic AI. Additionally, these findings are not synthesised into a model that would enhance the understanding of holographic AI or succinctly summarise its characteristics.

The objective of this doctoral project is to enhance the understanding of holographic AI. Based on a systematic literature review, this thesis constructs a novel model for developing holographic AI that exhibits humanlike performance, integrating attributes such as appearance, voice, and behaviour. It establishes an overall taxonomy for such agents and identifies their key characteristics. Moreover, the thesis examines the design of holographic AIs with physically plausible features and coherent behaviours to validate the model. Building upon the foundational model, the requirements for educational holographic AIs, including instructional gestures, have been analysed in depth. Experiments regarding the development of instructional gestural animations have been conducted, which have the potential to expand the elements of the model. To determine which factors can influence learners' trust, this study considers how a trusting relationship between children and holographic AI might be fostered through AI features, examining the holistic user experience and trustworthiness of the AI. Consequently, this study proposes a set of design guidelines and recommendations for holographic AI. These guidelines aim to provide designers and system developers with a better comprehension of the overlooked aspects of holographic AI, guiding the selection of appropriate components based on its type and intended functions. The insights from case studies and experiments are detailed, suggesting the design of a pedagogical holographic AI that is capable of fostering a sense of trust in children. Furthermore, the study conducts a comprehensive investigation and enhancement of the holographic AI model through case studies and experimentation, assessing its user experience and feeding back into the model's refinement.

Throughout this thesis, detailed findings from empirical studies will be presented while proposing a series of recommendations for developing human-holographic AI interaction. This chapter also offers an overview of the thesis discussing the motivation for this research its scope and contribution to the field.

1.2 Research Motivation and Objectives

For holographic AIs, anthropomorphism plays an important role in mimicking human performance, from appearance to thinking to a model for decision-making. Such a holographic AI undergoes changes in virtual interactive space, generating visual and communicative plausibility and immersion. Therefore, it can provide a series of services in various domains. However, there does not yet exist a harmonised model of holographic AIs which is grounded in an anthropomorphic taxonomy, and there are certain features of holographic AIs which differ from those of VR and screen-displayed agents. VR agents only recognize virtual objects in the virtual surrounding, screen-displayed agents can barely interact with users. Holographic AIs employ multiple interaction dimensions to deal with both virtual and physical contexts, and the users in real-time.

In order to arrive at a feasible proposed model for holographic AI design, the processes of creating a holographic AI be examined. It should be noted that holographic AIs are computer-generated 3D dynamic graphics, consisting of representation (appearance) and behaviour (animations), it is necessary to find a semi-automatic means of generating user's own 3D avatars that can be employed in specific AR domains.

Holographic AI interactivity should in its behaviour be able to offer an illusion of physicality. This is especially critical for intelligent tutor systems, where, for example, instructional gestures can be utilised to express complicated and abstract definitions to help students understand these types of concepts. However, not all gestures are beneficial to learning. If a holographic AI can only perform basic interactive gestures, such as greeting, opening arms, and closing, its behavioural animation lacks affordance in that the holographic AI does not react seamlessly to physical contexts by its subsequent behaviours. Further, Hostetter (2011) found that too much overlapping information synchronously expressed by co-speech gestures reduce learning gains. Too much gestural expression accompanying utterances in a learning space may also produce a negative influence. For that reason, this thesis focuses on key instructional gestures which should be employed by an educational holographic AI.

In terms of interaction, trust can be a crucial factor in deciding whether a holographic AI induces a positive user experience. The concept of trust introduced in this thesis includes cognitive and affective aspects of HCI (Sousa, Lamas and Dias, 2014). Cognitive-based trust relates to the capacity of a holographic AI to direct users in the completion of tasks, and therefore is a rational perspective， and affective-based trust concerns users' emotional reactions. By measuring trust, the researcher essentially measures the number of degrees of willingness of a user to interact with a holographic AI, as well as propensity of engagement. Further, children's sense of trust can reflect children's attitude towards this recent technology, and the technology's acceptability.

Cummings et al. (2007) proposed the FINER criteria for research questions, which encompass feasibility, interest, novelty, ethics, and relevance. This suggests that the current study is feasible, engaging, innovative, safe, and has the potential to contribute meaningfully to the scientific community. Moreover, effective research questions should be articulated with clear motivations and defined research objectives (Thuan et al., 2019), enabling these questions to be addressed and operationalized through appropriate methodologies and resources, thereby fostering the advancement of pertinent research areas. Consequently, the goal of this study is to develop holographic AIs from their initial conception to their subsequent evaluation via case studies and empirical research. Specifically, it investigates the nature of holographic AI, its creation process, and the factors affecting user perception. Using the model derived from this study, holographic AIs are constructed within this study, which serves to evaluate the factors influencing user trust. Additionally, this case study validates the model and refines it, elaborating on its components. Therefore, research questions (RQs) guiding this project are presented below:

- RQ1: What elements and design dimensions constitute the holographic AI?
- RQ2: How to create an anthropomorphic holographic AI in practice, following this model?
- RQ3: What key instructional gestures should be used by an educational holographic AI?
- RQ4: What factors affect the user's sense of trust towards an educational holographic AI?

To achieve these aims, diverse types of holographic AIs in previous studies are analysed in order to arrive at a formal definition and a model for the construction holographic AIs, then building a novel taxonomy with explicit characteristics. Consequently, Chapter 2 delineates an extended avenue of inquiry by incorporating a subsidiary research question, denoted as RQ1-1. This extension is necessitated by the recognition that, within the domain of holographic AI systems, it is advantageous to discern the distinctions in their application-specific implementations. This inquiry contrasts with the identification of shared components that are ubiquitous across the spectrum of holographic AI systems.

Based on this model, the overall stages of holographic AI creation can be illustrated, including 3D computer-generated graphics, animations, and natural language processing, which enables to address RQ2 as well as provide a series of recommendations.

To advance an area of the model, i.e. behaviour, an experiment is designed to identify key instructional gestures that can be used in holographic AIs. Results from the experiment conducted are reported, using AR glasses and motion capture. Chapter 4 expands upon RQ3 by formulating additional, more precise hypotheses concerning gestural generation, aiming to examine how these non-verbal cues can convey the holographic AI's intent.

To investigate the validity of the overall approach and the holographic AIs created as part of the investigations, a second experiment is conducted, pulling out a key aspect, i.e. trustworthiness, of those new types of AIs. The experiment focuses on learners' sense of trust, results from the empirical study are reported in Chapter 5 to answer RQ4. For this, a new metric scale for measuring trust was developed, which can investigate the influencing factors that establish trust. Children's perceptions regarding trust were investigated, leading, finally, to the discussion of a series of recommendations for enhancing trust, as well as integrating findings into the initial model.

The structure of the thesis is as follows: The initial model, based on a literature review, is detailed in Chapter 2, addressing RQ1. Subsequent validation through case studies and experiments is conducted in Chapters 3, 4, and 5, tackling RQ2, RQ3, and RQ4 respectively. Each chapter refines the model in light of the empirical findings. Finally, the mature model is presented in Chapter 6, revisiting RQ1.

1.3 Scope of the Thesis Research

The research on holographic AI presented in this study is inherently interdisciplinary, rooted primarily in digital media and computer science. Moreover, it intersects with game design, MR, AR, and HCI. The holographic AI is designed to communicate through a variety of modalities, including visual appearance, body and facial animations, interactions both virtual and physical, as well as natural language processing.

This thesis therefore presents an overall approach for the development of holographic AIs in the field of digital media or digital media production, ranging from 3D modelling, animation, dialogue management for natural language processing, physical recognition to user experiment measurement. Although the development of holographic AI could be broadened to include a remote and synchronous collaboration with other users, this research only focuses on one-to-one interaction with the holographic AI.

The input channel of a holographic AI built in this thesis includes user's input, i.e. speech to text service, and output channels consists of vision (i.e. appearance, body movements, gestural animations, emotions, lip-sync animations), user recognition, and utterance (i.e. dialogue management and text-to-speech).

The aims of this thesis are to investigate a holographic AI in an intelligent tutoring system, and the effects of trust on learners' perception to inform the development of holographic AIs.

The limitation of the scope of this investigation is that the holographic AI cannot yet directly recognise physical objects and manipulate physical objects flexibly and openly, though a model for that is included in the thesis, but it can interact with users and virtual objects. It will be shown that the holographic AI can exhibit plausible behaviour to react to both virtual and physical surroundings based on a predefined scenario.

The research with respect to prototyping an executable holographic AI is based on a novel model. Factors influencing trustworthiness of holographic AIs is also investigated. In order to ensure the application's validity, i.e. showing that the holographic AI can be used in life-like settings, the thesis does not use controlled experiments such as 'Wizard-of-Oz' (Woz) testing, but conducts empirical experiments. Other than popular Woz studies, therefore, this thesis can observe current shortcomings of holographic AI technology much more precisely, outlining a path for future development and propose improvement aspects and identifying development trends.

1.4 Contribution

Prior research has not provided an exhaustive analysis of the development and deployment of holographic AI systems. This study initiates a thorough examination of the definitions and attributes of holographic AIs, aiming to facilitate a detailed investigation into their component parts. For instance, conventional VR and screen-based agents fail to meet the unique criteria and functionalities inherent to AR technologies such as spatial mapping. Moreover, the characteristics of preceding virtual agents are clearly delineated from AR agents. This distinction necessitates a deeper understanding from users and designers regarding the rationale for the adoption of holographic AIs. For example, the work of Norouzi et al. (2020) and Ali et al. (2019) falls short in fully delineating the classifications and unique attributes of holographic AIs, as well as in thoroughly examining their shared or distinct functions and features. Therefore, a model and taxonomy of holographic AIs are proposed, also to satisfy the requirement of specific application domains, so that different dimensions and elements can be selected based on the holographic AI features and a typology can be mapped out for specific feature combinations and settings required for particular applications.

This study scrutinizes case studies that delve into the design, creation, reconstruction, and intelligence enhancement of holographic AIs. A 3D scanning technique is utilized to generate anthropomorphic holographic AIs, which are then assessed in comparison to conventional creation methods. Additionally, this study presents an analysis of the processes involved in producing facial and lip-sync animations. The amalgamation of animation techniques, natural language processing, user movement tracking, and spatial awareness is leveraged in the development of interactive modalities, thereby enabling a comprehensive set of recommendations for future holographic AI advancements. The goal is to equip developers and designers with the knowledge required to conceptualize and implement holographic AIs within an MR environment. Furthermore, the proposed model undergoes refinement and extension within this study, with a focus on removing ambiguous and superfluous elements.

In order to develop holographic AI animations to enhance behavioural realism, it then provides results from an extensive investigation on what key instructional gestures that can be used in holographic AIs, especially for intelligent tutoring systems. The holographic AI instructional hand movements can enrich the variety of representational gestures, which can match the holographic AI's utterances and intentions. This study discovers novel categories of gestures in instruction, which is beneficial for animation designers to select appropriate gestures for an educational holographic AI or training system, aiming at promoting its motivation and emotional conversation, rather than repeatedly performing limited behaviours.

Trust is critical in forming both transient and enduring relationships. Nevertheless, previous studies have not concentrated on a trust questionnaire specifically devised for holographic AI systems. For example, Kim et al. (2018) concentrated on trust in technology to evaluate the safety of holographic AI systems. While holographic AIs are a part of technological progress, their unique interactive methods, environmental integration, and visual presentations demand a specialised trust metric. Therefore, this study appraises user experiences specifically within the holographic AI context.

To investigate the validity of holographic AIs, the key element of trust in holographic AIs is investigated, for which a new questionnaire is developed. This (re-)defines what trust means and how it can be measured, and it is applied directly to a holographic AI culminating from a series of prototypes, assessing the found sense of trust of learners towards this holographic AI. Each dimension of the model enables analysis of what contributing factors can influence user perception. While, children's trust perception is different to adults, this thesis makes six recommendations based on the collected feedback from young learners.

Overall, the aim of this study is to construct a comprehensive model with the intent of offering design recommendations that can bolster the development of holographic AIs for developers, designers, and users alike. This study presents a set of guidelines derived from case studies and empirical research, intended to reveal previously unconsidered factors. For instance, the use of representational gestures is proposed to enrich the diversity of gestural interaction, thereby facilitating the effective conveyance of the holographic AI's intents and motivations. The trust scale is developed to examine factors influencing user experiences, with an emphasis on discerning the particular needs of children to amplify their trust in holographic AIs.

The principal contribution of this study lies in the establishment of a theoretical framework for interactions between human users and holographic AI systems, as well as in the development of a trust measurement model within MR environments. Additionally, this study advances the creation of educational holographic AIs, fostering a symbiotic relationship between such agents and the learning experience.

The findings and conclusions further refine and expand upon the model for holographic AI and offer recommendations to surmount the limitations in exploiting body language performance and sustaining user engagement. While this study primarily focuses on the design and study of educational holographic AIs, the results have broader implications for physical robots and various MR/AR/VR agents. Instructional gestures, for example, can be adapted for use in both educational and robotic contexts. The development of holographic AIs is an interdisciplinary endeavour, intersecting HCI, artistic design, and user experience. This study synthesizes these diverse elements into a succinct set of design guidelines for interactions between humans and holographic AI, which will significantly contribute to future research aimed at enhancing interactions with screen-based, AR/MR agents, and fostering enduring interactive relationships.

1.5 Organization of the Thesis

This thesis explains a holographic AI’s model, and holographic AI creation based on this structure, and what key instructional gestures can be used by holographic AIs. Moreover, it investigates trustworthiness of the holographic AIs and reports findings.

Chapter 1 provides a brief introduction to the impetus for developing holographic AIs. An overview of the thesis was provided with a conceptual structure, which illustrates the rationale for investigating holographic AIs.

Chapter 2 proposes a model for holographic AIs on the basis of a systematic literature review. It also presents a taxonomy and characteristics of holographic AIs based on different application domains and user perceptions.

Chapter 3 describes the processes of a holographic AI creation, based on the model that was proposed in Chapter 2. It also compares two ways of creating holographic AIs. In addition, it details the generation of body and facial animations using motion capture. This chapter discusses multiple interaction approaches using natural language processing and user tracking techniques. Based on this experience, this chapter concludes with a series of recommendations, and the initial model is refined.

Chapter 4 reports on an experiment, the aim of which is to identify key instructional gestures that can be utilised by holographic AIs. It proposes a methodology for specifically recording and capturing instructional movements, and re-categorises representational gestures (i.e. iconic, deictic, emblematic, metaphoric, transformational, and mimicking gestures). The chapter concludes with a proposal for a holographic AI with instructional gestures and natural language processing, which allows to extend the model regarding performance of body language.

Chapter 5 proposes a new metric scale for measuring trust, targeted at holographic AIs. It presents 5 dimensions (competence, integrity, benevolence, compassion, and relationship) and 11 factors of trust. Based on the scale, this chapter reports on an experiment designed to measure children's sense of trust towards the holographic AI, the design of which is detailed in previous chapters. The chapter concludes with six recommendations for developing and improving trust in the relationship between children and holographic AIs. It further enhances the model with respect to user experience.

Chapter 6 draws conclusions on the findings of previous chapters and the experiments, and considers future work stemming from the research findings presented in the main chapters. Unique contributions to the field, as well as limitations of, this doctoral research are discussed, and recommendations for further research are prescribed.

Chapter 2 Systematic Literature Review and Model

The purpose of this chapter is to review the current state of the art in Augmented and MR agents. In the first section, a definition of virtual agents and Augmented and MR agents followed by a comparison of traditional virtual agents for VR and screen-display will be provided. This chapter will also illustrate how artificial in AI can be utilised for building intelligent agents in AR.

2.1 Introduction

Figure 2.1. An example of holographic AI (own graphic)

Virtual humans mimic both the appearance and behaviour of real humans. They are often defined as ‘autonomous agents’ (Rizzo et al., 2016) or ‘artificial agents’ (Traum, 2009) in that they possess an anthropomorphic body, express a human-like range of emotions, and engage in natural communication (both verbally and non-verbally).

For example, NEON has developed realistic AI assistants that appear on a screen (NEON, 2023), and the KRAFTON company has created a virtual female with a hyper-realistic appearance (KRAFTON, 2022), the latter which has become a digital celebrity. The capabilities of traditional virtual humans, however, often are inhibited by the restrictions inherent in the delivery devices currently available, even though screen displays can recognize the user’s behaviours by sensors. While, VR cannot capture information concerning real-world surroundings.

By contrast, AR can overcome this limitation by sharing both the real and digital spaces with the user directly. Guided by spatial mapping, virtual characters can obtain real-time information about the physical surroundings while generating both corresponding virtual representations of world and objects. AR endows virtual agents with a certain degree of ‘space intelligence’, i.e. physical awareness, enabling them to respond and react to external, real-world events.

AR agents’ sense and project into the physical world to process and display information in real-time (see Figure 2.1). In this context, the term ‘holographic’ is often used substituted for AR. This term should not be confused with holograms, which are different from holographic displays and user interfaces. Holograms are the resulting 3D images produced by light refraction that reconstitute the reflection as if it came from real objects. Holographic agents require a specific delivery system, typically a waveguide display, to realize their 3D embodiment and to extend the users’ real-time experience of their physical surroundings (K. J. Kim et al., 2020; Carrozzi et al., 2019). Such devices include Microsoft’s HoloLens, Lenovo’s Think Reality glasses, Magic Leap, or others of a growing class of AR glasses. In order to distinguish from VR agents, it is preferable to use the term” holographic AI” over “virtual agent” or “virtual human” (Huang, Wild and Whitelock, 2021). Essentially, a holographic AI provides a humanoid interface through which the user interacts with physical and virtual surroundings in the mixed world.

Although the number of studies regarding AR assistants has grown substantially over the last decade, academia is divided over the right terminology, and the differences between VR humans and AR agents are often blurred, with these terms referring to the same foundational definitions of ‘embodied agent’ (Techasarntikul et al., 2020), ‘intelligent virtual embodied agent’ (Iqbal, Mangina and Campbell, 2019), or ‘virtual human’ (Lampen, Lehwald and Pfeiffer, 2020).

For example, in researching traditional virtual agents, Knote et al. (2019) conducted a systematic literature review and a cluster analysis, from which they identified five categories of intelligent personal assistants (i.e. adaptive voice (vision) assistants, chatbot assistants, embodied virtual assistants, passive pervasive assistants, and natural conversational assistants). Their cluster analysis, however, does not clearly single out characteristics. Given that both natural conversation assistants and adaptive voice (vision) assistants use natural language processing, Knote et al. (2019) merely postulate that adaptive voice assistants combine it with visual interaction, not acknowledging the complexity of natural conversation assistance.

Huang, Wild and Whitelock (2021) have proposed an additional early taxonomy for AR agents, and have conducted a comparative analysis exploring nine AR anthropomorphic agents, including game characters, simulation agents, chatbots, and intelligent tutors. However, while the study is informative, it samples only nine commercial products, the findings of which cannot robustly support the generalization of the extracted features of holographic AIs.

Norouzi et al. (2020) have presented a systematic review of work covering the holographic AIs’ appearances, behaviours, spatial interactivity, displays, and social distance (50 papers in total). Their study also considers future directions in the development of holographic AIs and head-mounted displays. However, although they consider the holographic AI as a companion which can collaborate with users and possess personality, their study focuses more on head-mounted displays, rather than the holographic AI itself and the advancement of its abilities.

Physical recognition is a critical characteristic, as mentioned in two studies regarding holographic AIs (Huang, Wild and Whitelock, 2021; Norouzi et al., 2020). However, scholars have yet to fully investigate and categories different situations in terms of the level of physical interaction or spatial recognition needed to avoid improper holographic AI performance. Another critical research avenue is the development of, and the reliance of holographic AIs on, the ancillary AI tool of physical-object management for tracking or influencing objects' states, and realising mutual conversion of dynamic and static states.

Additionally, a holographic AI can work only for one application, which means it cannot fully transit multiple scenarios. If the holographic AI cannot express a comprehensive set of human emotions, or automatically produce its own storylines and experience, it will fail to attract and maintain audiences' interests, and be relegated as a short-lived virtual influencer or product.

Therefore, there is a need for an updated and comprehensive review which provides an overview of the current limitations of holographic AIs, and novel development tendencies in the field. In order to address the knowledge gap and answer the first research question, this chapter applies a systematic review of specific characteristics and corresponding implementations of holographic AIs.

The purpose of a systematic literature review is to identify, select, evaluate, interpret, and analyse sources of evidence (Moher et al., 2015). It classifies the background, conditions, and features of the phenomenon in question, and provides transparency and comprehensiveness. To avoid an ambiguous characterization of holographic AIs and their application scope, this chapter builds a taxonomy by exploring representative features by the way of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).

This investigation reviews published articles regarding AR embodied agents during the years 2009-2022. The critical objectives of this study are to consider explicitly what AR agents are, explore their integration in high-tech applications, and then summarize the current status of holographic AI research and development.

This study is dedicated to constructing a model that captures the distinguishing characteristics of holographic AI systems. It aims to explore the capabilities of these agents, including their visual representations, interaction modalities, and the overall user experience. In this context, it is deemed beneficial to devise a taxonomy of holographic AI to help further understand its distinct primary components. Within the range of holographic AIs, it is beneficial to understand how holographic AIs differ in their application, as opposed to which components are common to all holographic AIs (RQ1). For this, RQ1-1 (see below) has been added.

This desktop-based research study seeks to evaluate the full range of holographic AI capabilities via a systematic approach, by identifying their working principles, exploiting their future development tendencies, and then presenting a set of recommendations.

The objectives of this systematic review are as follows:

- What elements and design dimensions constitute the holographic AI? (RQ1)
- What are categories of the holographic AI based on this model for creation? (RQ1-1)

It should be clearly mentioned that there are other fields of study intersecting of AI with AR, which lie outside the scope of this chapter. Most notably, probably, one should mention Deep Fakes and AR filters in this context, where AI technology is used to augment the appearance and voice of a physical person to make them appear to be someone else (much like AR filters). In this case, the difference to what is studied here is that AI is understood in its narrower definition of solving specific problems with machine learning, while here it aims more at the production of general intelligence.

The rest of this chapter is structured as follows. In Section 2.2, the methodology of the review is described, and in Section 2.3, a summary of the results is presented, mainly highlighting the holographic AI features. Section 2.4 summarizes how studies investigate user experience. The findings are discussed in Section 2.5, and Section 2.6 summarizes holographic AI's development trend at the end of this chapter.

2.2 Methodology of the Review

A comprehensive evaluation of the concepts, features, and categories of a phenomenon by means of a systematic review can provide clear insights into available data, and also serves to identify existing research gaps and future development possibilities. The quality of a systematic reviews is dictated by the coverage and quality of the primary studies considered: the more fragmented evidence, the harder it is to extract and support generalizable results. There are three pillars of a systematic review: keywords, databases, and screening criteria. To avoid information retrieval and relevance issues, advanced literature search engines including Harzing's Publish or Perish tool were employed to collect all relevant studies, and the PRISMA methodology was applied (Page et al., 2021).

2.2.1 Identification

The goal of this review was to select studies that either explicitly investigate visual intelligent agents embedded into AR, or which describe general features of holographic AIs. The following keywords of article titles were considered: ("augmented reality" OR "mixed reality" OR "extended reality") AND ("agent" OR "intelligent agent" OR "assistant*" OR "embodiment" OR "virtual human" OR "virtual agent" OR "embodied").

In an effort to identify more relevant papers, an enlarged keyword search employing the aforementioned keywords, plus "virtual human", "augmented reality", "conversational agent", "mixed reality", and "hologram", was conducted.

The electronic databases utilised in this review included ACM DL, CEUR-WS, Frontiers, Google Scholar, IEEE Explore, IMD Lab, ISCA, MDPI, ScienceDirect, Springer Link, and SciTePress. The review covers studies publications from 2009 to 2022, including earlier reviews (Cipresso et al., 2018). In total, 321 papers were retrieved, and 5 papers are not free to read were discarded.

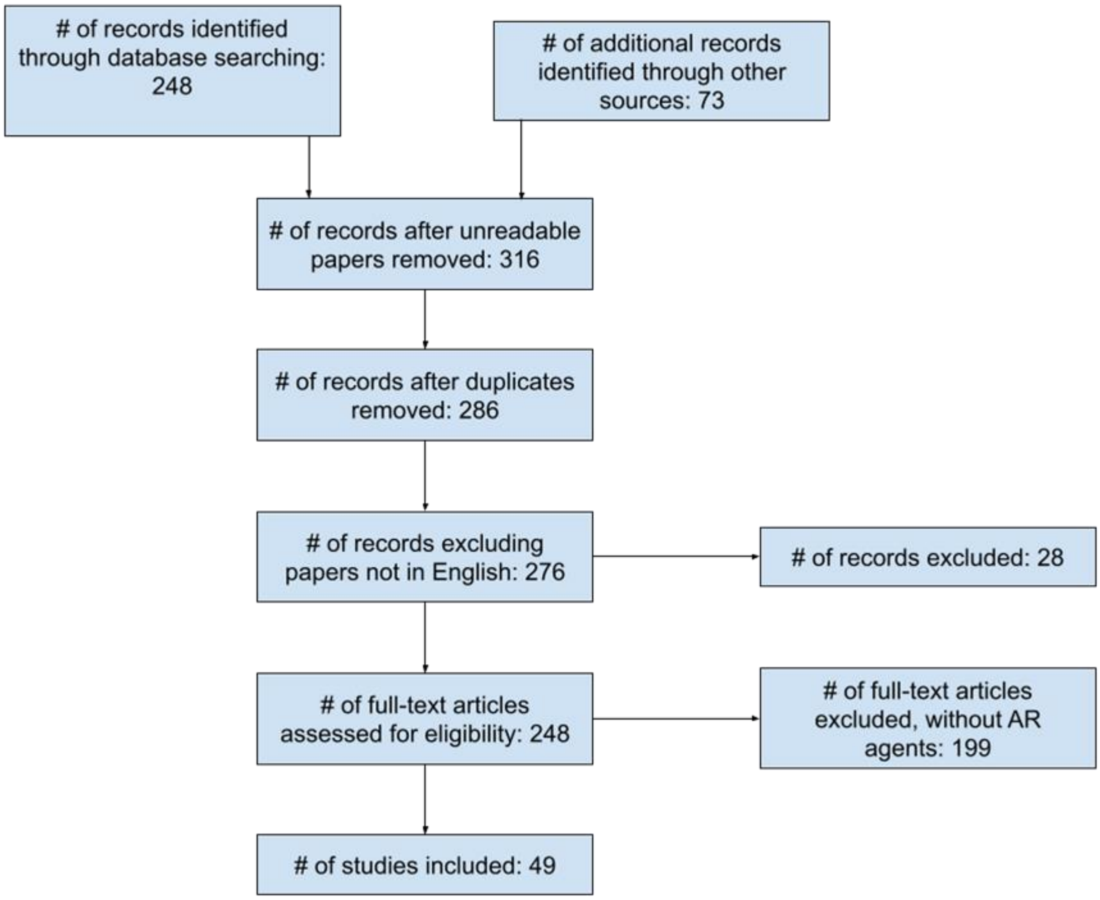


Figure 2.2. PRISMA flow

2.2.2 Study selection

On the screening stage, 30 duplicates, as well as 10 articles that were not published in English, were eliminated, leaving 276 studies. The eligibility criteria are structed upon the definition of holographic AIs provided by Huang, Wild and Whitelock 2021). Abstracts satisfying the following criteria were selected for review: (1) studies clearly revealing visual agents in an AR environment; (2) studies concerning animated holographic AIs; (3) abstracts for journal articles, conference papers, and doctoral dissertations.

In total, 276 abstracts were considered eligibility in accordance with the above criteria. Discarded studies were re-evaluated, and reasons for their exclusion were discussed below, ensuring no essential literature was missed.

Based on the preferred reporting items for systematic reviews and meta-analyses 2015 statement, Figure 2.2 shows the PRISMA flow diagram.

Twenty-eight studies that did not specifically concern AR technology were excluded. A further 199 studies were excluded as they did not use or investigate visible 3D agents in AR (n=79), or they concerned voice assistants (n=44), AR animation (n=25), AR shopping systems (n=26), or AR learning systems (n=18), or they were Bachelor and Master degree level-based dissertations (n=7).

This left 49 eligible full-text papers, including one PhD dissertation (Kim, 2018a) that consists of four studies (Kim et al., 2019, 2016, 2018; Kim, Bruder and Welch, 2017) with two additional articles published (Schmidt, Nunez and Steinicke 2019; Schmidt, Ariza and Steinicke, 2020).

2.3 Result

As mentioned, all studies in line with the selection criteria were published between 2009 and 2022 (see Figure 2.3), and concern virtual characters in AR. It is indicated in Figure 2.2 that the 2010s witnessed a general increase in academic studies on holographic AIs, a trend which appears to have been arrested during the Covid-19 pandemic: 77.6% of the studies were published in the last four years (from 2022 to 2018), including 7 studies in 2018, and nearly double that in 2019. Additionally, 49 studies concern the development of wearable smart glasses for observing holographic AIs, and 10 studies concern the application of mobile AR in screen display.

The selected studies were subsequently analysed in depth in terms of the following categories: definitions of holographic AIs, features of holographic AIs, interaction methods, areas of interest, and user experience in terms of interaction with holographic AIs.

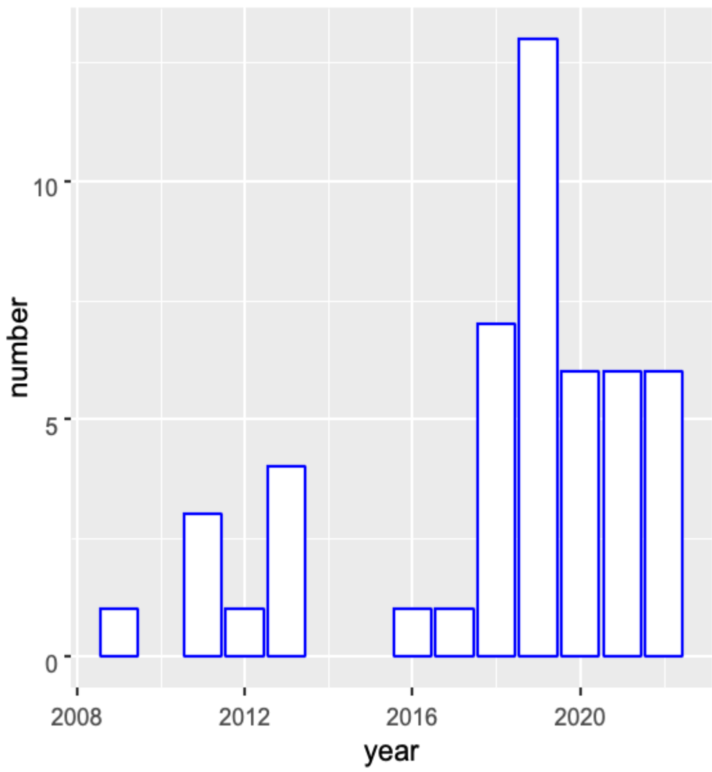


Figure 2.3. Publications retained by year

2.3.1 Definition, model, and taxonomies of holographic AIs

Several of the papers use the following words “virtual human” or “virtual agents/assistants/animals” to indicate the agent having a visible body in AR, and frequently employ the words “embodied (conversational) agents” in their concepts and titles to replace “virtual agents/ humans”. Some of them suggest that the holographic AI’s visual representation is based on (3D) computer-generated graphics sharing the same space with the user. Some definitions reflect specific interactive abilities, such as environmental understanding, real object recognition, translation, verbal communication, remote collaboration, translation, and educational training (n=31, see Table 2.1). Importantly, Table 2.1 shows that some studies explain an interactive environment by using the phrases “in the physical or real space”, “interact with the real world”, “share the physical space”, “integrate into a virtual or physical space”, “project into the physical world”, and “in the MR environment”. Only three studies propose that the holographic AI delivers plausible interactions by imparting on the human user a subjective illusion of physicality in the real world and the feeling of being together. The holographic AI overlays on the physical environment via wearable headsets or smartphones, and it is neither physically present in the environment nor sharing the same (real) space as the user; rather, it shares the same MR space.

The concept of holographic AIs is often loosely defined. These studies only describe one or two features of holographic AIs, such as appearance, abilities, or functions, but do not comprehensively define what holographic AIs are. For example, Lee et al. (2021) defined the virtual human in AR as a humanoid computer graphical representation that can share the real environment with users. This definition reflects two elements: visible humanlike appearance, and interactive space. Although Lee et al. (2021) provided more details later regarding the holographic AI’s abilities, their general definition does not highlight its key features. Among the papers, the majority of definitions either suggest that the holographic AI has embodiment in the real world, or only focus on specific interactive capabilities. For instance, the holographic AI is capable of verbal and non-verbal communication (Wang, Smith and Ruiz, 2019), understanding the user’s context (K. J. Kim et al., 2020), or performing real-time interaction (Verma et al., 2020). A definition focusing on a single functionality or characteristic makes it rather difficult for researchers to explain why users need these AIs with AR support. Such a definition fails to explain the particularity of an AR interactive environment, nor can it distinguish the AIs from virtual agents in VR or screen displays. For example, Nasution et al. (2020) created a translation system for a holographic AI that can teach English; however, with this arrangement the holographic AI could be replaced by text-input. Besides, the term ’embodied agents’ or ’embodiment’ consider coexistence with the user in a place, but cannot accurately indicate the spatial levels of interaction, such as virtual interaction, mixed surrounding interaction, or physical awareness and interaction.

The holographic AI and virtual human/agent have common features in appearances, verbal/non-verbal abilities, or application domains. However, holographic AIs can generate the illusion of physical presence in the real world because of their visualization and performance in multiple interactive spatial dimensions. Physical and virtual interactive spaces are not in the same horizon, thus the holographic AI or virtual object cannot occupy the real environment since it lacks physical instantiation, and also because it does not completely rely on the virtual social context. As mentioned above, these definitions cannot convey more information in terms of certain aspects of difference between different holographic AIs, even though the common feature among holographic AI is a certain degree of interactive ability in the physical surroundings. Therefore, a comprehensive definition should clearly present what the holographic AI looks like, the holographic AI’s abilities, where it displays, and its aims. Table 2.1 highlights the unclear and contradictive definitions (Chu et al., 2012; Kim 2018b; Park and Jeong 2019; Lee et al., 2019; Pimentel and Vinkers 2021; Norouzi et al., 2022; Huang et al., 2022).

Description of virtual agents in AR	**Definitions**	**Studies**
Virtual human	is a humanlike computer graphics manifestations that gives the illusion of physical presence.	Kim et al. (2021b)
Virtual human	is a 3D computer-generated embodied agent in a physical environment.	Pimentel and Vinkers (2021)
Virtual human	is a human like computer graphics representation that can display in a virtual surrounding or share a real interactive world with the user.	Lee et al. (2021)
Virtual human	is an intelligent virtual agent that is able to produce social influence by its presence in the shared interactive surrounding with the users".	Kim (2018b)
Virtual human	is a computer graphics with animations that can generate an illusion of physicality.	Kim (2018a)
Virtual human	can produce an illusion of physicality in the physical surrounding world by its appearance and their behaviour.	Kim, Bruder, and Welch (2017)
Virtual human	has humanlike representation that is able to perceive and interact with users verbally and nonverbally.	Hartholt et al. (2019)
Emotive virtual human	is a virtual patient that can help medical students practice interviewing.	Zielke et al. (2018)
Virtual human agent	creates a social context in which users can interact or collaborate with it in the same environment.	Huang et al. (2022)
Intelligent virtual agent	has appearance and animations that is able to communicate with users in a natural way by recognizing interactive environment and influencing physical objects.	Kim et al. (2018)
Intelligent virtual assistant	is an intelligent collaborative entity that can verbally command and identify user's context to be	K. Kim et al. (2020)

Virtual agent	can improve the capability of a user to interact with the real surrounding by MR devices.	Lang, Liang and Yu (2019)
Digital personal assistant	is a speech input and output interface that is able to interact in real-time, identify and finish user commands.	Verma et al. (2021)
Indoor dialog agent	can interact with nearby users in a real-world space.	Park and Jeong (2019)
Virtual conversational agent	shares the same space with users and can control physical objects at the very position.	Miyake and Ito (2012)
Desktop assistant	has a visible body allows it to share the same physical environment with users, allowing it to move freely in all directions.	Chu et al. (2012)
AR machine translation agent	is able to translate a text from Indonesian to English by voice.	Nasution et al. (2020)
Interactive agent	allows to experience flower gardening over a real book.	Oh and Byun (2012)
Blended agent	is able to control a real object associated with its location and surface material.	Schmidt, Nunez and Steinicke (2019)
Intelligent blended agent	can react to both real and virtual objects.	Schmidt, Ariza and Steinicke (2020)
Virtual animal	is a computer-generated graphics that can be projected into a real or digital world.	Norouzi et al. (2019)
Embodied agent	is computer animated characters that can be applied in various applications across a variety of display modes and setups.	Norouzi et al. (2020)
Embodied conversational agent	is a natural speech interface with humanlike appearance.	Wang, Smith and Ruiz (2019)
Embodied conversational agent	utilizes 3D VR and AR technologies to provide tangible benefits for human-to-human communication and interpersonal skills development.	Chetty and White (2019)

Embodied conversational assistant	mimics real human voice, appearance, and behaviour.	Reinhardt, Hillen and Wolf (2020)
3D embodied agent	is able to identify user' requirement and perform interaction in the real world.	Norouzi et al. (2022)
3D virtual walking partner	has humanoid appearance to provide exercise services.	Yoo and Tanaka (2022)
Mini-Me	is a novel adaptive avatar for remotely collaborating on MR with redirected gaze and gestures.	Piumsomboon et al. (2018)
MiRa	has embodiment in the MR surrounding.	Holz et al. (2011)
AuARs	has embodiment in AR environment.	Campbell et al. (2014)
Holographic AI	is a humanoid user interface that is able to identify and interact with real and virtual surroundings.	Huang, Wild and Whitelock (2021)

Table 2.1. Definitions

In Table 2.1, some of the studies use virtual humans/agents in VR and screen displays to describe embodied agents in AR, and their definitions also apply to the traditional concepts applied to virtual humans/agents. Even though virtual humans and holographic AIs have similarities, their interactive spaces and extended capabilities in terms of physical and virtual recognition are different. By explicitly defining holographic AIs, users can understand better whether a holographic AI can react to the physical context.

In terms of taxonomical analyses, Holz et al. (2011) introduced agency, a sense of co-presence, and interactive ability, which was then subdivided into weak and strong levels. They cite over 25 projects in their exploration of four types of agents based on the relationship between the physical and the virtual environment. Stronger virtual embodiment focuses on the virtual domain. Virtual agents with spatial awareness can deliver contextual information about the real world, but cannot influence their surroundings. Robot agents with stronger corporeal presence integrate into virtual environments for navigation and positioning. MR agents have both strong real and virtual co-presence, and they have virtual representations showing on screen-displays on real robots (ibid).

This taxonomy, however, includes diverse agents that are not fully virtual, which lie outside the scope of the holographic AI. As mentioned above, Huang, Wild, and Whitlock (2021) presented a comparative survey of how holographic AIs are developed. By inputting keywords such as “HoloLens virtual human”, “holohuman”, and “holographic system”, the authors identified relevant products on Google and YouTube. The nine agents they identified can be categorized into four usage types: game characters, simulation AIs, chatbots, and intelligent tutors. These four types of holographic AIs differ greatly in terms of appearance, behaviour, intelligence, and responsiveness. For example, the AR game character uses storylines and performs various animations.

Norouzi et al. (2020) have classified 50 papers concerning holographic AIs using head-mounted displays based on research categories, application areas, and appearances. Their taxonomy of holographic AIs includes human, animal, robot, and other kinds. Other types of the holographic AI (fourth in the list) include cartoons, monsters, and humanoid objects, and this vague category ensures that these AR virtual objects have interactivity, or are merely decorations in the AR environment, such as anthropomorphic sun. The holographic AI's appearance is only one branch of the classification, which cannot differentiate application domains or affordances. The authors also classify holographic AIs into application areas, including assistive/collaborative, entrainment and interactive media, healthcare, and training. Healthcare-related holographic AIs are design to improve the health of specific users, or serve as a teaching tool. However, the study does not specify the differences between the health and training groups, and both types of holographic AI can also foster a collaborative relationship.

In terms of technical architecture, Chetty and White (2019) have proposed a multilevel architecture of a sense-think-act cycle, including a cognitive level, sensor-fusion level, and an environment simulator level. The cognitive level applies decision-making in accordance with user goals. The sensor fusion in the second level integrates inputs and outputs from a variety of sensors. The environment simulator level ensures the holographic AI is able to interact with different contexts, environments, and other agents. The model relies on a reasoning model of belief-desire-intention (Wadsley and Ryan, 2021), but the authors do not explain how this model contributes to the cycle they propose, or how the sensor fusion level performs a transformation from input to output.

Similarly, a process of creating holographic AIs with full-sized cartoons and animations is presented by Ali et al. (2019). Interaction ensures that the holographic AI can identify objects the user gazes at, as well as express body language and emotions. Scalability incorporates anchor maps based on image recognition and physical tracking to build an interactive area for the tracking of real objects. Empowered with these qualities, the holographic AI can recognize real flowers and perform face-to-face communication. This architecture fits the model proposed by Huang, Wild and Whitelock (2021).

Chu et al. (2012) postulated that the holographic AI has four elements: static and dynamic appearances, physical understanding, and user behaviour recognition. The static visual appearance emphasizes naturalism/stylization and the size of agents. Animation in dynamic, while visual appearance should be consistent with static appearance. Physical awareness follows physical norms and avoids conflicts with real objects. User action recognition conducts gesture interaction and deictics expressions. However, their study does not prove insights into why this structure should be rational. Two of the examples AR desktop agents in this study do not exhibit these elements, and neither agent is visible.

In these studies, the taxonomy of the holographic AI is dependent on application domains. Holographic AIs in the collaboration, training, scenario simulation, and intelligent tutor system are instructive and educational. The simulation agent utilises memetic events or plots to potentially guide users to achieve an experience, and to learn strategy and risk assessment knowledge in different situations. The holographic AI in the simulation does not strictly have to be a virtual teacher, but can be any characters which triggers tasks. The holographic AI is an intelligent tutor system provides the one-to-one tutoring (Jiménez et al., 2018). Therein, the holographic AI is a virtual teacher or coach, which no longer requires a specific surrounding to drive storyline occurrence. Norouzi et al. (2020) proposed that a collaborative holographic AI performs assistive functions (examples include personal agents and virtual coaches), while a holographic AI in a training is designed to teach a specific skill. Their classification relies on the holographic AI's roles instead of application domains; thus, a training agent may exist in both simulation and intelligent tutor systems, while a personal assistant performs a particular function, and might even be a chatbot.

The model of the holographic AI in the above three studies relies on the interactive circulation levels and key elements. Each element is not independent. The user's behaviour and context are stimuli for the holographic AI's perception and reception of interactive information, the transmission of which is governed by a decision tree.

2.3.2 Characteristics

Table 2.1 highlighted 31 definitions which capture distinctive features of the holographic AI, such as appearance, interactive surrounding, and abilities; these do not form a robust holographic AI component architecture. Therefore, the aim of this section is to investigate inclusive characteristics identified in 49 studies. The findings are grouped according to the four key characteristics of persona, intelligence, conviviality, and senses.

Regarding persona, aspects of parametric appearance incorporate size-mapping and realism to represent what holographic AIs will look like. Procedural animation refers to behavioural aspects such as facial and body animation. Intelligence can be defined as ways of interaction, and refers to the abilities which holographic AIs possess. These include environmental responsiveness, in that the holographic AI is able to interact with virtual and physical environments. Conviviality is normative sociability, referring to the ability of an agent to socialise, but aligned to an underlying purposiveness. Senses refers to the human senses augmented by the intelligent agent, i.e. its supported interaction modalities (Kim et al., 2021).

2.3.3 Persona

In terms of psychology, the term persona means a person's social face that performs in the real world (Jung, 2016). It also can present a person's personality (Leary and Allen, 2011). In the conversational agent, the term is defined as "elements of identity" (Li et al., 2016), such as profiles and ways of expression. Wang (2014) claim that a virtual persona consists of a humanlike face, behaviour, and character's background. Therefore, the persona of the holographic AI is able to define its external characteristics that can be directly observed. According to the definition of persona, this section focuses on appearances and behaviours.

2.3.3.1 Parametric appearance

Holographic AIs with realistic human-shaped appearance are widely use (see Table 2.2). According to studies' aims and functions of the holographic AI, seven studies examine how body representation, realism, size, and the presence of holographic AIs affect user experience (Kim et al., 2018; Li et al., 2018; Wang, Smith and Ruiz, 2019; K. Kim et al., 2020; Reinhardt, Hillen and Wolf, 2020; Mostajeran, Reisewitz and Steinicke, 2022; Norouzi et al., 2022).

Regarding the influence of the holographic AI's embodiment, five studies proved that the holographic AI with more humanlike features is able to enhance social presence or user perception. For example, Kim et al. (2018) have discussed whether a holographic AI requires a body representation. Their study compares a voice assistant, a holographic AI with gestures and speech, and a holographic AI with speech, gestures, and locomotion. These three types of holographic AIs interact with physical objects via Wi-Fi, such as turning a lamp on/off, to help the user check the surroundings. They found that the holographic AI with gesture and locomotion achieved the best user experience, since this assistant can increase the user's confidence. Similarly, Reinhardt, Hillen and Wolf (2020) also compared invisible voice assistants, a full-sized wireframe-shaped agent, and life-sized realistic holographic AIs. They assigned short communication tasks, such as asking time, telling a joke, and sending a message. The realistic holographic AI was found to trigger better visual attention and provide more social cues. Mostajeran et al. (2022) classified levels of anthropomorphism to range from voice assistant, virtual lip, virtual head, holographic AI with upper-body, and full-body but small-sized to measure presence. Authors find that the level of realism correlates positively with social presence and cognitive performance, but negatively with response time. K. Kim et al. (2020) further validated whether users prefer to complete a task alone, work with a voice assistant, or collaborate with a holographic AI. Their experiment was an AR desert survival task, whereby the user placed physical image markers of the survival items in order. The voice assistant and holographic AI help users to achieve a higher score, but with the holographic AI with facial expressions and behaviours it was found that it could help reduce workload and produce higher social presence. Li et al. (2018) also proved that the user prefers to interact with humanlike holographic AI compared to the robot.

On the contrary, two studies demonstrated that the humanoid holographic AI may negatively affect user experience. Wang, Smith and Ruiz (2019) investigated the impact of sizes of holographic AIs by comparing voice assistance tool with full-sized and mini-sized holographic AIs. They found that the full-sized holographic AI produced the uncanny valley effect, whereby the participants believed that size was too humanlike, even though the holographic AI's performance is not as same as that of a real human being. Therefore, the mini-sized holographic AI enjoyed a higher acceptance rate. a holographic AI with a highly realistic human appearance can leave the user feeling stressed. For example, in a study of interactivity and user affection, Norouzi et al. (2022) compare a virtual dog and a virtual female. Both types of holographic AIs acted as companions with users by performing facial and head animations, such as smiling and nodding. The participants took on a heart rate monitor and were required to finish a subtraction task under two different conditions. However, it was observed that although the virtual human was more interactive, the users experienced greater difficulty with completing the task with the virtual female and felt that the virtual dog was more supportive. The participants claimed that the virtual female's behaviour was too distracting, while the presence of the virtual dog itself engendered a sense of comfort.

Appearances	**Studies**
Life-sized and human-like	Obaid et al. (2012); Campbell et al. (2014); Kim (2018a); Peters et al. (2018); Kim et al. (2018, 2016); Kim, Bruder and Welch (2017); Kim (2018b); Li et al. (2018); Lee et al. (2018); Hartholt et al. (2019); Zielke et al. (2018); Wang, Smith and Ruiz (2019); Randhavane et al. (2019); Miller et al. (2019); Kim et al. (2019); Lee et al. (2021); Schmidt, Nunez and Steinicke (2019); Schmidt, Ariza and Steinicke (2020); Reinhardt, Hillen and Wolf (2020); Kim et al. (2021b); Pimentel and Vinkers (2021); Huang, Wild and Whitelock (2021); Mostajeran et al. (2022); Norouzi et al. (2022); Mostajeran, Reisewitz and Steinicke (2022); Yoo and Tanaka (2022); Wolf et al. (2020); Wolf et al. (2022); Huang et al.(2022)
Mini-sized and human-like	Zhou et al. (2009); Wang, Smith and Ruiz (2019); Nasution et al. (2020); Huang, Wild and Whitelock (2021)
Life-sized robot	Li et al. (2018); Peters et al. (2018); Lang, Liang and Yu (2019)
Mini-sized robot	Lang, Liang and Yu (2019); Verma et al. (2021)
life-sized cartoon	Norouzi et al. (2019); Ali et al. (2019); Huang, Wild and Whitelock (2021); Norouzi et al. (2022)
Mini-sized cartoon	Oh and Byun (2012); Aramaki and Murakami (2013); Miyake and Ito (2012); Park and Jeong (2019); Chahyana and Yesmaya (2020); Li et al. (2021); Kim et al. (2021a); Huang, Wild and Whitelock (2021);
Life-sized simplified wireframe	Reinhardt, Hillen and Wolf (2020)

Table 2.2. Appearance

2.3.3.2 Procedural animation behaviour

Most of the studies indicate that holographic AIs exhibit animations (see Table 2.3) or mention the importance of behaviours (Chu et al., 2012; Chetty and White, 2019; Norouzi et al., 2020; Huang, Wild and Whitelock, 2021). Nevertheless, not all of them explain what animations they created (Nasution et al., 2020) or identify the differences between original postures and dynamic states, such as animations of activating models (Verma et al., 2021).

In terms of body animations, gestures (n=31) account for the most, including hand wave, thumbs up, pointing, writing, playing golf, and moving tokens. ‘Idle’ standing and ‘walking’ are the most popular basic animations.

Facial animations fall into the category of display of emotions and lip-syncing. However, most papers declare that holographic AIs have emotions, but do not exemplify emotional animations. Pleasure indication such as smile expressions are very frequently used for the holographic AIs (n=4). Lip-syncing animations depend on whether the holographic AI can talk. Only one study (by Ali et al., 2019) features a holographic AI with a diversity of facial expressions (over 10) linked to the content of the speech, which exude different degrees of angry, fear, and sadness. Importantly, none of the studies below describe the smoothening of transitions between animations, especially when the holographic AI transforms from idle state to, for example, the display of an emotional reaction to its users.

Behaviours	**Studies**
Body gestures	Obaid et al. (2012); Campbell et al. (2014); Kim et al. (2016); Piumsomboon et al. (2018); Kim, Bruder and Welch (2017); Kim et al. (2018, 2019); Wang, Smith and Ruiz (2019); Kim (2018a); Kim (2018b); Li et al. (2018); Lee et al. (2021); Lang, Liang and Yu (2019); Kim et al. (2021b); Schmidt, Nunez and Nunez (2019); Ali et al. (2019); Reinhardt, Hillen and Wolf (2020); Schmidt, Ariza and Steinicke (2020); Kim et al. (2020a, 2021a); Oh and Byun (2012); Li et al. (2021); Huang, Wild and Whitelock (2021); Mostajeran, Reisewitz and Steinicke (2022)
Walking	Aramaki and Murakami (2013); Piumsomboon et al. (2018); Peters et al. (2018); Kim et al. (2018); Kim (2018a); Lee et al. (2018); Miller et al. (2019); Randhavane et al. (2019); Schmidt, Nunez and Nunez (2019); Schmidt, Ariza and Steinicke (2020); K. Kim et al. (2020); Kim et al. (2021b); Kim et al. (2021a); Yoo and Tanaka (2022)
Climbing	Kim et al. (2021b)
Jumping	Zhou et al. (2009); Lee et al. (2018)
Falling down	Zhou et al. (2009)
Sitting	Miller et al. (2019); Norouzi et al. (2022)
Idle	Miyake and Ito (2012); Lee et al. (2018); Miller et al. (2019); K. Kim et al. (2020); Kim et al. (2021b); Li et al. (2021); Huang et al. (2022); Mostajeran et al. (2022); Kim, Bruder and Welch (2017)

Head motion	Randhavane et al. (2019); K. Kim et al. (2020); Norouzi et al. (2022)
Greeting	Wang, Smith and Ruiz (2019); Randhavane et al. (2019); Pimentel and Vinkers (2021);
Motion tracking	Wolf et al. (2020, 2022)
Animal's behaviours	Norouzi et al. (2019); Chahyana and Yesmaya (2020); Huang, Wild and Whitelock (2021); Norouzi et al. (2022)
Gaze	Peters et al. (2018); Kim (2018a); Kim et al. (2019); Randhavane et al. (2019); Schmidt, Nunez and Nunez (2019); Wang, Smith and Ruiz (2019); Reinhardt, Hillen, and Wolf (2020); Schmidt, Ariza and Steinicke (2020); Pimentel and Vinkers (2021); Norouzi et al. (2022)
Speaking	Miyake and Ito (2012); Kim et al. (2016); Kim, Bruder and Welch (2017); Kim et al. (2018, 2019); Kim (2018a); Kim (2018b); Hartholt et al. (2019); Miller et al. (2019); Schmidt, Nunez and Nunez (2019); Wang, Smith and Ruiz (2019); Reinhardt, Hillen, and Wolf (2020); Kim et al. (2021b); Schmidt, Ariza and Steinicke (2020); K. Kim et al. (2020)
Facial expression	Kim (2018a); Schmidt, Nunez and Nunez (2019); Kim et al. (2019); Schmidt, Ariza and Steinicke (2020); Pimentel and Vinkers (2021); Kim et al. (2016, 2018); Kim, Bruder and Welch (2017); Kim (2018b); Zielke et al. (2018); Li et al. (2018); Ali et al. (2019); K. Kim et al. (2020); Li et al. (2021); Huang, Wild and Whitelock (2021); Huang et al. (2022)

Table 2.3. Behaviour

While, out of 49, only two studies explored that behaviours affect user experience (Li et al., 2018; Obaid et al., 2012). For example, open-arms postures of holographic AIs foster within users a greater willingness of interaction, compared to closed-arms animations (Li et al., 2019). Obaid et al. (2012) conducted an investigation into whether a holographic AI with different culture-specific behaviours can influence users' physiological response. They created two holographic AIs that simulate German and Arab cultural stereotype behaviours. They found that when either of the holographic AI did not perform the behaviours in line with users' cultural norms, the resulting psychological arousal of the users was higher.

In order to investigate a system of friendly performance, Randhavane et al. (2019) developed animation algorithms based on motion capture, and a hierarchical skeleton that ensure positions of each joint. The algorithms trigger gestures and gaze to match corresponding contexts. The friendly model consists of different degrees of arm opening (e.g. waving, arm-crossing), head movement, eye gazing, and walking gaits. In order to achieve a standard friendliness-based measure of gait, the participants in the experiment rated three levels of friendly walking, and calculated the holographic AI’s position. They reported that animations considered by the participants as more friendly generate better social presence, and result in higher levels of participants’ confidence in the holographic AI’s spatial understanding.

2.3.3.3 A summary of persona (P)

This study includes in the aspect of emotional expressiveness, because focus is on appearance aspects and character expressiveness. Therefore, investigation into the effects of the holographic AI’s persona is conducted by Kim et al. (2018), Li et al. (2018), Wang, Smith and Ruiz (2019), Obaid et al. (2012), and Randhavane et al. (2019). Their studies outline the functions of such agent. In contrast, the remaining studies employ visual representations or animations without comparative analyses of different styles or explanations of the selection of a particular holographic AI. Instead, they primarily focus on other facets of the virtual agent, such as natural language processing or the comprehension of spatial or physical objects, which leads to different objectives of the holographic AI utilization.

Based on the studies’ objectives and findings, the visual realism of the holographic AI can improve user perception. However, it is important to note potential shortcomings, such as distractions or an uncanny valley feeling, especially with the life-sized holographic AI (Norouzi et al., 2022; Wang, Smith and Ruiz, 2019) since this agent with a wider amplitude of movement attracts more visual attention in the limited interactive space, which potentially amplifies the effect of mechanized animation. Besides, Norouzi et al. (2022) prove that the user is stressed when interacting with this type of appearance. On the other hand, the humanlike but mini-sized holographic AI can alleviate the user’s pressure in collaboration (K. Kim et al., 2020). It allows users to focus on the task itself or other places that require more visual attention when he/she interacts with a mini-sized holographic AI with unobservable animations. However, the miniature holographic AI is inappropriate for all scenarios, such as face-to-face educational systems, where decision-making relies on the agent’s reaction.

The visual representation of the holographic AI has an impact on the user's perception. For example, voice assistants or wireframe-shaped agents cannot exhibit facial expressions (Reinhardt, Hillen and Wolf, 2020). Lee et al. (2018) demonstrate that the interactive impact of a robotic virtual agent is comparatively lower than that of a humanlike counterpart, attributed to users' stronger inclination to engage with real humans or robots rather than virtual agents since the latter tend to be conceived as artificial. The absence of emotions in robotic appearances further contributes to this preference. Moreover, the animal appearance is more acceptable than that of the humanoid virtual agent (Norouzi et al., 2022). Despite the virtual dog's ability to simulate human-like emotions, the user perceives it as devoid of judgment. The holographic AI’s execution of random emotional expressions, such as nodding and smiling, regardless of the accuracy of the user's answers to mental arithmetic questions, contributes to an additional sense of pressure when engaging with the humanoid holographic AI. Conversely, in the experiment conducted by K. Kim et al. (2020), the holographic AI collaborates with the user to assist the user in task completion rather than merely observing the user.

Vision is the fundamental subsystem of perception for sensing the environment (Thalmann, Musse and Kallmann, 2000; Huang, Wild and Whitelock, 2021). Body language behaviours should be able to represent how the holographic AI reacts to the user. Both believable appearance and realistic animations are indispensable to creating an element of realism (Capin et al., 1997). Further, the user's perception of interactive space is also influenced by behaviour. However, the studies mentioned above narrow their focus to a restricted set of animations for the holographic AI, such as idle, standing, walking, and arm-opening. Consequently, the visible behaviour of the holographic AI tends to exhibit a mechanical quality. The execution of natural behaviours in the physical environment should be initiated by the user's command, as remaining periods of idleness may have a detrimental impact on the user's emotional state within the context of the virtual agent's consciousness. In interpersonal communication, body language is often accompanied by speech, and individuals must ensure that their actions, particularly gestures, align with the context in which they occur. However, the existing research has not considered the semantic interpretations of gestures or their underlying motivations. The generation of gestures holds a crucial role in academic research. For example, the only interpretation of arm-opening (Li et al., 2018; Randhavane et al., 2019), 'holding the paper' gesture (Kim et al., 2019), or emotional gestures (K. Kim et al., 2020) is that the holographic AI is dynamic. The pointing gesture is represented in the study by Piumsomboon et al. (2018). To date, key gestures in holographic AI interaction constitutes one AI research area with considerable scope for further research and development.

In addition, facial animations reflect emotions or affective aspects of consciousness, and yet the holographic AI design in these studies have tended to overlook emotional reactions, and even basic facial animations, such as gaze and smiling, fail to meet basic requirements for embodiment and social interaction. The impact of dialogue and narratives in evoking emotional responses has been well acknowledged (Wright and McCarthy, 2008). Dialogue and storylines trigger emotional expressions, but none of the studies explain how to transition emotions in a realistic, believable way.

2.3.4 Intelligence

Intelligence refers to the way in which the holographic AI interacts with the user and context to achieve goals (Kim and Im, 2023; Huang, Wild and Whitelock, 2021). Out of 49 studies, 16 focus on intelligence development, such as physical-object recognition and spatial understanding, natural language processing, learning systems, computer vision, AR plugin development, and the user's motion tracking (see Table 2.4). These studies demonstrate the main methods employed for intelligence development through their titles and abstracts. For example, Kim et al., 2021a use ways of silhouettes to develop physical-object recognition, as reflected in its study title that emphasizes the enhancement of realistic interactions.

Intelligence	Studies
Spatial understanding	Lang, Liang and Yu (2019)
Physical-object recognition/interaction	Kim et al. (2021a); Zhou et al. (2009);
Natural language processing	Miyake and Ito (2012); Park and Jeong (2019); Nasution et al. (2020)
Learning systems	Oh and Byun (2012); Zielke et al. (2018); Hartholt et al. (2019); Li et al. (2021); Huang, Wild and Whitelock (2021)
Computer vision	Verma et al. (2021)
AR plugin development	Campbell et al. (2014)
Synchronization with the user's behaviours	Piumsomboon et al. (2018); Wolf et al. (2022, 2020); Yoo and Tanaka (2022)

Table 2.4. Intelligence

Figure 2.4. An example of the holographic AI with spatial understanding (Lang, Liang and Yu, 2019).

2.3.4.1 Spatial understanding

Lang, Liang and Yu (2019) have developed an enhanced holographic AI with spatial understanding, in order to locate appropriate position and orientation (see Figure 2.4). The system recognizes the geometry of the scene, and generates textures of 3D objects by applying spatial mapping and capturing a video steam. A segmentation mask produced from the steam video is used for separating colours, classifications, and boundaries, after which the key objects are mapped to the corresponding 3D models. In their study of spatially aware placement algorithms for positioning the holographic agent, the authors compared a novel approach for placement based on visibility cost, with a so-called 'traditional' approach whereby the user identifies first the relevant plane, and the agent places randomly (with orientation facing the user in angles of -30°to +30°), with positioning directly in front of the user. In their experiments, they demonstrated that the placement based on visibility cost is considerable superior, with the visibility cost calculated from visual occlusion of key objects by the virtual agent, penalizing additionally if the occlusion is more central rather than peripheral. Moreover, spatial placement takes into accounts the visual cost, penalizing whenever the agent's placement is too close or too distant.

2.3.4.2 Physical-object recognition/interaction

In order to enhance holographic AI interaction with physical objects, Kim et al. (2021a) have proposed an approach using silhouette meshes. The information of real objects is extracted from images of real animal dolls. Following the collection of images, the input size and pixel accuracy of datasets is optimized to reduce network cost. A corresponding segmentation mask consisting of real doll silhouettes enables substitution of virtual surfaces for the real object. A mobile phone camera detects the interactive objects, and it then produces bounding areas and their points in a virtual plane. Then, the volume of real dolls is calculated, and the value of the distance between the camera and the virtual floor, a parameter, is used to position the silhouette mesh. In addition, the system identifies pre-defined dolls to avoid collision and mutual occlusion. The study by Kim et al. (2021a) have assessed spatial recognition, the sense of presence, perceived naturalness, object occlusion, and physical interaction ability. The study proved that the silhouette mesh methods provide realistic interaction with the real object, and improves spatial presence as well.

Zhou et al. (2009) developed a holographic AI whose physical-object awareness consists of tracking fiducial markers pasted in real-life cubes. The holographic AI reacts to the user action and follows physical norms (for example, if the user flicks the cube, it falls down). However, although this holographic AI offered better user experience compared to that of the other holographic AI in the study which lacked physical interaction, the authors did not describe how relevant user perception data could be gathered.

2.3.4.3 Natural language processing

Figure 2.5. Translation chatbot (Nasution et al., 2020)

Three studies focused on natural language processing (Nasution et al., 2020; Park and Jeong, 2019; Miyake and Ito, 2012).

In the study by Nasution et al. (2020), the holographic AI in mobile AR bidirectionally translates from Indonesian to English, and it does not require an image marker to position itself (see Figure 2.5). The translation system has two categories. The corresponding output words depend on whether the user input can be recognized in a special category.

Likewise, an indoor dialogue agent has been developed by Park and Jeong (2019). Unfortunately, neither holographic AI can perform natural language processing via a speech-input service.

Miyake and Ito (2012) developed virtual conversational agents that can control home devices via voice commands. The smartphone camera points at an image marker of a target object to project the holographic AI into the physical surrounding, and interact with it through dialogue. The network manipulates the home device, such as turning a TV or air conditioner on or off. The study also compares a voice assistant with the holographic AI. One clear advantage of the holographic AI is that its existence enhances fluency, as it has the ability to increase accuracy of response.

2.3.4.4 Plugin development

Campbell et al. (2014) explained the motivation for employing agents in AR. Probabilistic estimation, for example, makes it possible to deal with uncertain situations in a stochastic environment. Dynamic and episodic environments require the agent to update plans for achieving goals, continually perceiving, and adjusting context in real-time. Campbell et al. devised an experiment which compared three types of navigation agents, which showed that a virtual character proved superior to an arrow system and a bubble navigation system. In order to achieve the shortest route and correct direction, they developed a plugin with multiple sensors which is able to identify the real environment. The study, however, used AR simulators, and relied on a VR system to evaluate HCI; further, the achieved experience was utilized in AR.

2.3.4.5 Synchronization with the user's behaviours

The user's motion tracking means that the holographic AI can directly reflect or preform the user's behaviours, rather than communicating with him/her. Therefore, the interactive feature of the user's motion tracking is that the holographic AI follows the user movement or interact with other partners or holographic AIs. There are four studies develop the user's motion tracking, including a remote collaborative platform (Piumsomboon et al., 2018), a walking system (Yoo and Tanaka, 2022), and two AR mirror systems (Wolf et al., 2020, 2022).

Mini-me (Piumsomboon et al., 2018) is a special case (see Figure 2.6) in that the users cannot directly interact with the holographic AIs. The authors have developed an asymmetric remote collaboration system, in which virtual avatars of VR users remotely guide local AR users in the same virtual and real environments. The AR user observes their own avatars, a life-sized avatar of the remote VR user, and a changeable sized avatar (Mini-me) which can follow the local AR user's gaze. When the local AR user gazes away from the life-sized avatar, the adaptive Mini-Me in the scene appears, and the adaptive avatar disappears when the AR user gazes back to the life-sized avatar. In the VR environment, the VR user can change viewpoints by scaling the avatar's size, and it aligns to the Mini-Me's position when it turns miniature. The experiments by Piumsomboon et al. (2018) compare the existence and non-existence of Mini-Me to evaluate social compresence and usability in asymmetric and symmetric tasks. In the asymmetric task is that the VR user navigates the AR user to pick up and place a tea box on a specific shelf, while the VR and AR users work together in the symmetric task. Consequently, the Mini-me system achieves a higher social presence, can save time, and reduce the difficulty of cooperative work.

Figure 2.6. MiniMe (Piumsomboon et al., 2018)

Other three studies report notable developments in user's motion tracking that allows the holographic AI follows the user's movement. For example, Yoo and Tanaka (2022) have designed a virtual walking companion to motivate physical exercise. The life-sized, realistic, and personalized avatar has two modes: walking with the holographic AI, or walking with a remote user's holographic AI. The holographic AI is able to avoid physical collision, recognize the user's speed, capture the remote user's camera position to implement synchronized movement, and track the user's gaze to adjust direction. However, the holographic AI only can interact in a pre-defined scenario, and so users cannot casually select a position for the holographic AI.

Wolf et al. (2020, 2022) have developed an AR mirror system for users with obesity and misperception of body weight. In the first study, the application relies on an AR and VR see-through devices, in which the user's pose and behaviour is tracked and displayed by an avatar. The system detects movements of the head, hand, hip, and feet, and then it generates a humanoid body pose. The authors then explore body weight perception in terms of the user's body-mass-index by analysing the difference between the avatar real body-mass-index and the user's feeling of the avatar's weight. The user wears VR and AR devices and follows instructions to perform movements, such as raising arms, waving, and stretching arms. The sense of spatial presence, however, is lower in the AR system, but the embodiment level is similar in both systems. In their second study, Wolf et al. (2022) proposed an improved AR system based on optical see-through. For this, they set up eight cameras on the ceiling to conduct markless motion capture, and head and hand movement recording by wearable AR headset so that the user can observes the avatar performs the same behaviour. However, the avatar cannot fully perform each corresponding movement, especially elbows and fingers, and even the new AR system has a lower sense of presence.

2.3.4.6 Learning systems

An intelligent tutoring system can be defined as a computer-based systems that provide a natural language interface and deploys adaptive educational strategies to support learners (Sottilare, 2018). Out of 16, 4 studies mention or develop learning systems (Oh and Byun, 2012; Zielke et al., 2018; Hartholt et al., 2019; Li et al., 2021).

For example, Hartholt et al. (2019) develop virtual humans to help users with Autism Spectrum Disorder (ASD) improve their job interviewing techniques. Their holographic AI uses an automated review phase to generate real-time feedback by analysing eye gaze, eye blinking rate, head orientation, speech delay, and speaking volume. The study, however, does not provide details on the accuracy of generated feedback.

Zielke et al. (2018) introduced an emotive virtual human system which aids remote clinical training of medical students. The holographic AI acts as a virtual patient which exhibits symptoms and is designed to improve learners' body language interaction. In addition, the holographic AI can mimic real teachers when providing feedback, while real tutors evaluate the student performance by remotely observing the student's performance.

Oh and Byun (2012) created an interactive virtual bird which teaches children how to grow flowers. When the user ignores the holographic AI's feedback and selects the wrong factors, the holographic AI adopts an anxious expression, and the flowers start to fade in colour. The holographic AI is superimposed on the pages of a physical book and tracks the pages of the book. In the experiment, it was observed that the child participants preferred to interact with the animated holographic AI, rather than a static one.

The FaceMe system is a 3D AR game displayed on a computer, which is designed to help children with ASD learn how to recognize and perform diverse types of facial expressions (Li et al., 2021). Wooden cubes labelled with image markers manipulate the emotions of a 3D character model. These help them remember, select, and follow the facial expressions that the holographic AI expresses. In this step, facial recognition assesses children's performance. The survey results in Li et al.'s study provide evidence showing that children can identify rude emotions whenever a holographic AI adopts an angry countenance.

2.3.4.7 Computer vision

In computer vision, the image frame undergoes a pre-processing stage to identify and extract the edges and contours of the marker image. Computer vision is essential for spatial understanding and physical-object awareness or interaction (Ghasemi et al., 2022). Nevertheless, the studies as mentioned above, such as the FaceMe system (Li et al., 2021), synchronization with the user's motion (Piumsomboon et al., 2018; Wolf et al., 2022, 2020; Yoo and Tanaka, 2022), the utilization of silhouette meshes (Kim et al., 2021a), and spatial understanding (Lang, Liang and Yu, 2019), fail to provide a comprehensive illustration of the application or advancement of computer vision technology. For instance, Lang, Liang and Yu (2019) directly employ spatial mapping. Kim et al. (2021a) utilize deep learning techniques to extract data pertaining to the silhouette of actual dolls, with a specific emphasis on understanding physical objects.

On the other hand, Verma et al. (2021) presented low-level OpenCV for marker tracking to avoid radial distortions. The holographic AI detects and recognizes real words, translates English to French, and executes voice commands. Although this method could be replaced with diverse plugins and technologies, the latter offers limited advantages. For example, Vuforia is an AR mobile development kit to track marker, which projects the holographic AI by a way of a markerless technique. Although the holographic AI proposed by Verma et al. can perform voice chat, physical word recognition, and translation, the authors do not explain how these functions work.

2.3.4.8 A summary of Intelligence (I)

The aforementioned experiments explicitly explore the development of holographic AI, encompassing eight distinct domains: spatial understanding, physical-object recognition/interaction, natural language processing, AR plugin development, computer vision, learning systems, adaptivity, and synchronization with users' behaviours. Additionally, some studies assess user experience in relation to intelligence development.

Natural language processing is one of the main channels of interaction of AI systems, and it relies on using a large corpus to find related and appropriate output. Natural language processing, however, lacks the capacity to enter into a debate with the user in such a way as to project a holographic AI that provides opinions conflict with those of the user.

In terms of physical interaction and awareness, the holographic AI can capture and reconstruct real environments and objects to place the holographic AI (Kim et al., 2021a; Lang, Liang and Yu, 2019). Although aforementioned studies declare that the holographic AI is able to recognize real surroundings and physical objects, this does not imply it can manipulate physical objects. Only five studies clearly display the holographic AI capable of performing physical-object interaction via Wi-Fi, sensors, and activated systems (Schmidt, Nunez and Steinicke, 2019; Schmidt, Ariza and Steinicke, 2020; Lee et al., 2021; Miyake and Ito, 2012; Huang, Wild and Whitelock, 2021). For example, Azuma Hikari (see Huang, Wild and Whitelock, 2021) can reciprocally exchange information on both the virtual and physical surroundings. Using multisensory technology, it can check room temperature, manipulate physical objects, and send messages to the user.

Moreover, the user model and adaptivity constitute primary characteristics of holographic AI (Aroyo et al., 2006). An adaptative holographic AI can maintain user models through engagement in processes such as storing, updating, removing, and encapsulating assumptions related to various aspects of the user's plans and tasks, preferences, goals, and knowledge levels. This functionality enables the implementation of customized services. The user model is crucial in educational systems as well, although only two studies introduce how adaptive systems can collect and analyse user information (Huang, Wild and Whitelock, 2021; Hartholt et al., 2019). Nevertheless, Hartholt et al. (2019) represent that a holographic AI as monitoring user performance and generating corresponding feedback. Besides, the user model and adaptivity can effectively handle unforeseen circumstances or predict the user's utterance and performance. This adaptivity can be achieved by the utilization of probabilistic estimate techniques (Campbell et al., 2014) or the implementation of predictive models (Skantze, 2021).

2.3.5 Conviviality

Conviviality aspects	Studies
Co-presence and social presence	Kim et al. (2021b); Kim (2018a); Kim, Bruder and Welch (2017); Pimentel and Vinkers (2021); Kim et al. (2019); Lee et al. (2021); Schmidt, Ariza, and Steinicke (2020); Reinhardt, Hillen, and Wolf (2020); Kim et al. (2016); Kim (2018b); Miller et al. (2019); Schmidt, Nunez and Steinicke (2019); Norouzi et al. (2019)
Social facilitation and inhibition	Mostajeran, Reisewitz and Steinicke (2022)
Proxemics	Aramaki and Murakami (2013); Li et al. (2018); Lee et al. (2018); Peters et al. (2018); Huang et al. (2022);

Table 2.5. Conviviality

Conviviality can reflect a quality of social interaction (Caire, 2010; Caire and van der Torre, 2009). Therefore, conviviality refers to the social, but purposive, ability of individuals to interact with others, typically targeted to satisfying their (and own) needs.

In relation to the dimension of conviviality, 20 studies investigate user experience, including co-presence, proxemics, and social facilitation and inhibition (see table 2.5). As previously indicated, the focus of these holographic AIs lies in intelligence to engage in physical-object awareness or natural language processing. However, their intended purpose is to gauge user perception rather than foster intelligence development. Consequently, this section does not delve into studies that assess intelligence through user perception.

2.3.5.1 A sense of co-presence

Two studies by Kim et al. (2021b) and Kim, Bruder and Welch (2017) examine the impact of conflict with physical objects on social presence. Kim et al. (2021b) created fade-in/out and flare-up effects and set out to determine which is able to improve the sense of co-presence, when holographic AIs conflict with the physical objects and real human. The study compares four conditions: two different visual effects, collision-free conditions, and overlapping with real objects. In the collision-free condition, the holographic AI recognizes real objects and avoids collision. As a result, fade-in/out provides a less noticeable but more comfortable feeling. The collision-free setting is found to be the most optimal. Although visual effects are able to influence user experience, they seem too deliberated. Earlier, Kim, Bruder and Welch (2017) also investigated overall user experience towards the holographic AI. In that study, the holographic AI sits in a wheelchair and moves around the room, and it either passes through real objects or avoids collision. The authors measured walking trajectory to observe whether the user passes through or avoids collusion with the holographic AI. The project results indicated that conflicts produce negative perception, but that the user might not necessarily perform natural locomotion behaviour to avoid conflicts.

Figure 2.7. Measuring the sense of co-presence (Pimentel and Vinkers, 2021)

Two studies investigate how physical-object and spatial understanding influence the sense of co-presence (Kim et al., 2019; Pimentel and Vinkers, 2021). Pimentel and Vinkers (2021) have considered how the reaction of a holographic AI impact the feeling of presence in a physical event. For example, the holographic AI in Figure 2.7 does/does not turn its head, and gazes at a real fallen broom. The sense of co-presence refers to how one feels 'there', how the holographic AI responds to this place, and whether the interactive environment feels like reality (Bulu, 2012). In addition, Kim et al. have considered how contextual responsiveness affects social presence (Kim et al., 2019). This study consisted of two experiments, although the first task applied a projection screen to display a virtual human, a technology which falls outside the scope of this thesis. One scenario showed a virtual sheet of paper and curtain fluttering in the breeze of a physical fan whilst the holographic AI interviews the user. The holographic AI either does or does not recognize this situation, in the former case by restraining the fluttering paper and looking at the fan. Both studies confirm that the highest sense of co-presence results from the holographic AI with spatial awareness.

Norouzi et al. (2019) designed virtual walking dogs in order to investigate the influence of the AR animal on user experience and behaviour. In their experiment, the participant is permitted to give commands, such as walking, sitting, digging, and drinking. The experiment consists of the following conditions: the experimenter wears/not wearing a headset, and exhibiting recognition/non-recognition of the dog, and the dog reacting/not reacting to the real human's foot movements. By comparing the walking path of the user in the existence and absence of the virtual dog, the authors observed that the participants assigned space for the virtual dog in a walking situation; that the presence of the virtual animal affects users' behaviour, locomotion levels, and walking speed; and, that the virtual dog exuded a higher sense of co-presence when it recognizes the experimenter.

Similarly, three studies focus on how physical-object interaction influences emotional responses of the user, co-presence, and perceived plausibility (Lee et al., 2018; Lee et al., 2021; Schmidt, Nunez and Steinicke, 2019; Schmidt, Ariza and Steinicke, 2020).

Lee et al. (2021) created a board game, within which, to provide a holographic AI with physical illusions, a magnetic actuator surface controls the physical tokens in a game, and a magnet attracts metal tokens. A motorized translation stage moves in parallel the magnet on a tabletop, and a tracking system conveys information of movement. In their first experiment, the holographic AI controls a virtual token, thus making comparisons with the moving physical token. The aim behind the physical interaction is to optimise the sense of co-presence and user experience, so that the user experiences a sense of expecting the holographic AI to perform like a real human being. The authors observed that the participants could notice a lag between the gesture of the holographic AI and the movement of the actual token, and so their study assesses the user's perception of latency with regard to directionality and magnitude. Slight latency is able to improve the level of realism.

Kim et al. (2016) conducted an experiment to determine the influence of the user's personality on social presence, in which they compared ignoring/inconsistency and requesting/consistency conditions. In the inconsistency condition, the holographic AI cannot recognize and avoid physical objects; with the opposite condition it can avoid collisions and asks for help from users to move real objects. Each experiment commenced with the holographic AI assessing the user's extraversion and introversion, in order to explore how personality influences social presence, and how the user evaluates interactive experience according to their emotions. The authors found that extrovert users are more likely to experience a higher level of social presence whenever the holographic AI asks for help, and that they maintain longer eye contact than introverts.

Lee et al. (2018) considered how visual factors influence the locomotion behaviour and proxemics of holographic AIs. Their study tested two visual settings: the user being able to observe central and unpigmented regions (unrestricted condition), whereby the user can see the holographic AI appear and disappear; and the user only seeing the central field (restricted condition). The primary obstacles are a real human and a full-sized holographic AI, and the secondary obstacles are virtual and real human movements. In the study, two levels of vibrotactile footstep feedback for the holographic AI were analysed, whereby the user could sense a vibration whenever the holographic AI jumps and walks. It was observed that the social distance with the holographic AI was large in the restricted condition, with less walking and speeding since the user interacts with the holographic AI infrequently, and the AI cannot change its movements to avoid the collision. The study also reports that in unrestricted conditions, the user glances less frequently at the real human, and the obstacle position affects spatial awareness. Further, the vibrotactile footstep feedback increases the sense of co-presence.

Schmidt, Nunez and Nunez (2019) and Schmidt, Ariza and Steinicke (2020) have developed a robotic ball and a thermal table with which a holographic AI can interact. The robotic ball simulates the pathway and movement of a real golf ball based on a script, and temperature changes trigger thermochromic ink to write on paper. Their experiments consisted of four situations related to interaction abilities: virtual/real golf ball interaction, and virtual/real writing. As a result, it proves that physical-object interaction can produce positive effect of the user's emotional responsiveness.

2.3.5.2 Social distance

Li et al. (2018) conducted an experiment on social distance and sense of presence in interacting with a real human, a real robot, a virtual robot, and a humanlike holographic AI. On the basis of user preference and willingness of interactive characters, the participants firstly ranked pictures of different postures, including smiling, open arms, and head orientation. Over the course of the experiment, the authors varied the user-to-agent distance at which the agent's behaviour would be triggered by the user's approach. (For example, the agent might activate arm stretching once the user comes within 2.5 meters.) The authors observed that interaction with the real human was connected with the highest level of social presence, interaction level, body language, and interdependence. In other words, the user is more willing to interact with animated humanlike agents with emotions, rather than with that bearing a robot appearance. They also found that a higher willingness for interaction was engendered by a wider openness of the agent's arms. The level of social interaction, however, did not appear to be closely dictated by the five postures (neutral, open, and closed postures, facing forward direction, and head orientation) and three emotions (positive/negative/neutral).

Aramaki and Murakami (2013) investigated the spatial relationship between users and holographic AIs in order to develop a communicative system, and to determine whether social distance influences the holographic AI's size and height. In the first experiment, the holographic AI's height was set at 0.20 meters, and it was placed an initial distance of 1 meter away from the users. In their experiment, the holographic AI walks closer or steps further away until the user feels comfortable enough to communicate with it. It was found that 0.43 meters was the preferred distance. The heights of the holographic AI in the second experiment were 0.5 meters, 0.15 meters, and 0.25 meters. During the experiments the holographic AI's height and size were altered in an effort to determine the optimum proportions from the users' perspective. It was observed that the most appropriate height and distance of the holographic AI with 18 cm and 0.7 meters.

Peters et al. (2018) explored social distance between real humans, full-sized holographic AIs, and two virtual robots of different sizes. Participants rated each agent's appearance, height, gender, likeability, and realism. The original distance is 3 meters, and the general stopping distance between agents and users was observed to be 1.23 meters.

Huang et al. (2022) have designed a scenario, where participants ask holographic AIs' help for directions in an AR gallery, in order to understand how the holographic AI's personal space affects users' behaviour and arousal. The users walk toward or pass through the holographic AI from the original distance of 2.5 meters until they feel comfortable enough to communicate with it. Almost 74% of the participants in this study preferred a distance range of 0.46– 1.22 meters, and a closer social distance when the anthropomorphic holographic AI appears. In addition, Huang et al. set obstacles, including asking participants to walk through the holographic AI to produce a value of physiological arousal caused by electrodermal activity. Passing through the holographic AI increases participants' skin conductance level, leading to physiological arousal.

2.3.5.3 Social facilitation and inhibition

Miller et al. (2019) have demonstrated that the present and absent of the holographic AI can influence social facilitation and inhibition, as well as user performance. In their experiment, participants can complete more easy anagram tasks when the holographic AI exists. While social inhibition is confirmed by participants solving more difficult tasks in alone environment. The second experiment in this study tested whether the user follows social norms when interacting with the holographic AI. In it, the holographic AI sits on one of two real chairs, and the user puts on an AR headset in order to select a chair – as opposed to the other condition, where the user takes off the headset after the hologram sits down, and then selects a chair. It was observed that the participants tended to follow social norms and avoided sitting in the same chair as the holographic AI.

Mostajeran, Reisewitz and Steinicke (2022) have recently investigated the effects of holographic AI social facilitation and inhibition as well. In their experiment, the holographic AI maintains eye contact with users, and follows their heads in two tasks. In the cognitive task, the participants had to quickly repeat lists of numbers, and it was found that they made more mistakes in the presence of the real human and holographic AI, which, respectively causes social facilitation and inhibition. The second coordination task required the participants to stand on a balance board either with two feet or on one foot. It was observed that the participants performed better when assisted by a real human, but struggled to maintain their balance on one foot in the presence of a real person and holographic AI.

2.3.5.4 A summary of conviviality (C)

The Woz methodology represents a technique wherein experimenters simulate AI by manually orchestrating behaviours and responses. In this approach, a human operator clandestinely manipulates the actions and responses of a virtual entity, thereby preserving the illusion of autonomous AI for the user. Within the scope of user interaction, 14 distinct studies have employed remote control to activate predetermined animations and dialogue choices (Aramaki and Murakami, 2013; Kim, 2018a,b; Li et al., 2018; Norouzi et al., 2019; Kim et al., 2016, 2018, 2019; Kim, Bruder, and Welch, 2017; Schmidt, Nunez, and Steinicke, 2019; Schmidt, Ariza, and Steinicke, 2020; Kim et al., 2021b; Pimentel and Vinkers, 2021; Li et al., 2021). In the domain of user experience research, it is evident that the holographic AI's capacity for perceiving dynamic events is not an intrinsic feature. Consequently, the primary aim of employing holographic AI in these studies is to evaluate user experience with respect to the perceived capabilities of the system, rather than to advance the AI's intelligence. Furthermore, while investigations into the effects of persona on user interaction have been identified and scrutinized in Section 2.3.3, this section avoids reiterating those findings to eschew redundancy.

In interactions with holographic AI, users tend to adhere to societal norms, actively avoiding discord or friction with the AI system. Moreover, there appears to be a user preference for smaller-scale holographic AI representations, with dimensions approximately between 0.15 and 0.18 meters (Aramaki and Murakami, 2013;

Wang, Smith and Ruiz, 2019), as evidenced by the utilization of miniaturized holographic AIs in the study by Aramaki and Murakami (2013). It is also noteworthy that the interpersonal distance between the user and the holographic AI is not uniform, varying from 0.43 to 1.22 meters, indicating variability in the spatial comfort levels of users during interaction.

Miller et al. (2019) and Mostajeran, Reisewitz and Steinicke (2022) have elucidated that the presence and absence of holographic AI can exert both facilitative and inhibitory effects on user interaction. In a distinct investigation, Kim et al. (2020) found that the incorporation of holographic AI can engender positive outcomes in the context of collaborative tasks and services. Users interact with a holographic AI characterized by augmented responsiveness, superior spatial awareness, and the capacity for physical-object interaction, which collectively foster a more pronounced perception of human-like qualities. However, it is noted that the holographic AI is limited to performing static animations such as standing or sitting.

A multitude of research endeavours have sought to delineate the myriad factors influencing the efficacy of holographic AI, including but not limited to physical-object interaction, persona recognition, and user personality traits, given the inherently hybrid interactive milieu in which the holographic AI operates. Yet, the influence of dialogue content on user experience remains an area shrouded in ambiguity. Prior research has also not satisfactorily explored how the intelligence or perceived sentience produced by holographic AI might facilitate the establishment of trust between the user and the AI entity. The social presence evoked by holographic AI can engender a semblance of tangibility within a co-occupied space, prompting inquiries into the user's trust perception towards the holographic AI. In light of these considerations, this dissertation intends to scrutinize the construct of trust as it pertains to user conviviality in Chapter 5, thereby endeavouring to rectify the identified research shortfall.

2.3.6 Senses

Both parts pertaining to intelligence and conviviality have addressed the use and impact of holographic AI perception, specifically in terms of spatial understanding and awareness of physical objects. However, these intelligence sections lay no stress on the specific perception abilities employed by the holographic AI. Consequently, it is challenging to comprehend the mechanisms through which the virtual agent acquires information and its capacity to generate a corresponding response for sustaining ongoing interaction. Therefore, this section discusses the senses that the holographic AI uses.

The human perception includes vision, audition, touch, smell, and taste, and modality applies a particular sense to receiving stimuli (Turk, 2014). In terms of HCI, modality can be seen as corresponding to the human senses. For instance, a camera can be considered as sight, and a microphone considered hearing. Both appearance and behaviour can be seen as visual effect, language is an auditory sense. Therefore, the sense can be regarded as a ‘modality’ of interaction that converts the provided information into manifested behaviour, mirroring the motivations and intentions of the virtual agent (Blanke and Metzinger, 2009).

One sensory modality of holographic AIs refers to a single input or output pathway, such as gesture recognition, natural language processing, and facial recognition. If used in isolation, it is unimodal (Nizam et al., 2018). For example, although the holographic AI can perform basic movements, but its animations do not factor in the user's behaviour or motivations, as noted by Nasution et al. (2020). A holographic AI system developed by Zhou et al. (2009) has the capacity to recognize physical objects, yet it lacks the ability to engage in more complex interactions with users.

Nevertheless, sensing the external environment does not rely solely on one channel, and multimodal interaction is prevalent. Take the FaceMe system as another example; it integrates holographic AI that relies on the implementation of natural language processing and facial recognition technologies, as documented by Li et al. (2021). Besides, research by Schmidt, Nuñez and Steinicke (2019), along with Schmidt, Ariza and Steinicke (2020), introduced a robotic ball and a thermal table that create the illusion of handwriting and playing golf, respectively. Kim et al. (2018) utilized Wi-Fi technology to control a lamp, thus creating the illusion of a holographic AI interacting with physical objects.

Natural language processing as key interaction modality is typically combined with animation, physical-object awareness, eye gaze tracking, position detection, or posture interaction. However, since animations or verbal communication cannot actually interact with the user, the performance of holographic AIs only reflects the current interactive situation or serves to improve immersion, which cannot influence the user behaviour. Interaction in human-computer relationship refers to the behaviours of two entities that can influence each other (Hornbæk and Oulasvirta, 2017). For example, the holographic AI in Lee et al. (2021) uses recorded audio to show the game result, whereby the holographic AI says ’yes’, when it wins.

Interaction modalities	**Studies**
Non-verbal communication interaction	Zhou et al. (2009); Holz et al. (2011); Campbell et al. (2014); Piumsomboon et al. (2018); Li et al. (2018); Miller et al. (2019); Pimentel and Vinkers (2021)
Verbal interaction	Miyake and Ito (2012); Oh and Byun (2012); Zielke et al. (2018); Hartholt et al., (2019); Wang, Smith and Ruiz (2019); Lang, Liang and Yu (2019); Schmidt, Nunez and Steinicke (2019); Ali et al. (2019); Reinhardt, Hillen, and Wolf (2020); Schmidt, Ariza and Steinicke (2020); Huang, Wild and Whitelock (2021); Kim et al., (2021a)
Physical-object awareness	Holz et al. (2011); Lang, Liang and Yu (2019); Schmidt, Nunez and Nunez (2019); Schmidt, Ariza and Steinicke (2020); Huang, Wild and Whitelock (2021); Kim et al. (2021a,b)

Eye gaze tracking	Ali et al. (2019); Hartholt et al. (2019)
Position detection	Park and Jeong (2019)
Posture interaction	Li et al. (2018)

Table 2.6. Interaction modalities

2.4 Validation Methodologies in the Literature

Experimental designs found in the literature concerning holographic AIs fall into two classes: comparative experiments, and controlled experiments.

The comparative experiment aims to determine which of the conditions is most likely to produce a positive or negative impact on user experience. For example, the independent variables could be hitting a golf ball and handwriting, and the dependent variables could be physical or virtual interaction (Schmidt, Ariza and Steinicke, 2020).

Several cited studies (e.g. Kim et al., 2016; Zielke et al., 2018; Wolf et al., 2020) use the same experimental design, whereby participants are assigned to a condition. For example, Miller et al. (2019) conducted two experiments. For the second experiment, they developed 4 conditions and 24 interactive orderings, and invited 60 participants to complete an ordering of the four conditions. The holographic AI's presence or absence is a between-subjects factor, while task-based difficulty versus ease is a within-subjects factor. The main difference is that time influences the effects of the former factor. In the first experiment, the participant randomly assigned one of the four conditions into sequence, in an effort to increase their anagrams puzzle scores over time. In the second experiment, the user randomly chooses one of conditions either wearing the AR headset or without the AR headset.

In randomised-block design, users (participants) are divided into different groups, and each group randomly assigned to testing a condition. One example is the study by Campbell et al. (2014), in which 3 groups of 54 participants test one of negative agents. Similarly, the study by Zielke et al. (2018) randomly assigned participants into three groups - monitor-based and VR, monitor-based and AR, and VR and AR - to compare the efficiency of delivery methods of medical interviewing.

There are 16 studies which have implemented within-subjects studies, whereby each participant repeatedly evaluates all conditions (Zhou et al., 2009; Miyake and Ito, 2012; Kim et al., 2018; Kim, Bruder and Welch, 2017; Li et al., 2018; Piumsomboon et al., 2018; Lee et al., 2018; Norouzi et al., 2019; Randhavane et al., 2019; Lee et al., 2021; Wang, Smith and Ruiz, 2019; Kim et al., 2021a; Huang et al., 2022; Mostajeran, Reisewitz and Steinicke, 2022; Norouzi et al., 2022; Mostajeran et al., 2022). For example, Kim et al. (2021a) have recently compared the silhouette mesh method with conditions of free-occlusion and faulty occlusion, by arranging for 30 participants to watch three videos' clips. Li et al. (2018) conducted an experiment that is similar to that of Miller et al. (2019). In this case, each participant could access a total of 24 possible trials, with four types of agents, and randomly chosen orderings. In another experiment, Lang, Liang and Yu (2019) divided participants into three groups of 10 randomly and tasked them with comparing each task in order. Each participant scored the holographic AI's performance in terms of location, direction, and overall impression.

A controlled experiment requires a controlled group to reduce or control the influence of factors other than the independent variable. In a recent experiment, Kim et al. (2021b) compared four conditions with two types of visual effects, with overlapping with the physical objects being the control condition. It consists of six steps whereby the holographic AI passes through different physical objects. Therefore, the 16 participants in the experiment need to repeat similar tasks with multiple levels of independent variables. Other independent variables, such as interaction time, collisions with virtual parts of the body, and gestures, were isolated. K. Kim et al. (2020) have conducted a within-subjects study, in which 36 participants completed a task three times. The baseline condition is that the user performs the survival task alone, and the voice assistant and holographic AI are the experimental groups.

The study by Pimentel and Vinkers (2021) analysed the users' actions by testing each condition. In their experiment the control condition is defined as no event occurrence in the real space, while a comparative condition is the holographic AI either ignoring or reacting to the falling broom. A mixed within-between design defines the interactive environment as a within-subjects factor, while holographic AI responsiveness is a between-subjects factor.

Kim et al. (2019) evaluated the physical/virtual responsiveness of the holographic AI using a within-subjects design with three conditions. Reinhardt, Hillen, and Wolf (2020) applied within-subjects design so that all 18 participants in their experiment could assess three tasks with different levels of realism of agents. The authors report that Latin square design can be used to avoid ordering of the three tasks, and that it is the same as the method by Miller et al. (2019). The problem with the Latin square method, however, is that it does not indicate what the control condition is.

Wolf et al. (2022) have conducted a controlled comparative study with between design to investigate optical see-through AR, video see-through AR, and video see-through VR technologies. The studies by Nasution et al. (2020) and Oh and Byun (2012) lack detailed reasoning on the assigning of participants, as well as the contents of questionnaires. Even though the study by Oh and Byun (2012) compared the levels of child satisfaction in the conditions of animated and static holographic AIs, it does not demonstrate the effectiveness of the holographic AI in completing a collaborative task. As mentioned, the two studies merely list the experimental results, without analysing the user experience in depth.

The methodologies of the studies include comparative/controlled experiments with within/between subject designs. Some studies do not clearly explain how long each experiment lasts (e.g. Lee et al., 2018; Lee et al., 2021; Kim et al., 2019; Norouzi et al., 2019; Reinhardt, Hillen and Wolf, 2020; Nasution et al., 2020; Kim et al., 2021b), or indicate whether the user has enough time to engage in interaction with the holographic AI. In the within-subjects design, each participant must test all conditions in the same scenarios with different conditions, a long and repetitive process which results in eye strain and depleting immersion, as reported by Reinhardt, Hillen and Wolf (2020). With the between-subjects design there exists the problem of individual variability, and the group lacks homogeneity. Another concern with the cited studies is that participant numbers vary greatly, from 9 (Miyake and Ito, 2012) to 65 participants (Pimentel and Vinkers, 2021).

These studies rely on questionnaires, interviews, open-ended questions, behavioural measurement, general feedback, and scoring for holographic AIs and user choice (Lang, Liang and Yu, 2019; Miller et al., 2019). Even though the questionnaires might have been extract from other studies, authors tend not to explain how and why they incorporate or combine certain statements or questions to make a new questionnaire, and the extracted questionnaires are not specifically created for the AR environment or holographic AIs.

For this review of applied methodologies, it was decided to exclude three papers (Kim, 2018a,b; Schmidt, Nunez and Steinicke, 2019), since they refer to the same studies (Kim et al., 2019, 2018, 2016; Kim, Bruder and Welch, 2017; Schmidt, Ariza and Steinicke, 2020). Moreover, Nasution et al. (2020) and Li et al. (2021) conducted surveys, but did not reveal how they collected data from the users.

Experimental design	**Studies**
Comparative experiments	Zhou et al. (2009); Miyake and Ito (2012); Campbell et al. (2014); Lee et al. (2018); Zielke et al. (2018); Piumsomboon et al. (2018); Wang, Smith and Ruiz (2019); Lee et al. (2021); Lang, Liang and Yu (2019); Norouzi et al. (2019); Kim et al. (2021a); Schmidt, Ariza, and Steinicke (2020) Kim et al. (2016, 2018); Kim, Bruder and Welch (2017); Li et al. (2018); Miller et al. (2019); Wolf et al. (2020, 2022)
Controlled experiments	Kim et al. (2019, 2020a); Reinhardt, Hillen, and Wolf (2020); Kim et al. (2021b); Pimentel and Vinkers (2021)
Questionnaire survey	Zielke et al. (2018); Lee et al. (2021); Schmidt, Nunez and Steinicke (2019); Schmidt, Ariza and Steinicke (2020); Kim et al. (2016, 2017, 2018, 2019); Li et al. (2018); Wang, Smith and Ruiz (2019); Reinhardt, Hillen, and Wolf (2020); Nasution et al. (2020); Kim et al. (2021b,a); Pimentel and Vinkers (2021); Miyake and Ito (2012); Oh and Byun (2012); Campbell et al. (2014); Piumsomboon et al. (2018); Norouzi et al. (2019); Lee et al. (2018); Wolf et al. (2020, 2022)
Interview	Li et al. (2018); Zielke et al. (2018); Norouzi et al. (2019); Lee et al. (2018); Reinhardt, Hillen, and Wolf (2020); K. Kim et al. (2020); Wolf et al. (2022)

Open-ended questions	Norouzi et al. (2019); Lee et al. (2021); Schmidt et al. (2020, 2019)
Behavioural measurement	Kim et al. (2016); Lee et al. (2018); Norouzi et al. (2019); Lee et al. (2021); Kim et al. (2021b)
General feedback	Piumsomboon et al. (2018); Kim et al. (2021b,a)
Questionnaires extracted from other studies	Kim et al. (2016, 2018, 2019); Kim, Bruder and Welch (2017); Li et al. (2018); Piumsomboon et al. (2018); Lee et al. (2018);Lee et al. (2019); Wang, Smith and Ruiz (2019); Schmidt, Ariza and Steinicke (2020); K. Kim et al. (2020); Norouzi et al. (2019); Reinhardt, Hillen and Wolf (2020); Kim et al. (2021b,a); Wolf et al. (2020, 2022)

Table 2.7. Experimental Design

2.5 Discussion

This section summaries the features of all holographic AIs pertinent to RQ1 and RQ1-1, it also analyses shortcomings through the lens of 49 identified articles and provides corresponding potential solutions and future development tendency.

The concept of the holographic AI depends on the areas of interest and functionality. Some concepts, however, do not explain differences between AR agents, VR virtual assistants, and screen-based agents. Over half of the studies provides inadequate definitions of interactive space. Although these studies emphasize spatial relationship or environmental aware behaviour, the holographic AI in these studies only performs virtual interaction or possess physical/spatial understanding. Furthermore, a holographic AI does not exist in the real world, nor does it share the same space with the user; instead, the holographic AI overlays on or embodies in the real world. The words ’virtual human’ and ’virtual agent’ cannot be equated, as they have different limitations in terms of appearances and behaviours. Thus, the concept of the holographic AI is an embodied virtual agent that exits in the MR world to play different roles, and which responds to and even influences virtual and physical surroundings.

2.5.1 Model of the holographic AI

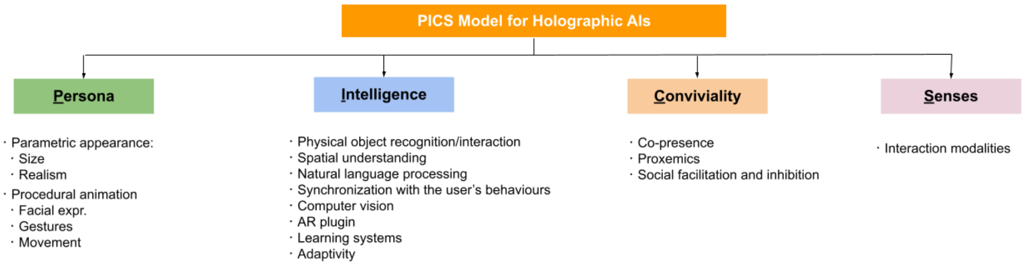


Figure 2.8. PICS model for Holographic AIs

The total of 49 studies were reviewed, showing that holographic AIs involve four major hallmarks that can be integrated into a novel model. This section proposes the following novel model for holographic AIs – persona (P), intelligence (I), conviviality (C), and senses (S) – PICS (see Figure 2.8).

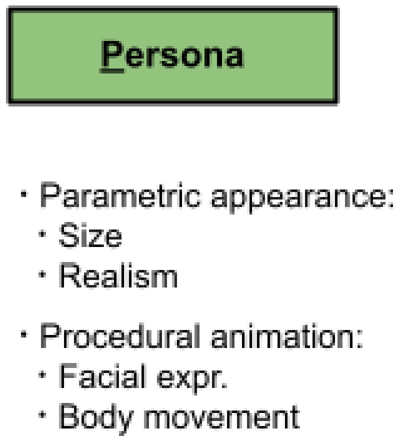


Figure 2.9. Persona

Persona identifies a specific human being (Burden and Savin-Baden, 2019), and is a manifestation of the holographic AI by way of its body and behaviour (see Figure 2.9). Visual identity portrays a person’s physical features, personality, and even abilities for social interaction. Most of the cited studies feature photorealistic, humanlike, animated, and life-sized virtual agents. One study (Wang, Smith and Ruiz, 2019) suggests that shrinking a humanoid holographic AI to miniature size helps to reduce the uncanny valley effect. However, such an agent is unsuitable for intelligent tutor systems in interview and medical training. In terms of behaviour, a holographic AI cannot solely perform standing and waiting, as it cannot observe differences. Gaze, lip-sync animations, neural, and positive and negative expressions are commonplace emotional expressions. While the studies highlight plausible behaviours of holographic AIs that can affect user perception, their holographic AIs do not show clear and manifold emotional expressions that are aligned with the body animations and contexts. Meanwhile, some animations cannot influence the user decisions and performance that do not belong to interaction modalities. None of the cited studies investigates which gestures can affect the holographic AI’s effect, or how users understand non-verbal interaction. For example, arm-opening and crossing, waving, and slightly arm gestures do not have communicative meanings; these basic gestures merely prove that the holographic AI is dynamic.

Persona and function belong in the same horizontal line for conveying precise body language and emotions. By contrast, monotonous emotional expressions (e.g. pleasant sensation) and movement (e.g. repeated gestures and standing) do not benefit the performance or identify potential cues. For example, although a holographic AI can recognize physical objects and perform corresponding behaviour, its facial expression remains constant. Even though the holographic AI intelligence does not depend on the appearance or animations, this lacklustre performance triggers within the user a lack of confidence in the holographic AI’s abilities.

Intelligence

- Physical object recognition/interaction
- Spatial understanding
- Natural language processing
- Synchronization with the user's behaviours
- Computer vision
- AR plugin
- Learning systems
- Adaptivity

Figure 2.10. Intelligence

The intelligence of a holographic AI refers to its abilities. Figure 2.10 shows that the dimension of intelligence involves computer vision, natural language processing, spatial mapping and understanding, responsiveness to physical-object and environmental interaction, and multimodal adaptivity. Computer vision is fundamental to identifying the user environment, detecting location, and overlaying virtual information on the real-world scene. Five studies use AR mobile (or mobile augmented reality) (Oh and Byun, 2012; Miyake and Ito, 2012; Park and Jeong, 2019; Nasution et al., 2020; Kim et al., 2021a), and two rely on markers (Miyake and Ito, 2012; Oh and Byun, 2012). For example, Oh and Byun (2012) identified key points on images for tracking pages of a book using an open library (Lepetit and Fua, 2006). Additionally, Schmidt, Ariza and Steinicke (2020) used computer vision to detect eye contact based on the facial features of the user. Their system collected the user images, and analysed the position of eyes and mouth to ensure their position. Then, the authors built a pre-trained model for each aligned image, so that the system could receive and categorize video frames for face recognition.

Natural language processing is the primary interactive channel. In the cited studies, proposed methods simulate specific events, dialogue contents, or answers, but they cannot satisfy a daily pattern of communication, especially when the speech is not intonated. In addition, the mobile AR chatbot agents described in the studies rely on dialogue and text-input, and do not take advantage of voice assistants in AR; nor do any studies explain why project chatbot agents project into the physical space. Further, pre-structured dialogue flow and voice output should attach a sentiment engine to trigger corresponding contexts, to stimulate within the user the feeling of being together with the agent in one space. Additionally, AR intelligent tutor systems simulate both patients and teachers, possessing more complex virtual human architectures. The demand for pedagogical holographic AIs, especially medical assistants, cannot be satisfied with basic appearances, facial and body animations, and predefined dialogues alone. These features do not enable virtual patients to fully express a painful situation by way of tone, rhythm, and even modal particles. The holographic AI needs to accurately simulate the patient's emotions, collect and analyse data of trainer's performance, and provide real-time feedback from users.

Spatial mapping and understanding allows the holographic AI to recognize physical objects, avoid conflicts, and project them into an appropriate place. It is not enough, however, to accurately capture the dynamic environment. For example, the holographic AI in the study by Lang, Liang and Yu (2019) scans and reconstructs the interactive space, but only identifies static objects. Therefore, even if the holographic AI reacts to physical surroundings and controls virtual information, it should mimic and perform multiple behaviours in order to conceal the weaknesses of not being able to actually manipulate physical objects. A number of researchers have developed their own methods and systems for optimizing holographic AI physical awareness, such as an optical tracking system, silhouette mesh, and AR navigation toolkit. Although the holographic AIs in some studies are able to interact with real objects or perform multi-modal interactivity, they only recognise fixed, real objects. Therefore, physical-object recognition is different from physical interaction. Holographic AIs capable of physical recognition only can recognize real objects, whereas physical interaction allows holographic AIs to manipulate physical objects, such as turning on a light or changing some other property. In terms of spatial awareness, the holographic AI recognizes contexts and performs behaviours, but it cannot execute physical interactions.

In recognition of these limitations, two categories are proposed: unidirectional interaction, and bidirectional interaction. The unidirectional interaction allows the holographic AI to identify, track, response to physical surroundings, but the real environment does not affect the holographic AI. On the other hand, bidirectional interaction occurs when physical and virtual objects are able to interact and influence one another. For instance, six studies proposed apparatuses and multiple systems enabling bidirectional interaction, such as a magnetic actuator surface (Lee et al., 2021), a thermal table and a robotic ball (Schmidt, Nunez and Steinicke, 2019; Schmidt, Ariza and Steinicke, 2020), Wi-Fi control(Kim et al., 2018), control centre of home devices (Huang, Wild and Whitelock, 2021), and voice control (Miyake and Ito, 2012). However, a holographic AI capable of bidirectional interaction might not be able to manipulate every physical object in the interactive setting. Further, this type of interaction is hard to implement, especially in the dynamic event. Moreover, the mutual conversion of objects from static and dynamic one is also a significant challenge posed by directional interaction. For example, when the user holds a real book to read, this action status changes from static to dynamic. The holographic AI may need to identify this behaviour and physical book by its position or the user hand gestures.

Adaptivity is defined as a holographic AI's capacity for self-adjustment in accordance with the user's preference and requirement. Such a holographic AI should position itself in an appropriate place and perform different animations based on spatial understanding and physical-object awareness therein (Lang, Liang and Yu, 2019). However, the methods in the cited studies regarding user experience measurement do not focus on adaptivity, but tend to employ Woz to remotely control the holographic AI. Nevertheless, for a pedagogical holographic AI, adaptability is crucial to the implementation of personalized training, adaptive teaching materials, and real-time feedback, as well as allowing students to expand their thinking. Although Zielke et al. (2018) proposed an emotive holographic AI for medical education, they failed to explain how the system aligns with the student's requirement, and did not even provide sample lessons. A similar example is the holographic AI proposed by Hartholt et al. (2019), which provides immediate feedback after training by analysing eye gaze, blink rate, head movement, and voice; however, the authors did not assess the accuracy of its feedback.

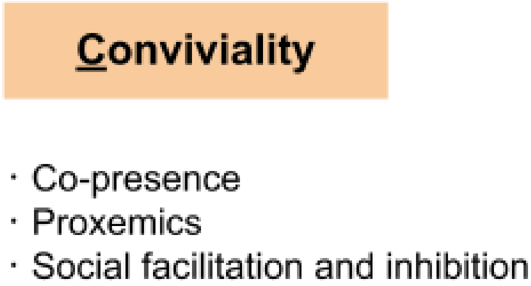


Figure 2.11. Conviviality

Conviviality is defined as the user's perception of social interaction and level of satisfaction with a holographic AI (see Figure 2.11). Over half of the cited studies conducted comparative and controlled experiments with different assignments being completed by participants. Photorealistic holographic AIs convey a stronger degree of user experience and even usability. It is obvious that users prefer to interact with a humanoid holographic AI with spatial understanding and physical-object awareness, as it conveys a stronger sense of togetherness. In proxemic experiments the holographic AI impacts on the user's locomotion and attention. Although the holographic AI is virtual, the user follows social norms to avoid collisions. Social distance is also hard to determine, since it depends on the size of the holographic AI. Besides, in normal social interaction, the user does not walk repeatedly towards a holographic AI, or a real human. Some studies do not indicate interactive time, and do not explain the reasoning behind the measurements chosen for the questionnaires.

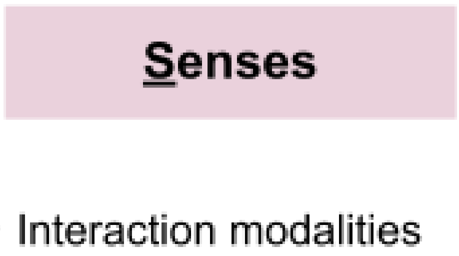


Figure 2.12. Senses

Senses in Figure 2.12 enable a holographic AI to perceive and react to the user and undergo context changes via different abilities. As mentioned before, such an agent relies on natural language processing, gaze tracking, gesture interaction, and physical interaction/awareness. Although interaction modalities are dependent on the holographic AI's performance, not all animations or dialogues need be interactive. For example, if a holographic AI only speaks to the user at the end of the interaction, it does not exert any such influence.

Generally, in the PICS model, persona is the external representation of the holographic AI, and intelligence is inner core that control its performance, interactivity, and responsiveness. Sense is the information acceptor, located at the periphery of intelligence to receive, select, and categorize interactive content. Conviviality establishes a relationship between the holographic AI and the user, such as social presence and proxemics.

2.5.2 Five categories of the holographic AI

Pivoted on the 49 cited studies, the above PICS model enables to contrast and group different types of holographic AI, relying on studies' aims and holographic AIs' main functions. Each type of holographic AI can match primary elements. For example, natural language processing is important for the chatbot, but not a key criterion for the game character. Therefore, a novel taxonomy of the holographic AI is provided below: it involves user avatars, simulation agents, intelligent tutor systems, game characters, and chatbots.

2.5.2.1 User avatars

The purpose of a user avatar is to develop consistent alignment with the user's actions, achieved through head-tracking or position-tracking techniques (Piumsomboon et al., 2018; Yoo and Tanaka, 2022; Wolf et al., 2020, 2022). However, these holographic AIs are constrained by their limited range of facial expressions, body language, and natural language processing capabilities. For example, in the Mini-me system, users in AR engage with VR avatars using air-tap gestures, but the VR avatar is restricted to pointing motions (Piumsomboon et al., 2018). In addition, the emotions of the user cannot be reflected through a virtual mirror (Wolf et al., 2020, 2022). Yoo and Tanaka (2022) created a walking avatar capable of emotional expression, such as victory and sadness, in competitive scenarios, but direct interaction with this avatar is not possible.

Therefore, the user avatar, including both life-sized and mini-sized representations, does not necessarily mirror the user's appearance. In terms of behaviour, this virtual agent can track the user's basic movements or execute predetermined animations under specified circumstances. Moreover, the Mini-me and walking system enable the user to engage in communication with other AR/VR users rather than merely interact with the user avatar itself.

Further, traditional avatars in VR or screen-displays have first-person and third-person viewpoints. First-person viewpoint shows only the hands and lower part of the body, whereas four papers detailing AR avatars present a third-person viewpoint. Therefore, gesture (or body) and environment synchronization are the main issues. In encountering remote collaboration or companions, the user may move physical objects or act out of pre-defined contexts, which can leave the user struggling to fully understand the local holographic AI's performance. Motion capture provides a one-to-one mapping of movement in order for the user to control avatars, but it has problems dealing with dynamic scenarios, especially outdoor activities, since it cannot label sensors everywhere. This type of holographic AI can utilize technologies of screen scanning and silhouette meshes. Spatial mapping identifies static objects and their distances, and silhouette meshes recognize human body shapes to avoid collision with other persons.

2.5.2.2 Simulation agents

The simulation agent mimics a real-life role in a specific scenario, and this is useful to assess user experience or develop training applications. According to these studies' main objectives and motivation, 27 papers used simulation agents as a tool to measure factors affect user experience, such as social presence, degree of realism, collaboration, physical-object/environment interaction, or social distance (Kim et al., 2021b; Reinhardt, Hillen and Wolf, 2020; Pimentel and Vinkers, 2021; Miller et al., 2019; Kim et al., 2018a; Kim et al., 2019; Kim et al., 2018; Kim, Bruder and Welch, 2017; Kim et al., 2016; Kim, 2018b; Li et al., 2018; Li et al., 2018; K. Kim et al., 2020; Norouzi et al., 2019; Lee et al., 2018; Kim et al., 2021a; Lang, Liang and Yu, 2019; Zhou et al., 2009; Chetty and White, 2019; Norouzi et al., 2022; Mostajeran, Reisewitz and Steinicke, 2022; Huang et al., 2022; Aramaki and Murakami, 2013; Randhavane et al., 2019; Peters et al., 2018; Obaid et al., 2012; Wang, Smith and Ruiz, 2019). While, 3 studies evaluated user feelings of the holographic AI by game approaches (Lee et al., 2021; Schmidt, Ariza and Steinicke, 2020; Schmidt, Nunez and Steinicke, 2019), which are simulation as well, since for users the holographic AI is game characters, but for researchers, it is a way for measurement, and all scenarios mimic real tasks.

This type of holographic AI simulates different appearances, behaviours, and degrees of intelligence in a specific scenario to compare different conditions. The other situation that required this agent is the simulation of a specific environment, where leaners achieve hands-on knowledge by using both virtual and physical instruments, and can use previous knowledge and practical experience to address problems (Andreu-Andrés and García-Casas, 2011; Vlachopoulos and Makri, 2017). For example, the holographic AI could simulate real patients' behaviour to describe symptoms, and then students need to diagnose them based on dialogue, the holographic AI's emotions, and provided additional information on virtual panels (Huang, Wild and Whitelock, 2021; de Barcelos Silva et al., 2020; Zielke et al., 2018). The simulation system has a navigation function to guide trainers in developing particular skills and experience. For example, the holographic AI can record and analyse the users' voice so that stress can be measured, and diction can be improved during a virtual interview based on human pose estimation (Voulodimos et al., 2018).

When employed for user experience measurement, the simulation agent can exhibit a certain level of substitutability. For instance, Kim et al. (2021b) utilize a life-sized and humanlike holographic AI to compare visual effects in conflict conditions. It is conceivable to substitute two holographic AIs and a voice assistant with alternative personas. In addition, this simulation agent can emulate real human characteristics and simulate real-life possibilities and tasks as part of an educational program (Ahmed and Sutton, 2017).

However, out of the 49 papers, none consider holographic AIs in connection with the Internet of things (IoT), neural networks, or fuzzy inference system. IoT can manipulate and transfer data of real objects and training performance to interaction devices. For instance, Ghorbani et al. (2022) have developed a serious game which publishes information detailing the movements of elderly patients for the attention of their caregivers, via IoT. A simulation agent utilising IoT should have resource management to enhance scalability (Ahmad et al., 2022). Thus, tasks can be selected by trainers based on their learning requirement. Learning processes and physical-virtual interaction should be recorded and data analysed to generate enhanced personalized services based on artificial neural networks. Further, fuzzy inference allows holographic AIs to generate emotional states and optimal behaviours (Liu, He and Song, 2008). Such a holographic AI in the simulation can play different characters to trigger plot development, and facilitate the learner's ability of decision-making and risk assessment. Therefore, the storyline setup and multimodal interaction approaches are critical requirements.

2.5.2.3 Intelligent tutor agents

Similarly, the holographic AI simulates a real teacher or coach to help learners gain knowledge. The intelligent tutor differs from the simulation agent. This intelligent tutoring system does not focus on narrative changes but on teaching content itself. The intelligent tutor holographic AI cannot offer practical experience. It can manage learning paths, provide one-on-one tutoring, and facilitate user-centred learning experiences (Zawacki-Richter et al., 2019; Churi et al., 2022). Alrakhawi, Jamiat and Abu-Naser (2002) proposes a model consisting of a student model, an expert model, a pedagogical model, and an interface. The assessment of students' academic achievement is conducted through a student model that incorporates learners' cognitive abilities, knowledge acquisition, and emotional disposition. A model of competence refers to a collection of instructional resources (Chang et al., 2020). The pedagogical model encompasses guidelines, problem-solving approaches, mechanisms that ensure requirement fulfilment among students, and assessment of student learning. Besides, the holographic AI is required to retrieve and analyse the data for implementing the users' emotion recognition and contextual responses, a process of machine learning (Joshi and Kanoongo, 2022). Additionally, the UI model involves verbal and non-verbal communication between students and the holographic AI. Huang, Wild and Whitelock (2021) introduce a virtual tutor for sports training that effectively meets four critical criteria. The system collects personal information (i.e., the student model) to generate corresponding exercises (i.e., teaching material), evaluates user emotions and performance outcomes (i.e., the pedagogical model), and utilizes the holographic AI as a UI.

The intelligent tutor system distinguishes itself from the simulation system in that, while simulation may facilitate user collaboration in teams, it does not directly offer instructional materials. However, learners should effectively apply acquired knowledge in their engagement with the simulation system.

2.5.2.4 Game characters

In the case of the game character, behaviours of players trigger the holographic AI’s reaction and storylines, and so it is necessary for the game character to have a pre-defined storyline which can activate victory or defeat, but not be compulsory for the simulation agent. For example, a holographic AI in GhosttPacer (Huang, Wild and Whitelock, 2021) is a company and navigator, rather than an opponent.

Game characters cannot support users and enable them to obtain a learning experience of a domain. The holographic AI in this game does not require a specific environment. Instead, spatial mapping scans the user’s surrounding and produces a game area, and head movement tracking and eye tracking establish the user’s position and orientation. However, users cannot perform certain movements in their interaction with the holographic AI, such as jumping. In view of the recent developments and increased availability of motion capture and smart gloves technologies, it may be predicted that wearable devices might improve AR playability.

In contrast, serious games serve distinct educational purposes (Ahmed and Sutton, 2017), with the primary objective of facilitating learning in a specific domain, such as the FaceMe application (Li et al., 2021). While Chahyana and Yesmaya (2020) created a pet simulator game that enables users to gain insights into pet behaviour, the primary focus of this game is on enjoyment rather than teaching. Serious games incorporate independent learning principles and a reward mechanism (Whittaker et al., 2021; Laamarti et al., 2014), setting them apart from the simulation or intelligent tutor system. For example, children use physical wooden markers to interact with the holographic AI (Li et al., 2021). Children employ a physical controller to select factors influencing flower growth (Oh and Byun, 2012). Therefore, the holographic AI acts as a navigational tool that indirectly guides the user through the provision of feedback.

2.5.2.5 Chatbot agents

As mentioned before, three studies developed the holographic AI with natural language processing (Nasution et al., 2020; Park and Jeong, 2019; Miyake and Ito, 2012). The chatbot agents used natural language processing primarily as an interactive approach, which is often used in translation systems (Nasution et al., 2020). The system relies on a database of words to match corresponding answers but does not perform any multi-modal interaction such as gestures or eye tracking. Although Huang, Wild and Whitelock (2021) introduced Azuma Hikari (2023) that can apply different body languages to control home devices and interact with users, the purpose of this holographic AI is to address and answer the user's questions through conversation rather than gestural response.

There is ongoing debate as to whether a virtual chatbot agent needs to be deployed in the real world if the holographic AI relies only on text input, and cannot perform reaction or responsiveness. Natural language processing also uses deep reinforcement learning and a Recurrent Neural Network (RNN) to provide responses based on past dialogues. For example, Chuang et al. (2023) have recently developed an educational chatbot in AR, which employs the ARCS model (attention, relevance, confidence, and satisfaction) to improve teaching materials and arouse users’ attention. The chatbot can track students’ learning processes, and provide an online cognitive service via IoT to generate relative news discovery (Sheth et al., 2019). An AR chatbot can also introduce and display virtual artworks (Guazzaroni, 2022). Therefore, the holographic AI may replace museum interpreters to recognize and introduce physical cultural relics. However, as mentioned before, such a chatbot cannot generate contradictory arguments or opinions to persuade users, as it lacks self-cognition.

2.5.2.6 A summary of taxonomy of the holographic AI

Holographic AIs	Natural language processing (including translation)	Synchronization with the user's behaviour	Real-life learningal role	Independent learning rules	Teaching content (one-to-one)	Victory /defeat	Storylines
User avatars		x					
Simulation	x		x				x
Game characters (including serious games)	x			x		x	x
Intelligence tutor systems	x				x		
Chatbots	x						

Figure 2.13. Comparison for five types of holographic AIs

This section analyses the differences and similarities among five categories of holographic AIs. (see Figure 2.9).

As mentioned before, other holographic AIs, aside from user avatars, can engage in conversations, although some depend on Woz to select appropriate dialogue. The user avatar is designed to represent the user's actions (e.g., running, walking, or gesturing). However, it is unable to engage in direct communication with the user but capable of conversations with other virtual partners.

Game characters are designed for entertainment, as opposed to holographic AIs in the serious game. Learning systems may incorporate intelligent tutor systems, simulation agents, or serious games. For example, simulation systems and serious games feature pre-defined narratives wherein the user's selection influences scenarios or the performance of the virtual agent. Engagement in these activities allows users to gain practical experience and apply acquired knowledge. Nevertheless, the simulation agent adheres to real-life social rules, whereas the serious game establishes independent interaction rules and reward mechanisms not bound by fundamental social interactions. For example, Huang, Wild and Whitelock (2021) introduced the HoloPatient application, wherein the conduct of a simulated patient is contingent upon the evaluation conducted by a student. Besides, the simulation agent can be utilized to gather and analyse user experience. In the serious game, learners interact with the holographic AI through a designated channel. For example, the holographic AI generates emotional animations to demonstrate whether the learner chooses appropriate elements to cultivate virtual flora (Oh and Byun, 2012). At the same time, the intelligent tutor system emphasizes knowledge and instructional content. In the simulation and serious game, the holographic AI facilitates collaboration among fellow students, whereas the intelligent tutor system offers one-to-one educational assistance.

The simulation agent, intelligent tutor system, and gaming character (including the serious game) are equipped with natural language processing capabilities. These holographic AIs may engage in behaviours or exhibit cues directly inside the context of games or simulations. However, chatbots primarily operate in a question-answering mode, and the virtual agent integrates into chatbot systems to offer users translation capabilities and personalized recommendations in response to their specific needs (Huang, Wild and Whitelock, 2021).

With this PICS model, the avatar can synchronize with the user's behaviour. The chatbot agent not only provides recommendations but also controls home devices through frequent question-answering dialogues. In the simulation system, the holographic AI assesses user experience and contributes to the learning system. Hence, it involves dimensions of conviviality, senses, and intelligence elements, such as physical-object recognition/interaction, spatial understanding, natural language processing, or adaptivity. The game character shares similar features, yet it prioritizes the delivery of entertainment through diverse storylines and personas. In contrast, the intelligent tutor system prioritizes personalized instruction that enables students to acquire knowledge through direct interaction with the holographic AI instead of relying on experiential learning.

On the other hand, future development in the holographic AI technology may entail the emergence of new types and categories. For example, simulation agents can serve a practical purpose, and chatbot services can generate follow-up course plans and solve students’ questions to update learning goals. The next few years might witness the rollout of chatbot holographic AIs for smartphones. Such an intelligent tutor system featuring a chatbot could teach and help students review theoretical knowledge, and simulations could provide practice sessions. Therefore, in foreseeable future, holographic AIs may automatically transmit data in diverse application domains and feature in derivative products. Future holographic AI research and development is likely to focus less on single characters and their functions, and more on intelligence integration to create a more human-like virtual character. Therefore, natural language processing, adaptivity, and interaction modalities can be used for this agent.

2.5.3 Validity and shortcomings of the PICS model

The objective of this study is to understand the essence of holographic AI and its distinct features in order to establish a model for such systems. The ensuing chapters will validate this model by demonstrating how holographic AI systems can be developed based on its structure. Given that holographic AI will be implemented in intelligent tutoring systems, specific elements of the model will be selected and adapted to meet particular requirements, such as learning goals and interaction approaches. For example, Chapter 3 details various methods of creation, encompassing aspects like persona, intelligence, and sensory capabilities. The holographic AI will create personalized workout animations, thus addressing the need for personalization and user-centric design. Section 2.4 shows that prior studies have utilized quantitative research methodologies to gather and analyse user experience data, a practice that is appropriately applied in Chapter 5 to investigate user experience.

The PICS model, constructed from 49 studies, has limitations in its scope, potentially omitting other relevant research or types of holographic AI, which limits its generalizability. For instance, it may not accommodate future enhancements that could be integrated into diverse holographic AI systems. Additionally, this study focuses on the effect of singular dimensions like intelligence or persona on user friendliness, overlooking the combined impact of multiple dimensions on user experience and the interplay between these dimensions or components. While Section 2.4 outlines approaches for measuring user friendliness, it does not investigate the nuances across different user demographics. There is also some overlap among these components, which might not be universally applicable to all types of holographic AI. For example, computer vision is essential for physical-object recognition, spatial mapping, and marker-based augmented reality. Besides, the aim of educational systems is to develop an application, not a capability that is transferable across multiple platforms. The conviviality dimension is currently composed of just three elements, each assessing a single influence.

This study does not delve into attributes, such as gender, career, hair, and clothing styles of holographic AI, to avoid overcomplicating the creation process. Norouzi et al. (2022) investigate the effect of a virtual canine and a female holographic AI on user experience but do not justify the choice of a cartoon-like appearance for the dog, the decision to use a female representation, or whether stereotypes exist that suggest female holographic AI might offer emotional support. Additionally, prior research has not primarily focused on rendering and compression techniques. Elements such as interaction processes, content, and devices have been overlooked, which emphasizes system interaction and programming rather than the holographic AI as a whole.

Future research will therefore investigate the potential influence of combined dimensions on user friendliness and explore additional dimensions and features.

This will help to develop a more holistic and nuanced understanding of holographic AI systems.

2.6 Summary

This chapter presents the review of 49 studies regarding AR agents published between 2009 and 2022. The literature has excluded ‘pure’ voice assistants, navigation systems, AR shopping agents, and motion tracking systems. Besides, this study does not analyse the holographic AI’s clothing, hair, careers, and gender. It also excludes robots, compression, rendering, interaction content, and interaction equipment, since these elements are beyond the scope of the holographic AI research.

This chapter has reviewed examples of holographic AIs in different areas of interest (i.e. conviviality and intelligence development). It has focused on concepts and features of holographic AIs, and has proposed the novel PICS (i.e. persona, intelligence, conviviality, and sense) comprehensive model for the design and study of holographic AIs.

The concepts of the holographic AI are not uniform in these studies. For example, the holographic AI detailed in the studies do not exist in the same real interactive space with the user, but instead employ MR merging both virtual and real surroundings in order to create the illusion of physicality, and provide the user a sense of togetherness. In short, a holographic AI is an embodied virtual agent or user interface in the MR that is capable of interacting with both physical and virtual environments.

The chapter then conducted a thorough analysis of the selected studies to identify and categorise specific characteristics of holographic AIs. For instance, the interactive approach postulates dividing holographic AIs between the two groups of unidirectional interaction and bidirectional interaction, based on physical-object/spatial understanding. It has been found that the user model and adaptivity are easily ignorable but critical factors affecting the design of personalized services for managing interactive multimodal features.

The integrated features of the holographic AI described in the 49 studies make up the PICS model, based on this, a taxonomy of holographic AIs consisting of user avatar, chatbots, game characters, simulation agents, and intelligent tutors has been proposed. The model can be used to guide the design of different types of the holographic AI; further details on holographic AI creation will be represented in the next chapter. In future trends, it is likely that the holographic AIs may provide an integrated service, that their functionality will not be limited in a single classification, but instead take the form of a mixed AI agent.

Chapter 3 Creation of the Holographic AI

3.1 Introduction

Following the analysis of the ‘PICS’ model and holographic AI features, this chapter describes the holistic processes for creating anthropometric holographic AIs, presenting viable approaches for each of the PICS dimensions of persona, intelligence, conviviality, and sensory modalities. Based on the model, this chapter considers contributions in the field in terms of the development of anthropometric holographic AIs, including a number of relevant case studies.

It is critical to show that this theoretical model (postulated from literature) can be applied in practice and is ready for use, and that the investigation of user experience can be explained under this foundation.

The first section of this chapter illustrates a novel process of creating ‘persona’ of the holographic AI, based on an effective and operable way of using 3D scanning technology and motion capture for generating animations.

Appearances are created via 3D scanning, automatically generating point clouds and meshes with textures. I show how to clean data of the scanned 3D model, and compare the difference of the approach to a manual way of creating an avatar. This part about appearances is documented in Section 3.1.1.

These 3D generated holographic AIs by 3D scanning should be equipped with behaviours, intelligent communicative abilities, and spatial awareness in order to implement interaction. Therefore, motion capture will be explained further in Section 3.1.2, using examples of rehabilitation exercises, lip sync, and facial expressions.

The created holographic AI is then embedded into different applications to investigate how to establish elements of intelligence (Section 3.2), conviviality (Section 3.3), and sensory modalities (Section 3.4).

The first was part of an exercise to create holoCARE (Wild, Loesch and Huang, 2019). holoCARE was a collaborative project aimed at building a personal virtual fitness coach for cancer surgery survivors, capable of instructing patients of the right rehabilitative workout exercises. For this, I used motion capture to create a rich inventory of workout exercises, while a bespoke app composed of a speech and gesture interaction decision tree with speech and gesture interaction for selecting the right exercises was built with the assistance of consultants.

This informed the second case study, where my animations were later integrated by the development team of MirageXR (Wekit ECS, 2022) into the core of the AR learning experience authoring tool and player. Together with my supervisor, Prof Wild, I co-authored DAIMON (see Section 3.2, intelligence), which formed the basis for this integration into MirageXR. This included an extension for procedural animations triggered by dialogue, as well as a plan for further extension with parametric procedural animations.

This extension towards parametric procedural animations motivates the investigation of instructional gestures in Chapter 4.

In terms of intelligence, Section 3.2 discusses persona, natural language processing, physical-object awareness and interaction. This part includes documentation of DAIMON and its technical based on IBM Watson (IBM, 2023), which facilitates integration with the 3D model. Originally, DAIMON was built as an extension of holoCARE (Wild, Loesch and Huang, 2019) to extend the more rigid keyword-based speech recognition and gesture interaction by implementing dialogue understanding services.

Therefore, this chapter is to demonstrate how to create an anthropomorphic holographic AI in practice, following this model (i.e. RQ2).

Section 3.2 represents ways of creating holographic AIs. Section 3.3 investigates the aspect of intelligence, especially social awareness and some special deliberations about the use of a user model in this context. Section 3.4 discusses the dimension of conviviality. Senses in the Section 3.5 discusses explains how to implement interaction modalities in physical and virtual environments. Finally, a summary and recommendations for developing holographic AIs follows in the Section 3.6.

3.2 Persona

Persona is what the holographic AI looks like, and it is the first impression regarding the user’s visual perception. The persona dimension refers to the two exterior traits of a holographic AI 3D model are its behaviour and appearance. Social attribute can determine the extent to which the character is motivating by its appearance (Baylor, 2009). Such agents can be full-sized or mini-sized, and are imbued with humanlike appearance and characteristics, as well as different levels of personification. In this context, the appearance consists of realism and size mapping.

There are two components of behaviour: expressiveness, and animation. The latter refers not to mere animations, but to subtle expressions which are carefully connected to both explicit and potential information contained in storylines and conversations. With game characters there is strong focus on body movements. According to Carroll (1996), emotions expressed in the game environment must possess a context-based functional capability. Chatbot agents are expected to exhibit realistic facial expressions (smiling, frowning, etc.) during the course of AI-user communication. Users’ choices, and dialogue content and events should dictate such animations. The appearance and engagement of intelligent tutors are built on behavioural structure scaffolding, which in turn consists of key visual cues (including nodding, gaze, gestures and other facial expressions) employed in affective feedback.

3.2.1 Parametric Appearance

Modern non-contact sensor technology-based 3D scanners are capable of in situ capturing of human body shapes and surfaces, and can provide efficient diversification and individuation of virtual characters. One example is the Occipital Structure sensor, which is used to capture textures and meshes, and which can gather feedback quickly on the iPad platform. In this project, the static scan process of the Occipital Structure sensor is combined with motion capture to develop a method of recording dynamic joint behaviours via tracking of body-attached markers.

In this context, the quality of 3D avatars is affected by the following factors: real actor poses (A-pose or T-pose), scan time, and both scanning equipment and background. Lighting is another critical factor: ideally, the lighting should appear natural and balanced, and neither too strong nor too dark, in order to reduce blurring, distortion and overexposure.

Nevertheless, there will always be a slight dissimilarity between the appearance of a real human actor and that of the reconstructed model.

This chapter details my efforts to strike a reasonable balance between AI hologram quality and the attendant technical effort, using the Smart Glasses (Microsoft HoloLens) equipment.

Workflow-wise, the scanned 3D model is exported into Unity (Unity, 2023), a game engine, to implement interactive behaviour.

One example of our Unity-generated 3D characters is provided in Figure 3.1 below. To ensure compatibility with the HoloLens, the number of polygons for each of the body models' avatars has been optimized to meet the HoloLens requirements. Otherwise, the failed meshes impact negatively on animations, and a model with high polygon counts can affect GPU and CPU speeds.

Figure 3.1. A reconstructed avatar (own graphic)

As detailed in Figure 3.2, the avatar reconstruction process includes 3D scanning, reconstruction and animation. This process of reconstructing an animated holographic AI is based on the persona of the PICS model.

The scanning environment comprises a capacious room, the ceiling of which contains five long-shaped light sources, ideally.

At this point it should be noted that certain parts of the body cast shadow under this intense lighting, particularly the nose and neck. To compensate for this problem, an LED light source mounted on the iPad was used to attain a more balanced level of brightness across the actor's face. The actor’s head and body were scanned separately, during which scanning time, the actor’s height and head characteristics were observed.

Figure 3. 2. The process (own graphic)

3.2.1.1 3D scanning

The recent advances in the field have facilitated the application of 3D scanning across a wide range of fields, namely entertainment, fashion design, and archaeology (namely digital capturing of cultural relics).

The different methods for constructing a virtual human have been reviewed by Peng et al. (2016). One approach is to scan the human body using a reconstruction facility, by taking multi-view colour photographs (front, side, back), and amending the shapes of 3D virtual humans to fit target models. Peng et al. also cited a full-body cloning method, the algorithm of which can define and construct a humanoid agent using graphics produced in low lighting conditions. A valuable tool is silhouette extraction, which can exploit data from the edges of the graph to convey textures, colours and greyscales. There are three main types of contour extraction methods identified by Peng et al. (2016): approaches using prior knowledge, gradient-based methods, level-set methods, and active contour-based methods. Bartol et.al. (2021) proposed an overview on the three different categories of 3D scanning technologies for 3D human shape construction: passive stereo, structured light, and time-of-flight imaging.

When the 3D human is created using passive stereo, images of the real body from multiple perspectives (or from two subsequent frames of a moving camera image) are used. Structured light is an alternative to that, able to avoid failed texture generation of the passive stereo approach, using observations of light pattern deformations to reconstruct 3D models. The third option, time-of-flight, measures return time of light projected onto the real human body to generate the virtual body.

Bartol et al. (ibid) also propose preparation, scanning, feature extraction, model fitting, and measurement extraction as five stages of 3D scanning.

The actor must wear labelled markers in order for their human shape to be identified, and must retain a specific pose until the scanning is finished so that a 3D point cloud is generated. The features of the actor’s body and silhouettes are extracted to calculate a 3D T-pose graphical model. The aim of the model fitting step is to generate a statistical body mesh. Thereafter, the body feature, T-pose model, images, and 3D scanning data are fed into the body measurement.

However, it should be noted that, owing to lack of interoperability among different methods and standard editing tool chains, few state-of-the-art methods are currently applied in industry. A professional environment likely demands hundreds of depth cameras, as well as green screens and balanced lighting, altogether a substantial assembly of resources which requires high financial investment.

In Figure 3.3, ItSeez3D is depicted, an iOS-compatible 3D scanning tool which runs on the Cloud, within which it processes data sourced from the Structure.io structured light scanner. There are three types of scanning that ItSeez3D can perform: body scans, bust scans, and object scans. For the T-pose to be captured, the person being scanned needs to maintain the T-pose throughout the scan, which usually takes between 1 and 2 minutes.

Figure 3.3. ItSeez3D and Structure sensor scanning body model of a female (photo: Mikhail Fominykh)

It can be seen in Figure 3.4 that certain artefacts in the 3D scanned model are similar to those of the human model (e.g. hair), while others are falsely coloured (e.g. clothing, furniture, wall in background). The whole human actor is scanned from head to toe, the human actor maintaining the same position throughout, and the two scans of the head are compared.

Nevertheless, collected data may be incomplete. As shown in Figure 3.5, data for the top of the actor's head are missing, the meshes of the head and ears clash, resolution is low, and the facial textures are distorted. The sensor also fails to recognise and capture in full the actor's extended fingers.

As illustrated in Figure 3.6, some data for the hands are missing, and redundant meshes of the fingers can collide. Following scanning, the high-quality 3D scan models were selected and used to recreate.

Using ItSeez3D, it is necessary to produce two body models of good quality, and assess the scan quality using the cloud render preview function, and not just the local preview function in ItSeez3D.

At least 2 high-quality head and hair scans should be collected. Standard models for hands and feet can be used from Wrap 3 gallery (Russian3DScanner, 2023), among other online open-source libraries. It should be noted that 3D scanning often fails to generate models which meet the quality bar for models on these websites, possibly due to the sensors used or the human actor failing to remain motionless throughout the scan.

Figure 3.4. Different scans of the same actor (Huang, Twycross and Wild, 2019)

Figure 3.5. Artefacts and mesh distortion for the ears,
when actor scanned in full body model (Huang, Twycross and Wild, 2019)

Figure 3.6. Scanning errors and artefacts of actor’s hands (Huang, Twycross and Wild, 2019)

3.2.1.2 Reconstruction

In order to ensure the 3D scan model can be animated, the failed meshes and textures should be cleaned. This process is called reconstruction.

3D scan errors like false colours and distorted limbs (e.g. hands) can be corrected with Photoshop, and 3D scan models and textures can be sanitized with Wrap 3.4 (2021).

A wide range of low-polygon models with corresponding UV maps (2D representation of the 3D model) are available on Wrap 3 (Russian3DScanner, 2023). Each of these models consists of a human body, with limbs and head of unique proportions. These templates can be matched with 3D scan models in terms of the sizes of body features (nose, eyes, feet, shoulders, etc.), and this matching process is used to reproduce avatar model with high polygon counts, as well as produce high-resolution textures using nodes. The stages required for head reconstruction using Wrap 3 are illustrated in Figure 3.7. As shown in Figure 3.7(1), SelectPoints is used to align the markets on the template model with those of the scanned model. Transformation of the scanned avatar model to the low-polygon model is conducted using the ‘wrapping node’ (Figure 3.7(2)), and automatic generation of the texture is achieved using the ‘TransferTexture’ node.

As shown in the figures below, it is successfully reconstructed the head and full body meshes separately. However, there is some distortion in the reconstructed hands, some incorrect colours in the texture, and distortion in the shape of the actor’s feet (Figure 3.8). These errors are addressed in the subsequent reconstruction stage.

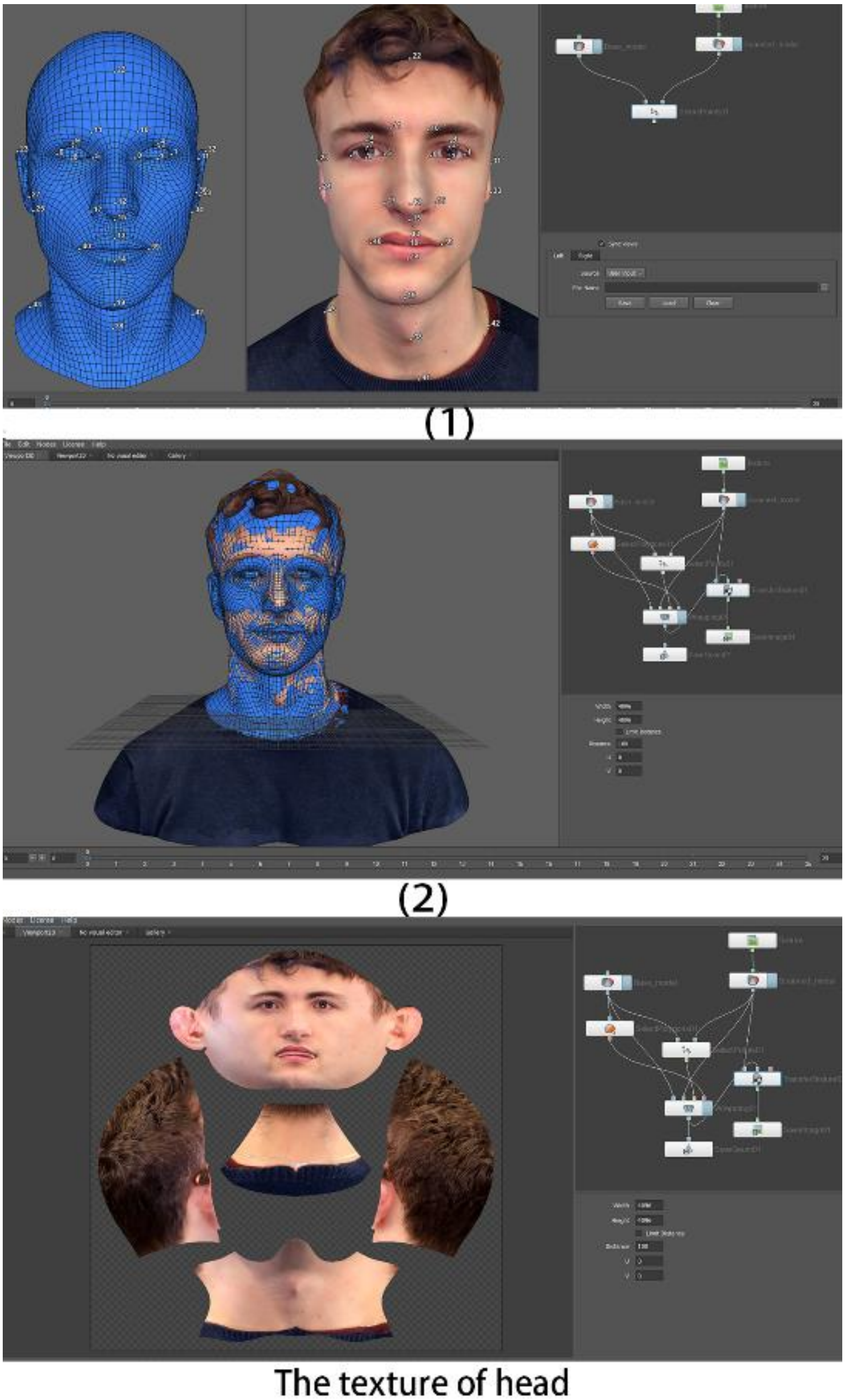


Figure 3.7. Reconstructing the head model in Wrap 3 (Huang, Twycross and Wild, 2019)

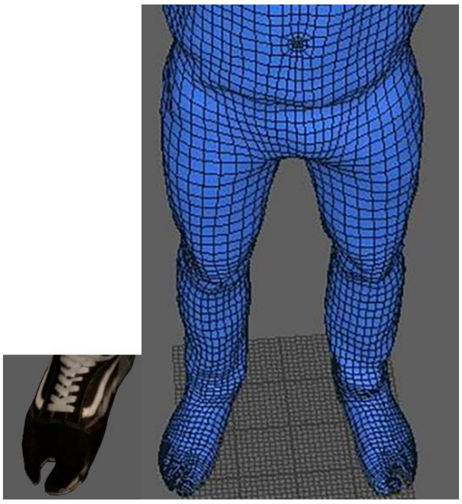


Figure 3.8. Since the full body model template does not wear shoes, meshes of toes cannot be transformed properly on actors wearing shoes (Huang, Twycross and Wild, 2019)

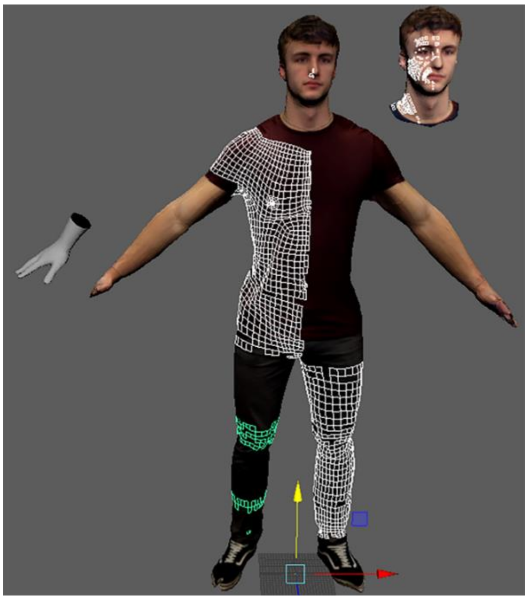


Figure 3.9. Reconstructed body and head models, as well as newly created hand models (Huang, Twycross and Wild, 2019)

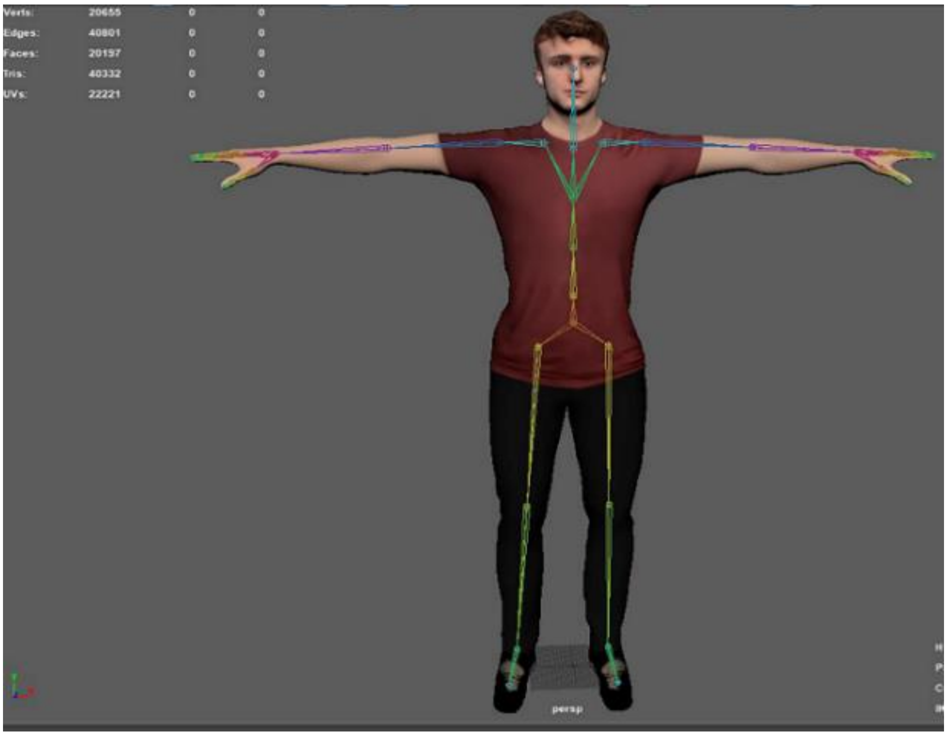


Figure 3.10. Body rigger for T-pose (Huang, Twycross and Wild, 2019)

As shown in Figure 3.9, the transformed template of the human body is composed of small quadrangles which are not aligned with one another. All points need to be merged together before being exporting from Wrap 3 into Maya. In the reconstruction, it is necessary to substitute 'raw' scan data for out-of-scale meshes.

For example, vestigial portions of the real body which normally are covered (head, hands, and shoes) may have to be replaced by the individual head and the newly created models of these body parts and enveloping clothing (e.g. shoes). In order to minimise the number of polygons, some unnecessary edges, such as navel, should be cleaned.

All meshes should be smooth, and failed colours are corrected by stamp copying over correct colours by Photoshop. Human body textures can be imported where relevant data is missing or requires embellishment. As shown in Figure 3.10, rigging is used to convert from A-pose to T-pose by rigging. However, although an integrated texture can be produced using Wrap3, the resulting size is smaller than that in the UV map, necessitating a time-consuming process of matching adjacent parts of the mesh not occupying neighbouring positions in the UV layout.

3.2.1.3 Comparing with manual design

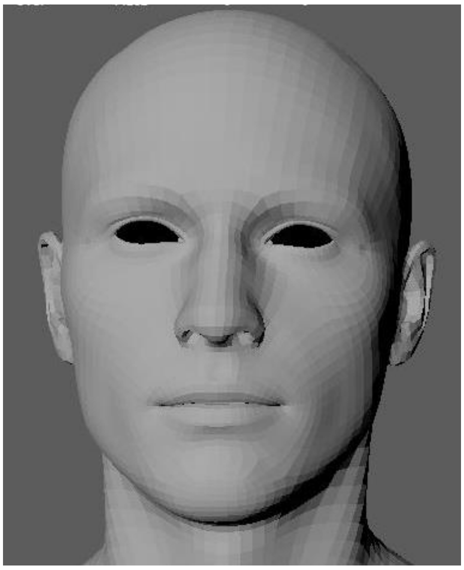


Figure 3.11. The base mesh model (own graphic)

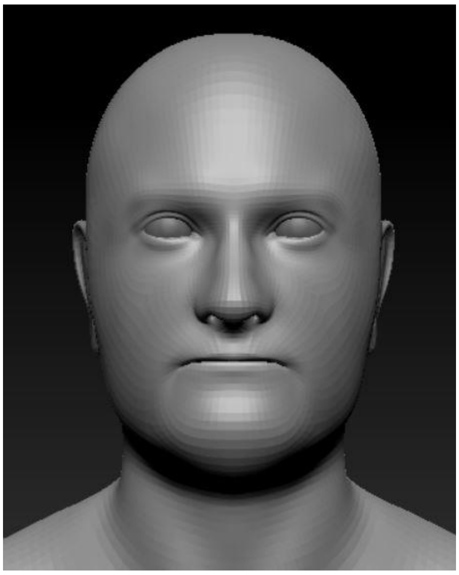


Figure 3.12. Reshaping the model (own graphic)

There are three steps of manually designing a realistic holographic AI shape: building a basic low polygon model; sculpting the low-polygon mode to a high-polygon model; transforming the high-polygon model to a low-polygon model with details.

This section determines which of these methods produces the best-quality holographic AI. In order to avoid repeating the description of steps for the reconstruction of the 3D scan model, this section will amplify how a realistic head for the holographic AI may be created, since the face is critical to delivering a person's appearance features during interaction.

In order to capture the structure of facial muscle, a male head model with 5,922 faces has been downloaded from the Wrap3.4 gallery. It provides a basic mesh of head and a clear facial skeleton (see Figure 3.11). This model then is imported to Zbrush which is used to sculpt and repaint a high-poly model. The shape of this mesh is different from the actor, thus for the basic model the 'Move' tool of Zbrush is used to re-shape the face, according to the real actor's pictures with different angles. This requires following human skeleton and muscle models, such as the directions of the orbicularis oculi muscle or brow ridge. It is unnecessary to subdivide the model until the shape of the base mesh is similar to the actor's appearance (see Figure 3.12). Next, the model is turned into an unfolded UV in Maya and Unfold3D. In order to provide more details in textures and consider the HoloLens CPU, the UV map cannot be over one tile, but if the screen output can over 2048*2048 pixels then it will show more details on 3-4 UV tiles (see Figure 3.13). Therefore, in this study, one UV tile of 2048 * 2048 size is used, and this parameter matches the requirement of the game character design (see Figure 3.14).

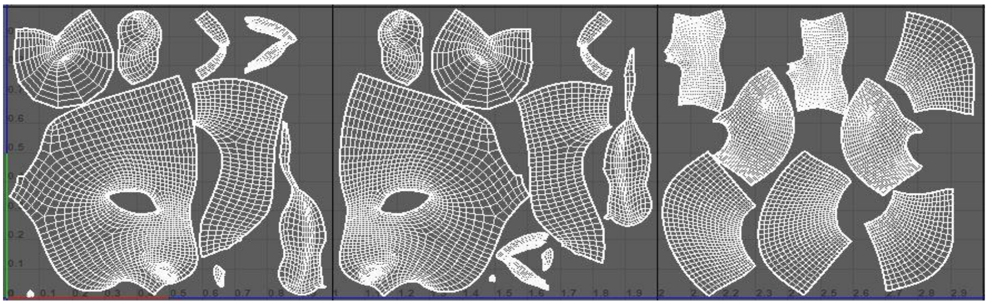


Figure 3.13. UV maps (own graphic)

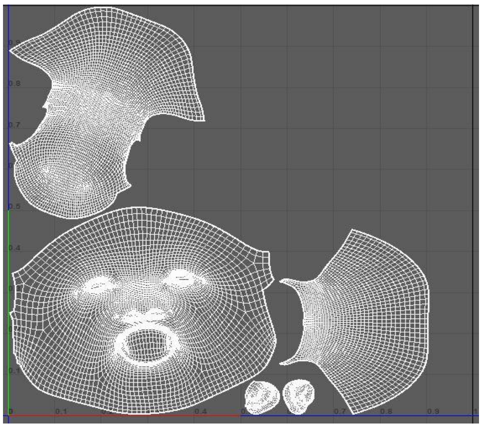


Figure 3.14. One UV map (own graphic)

The actor's model then needs to be enriched with more details, such pore, wrinkle, moustache, and hair. First, in order to create realistic and well-proportioned pores in the face, the model should be subdivided four or five times to generate a smooth and high-poly model for creating its skin. The skin pore is a primary detail for realistic texturing. The skin pore density is higher around the two sides of the nose, and the mouth. The size and coverage of pores depend on age, gender, sunlight situation, and diet (Flament et al., 2015). Males have bigger and deeper pore size and diversity. Therefore, three types of pore brushes were created for different parts of the face (see Figure 3.15).

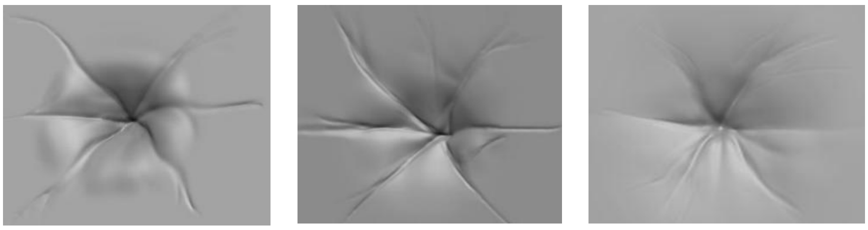


Figure 3.15. Three types of pore brush (own graphic)

These skin details need to be checked via high-resolution pictures to understand the growth directions of the muscle and wrinkle. Finally, a high-poly model is produced (see Figure 3.16).

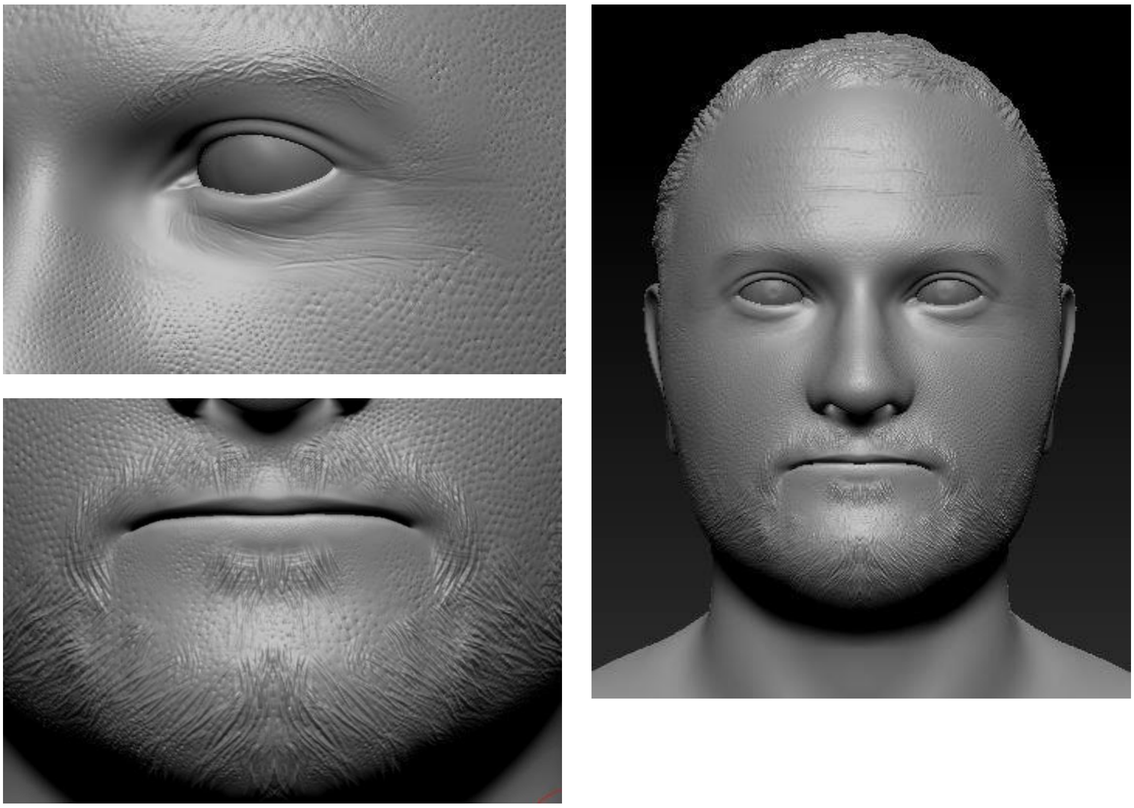


Figure 3.16. The high-poly model (own graphic)

In order to avoid excess polygons, the hair and moustache were not created. The details in the high poly model serve to bake a normal map to the low-polygon model again. Again, the high-polygon model needs to produce a normal map for the low-poly model, as well as create textures for the low-ploy model in Substance Painter (see Figure 3.17 and 3.18).

Figure 3.17. Normal map (own graphic)

Figure 3.18. Texture (own graphic)

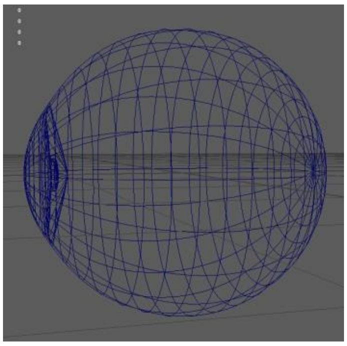


Figure 3.19(1). The eye model (own graphic)

Figure 3.19 (2). Iris (own graphic)

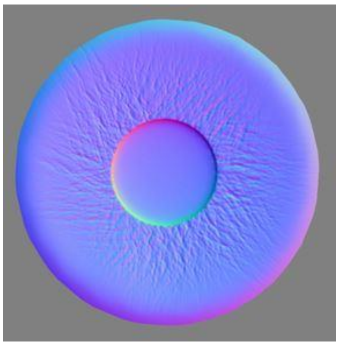


Figure 3.19 (3). Normal map of the eye (own graphic)

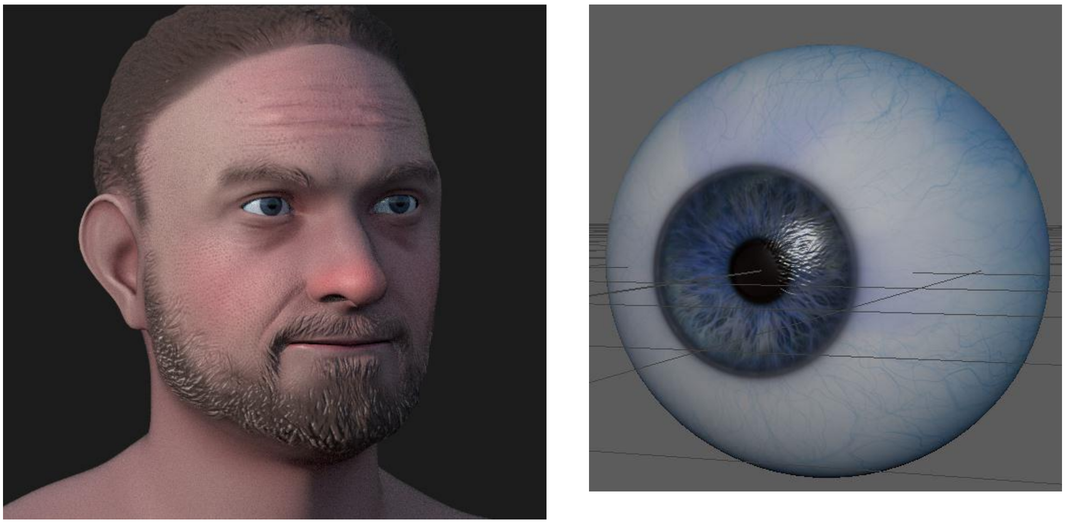


Figure 3.20. The final low poly models of head and eyes (own graphic)

Regarding the aspect of eyes – an eye is not just a ball, but is composed of iris, pupil, and sclera (see Figure 3.19). Therefore, after these elements have been constructed, the whole irises are exported to Zbrush to generate the texture and bake its normal to low-ploy model.

Then the low poly model of the head and eyes is rendered in Maya (see Figure 3.20). In total, the head model has 15260 polygons.

Based on this method, two full body models of holographic AIs with adjustable size and cartoon appearance have been created (see Figure 3.21).

Figure 3.21 (1). Sarah

Figure 3.21 (2). Hanna

In order to create clothes for both holographic AIs, the fashion design software tool, Marvelous Designer was selected. For example, the character model, Sarah, was imported into this software, so that a dress's size and shape can be created based on her height and weight. After the six parts of the dress are connect to one other, a high-ploy model of the dress is generated (see Figure 3.22). Then, the dress needs to be transferred to the low polygon model with unfold UV map again.

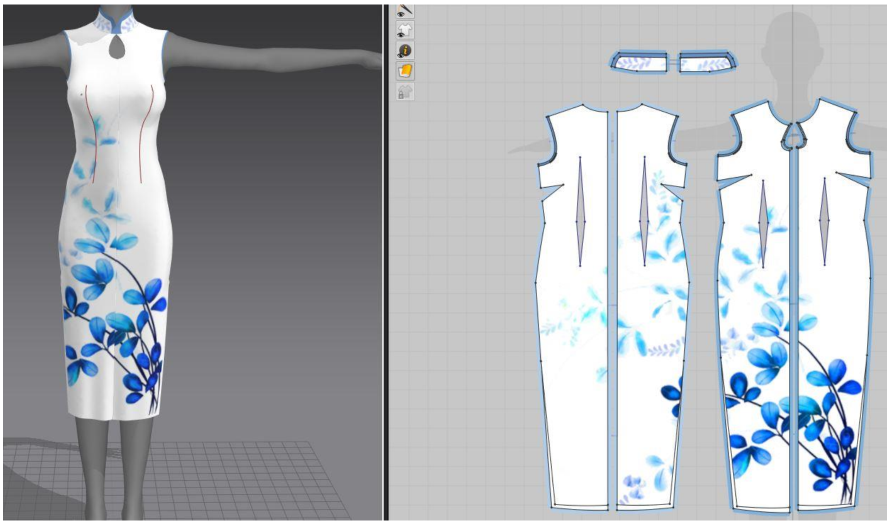


Figure 3.22 The dress design

There now exits 3D scanning technology aimed at users whose experience of art and painting is limited, which is capable of directly generating a high-polygon count model. However, 3D reconstruction of the scanning model is more complex. The 3D scanning firstly generates 3D cloud points followed by disordered triangular faces, which cannot be used directly in animations. This requires re-arranging of failed faces, as well as transfer to square meshes for animation.

Unlike the realistic 3D scan avatar, the manual way of designing the holographic AI's appearance can be stylized, compared to the realistic 3D scan avatar.

High/medium polygon counts and high-resolution affect the rending and processing speeds. In social media, virtual influencers with high-polygon counts are used the same way as for the film and television industry in order to create a robust sense of realism.

3.2.2 Rigging

The animation skeleton, along with body shapes hosting facial expressions and lip-sync, are generated in order to produce the final holographic AI avatar. The interaction modalities and social interactions of the holographic AI avatar are contingent on these foundational elements, which in turn are composed using Autodesk Maya.

3.2.2.1 Blend shapes for facial animations

Figure 3.23. Reference of phonemes (Huang, Twycross and Wild, 2019)

Basic facial expressions include anger, disgust, fear, happiness, sadness, and surprise (Sato et al., 2019). From these, main phonemes such as AI, E, O, U, F, L, M, TH and WQ are constructed (Figure 3.23).

Maya is used to deform the body geometry to create specific shapes. The mesh shape is altered using the grab tool, and the lip is separated using the smoothing tool. The technology captures the ever-changing shapes of the masseter and chin caused by muscular movements around the mouth during speaking (Figure 3.24). In order to blend the smiling face of the base model with the target model, the original character model substitutes as a base object.

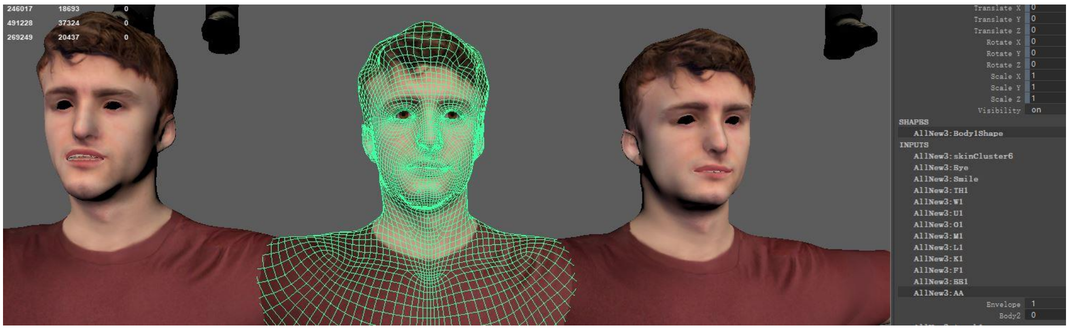


Figure 3.24. Blend shapes. The left and right models are base objects, and the middle one is the target (own graphic).

Figure 3.25. LipSync (Huang, Twycross and Wild, 2019)

Using LipSync pro 1.5 (Unity, 2018), which can reset previously created facial expressions and support a blend shape system, facial animations can be edited, and corresponding lip shapes can be imported (see Figure 3.25). The program offers four emotions, nine phonemes, and an eye manager responsible for eyeball movement and blink rate. It can be used to simulate emotion-specific facial expressions such as smiling, eyebrow-raising, and frowning. Certain eyeball movements such as looking up and down are controlled by the blend shape, and need to be aligned with the appropriate phonemes.

However, in the setup the eye trigger is set to ‘random looking’, thus bypassing the more complicated animation settings aforementioned. Using LipSync, phoneme markers are added manually; alternatively, corresponding phenomes are generated via integrated SoX sound exchange. Further, three scripts are used: lip script; the ‘record audio’ script; and the ‘save audio’ script, whereby the human actor records their voice in Unity, and the supported audio format in Unity is produced. The voice recording, stop capture and sound check steps are carried out using a user interface. Afterwards, the lip script is connected to the plugin, the audio script is saved, and the audio source is found. These steps precede automatic generation of lip animation by the character model. The sound recording process is illustrated in Figure 3.26.

```
    Debug.Log("starting");

    Microphone.GetDeviceCaps(null, out minFreq, out maxFreq);
    if (minFreq == 0 && maxFreq == 0)
    {
        maxFreq = 44100;
    }

    if (!Microphone.IsRecording(null)) {
        goAudioSource.clip = Microphone.Start(null, false, 120, maxFreq);
    }
}
```

Figure 3.26. The script for capturing audio

Similarly, SALSA LipSync is another means of implementing facial animations in Unity. It does not demand manual addition of nodes of corresponding phonemes. Instead, it automatically generates lip animations based on audio, and this plugin also guides eye movement (see Figure 3.28).

Figure 3.27 shows users can adjust parameters, such as animation timing, customized open-mouth size, and level of smoothness, until the interaction becomes natural.

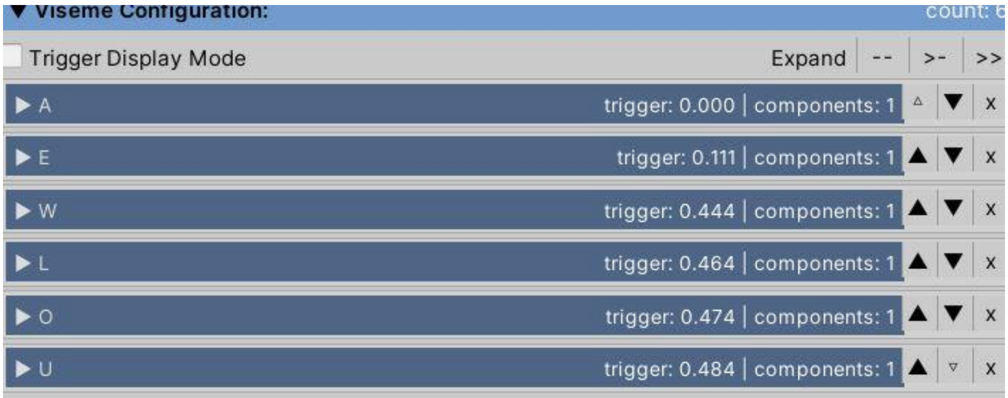


Figure 3.27. SALSA Lipsync sets up visemes

MirageXR uses ThreeLS, which in turns provides three blend shapes: kiss, lips closed, and mouth open. ThreeLS is concise and is capable of configuring parameters easily as the visual parameters link to frequency-band specific energies, enabling the triggering of weights of blend shapes (Llorach et al., 2016). Therefore, although it only offers three blend shapes, these can be resized. Similarly, vocal tract length, smoothness, and degrees of sensitivity enables to altered in order to control length, smoothen animations, and eliminate environment noise (see Figure 3.29). However, the values of blend shapes cannot account for compressed lip, such as the M viseme.

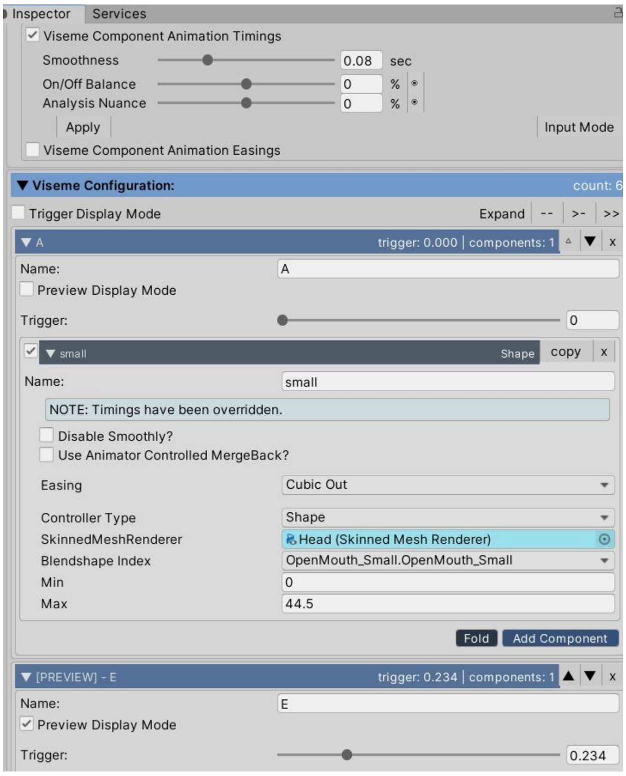


Figure 3.28. Parameters of SALSA

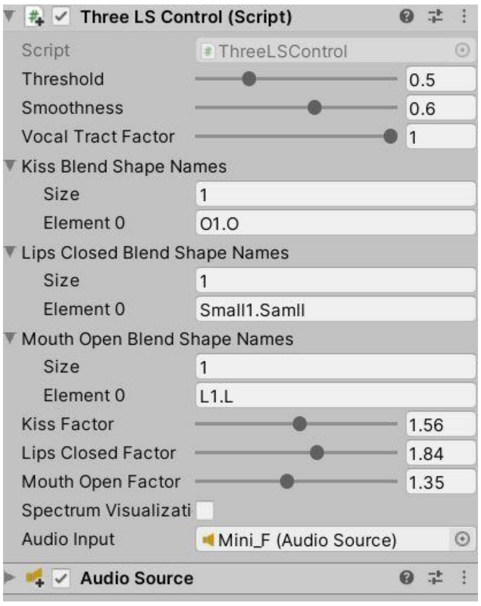


Figure 3.29. ThreeLS

Generally, there are nine categories of phonemes (AI, E, U, O, CDGKNSTHYA, FV, L, MBP and WQ), and each type brings with it multiple shapes of lip. This range is sufficient for lip motion, whereby each category has its own diverse visemes, such as F and V.

The scripts of Lipsync can automatically generate corresponding with recorded audio sources. ThreeLS offers three blend shapes; however, regarding tongue movement, facial expressions and the F, L, TH visemes, its output is lacking in detail. SALSA LipSync is more flexible, but it cannot be used to automatically mark-up which specific visemes are needed in an audio sequence to match a specific phoneme. For example, if the avatar pronounces 'we', the lip shape should be adjusted 'w' viseme by a lip controller in the plugin.

Facial animation is a key component in human-to-human interaction, and AI technology can to a certain extent capture and interlink different emotions and facial expressions.

However, the result might never feel as natural as that of real human-to-human interaction, not least because of subtle, less observable changes such as the generating of wrinkles by different emotional responses.

3.2.2.2 Body animation

Prior to recording, the actor dons a black suit with 57 markers on the key joints spread evenly on its surface (Vicon, 2023), and cameras are warming up. Exporting of body animations can take between 4 and 5 hours.

Almost all animations appear natural and show smooth movements, albeit with small flaws such as incorrect finger movements in keyframes. One of the experiments featured the chair dip, a rehabilitative sports exercise involving a chair, upon the edge of which the actor places their hands. The adjacent keyframes were used to replace ones with false movements, keyframes were joined to form a reasonable trajectory of motion by applying slight orientation adjustments. Delivering a realistic simulation is particularly difficult if the action being interpreted consists of fast non-smooth movements within a short time period. Among the 21 animations, painted skin weights (softening factors applied using a paint brush) are required only for the T-pose character model.

The final rigged avatar shown in Figure 3.10 consists of two textures and 20,197 polygons, involving a transparent texture for parts of the eyes and eyelashes, and head and body textures (both of which are high-quality resolution size (4096 x 4096 pixels). In order to preserve facial expressions and animation, the fbx exported format is used. The avatar with 21 body movements and its facial blend shapes were imported into Unity (see Figure 3.30).

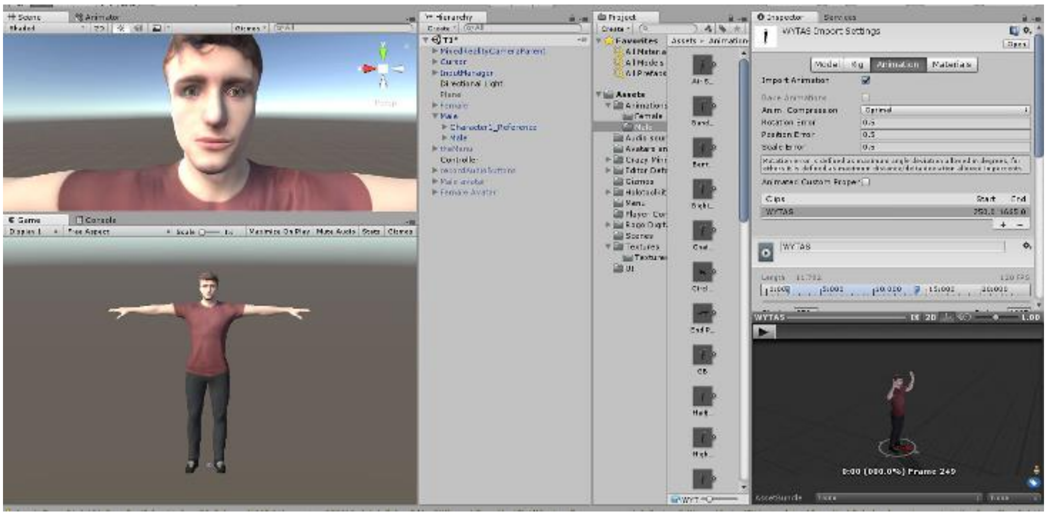


Figure 3.30. The 21 animations and character are imported into Unity (Huang, Twycross and Wild, 2019)

These capture the flexible movements of the legs and fingers, and demonstrate the ability of Unity to work with over 20,000 polygons in this case, and deliver an accurate representation of the actor’s actions. However, fidelity is limited due to the hair and clothes not being reproduced, hence the absence of movement during the rendering. This incomplete motion capture can be compensated by copying the right keyframes, although at the expense of smooth arm movement. The armpits are the most unnatural, less realistic regions, appearing distorted whenever the arms move. It explains the limited capacity of painted skin weights to prevent distortion of the cuff during arm movement, as the clothes and body are integrated into one model. Further, for capturing high-precision exercises, 12 Vicon cameras are insufficient, particularly when the hands cover markers on shoulders, or when the arms stretch round the back.

In summary, the holographic AI should be capable of delivering high resolution textures, smooth animation, and clean meshes, which altogether convey a realistic and compelling representation. The resulting virtual model appears realistic, but owing to reconstruction and editing factors, shares some dissimilarities with the original model. The positioning of some markers may lead to missing trajectories, limiting the fidelity in animation. This may prove to be a significant problem when recording movements.

3.2.3 Procedural animations

The previous section has illustrated the processes for developing the holographic AI's appearance and behaviours, and this section will present three case studies of different ways of generating animations for the holographic AIs, the outputs of which also serve as an open resource for the research community.

Three cases are presented. The first one focuses on generating professional training exercises, created as part of the holoCARE project (Wild, Loesch and Huang, 2019), where a smart glasses application supports cancer-surgery survivors in choosing and then following a suitable set of exercises. This case uses optical, camera-based motion capture with the Vicon system. This case is presented in Section 3.2.3.1.

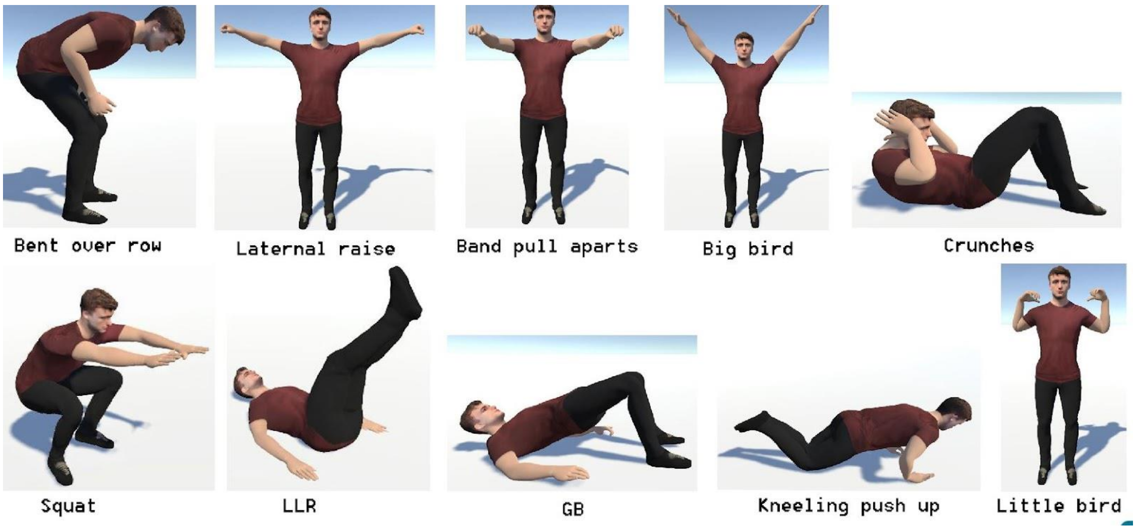
The second presents the extraction of standard animations for MirageXR character models from a public database of full body animations, Adobe Mixamo (Mixamo, 2023). This case is presented in Section 3.2.3.2.

The third case shows how to use an IMU-based motion capture suite, Rokoko (Rokoko, 2023). This case is presented in Section 3.2.3.3, while the outputs are further documented in Chapter 4.

3.2.3.1 Extracting a standard set of animations for rehabilitation exercises

In order to generate professional training exercises for the holoCARE project (Wild, Loesch and Huang, 2019), a female sports student and a male sports student helped me to record animations using Vicon cameras. One product design criterion is that cancer survivors are able to select a matching holographic AI, whose programmed movements are compatible with the user's age, gender and health information. Users do not need to use additional fitness equipment; rather, they can observe and follow the holographic AI.

This section considers the capture of all exercise motions (shown in Figure 3.31). Each exercise takes 5-10 second per time. After workout, the holographic AI will ask the user to provide feedback by asking questions concerning their feelings, and perceived difficulty of the exercise level.



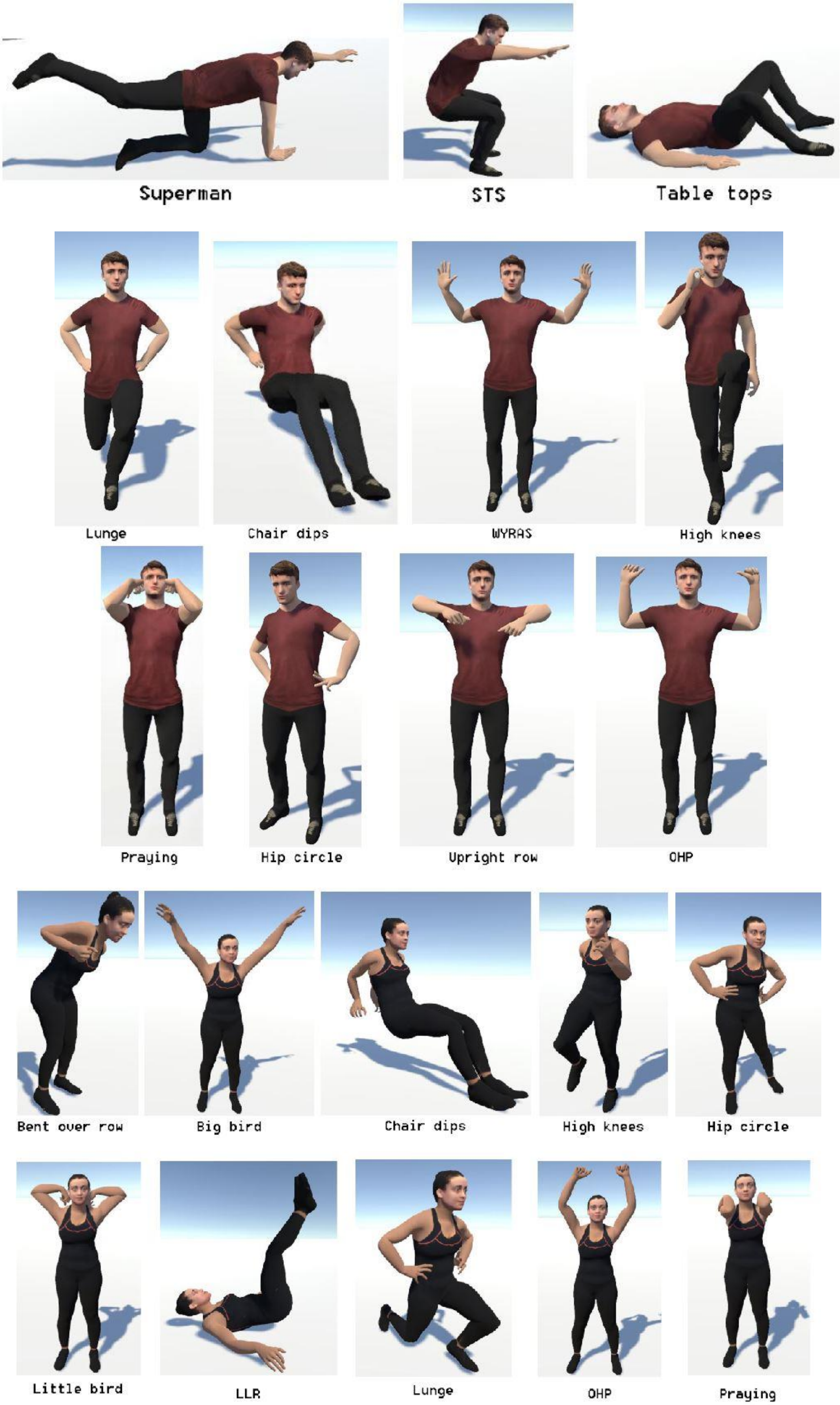

Superman
STS
Table tops
Lunge
Chair dips
WYRAS
High knees
Praying
Hip circle
Upright row
OHP
Bent over row
Big bird
Chair dips
High knees
Hip circle
Little bird
LLR
Lunge
OHP
Praying

Figure 3.31. Motion capture by Vicon

3.2.3.2 Standard animations for virtual characters

During a discussion with the wider project team, it was decided to include the following set of animations in MirageXR, extracted from the free repository Mixamo (Mixamo, 2023). Mixamo is an Adobe product which provides free animations of standard actions such as standing idle, waving, jumping, or fighting. These are useful, when, for example, the holographic AIs needs to introduce themselves. During such introduction, it should adopt the talking or standing-idle animations (see Figure 3.32).

Pose	Description
	Idles: Talking, relaxing, Duration: 30 frames per second

Figure 3.32. Standing-idle

3.2.3.3 Rokoko motion capture

Figure 3.33. Rokoko suit

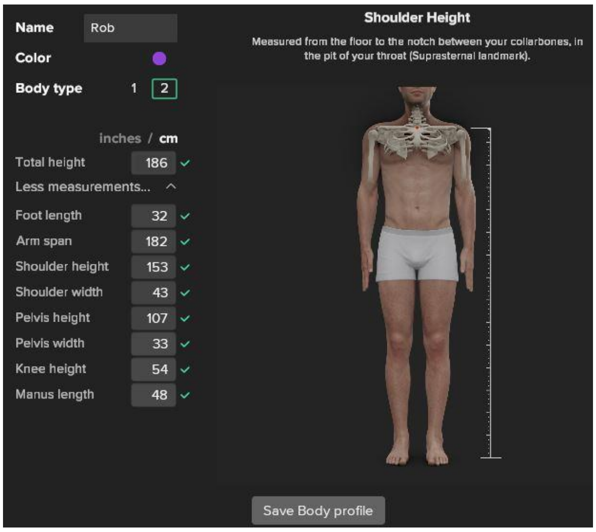


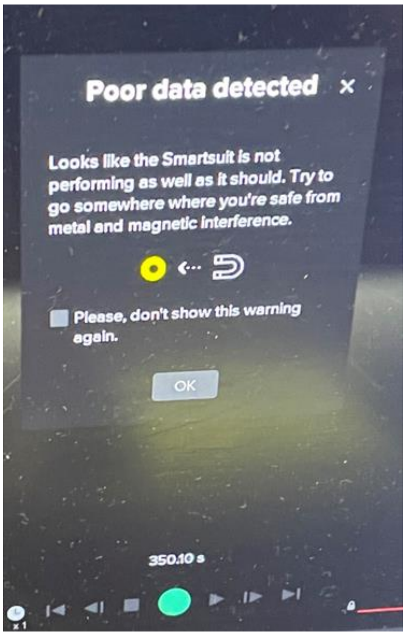
Figure 3.34. User profile

Motion capture can also be conducted using the Rokoko motion capture suit (see Figure 3.33). This results in high frequency joint and limb movement/orientation/position data (~100Hz for 19 sensors). These sensors are labelled on the key joints of the suit, such as shoulder and hip.

In order to review and adjust failed joint animations, it is necessary to record reference video (and audio) in parallel, providing a third person view to compare motions captured with the context in which it was delivered.

Users can set up their own avatars in Rokoko to configure the motion capture (see Figure 3.34). Rokoko motion capture also produces facial animations, via Rokoko remote, but this is possible only on the iOS platform.

Rokoko achieves the same stages of animation as the Vicon, from wearing the suit to cleaning data, but with body-worn ('inside') Rokoko sensors instead of over 40-63 markers plus cameras for tracking. It is not necessary for users to decide where markers should be placed, as body parts will not be covered by others, and the suit is portable. There are, however, shortcomings with the sensor-based mocap suit. Rokoko sensors are sensitive to magnetism (see Figure 3.35), which leads to inaccurate raw data. Moreover, the hub can easily disconnect from the power bank, sensors can get damaged more easily than room-mounted cameras, leading to a malfunctioning suit.



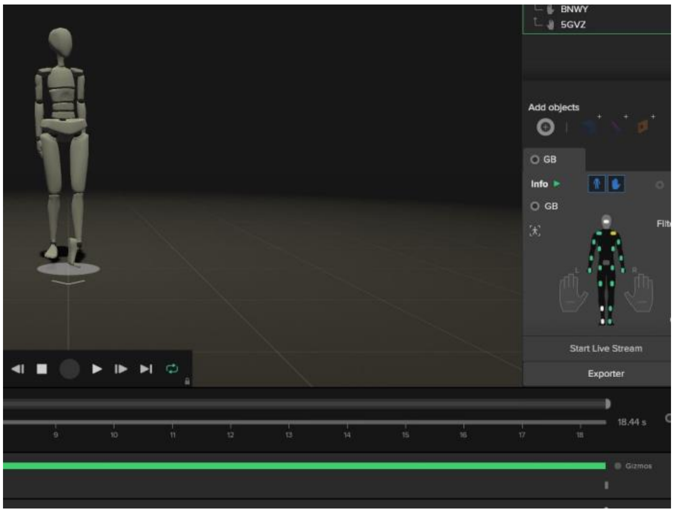


Figure 3.35. Magnetism issue

Figure 3.36. Broken sensors

Figure 3.36 shows three sensors show up with status ‘white’ in the control panel on the right. This indicates that the corresponding sensors on these key joints have disconnected, and no data is recorded for movement of head, left lower leg and foot. In this situation, this further impedes processing as it affects the stabilisation algorithms that try to keep feet on the floor. Both feet should touch the ground or leave the ground at the correct moment if all sensors are functioning properly. Upon failure, it is required to clean these data manually in Maya or Motion Builder. For example, the Rokoko animation is imported to Maya in order to create a new animation layer for an adjusted joint and add keyframes, thus the original data need not be overwritten. Then the failed joint’s rotation and position can be edited and covered by the new keyframe.

The Rokoko suit has high requirement in terms of environment. Metallic objects, such as legs of tables, chairs, cameras, or a watch, can visibly affect the quality of capture, possibly irreversible, or with the consequence that it takes a long time to fix resulting problems.

3.2.4 Recommendations for persona configuration

Based on the PICS model, the persona dimension includes attributes such as size and realism, and the animations consist of facial expressions and body movements.

All cases use HoloLens to display created holographic AIs. Since these assistants are employed in the intelligent tutor system to display exercise animations, so they have life-sized and humanlike appearances and behaviours. MirageXR also has both a mobile and HoloLens versions, where the holographic AI’s size can be adjusted to fit the surrounding or display device of the user. The holographic AI with humanoid and stylised appearances can generate corresponding behaviours based on MirageXR scenarios that can crated by users and target clients.

In addition, the antecedent studies have proved that the photorealistic appearance is able to enhance the sense of co-presence, and these holographic AIs also perform animations for interaction, even though they cannot be fully as natural as human being. However, there is a relationship between persona and the so-called 'uncanny valley'. The uncanny valley theory claims that, generally, users' positive perception of virtual humans improves with how much they resemble real human's regarding appearance and behaviour. If, however, a virtual human appears almost, but not exactly like a real human, this elicits a feeling of uncanniness, and it finds those virtual humans that are close to real but just not simply eerie. In the PICS analysis, the presence of the holographic AI can cause inhibiting effect. For example, the holographic AI with realistic appearance but mechanized or unmatched animations cannot facilitate its effect of performance and intelligence, rather distraction.

Wang, Smith and Ruiz (2019) suggested in their analysis in Chapter 2 that life-sized holographic AI might induce a sense of the uncanny valley, whereas smaller-sized AI typically does not. Nonetheless, the functionality and purpose of the holographic AI should take precedence. For example, a miniaturized holographic AI might be more suitable for use as a personal chatbot on an office desk, appearing more accessible; however, the range of motion for a life-sized holographic AI would be more extensive, taking up more space within the MR environment and possibly interfering with physical objects. The practicality of a smaller holographic AI is not always assured. For example, a holographic AI portraying a virtual patient may be too small for learners to discern its reactions, thus hindering their perception of the agent's responses.

Therefore, a life-sized humanoid holographic AI is created to demonstrate exercises, enabling users to observe and mimic its movements in Section 3.2. If this holographic AI was to adopt a cartoonish or miniaturized form to circumvent uncanny valley effects, users might be unable to execute the exercises correctly, potentially leading to accidents and ineffective instruction. It is also crucial for the humanoid holographic AI to avoid using overstated animations, especially if they do not aid interaction, as this could distract the user.

In a different approach, MirageXR presents a holographic AI feature known as "ghost tracks" (Huang, Wild and Whitelock, 2021). This particular holographic AI is designed with only a basic upper body with simplified gestural animations to assist users in aircraft maintenance tasks. Its design and animations do not detract from learners' performance because the application prioritizes real-time visualization and learning feedback. The simplistic design and unobtrusive animations are intentional, ensuring that learners remain focused on the task at hand.

Based on above reconstruction of holographic AIs, a recommendation of designing persona shows below:

- The size and appearance of the holographic AI should depend on the function and goal. Photorealistic appearance and animations benefit the sense of co-presence, but can also lead into the uncanny valley and distraction.

3.3 Intelligence

AI endows virtual humans with human likeness, and controls their performance on the basis of contextual understanding. Therefore, this section discusses how AR agents can be embedded into user's life. It focuses on verbal interaction, which is the main communicative approach for real humans, as well as spatial understanding and interaction.

This section explores the use of natural language processing, dialogue management, and physical-virtual interaction (discussed back in Chapter 2) for improving usability and implementation ability of the holographic AI. Specifically, it will discuss (1) the use of a question-answering state machine to develop customized dialogue content, and (2) adaptation to surrounding space and embedded activity.

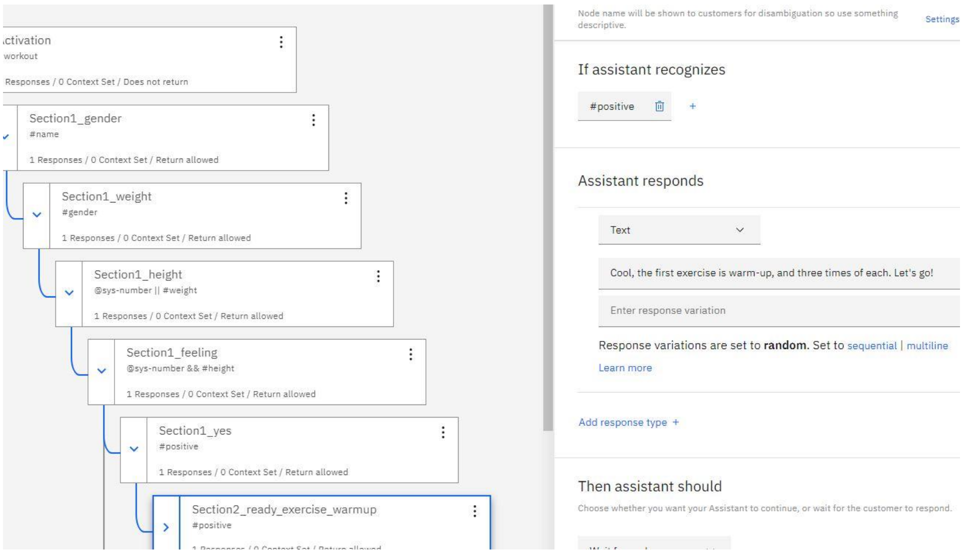


Figure 3.37. IBM Watson: dialogue tree example

3.3.1 Natural language processing

IBM Watson is a cognitive system that is able to, among others, simulate human sound, and is capable of decision-making and answering question using natural language. It allows users to save their information into a repository, so that it can capture key words and automatically select corresponding responses. It provides a chatbot service which can hold users' key words and the holographic AI's answers in a question-response dialogue tree. This service is based on deep neutral networks incorporating with voice signals (Schmidt, Ariza, and Steinicke, 2020). It analyses users' intents based on their voice input and transcript, and offers additional support to identify dates, time, numbers, or letters.

IBM Watson is designed to consider the possibility of answers pertinent to the user's input. Proper dialogue management is important for discerning intent from the raw speech-to-text input.

For example, if the user utterance is "I want to work out", the keywords group containing "work out" and "exercise" is activated, recognizing the user's intent, if it has been featured in the currently active nodes of a dialogue tree. The activated node then helps find a relative output, the response, of the assistant. Next, child nodes in the dialogue tree are triggered by further recognised user intents, the dialogue unfolds sequentially from the nodes to their branching.

Besides, the holographic AI enables to perform specific animations based on IBM Watson by capturing a key word (see Figure 3.38).

```
                                break;
    case "waving":
      myAnimator.Play("waving");
      break;
                        default:
                                break;
                }
```

Figure 3.38. IBM triggers animations

The dialogue tree only can provide a basic information base of the holographic AI, as all conversation content is pre-defined. The dialogue tree also can be interrupted due to failed recognition, or the environment being noisy.

IBM's AI discovery services could provide an extension here. They could be utilised to extend the corpus and to facilitate decision-making about what to include into the knowledge base and what to discard, identifying insights. The audio output of the holographic AI is also synthesized via text-to-speech in IBM Watson, so that the holographic AI finds appropriate voice models with different languages and gender to fit their appearance and characteristics.

```
private string versionDate = "2019-02-28";//"2021-11-27"
[Tooltip("The assistantId to run the example.")]
[SerializeField]
private string assistantId = "b392e76[redacted]
private AssistantService service;
private DaimonManager dAImgr;
```

Figure 3.39. Assistant ID

The Unity engine builds real-time 3D projects for AR platforms, and integrates with IBM Watson using the Watson SDK to perform the communicative and animated holographic AI. In the Unity project, it needs to utilize Watson SDK and must be first authenticated applying an API key and assistant ID. The service credentials (API key), endpoint URL, and assistant ID can connect the application with the web service (see Figure 3.39).

In order to add a level of verbal interaction and translation service to the holographic AI, it needs five scripts: For activating holographic AI ('Daimon' manager script), dialogue service, speech input service, speech generation (output) services, and, optional, second language translation services.

The activation script manages the holographic AI's animator trigger, and speech input and output scripts in Unity. Figure 3.40 also displays how the script controls the holographic AI's position of look by 'lookTargets', thus the holographic AI can recognize real objects and gaze at it.

In the dialogue service, it connects other two speech to text, text to speech scripts, a user profile for recording health information or preference, and series of animations of exercise (see Figure 3.41).

In this example, the aim of the holographic AI is to induce users into doing more exercise, and to let them follow the holographic AI's performance.

```
private SpeechOutputService mySpeechOutputMgr;

public GameObject myCharacter { get; set; }
private Animator myAnimator;

public GameObject[] lookTargets { get; private set; }

private Dictionary<string, object> _context = null;
private bool _waitingForResponse = true;
```

Figure 3.40. Daimon manager class

```
[SerializeField] private SpeechInputService mySpeechInputMgr;
[SerializeField] private ExerciseController ExerciseController;

// - - - - - - - - - - - - - - - -

[SerializeField] private string Name;

public enum gender
{
    male, female
}

public gender Gender;

public int Age;

public double bmi;
public int Height { get; private set; }
public int Weight { get; private set; }

public enum cancers
{

}
public cancers CancerType;

public enum Frequency
{
    threeMonths,
    threeToSixMonths
```

```
// list of exercises; we remove them bit by bit
[SerializeField] private List<string> Exercises;

public void Reset()
{
    Exercises.Clear();
    Exercises.Add("A1");
    Exercises.Add("A14");
    Exercises.Add("A15");
    Exercises.Add("A21");
    Exercises.Add("B");
    Exercises.Add("B11");
    Exercises.Add("B12");
    Exercises.Add("B13");
    Exercises.Add("B14STS");
    Exercises.Add("B22");
    Exercises.Add("B32");
    Exercises.Add("B33");
    Exercises.Add("B41");
    Exercises.Add("B41Y");
    Exercises.Add("B42");
    Exercises.Add("B51");
    Exercises.Add("B51");
    Exercises.Add("B52");
    Exercises.Add("B61");
    Exercises.Add("B62");
    Exercises.Add("B72");
    Exercises.Add("B82");
    Exercises.Add("B83");
    Exercises.Add("B86");
    Exercises.Add("B1230");
    Exercises.Add("B4260");
    Exercises.Add("B7230");
    Exercises.Add("B8330");
    Exercises.Add("B8630");
    Exercises.Add("B62120");
    Exercises.Add("Plank");
```

Figure 3.41. User profile and lists of exercises

The speech to text service enables speech content to be transcribed and print out in different languages, before being transmitted to the dialogue tree. Written text is converted to audio by means of the text to speech function. In the DAIMON implementation, the holographic AI can translate English to German and Chinese using translation service (see Figure 3.42). However, the speech to text function can only accommodate articulation of a language in a certain accent, and provincial dialects will impair interaction.

```
private LanguageTranslatorService languageTranslatorService;

public string translationModel = "en-de";
public string versionDate = [redacted];
//public string apiKey = "";
//public string serviceUrl = "https[redacted]

public string lastTranslationResult = null;
```

Figure 3.42. Translation script

Furthermore, SALSA lip-sync can work with the Watson services in Unity, and so the holographic AI can simulate human social interaction using facial animation.

The key speech features will be mapped to lip parameters when the audio source has been analysed (Llorach et al., 2016), and then facial blend shapes will be driven and synchronized voice output. Figure 3.43 displays a workflow of natural language processing and facial animations.

Lip and facial expressions represent a person's attitude and identity (Dineshshankar et al., 2013). By combining natural language processing and lip sync, users can engage in hands-free interaction with their interlocutors, which satisfy two sensory. Further, IBM Watson provides tone analysis by capturing emotional written text, but it only recognizes tone, and does not engage in simulating and responding to proper intonation and speech rhythm.

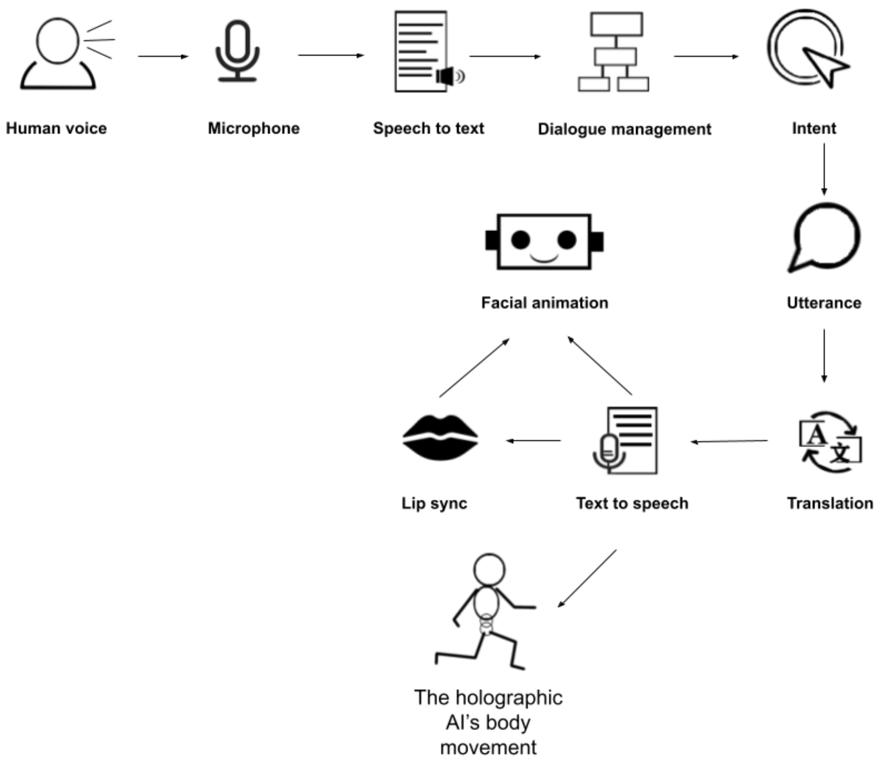


Figure 3.43. Workflow (own graphic)

3.3.2 *Spatial understanding and interaction*

Spatial awareness is one of the key features of AR as it enables the holographic AI to recognize physical surroundings and objects. As mentioned back in Chapter 2, although unidirectional interaction only enables the holographic AI to build an understanding of physical objects and environment, it can prevent out-of-control situations. By contrast, with bidirectional interaction a pre-defined scenario for managing physical objects is required, and their states are changed using additional sensors. This sub-section illustrates how the holographic AI in MirageXR dynamically recognizes the user's surroundings and executes unidirectional interaction.

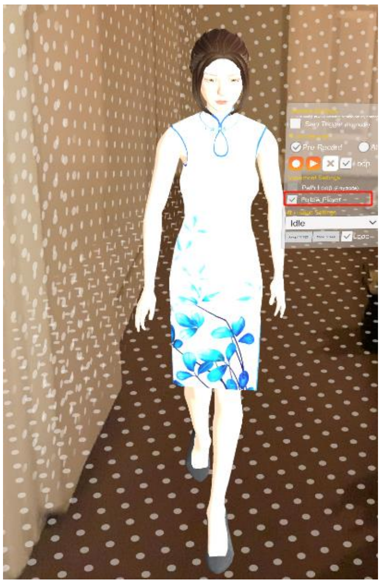


Figure 3.44. ‘Follow Player’

The holographic AI is capable of detecting and following the user’s position, if ‘Follow Player’ switches on (see Figure 3.44). The aim of spatial mapping is to create a surface for placement of holographic AIs, enabling users to put virtual objects in their places. The holographic AI shown in Figure 3.43 can recognize the wall and floor by way of spatial mapping. Then, the relationship between the data of the camera and that of a known object in the scene are calculated, and rotation between the captured object and the camera are executed (Oufqir, El Abderrahmani and Satori, 2020). Therefore, it first calculates the distance between the camera’s coordinate position and the holographic AI’s location. The holographic AI’s new direction is updated continually, and it triggers walking animation when the distance is over a parameter, so that the holographic AI can follow users. Similarly, users can mark anchors in different directions and locations, and ask the holographic AI to follow the trajectory (see Figure3.45).

```
private void Start()
{
    _cameraOffset = transform.position - PlayerTransform.position;
    _newPos = PlayerTransform.position + _cameraOffset;
    _lookPos = PlayerTransform.position + lookOffset;
}
```

Figure 3.45. The script for calculating distance between the holographic AI and the user

The holographic AI in MirageXR cannot fully perform physical-object understanding (i.e. bidirectional interaction), but this application pursues a personalized open source of an AR training system.

3.3.3 Recommendations for intelligence

Intelligence refers to the holographic AI’s internal abilities that the user perceives its function after interaction, such as spatial understanding and natural language processing. In addition to adapting to explicit and implicit user demands, it is able to execute the tasks intended by the user.

The unidirectional interactive approach is reliant on physical-object recognition and spatial understanding. The holographic AI should be able to avoid colliding with physical objects by overlaying into an appropriate position in real time. In AR, physical-object interaction is reliant on simulation of a specific event, as well as an additional equipment, it can enhance an illusion of physicality, but it is harsh to deal with unexpected context.

The holographic AI discussed in this chapter possesses spatial awareness, enabling it to recognize and track a user's movements. However, it cannot physically interact with or affect real-world objects, as it is designed for unidirectional interaction, which is determined by its intended use.

Chapter 2 reveals that previous holographic AI systems provided standard services without the ability to tailor experiences to meet the individual needs of users. With advancements in natural language processing and adaptive learning technologies, the holographic AI can now generate personalized workout animations. Furthermore, natural language processing facilitates the control of specific body movements during these exercises. For example, if the user’s utterance is ‘Hi’, the holographic AI performs waving or greeting animations with smiling facial expressions, and then conducting next performance based on other conditions.

Therefore, according to the reconstruction of holographic AIs and the PICS analysis, two recommendations of intelligence show below:

- Verbal communication is a way to collect user information and provide customized services by adaptivity and user models.
- The bidirectional and unidirectional interactions depend on whether the holographic AI is required to manipulate real objects. Although the bidirectional interaction can reinforce the sense of co-presence, the holographic AI only can influence pre-defined objects.

3.4 Conviviality

People prefer to interact with holographic AIs with more humanlike features, and they have a tendency to treat virtual humans in the same way as they do towards real people, in keeping with social norms (Miller et al., 2019). However, users are very sensitive to artefacts, and every single unnatural performance by the holographic AI can undermine social interaction. This problem may be magnified by face-to-face communication in AR. Although previous studies have considered various factors that could affect user experience, some aspects have generally been overlooked.

User experience a sense of co-presence when they feel that the holographic AI coexists with them in the same space. This feeling stems from the illusion of physicality, since the holographic AI is not in the real world. Therefore, past studies have tended to focus on physical-virtual, bidirectional interaction, and various transitional approaches in the MR environment. Users can perceive the holographic AI is ‘being’ with them, as this sensation is from both physical and virtual interaction. Virtual presence refers to the way in which the holographic AI can interact with the virtual objects, and physical presence means that the holographic AI can recognize or interact with real objects. For example, if the user delivers a virtual object to the holographic AI, the holographic AI needs to firstly track this virtual object, and then catch it from the user's hand in a natural manner. Further, the holographic AI should exhibit social awareness in both virtual and physical situations. For instance, MirageXR provides a service, with which users can deposit virtual icons in an effort to determine where the holographic AI stands, and watch the holographic AI walk through the pathway of markers.

Another important facet of the holographic AI is warm communication. For example, the holographic AI in MirageXR can be a chatbot which engages in a 'warm-up' at the beginning of interaction. The holographic AI expresses empathy towards the user by asking inviting questions like 'How are your feelings today?'. Although there is no evidence to prove that the warm dialogue can enhance the possibility of social interaction, it generally offers the user a better experience than output consisting of cold utterances.

To build a user model, the holographic AI can collect the user's data via natural language processing to produce a set of preferences. For example, the holographic AI in the holoCARE (Wild, Loesch and Huang, 2019) first needs to understand user's requirements, health conditions, age, weight, etc., creating a user model that saves and updates this data. For adaptivity, it then extracts the data again, conducts self-learning, and makes decisions on which exercises to recommend for the user and stated needs.

The holographic AI should exhibit social awareness by being imbued with the ability or initiative to stimulate the user's interactivity by behaviours, language, and determination of the user's preference. However, although such a holographic AI may appeal to those with no experience in AR or VR, it is possible that those users might lose interest in engaging in such further interaction.

3.4.1 Recommendation for conviviality

As the holographic AI operates within a MR environment, distinct from VR or the physical world, most studies in the last chapter focus on a single factor affecting co-presence or social presence. The research seeks to determine how the recognition of real-world objects can influence the user-friendly nature of the system, assuming that users are allowed to engage with the holographic AI. However, few studies consider whether users are willing to interact with the holographic AI, a perception crucial for building a positive and lasting relationship. There is also a shortage of research addressing the measurement of holistic user experiences.

Mayer et al. (1995) define trust as the willingness to be vulnerable, a key determinant of a user's readiness to engage with the holographic AI. While Kim et al. (2016) measure trust, their focus is on safety rather than the user's disposition or attitude. If users do not trust the holographic AI, they may be reluctant to accept or continue using it. Therefore, Chapter 5 highlights the importance of trust in human-holographic AI interactions. The following recommendation is proposed:

- Conviviality should represent the interactive quality and overall user experience with the holographic AI. Although holographic AIs differ in characteristics, they should aim for a comprehensive user experience.

3.5 Sense

	Input modalities (users – holographic AIs)	**Output modalities (holographic AIs – users)**
Real world	• Temperature capturing, • Ray tracking, Spatial mapping, Spatial/physical-object understanding	• Physical-object management / influence • Avoiding collision
Virtual world	• Virtual objects interaction, • Virtual and physical information connection	• Information of the virtual context update
User	• Gaze tracking, • Voice input, • Head tracking, • Location tracking	• Animation, audio, natural language processing, eye contact

Table 3.1. Input and output modalities in AR

Senses are related to interactive approaches, which are characterized by input and output modalities. Barakonyi and Schmalstieg (2005) proposed a table listing input and output modalities in the physical and virtual worlds. This section expands their table and its contents by dividing it according to three situations: real world, virtual world, and the user (see Table 3.1).

The term input modality concerns the cues which help the holographic AI recognize the user, such as their speech and dialogue, user gestures, eye gaze, space proximity (for following the user), or head tracking (the user to the holographic AI).

The term understanding refers to the holographic AI's ability to process data from the user during conversation and determine the user's intentions.

The term output modality concerns the holographic AI's responses to the user, which in turn are composed of audio output, animations, eye contact, facial expressions, speech, and dialogue. In this vein, the term responsiveness is defined as the ability of the holographic AI to react intuitively and quickly both to the digital and physical spatial surroundings of the user, and to perceived activities, and therefore its ability to maintain a sense of corporeal presence (Campbell, 2014). The main ways by which users and holographic AIs interact are via gestures and speech, and so the measure of a holographic AI's interactive ability is its ability to utilise data supplied by the user in such a way that its output matches the user's needs (as discussed back in Section 3.2: Intelligence).

The Input modality of a holographic AI should be such that it can track pre-defined real objects based on the objectives of the application, such as temperature, ray tracking, physical-object awareness, or spatial understanding.

One notable example is the ‘ghost tracks’ IoT (Internet of Things) tool developed by MirageXR, designed to facilitate XR-based learning by allowing experts to provide learners with holographic guidance in the form of ghost tracks with the use of smart glasses containing built-in sensors, which enable capture and anchored replay. The MirageXR tool contains floating agents capable of recognising objects, which can tutor learners by demonstrating tasks in the virtual environment.

The term virtual input modality refers to virtual-object management and connection between real and digital contents. It can be activated by the user’s input and input tools, such as gesture and the voice controller.

As mentioned in the previous section, speech and dialogue are integral to natural language processing as they capture the user’s ways of expression. Not only can a user’s gestures replace the clicks of a computer mouse, but they can also supply a virtual object with a more specific, user-defined gesture as a virtual output modality. For example, Wang et al. (2021) have developed the GestureAR system, which uses a visual programming interface that enables different types of hand gestures to match virtual objects’ performances, such as real hand can grab a virtual cup.

Section 3.3.2 considered how the ability of the holographic AI to follow the user’s footsteps, i.e. spatial awareness. Previous studies have investigated social distance between the user and holographic AI, and have applied different fixed values for determining a proper distance. For example, introverted people either prefer to maintain a certain distance from the holographic AI, or prefer to stand closer to it. Therefore, the space trigger provides a satisfactory and user-defined approach for avoiding too-near or too- far interaction.

Gaze tracking is also critical in input and output modalities. It can be used to avoid unintended behaviours and improve accuracy of eye movement (Papadopoulos et al., 2021). In the holographic AI to user interaction, eye contact is imperceptible yet important element of social interaction which enhances the user experience, and it has been demonstrated that eye movement is related to cognitive activity (Nikolaev, Pannasch and Belopolsky, 2014). Therefore, eye tracking enables the holographic AI to identify the user’s gaze orientation and garner data of the user’s visual attention. Eye tracking is based on saccade, fixation, and both time and space are factors of gaze duration at attractive objects (Lai et al., 2013). Kapp et al. (2021) have developed an open-source AR toolkit and R package for data analysis. HoloLens 2, which tracks the user’s eyes using infrared cameras, gathers the gaze data (including gaze position, gaze point, and reference), and transfers the data to Unity 3D where it is recorded, and its accuracy is evaluated. During the holographic AI’s interaction, eye tracking information based on the user’s gaze position is gathered. For example, Pfeuffer et al. (2021) have developed aRtention for gaze input. The additional interactive interface displays context/object – based information. The aRtention system has three dimensions: the first considers the volume proportion of the virtual information panel (transmission of virtual and physical information); the second enables the user to browse different levels of information via gaze interaction; and the third is task transitions, which can adapt to content based on the user’s choices via expanded information. aRtention system can serve as an intelligent tutor system, as trainers using the program need not have to rely on book reading or other physical resources to convey knowledge and skill.

3.5.1 Recommendation for senses

While sensory perception is an aspect of intelligence, this discussion centres more on how holographic AI integrates various capabilities to perceive and respond to the context of interaction. Current AR technologies focus on different interrelated input and output modalities, and these multiple and mixed interactive approaches affect conviviality. However, previous holographic AIs limited interactions to one or two senses, such as natural language processing, spatial awareness, or object recognition, without addressing how the AI could simultaneously process information from user motivation, and from both virtual and physical environments, taking into account diverse contextual elements to achieve optimal results.

Past studies have typically focused on one-to-one interactions, which means the holographic AI may not be designed to recognize and meet user-centred interactive needs and services, particularly in educational settings. For instance, a holographic AI that offers a uniform learning curriculum fails to address varying levels of student proficiency. Advanced students might engage in redundant review of known material, while beginners might struggle with the pace and complexity of the content. Thus, social awareness is proposed as an additional factor to enhance user experience. The development of a user model, as described in Section 3.3.1, aims to identify and satisfy user requirements through adaptivity. Additionally, students may not have a clear understanding of their objectives within a specific domain. Hence, the holographic AI should estimate and refine the user model by gathering and analysing user information and responses to facilitate a broad range of interactive and user-centric content.

On the basis of the above discussion on interaction modalities, and in view of recent developments such as MirageXR, the following a recommendation is proposed:

- The user model and adaptivity act as interfaces for interaction modalities, in that they capture user input within the user model. This model is designed to self-update throughout the interaction to enable adaptive responses, thus generating pertinent context in real-time.

3.6 Model Revision

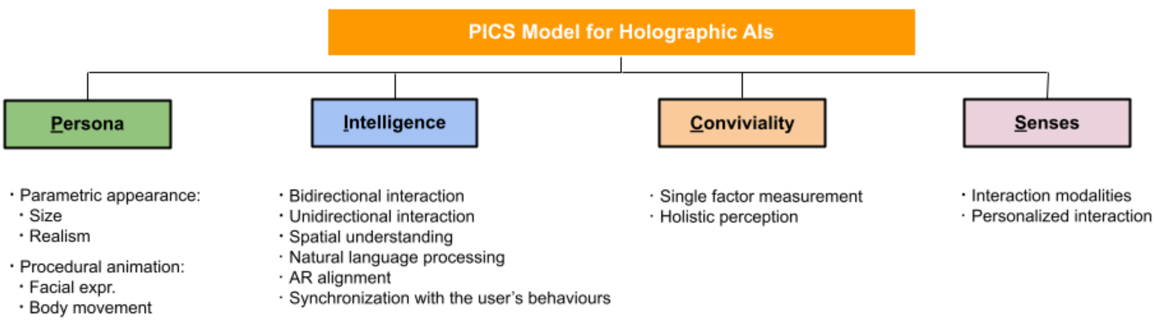


Figure 3.46. The refined model

According to this case study of creating the holographic AI, the previous PICS model is refined (see Figure 3.46).

The dimensions and components of the persona are taken into account when creating the holographic AI. Although 3D scanning and motion capture can generate realistic appearances and animations, the depiction of Sarah (refer to Figure 3.21) exhibits a cartoon-like portrayal with human characteristics by using traditional methods. The dimensions of the holographic AI can be tailored to meet user specifications. Although the body animations are contingent on the AI's functions and goals, this does not take into account whether the AI's verbal output aligns with its non-verbal cues.

In the prototype model, learning systems are suitable for use in intelligent tutor systems, simulation agents, and educational games, with a focus on application development rather than mastering a particular ability. Consequently, this aspect has been omitted, yet it has been examined within the context of the holographic AI's classification.

Computer vision is essential for recognizing physical objects, as well as for spatial mapping and understanding. Chapter 2 details how OpenCV has been implemented in marker-based augmented reality (Verma et al., 2021). Nevertheless, this facet was deemed redundant and overly broad in the initial PICS model, complicating the selection of specific intelligence components. For instance, both marker-based and marker-less AR are applicable to mobile devices. Marker-less AR permits the placement of virtual items in any location without markers. A trigger image can track the holographic AI, yet its visibility is lost when the camera angle shifts away from it. This case study utilizes spatial mapping to exhibit the holographic AI, which can be better described by AR alignment as it synchronizes.

Campbell et al. (2014) state that holographic AIs can alter their state in a dynamic interactive space; otherwise, they remain static. The holographic AI resides in an MR environment, where virtual space can be easily manipulated, unlike in the real world. Prior studies have used sensors and controllers to interact with specific real-world objects. Thus, this interactive setting is deterministic, with each action being predictable (ibid). For instance, the holographic AI is capable of moving real tokens when following a pre-established interaction protocol (Lee et al., 2021). However, the classification of environments as random/deterministic or dynamic/static falls short in describing whether the holographic AI can perceive or influence a hybrid reality. Nonetheless, the prior chapter utilizes the concepts of bidirectional and unidirectional interactions to discuss physical-object recognition or interactions. Holographic AIs with bidirectional interaction capabilities can manipulate and influence physical objects, but should first recognize them. Unidirectional interactions, on the other hand, are limited to recognizing physical objects. For example, the holographic AI in MirageXR can locate the user. While it cannot manipulate physical objects, it can respond to stochastic behaviours by tracing the user's movements. As a result, bidirectional/unidirectional interactions replace physical-object recognition/interaction.

The primary mode of engagement is natural language processing. This system encompasses dialogue management, language translation, and speech recognition and synthesis. The holographic AI system extracts predefined keywords to generate appropriate responses in line with the set topic and conversational context. If a user provides irrelevant input, the holographic AI may not ascertain the user's intention. Therefore, the system should possess the capability to detect such deviations and steer the dialogue back to its intended trajectory.

Conviviality denotes the quality and user experience of the interaction, a topic that will be explored in Chapter 5. Prior research has honed specific capabilities to assess whether the holographic AI can enhance user perception by comparing or manipulating variables. Consequently, the elements or factors that augment user experience are varied, encompassing physical-object interaction and recognition. These studies might be viewed as single-factor assessments since they evaluate the influence of the holographic AI's cognitive abilities or aesthetic on interactive engagements. However, a comprehensive evaluation of the holographic AI is also vital to determine its acceptance, reliability, and relationship with the user. The concept transcends a focus on a singular capability, function, or representation of the holographic AI. Therefore, conviviality involves both single-factor assessments and an overall perception.

Sensory detection examines how the holographic AI perceives and responds to the environment and users, including eye tracking, facial recognition, and position sensing. Additionally, the holographic AI can utilize natural language processing to compile and refine a user model that enhances adaptability, thus providing a tailored and user-centric experience. As a result, adaptability extends to sensory detection and merges with the user model to facilitate personalized interactions. The differentiation between intelligence and sensory detection is in the manner of interaction with the user and the method of information exchange; intelligence, conversely, relates to the ability to perform tasks and solve problems. Therefore, a holographic AI can amalgamate sensory detection with varying levels of intelligence. For instance, Azuma Hikari is a holographic AI that engages in bidirectional interaction by recognizing specific hand gestures to control lighting (Huang, Wild and Whitelock, 2021), and learning feedback can be derived from integrating semantic analysis with eye tracking (Hartholt et al., 2019). Therefore, senses include interaction modalities and personalized interaction.

In conclusion, the initial model is refined based on this case study. However, not all components are chosen for the development of the holographic AI in this chapter, such as bidirectional interaction, AR alignment, or co-presence measurement.

Besides, the logical extension of the work presented here next would be to investigate further means and capabilities to fully implement unidirectional or bidirectional environmental interaction, equipping the holographic AI with IoT functionality.

In referring to the PICS model and case experience, the elements of the model are optional, depending on the aim of application. However, they should be applied in such a way that the user is convinced of their need for the holographic AI and AR. Therefore, the following five recommendations are proposed:

- Persona: The design and scale of the holographic AI must align with its intended purpose and objectives. Achieving a balance between photorealism and behaviour is crucial to avoid the uncanny valley effect, which may otherwise disrupt user engagement.

- Intelligence: For holographic AI, natural language processing serves as a critical method for initiating animations and gathering user data. It should also be adept at recognizing and rectifying erroneous inputs.
- Intelligence: Both bidirectional and unidirectional interactions can handle a limited array of predetermined objects in real-time, bidirectional interactions necessitate a designated space for the management of physical entities.
- Conviviality: The user experience with holographic AI is influenced by multiple factors, such as the degree of interactive capabilities and the AI's persona. However, it is essential to consider the overall user perception, which includes trust, usability, and acceptance.
- Senses: Multimodal interaction must cater to user preferences and requirements, ensuring a fluid integration of the virtual and physical realms to deliver personalized experiences.

3.7 Summary

This chapter builds upon Chapter 2, which details the elaboration of the PICS (persona, intelligence, conviviality, and senses) design model to form a comprehensive process for designing a holographic AI, including 3D scanning, reconstruction, animation creation, and intelligence development. Additionally, the case study refines and validates the preceding PICS model by restructuring and consolidating similar or redundant elements. The result is a more succinct and focused model. These considerations have been incorporated into the proposed guidelines to underscore the crucial aspects of holographic AI development.

The first section of this chapter has examined the process pipeline and approaches for 3D scanning and reconstruction of body movements, specifically facial expressions. A high-quality reconstruction model is capable of scanning a large flux of data, and delivering textures with no distortions or missing meshes.

The kind of 3D reconstructed character described earlier in this chapter is different from traditional humanoid ones used in games and social media, such as Lil Miquela, a virtual influencer. Reconstruction is complicated task: sculpting high-poly models with skin textures is a tenuous process during which some data are lost, and the resulting the high-poly 3D model might be missing certain details or structures. For example, eye parts include the lower/upper eyelids, palpebra, lacrimal gland, lacrimal sulcus, and lacrimal sulcus, and 3D scanning might fail to capture all these tiny structures. The approach of directly baking normal maps and ambient occlusion maps from 3D scan models does not circumvent the problem as it yields a blue graphics for the normal map and a wrong white image for the latter kind.

As detailed above, two methods for generating animations have been developed: an optical, camera-based system (Vicon, 2023) working with key markers, and a body-worn smart suit system (Rokoko smart suit) for sensor-based capture. The quality of animations from the multiple cameras are more stable, but it encounters the problems of some markers being covered by others.

The sensor-based capture of the Rokoko suit is more portable, but it has strict magnetism requirements. For example, if the suit touches metal objects, legs, or arms of the avatar will be distorted, and it can be difficult to clean up later requiring the failed parts of body movements to be reconstituted (key) frame by (key) frame.

Three Unity plugins were compared for producing facial animations and lip sync, and found that SALSA works better than the others, but has licensing restrictions. It provides more optional parameter of blend shapes, integration with emotions, eye blinking, and lip movements.

Intelligence components include natural understanding and spatial understanding, given that verbal communication is the main human interactive approach, and spatial understanding is key feature of AR. As regards to natural language processing, MirageXR applies dialogue management, speech to text, text to speech, and translation services.

This chapter has described the development of a distance trigger which enables the holographic AI to follow the user's footsteps via spatial understanding, helping the holographic AI find the user, offering a proof of concept how spatial personalisation can be performed. This could be further developed to reduce issues of social distance in connection with users' different personalities and/or cultural backgrounds. Although the MirageXR holographic AI's ability to interact with users is basic, it focuses on user-centred tasks.

The development of the holographic AI need not be limited to a predefined pattern. For example, if the simulation agent relies on a pre-trained scenario with various possibilities in terms of results, users, however, cannot conduct self-made conditions, so tutors cannot build a scenario that is based on students' learning situations.

Further, the types of holographic AIs in MirageXR are mixed: each one can be an intelligent tutor and chatbot simultaneously. Users can self-create different scenarios based on their requirements. For example, the holographic AI could teach primary Year 4 mathematics, guide students to positive behaviour, teach trainers to check and maintain precision equipment, or provide electrocardiogram training.

Conviviality is social interaction, linking with dialogue understanding, but serving a different purpose. Dialogue can be utilised to implement warm-ups and icebreakers, to make users feel warm their interaction with and social awareness of the holographic AI.

Senses express the system's input and output modalities, concepts which involve both physical and virtual worlds, and the user. The input modality of the holographic AI can be defined as its understanding and awareness, and is contingent on users' detection functions, such as capturing utterance, user gestures, gaze trigger, space trigger, etc. The output modality is its responsiveness/ reaction, the way the holographic AI responds to user input in terms of traits such as animations, eye contact, facial expressions, etc.

Chapter 4 Extracting Teaching Gestures for Animation: An Experiment

4.1 Introduction

In previous chapters, the concept, model, features, and creation of holographic AI have been discussed. However, the previous studies that are discussed in Chapter 2 do not provide that their holographic AIs can perform diverse gestures as a non-verbal interaction, even though they have animations, especially in the pedagogical domain. As an example, the holographic AI can merely perform basic animations such as standing, breathing, and pointing, to measure user perception, or prove animation can influence social presence. Therefore, this chapter investigates the influence of gestures on social awareness, academic achievement, and user experience, and explores the different types of gestures that could be useful for the holographic AI, within a pedagogical domain. This chapter, however, does not develop the holographic AI's model since it is beyond the scope of the research.

A gesture is a primary method of expression that is spontaneously generated, and that may even replace speech in communication due to its relationship with language generation and perception (Wagner, Malisz and Kopp, 2014). Non-verbal language also reflects speakers' mental processes, linguistic organization, and level of cognition (Lyons and Semantics, 1977; Goldin-Meadow and Brentari, 2017). Therefore, gestures can be considered ‘visible voices’. In contrast to other movements, gestures serve informative and interactive functions (Lyons and Semantics, 1977). As an example, the size of an object can be schematized (represented by) figures or hands, and the orientation of an object can be described using pointing gestures. Co-speech gestures, which include spoken utterances accompanied by gestures, are a common form of communication. So et al. (2014) demonstrated that meaningful co-speech gestures help children and adults remember more words. Further, gestures also serve unvalued functions, such as habitual gestures, which may have potential functions or linguistic cues that can reflect a person’s habits of thinking and movement (i.e. habitual thinking), cognitive ability, personality, and emotions.

In games and intelligent personal assistants, gestures are frequently used to enhance the user’s sense of immersion. One such program is NEONS, a virtual assistant that mimics the speech of humans. In a promotional video, NEONS employs motion capture to generate a series of co-speech gestures. However, NEONS lacks the capacity of producing abundant gestures when speaking, despite learning from real human emotions. Gestural animations are also critical for virtual characters, especially for intelligent tutor systems. Such animations should incorporate human body affordances (Cassell, Vilhjálmsson and Bickmore, 2002). Therefore, an animated virtual teacher requires not only speaking abilities but also body movements to enhance natural face-to-face interaction, rather than displaying only head movements and facial expressions, especially when highlighting key learning points. For example, when a teacher asks students to look at a number accompanied by a pointing gesture, which can attract student’s attention.

The co-speech entails multiple dimensions, since meaning and timing are co-produced (Abner, Cooperrider and Goldin-Meadow, 2015). Therefore, if speech rhythms and gestures simultaneously appear in a moment, which helps grasp key messages or serve to filter unmeaningful messages in both physical and virtual environments. For instance, when people portray the shape of an object, their spoken utterance and gestures may occur in parallel, or be backward in thinking, depending on the way of human user is searching for selecting key information. Although the virtual human attempts to simulate human performance speech, and the delays or pauses undermine its performance. While, Piwek (2014) claimed that a humanoid character with animations can improve familiarity regardless of whether it is natural, rather than producing uncanny valley effects. However, if the semantics of gestural animations cannot match the speech content, and the virtual human can only open their arms while speaking, this will create confusion and an uncanny feeling, since a real person generates different gestures in speech.

Therefore, this chapter reviews past studies concerning gestural investigation. It also reports the findings of experiment, the purpose of which was to collect data in relation to gestures, classify key holographic AI gestures, and explore how gestural animations can be incorporated into a pedagogical holographic AI.

This chapter is organized as follows: Section 4.2 reviews previous studies; Section 4.3 describes a methodology for data collection in scope of key gestures; Section 4.4 details the experiment set up and data collection, and Section 4.5 presents the data analysis and findings. Section 4.6 discusses the experiment in detail, and Section 4.7 provides what gestural animations can be produced and used for a holographic AI, and limitations of this study represents in Section 4.8. Then Section 4.9 summarizes the findings of this chapter.

4.2 Previous Studies

A gesture can emphasize meaning, guide movement in real-time and through space, or even be used for more abstract tasks (Goldin-Meadow, 2011). Gestures also serve a critical function in education. It is known that gestures can be utilised in teaching to positively influence both learning outcomes and learner satisfaction (Lester et al., 1999). The virtual teacher in an intelligent tutoring system can help learners obtain and review knowledge. Although a virtual human can mimic real teachers regarding speech and behaviour, virtual teachers today often use rather repetitive animation loops with a limited repertoire, such as standing for a long time, pointing in a specific direction, or crossing arms, such as FaceMe system (Li et al., 2021). Such mechanistic behaviour cannot adequately mimic the performance of a good teacher or trainer, and insight is needed into a wider repertoire of gestures and the context and conditions under which they can be deployed to support understanding and guiding of student attention, rather than distracting them (Kajopoulos et al., 2021). This finding suggests that an explicit investigation into how the real teacher’s gestures can be transferred to a virtual humans’ behaviour, which does not merely rely on simulation, should be undertaken.

However, there have been studies which specify exactly what these key instructional gestures are, and which gestural animations can be deployed by a holographic AI to improve the user’s sense of trust and experience. Therefore, in order to understand the taxonomy of gestures and how to encode them, Section 4.2.1 firstly reviews types of human gestures and their distinction, and then consolidates them into an investigative framework. Section 4.2.2 considers which types of gestures are frequently used in teaching, and whether they can affect learning outcomes. Section 4.2.3 considers how a virtual human could apply gestures to facilitate interaction. Section 4.3.4 reviews how to analyse data of gestures by coding. Based on the review of previous studies, the research questions and knowledge gap are outlined in Section 4.2.5, a brief plan to fill the gap is provided in Section 4.2.6.

4.2.1 Gesture forms

In the development of non-verbal communication in virtual humans and holographic AIs, many technologies capable of generating a series of high-fidelity animations have been demonstrated in the research field. However, although some researches have claimed that the gestures exhibited by their holographic AIs are meaning in interaction (Rebol, Güti and Pietroszek, 2021), these do not entail reusable and reliable gestures characteristic of human movement in speech. In considering gesture generation and the gestural relationship in terms of spatial information, this section details the concept, taxonomy, and framework of gestures.

A gesture can be defined as 'visible action' in an utterance (Kendon, 2004), whereas an action occurs, for example, when a person moves an object. By altering the amplitude of the arm and hand, a gesture describes an object's properties (Novack et al., 2014). Therefore, it is defined as a visible/schematized language that implies an intent via particular movements of the hands and fingers.

In terms of gesture taxonomy, Abner, Cooperrider and Goldin-Meadow (2015) summarize two approaches. The first is based on an articulator for generating hand movements, such as those used in sign language. During conversation, it integrates other body parts together with the production of speech sounds. Gestures, for instance, are often used to indicate agreement or disagreement in conjunction with head movements. A second way is to use a gestural function in conversation which is interactive and representational.

In the interactive type, gestures are used to manage speech or dialogue between listeners and speakers (Kendo, 1995). Attitudes, emotions, or ideas are conveyed through them, but the interaction gesture is not associated with communication topics (Bavelas et al., 1992). Listeners give feedback or corresponding gestures, such as handshakes, high-fives, and invitations, to the speaker.

A representational gesture, on the other hand, conveys information about an object, activity, or environment. For example, people use their hands and fingers to indicate an object’s shape or size. Table 4.1 lists the types of representational, namely deictic gestures, iconic gestures, metaphorical gestures, and emblems (Bernard, Millman and Mittal, 2015; Abner, Cooperrider and Goldin-Meadow, 2015).

Code	Description	Example
Interactive	Managing speech, maintaining the social system	Greeting gesture, high-five
Deictic	Pointing at a location, position, time, and object	Index finger points at a direction.
Iconic	Depicting a physical object’s property	A vase’s shape: open hands and perform curl.
Metaphoric	Describing metaphor concepts	‘Pointing upwards gesture’ means increase.
Emblematic	Widely common, ‘cultural’ gestures that may lose their original meaning	‘Ok’ gesture and silence hand gestures
Beat	Pragmatic gestures that emphasise the flow of speech	Hands/ index finger goes up and down in speech
Cohesive	As repeated gesture, it can be used to cohere with the interrupted narrative by creating the same gesture in the original position	The same gesture is created again in the same position for resuming the utterance, e.g. using two finger quotation marks to mark the beginning of a quote and the end.

Table 4.1. General gesture types

Deictic gestures involve pointing at objects, directions, people, or abstract concepts such as time (Krauss, Chen and Gottesman, 2000). This type of gesture uses spatial expressions that can draw the user's attention by replacing spoken determines, i.e. 'this', 'that', 'there', or 'here'. A typical deictic gesture is a pointing gesture. Forefinger pointing is the most common and habitual gesture used when pointing at a referent (Navas Medrano, Pfeiffer and Kray, 2020). This type of gesture has been divided into three branches by Hassemer and McCleary (2017): pointing at objects, pointing at locations, and pointing at directions. The deictic gesture relies on whether the interlocutor shares space with the demonstrator. With the pointing gesture, the speaker can map out places in his/her mind and assigns these different spaces to the place to which he/she intends to refer. The place and object are taken as referents by the speaker when conversing with others, and these referents are processed by the way of spatial information. The pointing gesture has an abstract meaning in this context, despite the fact that the place pointed out is concrete. As it depends on context interpretation, and the referent may not be physical (Gunter, Weinbrenner and Holle, 2015).

The iconic gesture can depict a concrete matter, representing its properties. As an example, people describe running postures by swinging their elbows and arms. In the context of conversation, it is useful for indicating relevant features by gestures. Kinesics features of the iconic gesture are associated with semantic features (Kopp, Tepper and Cassell, 2004). For instance, a person might say "water velocity is fast" non-verbally by simultaneously waving their arms and hands. The iconic gesture expresses a concept of the referent, while its Kinesics marker corresponds to its physical characteristics (ibid).

The metaphorical gesture is used to describe concepts with a figurative, symbolic expression (Bernard, Millman and Mittal, 2015). Metaphors are words that are used not with their literal meaning, such as "night owl" or "spilling the beans". For example, the pushing forward movement of the hand gestures "opening the heart". This pushing gesture is a metaphorical way of expressing one's feelings. Furthermore, such a gesture can also describe a specific orientation of time. The left hand could indicate the past, while the right hand indicates the future (Lhommet and Marsella, 2016). Therefore, metaphorical gestures also can be used to concretize abstract concepts (Kircher et al., 2009; Lhommet and Marsella, 2016). However, the concept of the metaphorical gesture is questionable since the taxonomy of gestures should be based on form and semantic meaning with speech and motor representation. Krauss, Chen and Gottesman (2000) considered the metaphorical gesture as an iconic type. A metaphor utterance is difficult to understand if the speaker only performs hand movements. Further, the classification of the gesture cannot rely solely on a signal representation. The metaphorical gesture should visualize potential features of metaphor expression, but it cannot be defined as iconic since an iconic gesture can directly describe an object's physical features.

The emblematic gesture represents a specific and widely understood concept, such as the upward extension of the thumb meaning 'good', and the 'hand-to-ear' meaning that the interlocutor is talking loudly.

McNeil (1995) classified two more types of gestures: beats and cohesive. The former is a pragmatic gesture that is used to maintain a conversation. As defined by Prieto et al. (2018), this hand performance is a 'non-referential hand gesture'.

The latter markers a referent's location and tracks the referent in a discourse (McNeil, 2005). When the same matter is mentioned again in the later conversation, it can present in the same location (Sekine and Kita, 2013). For example, the listener interrupts an utterance, and the speaker retains the original gesture. When he or she resumes the suspended story, the same gesture in the same position is created. In this way, the speaker is able to bring the narrative together.

In a closer examination of the gesture generation process, Kendo et al. (1980) proposed three steps: preparation, stroke, and retraction. The first step is to prepare the hand movement. The gesture then performs in its active phase. In the retraction, the gesture reverts to a previous direction or prepares the next gesture. The hand movement can hold for a while in the gestural phase, which is the pre-stroke posture (Kita, van Gijn and van der Hulst, 1998).

To investigate how gesture generation affects the cognitive process, Kita, Alibali and Chu (2017) proposed a theoretical framework of the representational gesture, with four elements: gesture activation, manipulation, packaging, and exploration of spatio-motoric information (see Table 4.2).

In terms of activation, gestures can trigger new motor commands or preserve existing spatio-motoric representations. It has been found that people gesture more when they describe paintings from memory than when they describe them from sight (Wesp et al., 2001). People's gesturing relies on the pre-exiting information which they have observed. In this way, the gesture can prevent pre-existing spatio-motoric representations from fading from memory. However, the mind constantly updates and adapts to changes in the surroundings and body, making it difficult to recall pre-existing spatial information and maintain its activation. This means that gestures can help activate, keep active, and even modify new spatio-motorically grounded representations. Speech content, for instance, can be visualized and changed using gestures, allowing for more spatio-motor descriptions. In addition, such a gesture also can prompt abstract and metaphoric concepts by activating spatio-motoric representation (Kita, Alibali and Chu, 2017).

During a conversation, information is manipulated in order to observe the referent from a variety of perspectives (Kita, Alibali and Chu, 2017). For example, people simulate rotation and place objects in the right holes, using co-thought gestures. Following this, they manipulate the object in spatio-motor representation by gestures that relate to the task. It is critical to note that the gesture does not directly map or offload the referent's property; instead, sensory-motor information is continuously transferred to the brain, which then effectively generates movement (ibid). People may produce gestures that differ from speech content and extract information from the perceptual dimensions of the referent (Alibali and Young, 2010).

Furthermore, complex meaning, when represented in utterances, can at times be overwhelming. Chunking gestures provide remedy here. Kita, Alibali and Chu (2017) explained that the packaging system of our brain can separate chunk information into units using spatio-motoric hints, which in turn allow people to gesturally recode information along chunks. For instance, when describing motion event, a person gestures 'rolling' and then 'down' to convey "rolling down the hill" (Mol and Kita, 2012).

In considering which information is useful or can be optimized for the task at hand, people evaluate relevant information and affordances. Failed and abandoned gestures are part of the exploration: such gestures cannot match utterances changes in a person's thinking (Kita, Alibali and Chu, 2017). For example, if the listener misunderstands, the speaker may generate a new form of gesture to re-organize spatio-motoric information.

Speech and gestures have different representational modalities but complement each other (Abner, Cooperrider and Goldin-Meadow, 2015). For example, gestures can encode visuo-spatial information that speech cannot. As part of an entity, speech is a discrete digital unit, whereas gesturing is a continuous form in the analogue world that can provide additional language information (Fuks, 2014).

The gesture can be used to eliminate speech misunderstandings or to convey the same concepts as a speech.

In previous studies, gestures are correlated with time and duration, spatial movement and information, and speech. The gesture is a phase of movement from preparation to retraction. The co-speech gesture is temporally aligned with time, serving as “a prosodic structure of language” (Abner, Cooperrider and Goldin-Meadow, 2015). The gesture's taxonomy is defined by its function. Different gestures can convey the speaker's key points or directly some actual meaning. A gesture can be used to explain ways of thinking, and how people process information into a spatial representation.

It presents a gestural taxonomy as summary of the above in Table 4.1, and the framework of cognitive process that affect gesture generation is presented in Table 4.2.

Code	**Description and features**	**Examples**
Activation	Gestures activate or maintain a spatio-motoric representation; communicative success depends on whether listener can select the intended pre-existing information from memory.	People may extract pre-existing information from the past to maintain activation or activate new spatio-motoric information that is not yet active.
Manipulation	The spatio-motoric information can be rearranged by the gesture to facilitate penetrative thinking for simulating the referent.	People need to gesturally rotate a Rubik's Cube
Packaging	Gestures can help separate information into chunks.	When people describe similar referents with different shapes, their gestures need to package this spatio-motoric information (e.g. counting fingers for different aspects: one, two, etc.)
Exploration	Gestures explore whether information was received information. It can be a trial-and-error process.	When a listener cannot understand t speaker’s speech or gestures, the speaker will give up the present performance, and try gestures again, and then seek to provide additional useful information.

Table 4.2. Types of gestures used for influencing thinking

4.2.2 Influence of gestures in pedagogy

This review presents what types of gestures used in a pedagogical setting and their influence in the students' learning outcomes.

Novack and Goldin-Meadow (2017) developed a gestural framework that explains how gestures represent functions, even though the meanings of the gestures are abstracted. They argued that gestures and actions serve their own aims, leading to different learning achievement. The authors conducted an experiment, in which they asked three groups of children to calculate mathematical equivalence questions, such as 3+7+2 = ___+ 7, using hand gestures or actions (Novack et al., 2014). The experimenter guides children in using gestures or actions. The first group picked up tiles with numbers, and the second pretended to move the tiles. The last group of children pointed at the two numbers using fingers forming a V-shape. It is claimed that only children who use the deictic gesture can deal with similar problems, because this gesture promotes penetrative thinking and directs attention (Atit, Gagnier and Shipley, 2015), and that the important information in relation to numbers can be packaged by the gesture. Here, gestures can enhance the learner's cognitive abilities and reduce distractions for students.

Matsumoto and Dobs (2017) investigated the influence of gestures in regard to learning a second language. They observed students simulating teachers' gestures, and teachers using gestures to explain temporal concepts. According to the authors, in grammar lessons, teachers use the metaphoric and abstract deictic gestures to compare past and future tenses, these gestures match the English metaphorization of time, and the abstract deictic gestures can be used to explain grammatical features. As an example, in the authors' experiment a teacher generates a metaphorical gesture to represent a time range by making a circular and continuous action, and then the deictic gesture points down to explain the present time by pointing at different positions of the relevant space. The gesture also represents transcription symbols, and it can compare past and future tenses. In this investigation that conducted by Matsumoto and Dobs (2017), students focused on and imitates the teacher's gestures to assimilate knowledge and enhance understanding. Therefore, the teachers' gestures are an essential tool that is able to adjust students' way of thinking and their expression. The gesture also serves to establish a visual image in communication, and repetitive movement can serve to highlight focus.

However, teachers' gestures alone cannot stimulate students into learning more knowledge. For example, Yeo et al. (2017) implemented four lessons in which teachers perform/did not perform gestures to teach equations and graphs. Afterwards, the student's learning outcomes were measured on the basis of whether they can successfully answer questions concerning slope and intercept. It was observed that the teacher's gestures alone did not enable students to achieve higher scores, in contrast to the verbal lessons. There exists the risk of hand gesture functionality being diminished where there is excessive overlap between gestures and speech (Hostetter, 2011).

Children's learning does benefit immediately from hand gestures (Cook, Mitchell and Goldin-Meadow, 2008). In the experimented by Yeo et al., the children were later asked to calculate the equivalence mathematical questions again, whilst the teachers represented handling ways by sweeping gestures from the left to the right of equations. The students mimicked the teacher’s behaviour, i.e. speech and gesture. Then the students complete a follow-up test four weeks after. This experiment is designed to measure the extent to which the students retained learned abilities when receiving gestures, speech, or both with co-speech gestures. The authors found the group receiving gestures and co-speech gestures retained the most knowledge.

In general, however, the teacher's gestures can guide students, and it is more of an indicator for securing the students' attention than a teaching method. To improve cognitive ability and interactional skills in speaking, students can simulate the gestures of their teachers to improve grammar knowledge. Although a relevant gesture is not always available, such gesturing can improve learning outcomes over time. Nevertheless, if gestures lack a prosodic structure, such as pausing, and do not align with time meaningfully, they are ineffective (Abner, Cooperrider and Goldin-Meadow, 2015).

Table 4.3 summarizes that the metaphoric and deictic gestures that are common in teaching hand movements, and that point out key content as well as explain abstract concepts using metaphors.

Type	**Description**	**Example**
Deictic	V-pointing gestures Sweeping, i.e. pointing one side to the other side	To focus on a specific referent
Abstract deictic	Representing time period	Refer to time
Metaphoric	Representing time frame	Time zones are represented by gestures drawn in circles.

Table 4.3 Gestures in pedagogy

4.2.3 Gestures in virtual humans

The purpose of this section is to review prior research on the factors of gestures that influence human-computer and virtual agent interactions.

Ferstl et al. (2021) analysed the relationship between the realism of a virtual agent’s persona (i.e. voice, behaviour, and appearance) and perceived co-speech gestures. They used motion capture to obtain animations, and adjusted gesture animations for the purpose of comparison. Examples in their study include a virtual human and robot performing natural movements with gestures and legs motions, synthetic movements with an idle static standing pose and natural gestures, and a scenario in which the robot’s gesture stroke has been reduced. The authors observed that perceived match between motion and voice has an impact on the character’s realism, especially gestures, while appearance and gestures have no mutual effect. Therefore, non-verbal interaction and voice drive realism since natural animation aligns gesture with the voice, whereas in the reduced condition the gesture disappears when stressed syllables appear.

Davis et al. (2021) investigated the influence of gesture frequency on the student's satisfaction and learning outcomes by using a virtual human which teaches on the causes of lightning. The teacher's gesturing consists of iconic, metaphoric, deictic, and beat representations based on a co-speech prosodic structure that is a relationship between gestural types and phases. The experiment is designed to measure the frequency of gestures and the video speed of a virtual human. The enhanced condition involves the agent performing 20 gestures for every 100 words, as opposed to 14 gestures for every 100 words in the control condition. The authors found that gestures with the enhanced or average frequency of use are associated with higher recall, satisfaction, and generalizability. The use of gestural animation can enhance users' visual stimuli and spatial awareness, which can improve memory and learning performance.

Nirme et al. (2020) recently designed a presentation performance including facial capture and audio to accurately control a single target stroke that synchronized with the appropriate stressed syllable of words. In their setup, through a process from gesture generation to gesture duration, the stressed syllable should appear. The experiment uses a marker-based optical motion capture system for capturing speakers' behaviour and generating animations, before selecting standard segments. In accordance with the three steps of gestures (i.e. preparation, stroke, and retraction), the animation is coded, and the gestures are analysed in the stroke stage. The study compares the original synchrony between speech and animations, with that in which the gestural stroke is advanced by 500 milliseconds, and that in which the gesture is delayed by 500 milliseconds. The authors reported that the pointing gesture appearing before or after the stressed words did not produce an unnatural situation, although speech–gesture asynchrony can lead to unnatural performance. If the speech is paused, the ongoing gestures should also stop or freeze (Graziano and Gullberg, 2018). Moreover, Nirme et al. (2020) argue that gestures that overlap with pauses seem unrealistic, while gestures that overlap with spoken content have less of a negative effect. However, neither of these studies provides detailed descriptions of the selected gestures.

A virtual human endowed with gestural animation can also improve the learner's memory by pointing to a specific element (Craig et al., 2015), as evidenced by the results reported by Novack and Goldin-Meadow (2017) and Cook et al. (2016). These studies compare and contrast three conditions: general gestures, specific hand movements, and non-performance. Craig et al. (2015) conducted an experiment, where a virtual human explains the formation of lightning, and then the participants complete retention tests. It was observed that because learners select and store relevant information by gesturing in multiple ways, gestures improve outcomes when visual content is spatially close to hand performance.

It has been demonstrated in previous studies that the hand movements serve as an additional avenue for the transmission of information. Hand gestures can be used to stress syllables, frequency, and consistency with speech. Although some scholars have advocated using the pointing gesture to improve retention and outcomes, it is known that teaching gestures should not be based solely on the movements of one hand during interaction with students. In addition, many simulations proposed and tested in previous studies tend to feature gestures that rely either on video displays or predictable questions that students can guess the answers to, such as calculations. Such a design cannot empirically realise key gestures that can facilitate holographic AI's development.

4.2.4 The way of coding gestures

In order to analyse data of the gesture, this section presents three pieces of related work:

Cook, Yip and Goldin-Meadow (2010) proposed a means of coding gestures when investigating whether co-speech gestures can facilitate memory. They provided a series of short videos to participants, including movements of animals, people, and objects. Then participants were instructed to recall and describe the contents of the videos. The information concerning gesture and utterance was transcribed and coded; in detail, the data included pauses, and non-functional gestures such as scratching and beat gestures. In discussing the results, the authors argued that if a participant can correctly describe motion and semantic information, which particular fragment is recalled; whereas, if the description does not correctly present the event, the recall is unsuccessful. In their study, the second coder independently coded the transcription and gesture to measure reliability.

Similarly, Atit et al. (2013) coded spatial gestures in order to explore iconic and deictic types in a speech concerning details on a geological map. Each gesture and accompanying dialogue were coded to evaluated whether the described activity referred to an object's property, such as size and direction. Atit et al. categorised the gestures concerning spatial relationships, e.g. pointing at a position, an index finger in space indicating 2D information by a line, a hand palm inferring a plane, a hand gesturing a 3D shape, an event being described using hand gestures, etc. The authors divided gestures into iconic, deictic, mixed type (with both iconic and deictic gestures), and other gestures, in their analysis of the proportion of spatial gestures. The authors claimed that mixed type gestures tend to indicate complex spatial information.

Stefanini et al., (2009) evaluated changes of speech and gesture in a simple naming task, in which children named graphics displaying objects, movements, and features. The communicative parts were coded, and the authors observed the children's ways of utterance, accuracy of answers, types of gestures, in an effort to determine the relationship between correct description and gestures. The authors' gestural coding is based on a classification that consists of deictic, representational, and other kinds of gestures.

These three studies provide approaches to characterize the data of gestures as a basic process, but the proposed categorisation of gestures as well as the research methodology is not directly applicable to the study of instructional, teaching gestures. Therefore, this study will develop its own, novel methodology to investigate the taxonomy of gestures in teaching that is illustrated in Section 4.3.

4.2.5 Research gap and questions

The teacher's behaviours can have a positive impact on the student's learning performance and achievement (Wang, 2020) since their gestures are sometimes used as a primary tool for instruction or. more often, appear alongside of speech (Clough and Duff, 2020). In terms of an intelligent tutoring system, the intelligent virtual tutor simulates the human voice as well as behaviours, such as standing, hand gestures, sitting, or talking, thereby aiming to engage trainees in learning. Immersive and augmented intelligent tutoring systems. However, these systems often focus on the teaching content, while neglecting more subliminal gesticulation.

Virtual teachers mostly apply one or two types of gestures, prominently deixis, i.e. pointing to objects. For example, FaceMe provides a virtual partner, who guides children to identify and perform different emotions (Li et al., 2021); however, the virtual agent only performs waving prior to interaction, and limited hand gestures are utilized for interaction. Barry, the virtual employee created by Talespin, helps users practice the action of terminating the employee's contract with immediate effect (e.g. 'firing' Barry, or 'laying off' during downsizing) (Katz, 2019). Barry's deictic and beat gestures (e.g. knocking the table and folding arms) reflect his emotions, although hand gestures are absent when he is sitting on a chair.

Few studies explain in detail the animations they employ for virtual teachers (Zielke et al., 2018). Early research studies promoted the ideal that the realism of an agent can be improved using deictics to direct spatial orientation (Lester et al., 1999). There remains the need, however, for a comprehensive investigation designed to help the researcher identify a wider, more generic set of gestures and body language for virtual teachers during virtual tutoring. Although virtual humans are in principle capable of mimicking human mannerisms, few studies have investigated the degree to which different gestures help express ideas, provide guidance, as well as support motivation.

Therefore, in view of the paucity of studies exploring the role of animated gestures, this chapter reports the findings of a novel experiment designed to identify and explore instructional gestures, with the aim of extracting an improved set of teaching gestures, which in turn can be used to animate virtual humans and facilitate the instruction of learners using AR.

4.2.6 A plan to fill this gap

The overarching purpose of this experiment is to ensure that only the virtual teacher can observe questions and answers without being influenced by the student, in order to determine exactly how the teacher can guide the student using gestures. The objectives are to identify key gestures, as well as determine whether the trainee can understand and follow the trainer's explanations of, and instructions for completing tasks. In order to ensure sufficient collection of data concerning gestural animation, each task has been conducted four times. Barrett, Foundas and Heilman (2005) claimed that gestures and speech are independent, which the former is an auxiliary function that becomes prominent whenever the content is tough to describe in words alone. Therefore, this study will compare concerning of co-speech gestures and non-speech gestures to observe whether the trainee can also understand gestures if the trainer cannot speak. More details of the experiment and analysis will be presented in the following sections.

This experiment employs the Rokoko Motion capture suit and Rokoko smart gloves to generate corresponding gestural animations and fingers' movement.

Fragments that the trainee misunderstands and that lead to failed guidance have not been discarded, since it is important to explore how the trainer re-manipulates spatial information. The corresponding animations are selected and coded based on the steps of the gesture generation (preparation, stroke, and retraction) and the taxonomy of representational gestures (i.e. deictic gestures, iconic gestures, metaphorical gestures, and emblems) to develop the holographic AI's instructional gestures. The experiment also analyses the trainer's ways of thinking and how he/she uses body languages to describe different tasks, according to the gesture taxonomy presented above (gesture activation, manipulation, packaging, and exploration of spatio-motoric information), which will support the generation of corresponding animations for the holographic AI.

In this way, the most frequently used gestures were selected and analysed, and then the teacher's motivation can be inferred, and resulting cognitive behaviour of the students can be validated by observing their progress in the tasks. Finally, this results in producing a comprehensive set of gestures for the holographic AI's.

4.3 Methodology

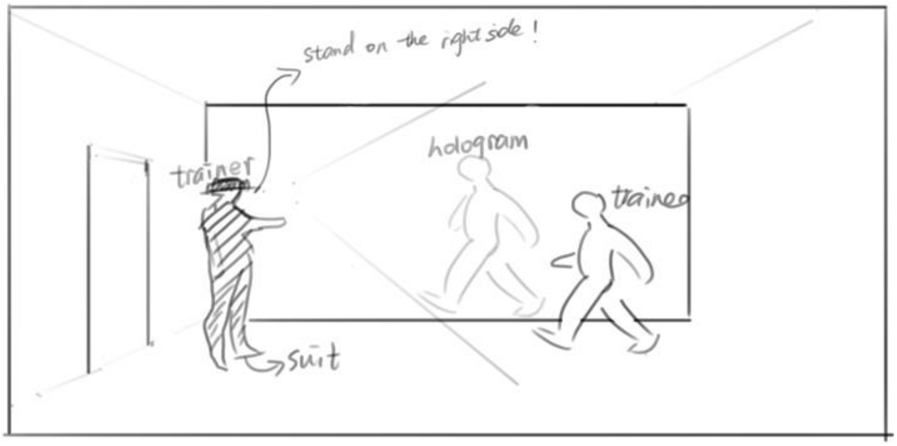


Figure 4.1. Navigation

This section explains the methodology devised for this study. Guided by the project's goal of developing trainers with more natural instructional gestures, this experiment tasked the participants with three tasks: one focused on navigation, one on assembly, and a precision task. To be more specific, the gesture has a navigation function that can translate messages from "exhibiting movement" to "communicating it to other persons", so that the listener can understand contents that the speaker does not verbally describe. The assembly stage is a collaborative and cognitive task, in which the trainer guides the trainee to build specific objects. The gesture generation is reliant on the way in which the trainer organizes matter. The precision task focuses on how the trainer can accurately use co-speech gestures to instruct the trainee.

It firstly applies motion capture technology to record and generate animation. The Rokoko motion capture suit has 19 sensors that record movement/orientation/position data of individual joints/limbs. In order to determine the relationship between speech and instructional gestures, each task includes a 'warm-up' trial, during which participants were not permitted to verbally communicate with the trainee, but during which the trainee could ask questions.

In the navigation task (see Figure 4.1), the trainer wears a HoloLens and the Rokoko Motion Capture Suit. Using the HoloLens 2, the participant acting as trainer is able to see a hologram of a human (in a specific location and pose in the room). The task for the trainer is to navigate the trainee using gestures and words so that the latter stand right where the hologram is, mimicking the hologram's body and limb positions as closely as possible (e.g. hologram stands facing towards the door with left arm extended to the ceiling). When the participant acting as the trainer is satisfied, the task is finished, and a picture showing the hologram and participant acting as trainee using the Microsoft HoloLens 2 MR Remote Capture is then taken. (The photograph supplies additional evidence of the task being completed successfully.)

In the second task, the trainee has to assemble a cardboard fort (see Figure 4.2). The trainer again dons the Rokoko Suit for data collection, and the HoloLens 2 shows a hologram of the cardboard fort. The task for the trainer is to instruct the trainee to assemble the cardboard fort from its pieces (cardboard walls and Velcro straps). The fort is composed of a specific shape and dimensions, which are inspected upon completion of the task. When the participant acting as trainer is satisfied, a picture is taken using the HoloLens 2 remote capture device, enabling the research to assess whether the task was completed successfully (and with what accuracy).

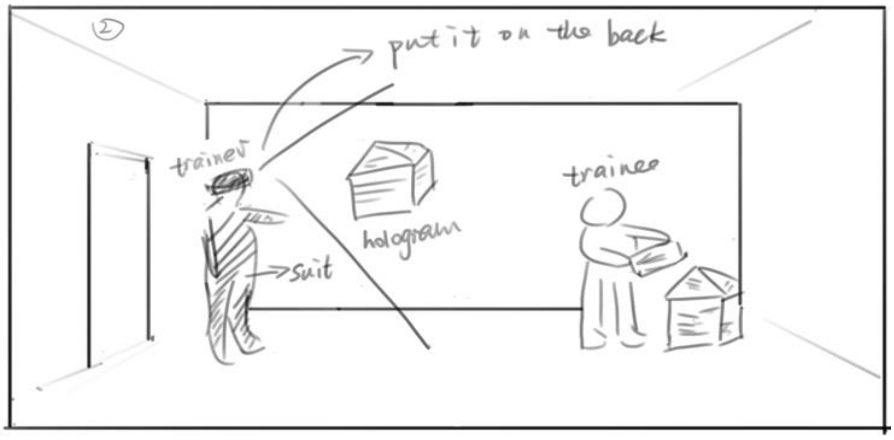


Figure 4.2. Assemble task

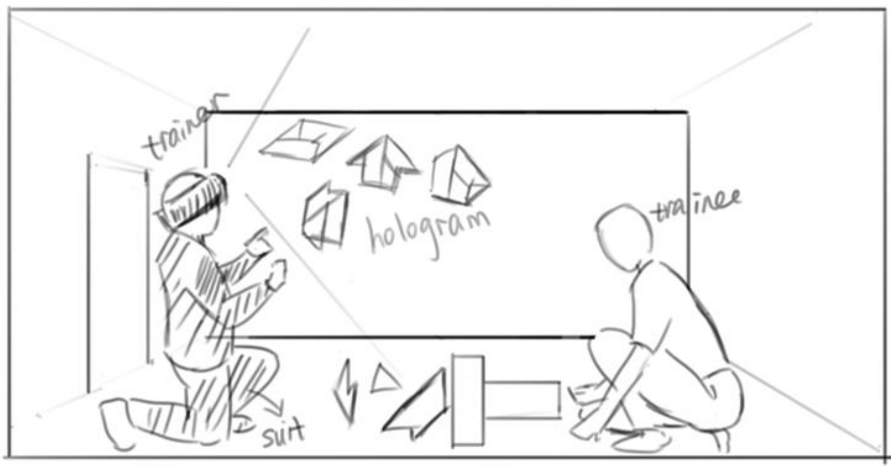


Figure 4.3 Precision task

In the precision task (see Figure 4.3), the participant acting as trainer wears the suit and HoloLens, and instructs the trainee on assembling the correct patterns in ‘Puzzle-T’ (also known as ‘Tangram’). Again, the outcome is checked with the help of a MR capture photo showing the result.

Each participant needs to finish each of the three tasks four times. During the course of task repetition, the object (the human hologram, cardboard fort shape, and puzzle-T pattern) is changed in order to determine which gestures are useful and worth repeating. During the first instance of each task, the trainer cannot verbally communicate with the trainee (i.e. warm-up trial). Once all sessions are completed, the recorded footage is used to analyse all raw data of animations. The aim of this analysis is to identify key movements that occur frequently in the trainer’s gestures.

4.4 Experiment

This experiment adopted a within-design, and the participants experienced all conditions: speechless trials in each task (i.e. warm-up), and co-speech trials.

All holograms were created using Maya. The models in the navigation task were downloaded from Mixamo (Mixamo, 2023). In order to ensure all models were static and had no animations, skeleton and key frames were deleted. The postures include standing, kneeing, and sitting (see Figure 4.4). Besides, the trainer could not verbally speak in each warm-up trial. The difficulty levels of the other subtasks are dictated by the position and direction of the limbs and trunk. For example, Task1_medium in Figure 4.4 requires the left side leg to hold the centre-of- gravity position, with the hip protruding slightly to the left, but Task1_easy only requires protruding slightly to the left-side hand pointing on the left side.

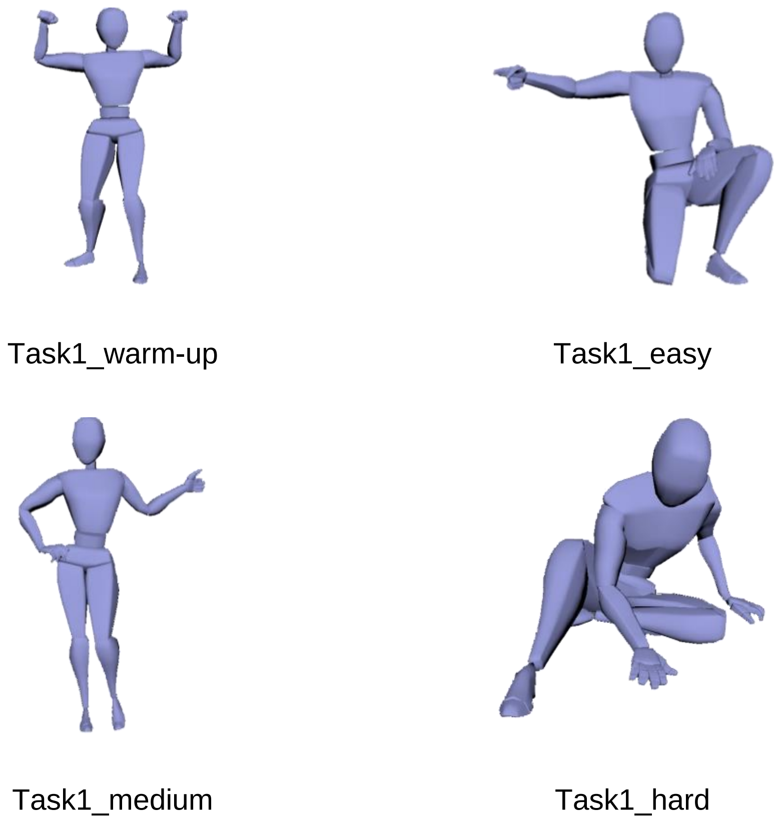


Figure 4.4. The navigation task

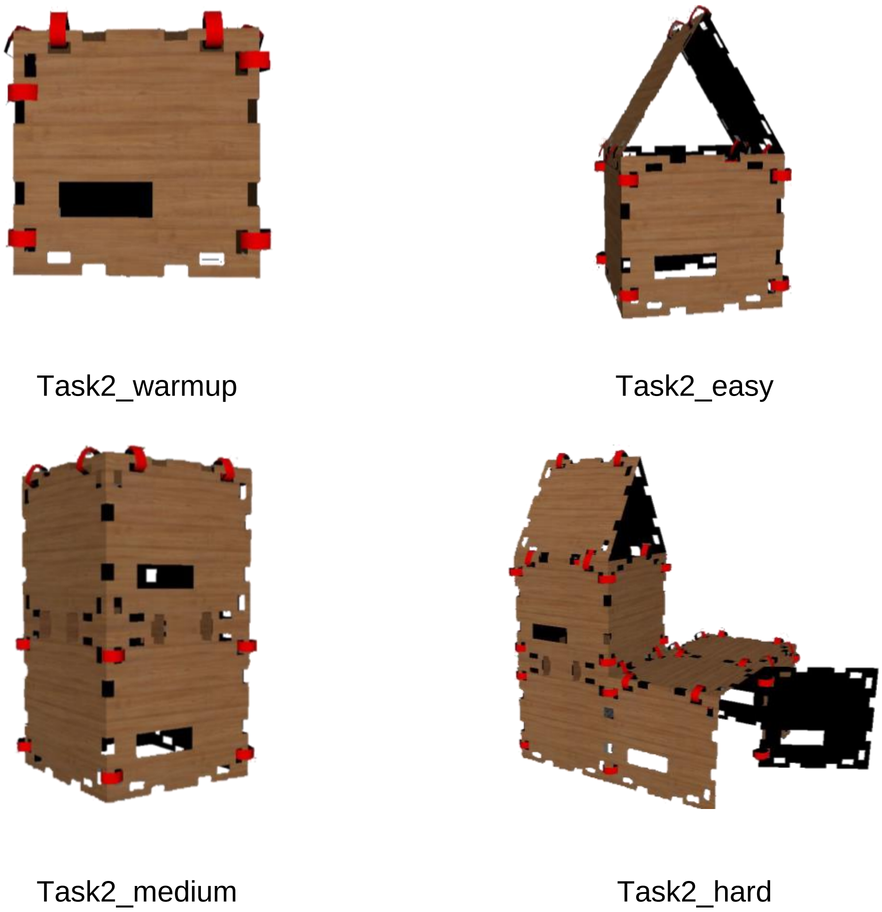


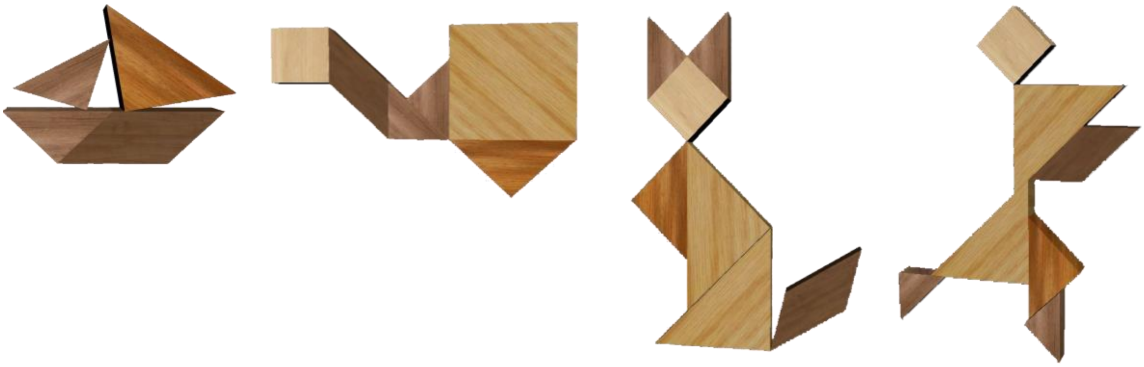
Figure 4.5. The assembly task

Figure 4.5 illustrates the assembly task. Based on real cardboard forts, the virtual house includes 12 panels and 4 boards with windows. As shown in Figure 4.5, the medium and hard levels require the use of Velcro and connectors (see Figure 4.6) for building the first floor. During the speechless trial the trainee is required to assemble a basic cube, whereas the assembling hard level requires all cardboard panels and connectors.

Figure 4.6. The connector for building the first floor

The final task uses three shapes, consisting of two small triangles (25cm * 12.5cm), two large triangles (50cm * 25cm), one medium-sized triangle (35cm* 17.5cm), a rhomboid (37.5cm * 17.5 cm), and a square (17.5cm * 17.5cm). The warm-up trial requires the trainee to set up a ship, a tobacco pipe shape (displayed for Task3_easy in Figure 4.7). The medium level task is assembling a cat (or fox) shape, and the hard level task is assembling a person-like shape. Only the trainer can see the hologram, while the trainee does not know what shapes he/she needs to build. All holograms were uploaded to HoloLens, and then displayed by 3D viewer.

Only the trainer can see the hologram, while the trainee does not know what shapes he/she needs to build.



Task3_warmup Task3_easy Task3_medium Task3_hard

Figure 4.7. The precision task

4.4.1 Material

This experiment uses motion capture technology, i.e. the Rokoko motion capture suit, to generate high-frequency joint and limb movement/orientation/position data (~100Hz for 19 sensors). Following completion of the experiment, the motion capture was evaluated in terms of context-specific matching of video footage (third person view) with the audio. Basic demographic data of the participants were collected (age bracket, gender), and measurements of limb lengths were used to configure the motion capture.

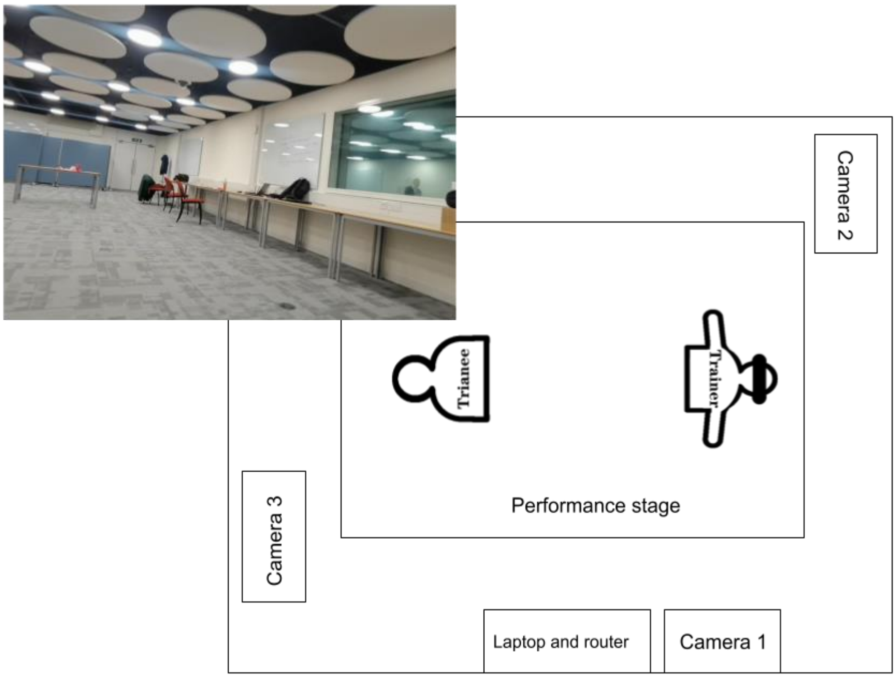


Figure 4.8. Space area for motion capture

4.4.1.1 Environment setup

The experiment was conducted in four different areas, including a drama studio and three different meeting rooms. The rooms range in size from 50 to 120 square meters, and they are arranged like square stages with cameras set up on each side. The three cameras recorded the participant's behaviours from different angles (see Figure 4.8).

This experiment used a HoloLens, medium and large-sized Rokoko suits, Rokoko smart gloves, two power banks for connecting the Rokoko devices, a laptop, two phones, and an iPad. Since the Rokoko devices cannot use public Wi-Fi, an additional Wi-Fi router was applied to connect to the Ethernet.

4.4.2 Procedure and participants

The experiment was conducted with 22 participants (4 females, 18 males) aged between 18-64. The trainer group had 3 females and 8 males.

When the participants arrived, the experimenter briefly described the aim of the study and protocol to the participants, and asked them to choose a role of either trainer or trainee. Then, both participants were asked to read the information sheet, sign the consent form, and fill out a demographic questionnaire regarding gender and age.

The Rokoko suit’s sensors label the key joints, thus in order to match the position of each key joint with each sensor the suit size must be fit the trainer. For this reason, the trainer's height, arm length, feet length, shoulder width, knee height, pelvis height and width, and manus length were recorded. These personal data were not collected and analysed, but were deleted. It only served the configuration of the motion capture and animation. Next, the experimenter set up the virtual holograms in position, and the trainer was asked to put on the HoloLens.

When the participant was ready, the trainer’s movement was recorded. After each task was finished, the trainer removed the HoloLens, and the experimenter prepared the next hologram.

4.4.2.1 Interaction scenario

The experiment consists of 12 trials (4 trials*3 tasks) in total, and the steps of interactive scenarios are as follow below:

- Step 1: The experimenter explains what the trainer has to do in the non-verbal (warm-up) and verbal (navigation) trails.
- Step 2: The warm-up trial of the navigation task: the trainer cannot verbally guide the trainee to stand near the hologram for the first time, but the trainee can ask questions.
- Step 3: The navigation task (second-fourth times): the trainer can verbally guide the trainee.
- Step 4: The experimenter instructs the trainee on assembling the real cardboard forts using non-verbal (warm-up) or verbal cues.
- Step 5: Assembly task warm-up: the trainer cannot talk with the trainee.

- Step 6: The assembly task (second-fourth times): the trainer can verbally guide the trainee.
- Step 7: The experimenter instructs the trainee on assembling the real cardboard forts using non-verbal (warm-up) or verbal cues.
- Step 8: Assembly task warm-up: the trainer cannot talk with the trainee.
- Step 9: The assembly task (second-fourth times): the trainer can verbally guide the trainee.
- Step 10: The experimenter instructs the trainee on assembling the real puzzles and their shapes prior to the task.
- Step 11: Precision task warm-up: the trainer cannot talk with the trainee.
- Step 12: The precision task (second-fourth times): the trainer can verbally guide the trainee.

One requirement of the navigation task on the part of the trainer was that he/she neither sat on the floor nor fully mimicked the hologram posture. If the trainer directly simulates the posture, none of the key gestures can be achieved. The second task requires that the panels with windows should be facing the right direction (always at the bottom). The trainer needs to select the correct size of triangles in the precision task. The experiment has no time limitations; rather, the session ends when the trainer agrees that the task has been done.

4.4.3 Measures and hypotheses

The pre-test was conducted prior to the experiment in order to assess the operability of the three tasks as well as the condition of the room in the drama studio, as magnetism affects the quality of Rokoko animations. This test took 146 minutes (from preparation to completion), and the three tasks took around one hour and 6 minutes. The trainer waited for the trainee to finish executing the previous cardboard assembly commands before providing instructions for the next step; however, this occupied too much time, especially with task2_hard (half an hour). Whenever the trainee made a mistake, the trainer would employ a ‘other way’ gesture to direct the trainee. If the trainer considered the cardboard fort to be stable, there was no need for the trainee to tape all Velcro.

In total, the experiment entailed the collection of (11 groups * 4 cameras * 3 tasks) videos recorded by HoloLens, two mobiles, and a iPad. The HoloLens’s video recording was used to deliver the trainer’s voice and, subsequently, generate the transcripts. For the different trials and tasks, the trainer's gesture that can be observed in clear resolution in videos were selected for analysis: a total of132 videos (11 groups * 12 trials) was obtained.

The qualitative software package, NVivo, was used to organise, store, and analyse data which were coded in terms of key segments and words. Transcripts of the trainer’s dialogue were also imported in order to observe the dialogue content and determine whether the speech and gestures are relevant.

4.4.3.1 Standards for segments and words selection

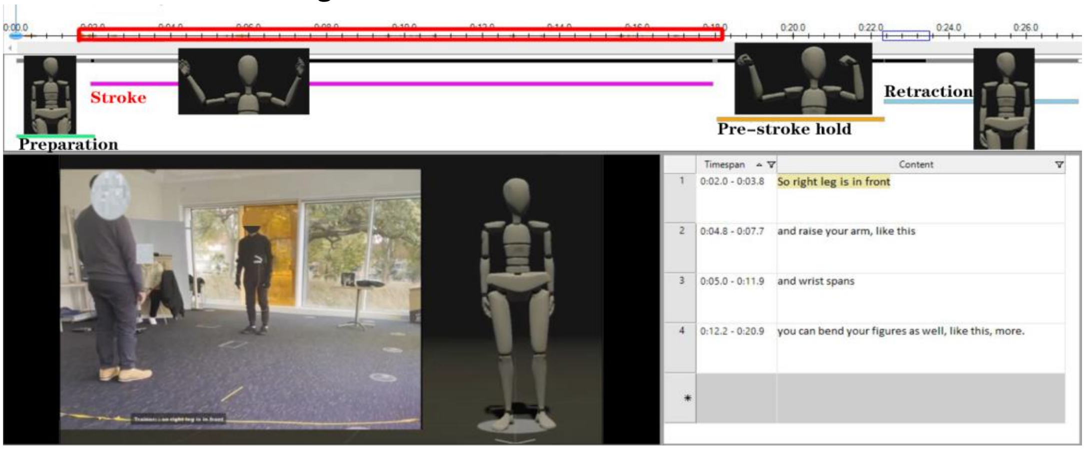


Figure 4.9. The step of gesture generation

According to the steps of gesture generation proposed by Kendo et al. (1980) and Kita, van Gijn and van der Hulst (1998), the steps of a stroke or pre-stroke hold consist of segments of key instructional gestures. For example, during the preparation step the trainer stands before the group and issues a preparing gesture for the navigation task, and then the trainer starts to generate gestures, keeping the hands raised, before deflecting them upon indicating the pre-stroke hold, and finally putting down the arms at retraction (see Figure 4.9). The fragment of the preparation and retraction steps were discarded. Although there were no meaningful gestures during the pre-stroke hold and retraction, it is necessary to review how the trainer jumps to the next step through these gestures, in order to avoid missing gestures. For example, when the trainer instructs the trainee on building the cardboard fort, he/she points at a particular position, and waits for the trainee to complete, before issuing the next step (see Figure 4.10). The gesture in the pre-stroke hold is not instructional; rather, it implies that the trainer is preparing to activate new spatio-motoric information via gesture generation.

3:49 - 3:51	3:52 - 4:08	4:09 - 5:30	5:31 - 5:42
Stroke	Pre-stroke hold	Retraction	Stroke

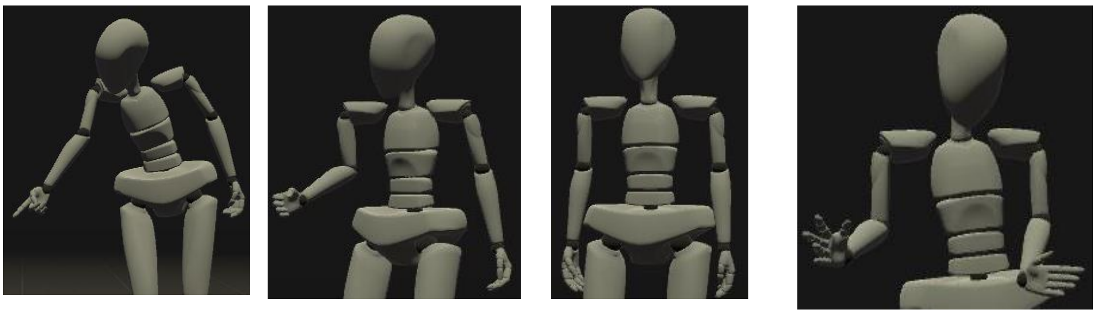


Figure 4.10. From stroke step to stroke step

Subsequently, the selected fragments were refined in terms of gesture taxonomy, i.e. interactive and representational. Figure 4.11 shows a distribution of gestural types in the first task by nodes. There is no gesture between the interactive and representational fragment. Further, the corresponding dialogue content of the trainer has also been coded in order to check whether the meaning of gesture can be matched with it.

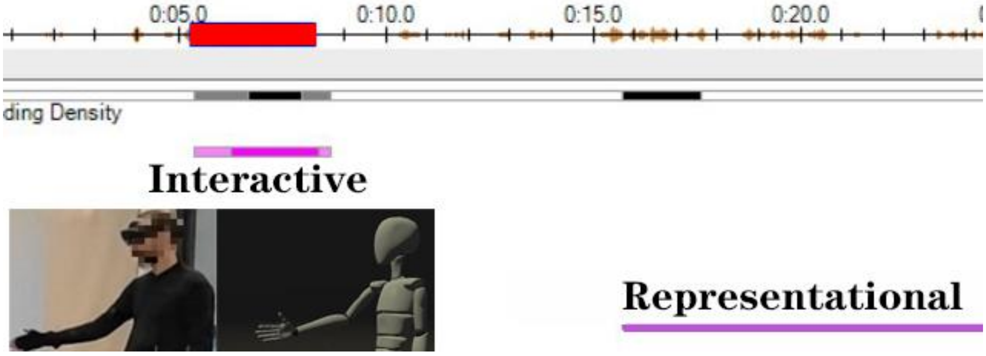


Figure 4.11. The interactive and representational gestures

4.4.3.2 Coding gestures

Gestures include hand and arm movements, although the term also refers to any body part motion (Lin, 2017). However, this study mainly focuses on hand movements, especially representational gestures, as the representational gesture has affordance, and the movement of lower limbs includes almost walking and standing and mimicking the posture of the first task in the experiment.

According to the process that proposed by Atit et al. (2013), the interpretation of gestural meanings should be based on the context and accompanying speech. Gestures should be coded on the basis of the type of the representational gesture, i.e. deictic, iconic, metaphorical, beat, emblematic, and cohesive (Bernard, Millman and Mittal, 2015; Abner, Cooperrider and Goldin-Meadow, 2015). Specially, the deictic gesture includes pointing at objects, pointing at locations, and pointing at directions (Hassemer and McCleary, 2017). Therefore, in this experiment the segments regarding deictic gestures have been subdivided.

After that, the gestures were coded, based on whether the trainee can correctly follow the instruction; is view of this, it is important to explore what factors can lead to failed instructional gestures, and how the trainer can change strategies. Therefore, both situations have been classified and coded.

Figure 4.12 illustrates a distribution of gestures in the assembly task of one of the groups. The empty places indicate periods during which there no instructional gestures, or when the trainer was waiting for the trainee to complete the sub-task as per the relevant instruction. During the experiment, instances when the trainee misunderstood or ignored the trainer’s instruction were recorded, and both situations were then analysed and compared. Fortunately, such occurrences were few in number.

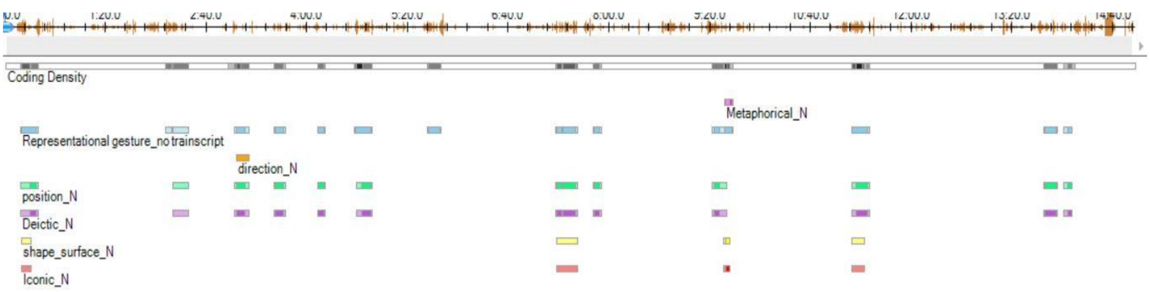


Figure 4.12. The distribution of gestures

During gesture coding, it was decided that each type of the representational gesture could be subdivided into different branches.

In terms of *deictic gestures*, the trainer used the index finger to point at a specific object and position, and the palm pointed to the object's direction. For example, in the Figure 4.13 (1) the trainer's palm instructs the trainee to put a panel on the left, and in Figure 4.13 (2) the trainer points at the panel to help the trainee select the right flat board. Therefore, the palm can be used to represent an orientation, as explained by Iverson and Goldin-Meadow (2005). Further, palm pointing can indicate position and direction simultaneously. In Figure 4.14, the trainer instructs the trainee to place a panel at a specific point, his palm moving down to imply that it should go straight.

Figure 4.13 (1). The palm pointed to the left side

Figure 4.13 (2). The index finger pointed to the panel

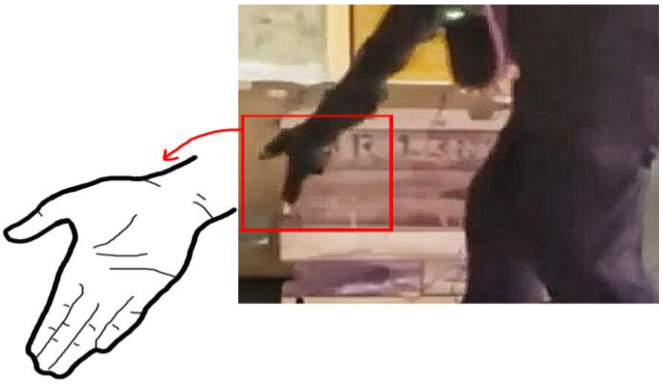


Figure 4.14. The palm indicating the referent's position and direction

The *iconic gesture* describes a referent's feature or property, thus indicating the shape, length, or size of an object. For instance, the trainer uses two hands to describe the roof's shape (see Figure 4.15(1)) in the assembly task. The thumb and index fingers can explain whether a referent is small. The trainer could create a far distance between left and right sides of hands to indicate 'big triangle' or stress 'big' (see Figure 4.15 (2,3)). If the trainer highlights 'long side', his/her palm would draw from top to downside.

Figure 4.15 (1). The roof shape

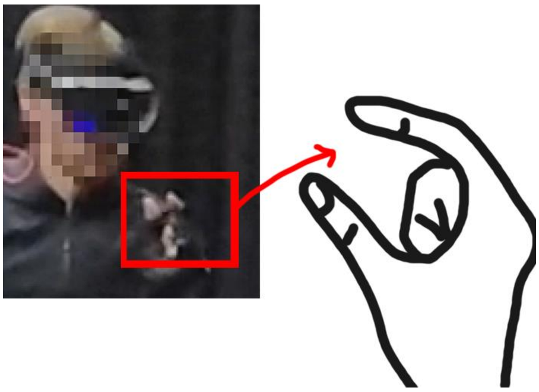


Figure 4.15 (2). Iconic_small size

Figure 4.15(3). Iconic_big size

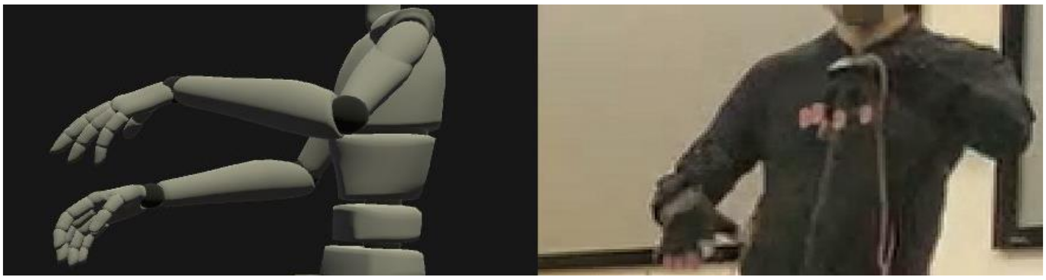


Figure 4.16. The spatial relationship of two referents

Besides, the iconic gesture also can present spatial relationship between two referents (Kandana Arachchige et al., 2021). Therefore, this type includes gestures indicating spatial position and referents' angles. For example, the trainer's two hands can mimic a spatial relationship between two referents. Additionally, the trainer uses two hands when asking the trainee to place his/her feet far apart. Since the cardboard house can be divided into several boxes in the hard level of the assembly task, the trainer emulates how the parts of the house are placed. Figure 4.16 displays the trainer indicating how base cubes should be positioned and connected by putting down the grasping hands gesture. It can be seen in Figure 4.17 that the trainer's palms mimic a right angle. The iconic gesture is based on the hologram representation to simulate a relationship of a set of objects, and it neither points to a direction or position, nor indicates an object's shape or size.

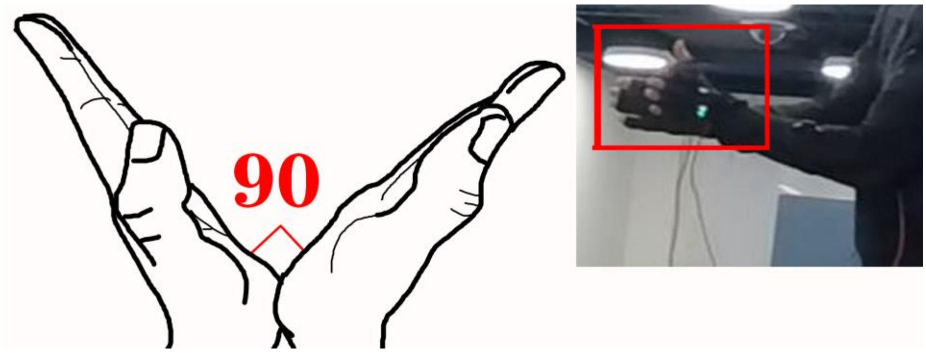


Figure 4.17. Mimicking a right angle

The *emblematic gesture* does not necessarily rely on speech as it can operate independently in expressing meanings (Matsumoto and Hwang, 2013), and it can be interpreted as a commonplace gesture. The emblematic gesture in this experiment includes thumbs up, waving index finger, crossed arms, palm forward, counting finger, and palm movement. In order to express the number of referents needed to be placed at a certain point, the trainer uses the fingers. To be more specific, the trainer will use the thumbs up gesture when in agreement with the trainee's behaviour (see Figure 4.18(1)), whereas the crossed arms or waving index fingers gestures indicate that the trainee is failing to follow the instructions correctly (see Figure 4.18(3)). The palm facing forward gesture is used to ask the trainee to stop, or to pause for a few seconds and allow the trainer to double-check the hologram (see Figure 4.18(4)). Extending four fingers can indicate the number of panels needed (see Figure 4.18(5)), and horizontally moving the arm means the trainee needs to restart or move out objects (see Figure 4.18(6)); these gestures occurred during the assembly and precision tasks.

Figure 4.18 (1). Thumb up

Figure 4.18 (2). Waving index finger

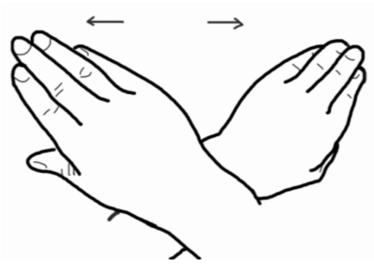


Figure 4.18 (3). Crossed arms

Figure 4.18(4). Palm facing forward

Figure 4.18(5). counting finger

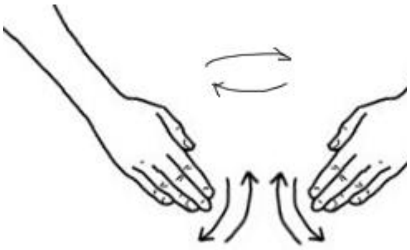


Figure 4.18(6). Restart/move out

Figure 4.19. The mimicking gesture in the navigation task

In addition to emblematic, iconic and deictic gestures, the trainer's hand movement can simulate the hologram's posture in the first task, i.e. perform *mimicking gestures*. For example, most of the gestures simulate the hologram's posture in the navigation task, even though the trainer cannot do fully the same as the hologram (see Figure 4.19).

Another type which occurred frequently in the experiment is the *transformational gestures*. This type is used to guide the trainee to manipulate referents, such as flipping and rotation. For example, in the precision task, the trainer's palm flips to instruct the trainee that the parallelogram needs to be turned over. The rotating wrist gesture indicates that the square needs to be rotated (see Figure 4.20). It also can express the other side of the referent by this gesture.

Figure 4.20. The transformational gesture

As examples of the *metaphorical type*, the trainer moves his palm from left to right side to imply that the cat hologram in the precision task resembles a mirror image. The hand put down gesture means that the trainee's knee required bending (see Figure 4.21), if the trainer suspends instruction, he or she points the index finger upward when asking for a few seconds to observe the hologram.

Figure 4.21. The example of the metaphorical type

The six sub-categories, or branches, of representational gestures are illustrated in Figure 4.22:

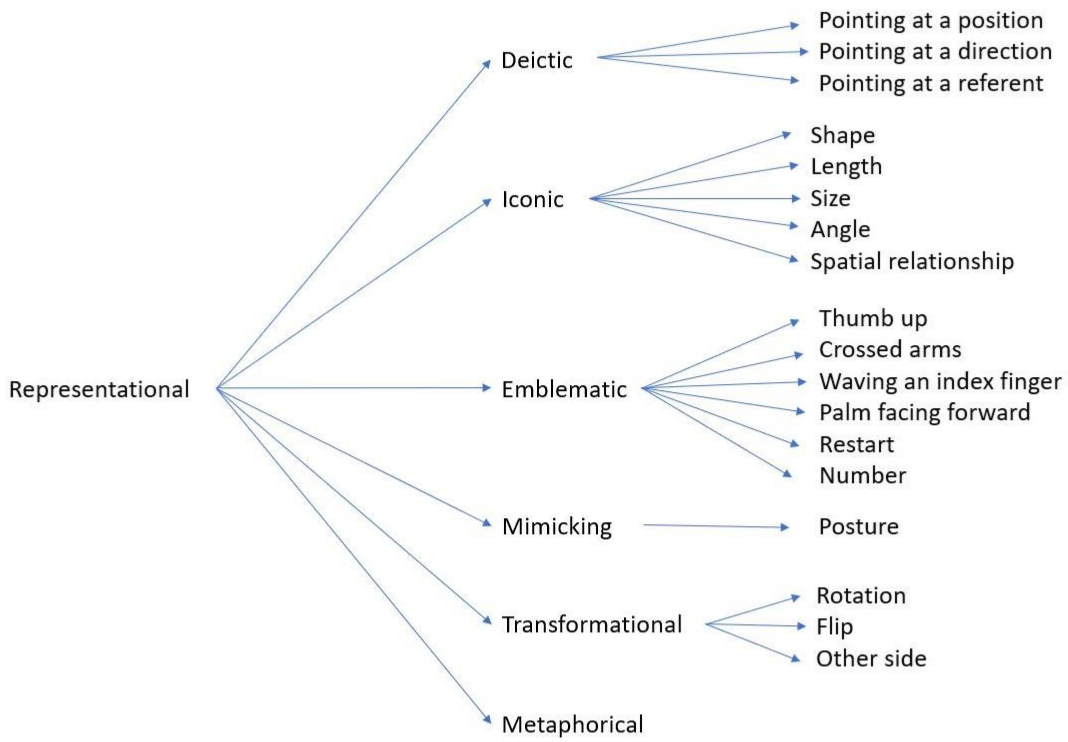


Figure 4.22. The branches of the representational gesture

As mentioned before, deictic gestures entail pointing at a position, direction, and a referent. The gesture of pointing at a position requires that the trainer is clearly asking the trainee to put an object on a location, such as closing a square, or placing an object on a specific location. A gesture of pointing in a direction reflects a direction, and not just a specific position such as turning left, but also aligning an object in a certain direction. Pointing at a referent might include referring to an object or a part of the body of the trainer or trainee. For example, the trainer might be pointing at his/her leg.

Iconic gesture in this experiment often describes the referent's shape, size, and length. When the trainer is describing the referent's length, he/she may use a comparable approach. For example, if the length is long, the trainer's arm movement range is large, or the distance of the two arms is longer. In addition, this gesture simulating spatial position reflects that different referent's relationship in space. Both hands of the trainer represent relative locations, rather than pointing at an orientation or a position. Similarly, this type of gesture also can describe angle, but in this experiment, the trainer employs both hands or fingers, and instructs the trainee to integrate two referents by mimicking an angle.

By registering the emblematic gesture, the trainee can directly comprehend the trainer's meaning, despite their cultural differences. A thumbs-up gesture can encourage trainees, but an arm crossed open, or finger waving implies “no” or “leave out”. When the trainer put (a) hand(s) forward, the trainee is instructed to discontinue a certain behaviour and wait for the next instruction if the trainer conveys that the instruction has not been followed correctly and he/she needs to restart again, the arms and hands are positioned on the same horizon and randomly moved around, which implies that the process is incorrect. For assembling cardboard forts, a counting finger is used to select the number of panels.

The mimicking gesture replicates the hologram's posture, but it is not static gesture. As an example, if a trainer holds his/her hands up, his/her arms move slightly up and down to show the position of the arms.

The transformational gesture can be used to manipulate the referent itself, but it does not establish a relationship between space and other objects, whereas iconic or deictic gesture can help construct a spatial position by referring to the other object's location. A gesture may also reflect a direction, for example, the trainer pointing on the left side when the referent should move to the left.

The metaphorical gesture is used to describe a physical object. For example, the trainer pointing at his/her elbow is employing a gesture which represents a corner of a shape.

4.4.3.3 Hypotheses

This study aims to explore which key instructional gestures can be utilized by an educational holographic AI (RQ3) to enhance the AI's non-verbal communication capabilities.

The experiment includes both non-verbal and speech-accompanied gesture trials to compare which can encourage the instructor to produce more gestures and to assess whether similar tasks are articulated differently, for example, warm-up and easy trials in task 2.

Previous studies have demonstrated that the pointing gesture can enhance a learner's understanding in education (Atit, Gagnier and Shipley, 2015; Matsumoto and Dobs, 2017). Therefore, the experiment may frequently observe deictic gestures.

Moreover, the first task investigates how a stance is transitioned, while the second and third tasks explore how the instructor sequences information and whether they employ consistent gestures to direct the learner. It is plausible that different tasks may influence the instructors' cognitive processes.

Therefore, based on the related work and the study method, the following hypotheses are proposed:

- H1: Participants generate more gestures during the speechless segments of the three tasks (warm-ups).
- H2: For the participants, deictic gestures constitute the key functional approach.
- H3: The three tasks differently affect participants' way of thinking.

These three hypotheses aim to identify which key instructional gestures are essential, as well as how to utilize these gestures to convey the holographic AI's motivation and instructional intent.

4.3.3.4 Reliability

Each of the gestures and accompanying speech was coded by the first-author, and, in order to measure interrater reliability, the concordance with two independent coders on 11.3% (15 trails) of videos that were randomly selected was measured. It should be noted that the second coder did not select corresponding verbal content, since timespans of speech content depend on occurrence and duration of gestures, and the warm-up trials had no transcripts.

Code	File	File Folder	File Size	Kappa
Representational gesture_no trainscript	20221012_Task1_easy	Files\\instructio	0:27.9 duration	0.9755
Representational gesture_no trainscript	20221214_Task1_warmu	Files\\instructio	0:38.6 duration	0.9262
Representational gesture_no trainscript	20221027_Task1_mediu	Files\\instructio	5:01.1 duration	0.8398
Representational gesture_no trainscript	20221012_Task2_hard	Files\\instructio	14:57.1 duratio	0.824
Representational gesture_no trainscript	20221205_01_Task3_har	Files\\instructio	3:42.3 duration	0.8139
Representational gesture_no trainscript	20221128_Task2_mediu	Files\\instructio	5:53.0 duration	0.8129
Representational gesture_no trainscript	20221109_Task2_easy	Files\\instructio	5:15.1 duration	0.8082
Representational gesture_no trainscript	20221205_01_Task1_har	Files\\instructio	1:38.5 duration	0.7971
Representational gesture_no trainscript	20221214_Task2_warmu	Files\\instructio	3:59.2 duration	0.7959
Representational gesture_no trainscript	20221016_Task3_mediu	Files\\instructio	3:57.9 duration	0.7921
Representational gesture_no trainscript	20221214_Task2_hard	Files\\instructio	11:52.5 duratio	0.7438
Representational gesture_no trainscript	20221129_Task2_easy	Files\\instructio	4:40.5 duration	0.735
Representational gesture_no trainscript	20221205_01_Task2_har	Files\\instructio	11:06.8 duratio	0.7334
Representational gesture_no trainscript	20221128_Task3_hard	Files\\instructio	4:44.1 duration	0.7219
Representational gesture_no trainscript	20221214_Task3_mediu	Files\\instructio	4:09.5 duration	0.719

Figure 4.23. The Kappa value

Cohen's Kappa was then used to calculate agreement on gestural existence (Cohen, 1960). Fleiss (1981) pointed out that the if the Kappa value exceeds 0.7, the coding is reliable. NVivo can directly measure Cohen's Kappa and produce a reliability value. It relies on the selected nodes as well as duration of segment.

A second rater coded all types of representational gestures, including deictic, iconic, metaphorical, transformational, mimicking, and emblematic gestures. Due to NVivo12's calculation of segment length, the timespans selected by the two coders differed. Therefore, both raters compared selected nodes in each fragment, and made sure the nodes were the same, but did not refer to the node’s duration.

The percentage of the agreement in identifying the representational gesture in each trial ranged between 0.72 and 0.98 (see Figure 4.23), and the mean value was 0.8. The Kappa values for different types of representational gesture all exceeded 0.7 (see Figure 4.24), with the exception of the emblematic gesture since this type of gesture appeared too swiftly, especially the thumbs up gesture, it was harder for the raters to find agreement here.

Kappa	Transformational	Deictic	Mimicking	Iconic	Emblematic	Metaphorical
20221012_Task1_easy	1	0.7599	0.984	1	1	1
20221012_Task2_hard	1	0.7311	0.9099	0.7609	0.8337	1
20221016_Task3_medium	0.8983	0.7256	1	1	0.8088	1
20221027_Task1_mediu	1	0.8559	0.8113	1	1	1
20221109_Task2_easy	1	0.7832	1	0.7682	1	1
20221128_Task2_medium	1	0.7413	1	0.7967	0.5355	1
20221128_Task3_hard	0.7444	0.771	1	0.7547	0.6323	1
20221129_Task2_easy	0.7614	0.8636	1	0.7821	1	1
20221205_01_Task1_hard	1	0.7	0.8563	1	0.8289	1
20221205_01_Task2_hard	1	0.8195	1	0.7037	0.6951	1
20221205_01_Task3_hard	0.8389	0.8216	1	0.7234	0.4173	0.7815
20221214_Task1_warmup	1	0.8532	0.9135	1	0.8014	1
20221214_Task2_hard	0.8017	0.7587	1	0.7715	0.4803	1
20221214_Task2_warmup	0.8127	0.7768	1	0.8676	0.6075	1
20221214_Task3_medium	0.7042	0.8871	1	1	0.8232	1

Figure 4.24. Kappa values of gestural types

4.4.3.4 Frequency of the gesture

NVivo can calculate the number of types of gestural occurrence by referring to the node, and it also includes the times recoded by all raters. Therefore, the second rater's history was deleted to avoid repetitive nodes.

In view of the satisfactory reliability measurement results, it is now possible to measure the frequency of each type of gesture according to the representational gesture branches to determine which gestures play a key role in the experimental tasks. Highly frequent or rare gestures are less likely to be meaningful or critical since they may be artefacts in coding, mis gesticulation, or involve emotional or non-functional gestures, such as akimbo and scratching.

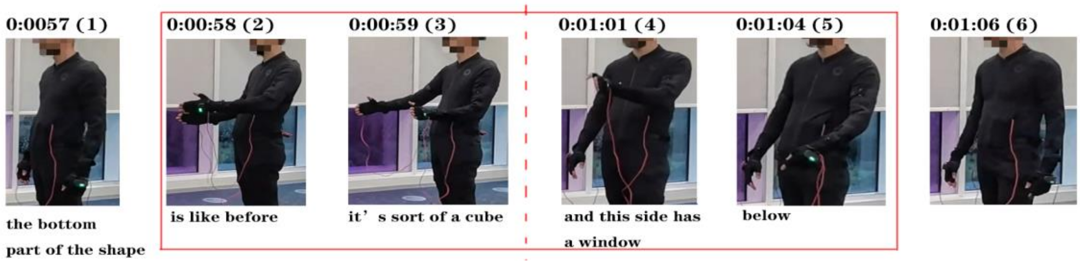


Figure 4.25. An example of selecting key gestures

The process of key gesture selection requires filtering non-instructional gestures that the trainee has misunderstood or not observed. A previous gesture is calculated once before the next meaning of gesture appeared, or the gesture is turned to pre-stroke hold and retraction. For example, in the second and third pictures of Figure 4.25 the trainer describes a shape of a cube that can be built by panels, and his two hands change orientation. This is one node of the occurrence of the iconic gesture. Then the fourth and fifth imagers of Figure 4.25 show that the trainer uses the deictic gesture to represent the orientation of the panel with a window. This is the second node of the gesture.

Then, it compared each type and its branch of gesture by the number of references. Figure 4.26 shows the distribution of all gestural types in 132 videos: it can be seen that pointing gestures occupy most of the area (53.8%).

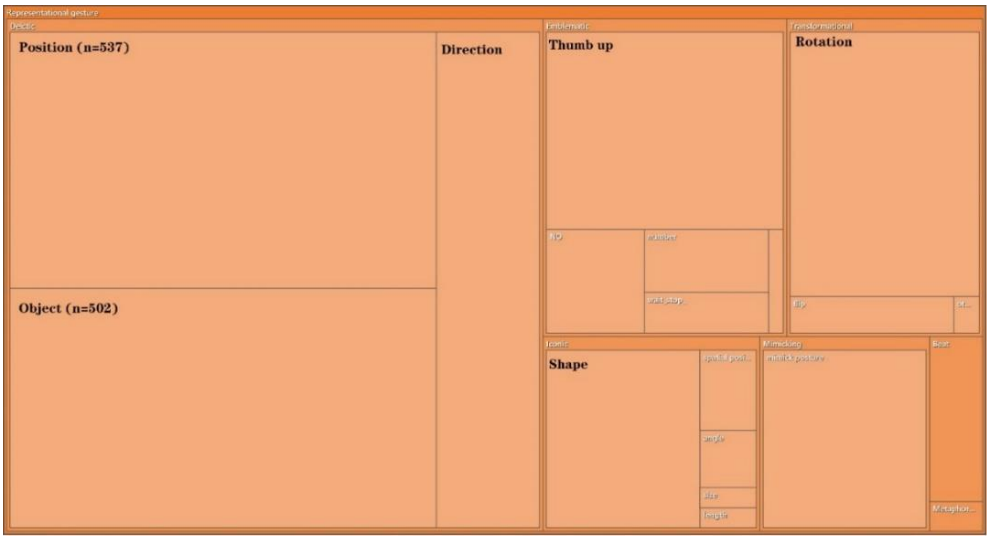


Figure 4.26. The distribution of all gestural types in 132 videos

4.5 Results

Following the review of the nodes of each type and branch of the representational gesture group, this section analyses the experimental data and identifies the key gesture. Since the trainer directly instructs the trainee in the experiment, only three trainers use the interactive gesture to express the task started or to imply emotion, such as inviting gestures and opening arms.

In total, 132 videos contained 2,348 nodes of the representational gesture group, which means different gestures appeared over 2000 times. The videos included 39 nodes that the trainee did not correctly follow, and 3 nodes where the trainee did not observe during instruction (see Table 4.4). In the experiment the trainer uses waving hands or index fingers to express ‘not correct’. This situation often appeared in the speechless trials of the assembly and precision task. The trainee was required to build 3D paper houses, for which the failure rate was higher. In performing this task, the trainee must rotate one of the cardboards to the correct orientation after selecting it. The failed gestures were not discarded in data analysis, since it is important to observe how the trainer changed instructional approaches.

Situations	The trainee did not watch the gesture that led to make mistakes.	The trainee observed the trainer’s gestures but misunderstood.
Nodes	3	36

Table 4.4. The number of nodes of failed gestures

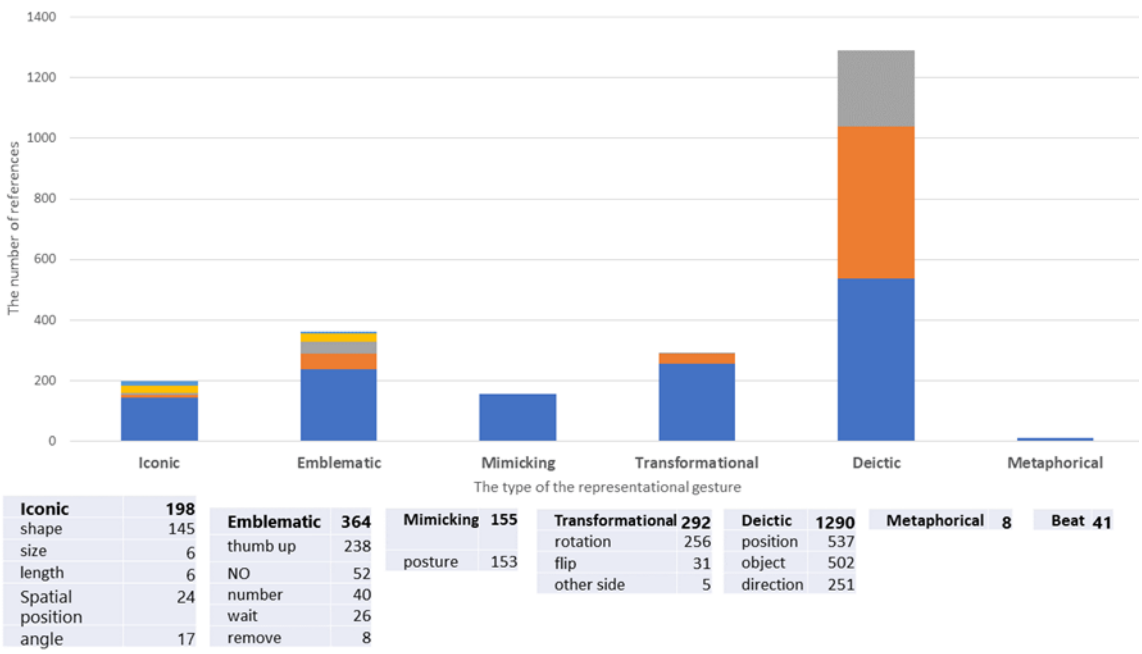


Figure 4.27. The number of each branch of the representational gesture

4.5.1 Data analysis

Figure 4.27 shows the numbers of gestures of each gestural branch in the representational type. This sub-section will investigate each type. The number of deictic gestures is the highest (n=1290), especially pointing at position and object.

Although there were more emblematic gestures than those of the transformational type, participants used gestures that showed rotation was more than just a branch of the emblematic gesture. The iconic gesture was also the type most often used to describe shapes of referents. Additionally, the least common gestures were metaphorical gestures (n=8).

4.5.1.1 The navigation task

Task1	Warm-up	Easy	Medium	Hard
Average duration (seconds)	68.09s	96.55s	107.73s	144.64s
Average number of gesture occurrence	9.36	7.45	5.91	9.27

Table 4.5. Average duration and number of gestures during each stage of the first task

In the first task, the trainer instructed the trainee to simulate the hologram's poses. This trial had a mistake in that the trainer used a wrong hologram that should have been on the second trail. Nevertheless, the goal of this task is to explore how trainers describe 3D holograms using body language, and so all the trainers mimicked part of the hologram's posture. Therefore, slight differences tended to occur. Table 4.5 shows the average duration of each of the four stages, and the average number of gestures generation in each one. Although it appears that the trainer generated more in the speechless condition, some trainers only took less than a minute, although one trainer took almost 2 minutes. Furthermore, the trainer used other types of gestures, such as the thumb ups, drawing circles with an index finger, or pointing at a direction and referent (see Figure 4.28). This navigation task also involved deictic gestures emphasising which body part should be oriented or posed. For example, the trainer's palm pats a leg when asking the trainee to bend the left knee, or the trainer asks the trainee to follow the instruction and do the same pose as the trainer.

	Warm-up	Easy	Medium	Hard
Sum	39	37	40	39
Mean	3.55	3.36	3.64	3.55

Table 4.6. Mimicking gestures

In addition, all mimicking gestures were generated during the navigation task (Figure 4.28). Table 4.6 shows the average and sum of each gesture generated during the first task. Since mimicking posture gestures are continuous gestures that trainers hold hands or arms for a long period of time, difficulty does not affect the number of mimicking gestures.

The navigation task was the first task in this experiment, in which trainers either mimicked the hologram's pose or verbally guided trainees. The standing direction between trainer and trainee was opposing, in that the left side of the trainer was the right side of the trainee. Those who generated more gestures firstly asked the trainee to stand the same direction of the trainer, and also used the thumbs up gesture to encourage the trainee during non-speech and hard conditions.

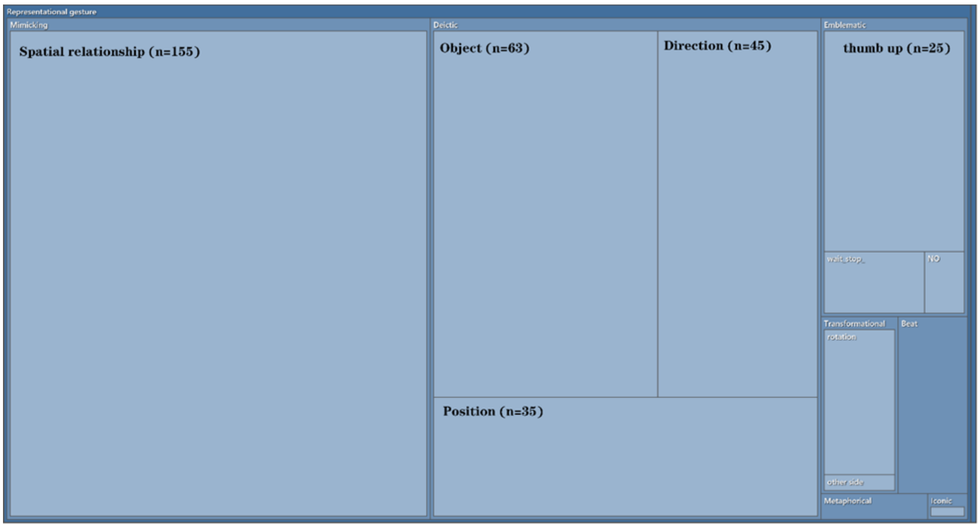


Figure 4.28. The distribution of types of gesture in the navigation task

4.5.1.2 The assembly task

The assembly task was the second of this experiment, whereby the trainer instructed the process of building cardboard forts based on virtual holograms. Table 4.7 shows the average duration and average number of gestural nodes for each of the four conditions. There were 869 nodes in this task, covering all types of representational gestures, except for mimicking gestures. It is evident that the degree of difficulty (level) affected the gestural generation. It can be seen in Figure 4.29 that the trainer generated more gestures in the non-speech trials, and then dropped some in later stages when the trainee was performing tasks similar to those earlier on in the experiment.

There is a flat roof on the warm-up level, and a triangle roof on the easy level. Hence, the trainer instructs the trainee to build the same base as before, with only the top being different. In the experiment, the trainer waits for the trainee to assemble the cardboard step by step, so the trainee’s progress determines the average duration. During the medium level, the trainer focuses on the orientation of the panel's window and uses standers to help the trainee build the second floor. While all the windows are positioned on the bottom in the last condition, the trainer must possess a well-organized ability to divide the whole house into blocks.

	Warm-up	Easy	Medium	Hard
Average duration (seconds)	251	296.82	385.82	885.27
Average number of gesture occurrence	29.27	11.55	12.18	26

Table 4.7. The average duration in the second task

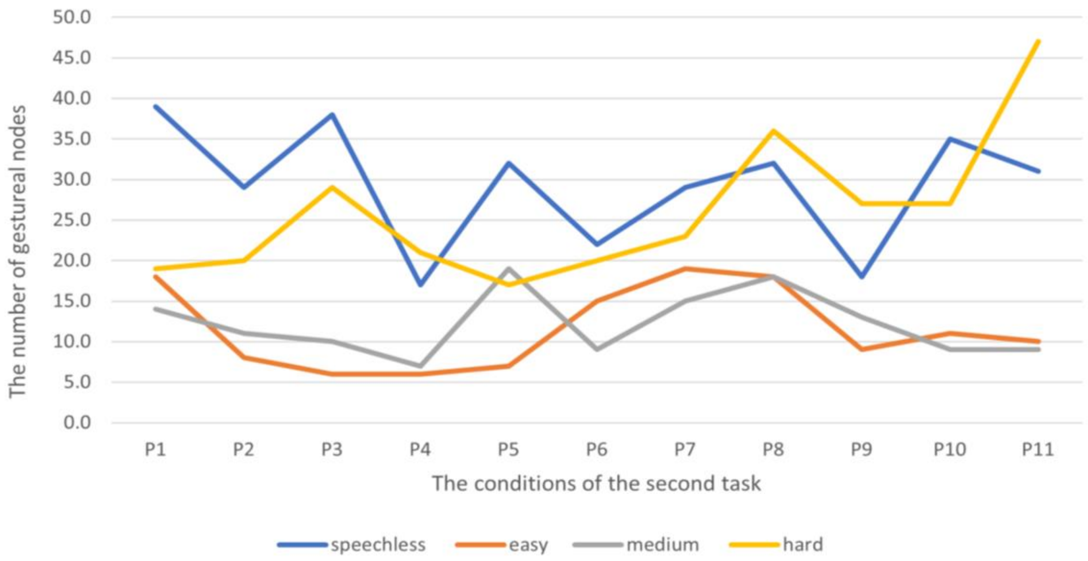


Figure 4.29. The variation tendency of four conditions in the second task

The deictic gestures were used most (n=456) in the second task (52.64%), and gestures pointing at a position and referent had 247 and 148 nodes, since in this task the trainer needs to direct the trainee's attention to the exact points for positioning the cardboard (see Figure 4.30). Interestingly, all trainers used iconic gestures to describe triangle roofs, and some also placed their palms to describe flat roofs. Furthermore, they also used emblematic gestures to encourage the trainee (n =110), and used counting-fingers to indicate the number of cardboard blocks needed (n=38).

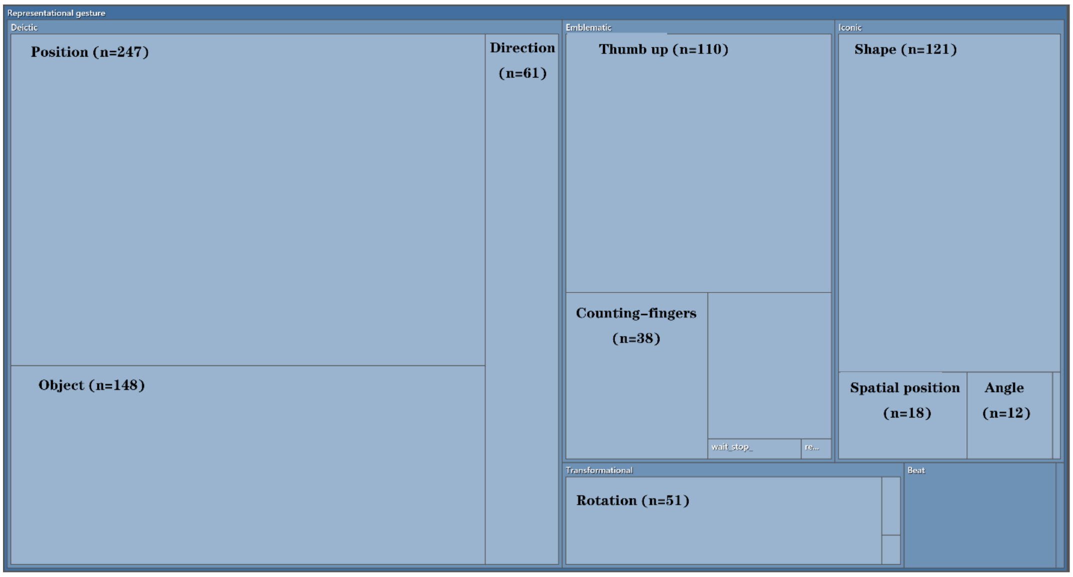


Figure 4.30. The distribution of the gesture in the assembly task

Under co-speech gesture conditions, some trainers briefly described the cardboard's information in advance. For example, a trainers explained cardboard fort shapes and their spatial position before instruction, and trainees did not assemble panels. Thus, two situations are worthy of examination: pre-instruction, and non-pre-instruction. There are 16 videos that show that some trainers provided pre-instruction, and 17 of them did not. It was found that the mean values of general nodes were 18.5 in the non-instruction situation and 14.5 nodes in pre-instruction situation, suggesting there may be a significant difference between those two situations in terms of the number of gestures. The pre-instruction situation is characterized by a lower gesture generation, as evidenced by a negative correlation coefficient (r=-0.20). Further, the time expenditure appears to have been negatively affected by the pre-instruction situation (r=-0.03), even though this correlation is not significant (p>0.1). Therefore, if a trainer provides each block’s position in advance, a trainee may spend less time completing this task. Moreover, while the average time on the pre-instruction situation was only 17.7 shorter. During the task, each trainer used deictic gestures or iconic gestures, i.e. gestures illustrating spatial position, signalling to the trainee which shapes should be built. The trainer’s two-hand movement and fingers can represent the different blockers of the cardboard fort, shapes, and position. In the non-description situation, the trainer had to explain to the trainee how many panels were needed directly, using finger-counting and pointing gestures. The trainer's meanings, however, may not be understood by all trainees. One trainer, for instance, in the hard condition, explained that the ground floor has an N shape, using the grasping hands gesture (see Figure 4.16 in Section 4.3). Therefore, during assembly, the trainer repeatedly explained the shape’s appearance.

Most trainers used palm-to-palm gestures to describe the orientation of the panel with the window and then performed the rotation movement to ask the trainee to put down the window at the bottom (see Figure 4.31). Although this gesture consists of iconic and transformational types, the trainer underlined the behaviour of rotation that the student should follow the instruction. There was a greater diversity of gestures depicting more completed referents in this task, compared to the navigation task.

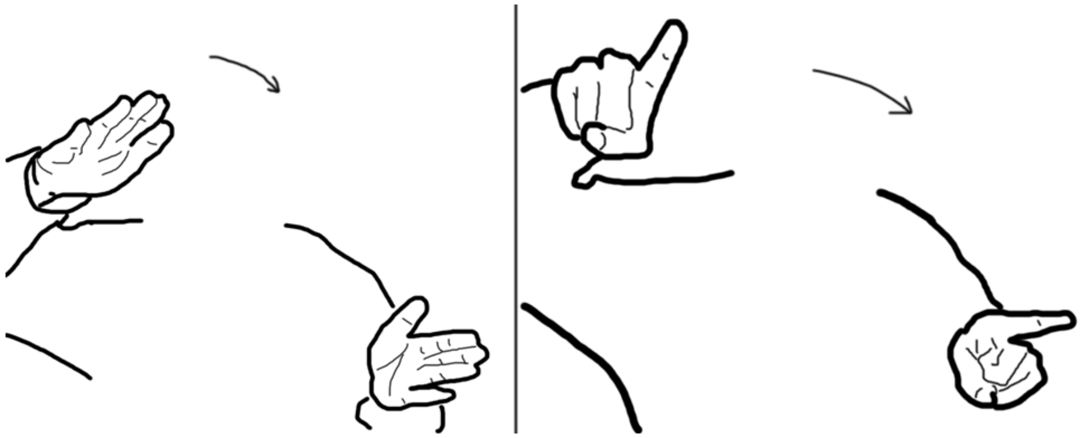


Figure 4.31. The palm-to-palm gestures

4.5.1.3 The precision task

	Warm-up	Easy	Medium	Hard
Average duration (seconds)	245.18	147.55	204.27	282.91
Average number of gesture occurrence	41.73	14.73	21.45	24.45

Table 4.8. The average duration in the third task

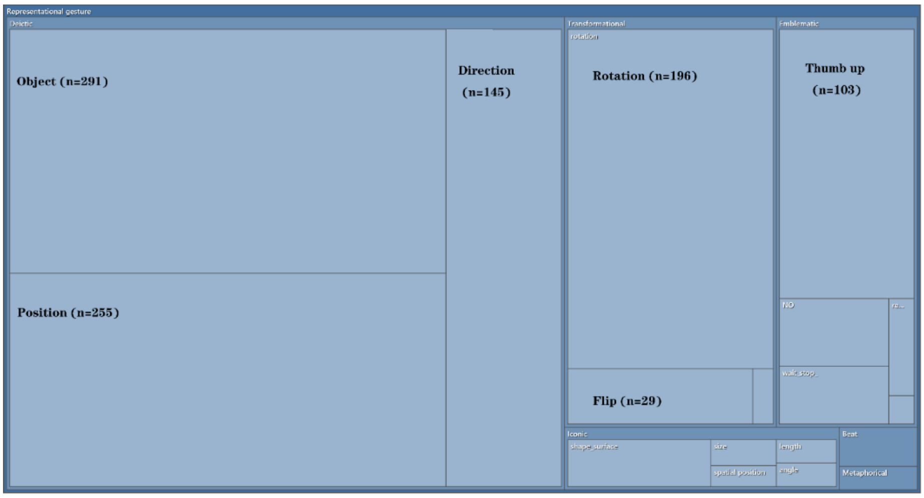


Figure 4.32. The distribution of the gesture in the precision task

In the last task, the trainee had to set up a specific puzzle-T structure. The average duration and average number of gestures for each stage of the precision task are shown in Table 4.8. The non-speech trial generated the highest number of gestures. Although the warm-up trial only needs four boards, the trainer needed to distinguish between three sizes of triangles, and correctly instruct the trainee to set up a ship shape. Over half of the trainers did not correctly emphasise size differences (n=6), resulting in the placement of a large triangle rather than a medium triangle. Further, a few trainers did not realise that two large triangles could form a big square under the easy condition.

In the precision task, the deictic gesture type had 691 nodes, including 291 nodes relating to pointing at referents and 255 gestures for pointing at a certain position (see Figure 4.32). The trainer's palm moves forward or downward to represent pointing in a particular direction, for when a board needs to be translated vertically or horizontally. The transformational gesture was also used frequently, especially rotation (n=145), but the hands were shaped differently. For this task the trainer need not perform the opening-arms and palm-to-palm gestures that were used in the assembly task, but can place the hands and palms down, and then use slightly rotated wrists. The trainer's fingers can also depict shapes via an iconic gesture, replacing the rotational gesture of the hands, so that the trainee can understand how to place the blocks correctly (see Figure 4.33).

Figure 4.33. The iconic gesture

Similarly, there are 9 videos that shows some trainers participating in this experiment explained the puzzle-T's shape prior to setting up (i.e. pre-instruction), and it was mainly during the medium and hard levels that the shape could be clearly identified. Those who provided pre-instruction generated fewer gestures (mean=21), while the direct instruction produced 3-4 more gestures, and spent more time. There was also a negative correlation between pre-instruction and gesture number (r=-0.28), as well as a negative correlation between the pre-instruction and gesture duration (r=-0.42, p=0.049).

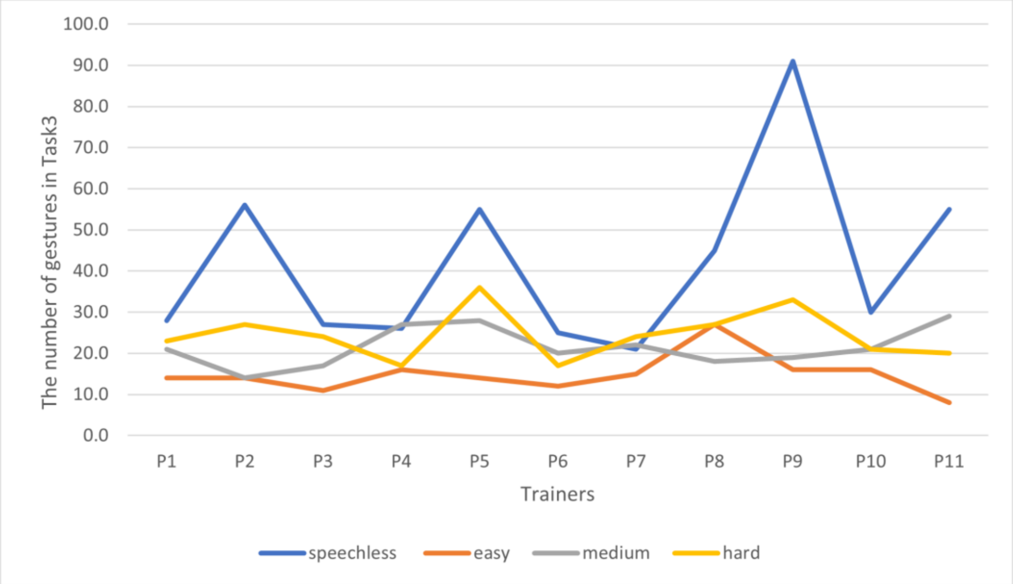


Figure 4.34. The variation tendency of four conditions in the third task

According to Figure 4.34, there occurred more gestures at the non-verbal level than at the hard level, since the squares in these two conditions shared the same orientation. Furthermore, the most metaphorical gestures were generated in this task (n=5). For example, the hand may be moved to the left side to indicate a previously used board; pointing at the elbow joint can indicate a triangle's right angle; and the hand being moved slightly from left to right (in the medium level) indicates that the two small triangles are mirrored.

The precision task requires imagination and therefore poses a greater challenge for the trainer compared to the assembly task since the puzzle-T shapes are abstract, and the trainer needs to consider how to rotate each shape to the correct angle. In this circumstance, if the trainer stands opposite the trainee, then the trainee is unable to follow the trainer's instructions. This difficulty is compounded if the trainer lacks experience in playing puzzle-T.

In the experiment, the trainers unconsciously simulated the hologram while performing the navigation task (n=147), even though the task did not allow performing the same pose. Nevertheless, this gesture cannot be considered a primary one, since this task has particularities that cannot be found in other tasks. Furthermore, the deictic gesture type was the most commonly used in assembly and precision tasks, especially when the trainers were pointing at a position or referent. The thumbs up gesture appeared frequently, even though this is not instructional.

The non-speech condition can facilitate more gestures which in turn function almost as a second language. More gestures were generated whenever the trainers needed to instruct with complex tasks, whereas familiar or similar trials require less gestural generation.

4.6 Discussion

The aim of this experiment is to investigate instructional gestures in order to generate corresponding animations for a holographic AI. Therefore, this section will discuss which types of gestures are appropriate for each type of situation, consider how gestures match utterance, and explore hypotheses.

4.6.1 Gestures

The deictic gesture consists of pointing at a position, a direction and a referent. In order to emphasise an important point and attract the trainee's attention, the index finger or hand palm can be used. As an example, the trainer points by way of explaining where the trainee must position the panels in the assembly task. Directing the index finger at the palm was the most frequent gesture used by the trainer when pointing at a specific direction, or when instructing the trainee to sit on the floor or place the referent on the ground, as these do not highlight a specific place. Trainers' palms can also be interpreted as flat objects. During the medium level of the assembly task, the trainer pretended to place his/her hand on top of the cardboard fort instead of pointing with his/her index finger. Therefore, the unfolded hand can be interpreted as the flat cardboard fort without a window, with the orientation of arm movement representing the position. Additionally, the behaviour of pointing is not static, especially when the trainer points in a direction to instruct the trainee to move an object forward, moving the hand up and down repeatedly until the trainee successfully corrects it. Importantly, the gesture with regard to pointing at an object may not refer to the object itself. In the precision task, the trainer should point at the panel and guide the trainee to put the other shape near the panel, thus this gesture implies a relative position as well. There was no abstract object being represented by the deictic gestures in this experiment (e.g. time).

An abundance of deictic gestures is noted in the second and third tasks, in contrast to the first task, which concentrates on body positions. The term "indicating a direction" might be ambiguous for instructing the placement of an object, while a specific location or referent provides clear guidance for the learner. In the first task, gesturing towards a direction is often used rather than specifying an exact location, especially when the instructor needs to communicate the orientation of a limb, such as left or right. This observation confirms the second hypothesis concerning the prevalence of deictic gesture production.

The emblematic gesture was the second most frequently used gesture (n=364), particularly thumbs up (n=238). However, over 75% of videos show generated thumb ups in speechless conditions, even though the thumb ups gesture is not instructional. It appears that inclusion of the thumbs up gesture produces a decrease in instructional time (17.8 seconds average). The thumbs up is an emblematic gesture (Andric, 2012), but also an interactive gesture (Curioni, 2020). It is a cultural mutual gestural vocabulary to convey thoughts and emotions.

In this experiment, the crossed arms gesture can represent the cardboard without a hole (see Figure 4.18), and the counting finger gesture supports the trainee by implying how many cardboard blocks are needed for each floor, after the trainer has pointed at the right type of cardboard.

The trainer employs the two hands' palms facing forward gesture when asking the trainee to hold for a second and look at the hologram, or whenever the trainee rotates or translates correctly to stop continuous rotation. By moving one hand from left to right, the trainer can imply that the referent needs to be removed. If two palms casually draw circles on the same plane, this means the referent needs to be reset (see Figure 4.18(3)). Similarly, the gesture of waving the index finger or hand (applied uniformly throughout the experiment) denotes an incorrect action by the trainee.

In the experiment, transformational gestures outnumbered iconic gestures. Two palms facing each other, and an unfolded hand make up the gesture of rotation. These two forms may depend on the size of the referent. Two palms or index fingers kept apart and facing each other imply the size of the cardboard, while the unfolded hand conveys the size of a puzzle-T board. Another transformational gesture which appeared regularly is the flipping gesture, whereby the trainer instructs the trainee to place the base cube with windows at the bottom by employing the gesture of two palms facing each other, as if to say “you would reverse the window at the bottom” (see Figure 4.35). Therefore, the transformational gesture can potentially convey information about the shape of the object.

Figure 4.35. The transformational gesture can imply the iconic meaning

Figure 4.36. The rotating gesture in the precision task

However, the gesture of rotating the referent can also be performed with the palms facing each other, or using parallel index fingers, since the rotation angle and the way in which the trainer stands can affect the gestural pose. This form may emphasize the shape that needs to be rotated extensively. In the precision task, one trainer did not generate unfolding hand movements to guide the trainee in rotating the panel: he stood in a position, and pointed two index figures pointed forward, and his arms moved towards a specific direction. The rotating angle increased along with bigger arm movement amplitude, as he guided the trainee to rotate 90 degrees (see Figure 4.36).

The flipping gesture appeared uniformity throughout this experiment. For example, the trainer puts down palms to describe the referent, and then rotates the wrist to put the palm upward. This is similar to the transformational gesture in that it depicts the referent’s adjacent side, such as the other leg or the other edge of a triangle. As an example, the trainer's index finger or palm jumps forwards whenever the trainee grasps the wrong side. Although this could replace the rotating gesture, the movement of gestural direction is vertical.

The difference between this gesture and the deictic type is that the former considers the position of the adjacent object, so it cannot represent a direction or position.

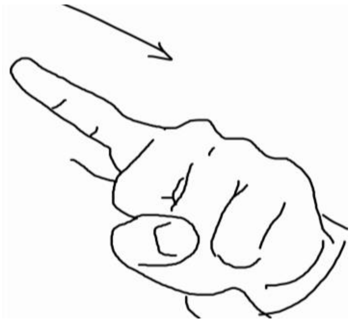


Figure 4.37. Example of the index finger drawing a line

As mentioned, the *iconic gesture* is used to describe the object’s shape, length, size, angle, and spatial position. In the experiment, the trainer employs this when mimicking the 3D basic cube, flat cardboard with window, roof, or 2D triangle. The iconic gesture can also depict a panel with the window in one specific level in the assembly task, which does not appear in other levels. In the non-speech level, the trainer first outlines that the panel has no window by drawing a rectangle with his/her index finger, before drawing a diagonal line with the same finger to instruct the trainee to choose the right cardboard (see Figure 4.37). In order to point out that the long side of the triangle should be closed against another shape, the trainer stretches both arms or draws a line. Therefore, the iconic gesture also can stress a key feature of the referent alone.

The iconic gesture can also illustrate a spatial position, and this representation can describe an overall arrangement as well as reflect the object's position and features, although it does not refer to a particular place or shape. For example, in the hard level of the assembly task the trainee has to build a castle. The trainer effectively splits this complex cardboard fort into blocks that are located in different positions, and his/her gestures represent the appearance of the high tower, how many floors it has, the position of each block, the shape of the next block, and the relative position of each block based on the tower.

Although the mimicking gesture appeared 155 times in the first task. Those may entail some complex poses, such as curving fingers and bent knees. In the speechless situation, the trainer can employ the deictic gesture to attract the trainee's attention and indicate the shape of the body part using mimicking gestures.

The metaphorical gesture was not frequently generated in this experiment, but it can imply previous behaviours and present performance via movements of the left and right hands. In the warm-up trial, the trainer points at his/her elbow to imply a corner of the triangle. The trainer's two hands' index fingers inscribe the letter 'X' to emphasize that two panels need to be integrated. The metaphorical gesture applies a particular characteristic of the hologram's position, shape, and spatial relationship.

The beat gesture in the experiment can help the trainer to organize expression and logical thoughts, and the duration of this gesture tends not to be long (around a few seconds). In the hard level of the navigation task, the trainer uses the beat gesture along with the speed of utterance to indicate that the trainee's left leg should put down on the side on the floor. When stressing the instructions "flat", "on the side" and "on the floor", his/her left-hand moves move downward.

Interaction with the trainer involves gestures that invite interaction, and an opening-arms gesture expresses the trainer's emotion.

The gesture's taxonomy is determined by its functions and semantic context. This experiment requires transformational gestures, not merely gestures of basic types. A transformational gesture is more complex than, e.g. pointing an index finger in a certain direction. In gestural representation, direction often refers to placing a referent on the left or right side, or moving it forward or backward, without providing a clear value. The transformational gesture, on the other hand, can represent by how many degrees something needs to be rotated.

4.6.2 Cognitive ability affects gesture generation

The trainer's gestures imply ways of thinking, cognitive and logic ability. This is especially the case with the assembly and precision tasks. In the warm-up tasks, even though the trainer cannot speak, gestures can successfully provide instruction by way of key information selection and organization in order to develop processes. However, in the warm-up level of the assembly task, the trainer does not indicate the basic shape in advance, but instead points directly at the panel. This alone demonstrates that cognitive ability need not rely on speech, but on gestures. This finding also supports the third hypothesis by demonstrating how varying levels of task difficulty can affect gesture generation and cognitive strategies.

Speechless trials require the least complexity; thus, they are considered elementary if verbal instructions are given by the instructor. Even though the warm-up trial shows a greater number of movements, it is completed more quickly. The more challenging trial prompts fewer gestures but takes more time to complete. The instructor and learner may become increasingly familiar with each other and their tasks over time, yet complex trials demand periods of contemplation and articulation. It does not imply that more time is required to produce a greater number of gestures at higher difficulty levels. Moreover, while the speechless and simple trials feature comparable configurations of cardboard structures, the instructor who produces fewer gestures may take longer to conclude the simple task, suggesting that the instructor needs to structure their verbal guidance more carefully. Therefore, the increased number of gestures noted in warm-up trials could support the first hypothesis.

If the trainer is permitted to offer verbal guidance, warm-up trails are simplified in every task. While the warm-up trial has the greatest quantity of motions, it requires a shorter duration. The challenging track produces a reduced number of motions, although requiring a greater amount of time. The process of familiarizing trainers and trainees with each other and their respective jobs typically occurs gradually. However, when faced with challenging difficulties, individuals often require more time for contemplation and articulation. This does not suggest that there is a correlation between the difficulty level and the time required to generate a greater number of gestures. Furthermore, it is worth noting that there are similarities between the warm-up and easy conditions in the assembly assignment. However, it is interesting to observe that the trainer who generates fewer motions required a longer time to finish the easy condition. This suggests that there may be a need for the trainer to improve their language organization in providing assistance. Therefore, the initial hypothesis can provide an explanation for the higher occurrence of gestures observed during warm-up trails.

Whenever a hologram similar in shape to a previous one appeared, the trainers would extract previous spatial information to generate similar gestures. However, the trainer did not repeatedly provide more gestures to instruct the trainee where the panel was; he/she applied the pointing gesture to remind the trainee to tape boards.

As discussed in the previous section, when the hologram was more complicated the trainer needs to spend more time in observing and figuratively cutting it into different blocks. For example, in the hard level of building the cardboard fort, most trainers instructed the trainees to assemble the foundation shape first before building the second floor, which horizontally split the house. Alternatively, the trainers firstly instruct the trainer to start from a base cube, as done previously, then assemble the second floor, and lastly build the right-side block. In the pre-instruction situation, the trainer's gesture continuously depicts the shape of the virtual castle and the location of the panel with the window. The trainee follows this instruction to assemble the highest part of the cardboard fort. Therefore, the information can be packaged into different units that not only represent the way of thinking but also gestural generation. On the other hand, there was an instance where one trainer described the foundation shape as being like the letter 'N', but the trainee did not fully understand this metaphor. If a concrete shape is described by way of an abstract object, the trainer's explanation may consume more time and entail more gestures. Additionally, the third hypothesis can be corroborated by observing how different tasks affect the way information is structured and manipulated through gestures.

Although the assembly task also involves hand movements regarding rotating and translating the object, the purpose of this particular task is to observe how the trainer's package spatial information. Information manipulation appeared frequently in this last task. The trainer needs to manipulate spatio-motoric information by gestures including rotating, flipping, pointing at a specific direction, and instructing each orientation and position of the shapes, one after another. If the trainer cannot recognise that two large triangles can make up a square, the trainee might decide to set up the end parts of the shapes.

Additionally, when the trainee does not correctly understand the instruction, the trainer might wave the hand or index finger. In the precision task, the trainer stands on the opposite side of the trainee, and might accidentally lead the rotation in the opposite direction. Therefore, the trainer's hand orientation moves to the other side to correct mistakes, or the trainer performs the same gestures with bigger amplitude of motion again. If the trainer finds that the shape cannot match the hologram, he/she resorts to using gestures which signal a restart. The data obtained in this experiment can be used to determine whether the trainer's thoughts can be dictated by the trainee. It explores the affordance of spatial information as well as practicability.

In addition, the gestural generation can be negatively affected by whether the trainer has described features of the referent prior to instruction, whether providing advance notice reduces the number of gestures and time expenditure necessary further down the line, as well as facilitate the trainee's imagination and his/her comprehension of the shape in question.

4.6.3 Co-speech gesture

This experiment also found that the gestural generation is quicker than verbal expression. This gap would be more distinct if the trainer has to retrieve words. The trainer may be able to organise spatial information more quickly by employing gestures. By way of example, the trainer might forget the name "Velcro", but can simulate its taped shape by drawing circles with his/her index finger. In this situation, the trainee can predict the trainer's meaning. The co-speech gesture supports lexical retrieval (Hadar, Dar and Teitelman, 2001), but the relative gesture also can directly replace the utterance in interaction. Further, the trainer can use different forms of gestures to describe similar functions of referents. The trainer who extracts the principal feature of Velcro using gestures might also show delays in lexical retrieval when describing the word "connector". The function of the connector and Velcro is similar in that both materials connect two panels, and so the trainer can generate the other form of gesture by vertically stretching two hands to illustrate what the connector is.

The trainer captures the key feature of the object, and stresses its key word along with corresponding gestures. In order to highlight a side of the basic cube that has no roof, the trainer need not simulate the cube shape, but employ a slow rhythm of utterance to emphasize "without roof", and simultaneously illustrating it using crossing-arm gestures. The same gesture can be produced again to underline the utterance "without", if the panel has no hole. The co-speech gesture is multi-modal communicative approach that conveys information by both schematization and speech. However, if the feature has been mentioned before, the trainer might choose to generate different forms of gestures. For example, all trainers' hands mimicked the triangle roof in the easy level of the assembly task, but in the hard condition, the trainers' palms pointed to the right cardboard blocks that needed to be assembled in the triangle shape and moved up to the top.

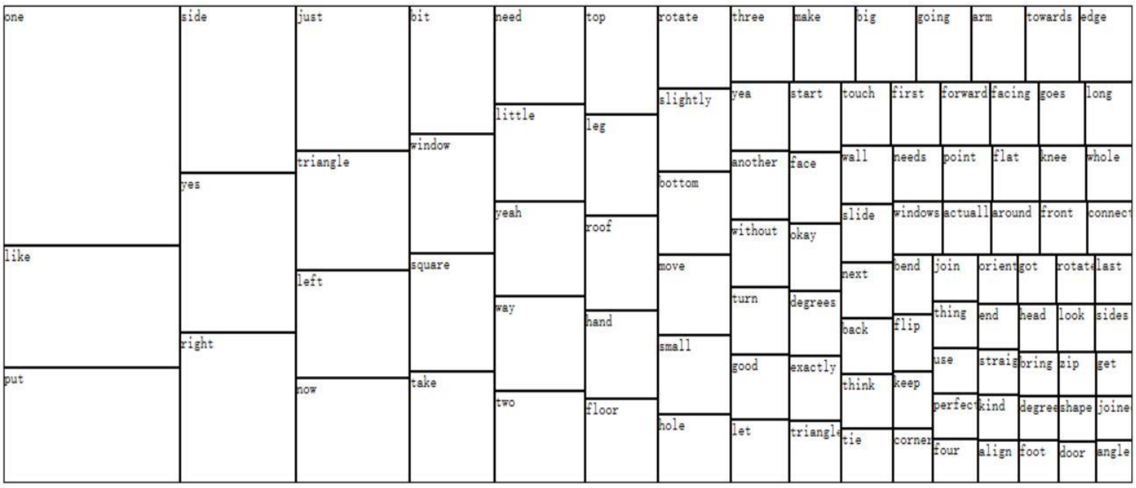


Figure 4.38. Word frequency in the experiment

In terms of word frequency, the word “one” was most frequently used, but it only appeared 5 times with the gesture (see Figure 4.38). By contrast, the word “side” was used often and accompanied by deictic or beat gestures to stress the referent’s position or direction. This can match the frequency of the gesture as well. Further, “side” is an expression of position or direction. If the trainer did not perform the gesture, the function of this word would lead to a misunderstanding, since the trainee does not know which side the trainer is referring to. The gesture in speech is an aided linguistic function for supplying key information via spatial representation. The word “like” occurred 110 times in this experiment. If the trainee did not follow the instruction, the trainer would use a “like this” gesture.

Frequently used gestures include deictic and emblematic gestures, mainly pointing at positions and objects. However, the thumbs up gesture was frequently used in each task. Almost all trainers used it for indicating correct performance; nevertheless, it lacks an instructional function, and is merely a supportive gesture. The transformational gesture was the second most frequently occurring gesture type in the experiment. Therefore, the second hypothesis regarding key gestures is tenable.

The three tasks focus on how the trainer uses spatial information via manipulation, package, and exploration. In the navigation task, the trainer simulates the hologram by mimicking and deictic gestures when directing the trainee to assemble the right body parts. The assembly task results demonstrate how the trainers package the information, while the precision task required the trainers to manipulate and explore spatial motoric information. The cardboard fort is a 3D concrete object, while the puzzle-T shapes are more abstract. Following completion of the tasks, the trainers and trainees opined that the precision task was more complex, since it required imagination, mathematical knowledge, and observational skills to explore how to rotate, translate, and make up these panels. The first and third hypotheses are therefore supported.

The non-speech conditions can facilitate gestural generation and replace utterance, but also demonstrate that the gesture is a linguistic aid in speech. When a similar shape appears later in the experiment, the trainer may ignore its features and not produce associated gestures. Further, the subsequent gestural generation can be altered by the trainer providing the trainee advance notice. If the trainer describes the shape prior to issuing instructions, the necessary number of gestures and time expenditure are reduced. The difficulty level of tasks can also influence gestural generation. For example, if the trainer adopts the same gestures for describing concrete and abstract objects, this may lead to negative outcomes and lead to more gestures being required further into the task.

4.7 The Gestural Animation

The scope of this study does not include the development of the holographic AI's model, as that is beyond the research's purview. Instead, the study focuses on examining instructional gestures that could be integrated into the functioning of an educational holographic AI.

Although the previous sections have analysed behaviours and utterance on the part of the trainer, not all co-speech gestures can be imported into a holographic AI. The three tasks in the experiment do not often appear in daily life (i.e. mimicking a person’s pose or building a specific pattern). However, they are designed to highlight the emergence of different types of gestures based on cognitive ability and various ways of thinking. Therefore, it is important to extract a series of gestures that can match the holographic AI’s persona.

The taxonomy of gestures consists of deictic, iconic, transformational, metaphoric, emblematic, beat, and cohesive types. Although cohesive gestures did not appear in the experiment, there occurred an example showing that the gesture can be appropriate during an interruption (the Rokoko suit disconnected, and the experimenter had to ask the trainer to do the collaboration pose, so the trainer used the same side hand to perform the same gesture to continue to instruct).

In order to ensure smooth motion, the animation selection is also based on the stage of gestural generation, i.e. preparation, stroke, pre-stroke, and retraction. The data analysis provides evidence showing that the deictic gesture, especially pointing at a position or object, is foundational, and that the form of this gesture consists of index finger and palm pointing. Therefore, the deictic gesture has been extracted from high-quality animation captured using the Rokoko suit and smart gloves.

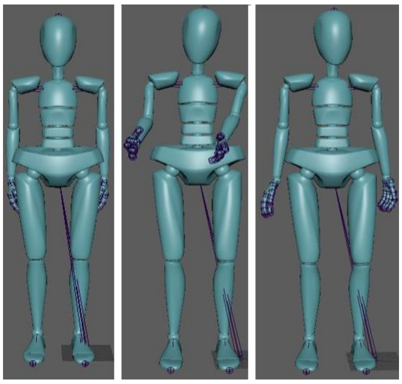


Figure 4.39 (1). The deictic gesture: two palms

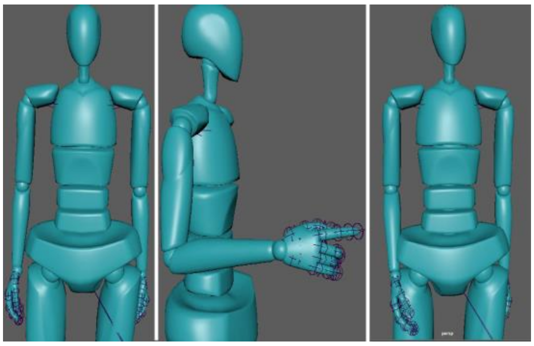


Figure 4.39 (2). The deictic gesture: the index finger

The trainer’s animations were extracted using the method of cleaning animation data, which was described back in Chapter 3. Figure 4.39 shows different forms of the deictic gesture. This gesture is used to point at an object, and was instrumental during the assembly task. If the object is on the ground, another pointing gesture is extracted from the precision task (see Figure 4.39 (3)).

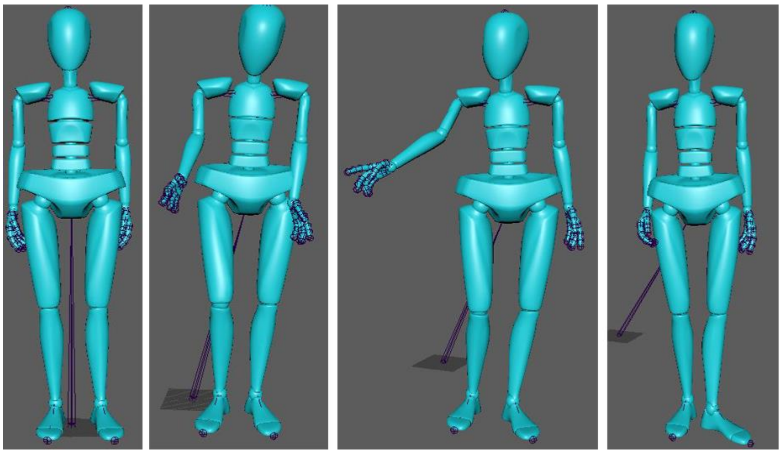


Figure 4.39(3). The deictic gesture: index finger points down

Meanwhile, the animation regarding the gesture of pointing is similar. For example, Figure 4.40 shows that the trainer requires the trainee to place a Velcro on a cardboard block.

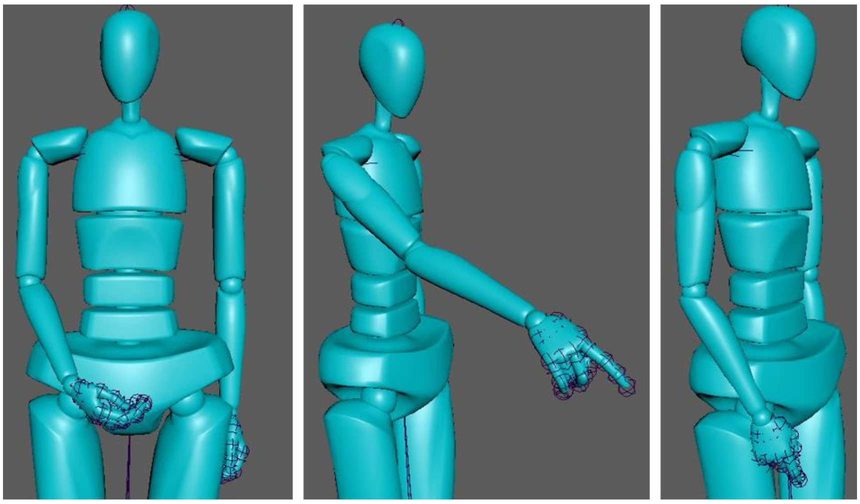


Figure 4.40. The gesture of pointing at the position

The animation of pointing in a certain direction is achieved. Some trainers preferred to use the palm to point in a direction, enabling them to employ distinct hand movements for pointing at a position and object (see Figure 4.41).

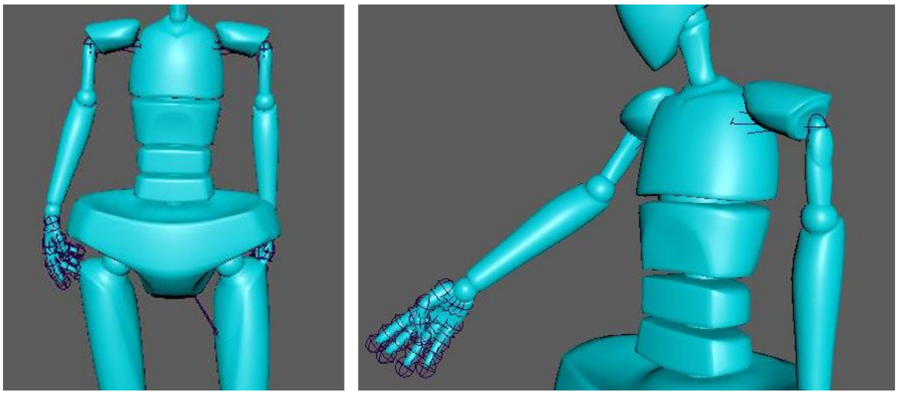


Figure 4.41. The gesture of pointing at the direction

The transformational gesture type includes rotation and flip gestures, which can represent the object's other edge or corner in the experiment (see Figure 4.42 (1,2,3,4)).

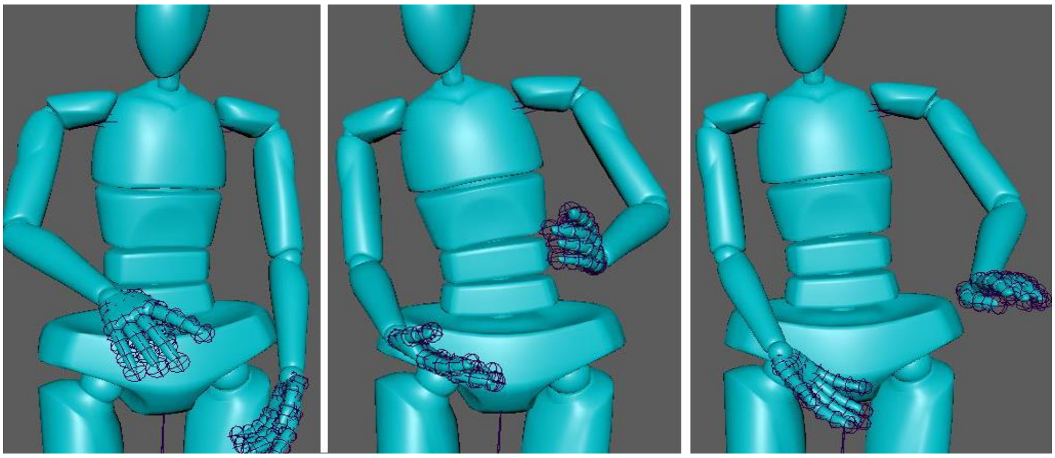


Figure 4.42 (1). The transformational gesture: rotation using palms

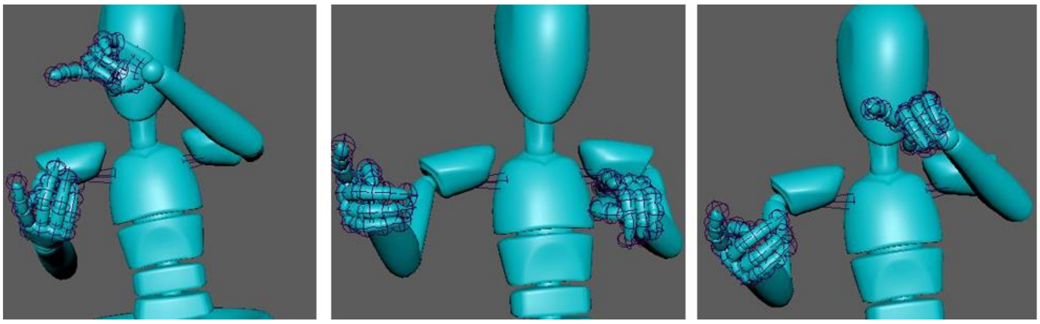


Figure 4.42 (2). The transformational gesture: rotation using index fingers

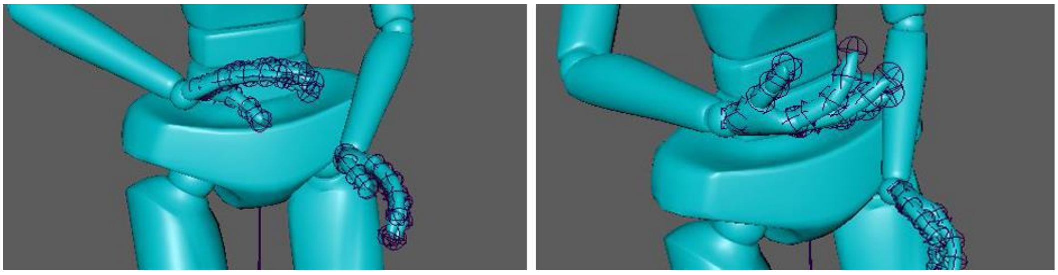


Figure 4.42 (3). The transformational gesture: flip

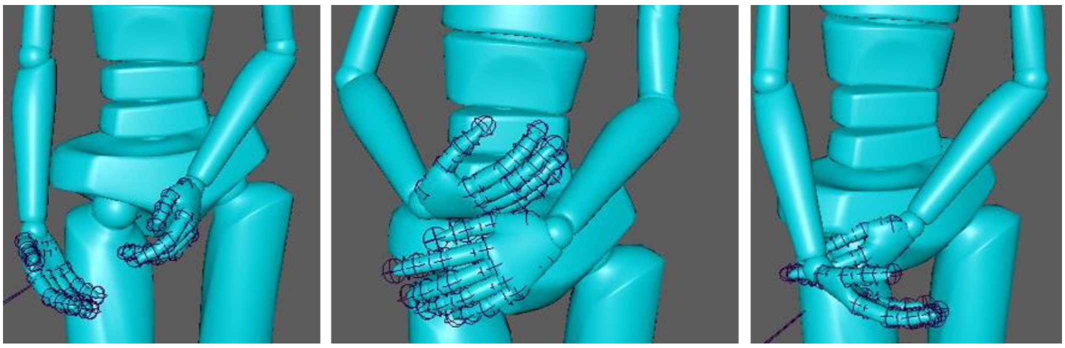


Figure 4.42 (4). The transformational gesture: other side

The iconic gesture should be based on the key object's feature. For example, although the triangle roof was often gestured in the experiment, it has particularity.

Thus, the gesture regarding the object's size and length can be widely utilised for the holographic AI's gesture (see Figure 4.43 (1,2)).

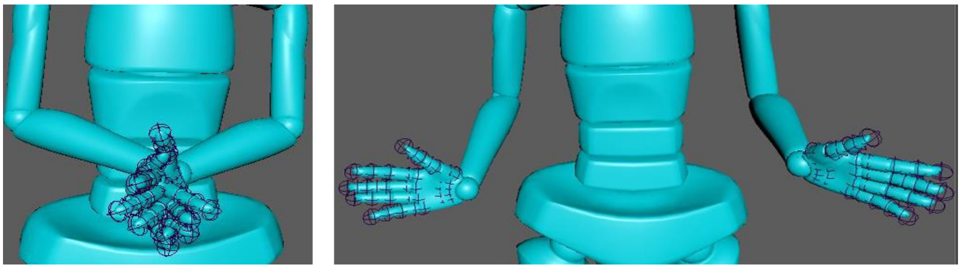


Figure 4.43 (1). The gesture described length

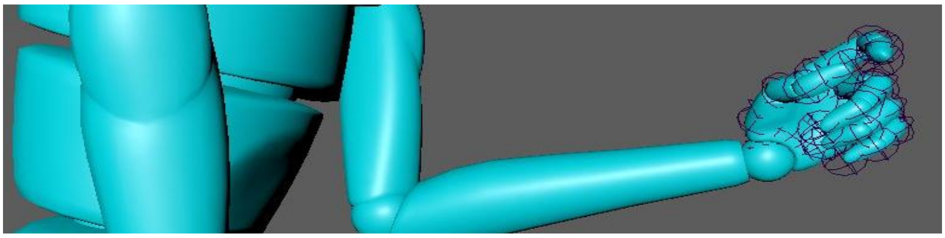


Figure 4.43 (2). The gesture described size

The thumbs up, index finger waving and palming forward gestures can convey agreement and disagreement, and they were also frequently used in the whole task, even though they are not instructional. The palming forward gesture in Figure 4.44 means holding on, whereas a waving index means incorrect performance. These gestures have supportive and reminder functions in communication.

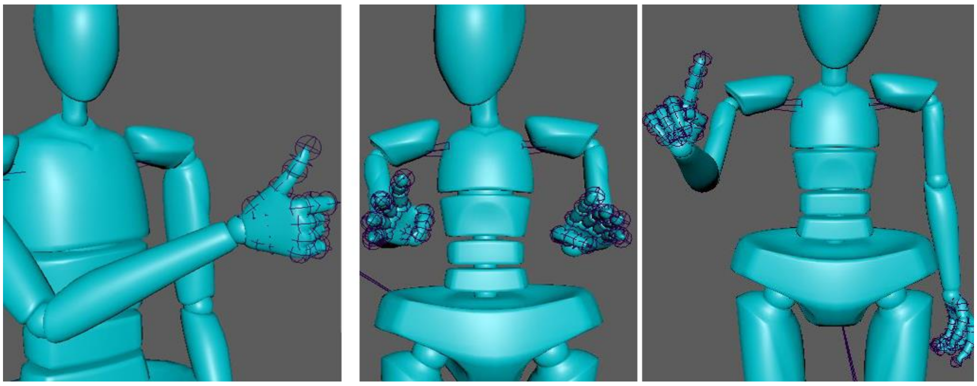


Figure 4.44. The emblematic gesture (from left to right): thumb up, palming forward, and waving index finger

Figure 4.45. The gesture in the holographic AI, including the deictic, transformational, iconic, and thumb up gestures

Figure 4.45 presents two deictic gestures, two transformational gestures regarding rotation by the palm and hands, an iconic gesture about the length, and thumbs up. The gesture is a successional movement, and the stage of stroke includes multiple gestural types and meanings. If the trainee successfully follows the instruction, the trainer's hand would change from a transformational gesture to thumbs up or the next gestural guidance, but not directly jump to the retraction or preparation. If the animation returns to the retraction step, it would be desultory since each animation only lasts a few seconds. The pre-stroke step allows keeping the gesture until the next command or utterance appears. These gestural segmentations need to create multiple possibilities based on the dialogue or the user's action.

The literature review in Section 4.1.3 suggests that holographic AI's realism can be enhanced by aligning animation, voice, and speech rhythm. A holographic AI should, therefore, evoke a deictic gesture when verbally expressing the side of a referent, and a transformational gesture when emphasizing the rotation of a referent. It is found in this experiment, however, that gestures can generate earlier than speech when a trainer describes an object with a similar function, which does not affect a listener's understanding. In order to increase behavioural realism, holographic AI can perform gestures earlier than utterances in such situation.

4.8 Limitations

This study examines the relationship between cognitive processes and the generation of gestures, with a particular emphasis on instructional gestures. Nonetheless, the experimental design and analysis are not without their limitations.

Due to the absence of teachers and students in the experiment, it is not clear whether innovative variations of gestures would emerge. Moreover, the experimental design incorporates three tasks that are seldom experienced in educational settings. These tasks were selected with the aim of optimizing the capture of a wide range of gestures for later frequency analysis. This could potentially trigger the manifestation of non-universal gestures, such as transformational and mimicking gestures. If additional tasks are incorporated, the experiment is likely to elicit other types of gestures.

Moreover, this study does not examine the potential impact of participants' educational backgrounds or teaching experiences. Since the cognitive abilities and effectiveness of learning are influenced by educational levels (Guerra-Carrillo et al., 2017). Consequently, participants assuming the position of instructors may exhibit distinct gestures compared to experienced teachers, and their limited teaching experience may necessitate increased gesture instruction. Furthermore, this experimental design does not rigorously control over the duration of time allocated for each task, nor does it account for the participants' prior experience with puzzle-T setups and cardboard fort construction assembly. These factors could lead to the repetition of gestures since experience may influence the guidance and gesture generation.

Cultural factors may have an influence on gesture generation. For instance, a Chinese participant thinks that excessive body language is impolite or friendly, whereas Italians prefer using their hands to express their thought.

This study does not provide clear evidence of how the instructor addresses and rectifies errors made by either the trainee or the instructor themselves. It solely captures partial instances of gestures associated with restarting or stopping actions. Consequently, it should be imperative for the system to produce visual representations, in the form of animations, that demonstrate the holographic AI's execution of distinct gestures within this context.

Therefore, this study should manipulate variables like cultural contexts, personal experiences, and educational backgrounds, it could enhance the classification of gesture types and reveal foundational aspects of gestural generation.

4.9 Summary

Although the PICS model outlines the role of body movement within the persona dimension, it lacks detailed insights into the instructional gestures that could be integrated into holographic AI systems. Hence, this chapter augments the PICS model by investigating the instructional gestures suitable for holographic AI, as illustrated in Figure 4.46.

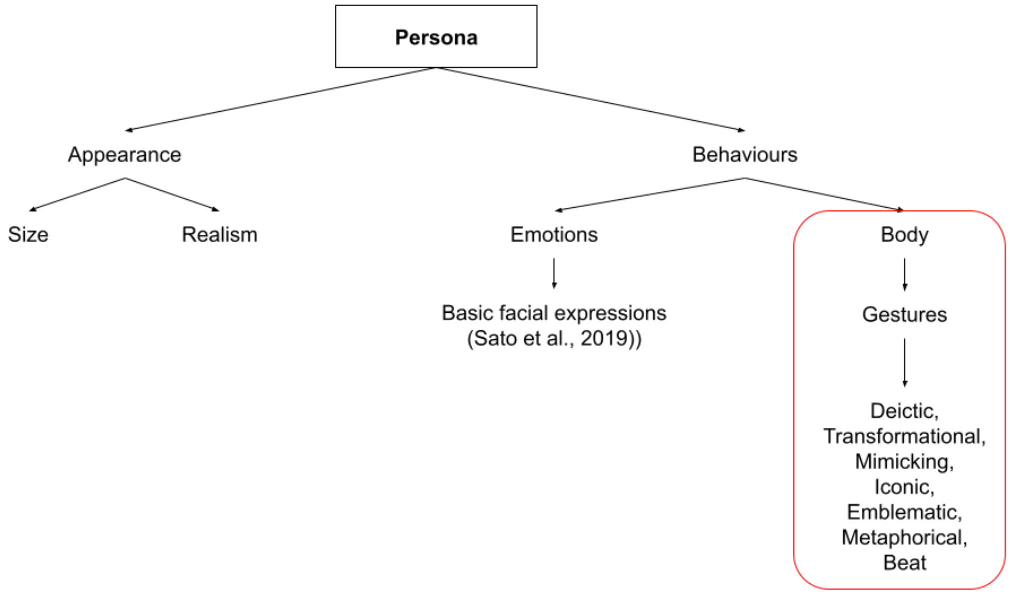


Figure 4.46. Instructional gestures in the PICS model

An experiment was implemented to collect and investigate data of gestural animations. It included three tasks: navigation, assembly, and precision. The experiment involved 22 participants divided into 11 groups. In each group or pair, one participant was a trainer who captured movements, and the other was a trainee. For the navigation task, each trainer instructed the trainee to mimic the hologram's pose. The assembly and precision tasks entailed building cardboard forts and puzzle-T patterns based on the trainer's instruction. Each task has two conditions, i.e. speechless and verbal, and the speech condition comprised three levels of difficulty.

This experiment recorded videos of the trainers' and trainees' performances, as well as the trainers' movements, via motion capture. Then, 132 videos were imported into NVivo12 to be coded by different gesture types, based on the taxonomy of the representational gesture, which includes iconic, deictic, metaphoric, emblematic, beat, and cohesive types. This experiment also focuses on transformational gestures, which are used to manipulate spatial information. Gestures relating to rotating, flipping, and expressing the other side of the referent are included. The transformational gesture neither refers to the object's position nor is a vague orientation. The purpose of the navigation task was to mimic the hologram's pose, thus requiring the trainer the unconsciously simulate, i.e. perform a mimicking gesture.

Each type of gesture occupies a different branch in the taxonomy. The deictic gesture consists of pointing at an object, position, and direction. The index finger is integral to the gestural form of pointing at a specific referent, and the palm also can point towards a position. The pointing gesture is not a static animation, such as pointing out and going back to standing. When a person uses this gesture, his/her forefingers may repeatedly point forward or draw circles to emphasize its importance or position. The object's size, length, and shape can be visualized by the iconic gesture. When the trainer gestures the object's size, the thumb and index finger form a ringlet circle to describe this size of area, while the length is a 2D line, which is different from the definition of size. This type of gesture also can represent the spatial position relationship of multiple objects, as well as describe an angle. In contrast to pointing at a direction, illustrating an angle does not require using the index finger to put the referent on an orientation, so the trainer must understand the two objects' positions in order to form an angle.

The emblematic gesture is characterised by cultural uniformity in that people of different cultures can identify the gesture's meaning, such as thumbs up. In order to convey right or wrong results, the trainer uses this hand movement to differentiate and pinpoint a key feature of the referent, such as the panel without a window. The metaphorical gesture requires imagination and verbal content to judge, thus this gesture did not appear in the non-speech trials, and the trainer used this to compare previous tasks. The trainer can use the beat gesture to organize utterances. Cohesive gestures are used to connect previous interpreted expressions; however, this this experiment did not display cohesive gestures.

On the other hand, a single gesture may convey multiple meanings. The trainer's hands describe the shape's size and then rotate the wrist to navigate the trainee to follow. In the assembly task, the trainer's hand mimics the cardboard's shape by opening arms and having the two palms facing each other, while the on-hand palm gesture represents the puzzle-T panel. In the medium level of the assembly task, iconic and deictic gestures mutually appear when the referent's feature can be represented by the palm and moved to a direction or position.

The results proved that the deictic gesture is the most frequently utilised type for instruction. The speechless and complex situations can facilitate the gestural generation, but if the trainer provides advance notice, such as describing the shape of the puzzle-T pattern prior to instruction, the required number of gestures will be lower. It is probable that the trainee can predict what the next shape is, and also its difficulty. The three tasks were also designed to examine the trainers' different ways of thinking via gestures. The assembly task focuses on how to package spatial information, while the puzzle-T is more abstract, requiring the trainer to manipulate and explore the information. However, if a concrete and visualized object is described in an abstract manner, the trainer needs to employ more gestures to explain its features.

In terms of co-speech gestures, when the trainers were permitted to speak the gestures turned into a linguistic aid. Each trainer verbally instructed each trainee, instead of relying solely on gestures, and if the same or similar elements appear later, each trainer did not need to focus on the corresponding gestures again or apply new spatial representation. Moreover, the word “side” was the most frequent one appearing alongside the deictic gesture. This word can depict a position and direction. Besides, the speed of gestural generation is quicker than that of verbal organization. People can recall the previous schematization and the referent’s information, and then its feature can be extracted by way of a gesture; whereas language organization is more complicated since it involves grammar rules, lexical retrieval, and logical thinking.

Following the data analysis, the corresponding animation was then selected based on the taxonomy of gesture and frequency. Considering that the animation should match diverse surroundings, the iconic gestures in regard to specific patterns or shapes, such as a triangle or square, are not reproduced. Besides, the branches of the deictic gesture type share no explicit differences, representing by animations. However, it has been observed that when a trainer points at a position or object, the shape of the index finger is clearer; whereas if pointing at an indefinite orientation, the index finger is loose, or the palm of the hand may be used for translation. When the holographic AI points at a position or object, the index gesture can be utilized, while the palm can refer to an orientation. It was also observed that during the speechless condition of the assembly task, if the trainer uses a single hand to manipulate spatial motoric information, the trainee may not be sure whether the cardboard needs to be flipped or rotated. Therefore, the hand animation should signify a clear difference in the gestural representation. In order to ensure gestural diversity, two hands spaced far apart from each other can represent a large-sized object, while one palm can reflect a smaller one. The iconic gestural animation should be based on the pre-defined object. If the scenario has a specific key virtual or physical object that is central to the topic of the training session, the corresponding shape of the gesture should be represented. The emblematic gesture, especially the thumbs up and waving hands, can express the holographic AI’s approval or disapproval, accompanied by head nodding. Inclusion of transformational gestures depends on whether the dialogue includes the object’s manipulation, and inclusion of mimicking gestures relies on whether the holographic AI needs to package spatial information or separate different blocks to describe a plan or shape in the complex situation. Metaphorical gestures can illustrate and compare previous events or behaviours in interaction via the movements of the left and right hands. In order to “plug the gap” within the speech output delay, gestural generation can be performed prior to utterance.

While transformational and mimicking gestures are able to convey how people or holographic AIs manipulate and package spatial information, the experiment's main focus has been on instructional gestures. Furthermore, past studies emphasize the importance of deictic gestures and their use in training and education, as well as their integration with other forms of gestures. Based on observation, the stage of gestural generation may help connect different types of gesture. When the trainee cannot immediately react, the trainer's gesture should occur in the pre-stroke or retraction stages, and he/she should either explain it again or provide other instructional methods. The trainer will generate the next gesture if the trainee has (not) not followed the instruction correctly. There may occur a pause between strokes. Nevertheless, if the trainee needs to spend a long time on the task, he/she will go through the stages of stroke, retraction, and stroke again. The user's reaction to the holographic AI is therefore crucial.

The holographic AI’s cognitive ability and intelligence level can be represented by the gesture. Although natural language processing is the main interactive approach, the gestural generation can be an aiding system that enhances the holographic AI’s persona. A series of gestural animations could satisfy the requirements of an educational holographic AI teaching mathematics, since this will likely include counting fingers, and 2D and 3D shape representation. Therefore, the next chapter will investigate whether children are likely to place their trust in such a holographic AI and its instructions.

Table 4.9 provides the utilization of different gestural animations for the holographic AI below:

Gestural taxonomy	**Gesture animation**	**Describe use**
Deictic gestures	Pointing at the position using the index finger.	Refer to a concrete location.
	Pointing at the object using the index finger.	Refer to a concrete virtual and physical object.
	Pointing to a direction using loose index finger or palm.	Translate an object or mention a vague orientation, accompanied by the utterance of “side”.
Transformational gestures	The wrist rotates and flips. The hand jumps forward.	To manipulate spatial information.
Mimicking gestures	Simulating a posture	To mimic the other one’s pose.
Iconic gestures	The index finger draws a shape based on the object.	To represent the object’s shape.
	The distance between thumb and index finger.	To describe size.
	Stretching two arms from clasped hand palm.	To represent 2D shape’s length.

	Both upper arm and lower arm aligned vertically, or align the two hands vertically.	To represent an angle using two hands.
	Putting grasping hands or palm down, and moving to different position.	To package and explore different blocks' spatial positions and relationship.
Emblematic gestures	Thumbs up, waving the index finger or hands, palm facing forward, counting fingers, crossing-arm movement.	These gestures have a prompting effect.
Metaphorical gestures	The hands move from left to right sides.	To represent past and present/ future time.
Beat	Gesture with rhythmic movement.	Such a gesture can help organize verbal language.

Table 4.9. The gestural taxonomy can be used in the holographic AI

Chapter 5 Trust towards Holographic AIs: An Experiment

5.1 Introduction

This chapter has a particular focus on trust: what trust is, what encourages a user to place their trust in a holographic AI, and whether and to what extent anthropomorphic 3D character models fitted with AI technology can elicit free-flowing, intelligent, spoken dialogue and engender within the user a sense of trust. More generally speaking, the aim of this chapter is to understand the relationship between human and computer in order to improve technology acceptance and the user experience.

Trust is an essential pivot of interpersonal relationships and thus forms the fabric of society. AI agents bear the semblance of real humans and may be regarded as social entities and actors (Borst and Gelder, 2015); in this context, trust to a large extent, defines a relationship between users and technological artefacts. Although HCI shares some of the similar principles found in interpersonal interaction (Reeves and Nass, 1996), such as emotional expressions and recognition, the requirement of trust in artificial agents is higher than that in human-human interaction.

For example, photorealistic, anthropomorphic 3D character models are widely used in immersive technology-based applications such as video games, training simulations, or augmented reality. These 3D characters can, for example, guide the user in solving a problem, or they can be used as a sort of user interface designed to deliver services. Especially in regards to critical services such as healthcare or education, technologists desire, of course, to develop tools which engender a high degree of trust.

The previous chapters have described holographic AIs with instructional, gestural animations. This chapter concerns a holographic AI that plays a role of a mathematics teacher that helps children identify different 3D and 2D shapes, such as a cube, cuboid, square, and rectangle.

As discussed earlier in Chapter 2, although holographic AI can positively influence children’s learning outcomes (Li et al. 2021, Oh and Byun, 2021), methodologically, user experience evaluations for holographic AIs lack a measure for trust. This is a gap in the literature and leaves open to test whether trust plays an important role for human computer interaction with intelligent tutoring systems, leaving it unclear whether trust is needed for providing a positive user experience.

To date, there does not exist a standardised scale for testing a human’s trust towards a holographic AI. Although Kim et al. (2018) have measured the sense of trust towards holographic AIs, their questions relating to measurement were extracted from the McKnight Trust Questionnaire (Mcknight et al., 2011). This questionnaire focuses on trust in information technology (IT), and offers a framework of trust in IT that differs from interpersonal trust. However, the authors measured students’ experience towards functions of Excel’s features, rather than directing at embodied virtual agents.

Hence, a novel methodology for measuring trust will be developed in this chapter and an according experiment will be presented that investigates the sense of trust towards this particular holographic AI, the geometry tutor, to exemplify what role trust plays in the users' experiences.

Section 5.2 reviews personal trust and relative potential factors of trust that influence HCI. Section 5.3 develops a specific questionnaire for measuring trust towards the holographic AI. Section 5.4 details the questionnaire adaptive to children and presents a pre-test. Section 5.5 offers the experiment design. Section 5.6 describes the experiment for analysing children's trust towards a pedagogical holographic AI. Section 5.7 provides the experimental results, and Section 5.8 discusses which factors appear to influence user experience. A summary of the findings of this chapter is provided in Section 5.9.

5.2 The Concept of Trust

This section sets out the definition of trust in the context of human-human communication and explores the differences between human-human trust and human-AI trust. Although the concept of interpersonal trust has been discussed, it is crucial to understand what the sense of trust generates between the human user and AI, identify factors of trust, and consider whether these factors based on human interpersonal relationships and can be utilized for holographic AIs. Another issue to consider is whether the requirements of trust may change along with social and science development.

5.2.1 Interpersonal trust

There are diverse definitions of interpersonal trust in different aspects. It can be an attitude (Jones, 1996; Helm, 2014), expectation (James, 2002), a psychological state (Rousseau et al.,1998), a belief (Reiersen, 2017). For example, trust as an attitude can be defined as a person's goodwill and capability in satisfying another's communicative requirements. In this context, the person is expected to have corresponding behaviours that can be relied on (Rotter, 1967), and which can help the other person achieve goals (Lee and Katrina, 2004). Dietz (2011) also defined trust as an expectation, such that the trusting individual is willing to take a risk and face its associated uncertainty. These definitions indicate that the interpersonal trust is characterised by a sense of expectation and hope on the part of the trustor based on the trustee's intention. Therefore, trust can be defined as an intention or willingness of someone to put oneself in a vulnerable situation (Mayer et al. 1995; McEvily and Tortoriello, 2011).

Further, trust can also be defined as an assumption based on competence, intention, and perception, i.e. the trustor believes the trustee will not thrust him/her into a risky situation (Pearce, 1974). Trust in this definition implies that the trustor predicts the trustee's behaviour. Elster (2007) proposed that trust is a unique action that is based on what the trustee does, which even when people do not trust a person, they might still perform such an action.

On the other hand, Hardin (2006) claimed that trust is not an action, but is instead a measurable indicator, a scale of belief in regard to expectation from the trustee, that it reflects what the trustor thinks of the trustee (Reiersen, 2017). Belief in the concept of trust refers to a certain degree of a person's trustworthiness, which can decide whether the next action will occur (Hardin, 2006; Chang et al., 2010). However, belief is one of source of trust, and trust itself is an integral behaviour within a personal relationship. For example, if a person does not trust another person owing to different stances, motivations, or attitudes, he/she will not act; accordingly, ipso facto, if the trustor's belief is that the trustee is trustworthy, then he/she will act. Therefore, the meaning of trust extends beyond that of a belief, and these multiple factors, such as belief and intentions, can translate into an action (Reiersen, 2017). Trust neither is a behaviour nor a belief, since there are other factors that affect action. Colquitt, Scott and LePine (2007) proposed that trust is a behavioural outcome, and that this action is derived from putting a person in a situation, in which perception of vulnerability precedes trust generation. Perceived vulnerability stems from an uncertainty of the trustee's motive and intentions (Kramer, 1999).

The definition of trust characterized by Mayer et al. (1995) became a broadly accepted and employed one (Rousseau et al., 1998; Lee and Katrina, 2004). Trust is willingness of the trustor to accept vulnerability based on a positive expectation that the trustee can perform the appropriate behaviour (Mayer et al., 1995). Moreover, Mayer proposed three essential elements of trust: competence, integrity, and benevolence (ibid) (see Figure 5.1).

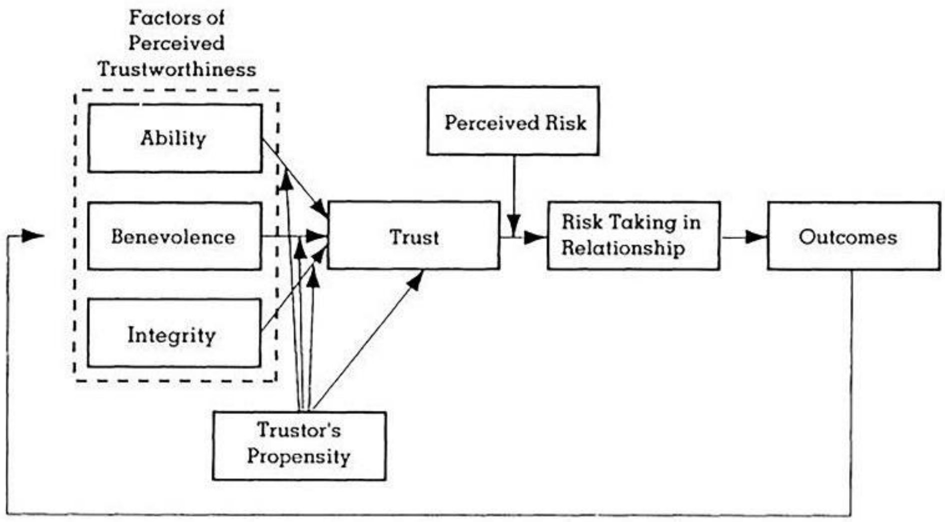


Figure 5.1. Trust model (Mayer et al., 1995)

Ability in interpersonal trust is the prior element, which requires that the trustee has a specific skill or competence to achieve a trustor's goals. The influence and perception of ability is intuitive (Mayer et al., 1995). Benevolence is the degree to which a trusted person intents to do good for the other. This is one of the most important aspects of trustworthiness, especially in the educational domain (Tschannen-Moranand and Hoy, 2000; Di Battista, Pivetti and Berti, 2020), since it should ensure that the trustee will not harm the trustor. A teacher's ability and benevolence are key for the students to perceive the sense of trust (Tschannen-Moranand Hoy, 2000). In terms of integrity, the trustee should follow a principle or social norms that are acceptable to the trustor (Mayer et al., 1995), such as honesty. According to Dobel (1999), integrity can be defined as a consistency in behaviour and promise. Therefore, the trustee's action should be aligned with his/her utterances and thoughts.

Although they consist of distinct dimensions, perceived trust requires a person should put a trustor's interest in priority and follows social norms to fulfil the other's goal based on a specific competence in a domain, and for the trustor, he/she will be vulnerable to take risk (Mayer et al., 1995).

The holographic AI mimics human characteristics – such as persona, sensory perception, and cognitive capacity – within a MR environment, equipping the AI with analogous attributes. Consequently, certain aspects of Mayer et al.'s trust model align with facets of the PICS model, particularly competence, which is associated with the holographic AI's demonstrated abilities. Integrity in this context refers to the holographic AI's consistent performance and the achievement of satisfactory results, gauged by the AI's adherence to its programmed commitments. However, the holographic AI inherently lacks empathy, even though it can simulate emotional expressions and offer personalized interactions. The holographic AI's intelligence is restricted to its programmed functions, limiting its capacity to rival human intellect in navigating complex, varied scenarios. Moreover, the functionality of the holographic AI is contingent upon user input, and it does not independently verify the correctness of user-provided data. The user's engagement is typically a one-way emotional response. Understanding the nature and dimensions of trust in relation to holographic AI is therefore paramount.

5.2.2 Trust towards virtual agents

Interpersonal interaction is not only based on verbal/non-verbal expressions and backgrounds connected with one's perception of another's potential cues, but also relies on subjective feelings and objective evaluation. In terms of HCI, holographic AIs have similar interactive approaches in terms of natural language processing, animations, and sense. Effective holographic AIs are treated as real humans, social entities and actors (Cassell et al., 2001; de Borst and de Gelder, 2015). Trust is also a critical determinant in the measurement of usability of technology (van Pinxteren et al., 2019). Although HCI shares some of the similar principles found in interpersonal interaction (Reeves and Nass, 1996), such as emotional expressions and recognition, the requirement of trust in artificial agents are higher than those in human-human interaction, because humans are sensitive to these anthropomorphic artefacts that are not identical to those of humans. Therefore, the sense of trust in a holographic AI may differ from interpersonal trust.

Kulms and Kopp (2018) define a user's trust in HCI as an attitude based on whether the agent is able to complete tasks based on the user's intention. Similarly, Lee and See (2004) also defined trust as an attitude to achieve a user's goal. The perception of the agent being trustworthy and able is based on cognition, and such cognitive trust is dictated by the agent's competence and functionality (Mcknigh et al., 2011). Further, it has been proved that emotional trust is also an important factor of trust (Lee and See, 2004). Emotional trust is different from cognitive trust. Users can experience cognitive trust when observing the virtual agent's performance, but not emotional trust (Seitz et al., 2021). On the emotional dimension, affective trust on the part of the user is determined by the concern and care shown by the agent (Borum 2010), which can be referred to as benevolence, i.e. a willingness to assist users independent of self-interest.

Therefore, the sense of trust towards a holographic AI also has similar factors: competence, and benevolence. Holographic AIs are said to be competent when they possess the abilities and knowledge necessary for executing and completing tasks. Benevolence indicates that HCI agents which execute tasks based on the user's interest are said to display this trait (Phillip et al., 2020). Besides, integrity can also refer to the interpersonal trust model. Holographic AIs which assume full responsibility and fulfil promises can be said to exhibit integrity (ibid).

The definition of trust exhibited by children is similar to that of trust exhibited by adults. The former is defined the child's confidence in a person's speech and behaviours, and it also the expectation that the trustee can fulfil promises (Imber, 1973). Attachment, personification, social realism, and humanoid requirements are the elements of children's preferences for virtual characters (Richards and Calvert, 2016). This represents the virtual character should have a human-like persona, social norms, emotions, behaviours, and competence. Although children's social presence is lower than that of adults during interaction with a robot, they treat it as 'human' instead of a toy (Guneysu and Arnrich, 2017). Further, children have high expectations regarding friendship in trust, and these expectations may be similar in their sense of trust in robots (Calvo-Barajas and Castellano, 2022). In terms of children's trust towards technology, trust occurrence is based on internal state, which in turn includes experiential and cognitive categories (van Straten, Peter and Kühne, 2020). The experiential state focuses on how children affectively interact with a robot, and includes the elements of engagement, enjoyment, and emotional arousal. The cognitive state is based on how children perceive the robot, i.e. person, and relies on co-presence, and support from the robot. The robot's responsiveness interaction is a factor that influences the sense of trust, and which contributes to the AI's realism and humanness (van Straten et al., 2020). Additionally, there are two dimensions of children's trust: social dimension, and competency (Calvo-Barajas and Castellano, 2022). Social dimension refers to moral features, such as honesty, and competency is the robot's performance, which affects cognitive trust. The authors designed a questionnaire based on these two dimensions to measure the sense of trust towards a virtual agent (ibid). It was found that the children aged 9–12 were able to evaluate both trust dimensions, and that they hold similar views in terms of trust towards and perception of the virtual agent over time.

The appearance and performance of technological artefacts can affect one's perception of AI features, and AI technology has an impact on trust generation (Gkinko and Elbanna, 2023). The holographic - human trust relies on interpersonal trust, and it requires a goal-oriented service for users. Children's sense of trust towards physical robots is similar to that of adults. Physical artefacts also require competence, social norms (i.e. integrity), support and concern (i.e. benevolence). However, few studies have considered how virtual humans or holographic AIs influence children's sense of trust, and mapped out associated factors. Further, scales for measuring children's trust tend to be derived from research on HCI rather than that focusing on virtual humans. Therefore, the following sections will develop a trust questionnaire for holographic AIs, and explore children's perception.

5.2.3 Research questions
This study aims to examine the factors that contribute to the development of trust in an educational holographic AI (i.e. RQ4). Trust can be defined as a set of different properties in different fields and contexts, and there is no uniform and accurate definition of trust towards holographic AIs. Therefore, it is crucial to comprehend the elements of trust in the holographic AI. Furthermore, existing research lacks a standardized instrument for assessing users' trust in such agents. Therefore, it is imperative to develop a metric that can quantify users' trust in holographic AI systems. Few studies have explored how holographic AIs influence the degree of trust and user experience. Although some studies have brought forth evidence of some factors significantly affecting the sense of trust, there is no agreement as to which factors are critical or have potential to increase the user's trust. These may be divided into three main aspects – holographic AIs and trust – upon which the following research and sub questions are posited:

- Definition - What is trust towards the holographic AI?
- Scale - How to develop a novel scale for measuring the sense of trust towards the holographic AI?
- Factors - What factors influence the degree of trust? (RQ4)

5.3 Developing a Questionnaire to Measure Trust

This section details the development of the questionnaire, the aim of which is to assess the user's trust in holographic AI. Since the questionnaires used in previous studies concerned trust specific to technology or human interaction, rather than holographic AIs, the development of this questionnaire addresses a knowledge gap in the field.

Trust in virtual agents can be defined as an attitude driven by the assumption that the agent is capable of and willing to fulfil the user's expectations (Kulms and Kopp, 2018). Although Lee and See (2004) defined a user's trust towards an automaton as the attitude based on the belief that the computer agent can help the user achieve goals, this definition might not be directly applicable to humanoid holographic AIs. For this reason, it is important to arrive at a proper definition of a user's trust towards a holographic AI before attempting to develop a tool for measuring the level of trust in this context.

It is also necessary to identify relevant constructs for predicting the level of trust. One simple approach is to adopt to traditional Likert scaling method. In this project, an 11-item Likert scale has been developed and validated. It is also worth considering whether these items are interrelated to, and may supplement, one another, in the design of a more innovative tool for quantifying the degree of trust in a holographic AI.

5.3.1 Related work
One recent study of trust in virtual reality-based agents was conducted by Gupta et al. (2019), who focused on cognitive load level, and in which they developed a subjective mental effort questionnaire designed for obtaining physiological sensor data including heart rate variability and galvanic skin response.

Data concerning human feelings can be obtained via surveys, interviews and competitive research tools; however, feelings constitute a subjective quality which cannot easily be quantified. The Likert scale is frequently used in psychological and social studies for interpreting numerically the severity and dynamic nature of people's feelings and opinions. One such tool is the trust scale, which assigns subjective statements to semantics, and converts people's attitudes and feelings to a rated value using the common five-point scale ranging from 'strongly disagree' (1) to 'strongly agree' (5) (Borum, 2010; Piemetel, 2010).

There are similar forms a Likert scale - the response scale, and statement-based scale – both of which are five-point scales. The statement-based scale has been outlined above, in that participants are asked to indicate the extent to which they agree or disagree with a particular statement. The scale separates the different tiers of (dis) agreement by assigning interval values, to obtain unbiased results. The response scale is similar, in that participants are asked to indicate their attitude towards a statement or entity: the scale ranges from 1 ('strongly unfavourable') to 5 ('strongly favourable'). Both scales include a central tier (3), giving the participant the option of expressing an undecided or neutral feeling or decision and thus avoiding bias (Kocaballi, Laranjo and Coiera, 2019; Bryman, 2012).

Using a Likert scale, Hanna and Richards (2019) studied users' impressions and comprehension of, and experiences with, a computer agent, in an effort to measure their experience of and interactions with computer agents, as well as gauge their feelings of trust in the virtual assistant. Their study involved 73 undergraduate students, and the authors monitored the participants' computer usage-based behaviours such as keystrokes and inputs. The three variables in their study were trust, performance and promise.

A study of the 'warmth' expressed by computer agents and its influence on user's trust was conducted by Kulms and Kopp (2018) using an altered Likert scale. Instead of focusing on the user's sense of agreement, their scale focused on positive qualities such as goodness, honesty, trustworthiness and good intentions. To reduce bias, a Likert scale-based study should involve a number of participants and obtain data sufficient enough to counter extreme opinions. The wording of statements in a Likert scale-based study must be clear and context-based enough to enable the user to select the correct choice. Although users' feelings cannot be quantified, it is possible to measure them crudely by observing their reactions, behaviours and decisions (Pimentel, 2010).

In order to develop holographic AIs which are demonstrably trustworthy, reliable and capable of making seemingly smart, intuitive decisions, it is necessary for researchers to determine exactly how trust facilitates human-AI interaction. To date, research and development (R&D) progress has been impeded by the wide range of definitions of interpersonal trust, and paucity of valid models for quantifying such trust. In view of these obstacles, in this study a panel of judges has been tasked with identifying and refining trust-related items by evaluating their validity and precision, together with identifying polarising items that reveal more distinctive characteristics of trust.

5.3.2 Methodology: a new scale for ‘trust’

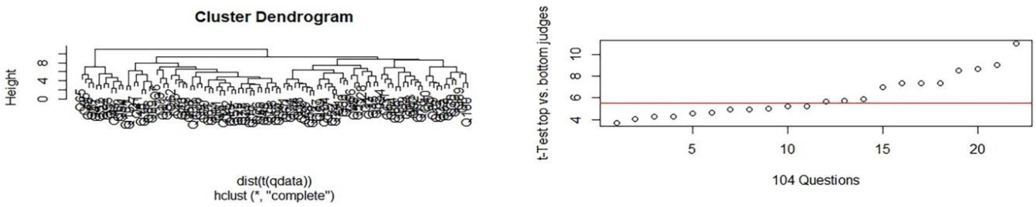


Figure 5.2. Correlation with sum scores (Huang and Wild, 2021)

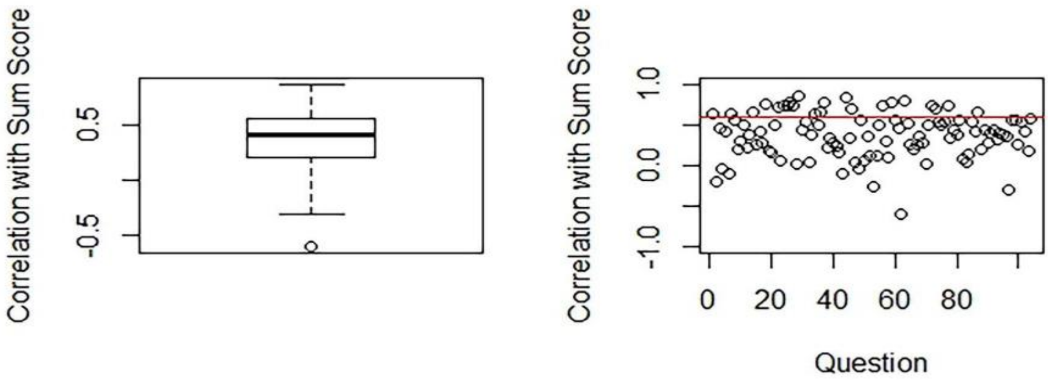


Figure 5.3. T-values of top and bottom quarter judges (Huang and Wild, 2021)

There does not appear to exist a validated tool for measuring a user’s trust in holographic AIs. Therefore, this study takes an ambitious step by proposing a novel scale for measuring trust in this context, based on the Likert scaling methodology (Trochim, 2021). This new scale employs items cited in previous studies, which have been discussed and extended via brainstorming sessions among the researchers (Huang and Wild, 2021) involved.

Subsequently, 15 'judges' were appointed to rate items using a scale from 'strongly unfavourable to the concept' to 'strongly favourable to the concept'. The process of item selection entailed the elimination of statements which did not appear to correlate closely with the sum scores of all statements, as well as interchangeable items. In order to form an administrative scale composed of optimal statements (e.g. ranked at>0.7), lower-value items are removed (Roberts, Laughlin, and Wedell, 1999). The most suitable items, i.e. those delivering polarized answers, were identified using a t-test of top and bottom-quarter answers.

A total of 104 statements relating to the user’s trust in holographic AIs was extracted from the literature, and scrutinized during the brainstorming sessions. The pool of judges then evaluated the polarizing and loading efficacies of the statements, and narrowed down the number of statements to a more administratively reasonable level.

Items whose correlation scores (between item and sum scores) were <0.6 were eliminated, and the final number of statements was 22 (see Figure 5.2). Using direction as a proxy for trust, constructs not directly loading onto the general direction of all statements were eliminated.

The t-test was employed to determine the mean of ratings given by the top quarter of judges (who assigned the highest scores). In an effort to determine the degree of polarization, these ratings were compared against those of judges occupying the lowest quarter (who assigned the lowest scores). The higher the t-value, the wider the difference between the views of the two groups of judges, and the better the capacity of the item to discriminate and separate people's views. This filtering process demands some intuition on the part of the researcher, as recommended by other scholars. It was decided to establish a t-value threshold of 5.5, above which 11 question items were isolated (Figure 5.3).

Cluster analysis was used to determine whether significant items were missing. This involved hierarchical clustering covering the Euclidean distances among questions, using hclust in R Stats package developed by the R core team (2021). The cluster hierarchy in the dendrogram was shrunk by clustering the questions into k=20 groups, and the level of homogeneity of the resulting clusters was analysed by way of visual analysis.

In an additional effort to detect missing items, the resulting groups were re-analysed manually. It was found that some of the eliminated groups appeared potentially relevant, but on closer inspection were lacking in correlation with sum scores (<0.6). The questions in these groups tend to focus either on multiple aspects including trust, or on aspects not specifically concerning trust (see Table 5.1).

#	Item
Competence	
1	The hologram is competent.
2	The hologram is very skilled.
Integrity	
3	I think positively about the hologram
4	The hologram answered my questions truthfully.
Benevolence	
5	I think the hologram wants to do good.
6	The hologram is benevolent.

Compassion	
7	The hologram feels real to me.
8	The hologram looks out for me.
9	The hologram was committed to helping me.
10	The hologram is compassionate.
Relationship	
11	The hologram and I created a good relationship.

Table 5.1. Final trust scale for measuring holographic AIs (Huang and Wild, 2021)

In summary, in order to arrive at a set of statements which elicit the most polarized responses in relation to the user's trust in holographic AIs, it is advisable to apply statistical tools to segregate 'raw' statements into groups so that statements which concern aspects other than trust can be eliminated.

Thereafter, following the method by Watts (2020), the questionnaire containing the remaining statements was pre-tested with the assistance of a small group of students (n=5). To ensure a uniform polarity, all the questions are phrased as positive statements, and each provides the participant with a 5-point scale of responses from 'strongly disagree' to 'strongly agree'. The purpose of the questionnaire is to explore the user's feelings and opinions associated with their interaction with a holographic assistant.

5.3.3 Discussion

The final selection consists of distinct items, all of which concern critical elements of trust, and which could yield more meaningful data for determining users' perceptions on the trustworthiness of holographic AIs. In line with earlier models explored in the literature review, these items can be grouped along the following constructs: competence, integrity, benevolence, compassion, and relationship.

Human action is founded on competence, which is also a fundamental determinant of performance (Wild, 2016), which in turn determines the user's sense of trust. Competence as a quality draw on knowledge and skills, which to a certain extent are domain-specific (Hager and Gonczi, 2009). Another instrumental component of trust is competence. According to McLeod (2020), "trust requires that we rely on others to be competent". For example, chatbots are expected to possess good communication skills. It is essential for an AI agent to possess the necessary competencies for determining the user's expectations and helping the user realize their goals. Statements #1 and #2 can be used to evaluate the user's optimism of the AI agent's competence and properties thereof.

In reference to Mayer et al. (1995), statements #3 and #4 in the above table concern integrity, one of the components of trust. According to the interpersonal trust model, integrity is whether an executor is able to adhere to principles that a trustor can accept (Mayer et al., 1995). In this context, the executor, also known as the holographic AI, while the trustor is the user. The set of principles pertains to the interaction protocols that users agree to when engaging with a holographic AI system. As a consequence, users perceive that the holographic AI adheres to predefined protocols to accomplish goals (#3). Additionally, the performance of the holographic AI can be synchronized with its responses, providing accurate answers to user inquiries (#4).

In reference to Mayer et al. (1995) as well as Sousa, Lamas and Dias (2014), statements #5 and #6 refer to benevolence, i.e. the disposition of the agent to doing good (Urbano, 2013).

Compassion in interpersonal trust fosters a mindset that encourages compassionate behaviour, resulting in optimism (Jones, 2019). Solomon and Flores (2003) define it as encompassing empathy and the understanding of others, indicating whether the holographic AI can express emotional awareness, such as care and concern, and recognize the user's needs (i.e., #7,8,9,10).

“Trust is a positive belief about the perceived reliability of, dependability of, and confidence in a person, object, or process.” (Tseng and Fogg, 1999). This belief relies on the above dimensions of trust, since the user believe the holographic AI can recognize his/her emotions, needs, or correctly answer questions. While trust is alignment of performance of the holographic AI with the user belief. The level of trust which the user places in a holographic AI is based on the extent to which they belief that the agent is capable of showing positive behaviour and intentions, and helping the user achieve goals. In short, a user is likely to trust an agent that is capable of forging a positive relationship.

Moreover, interpersonal trust is moderated by propensity to trust (Mayer et al., 1995). In essence, the user's readiness to place themselves in a vulnerable and potentially risky situation hinges on the expectation that the agent will perform as anticipated. This risk encompasses both safety and privacy concerns.

According to Sousa, Lamas and Dias (2014), trust originates from the predisposition to engage, an intention that precedes interaction. As stated in #11 of Table 5.1 (and depicted in Figure 5.4), the relationship between the user and the agent is predicated on subsequent and continuous interaction behaviours. Thus, this relationship can be viewed as an outcome. In interpersonal interactions, trust and relationship are mutually reinforcing. However, the default protocol for the holographic AI is to trust the user, leading to a unidirectional perception from the user's viewpoint. Hence, trust in holographic AI differs from interpersonal trust. Trust can be initiated by users who are open to forming connections with agents (Mayer et al., 1995), but it is presumptuous to expect a positive relationship before interacting with the holographic AI. A sustained relationship suggests a robust trust developed over time (Vanneste, Puranam and Kretschmer, 2013). It is seen as the culmination of trust built through interaction. Moreover, trust in HCI is crucial for forging a successful relationship (Salanitri et al., 2015; McKnight et al., 2011). A relationship may falter if the user distrusts the holographic AI. Thus, trust can be characterized as a belief wherein the user is confident that the holographic AI will assist them in achieving objectives with sincere intentions and actions, fostering a positive connection (i.e. an outcome).

Additionally, trust development is an ongoing, dynamic process (de Visser et al., 2016). As such, the calibre of trust can shape this relationship in future interactions, with current trust levels being influenced by past experiences (Mayer et al., 1995).

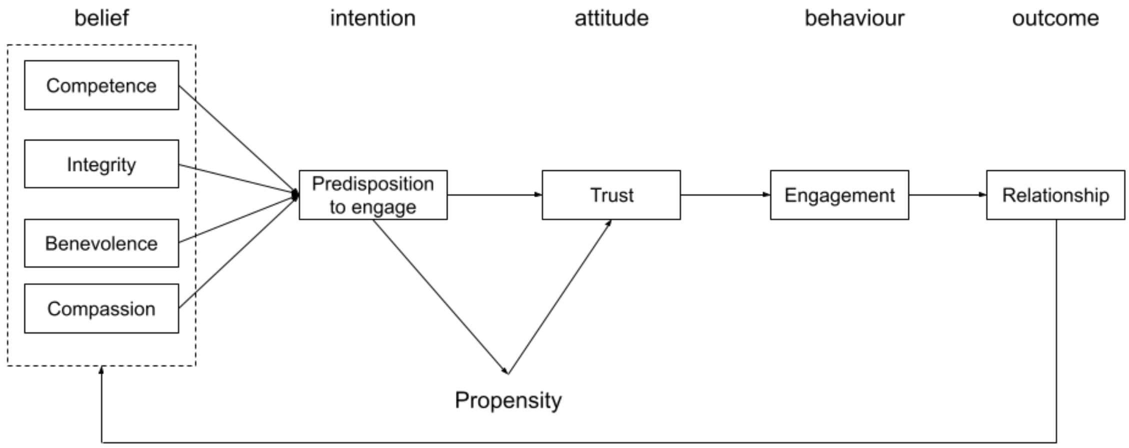


Figure 5.4. The novel model of trust

Conversely, factors such as acceptance and usability also impact engagement (Dahleez et al., 2021; Philip et al., 2020), yet these elements differ from trust. Acceptance evaluates or predicts the extent to which a user will adopt or utilize an AI system (Böhm and Stein, 2020). While trust is rooted in the belief that the holographic AI will demonstrate beneficial intentions and behaviours to achieve objectives, acceptance is a behavioural intention signifying user endorsement of the AI (Kelly et al., 2023). This indicates that adoption may not always be voluntary, as users must acknowledge the AI's limitations (ibid). Moreover, there is a scarcity of research exploring the interplay between acceptance and trust in holographic AI, as well as the sequence in which these factors emerge. Therefore, continued exploration of these dimensions is critical.

Usability measures the holographic AI's ease of use by referring to how intuitively users can interact with the AI to fulfil their goals (Issa and Isaias, 2022; Benyon et al., 2006). Usability focuses on the holographic AI's ability to complete tasks effectively, efficiently, and to users' satisfaction, distinguishing it from trust (Issa, 2007). Usability can be objectively measured, whereas trust is more complex, encompassing both objective and subjective assessments. Although trust and usability are interrelated (Salanitri et al., 2015), future research could delve into usability to further elucidate trust and the PICS model.

Figure 5.4 presents a template for a testable structural equation model, with the observable variables of competence, integrity, benevolence and compassion predicting the latent variable of trust. For some items such as compassion, it may be possible to substitute a more direct measurement tool for a Likert scale. For example, sentiment analysis of dialogue transcripts could yield more accurate data on affective trust expressed by the user, and the corresponding actions of AI. Further insights might be gained by studying facial expressions and prosody of speech for proxies and specific indications of affective trust, using laboratory tools such as EEGs and imaging technology, which in turn are capable of determining inner neurological states associated with specific answers to questions.

5.3.4 Summary

This section has proposed the construction of a new metric scale, an extended model of trust, which conceptualizes and to an extent quantifies the level of a user’s trust in holographic AIs. As detailed above, 104 statements were identified, and scrutinized by a panel of judges on the basis of duplication, relevance and polarization.

This section examines the parallels and contrasts between the newly proposed trust model and the interpersonal trust model suggested by Mayer et al. (1995). The capability of the holographic AI is linked to its task performance efficiency. Integrity relates to the user's belief that the holographic AI complies with a set of established interaction protocols. Benevolence reflects the holographic AI's good intentions towards the user. Additionally, compassion is absent from the Mayer et al. interpersonal trust model. Users might doubt whether the holographic AI can accurately perceive their emotional needs and exhibit corresponding behaviours through interactive modalities and tailored services. If users are convinced that the holographic AI can offer the necessary services along with emotional support, trust may be fostered, enhancing interaction and relationship building.

The remaining 11 statements serve as a comprehensive model for predicting trust in the HCI context. It should be mentioned that trust between a human user and holographic AI is different from human-human trust, as well as that between a human and a non-anthropomorphic technology such as websites and banking apps. In terms of enabling the user to achieve goals, the key dimensions of trust in holographic AIs are benevolence, competence, commitment, empathy and integrity, all of which are considered in the Likert scale-based questionnaire developed in this study.

5.4 Adaptation of the Scale for Children

Mellor and Moore (2013) devised an experiment involving children aged 6 -13, and measured differences between yes/no responses and 5 - point Likert format-based responses. From their observations they argued that children are capable of completing a 5-point Likert questionnaire with both physical and abstract tasks to indicate their degree of agreement and disagreement. In order to accurately measure the sense of trust, the questionnaire in this project employs a 5-point Likert format in the form of multiple-choice questions.

A pre-test was conducted with 5 students to test the validity of the questionnaire items. All questions are phrased as positive statements to ensure a uniform polarity, and are all multiple-choice questions, the answers to which range from “strongly agree”, through “agree”, “undecided”, “disagree”, to “strongly disagree”, thus providing evidence of user experience with holographic AIs.

The administrability of the questionnaire was verified with this pre-test. Following the positive pre-test, it was considered ready for investigating the trustworthiness of the holographic assistant created for the experiment, providing insights on interaction and testing the sense of trust developed.

Nr	Items
1	Sarah is clever.
2	Sarah knows what she is doing.
3	I like Sarah.
4	Sarah does not lie. (Reverse Polarity)
5	I think Sarah wants to do good.
6	Sarah is kind.
7	Sarah feels real to me.
8	Sarah looks out for me.
9	Sarah wants to help me.
10	Sarah is caring.
11	Sarah is my friend.

Table 5.2. Child-friendly version of the metric scale for measuring 'trust'

To back up the quantitative scale with more qualitative insights, four open-ended questions were added to investigate children's feedback regarding advantages and shortcomings of holographic AIs (see Table 5.3). The first, third, and fourth questions concern the direct user experience, while question 2 explores implications regarding sharing of personal information.

1. Would you like to share your story or life with Sarah? And why?
2. Do you think Sarah can keep your secrets
3. What did Sarah do well?
4. What did Sarah not do so well?

Table 5.3. Open-ended questions

In order to make sure that children can understand the level of agreement and disagreement, and the meaning of word 'undecided', the second pre-test was conducted with 7 children aged 9-13. It was found that the children could understand each degree of choice, and when they were not sure of their feelings they selected 'undecided'.

5.5 Experiment Design

As mentioned, this experiment has been designed to explore children's sense of trust towards holographic AIs. This section describes the process of designing teaching content for young children, so that a holographic AI can teach them to identify 2D and 3D shapes, such as a cube and cuboid.

5.5.1 Material

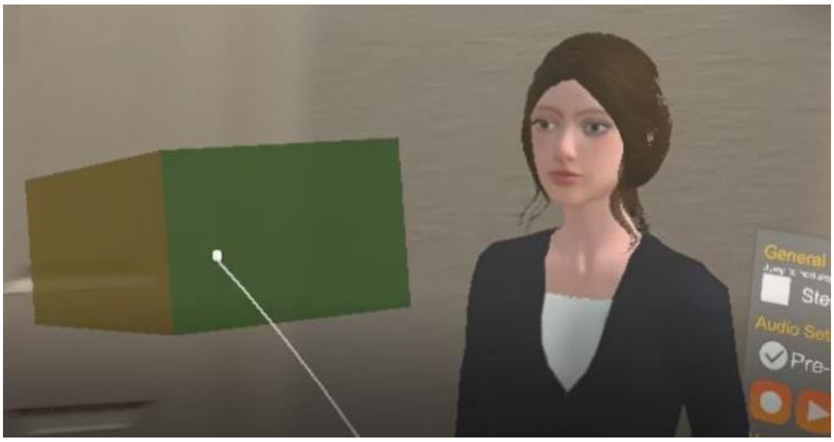


Figure 5.5. The holographic AI in the trust experiment

Traditional methods deployed in the curricula for young children sometimes struggle to effectively stimulate students' interest and curiosity. With immersive technologies, novel opportunities arise, that, if validated, can positively impact on engagement - at scale. Moreover, there are novel interaction techniques possible, that may be beneficial, particularly for knowledge that involves spatial reasoning and imagination (Baumgartner, Ferdig and Gandolfi, 2022). This experiment informs the design and implementation of a 3D character model by evaluating iteratively its ability as a holographic guide that can deliver trustworthy user interface orchestration and engage learners with the curricular learning content.

Communication is key for establishing a cooperative relationship between an AI hologram tutor and child. Such a holographic AI should be capable of responding to students, predicting their needs, teaching a selection of mathematical subjects, and recording, analysing and reacting to children's study progress, in order to help these learners solve problems after class. However, Chapter 3 focuses on intelligent tutor systems and workout teaching, and this holographic AI is used for simulation to measure user experience, thus both holographic AIs have difference aims and functions.

5.5.1.1 Pedagogical requirements

This experiment invited two professors with expertise in teaching mathematics and delivering online tuition to outline criteria for satisfying young students' requirements. They suggested that the application should not begin with calculations or formulae, but with identifying different 3D shapes and 2D shapes, and training children to classify and analyse shapes of objects in real life. They advised that children need to learn mensuration. For example, once they understood the differences between a cube and cuboid, the holographic AI should then ask them how many cubes can be fitted into a cuboid. Therefore, the holographic AI should enable children to attain a basic knowledge of each 3D shape by counting the number of faces, vertices, and edges. During that period, the children should be tasked with identifying the shapes and features of 3D shapes and corresponding 2D shapes; this way, the holographic AI will help children to understand the relationship between 2D and 3D shapes (see Figure 5.5). Steps of comparison and analysis require children to consider how a 3D shape can be established by the other 3D shapes. Finally, children need to measure the volumes of 3D shapes, a task which also does not involve learning equations. The teaching contents prescribed by the professors have four aims:

- Identification---Studying basic knowledge of 3D shapes (e.g. cube and cuboid) by counting the numbers of faces, edges, and vertices.
- Relationship ---- Identifying each face and shape of a 3D shape.
- Comparison and Analysis ---- Comparing different 3D shapes.
- Measurement--- Measuring length and volume, but without involving function.

5.5.1.2 Holographic AI and dialogue management

In this experiment, MirageXR (Wekit ECS, 2022), an AR application for a personalized training system, was used to present holographic AIs and interaction. A holographic AI with stylized appearance and humanlike behaviours was developed in MirageXR, in order to prevent the uncanny valley effect (see Figure 2.1 and Figure 5.5). The holographic AI can perform facial expressions, lip-sync, and body gestures. The size of the holographic AI in this experiment is generally of a child's height, since a child cannot observe fully an adult-sized holographic AI. The holographic AI also follows and stays with each child user via the 'Follow Player' function. Two shapes – the cube and cuboid – were created in Maya, and were imported into the application.

This experiment neither uses pre-recorded speech nor the Woz paradigm to adequately respond to children, since it cannot satisfy the ecological validity characteristic of real-life interaction (van Straten et al., 2019).

According to the above teaching outline, dialogue management was established for interaction. This dialogue management consists of 5 steps: (1) activation and preparation; (2) teaching the children the key features of the cube and cuboid; (3) identifying squares and rectangles; (4) assembling a cuboid by cubes; and (5) prediction and measurement, i.e. putting small cubes into a cuboid.

This experiment used IBM Watson to create dialogue management and speech input and output services (more details provided in Chapter 3). Firstly, a holographic AI can be activated by the utterance, "I want to learn maths/3D." The holographic AI then asks the children whether they have studied 3D shapes before. If not, the holographic AI would explain what 3D means. After the children indicate they are ready to study 3D shapes, the holographic AI commences teaching (see Figure 5.6).

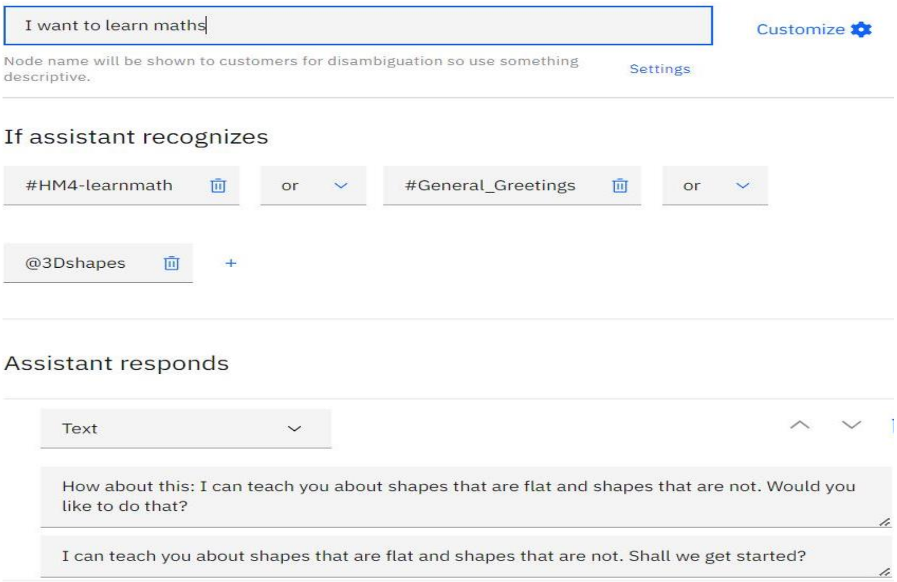


Figure 5.6. Activation and preparation

Third, given that the children are interested in learning 3D shapes, the holographic AI will provide step-by-step instructions by asking them how many faces, edges, and corners a cube has. The holographic AI will provide the right answer if the children do not. The holographic AI then instruct children on 2D shape, what a square is. It will ask the children to find features, and if children do not know then the holographic AI will encourage them and offer some tips. Upon completing the cube and square learning, the holographic AI will ask the children whether they would like to continue studying the shape of a cuboid (see Finger 5.7). All interaction steps are the same as those in the cube learning, except that the holographic AI (whose name is Sarah) will ask the children to state whether the edges and faces of the rectangle are the same as those of the square.

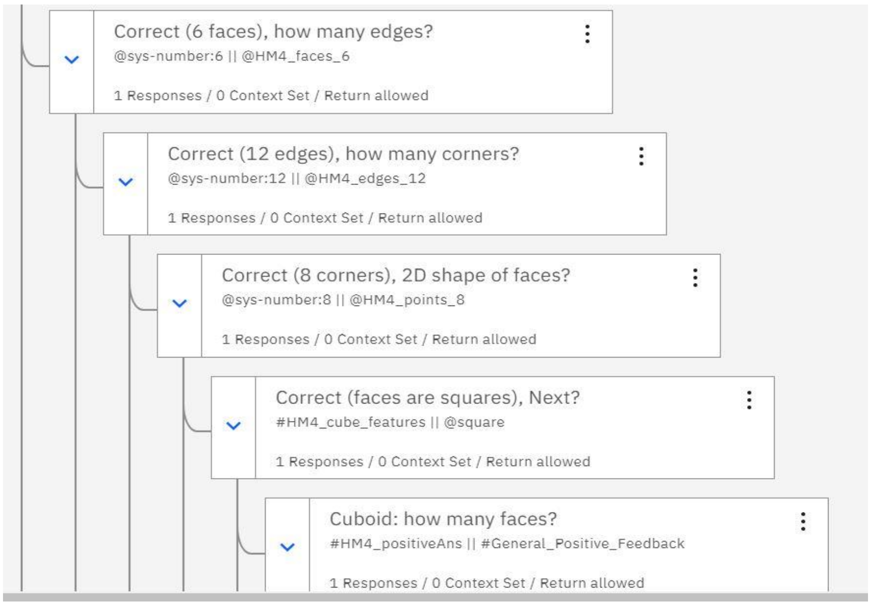


Figure 5.7. Identification

After identification learning, Sarah will ask the children to join cuboids or cubes together by controlling and moving virtual cubes and cuboids, a process which can facilitate interactivity. However, there is no script to support the holographic AI's responses as to whether the results are correct.

Finally, the children need to calculate how many cubes can be put into a cuboid. The key answer is still a number.

The holographic AI's responses are based on key word capture, whereby the 3D and 2D shapes' features are the key. For example, if a question asks the children to count the number of faces, "six" is key. It is a recognisable fact that other numbers are wrong, and the holographic AI will respond to those by asking the children to make another attempt to answer the question (see Figure 5.8). The words "equal" and "same" are key words in the identification of a particular feature of the square.

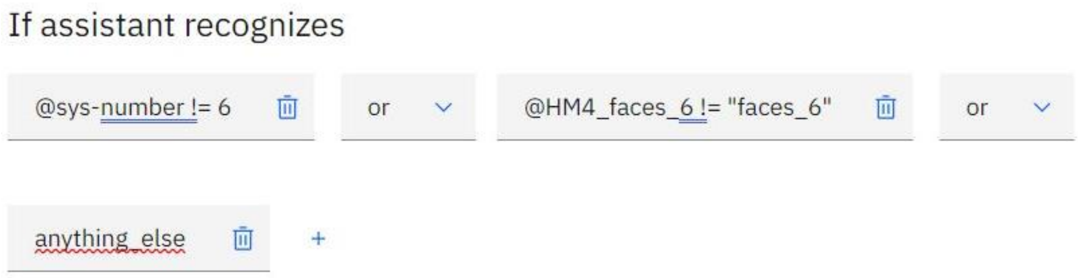


Figure 5.8 An example of key word management

There are five possible responses to each question: correct and incorrect answers, unknown (the child is unsure how to answer the question), asking questions, and withdrawing from interaction. It is important to ensure the interaction's smoothness and continuity. The correct answers can trigger the next steps of tasks, and one more chance will be provided if the answer is incorrect. The holographic AI should provide both correct answers and explanation. If the child either does not know the answer or feels unsure about his/her answer, the holographic AI provides some tips in an effort to encourage the participant. Children will also ask questions if they do not know a concept, such as right angles. The last possibility allows children to withdraw interaction at any time.

5.5.2 Environmental setup

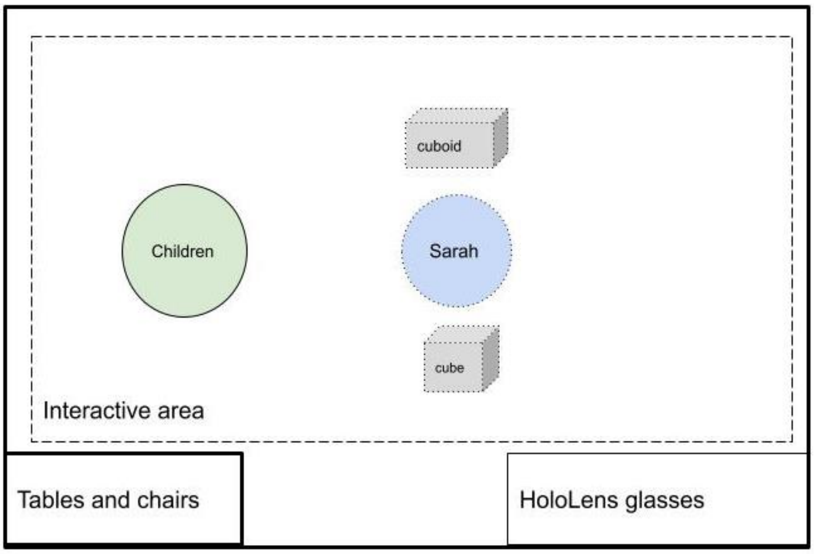


Figure 5.9. Experiment setup

The experiment was conducted in the Brookes Science Bazaar 2023 at Oxford Brookes University. The experimental space was around 10 square meters to ensure sufficient interactive area. This space was fitted to common furniture such as desks and chairs (see Figure 5.9). The following equipment was used for this experiment: chairs and tables for children and guardians to fill out the questionnaire; four HoloLens glasses; and two iPads.

It is noted that very young children are incapable of observing fully the virtual surrounding using Microsoft HoloLens as the interpupillary distance (IPD) of the HoloLens is around 63mm, whereas that of a child aged 8-11 years is 58mm-60mm. Therefore, efforts were made to adjust the IPD of the HoloLens's interpupillary calibration application, but if the children still could not see the whole virtual surrounding, the HoloLens was replaced with an iPad.

5.5.3 Participants

This experiment recruited 47 children in total from Science Bazaar 2023, Their ages range from 5 to 13 (mean=9.02, SD=2.2) (see Figure 5.10). There were 23 boys and 24 girls. In a demographic survey, participants were also asked if they had used AR and virtual agents (including voice assistants) before. There are two questions with degrees of usage (1-I don't know to 4 yes, often). Out of the 47 participants, only 15 had experience in using AR (Mean=2, SD=0.64), and 10 previous experience of play with virtual agents (Mean=2.13, SD=0.74).

Figure 5.11 shows the distribution of experience in using AR and virtual agents. Only 4 children have used AR and virtual agents before. All 47 participants were able to wear HoloLens as well as observe virtual objects, and they received panda keyrings as compensation. This experiment was approved by the Open University's ethics committee.

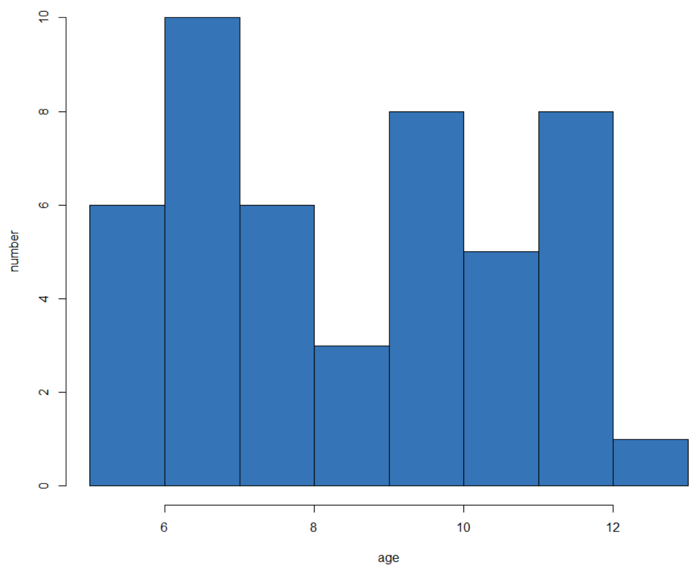


Figure 5.10. The numbers of participants by age

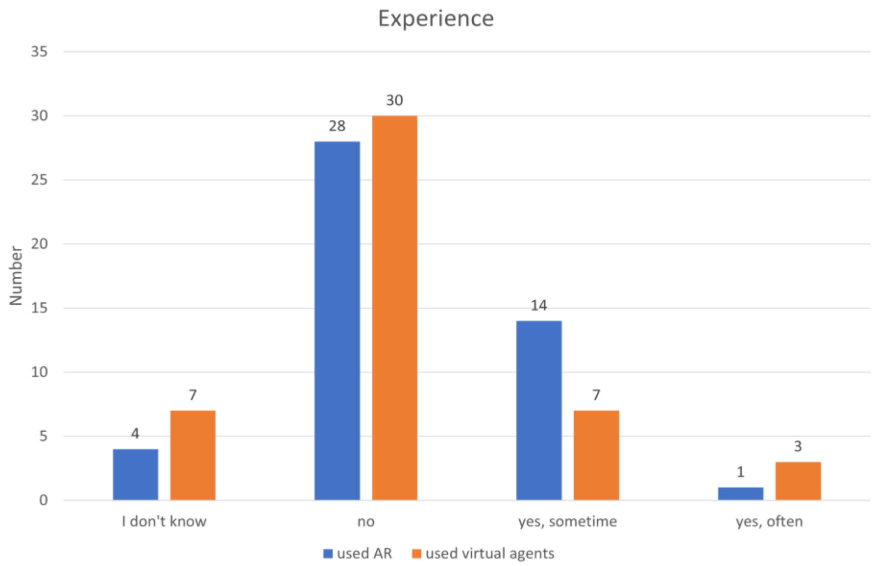


Figure 5.11. The numbers of children with different levels of experience of AR and virtual agents

5.6 Methods

This experiment has no uniform conditions: the experimenters set up and calibrated an interactive scene for each participant prior to interaction. In order to ensure that the children could quickly understand how to use AR and interact with the holographic AI, the interactive approach was mainly based on speech.

5.6.1 Interaction scenario

Following calibration, Sarah was projected into a position in the real word, and virtual cubes and cuboids were placed near the holographic AI so that the children could quickly observe them. Sarah's height was similar to those of the children (less than 1.5 metres). The virtual cubes' and cuboids' sizes were also smaller so that the children could easily control them. The first and last steps involved cubes of different colours for the purpose of distinction. In the measurement step, the holographic AI asked the children to use virtual white cube.

Figure 5.12 shows the interaction process. A child dons the HoloLens and observes the holographic AI performing greeting gestures and walking towards the participant.

A child dons a HoloLens and can see a holographic AI performing greeting gestures and walking towards the participant. The holographic AI then stood on the floor and waited to be activated.

As described before, the holographic AI then asks the child whether he/she is interested in learning about 3D shapes; if the child answers "yes", the holographic AI would continue to ask the participant to define 3D. Otherwise, the child is asked what he/she else would like to learn, and the interaction is discontinued.

During interaction, the steps of identification and relationship are intersected. For example, after the children are taught to recognise a cube, the holographic AI then instructs them on identifying 2D shapes, e.g. the square. Then the holographic AI teaches them about the cuboid and rectangle.

After that, the step of comparison and analysis is performed, whereby the children are asked to construct a cuboid or cube by dragging virtual objects. For the measurement step the children drag cube surges into a cuboid.

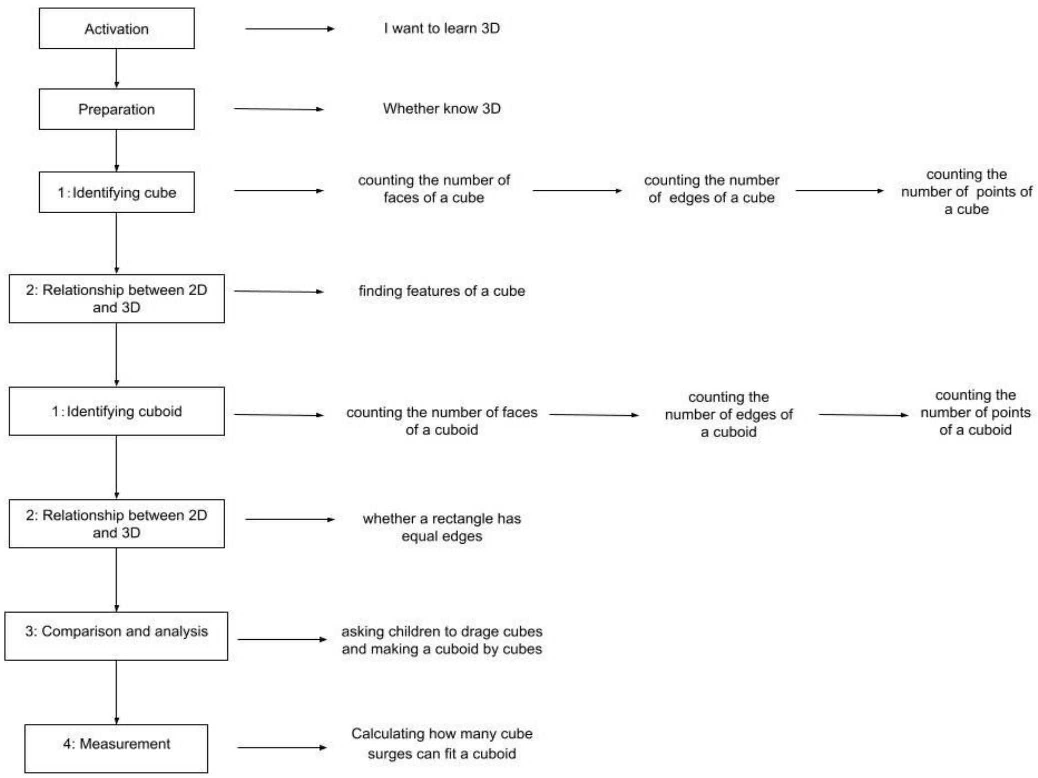


Figure 5.12. The interaction process

5.6.2 Procedure

The experimenters provided the participants with an introduction on holographic AIs, the MirageXR application (Wekit ECS, 2022), and the aim of this study. Afterwards, the experimenters asked the children whether they would like to participate in this investigation and fill out a questionnaire. Since the participants were aged under 16, they had to obtain parental consent, and the researchers need to ensure that the children and parents understood the nature and purpose of this research project. Potentially, mature participants would merely read an information sheet and sign a consent form, which would be returned to the researchers. The parent information sheet provided an ‘opt out’ section for parents who, on reflection, did not approve of their children participating in the research.

The experiment was conducted with the consent of all participants’ parents and guardians. The experimenter helped each child put on a HoloLens, and advised the child on how to use AR and interact with Sarah. During interaction, the holographic AI verbally instructed the child to independently identify 3D and 2D shapes in answer to the questions. After the experiment, each child was guided in filling out a questionnaire measuring his/her sense of trust towards Sarah. The experimenter endeavoured to help each child understand each question and, upon completion of the questionnaire, awarded him/her with a keyring thanking him/her for participating in the study.

5.6.3 Measurement

The questionnaire has been designed to measure children's trust towards holographic AIs via four dimensions, which have been presented back in Section 5.2. This questionnaire used a 5-point Likert scale (1 = "strongly disagree", to 5 = "strongly agree"), and four open-ended questions.

In this experiment, the sense of trust can rely on competence, integrity, benevolence (Mayer et al., 1995; Huang and Wild, 2021), compassion, and whether children and the holographic AI can establish a positive relationship (Huang and Wild, 2021).

Competence can reflect an objective opinion based on whether the holographic AI helps children achieve learning outcomes, and integrity reflects whether the holographic AI can provide correct teaching content and responses. Benevolence and compassion indicate an intention and motivation. The holographic AI is intended to prioritize users' needs, respond to users' emotions and provides support. Therefore, this questionnaire includes objective and subjective opinions in an effort to evaluate children's trust, and does not rely on other similar trust scales which relate to technology, automation, or virtual agents/humans.

The open-ended questions are designed to elicit detailed reasons concerning the children's perceptions. Additionally, the experiment's measurement also investigates the correlations among the children's sense of trust, and their age, gender and experience.

Afterward, 47 questionnaires are divided into two conditions for comparison between those who have not used AR/virtual agents and those who have.

5.6.4 Hypotheses

Chapter 2 has demonstrated how the intelligence of the holographic AI augments the user experience. Hence, competence could be a key factor influencing trust. Furthermore, Chapter 3 suggests that employing natural language processing could improve user interaction. Although the holographic AI does not possess intrinsic benevolence, it is programmed to demonstrate benevolent actions by meeting user requirements (Phillip et al., 2020). Past studies have indicated that benevolence from educators can help cultivate trust (Landrum, Mills and Johnston, 2013). This aspect is significant in educational contexts, suggesting that benevolence may affect trust.

Based on the aforementioned analysis, the following hypotheses are proposed:

- H1: Competence is a main influence factor in the sense of trust.
- H2: Benevolence can affect the sense of trust.

5.7 Results

This section analyses the correlations between the dependent variable (children's trust in holographic AIs), and independent variables such as their age, gender, and experience in using AR and virtual agents.

It should be noted that the experiment is based on five questionnaires, and that there were two repeated questions; therefore, the overlapping data are neglected. All selected data were analysed using the statistical software package, R studio, which omits invalid values, rather than replacing them with zeroes.

In order to measure whether children’s age is able to influence their sense of trust, the participants were divided into two broad groups: children aged 5–9 years (n=25) and those aged 10–13 (n=22). The values in the demographic questions were also based on how often participants used AR and virtual agents, from “Often” (=4) to “I don’t know” (=1).

The values of answers to questions indicate the degree of agreement, from “Strongly agree” (=5) to “Strongly disagree” (=1); “Undecided” equals 3.

In order to investigate the reliability of the questionnaire, the scores of the participants’ responses were scrutinised using the Cronbach’s alpha test, the result of which is (α = 0.78), which is greater than 0.6.

Dimension	Competence	Integrity	Benevolence	Compassion	Relationship
Scores	(#1 + #2) /2	(#3 + #4) /2	(#3 + #4) /2	(#7+#8+#9+#10) /4	#11

Table 5.4 Equations of each dimension

Each value of a dimension adds a weight based on the number of questions in each dimension for measuring the sum scores of trust and their correlation values. For example, the dimension of competence consists of 2 questions, thus the score of competence equals the values of the sum scores divided by 2. Since Compassion has 4 questions, its score equals the sum of scores of questions, divided by 4. The sum score of trust is equal to the total of all dimensions’ sum scores. Table 5.4 shows equations of each dimension.

5.7.1 Quantitative analysis

This section consists of two subsections, which focus on whether experience in used technologies can affect the sense of trust, and whether the relationships between trust and dimensions are significant. This section will investigate which of the 11 factors influences trust.

5.7.1.1 Experience in using AR and virtual agents

The participants delivered measured responses based on the 5-point Likert scale. Table 5.5 presents mean scores and standard deviations of each dimension, and whole trust (sum of results for all dimensions).

	Competence (#1and #2)	Integrity (#3 and #4)	Benevolence (#5 and #6)	Compassion (#7 to #10)	Relationship (#11)	Trust
Mean	3.76	3.62	3.74	3.43	3.26	17.46 (M=3.49)
SD	0.76	0.72	0.87	0.70	1.04	2.96

Table 5.5. Mean scores and standard deviations of each dimension and whole trust

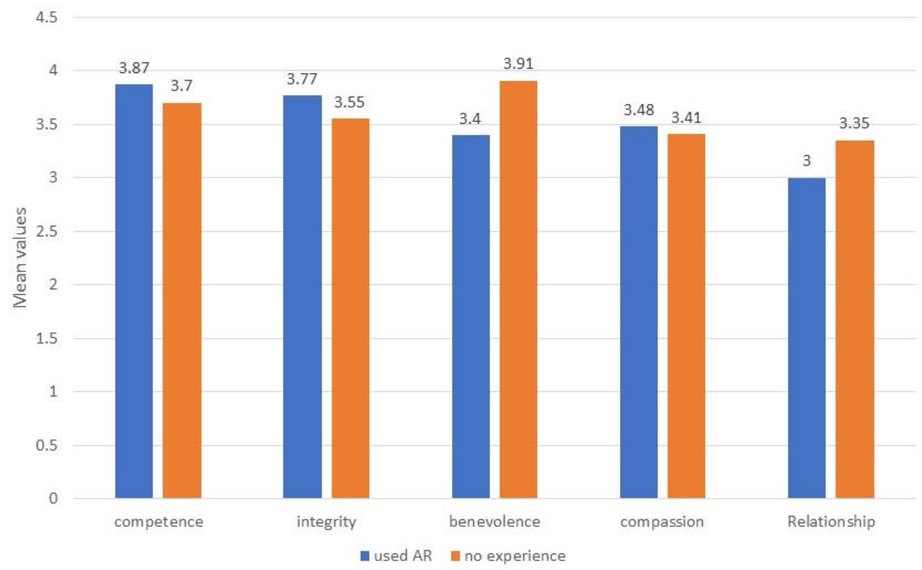


Figure 5.13. Differences in scores between participants with or without AR experience

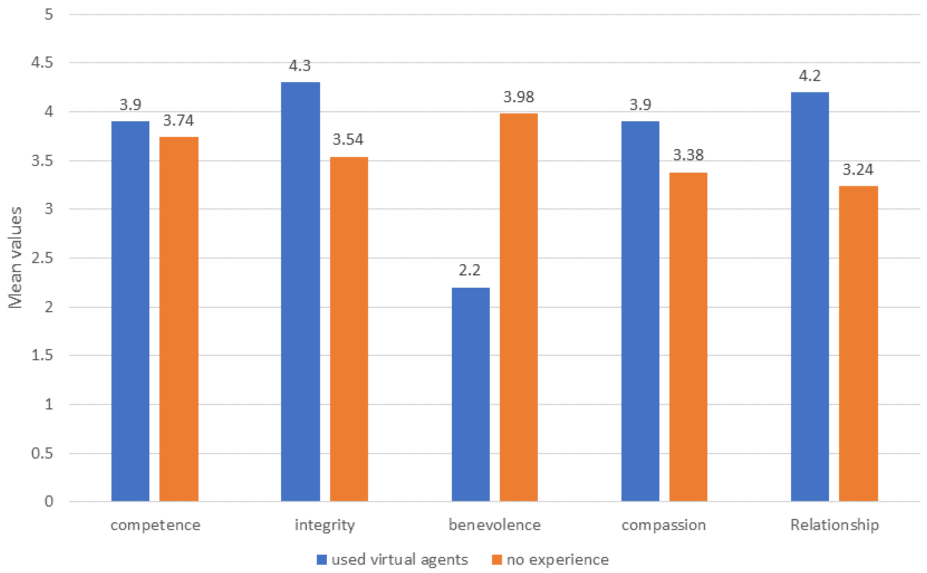


Figure 5.14. Differences in scores between participants with or without experience of using virtual agents

In order to ensure whether experience in usage of AR and virtual agents can affect the dimension of trust, the questionnaires questionnaire responses were split into (i) groups of participants having experience in using AR, and those with no experience in using AR; and (ii) groups of participants who have used virtual agents, and those with no experience of virtual agents. However, the average of variance cannot be calculated since the numbers of participants in the experience/non-experience groups (for AR and virtual agents) are not equal. Therefore, mean values were calculated in order to determine whether there are meaningful differences between these conditions.

There were 15 participants with previous experience of using AR, and 32 with no such experience, thus the average scores of each dimension were calculated by each condition. It can be seen in Figure 5.13 that the participants who had previously used AR generally graded better in terms of objective aspects of the holographic AI, such as competence and honesty. However, subjective feelings (benevolence and compassion) were lower among the children with previous experience of AR.

Out of 47 children, 10 had experience in using virtual agents, and 37 had not. The mean value of each dimension was measured (see Figure 5.14). Children who had experience in usage of virtual agents generally gave higher gradings for competence and integrity. Again, for participants with previous experience of virtual agents, the mean value of perceived benevolence was lower.

From these results it might be argued that prior experience of AR and virtual agents can affect levels of perceived dimensions. Such experience could lead to higher scores for competence and integrity, but a lower score for benevolence.

5.7.2 Correlation tests of trust, dimensions, and factors

The correlations among the dimensions, 11 factors and participants' demographic characteristics were investigated.

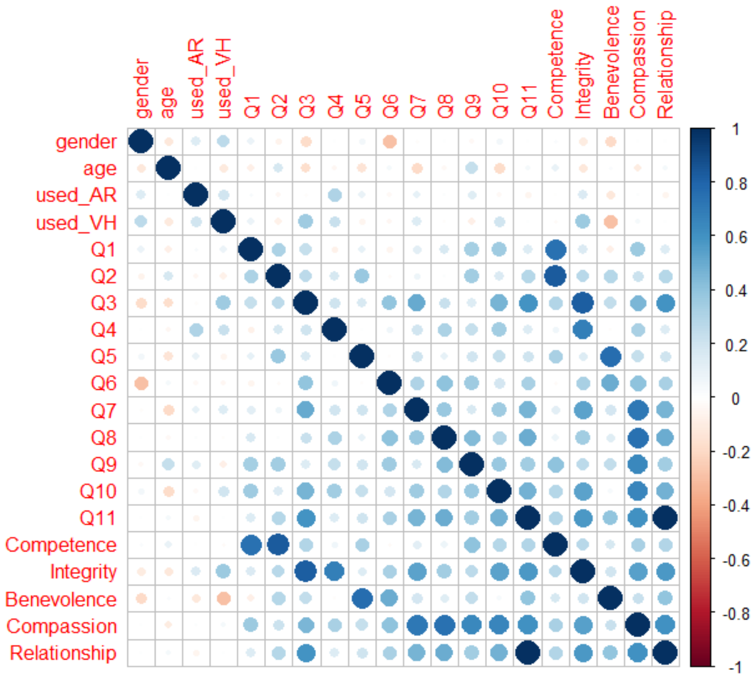


Figure 5.15. Correlations between each dimension, question, and children's demographic characteristics

The correlation between sense of trust and age was none (r=0.03), similar (though slightly negative) where participants were boys (r=-0.14). One possible reason is that the boys have more experience in using AR than the girls in the participant group.

Previous experience of virtual agents and AR also present a weak, negative correlation with level of perceived trust, especially those who had experience in using virtual agents (r=-0.26, p=0.08). It could be that young participants with interactive experience have a stronger tendency to recognise that virtual agents are not physical, and they also may compare this experience with real social communication.

Each of the five dimensions was correlated with each demographic characteristic of the children (see Figure 5.15). It is found that the dimensions of competence, integrity, compassion share negative correlations with the children's age and gender. These dimensions may also be affected by children's social experience.

Only perceived benevolence shares a positive, but rather weak correlation with age (r=0.25); the correlations between age and the other dimensions were found to be insignificant. The dimension of benevolence is negatively correlated with experience with both AR and virtual agents.

Perceived competence and experience in usage of virtual agents shares no correlation (r=-0.07).

The dimensions of compassion and integrity positively correlate weakly to experience of AR and virtual agents; this positive trend is stronger with perceived integrity (r=0.34, p =0.01). From this it may be claimed that the participants believed the holographic AI would neither harm them nor lie to them.

The following correlations between each question and the participants' demographic characteristics were obtained:

There appeared a significant weak negative correlation with age and #10 (Sarah is caring) (r=-0.3, p=0.04).

Item #7 ("Sarah feels real to me") also shares a weak negative correlation with the children's age (r=-0.21, p=0.13).

Item #6 ("Sarah is kind") negatively correlates weakly with gender (r=-0.31, p= 0.05)

In terms of experience in AR usage, "Sarah does not lie" (#4) shares a weak positive correlation (r=0.27, p=0.04).

The item "I like Sarah" (#3) also presents a weak positive correlation with experience in usage of virtual agents (r=0.33, p=0.01).

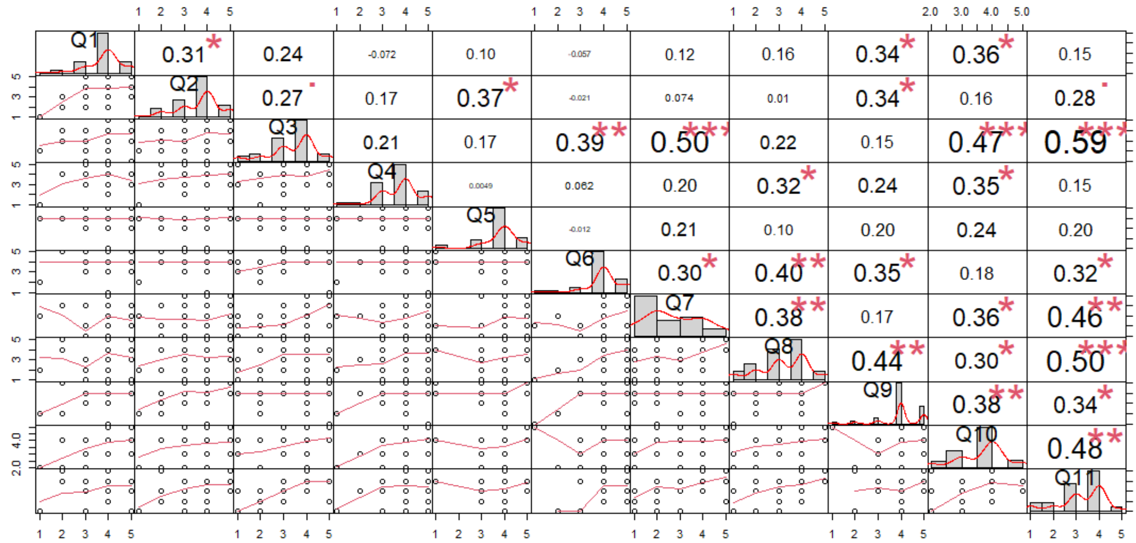


Figure 5.16. Correlations among the questions

Figure 5.16 presents a holistic correlation between each dimension, question, and children's demographic information.

It could be suggested that subjective feelings decrease with age, perhaps because older children may have more experience of as well as better cognitive ability to interact with holographic AIs. The statement “Sarah doesn’t lie” is the objective perception that Sarah provides truthful responses.

Figure 5.17 presents the correlation value and its level of significance for each pair of questions. Competence (#1, #2) is positively correlated with compassion (#9, #10). The question regarding the holographic AI’s skill (#2) is positively correlated with integrity (p=0.072), which implies that competence has a certain degree of driving force in stimulation of subjective feelings.

The dimension of integrity (#3) has a significant positive correlation with compassion (#6) (p<=0.001) as well with benevolence (#10). Item #3 evaluates a consistency of Sarah’s performance, which can be affected by levels of caring and kindness.

Perceived compassion (#7, #8, #9, #10) shares a significant positive correlation with relationship (#11) (p<=0.001), especially item #8 (“Sarah looks out for me”), an indication that if the holographic AI can take the point of view from the children’s requirements, then a positive relationship can be established.

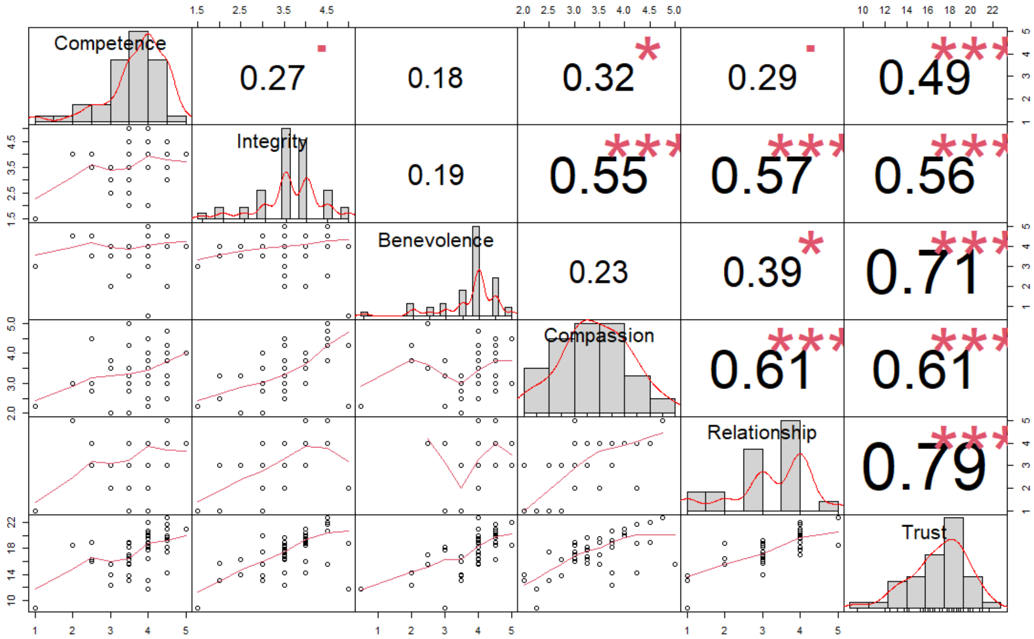


Figure 5.17. The correlations between each dimension and trust

Additionally, Figure 5.17 shows the correlations between each dimension and trust. For example, competence was found to be significantly correlated with integrity (p=0.065), compassion (p=0.028), and relationship (p=0.058). Relationship was found to correlate with the dimensions of benevolence, compassion, and integrity. Relationship and benevolence share stronger positive correlations with trust (r=0.79 and 0.71), compared to the other dimensions. Despite the highest correlation value for the relationship in Figure 5.17, benevolence and compassion are ranked as the second and third most influential factors in trust. Therefore, competence is not the sole determinant of trust. The first hypothesis remains unconfirmed, while the second hypothesis receives partial support, as children seem to prioritize compassion.

5.7.3 Qualitative analysis

Out of the 47 returned questionnaires, 5 were incomplete (participants neither provided feedback nor only answered the first question). These data were discarded.

To the first question, “Would you like to share your story or life with Sarah?”, 14 participants answered yes, in which 11 gave reasons. To the second question regarding whether Sarah can keep secrets, 21 answered “Yes” and 17 said “No”.

There were 20 participants who believed that Sarah can keep secrets, of which 9 participants had no experience in both AR and virtual agents, and 3 has previously used AR and virtual agents.

The aims of the third and fourth questions were to investigate which aspects of Sarah appear to be successful, and which should be improved. There were 30 children who found the teaching content helpful, and 4 children who found the natural language processing useful. Speaking and listening were the main issues in this experiment (n=27), since this Bazaar activity attracted numerous people to take part, the interactive space was noisy, thus the holographic AI could not accurately capture every participant’s voice.

5.8 Discussion

This section will consider how the factors mentioned above might influence the children’s sense of trust towards the holographic AI in the experiment, and the mutual effects of the dimensions. It is hoped that the findings in this chapter may inform future research pathways and development.

Overall, the mean values for competence and benevolence were higher than those of the other dimensions, while the correlation between trust and benevolence was found to be significantly positive.

5.8.1 Factors which influence trust

Competence involves ability and skill. The competence of a holographic AI can be defined as its ability to complete tasks in a specific domain (like teaching mathematics), and its skill is characterised by appropriate corresponding strategies for addressing a user’s issue, such as natural language processing and decision-making. A holographic AI’s empathy and motivation are dictated by its competence and skill. A specific AI ability can directly trigger the children’s positive attitude towards artefacts and foster within them a sense of the AI’s intention being beneficial.

Integrity consists of consistency and truthfulness. However, it was found that these two aspects did not share a statistically significant correlation. Item #3 (“I think positively about the hologram”) assesses consistency of performance in the holographic AI. If the holographic AI’s intention is not aligned with its behaviour and utterance, the resulting cognitive perception is unlikely to be positive. Therefore, it is important to explore whether the holographic AI’s behaviour can follow a real human’s interactive principles (social norms) as well. Item #4 (“The hologram answered my questions truthfully”) evaluates honesty.

Perceived integrity is positively correlated with intention as well as emotional responsiveness, especially compassion. In the experiment, it was found that the children held the view that the holographic AI’s correct answers could help them deal with mathematic problems, which perhaps explains their positive perception towards the holographic AI in terms of its capacity for empathy and caring. In addition, the holographic AI’s skill also can trigger perceived consistency of performance.

Benevolence in this context refers to whether a holographic AI’s performance is sincere and altruistic. However, Figure 5.16 represents that the results for items #5 (“I think Sarah wants to do good”) and #6 (“Sarah is kind”) share a negative, albeit insignificant, correlation. Since #5 is designed to investigate the holographic AI’s intention, #6 focuses on the result of this intention, i.e. whether the holographic AI expresses kindness.

Perceived benevolence is found to be significantly related to the dimension of compassion, and therefore yield a beneficial relationship with the children, which has a similar effect in that children trust benevolence performance in learning. (Landrum, Mills and Johnston, 2013). The scope of the statement “Sarah/the hologram feels real to me” is not limited to displaying humanlike appearance and behaviour, but also encompasses benevolent traits such as recognising children’s requirements, correcting errors, and encouraging and supporting children to study 3D shapes, in such a way as to facilitate ‘real’ feelings of perceived caring and support. Moreover, a positive disposition allows children to perceive Sarah, the holographic AI, as being motivated by altruism and kindness rather than by self-interest or hidden agendas. The holographic AI honours children's preferences and consent by inquiring about their interest in learning 3D shapes. Thus, the holographic AI consistently exhibits user-centric behaviour, which nurtures a sense of trust.

Compassion also can reflect the holographic AI’s intention and is characterised by positive emotional concern such as empathy. The difference between compassion and benevolence is that compassion involves recognizing and mitigating the user's distress or offering support by encouraging them to learn from errors instead of casting blame (Andersson et al., 2021). This may give children a sense of empathy and understanding. As mentioned before, questions in the dimension of compassion share a positive correlation with each other, and all four items also correlate to the dimension of relationship.

The relationship is a product of trust, ensuing from interactions. In this context, it can be defined by the willingness of the participants to establish and maintain this virtual social interaction. If the holographic AI can identify children’s emotions, resonate with these, and also take into account the children’s perspective in its reactions, a positive interactive relationship would precipitate. For example, item #8 (“Sarah looks out for me”) and item #10 (“Sarah is caring”) share significant positive correlations with relationship ($r=0.5$ and 0.48, $p<=0.001$). Besides, a relationship is not just a result of trust, but also a factor of influence. They both have a mutually beneficial effect. For example, if the user is impressed with the holographic AI's abilities and functions, then the user believes that the agent will be able to fulfil its promises, so that they can form a relationship at the beginning. However, if the user discovers that the product is not able to achieve goals in subsequent interactions, this relationship may negatively impact the user's sense of trust.

On the other hand, the results also include negative correlations, even though they are not statistically significant. For example, competence might not correlate with truth. There were three participants who believed that the holographic AI can lie to them. One possible explanation is that the holographic AI did not always capture each participant's voice correctly in the experiment, and may have responded immediately to voices from other children.

The young children probably utilized different ways of thinking when dealing with these questions, and they might not have distinguished the differences between the questions, which in turn are scaffolded on the dimensions. For example, it was found that the results for #5 (regarding benevolence) did not correlate with any other factors, that the correlation values were low ($r=0.1$-0.24). Besides, children tend to focus more on emotional responsiveness than on competence.

In terms of each dimension's correlations, it is found that competence shares higher and more significant correlations with integrity, compassion, and relationship, since these elements derive from AI technology. Although benevolence also shares a positive correlation with competence, its value is statistically insignificant ($p=0.231$). From the perspective of the children, perceived benevolence not only can be triggered by the holographic AI's intelligence. As mentioned before, benevolence reflects the holographic AI's intention, which is the aim of this application to teach geometry, and it is also more of an attitude (affective performance) than ability on the part of the holographic AI. The correlation value of integrity and benevolence is low and therefore insignificant ($r=0.19$).

Importantly, relationship is an outcome of trust in that trust engenders more interactive behaviours, therefore it is understandable that the correlation between the two variables is high and significant. Benevolence shares the second-highest value of correlation with trust ($r=0.71$); this reflects a similar finding by Tschannen-Moran and Hoy (2000). Since relationship is an outcome of trust, it cannot be considered an influencing factor. This pattern alone suggests that the children in the experiment would prefer to interact with a benevolent holographic AI compared to smart one, that benevolence has a critical influence on the user's sense of trust. However, according to the PICS model, this emotional expression is based on a certain degree of senses and intelligence. The value of perceived compassion shares the third-highest correlation with the sense of trust ($r=0.61$), indicating that this trait also plays a critical role in the establishment of trust in a relationship. Tschannen-Mora and Hoy (2000) proposed that a person should rely on another's good intention for the most interest, but not to the extent of exploiting the person's vulnerability to implement self-achievement. Benevolence in this context could be defined as the capacity of a holographic AI to follow interactive rules while communicating with a user. The experimental results suggest that the children subconsciously knew that the holographic AI harboured good intentions prior to the interaction with them. During the interactive period, even though participants made mistakes, the holographic AI did not criticise or blame them, but instead encouraged them by providing cues and explanation. Therefore, out of the 47 children, 42 either agreed or strongly agreed that the holographic AI is kind, and only one disagreed. Moreover, even though there were 5 missing data for #5, 35 of the children answering the question said they agreed or strongly agreed with the statement.

Figure 5.17 shows correlations between these five dimensions and trust, indicating that the strength of a specific dimension can enhance trust perception. Competence engenders trust in human-AI interactions, allowing a holographic AI to execute tasks and control its expressive responses and intentions. However, for children, subjective expression—particularly in the form of benevolence and compassion—is more critical. Thus, the data does not confirm that children prioritize competence more highly. Conversely, compassion is identified as a significant influential factor. As a result, neither hypothesis can be conclusively verified.

5.8.2 Other factors influencing trust

Referring to the demographic survey, it was found that the degree of perceived competence appears to decrease along with the children's age. As noted earlier, older children generally have greater social experience and a certain degree of knowledge and cognitive judgement, and are more capable of distinguishing between a real and virtual social interaction. This negative influence is pronounced in the responses to the question as to whether Sarah is caring ("The hologram is compassionate"). According to the results, the dimension of integrity, the holographic AI's intention, and perceived real interaction also can be affected by age.

Children's gender also appears to impact on factors of trust. For example, perceived benevolence was lower among the boys who interacted with the holographic AI. Gender could also impact on the dimensions of compassion and relationship, and perceived consistency of holographic AI performance.

Referring to experience in using AR and virtual agents, the children who have used AR before sense a stronger feeling of honesty in the holographic AI's responsiveness. It is likely that these children will be comparing current with previous experiences in their evaluation of holographic AI competence, performance consistency and compassion, since an increased trend of perceived benevolence appeared if children had no experience in using AR. Similarly, perceived compassion is generally lower among the children with previous experience in using virtual agents, albeit not to a statistically significant degree. It is also found that perceived benevolence is lower among the children with previous experience of interacting with virtual agents. However, perceived consistency of performance in the dimension of integrity appears to increase with experience in usage of virtual agents. Thus, the children with this experience appear to hold more positive views on the interaction with the holographic AI, which explains the significantly positive correlation between the dimension of integrity, and experience in using virtual agents.

Generally, children's age, gender and experience appear to have slight impacts on the dimensions of trust. The children appeared to judge the holographic AI rationally and objectively based on previous experience (e.g. perceived benevolence among those with previous experience of using virtual agents is lower). It may be that a child's sense of trust in holographic AIs might decline with increasing experience.

5.8.3 Children's suggestions

The open-ended questions presented whether the children would like to interact with Sarah in much the same way as with their friends by sharing stories and secrets in daily life. Some of the children said they might, given that they considered Sarah as reliable, kind, helpful, caring, and trustworthy. One child opined that Sarah seemed to care for her in a motherly way. Five children said they would not bother sharing stories and secrets with the holographic AI, pointing out that Sarah is not a real human. One child said he would not do so as he had never encountered Sarah before. More perceptively, some of the children opined that it would be unwise to share personal secrets with the holographic AI as an AI program can potentially be hacked, and it relies on a database. One child claimed that too many verbal utterances might undermine the ability of the holographic AI to keep secrets.

The teaching content (mathematics) facilitated the children's user experience. Overall, the children thought that Sarah provides good explanations, and is capable of recognising and correcting mistakes, as well as setting an appropriate level of difficulty in line with the children's requirements. For these reasons, one child commented that Sarah can "look out" for him. Further, the experiment also demonstrates that speaking is an important way of interaction. Some children commented that verbal interaction is one of the best facets of Sarah's performance. However, as mentioned, the interactive space was noisy, and under these conditions the holographic AI cannot clearly and accurately capture each child's voice. The children noticed that Sarah sometimes tracked other people's keywords, and most of them commented that Sarah misunderstood some utterances, or sometimes said she did not understand what a particular user had just said. Interestingly, a boy provided a suggestion that he hopes that Sarah could hold his hand and touch him during interaction.

According to the PICS model, competence can correspond to intelligence, which can influence the degree of perceived integrity, compassion and the holographic AI's intention in the aspect of benevolence. Despite this, the value of correlation between competence and trust in the experiment was below 0.5. For children, emotional support may be more important, especially for those who are 5–9 years old. Older children are more likely to recognise that a holographic AI is an artefact, that it cannot generate a real feeling between AI and user. Further, children may prefer to interact with a character with whom they are familiar, which has abundant story plots and personality to support interaction. This requirement is particularly crucial with long-term utilisation. In terms of gestures, it may be necessary to deliver emotional responsiveness consistent with the holographic AI's intentions, such as touching a child's head, high-five gestures, fist bump, and virtual hug.

5.8.4 Trust model of the holographic AI

Mayer et al. (1995) developed a model that emphasizes the importance of interpersonal trust in creating reciprocity and mutual benefit. This model recognizes the ongoing and dynamic nature of relationships between individuals. The elements of the human trust model can benefit to understand the sense of trust towards AI (Dzindolet et al., 2003; Hoffman et al., 2013). Both definitions of trust highlight the belief that user (trustor) expects the AI (executor) to meet their requirements. Therefore, competence, benevolence, and integrity have similar meanings in the context of trust towards holographic AI. However, the attributes and features of AI agents do not allow for the direct application of interpersonal trust. Trust towards holographic AI is unidirectional, with the default being to trust the users. This makes it difficult to establish a reciprocal relationship with AI systems (Ishowo-Oloko et al., 2019; Karpus et al., 2021). As a result, interpersonal trust cannot be fully employed in the AI system. For example, although the holographic AI can address specific tasks through possessing the competence that consists of integrative abilities (competence) and specific skills outlined in the AI trust model, it cannot reap rewards from users.

Furthermore, the human-centred trust model prioritizes user perception in the digital world, taking into account factors such as information transparency, ethics, cultural influences, explainability, security, and accountability (Gebru et al., 2022; Sousa et al., 2023; Scharowski et al., 2023). This model has a broader scope, making it challenging to explain trust towards specific types of AI agents. Sousa et al. (2023) argue that anthropomorphic AI can facilitate emotional trust. However, this perspective overlooks how holographic AI can potentially provide subjective support through its intelligence or competence, especially in the context of pedagogy. Moreover, previous studies and the human-centred trust model also demonstrate the provision of customized services to reflect the AI system’s ability to fulfil user expectation, they fail to elucidate the specific element or behaviour that can truly foster a meaningful connection between the holographic AI and users.

The proposed trust model towards holographic AI explores the importance of compassion, which is not present in Mayer et al. (1995)’s model, as well as benevolence. Users not only expect holographic AI to achieve goals, but also to provide warmth and understanding, even when he/she provides incorrect information. Since holographic AI lacks emotions, it is necessary to employ different dimensions of the PICS model to create an illusion of emotional expressions. Additionally, both interpersonal trust and human-centred trust consider the vulnerability and risks that users may face, which are not addressed in trust towards holographic AI. According to the definition and dimensions of trust, if holographic AI aims to establish a satisfying relationship with good intentions and motivations, it can avoid the possibility of negative situations.

5.8.5 Recommendations

Based on the children’s feedback and suggestions, six recommendations for improving children’s sense of trust towards holographic AIs are proposed:

- The holographic AI should have its own character’s background prior to interaction in order to improve intention and motivation. However, this does not mean the holographic AI should always talk; rather, the interaction should take the form of story sharing to enhance engagement.
- The interactive context should pay attention to emotional responsiveness, especially benevolence.
- A positive relationship depends to a significant extent on compassion. Therefore, in order for it to exhibit empathy and consistency of behaviour

and utterance, the holographic AI should take the perspective of children into account.

- Explanations for dealing with learning problems is important in intelligent tutor systems, therefore, the holographic AI not only should provide basic information, but also reasoning and solutions.
- Speaking and animations are the main interactive approaches. The personal service provided from the holographic AI should accurately capture a user's voice.
- Animations should be consistent with the holographic AI's intention to express emotional support.

5.9 Validity

The primary aim of the initial study is to develop a novel metric tool and subsequently evaluate the user experience. This involves utilizing the Likert scale methodology, where experts in the field of HCI assessed the relevance and distinctiveness of items related to trust. As a result, the tool has content validity. According to Yaghmaie (2003), content validity refers to the degree to which a measurement instrument accurately represents the concept it is intended to measure. Expert judgment and cluster analysis are crucial for determining the validity and comprehensiveness of statements about trust in effectively capturing the construct's scope under examination.

The second study utilized the scale to assess children's perceptions of trust towards holographic AI. In this experiment, the settings, scenarios, and variables of the holographic AI could accurately represent real-world interactions (Hartson and Pyla, 2019), thereby possessing ecological validity. Consequently, the findings from the experiment could have practical implications. Additionally, conclusion validity relates to the extent to which the conclusions drawn from an experiment are justified (García-Pérez, 2012). Therefore, this study demonstrates conclusion validity as it is based on the analysis of correlations between each dimension and trust.

5.10 Limitations

This study has limitations concerning the trust scale and experimental design. A pre-test was conducted to confirm the validity of the trust scale, and terms from 11 statements were adapted for children to ensure their comprehension. However, the study does not examine the degree to which these altered statements accurately convey the meanings of their original versions. For instance, item #3 ("I think positively about the holographic AI") indicates a positive attitude, which does not necessarily mean the user likes the agent. The statement "I like Sarah" for children might not align with the holographic AI's rules of interaction, performance, and achievement, as the dimension of relationship seems to have a minor correlation with it. Furthermore, while children who like "Sarah" may develop some form of relationship, this does not necessarily evolve into friendship. Children might struggle to comprehend the nature of their interaction with the holographic AI, as open-ended questions have shown that children perceive the AI as a mother or assistant.

Additionally, the engagement with the trust model is based on concepts from Sousa, Lamas and Dias (2014) and Mayer et al. (1995), but both studies do not explore what other factors might influence engagement. The trust model suggests that perception of trust can be affected by prior experiences and outcomes but is limited to this context.

Given that the purpose of this study is to assess the comprehensive user experience with holographic AI, rather than concentrating on capabilities, appearances, or perceptions, the categorization of holographic AI groups based on varying levels of trust was not used for comparative analysis.

The validity of the conclusion is affected by the 10 missing values. The data analysis does not classify different child groups, which makes it difficult to derive robust correlation values to support the conclusions, even though it provides correlation values between sense of trust and ages, gender, and experience using virtual agents and VR and AR. For example, older children who have acquired knowledge of geometry may exhibit reduced interest towards instructional material, Furthermore, cultural differences and educational backgrounds may also influence the sense of trust.

Therefore, to enhance the validity of future user experience research, control factors should be taken into account. Additionally, a new scale tailored for children should be developed to ensure that statements accurately reflect their intended meaning.

5.11 Summary

As presented in Chapter 2 and 3, there are few studies focusing on a holistic perception of holographic AI in conviviality dimensions such as trust (see Figure 5.19). It is essential to establish and maintain trust in interpersonal relationships as well as in HCI. Therefore, the purpose of this study is to investigate the factors that influence the perception of trust towards the holographic AI, which extends the conviviality dimension of the PICS model (Figure 5.19).

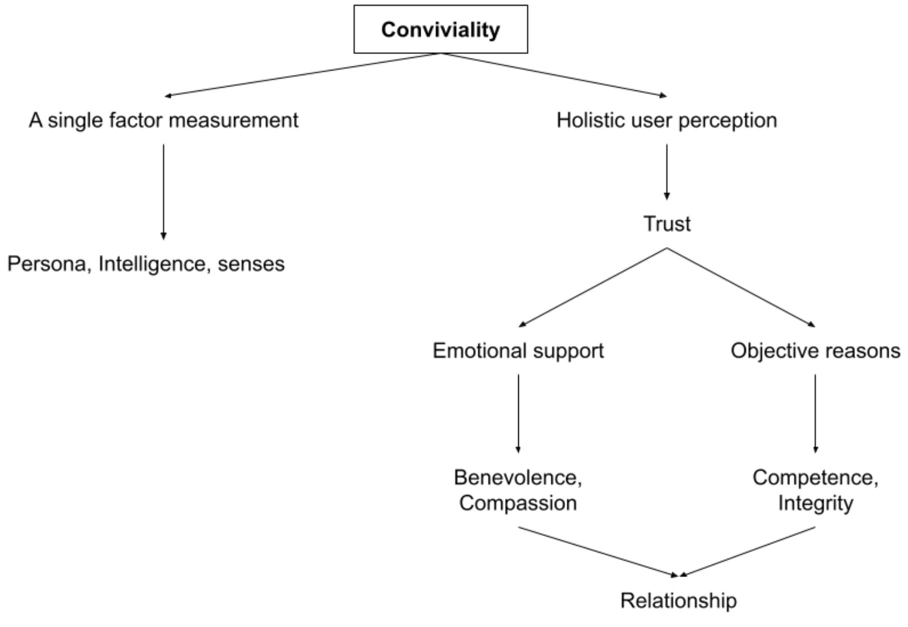


Figure 5.19. Holistic user perception in the PICS model

In order to investigate and measure the sense of trust towards holographic AIs, this chapter firstly detailed a new scale with 11 items from 104 statements that related to trust. This is based on 104 related items that related to trust, which in turn were reviewed by 11 judges for the purposes of selection, refinement, and analysis of precision. Therefore, the final selected items have distinctive features in regard to users' sense of trust, and can indicate how people regard holographic AIs in terms of competence, integrity, benevolence, and compassion, and therefore could be used to investigate whether the human user can experience interaction and enter into a trust-based relationship with a holographic AIs. Besides, the definition of trust in this context refers to the perception as to whether the holographic AI offers support to and helps the human user achieve goals via positive and altruistic intention.

Using this metric scale, a questionnaire designed to fit with children's level of understanding was produced in order to measure children's sense of trust towards holographic AIs. Although previous studies have investigated trust towards holographic AIs, they have not directly and mainly focused on perceptive trust, but instead have tended to focus on the influences of social presence and safety on trust (Kim et al., 2018). Therefore, an experiment was conducted in order to investigate what the primary factors of trust for children are in this context, and how the children's user experience and intelligent tutor system might be improved. This study then compared and analysed correlations among the dimensions of trust and the question item. It was found that the children can identify differences between real and virtual interaction. The children focused more on emotional responsiveness of holographic AI, especially benevolence and compassion. Benevolence and compassion are connected with a holographic AI's intention, and might significantly influence whether children can establish a positive relationship with holographic AIs. However, among children, past experience in using virtual agents might impact negatively their sense of trust. Based on the children's suggestions and feedback, this chapter has concluded with six recommendations for improving children's feeling of trust in holographic AIs.

Chapter 6 Conclusion and future plan

6.1 Introduction

The thesis has explored the domain of holographic AI, it has presented the entire process of developing holographic AIs, including 3D creation, body and facial animations, dialogue management, user recognition, instructional gestures, and trust measurement.

This chapter draws conclusions on the findings in this thesis in relation to holographic AI development, and considers its empirical contributions, discussing also its limitations and future research directions now possible.

6.2 Answering Research Questions

This section summaries the research findings in answer to the research questions and aim of this thesis.

Question 1 (RQ1). What elements and design dimensions constitute the holographic AI?

This study employs the PRISMA to conduct a thorough investigation into the understanding of the holographic AI and its underlying model. It collects a total of 49 studies to analyse and summarize the characteristics of the holographic AI based on their research goals and functionalities.

Out of these 49 studies, 30 employ or analyse life-sized and human-like holographic AIs; seven of these studies show how different appearances can impact user experience (refer to Table 6.1). The forementioned appearances encompass life-sized, mini-sized, humanlike, or cartoon styles. Most holographic AIs can perform animations, including emotional expressions and body movements, such as walking, jumping, or idling (see Table 6.2). Accordingly, appearance encompasses size and realism, and facial and body animations compose behaviours. Appearance and behaviours define a holographic AI's external attributes, personality, and job, i.e., its persona, explaining in Section 2.3.3.

Persona	**Studies**
Appearance	
Life-sized and human-like	Obaid et al. (2012); Campbell et al. (2014); Kim (2018a); Peters et al. (2018); Kim et al. (2018, 2016); Kim, Bruder and Welch (2017); Kim (2018b); Li et al. (2018); Lee et al. (2018); Hartholt et al. (2019); Zielke et al. (2018); Wang, Smith and Ruiz (2019); Randhavane et al. (2019); Miller et al. (2019); Kim et al. (2019); Lee et al. (2021); Schmidt, Nunez and Steinicke (2019); Schmidt, Ariza and Steinicke (2020); Reinhardt, Hillen and Wolf (2020); Kim et al. (2021b); Pimentel and Vinkers (2021); Huang, Wild and Whitelock (2021); Mostajeran et al. (2022); Norouzi et al. (2022); Mostajeran, Reisewitz and Steinicke (2022); Yoo

	and Tanaka (2022); Wolf et al. (2020); Wolf et al. (2022); Huang et al.(2022)
The effect of different appearances	Kim et al. (2018); Li et al. (2018); Wang, Smith and Ruiz (2019); K. Kim et al. (2020); Reinhardt, Hillen and Wolf (2020); Mostajeran, Reisewitz and Steinicke (2022); Norouzi et al. (2022)
Behaviour	
Emotional expressions (including speaking)	Kim (2018a); Schmidt, Nunez and Nunez (2019); Kim et al. (2019); Schmidt, Ariza and Steinicke (2020); Pimentel and Vinkers (2021); Kim et al. (2016, 2018); Kim, Bruder and Welch (2017); Kim (2018b); Zielke et al. (2018); Li et al. (2018); Ali et al. (2019); K. Kim et al. (2020); Li et al. (2021); Huang, Wild and Whitelock (2021); Huang et al. (2022)
Body animations	Obaid et al. (2012); Campbell et al. (2014); Kim et al. (2016); Piumsomboon et al. (2018); Kim, Bruder and Welch (2017); Kim et al. (2018, 2019); Wang, Smith and Ruiz (2019); Kim (2018a); Kim (2018b); Li et al. (2018); Lee et al. (2021); Lang, Liang and Yu (2019); Kim et al. (2021b); Schmidt, Nunez and Nunez (2019); Ali et al. (2019); Reinhardt, Hillen and Wolf (2020); Schmidt, Ariza and Steinicke (2020); Kim et al. (2020a, 2021a); Oh and Byun (2012); Li et al. (2021); Huang, Wild and Whitelock (2021); Mostajeran, Reisewitz and Steinicke (2022); Zhou et al. (2009); Miller et al. (2019); Wolf et al. (2020, 2022); Norouzi et al. (2019); Chahyana and Yesmaya (2020); Huang, Wild and Whitelock (2021); Norouzi et al. (2022); Peters et al. (2018); Reinhardt, Hillen, and Wolf (2020); Pimentel and Vinkers (2021);

Table 6.1 Persona

A total of 16 studies are dedicated to enhancing intelligence, which refers to the functions of the holographic AI. The objective of these studies is to develop computer vision, natural language processing, spatial understanding, learning systems, AR plugin, and synthetization with the user’s movements (see Table 6.2).

Intelligence	**Studies**
Spatial understanding	Lang, Liang and Yu (2019)
Physical-object recognition/interaction	Kim et al. (2021a); Zhou et al. (2009);
Natural language processing	Miyake and Ito (2012); Park and Jeong (2019); Nasution et al. (2020)
Learning systems	Oh and Byun (2012); Zielke et al. (2018); Hartholt et al. (2019); Li et al. (2021); Huang, Wild and Whitelock (2021)

Computer vision	Verma et al. (2021)
AR plugin development	Campbell et al. (2014)
Synchronization with the user's behaviours	Piumsomboon et al. (2018); Wolf et al. (2022, 2020); Yoo and Tanaka (2022)

Table 6.2. Intelligence

In order to measure quality of interaction, 20 studies examine co-presence, social distance, and social facilitation and inhibition. Therefore, conviviality is used to measure user perception of holographic AIs, such as their ability to establish a sense of co-presence or the ability to adhere to social norms and social distance, investigating the relationship between holographic AIs and users.

Conviviality	**Studies**
	Kim et al. (2021b); Kim (2018a); Kim, Bruder and Welch (2017); Pimentel and Vinkers (2021); Kim et al. (2019); Lee et al. (2021); Schmidt, Ariza, and Steinicke (2020); Reinhardt, Hillen, and Wolf (2020); Kim et al. (2016); Kim (2018b); Miller et al. (2019); Schmidt, Nunez and Steinicke (2019); Norouzi et al. (2019); Mostajeran, Reisewitz and Steinicke (2022); Aramaki and Murakami (2013); Li et al. (2018); Lee et al. (2018); Peters et al. (2018); Huang et al. (2022);

Table 6.3. Conviviality

Moreover, the holographic AI's intelligence is exploited to execute tasks or resolve problems. Section 2.3.6 defines senses that are about interaction modalities, how a holographic AI utilizes multiple interaction approaches, and how react to users and which ways of receiving and reacting to contextual information from the mixed surrounding they provide, such as natural language processing, physical-object awareness, eye gaze tracking, position detection, or posture interaction (see Table 6.4).

Senses	**Studies**
Non-verbal communication interaction	Zhou et al. (2009); Holz et al. (2011); Campbell et al. (2014); Piumsomboon et al. (2018); Li et al. (2018); Miller et al. (2019); Pimentel and Vinkers (2021)
Verbal interaction	Miyake and Ito (2012); Oh and Byun (2012); Zielke et al. (2018); Hartholt et al., (2019); Wang, Smith and Ruiz (2019); Lang, Liang and Yu (2019); Schmidt, Nunez and Steinicke (2019); Ali et al. (2019); Reinhardt, Hillen, and Wolf (2020); Schmidt, Ariza

	and Steinicke (2020); Huang, Wild and Whitelock (2021); Kim et al., (2021a)
Physical-object awareness	Holz et al. (2011); Lang, Liang and Yu (2019); Schmidt, Nunez and Nunez (2019); Schmidt, Ariza and Steinicke (2020); Huang, Wild and Whitelock (2021); Kim et al. (2021b,a)
Eye gaze tracking	Ali et al. (2019); Hartholt et al. (2019)
Position detection	Park and Jeong (2019)
Posture interaction	Li et al. (2018)

Table 6.4. Senses

Compared to traditional VR and screen-displayed agents, holographic AI augments perception of and interaction with the real and virtual worlds, and enhances multimodal adaptivity to process mixed information from users and context. For example, a holographic AI is able to combine information concerning the user's reaction and interactive spaces to make a decision and generate corresponding performance.

In general, on the basis of the findings of the systematic literature review, the initial model gathers and re-organizes features based on aims of literature and functions of the holographic AI. Therefore, this thesis has proposed a novel model titled PICS (see Figure 6.1), as it includes persona (P), intelligence (I), conviviality (C), and sense (S).

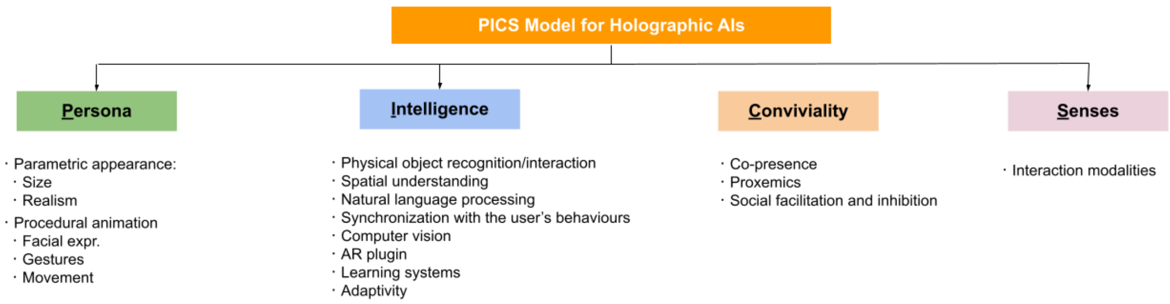


Figure 6.1. The initial PICS model

However, this study does not explore the identities of holographic AIs, including gender, occupations, hairstyles, and clothing. It also does not analyse learning materials, storylines, and interaction content and equipment. The initial model incorporates characteristics from previous studies on the holographic AI. However, components are unclear or repeated. Thus, in the next chapter, the model will be validated and refined through the creation of holographic AIs.

Question 1-1. Taxonomy of holographic AIs

The proposed taxonomy in Section 2.5.2 categorizes holographic AIs as user avatars, games characters, simulation agents, chatbots, and intelligent tutors in accordance with the functions of the holographic AI and the model.

The difference between these five types of holographic AIs has been identified (see Table 6.5). For example, a holographic AI simulation agent can mimic different roles setting it apart from the configuration of PICS features needed for a holographic AI in the intelligent tutor system, such as being a patient or an interviewee. A holographic AI in a simulation serves a navigation function by instructing learners to achieve learning through practical experience, such as decision-making, in a specific scenario, rather than directly providing teaching content. Whereas a holographic AI as an intelligent tutor is a virtual teacher or coach that merely dispenses knowledge, without storylines or narratives. However, the design of intelligent tutors focuses on whether learners can grasp knowledge (Zawacki-Richter et al., 2019; Churi et al., 2022). Game characters focus on vision stimuli and event triggers that depend on players' reactions, such as exaggerated animations. Player's decisions activate a holographic AI's animations or a storyline accordingly. In serious games, the holographic AI is employed for educational purposes (Ahmed and Sutton, 2017), but the game character is used for entertainment. The distinction between the serious game and simulation system lies in the fact that in educational games, the holographic AI adheres to independent educational rules (Whittaker et al., 2021; Laamarti et al., 2014), as opposed to imitating real-world social principles. Chatbot agents focus on natural language processing and translation, while a user avatar can synchronise the user's behaviour via position tracking.

Types of holographic AIs	**Studies**
User avatar	Piumsomboon et al. (2018); Yoo and Tanaka (2022); Wolf et al. (2020, 2022)
Simulation agents	Kim et al. (2021b); Reinhardt, Hillen and Wolf (2020); Pimentel and Vinkers (2021); Miller et al. (2019); Kim et al. (2018a); Kim et al. (2019); Kim et al. (2018); Kim, Bruder and Welch (2017); Kim et al. (2016); Kim (2018b); Li et al. (2018); Li et al. (2018); K. Kim et al. (2020); Norouzi et al. (2019); Lee et al. (2018); Kim et al. (2021a); Lang, Liang and Yu (2019); Zhou et al. (2009); Chetty and White (2019); Norouzi et al. (2022); Mostajeran, Reisewitz and Steinicke (2022); Huang et al. (2022); Aramaki and Murakami (2013); Randhavane et al. (2019); Peters et al. (2018); Obaid et al. (2012); Wang, Smith and Ruiz (2019); Lee et al. (2021); Schmidt, Ariza and Steinicke (2020); Schmidt, Nunez and Steinicke (2019); Hartholt er al. (2019); Huang, Wild and Whitelock (2021); Lee et al. (2019); Schmidt, Ariza and Steinicke. (2020); Schmidt, Nunez and Steinicke (2019)
Intelligent tutor systems	Huang, Wild and Whitelock (2021)
Game characters (including serious games)	Huang, Wild and Whitelock (2021); Li et al. (2021); Oh and Byun (2012)
Chatbot agents	Nasution et al. (2020); Park and Jeong (2019); Miyake and Ito (2012)

Table 6.5. Taxonomy of the holographic AI

However, the holographic AI taxonomy is not universal since it only examines a small number of holographic AIs. In the future, this agent may be capable of performing multiple functions.

Question 2 (RQ2). How to create an anthropomorphic holographic AI in practice, following this model?

The PICS model and taxonomy of the holographic AI provides a guideline for creating a holographic AI, so that the second study created an intelligent tutor system as an example. However, traditional 3D modelling takes time in prototypes design and 3D creation, thus it uses 3D scanning technology that can implement a semi-automatic method, especially it can directly generate a high-polygon model and texture. This study also proposes processes for dealing with disordered meshes of 3D scanned models. Then the thesis compared semi-automatic methods with the traditional approach for creating 3D characters: it has found although 3D scanning can generate 3D avatars quickly, correcting failed meshes is time-consuming, and 3D scanning could result in the appearance of the avatar almost resembling an actor, but not fully, which could lead to the Uncanny Valley effect. The traditional method is more flexible when designing different styles of appearances as it provides ways of creating facial animations and body animations using blend shapes and motion capture technology.

Implementation-wise, this thesis developed ways to for the holographic AI to verbally interact with users using IBM Watson services, namely, dialogue management, speech-to-text, text-to-speech, and translation, so that the resulting holographic AI is able to parse and understand users' utterances, and produce corresponding responses. in addition, specific speech content on the part of the user can trigger an animation or performance on the part of the avatar.

Regarding animation, the thesis explored ways of using motion capture technologies ranging from working with the Vicon system using optical camera tracking to a Rokoko motion capture suit with body-worn sensors. While the camera-based system was found to provide stable movement, it cannot record covered markers. The motion capture suit resolves this, but is sensitive to magnetic objects, which can lead to noise, data loss, and disconnection.

The sensory awareness of the holographic AI employed spatial understanding. The holographic AI can follow the user's position. However, it should be noted that this particular 'follow the user' script was created by the MirageXR team, following discussions and informed by the research conducted for this thesis (Wekit ECS, 2022). Besides, the created holographic AI cannot manipulate real objects or influence states of the MR environment, but it does not limit to places.

User model and adaptivity are necessary for recording user preferences in order to automatically update interactive information and generate corresponding context in real-time.

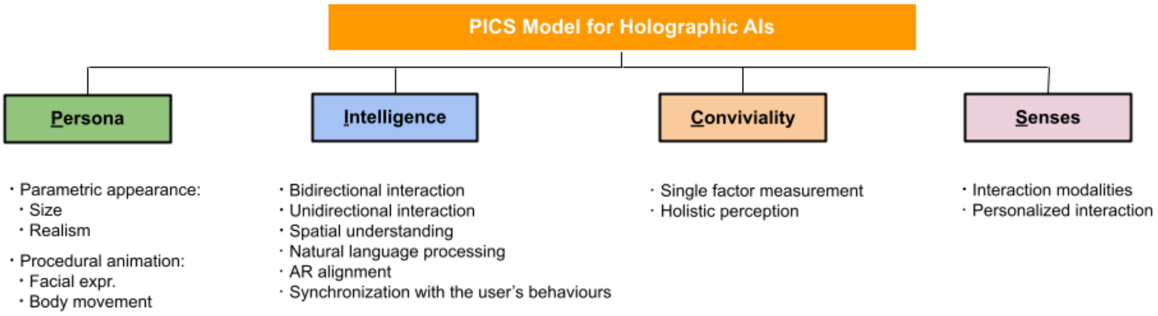


Figure 6.2. The refined model

Furthermore, some elements of the initial model are replicated as they integrate characteristics of the holographic AI. Figure 6.2 presents a refined PICS model, which omits redundant and ambiguous elements, including computer vision, physical-object understanding/interaction, learning systems, and user motion tracking, in accordance with the case study of its creation. The advancement of computer vision is crucial for the recognition/interaction with physical objects or for spatial mapping. In Chapter 2, computer vision is explored for projecting the holographic AI into the real environment. However, this concept is later supplanted by AR alignment, which refers to the processes involved in aligning digital and real-world environments. The distinction between bidirectional and unidirectional interaction determines whether the holographic AI is capable of recognizing or manipulating real-world objects. In addition, learning systems emphasize the development of educational applications, rather than a specific ability that can be generalized to other holographic AIs. Therefore, this aspect is excluded but examined in the taxonomy of holographic AIs. Moreover, previous studies have primarily examined conviviality in relation to a specific feature of the holographic AI, particularly how the level of physical-object recognition/interaction might enhance the user's sense of presence. The elements of conviviality should be replaced by single-factor measurements and a holistic perception of the user. Personalized interaction has been incorporated into the sense dimension since interaction modalities serve not only to respond to the user and the interactive context but also to offer customized services based on a user model and adaptability.

Question 3 (RQ3). What key instructional gestures should be used by an educational holographic AI?

An experiment was devised in order to collect more gestures and observe the different types of gestures used by the trainers to direct the trainees in completing the three tasks (navigation, assembly, and precision).

The objective of the navigation task is to simulate body positioning, while participants in the assembly task construct cardboard structures and arrange various shapes of panels in the precision task. These three tasks can reveal the instructor's organizational, collaborative, cognitive, and mimicry skills. Furthermore, the instructor is prohibited from providing verbal instructions to the trainee during each speechless trial, allowing for an analysis of the instructor's gesture production. The employment of pointing gestures can effectively communicate information about a specific object, location, or direction, which, in turn, can improve learners' performance (Atit, Gagnier and Shipley, 2015;

Matsumoto and Dobs, 2017). Therefore, this experiment presents three hypotheses.

- H1: Participants generate more gestures during the speechless segments of the three tasks.
- H2: For the participants, deictic gestures constitute the key functional approach.
- H3: The three tasks differently affect participants' way of thinking.

The experiment utilized a Rokoko motion capture suit to accurately record gesture and finger movements. Drawing upon studies of representational gestures (Bernard, Millman and Mittal, 2015; Abner, Cooperrider and Goldin-Meadow, 2015) and cognitive processes that influence gesture production (Abner, Cooperrider and Goldin-Meadow, 2015), the gathered data can be selected and annotated. These tasks aid in understanding the motivations behind gesture production and the various methods of referencing an object. The type of representational gestures mirrors the meaning behind the gestures.

A comprehensive array of representational gestures was identified and annotated in this experiment. These gestures encompass deictic, iconic, metaphorical, and beat gestures (Bernard, Millman and Mittal, 2015; Abner, Cooperrider and Goldin-Meadow, 2015). Transformational and mimicking gestures are novel types of representational gestures since both types of gestures can convey information. Transformational gestures are used when instructing a trainee to flip and rotate an object, and concern the manipulation of spatial information. The jumping index fingers or hand gesture is used when referring to the other side of the object. Transformational gestures can appear along with iconic gestures. For example, a trainer will firstly generate iconic gestures to imply a referent's size, and then rotate arms to represent the shape needs to be rotated. Besides, mimicking gestures tend to schematise a person's posture.

The most frequently occurring type of gesture in the experiment was the deictic gesture (existing 1290), whereby a trainer's palm or index fingers can express a direction, position, or object. This type of gesture was able to highlight a key point to attract the viewer's attention. If a trainer points in an unclear direction, the index finger is loose, or he/she directly uses a hand palm. Therefore, the second hypothesis can be supported.

Other types of representational gestures also appear in this experiment. Iconic gestures describe an object's features, such as size, shape, length, and spatial relationship. For example, a trainer may use an index finger to draw a long line to represent a referent's longest side. Or the trainer may stretch out his/her hands and position the palms of the hands in such a way that they face each other, in an attempt to indicate the subject's shape and size. Or the trainer may use gestures indicating spatial direction and position when instructing the trainee to organise and assemble blocks in different positions.

The emblematic gesture refers to a more specific but widely known category of gestures. For example, thumb ups implies that the trainee's actions are correct, whereas the waving index finger indicates they are wrong.

Metaphorical gestures can describe metaphorical concepts. For example, a left hand moved left refers to past behaviour, while a right hand moved right can express current or future behavioural plans.

Beat gestures can help trainers organise their thoughts and utterance.

Moreover, a single gesture can also express multiple and mixed meaning. A trainer’s hands can simultaneously represent an object’s size, and he/she can then his/her wrist when guiding a listener to select a right shape and rotate it to a certain degree. In terms of co-speech gestures, the most frequently used word in the experiment was ‘side’, it appeared along with deictic gestures to highlight an object’s orientation and position.

Gesture production can also reflect the instructor's cognitive processes, substantiating the third hypothesis. When the instructor discusses a complex array of objects or events, they segment information into discrete segments using gestures (Abner, Cooperrider and Goldin-Meadow, 2015). According to the analysis of gestures, the trainer who described an object's shape prior to offering guidance used fewer gestures compared to those who did not provide such preliminary instructions. It was also observed that if the instructor uses an abstract term to describe a concrete shape, this can lead to confusion or misunderstanding. If the trainee fails to comprehend an instruction correctly, the instructor may produce gestures repeatedly.

Regarding the first hypothesis, the warm-up trials tend to elicit more gestures since these trials are conducted for the first time, and there may not be an implicit understanding between the instructor and the trainee. Moreover, although the warm-up and easy trials involve similar cardboard structures, the instructor took longer at the straightforward level. This could suggest that the instructor produces fewer gestures but takes more time to structure their language.

Using the HoloLens and the Rokoko Motion Capture Suit, corresponding gestural animations were selected, and their animation files generated. These animations can be divided into gestures pointing at a specific object (usually index finger), and those pointing at an orientation (using palm and loose hands). Similarly, it was also found that when a described the size of the panel, he/she employed iconic gestures using two hands, and emphasised a small size with only a hand palm. Therefore, this feature should also be highlighted in holographic AI animations.

Additionally, when a holographic AI needs to explain a complicated arrangement, it should indicate how spatial information can be packaged into different units by way of iconic gestures. The holographic AI should perform diverse gestural forms when describing a similar or identical referent. In the experiment, although the trainers extracted previous spatial information and generate similar gestures, they generally simplified or reduced the gestural generation if the information appeared again, by using deictic gestures. On a final note, although the thumb up gesture is not instructional, it is a mean of encouragement. Likewise, waving an index finger or hand is not instructional, although these gestures signal that the viewer/trainee needs to correct a mistake.

Figure 6.3 presents that this study fills the gap of instructional gestures and extends the dimension of persona in the PICS model.

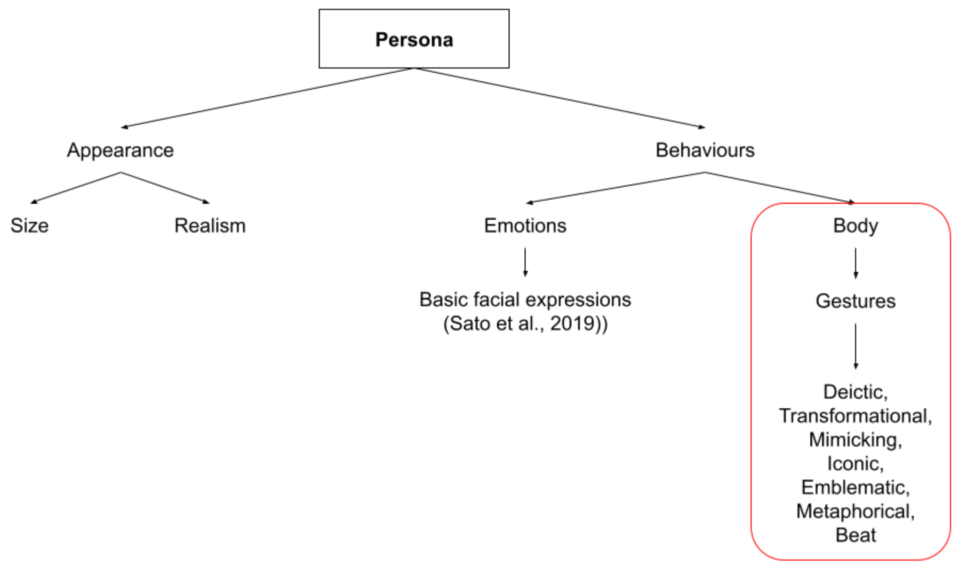


Figure 6.3. Instructional gestures extend persona in PICS model

On the other hand, the types of instructional gestures are informed by representational gestures. It was found that methods of instruction can influence gesture production and the learner's performance. However, due to the unique design of the experiment, limited data are available to draw conclusions about this finding. Additionally, the types of gestures observed in this experiment are based on three tasks, so other gestures might be observed if the tasks varied. Moreover, the study does not investigate how educational backgrounds and cultural influences impact gesture production, and the participants are from the general population rather than a specific profession.

Question 4 (RQ4). What factors affect the user's sense of trust towards an educational holographic AI?

Trust is a fundamental aspect of both interpersonal relationships and HCI. Prior research has applied the constructs of interpersonal or automation trust to characterize the trust between users and holographic AIs (Kulms and Kopp, 2018; Phillip et al., 2020). Yet, it is critical to acknowledge that user interactions with holographic agents may deviate from those with human counterparts. Consequently, this study is dedicated to elucidating a specific definition of trust in the context of holographic AI.

- Definition - What is trust towards the holographic AIs?

Kim et al. (2018) investigate whether users trust holographic AI in different situations of physical-object interaction, but the study focuses on its security. In order to measure a user's sense of trust towards holographic AI, this study develops a new scale. Therefore, a subordinate RQ was proposed:

- Scale - How to develop a novel scale for measuring the sense of trust towards the holographic AI?

Utilizing the Likert scaling technique, this research gathered and selected statements related to trust in virtual agents (Trochim, 2021). The study undertakes the compilation of hundreds of statements to investigate the notion of trust and to formulate a new scale tailored for holographic AI, drawing on the trust model proposed by Mayer et al. (1995) and the implications of the trust statements. A novel trust model has been developed to accommodate the unique dynamics of holographic AI.

While Mayer et al. (1995)'s model is traditionally employed to assess interpersonal trust, it is also applicable in the context where holographic AI emulates human interaction within a MR setting, thereby assisting in defining the meanings of each trust dimension.

The dimension of competence indicates the holographic AI’s abilities and specific skills, how it can use these abilities to help achieve a goal.

Integrity refers to truthfulness, i.e. whether the holographic AI can tell the truth. Besides, according to the interpersonal trust model. Besides, the concept of integrity is pivotal in evaluating whether the holographic AI upholds consistent interaction principles that users find credible (Mayer et al., 1995). It means the holographic AI’s performance should match its intention to achieve goal, in order to generate a positive attitude towards such agent.

Benevolence reflects a user’s affect aspects of trust. This dimension can be used to assess whether a holographic AI intention is good towards a user, e.g. shows kindness (Sousa, Lamas and Dias, 2014). For example, the holographic AI wants to do good by respecting children’s willingness.

Compassion also is affective-based trust. This process entails the holographic AI expressing concern, empathy, caring, and promising. Although this dimension does not exist in the interpersonal trust model (Mayer et al., 1995), it can yield optimism (Jones, 2019). For instance, the holographic AI encourages children to correct mistakes instead of blaming them, so that they are able to feel a sense of understanding. On the other word, the holographic AI's compassion is based on its ability to recognize the user's needs and provide them with personalized services.

The dimension of relationship is an outcome of trust, i.e. whether the user and holographic AI can establish a positive relationship. Moreover, the holographic AI inherently operates under a default assumption of trust towards the user, creating a unidirectional perception of trust, which is distinct from the reciprocal nature of interpersonal trust. Additionally, a robust trust dynamic in human-holographic AI interactions can foster a positive relationship, as evidenced by findings from Salanitri et al. (2015), McKnight et al. (2011), and Vanneste, Puranam and Kretschmer (2013). Therefore, this relationship is an outcome.

Generally, trust is an attitude that the user believes the holographic AI can provide services with positive intentions and behaviours. Therefore, trust between holographic AI interaction and interpersonal trust is different. The user hopes that this agent can provide both objective and emotional support. As depicted in Figure 6.4, the basis of trust is the predisposition to engage, which is established before the engagement itself occurs (Lamas and Dias, 2014). Based on holographic AI interaction and user engagement, a relationship is generated. Further, these five dimensions with 11 statements can be adopted to assess the factors that influence the sense of trust using a Likert scale (Hanna and Richards, 2019).

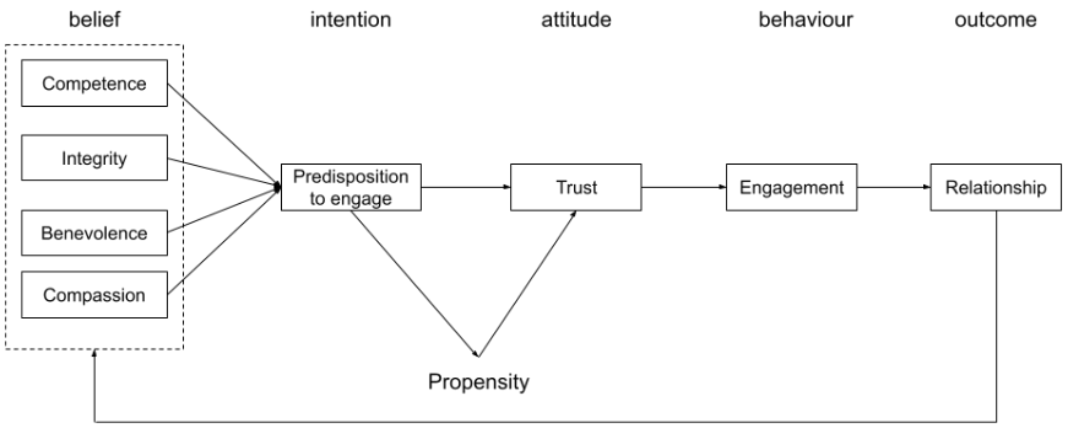


Figure 6.4. The refined trust model

Then a holographic AI is used to help children to identify 3D shapes in order to collect and analyse user experience. Chapter 2 has demonstrated that the dimension of intelligence can influence user experience, thus the first hypothesis is proposed as follow:

- H1: Competence is a main influence factor in the sense of trust.

Besides, emotional support, such as benevolence, can positively influence children's sense of trust in education (Landrum, Mills and Johnston, 2013), thus the second hypothesis is proposed below:

- H2: Benevolence can affect the sense of trust.

The correlation between competence and trust is 0.49, indicating that competence exerts a certain impact on trust ($p<=0.001$), while the correlations between trust and compassion is 0.61 ($p<=0.001$) or benevolence is 0.71 ($p<=0.001$). The perception of benevolence indicates whether the holographic AI can understand children's needs, correct their mistakes, and provide encouragement in a way that stimulates a sense of 'realism'. As a result, competence is not the main factor, benevolence and compassion are.

However, the holographic AI's competence (i.e. intelligence and senses in the PICS model) can manage its subjective expressions, thus this set up on the basis of competence by natural language processing or senses. Therefore, both compassion and benevolence can be influenced by its intelligence ($r = 0.32$ and 0.29). Therefore, children may be able to determine whether the holographic AI's behaviour is positive.

In addition, this study also illustrates a negative correlation with the children's age and subjective feelings ($r=-0.3$, $p=0.04$), in that the older children with mature cognition can recognize that the holographic AI is not real and cannot actually generate emotions like human beings. Besides, the open-ended questions also can prove this, because they think the holographic AI relies on database that can be hacked.

On the other hand, perceived competence might share an inverse relationship with children's age ($r=-0.04$). The experimental results also indicated that the children's trust may be also affected slightly by gender, in that the boys experienced less emotional responsiveness.

The children with experience in using AR and virtual agents tend to possess a higher sense of perceived integrity (r=0.12; r=0.34), but less affective responsiveness (r=-0.15; r=-0.49).

Figure 6.5 represents that this study extends the dimension of conviviality.

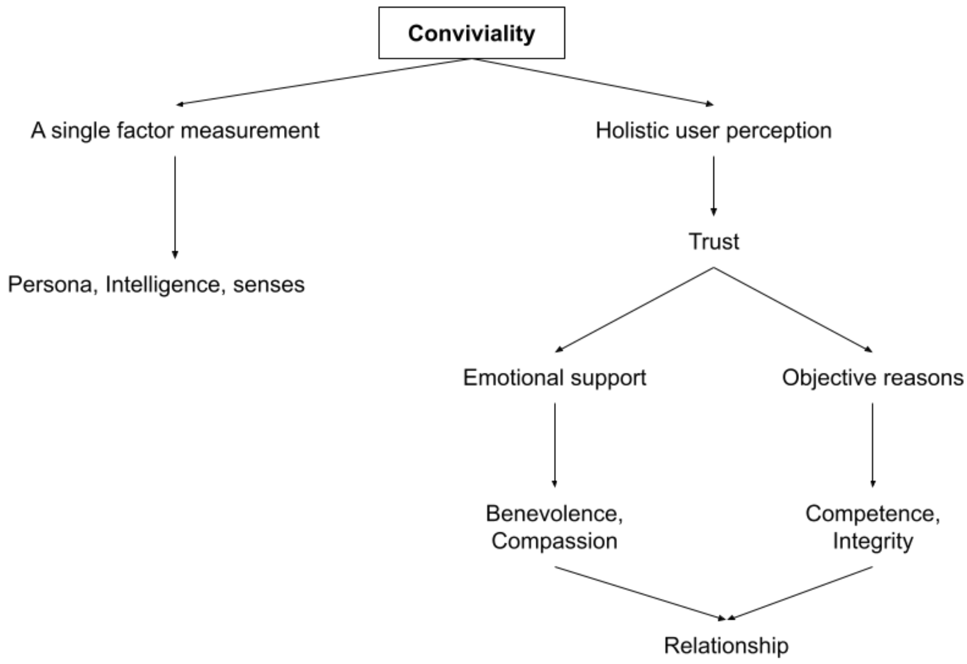


Figure 6.5. Holistic use perception in the PICS model

In the open-ended questions, only 14 children out of the 42 children said they would like to share a story with the holographic AI, since they perceived 'Sarah' as kind and helpful. Those who said they would not share secrets or stories indicated that they knew the holographic AI is not real, that the holographic AI cannot make a decision. Twenty-one children said they did not believe the holographic AI can be trusted to keep secrets; when asked why, they said that a holographic AI relies on databases, and it can be hacked.

Although the trust scale can offer content validity, adaptation for children is not examined, such that it is not likely to accurately represent actual meanings of statements. This study does not take into account educational backgrounds, cultural differences, and experiences in using AR/VR or virtual agents. Furthermore, trust is a continuous and dynamic process that necessitates increased engagement with the holographic AI to monitor its evolution.

6.2.1 Summary

This study delves into holographic AI, drawing on insights from literature reviews, case studies, and experimental research. Chapter 2 introduces the initial model, Chapter 3 refines it, and Chapters 4 and 5 offer the expanded version. As a result, a series of recommendations and guidelines are formulated for designers to comprehend the nature of holographic AI, the process of creating agents, and the additional factors to consider.

Chapter 2 featured a systematic literature review of holographic AIs, the aim of which is to identify the defining features of holographic AIs and produce an initial PICS model. This chapter then proposed a taxonomy of holographic AI features based on the PICS model, which can be used to arrive at an appropriate design structure.

In accordance with the PICS model, Chapter 3 presented a process for creating the necessary components of a holographic AI, including 3D modelling, facial and body animations, natural language processing, and spatial understanding. The study presented examples of holographic AIs developed by selecting features from the PICS model that align with the goals of specific applications. The initial model is refined based on the experience of creation to remove redundant and ambiguous elements and to uncover overlooked aspects of body language and user experience.

Chapter 4 presented an experiment using motion capture, the purpose of which was to observe and inventories key instructional gestures, as well as document gesture types not found in the literature. The results amount to an enriched taxonomy of representational gestures. It was found that the deictic gesture was the most frequently type occurring during the training sessions. Nevertheless, it is necessary for a holographic AI to perform diverse gestural forms based on different contexts. Therefore, the refined PICS model is extended in terms of persona dimension.

Finally, trust in the human-holographic AI context is a measure of the extent to which a holographic AI and its function can build a relationship with the user. To this end, an experiment for trust measurement was devised. This thesis has sought to describe in detail children's trust towards holographic AIs, and consider what scales may be employed for measuring this sense. Chapter 5 presented a Likert scale which constitutes a novel metric scale, redefines Trust, and which has been used to the proposed model of children's trust towards holographic AIs. The experimental results suggest that benevolence and compassion can influence perceived trust by understanding and recognizing children's needs, encouraging them to correct mistakes, and respecting their willingness, but it is based on whether the holographic AI is able to achieve goals by its competence. This chapter provided six recommendations for improving the user's perception of trust. The study also examines a holistic user experience, which can be used to extend the conviviality dimension of the PICS model again, since previous studies have assessed a single influence factor.

This thesis contributes to the development of holographic AI in AR and the relationship between users and such agents. Children rarely interact with holographic AIs, despite their similarities with virtual humans. In some cases, children chat with the holographic AIs according to social norms, but they do not ask questions or extend their communication. Although some children did not express uncanny valley feelings, others could sense that the holographic AI was not realistic, which may influence their sense of trust.

In this context, trust towards a holographic AI can be defined as an attitude based on the belief that the holographic AI can achieve goals with beneficial intentions and establish a relationship.

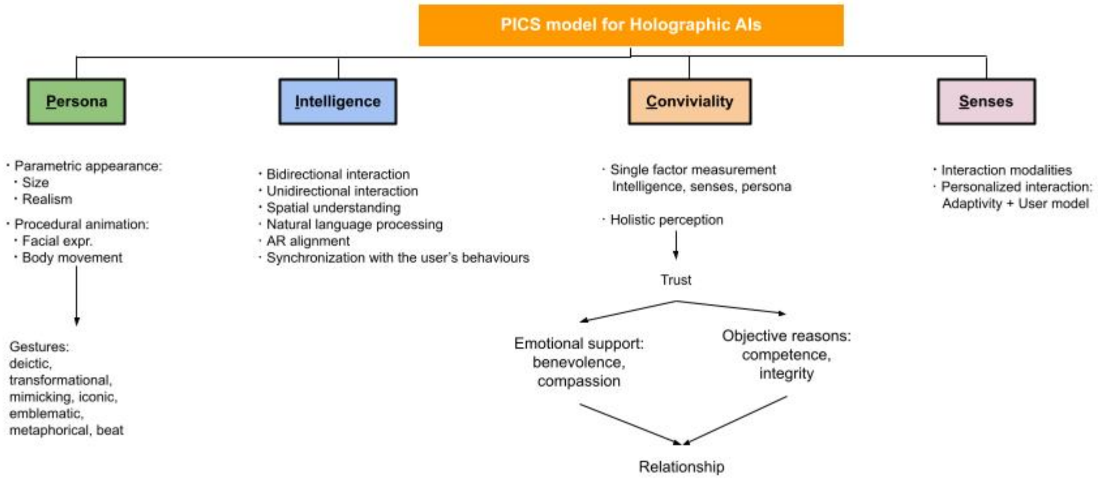


Figure 6.6. The final PICS model

Holographic AIs	Natural language processing (including translation)	Synchronization with the user's behaviour	Real-life learningal role	Independent learning rules	Teaching content (one-to-one)	Victory /defeat	Storylines
User avatars		x					
Simulation	x		x				x
Game characters (including serious games)	x			x		x	x
Intelligence tutor systems	x				x		
Chatbots	x						

Figure 6.7. Differences and similarities of types of holographic AIs

The objective of this study is to establish a comprehensive model for understanding the attributes of holographic AIs. Figure 6.6 illustrates the final iteration of the PICS model, where expanded components address the gaps in gesture performance and user perception. The persona dimension is characterized by animations and appearances. Designers may choose from an array of body languages to enrich user interaction. Moreover, the size and level of realism of the holographic AI are tailored to its intended functions. The AI's capability is signified by its intelligence. Conviviality is about not only examining individual characteristics but also assessing the overall user experience.

A taxonomy of holographic AI, grounded in literature analysis and the PICS model, is suggested. This taxonomy differentiates various types of holographic AIs, particularly emphasizing educational holographic AIs (as shown in Figure 6.7). The taxonomy's goal is to aid designers and system developers in selecting the suitable dimensions or features for crafting a holographic AI and to guide them in effectively integrating the agent's intention, motivation, and emotional support to improve interaction quality.

Techniques such as 3D scanning, reconstruction, and animation permit users to semi-automatically generate their own avatars or humanoid agents for real-world projection. This study showcases examples of educational holographic AIs. Readers can explore the contrasts between intelligent tutor systems (like the holoCARE holographic AIs discussed in Chapter 3) and simulation agents (such as Sarah highlighted in Chapter 5). Gesture studies enhance the educational holographic AI's functionality, allowing animators to select from a variety of gesture animations to vividly represent events or concepts in educational contexts, thus increasing their affordance.

Previous research has overlooked the significance of sensory perception, especially tailored interactions. Therefore, the development of holographic AIs relies on user model and adaptability to recognize users' needs and to deliver suitable exercise animations and instructional content. This includes respecting users' preferences and demonstrating social awareness. Furthermore, scaffolding /learning (also known as ‘personalisation’ / ‘adaptation’): This is personalisation and adaptation. Initially, this was included in intelligence, but in the final model this ended up under senses, stressing the input output modality aspect.

The trust model for holographic AI diverges from the traditional interpersonal trust model. This novel model encompasses five dimensions, including competence, integrity, benevolence, compassion, and relationship, which redefine the concept of trust in the context of holographic AI. It exemplifies the process of trust formation through interactions as described by the PICS model. Moreover, this study contributes to the creation of a novel scale designed to measure the overall user experience, emphasizing the elements that shape the perception of trust. While the holographic AI's competence governs its functionalities, inclusive of its intelligence and sensory inputs within the PICS model, children tend to concentrate more on the AI's emotional expressions, such as benevolence and compassion. They prefer these agents to exhibit warmth and avoid coming across as mechanical. Developers should expand their focus beyond the technical aspects of AI development to also encompass body language and social awareness, which can enhance the perceived subjective behaviour of such agents.

This study provides an in-depth examination of design principles and recommendations for crafting holographic AIs, with a special focus on the educational sector. The guidelines and suggestions stem from a comprehensive review of pertinent literature, knowledge acquired from case studies, and empirical research. Equipped with the ability to address personalized learning requirements, the holographic AI offers diverse instructional techniques to bolster engagement and interaction. Through an exploration of gestures and user experience, the holographic AI can assist in achieving objectives with positive intent and motivation, thus improving learners' understanding of material and fostering a satisfying relationship.

6.3 Empirical Contributions

This section highlights the unique contributions by the two empirical experiments to the canon of knowledge for holographic AIs.

6.3.1 Instructional gestures

The experiment in instructional gestures invited 22 participants to form 11 units (in pairs), so that one participant acts as the trainer, and the other one is acting as the trainee.

Three aspects were measured in the experiment: navigation, assembly, and precision. In each task, there are four trials, each with three conditions of co-speech gestures and a non-verbal condition. In total, 132 task samples were recoded.

The total number of representational gestures generated was 2348 nodes. The reliability value exceeded 0.7.

In total, 1290 deictic gestures, including 537 pointing position gestures, 502 pointing objects gestures, and 251 pointing direction gestures, were observed.

The number of emblematic gestures totals 364: these included 238 thumbs up; gestures, 40 counting-finger gestures, and 40 gestures related to waving index fingers or hands.

A total of 292 transformational gestures have been recorded, of which 256 involve rotation and 31 involved flipping.

A total of 198 iconic gestures were recorded, including 145 describing objects' shapes and 45 describing spatial positions.

There were 155 instances of mimicking gestures, and 41 instances of beating gestures.

Only 8 metaphorical gestures were observed.

On average, the navigation task took 68.09 seconds and generated 9.36 gestures. On average, it took 116.30 seconds for the co-speech gestures to be generated, of which 7.55 gestures were generated.

In the assembly task, the average number of gestures was 29.27 and the average time was 251 seconds. Under the co-speech condition, the average time was 522.64 seconds, and the average number of gestures was 16.58.

During the non-verbal cognition condition in the precision task the average time was 245.18 seconds and average number of gestural signals was 41.91. In trials with co-speech gestures, trainers on average spent 211.58 seconds and produced 20.21 gestures.

6.3.2 Children’s trust towards holographic AIs

In the trust experiment, 47 children between the ages of 5 and 13 took part, and all of them completed the questionnaire. Twenty-five of the children were aged 5-9, and 22 of them were aged 10-13. Out of a total of 47 children, there were 23 boys and 24 girls. A total of 15 children had experience with AR, and 10 had experience with virtual agents, and only 6 children had previously used both technologies. The statistical findings based on the children’s responses are as follows:

The mean value of perceived competence is 3.76, that of integrity is 3.62, that of benevolence is 3.74, that of compassion is 3.43, that of relationship is 3.26, and that of trust is 3.49.

The correlation between experience in using virtual agents and perceived integrity is 0.34 ($p=0.01$).

The correlation between experience in using virtual agents and perceived benevolence is -0.49 ($p=0.04$).

The correlation between experience in using virtual agents and trust is -0.26 ($p=0.08$).

The correlation between the Likert-based responses to statement #3 (“I think positively about the hologram” / “I like Sarah”) and experience in using virtual agents is 0.33 (p=0.01).

The correlation between the Likert-based responses to statement #4 (“The hologram answered my questions truthfully” / “Sarah doesn’t lie”) and experience in using AR is 0.24 (p=0.04).

The correlation between the Likert-based responses to statement #6 (“The hologram is benevolence” / “Sarah is kind”) and age is -0.31 (p=0.05).

The correlation between the Likert-based responses to statement #7 (“The hologram feels real to me” / “Sarah feels real to me”) and experience in using AR is 0.13 (p=0.5).

The correlation between the Likert-based responses to statement #8 (“The hologram looks out for me” / “Sarah looks out for me”) and relationship is 0.5 (p=0.0008).

The correlation between the Likert-based responses to statement #10 (“The hologram is compassionate” / “Sarah is caring”) and relationship is 0.48 (p=0.0014).

The correlation between benevolence and trust is 0.71.

6.4 Limitations

As discussed in Section1.2, this thesis focuses on the domain of holographic AIs in order to provide a comprehensive investigation into users’ perceptions and the consistency of holographic AI performance, especially holographic AI instructional gestures and associated sense of trust. This thesis has revealed that the proposed PICS model could serve as a design structure that highlights the key elements that make a holographic AI seem engaging and trustworthy. A holographic AI’s instructional gestures should align with its intention and context. Further, affective-based trust is critical for fostering children’s sense of trust towards holographic AIs.

However, this research project has its limitations. The systematic literature review may have overlooked significant previous studies and R&D milestones regarding holographic AIs. To be more specific, although the results of the systematic review were selected on the basis of word frequency and cluster analysis to ensure relevance and accuracy, only 49 studies were chosen. This model relies on previous findings to gather and re-organize the holographic AI’s features, thus dimensions and elements cannot represent possibilities of future development to produce a general model for all types of holographic AIs, since AR technology is wildly developing. Several aspects of the initial model proposed are not clearly delineated. For example, the domain of computer vision encompasses spatial understanding, recognition of physical objects, and projection techniques. The concept of conviviality should not be constrained to just three components. Moreover, although learning systems incorporate a broad spectrum of ideas, this study does not delve into the educational content delivered by holographic AI, the interaction processes involved, or their application within educational holographic AIs. Besides, this study does not discuss the holographic AI’s styles of appearance, clothing, jobs, rendering, or compression. The classification of holographic AI is also founded on a limited number of prior studies, despite the study's suggestion that future agents might be integrated and utilized across various applications.

Pursuant to the PICS model and classification, the study outlines various methods for creating holographic AIs, ranging from 3D modelling to the cultivation of intelligence. However, it primarily addresses intelligent tutor systems, which may not be suitable for all types of holographic AIs. The created holographic AI cannot perform bidirectional interaction, i.e. manipulate real objects. Besides, although it explains how to use 3D scanning to reconstruct a user avatar, the second method of traditional 3D modelling in Chapter 3 only presented the head of a stylized holographic AI, which was compared with a 3D scanned holographic AI.

Key instructional gestures are explored for the holographic AI in Chapter 4. Referring to the experiment of instructional gestures, the Rokoko motion capture suits offer the advantage of portability, and so are not restricted to a fixed motion space. However, it has its technological limitations, since the suit is sensitive to magnetic objects – which could not easily be avoided in the experimental rooms – thus sensors of the Rokoko suit often disconnected, and two sensors on the left side of the leg malfunctioned in some groups. The nodes of gestural generation in data analysis also may have missing values, especially for the thumbs up gesture, since these emblematic gestures tended to be generated too quickly. Although the total nodes coded for gestures generated exceeded 2000, the experiment involved only 11 participants, and numbers of gestural forms produced may vary with culture, age, and gender, for example, this experiment only collected movements of three females.

Regarding data analysis, the instructional methods could have influenced the generation of gestures, given that the analysis employed the average number of gestures rather than a correlation coefficient. Consequently, these findings can only be regarded as inferential. The experimental design paves the way for the discovery of new types of gestures, implying that alternate tasks could yield different gestures. Apart from cultural and educational experiences, the experience of participants might also impact the gestures they use. For example, the study does not consider whether prior experience in assembling cardboard models or completing puzzle-T tasks could affect gesture production.

According to Chapter 2, previous studies focus on how a specific aspect affects user experience, such as degrees of realism and intelligence, thus the empirical experiment on trust in holographic AIs among children measured the holistic situation of the holographic AI in Chapter 6, and did not investigate all aspects of a holographic AI, such as the level of anthropomorphism or utterance. In terms of scale adaptation, the transformation of the trust scale to be more "child-friendly" has not undergone revalidation. It could be improved with input from language experts to ensure that its simplification does not compromise its precision. As mentioned above, it had 10 missing values that were excluded from analysis. However, according to received wisdom, a sample of 47 people is unlikely to be representative of the human population, and the analysis thereof is likely to contain a sizeable degree of bias.

Given that this experiment employs a correlation coefficient for comparison, which serves as a primary factor, it is challenging to assert with certainty that prior experience with AR/VR and virtual agents can influence perceptions of a holographic AI's competence. Variables such as participants' educational levels, cultural backgrounds, experiences with recognizing 3D shapes, gender, and age have not been meticulously controlled or analysed within this study. These factors could influence the outcomes. If older children are already familiar with the content, their level of engagement might diminish. There may be an inherent assumption that the holographic AI is trustworthy when introduced by the experimenter. The proposed trust model for holographic AI does not account for engagement, despite its critical role in the development of trust.

6.5 Future Research Directions

This project has been an endeavour into understanding how human-agent interaction enhances the creation of augmented reality and improves the user's sense of trust. It is hoped that the findings of this thesis will contribute significantly to future holographic AI research. There exist a number of avenues along which holographic AIs could be improved in terms of realism and trust:

One future research direction is to improve the anthropomorphism of holographic AIs to avoid the Uncanny Valley effect and facilitate behavioural realism, especially in the aspect of persona (appearance and behaviours). Although there now exist holographic AIs with humanlike appearance and animations, in many cases their performance tends to be mechanical. The study will further explore how the visual representation of holographic AI may be affected by societal stereotypes, including aspects like appearance, hairstyle, attire, and perceived professional roles. In Unreal Engine (Unreal Engine, 2023), a game engine, users can create photorealistic virtual humans by assembling different shapes and colours of eyes, lips, skin, hair, etc. However, this application has a limitation when it comes to low polygon counts. For this reason, it will continue to seek a way to create high-quality holographic AI for AR smart glasses. The holographic AI will enhance interactivity by analysing more types of representational gestures, exploring relationships between cognition, body language, and emotional expressions. It is also critical to observe presentation performance to investigate how people explain an event when they are unable to see referents and rely on memory or imagination. Interactive gestures for holographic AIs will be identified, expanding the persona's dimensionality. Variables such as cultural background, educational level, gender, and prior experiences will be factored into experimental designs.

The second future research direction could be to implement natural language processing and enrich dialogue management. While IBM Watson is capable of providing basic verbal interaction, the utterance of its holographic AI appears delayed, and its speaking rhythm is not natural. Open AI's ChatGPT (OpenAI, 2022) relies on a language model and a large amount of text data that can implement free talk, which causes ChatGPT's responses may not be accurate. Therefore, there remains the need for researchers to develop deep reinforcement learning and a Recurrent Neural Network for holographic AIs. Besides, co-speech gestures could be used to explore how holographic AI can match appropriate gestures and more effectively interact with users.

Bidirectional interaction is the third direction. In order to recognize and manipulate real objects, the holographic AI should integrate with additional sensors and actuators. Therefore, in future, the holographic AI with physical-object influence will be developed, for example it can move 3D shapes by hands to instruct children to learn geometry.

A fourth direction is investigation into user perception. Although this thesis has measured children’s sense of trust towards one holographic AI, trust itself cannot be fully gauged using only a short simulation. For a fuller analysis of perceived trust, the holographic AI in question needs a plausible story background, personality, competence, emotional responsiveness, and consistency of performance, so that more practical benefits can be realised from improved perception. Besides, the trust scale will be refined to ensure that children can comprehend the meaning of statements accurately, taking into account factors such as educational level, age, gender, and previous exposure to AR/VR and virtual assistants. Since the proposed trust model does not specifically address user engagement, subsequent studies will assess usability, acceptance, and user expectations to broaden the dimension of conviviality within the PICS model.

Fifth, future research should focus on an integrated holographic AI, rather than limiting holographic AIs to a specific application or function. For example, a holographic AI might be able to play the role of a virtual teacher providing one-to-one learning content, and also play the role of a simulation agent implementing practical training following teaching. A real teacher could design a scenario based on the student’s requirements, in which a holographic AI’s design no longer is confined to uniform interaction. Besides, Tsovaltzi et al. (2012) identified that learners can achieve more outcomes through correcting erroneous examples, thus, holographic AIs should guide and courage learners to consolidate achieved knowledge and skills.

Advancements in educational holographic AI are transforming and diversifying educational methods and channels, facilitating immersive and MR environments that support personalized teaching approaches and integrate AI-powered assessment tools, instant feedback, and tailored teaching content. Therefore, future research will persist in advancing holographic AI technology and refining the PICS model. In the realm of educational technology, discussions will also revolve around how the integration of interactive processes and design can enhance learners' achievements and practical experiences.

References

Abner, N., Cooperrider, K. and Goldin-Meadow, S. (2015) ‘Gesture for linguists: a handy primer’, Language and Linguistics Compass, 9(11), pp. 437-451.

Achenbach, J., Waltemate, T., Latoschik, M.E. and Botsch, M. (2017) ‘Fast generation of realistic virtual humans’, Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. Gothenburg, 8-10 November. New York, NY: Association for Computing Machinery, pp. 1-10.

Ahmad, S., Umirzakova, S., Jamil, F. and Whangbo, T.K. (2022) ‘Internet-of-things-enabled serious games: a comprehensive survey’, Future Generation Computer Systems, 136, pp. 67-83.

Ahmed, A.; Sutton, M.J. (2017) ‘Gamification, serious games, simulations, and immersive learning environments in knowledge management initiatives’. World J. Sci. Technol. Sustain. Dev. 2017, 14, pp. 78–83.

Ali, G., Le, H.Q., Kim, J., Hwang, S.W. and Hwang, J.I. (2019) ‘Design of seamless multi-modal interaction for intelligent virtual agents in wearable mixed reality environment’, Proceedings of the 32nd International Conference on Computer Animation and Social Agents. Paris, 1-3 July. New York, NY: Association for Computing Machinery, pp. 47–52.

Almufarreh, A. and Arshad, M. (2023) ‘Exploratory Students’ Behavior towards Massive Open Online Courses: A Structural Equation Modeling Approach’, Systems, 11(5), p. 223.

Alrakhawi, H. A., Jamiat, N., Abu-Naser, S. S. (2023) ‘Intelligent Tutoring Systems in Education: A Systematic Review of Usage, Tools, Effects and Evaluation’. Journal of Theoretical and Applied Information Technology, 2023, Vol. 101. No. 4, pp. 1205-1226.

Andersson, C., Stenfors, C.U.D., Lilliengren, P., Einhorn, S. and Osika, W. (2021) ‘Benevolence – Associations With Stress, Mental Health, and Self-Compassion at the Workplace’. Frontiers in Psychology, 12.

Andreu Andrés, M. A., & García Casas, M. (2011) ‘Perceptions of gaming as experiential learning by engineering students’, International Journal of Engineering Education, 27(4), pp. 795–804.

Andric, M. and Small, S. (2012) ‘Gesture’s neural language’, Frontiers in Psychology, 3, p. 99.

Aramaki, R. and Murakami, M. (2013) ‘Investigating appropriate spatial relationship between user and ar character agent for communication using AR WoZ system’, Proceedings of the 15th ACM on International Conference on Multimodal Interaction. Sydney, 9-13 December. New York, NY: Association for Computing Machinery, pp. 397-404.

Aroyo, L., Dolog, P., Houben, G.J., Kravcik, M., Naeve, A., Nilsson, M. and Wild, F. (2006) ‘Interoperability in personalized adaptive learning’, Journal of Educational Technology & Society, 9(2), pp. 4-18.

Atit, K., Gagnier, K. and Shipley, T.F. (2015) ‘Student gestures aid penetrative thinking’, Journal of Geoscience Education, 63(1), pp. 66-72.

Atit, K., Resnick, I., Shipley, T., Ormand, C., Manduca, C., Goksun, T. and Tikoff, B. (2013) ‘Spatial gestures point the way: a broader understanding of the gestural referent’, in Knauff, M., Pauen, M., Sebanz, N. and Wachsmuth, I. (eds.), Proceedings of the 35th Annual Meeting of the Cognitive Science Society. Berlin, 31 July-3 August. Austin, TX: Cognitive Science Society.

Azuma Hikari (2023), Available at: https://www.gatebox.ai/ (Accessed: 15 August 2023).

Barakonyi, I. and Schmalstieg, D. (2005) Exploiting the physical world as user interface in augmented reality applications. Available at: http://www.barakonyi.net/papers/arpuppet_vr05_3dui_ws.pdf (Accessed: 15 March 2023).

Barakonyi, I., Psik, T. and Schmalstieg, D. (2004) ‘Agents that talk and hit back: animated agents in augmented reality’, Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality. Arlington, VA, 5 November. Piscataway, NJ: IEEE, pp. 141-150.

Barrett, A.M., Foundas, A.L. and Heilman, K.M. (2005) ‘Speech and gesture are mediated by independent systems’, Behavioral and Brain Sciences, 28(2), pp. 125-126.

Bartol, K., Bojanić, D., Petković, T. and Pribanić, T. (2021) ‘A review of body measurement using 3D scanning’, IEEE Access, 9, pp. 67281-67301.

Bates, E. (1976) Language and context: the acquisition of pragmatics. New York, NY: Academic Press.

Bavelas, J.B., Chovil, N., Lawrie, D.A. and Wade, A. (1992) ‘Interactive gestures’, Discourse Processes, 15(4), pp. 469-489.

Baylor, A.L. (2009) ‘Promoting motivation with virtual agents and avatars: role of visual presence and appearance’, Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1535), pp. 3559-3565.

Benyon, D., Turner, P. and Turner, S. (2006) Designing interactive systems: people, activities, contexts, technologies, Harlow, Munich, Addison-Wesley.

Bernard, J.A., Millman, Z.B. and Mittal, V.A. (2015) ‘Beat and metaphoric gestures are differentially associated with regional cerebellar and cortical volumes’, Human Brain Mapping, 36(10), pp. 4016-4030.

Bernath, M.S. and Feshbach, N.D. (1995) ‘Children’s trust: theory, assessment, development, and research directions’, Applied and Preventive Psychology, 4(1), pp. 1-19.

Bhattacharya, R., Devinney, T.M. and Pillutla, M.M. (1998) ‘A formal model of trust based on outcomes’, Academy of Management Review, 23(3), pp. 459-472.

Blanke, O. and Metzinger, T. (2009) ‘Full-body illusions and minimal phenomenal selfhood’, Trends in Cognitive Sciences, 13(1), pp.7–13.

Böhm, T. and Stein, J. (2019) ‘From HCI to HRI: About Users, Acceptance and Emotions’, Human Systems Engineering and Design II Advances in Intelligent Systems and Computing, pp. 149–153.

Bönsch, A., Radke, S., Overath, H., Asché, L.M., Wendt, J., Vierjahn, T., Habel, U. and Kuhlen, T.W. (2018) ‘Social VR: how personal space is affected by virtual agents’ emotions’, Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Tuebingen, 18-22 March. Piscataway, NJ: IEEE, pp. 199-206.

Borum, R. (2010) The science of interpersonal trust. Mental Health Law & Policy Faculty report No. 574. Available at: https://digitalcommons.usf.edu/mhlp_facpub/574/ (Accessed: 15 March 2023).

Bourgeois, P. and Hess, U. (2008) ‘The impact of social context on mimicry’, Biological Psychology, 77(3), pp. 343-352.

Bradley, M.M. and Lang, P.J. (1994) ‘Measuring emotion: the self-assessment manikin and the semantic differential’, Journal of Behavior Therapy and Experimental Psychiatry, 25(1), pp. 49-59.

Bryman, A. (2012) Social research methods. 4th edn. Oxford: Oxford University Press.

Bulu, S.T. (2012) ‘Place presence, social presence, co-presence, and satisfaction in virtual worlds’, Computers & Education, 58(1), pp. 154-161.

Burden, D. and Savin-Baden, M. (2019) Virtual humans: today and tomorrow. Boca Raton, FL: CRC Press.

Caire, P. (2010) ‘How to import the concept of conviviality to web communities’, International Journal of Web Based Communities, 6(1), p. 99.

Caire, P. and van der Torre, L. (2009) ‘Convivial Ambient Technologies: Requirements, Ontology and Design’, The Computer Journal, 53(8), pp.1229–1256.

Calvo-Barajas, N. and Castellano, G. (2022) ‘Understanding children’s trust development through repeated interactions with a virtual social robot’, Proceedings of the 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). Napoli, 29 August-2 September. Piscataway, NJ: IEEE, pp. 1451-1458.

Campbell, A.G., Stafford, J.W., Holz, T. and O’Hare, G.M.P. (2014) ‘Why, when and how to use augmented reality agents (AuRAs)’, Virtual Reality, 18(2), pp. 139-159.

Capin, T.K., Noser, H., Thalmann, D., Pandzic, I.S. and Thalmann, N.M. (1997) ‘Virtual human representation and communication in VLNet’, IEEE Computer Graphics and Applications, 17(2), pp. 42-53.

Carroll, J.M. and Russell, J.A. (1996) ‘Do facial expressions signal specific emotions? Judging emotion from the face in context’, Journal of Personality and Social Psychology, 70(2), pp. 205-218.

Carrozzi, A., Chylinski, M., Heller, J., Hilken, T., Keeling, D.I. and de Ruyter, K. (2019) ‘What’s mine is a hologram? How shared augmented reality augments psychological ownership’, Journal of Interactive Marketing, 48(1), pp. 71-88.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H. and Yan, H. (2001) ‘More than just a pretty face: conversational protocols and the affordances of embodiment’, Knowledge-Based Systems, 14(1), pp. 55-64.

Cassell, J., Vilhjálmsson, H.H. and Bickmore, T. (2001) ‘BEAT: the behavior expression animation toolkit’, Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, CA, 12-17 August. New York, NY: Association for Computing Machinery, pp. 477-486.

Catrambone, R., Stasko, J. and Xiao, J. (2002) ‘Anthropomorphic agents as a user interface paradigm: experimental findings and a framework for research’, in Gray, W.D. and Schunn, C.D. (eds.), Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society. Fairfax, VA, 7-10 August. Mahwah, NJ: Lawrence Erlbaum, pp. 1-6.

Chahyana, J. and Yesmaya, V. (2020) ‘Virtual pet simulator game using augmented reality on Android platform’, Journal of Physics: Conference Series, 1566(1), p. 012088.

Chan, J.C.P., Leung, H., Tang, J.K.T. and Komura, T. (2011) ‘A virtual reality dance training system using motion capture technology’, IEEE Transactions on Learning Technologies, 4(2), pp. 187-195.

Chang, L.J., Doll, B.B., van ‘t Wout, M., Frank, M.J. and Sanfey, A.G. (2010) ‘Seeing is believing: trustworthiness as a dynamic belief’, Cognitive Psychology, 61(2), pp. 87-105.

Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H. and Tian, M. (2019) ‘An overview of augmented reality technology’, Journal of Physics: Conference Series, 1237(2), p. 022082.

Chetty, G. and White, M. (2019) ‘Embodied conversational agents and interactive virtual humans for training simulators’, Proceedings of 2019 International Conference on Auditory-Visual Speech Processing. Melbourne, 10-11 August. Melbourne: International Speech Communication Association, pp. 73-77.

Chu, C.T., Lugtenberg, G., Yamamoto, G., Taketomi, T., Miyazaki, J. and Kato, H. (2012) ‘Augmented reality agent as desktop assistant’, Proceedings of First International Symposium on Socially and Technically Symbiotic Systems. Okayama, 29-31 August. Kyoto: Socially and Technically Symbiotic Systems.

Chuang, C.H., Lo, J.H. and Wu, Y.K. (2023) ‘Integrating chatbot and augmented reality technology into biology learning during COVID-19’, Electronics, 12(1), p. 222.

Churi, P. P., Joshi, S., Elhoseny, M., & Omrane, A. (Eds.). (2022) ‘Artificial intelligence in higher education: A practical approach (1st ed.)’, CRC Press.

Cipresso, P., Giglioli, I.A.C., Raya, M.A. and Riva, G. (2018) ‘The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature’, Frontiers in Psychology, 9, p. 2086.

Clough, S. and Duff, M.C. (2020) ‘The role of gesture in communication and cognition: implications for understanding and treating neurogenic communication disorders’, Frontiers in Human Neuroscience, 14, p. 323.

Cohen, J. (1960) ‘A coefficient of agreement for nominal scales’, Educational and Psychological Measurement, 20(1), pp. 37-46.

Colquitt, J.A., Scott, B.A. and LePine, J.A. (2007) ‘Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance’, Journal of Applied Psychology, 92(4), pp. 909-927.

Congdon, E.L., Novack, M.A., Brooks, N., Hemani-Lopez, N., O’Keefe, L. and Goldin-Meadow, S. (2017) ‘Better together: simultaneous presentation of speech and gesture in math instruction supports generalization and retention’, Learning and Instruction, 50, pp. 65-74.

Cook, S. W., Friedman, H. S., Duggan, K. A., Cui, J. and Popescu, V. (2016) ‘Hand Gesture and Mathematics Learning: Lessons From an Avatar’, Cognitive Science, vol. 41, no. 2, pp. 518–535.

Cook, S. W., Yip, T. K. and Goldin-Meadow, S. (2010) ‘Gesturing makes memories that last’, Journal of Memory and Language, vol. 63, no. 4, pp. 465–475.

Cook, S.W., Mitchell, Z. and Goldin-Meadow, S. (2008) ‘Gesturing makes learning last’, Cognition, 106(2), pp. 1047-1058.

Craig, S.D., Twyford, J., Irigoyen, N. and Zipp, S.A. (2015) ‘A test of spatial contiguity for virtual human’s gestures in multimedia learning environments’, Journal of Educational Computing Research, 53(1), pp. 3-14.

Cummings, S., Browner, W., Grady, D., Newman, T. and Hulley, S. (2007) ‘2 Conceiving The Research Question’, Designing Clinical Research.

Curioni, A., Knoblich, G.K., Sebanz, N. and Sacheli, L.M. (2020) ‘The engaging nature of interactive gestures’, PLoS One, 15(4), p. e0232128.

Dahleez, K., ElSaleh, A., Alawi, A. and Abdelfattah, F. (2021) ‘Student learning outcomes and online engagement in time of crisis: the role of elearning system usability and teacher behavior’, The International Journal of Information and Learning Technology, vol. 38, pp. 473-492.

Davis, R.O., Wan, L.L., Vincent, J. and Lee, Y.J. (2021) ‘The effects of virtual human gesture frequency and reduced video speed on satisfaction and learning outcomes’, Educational Technology Research and Development, 69(5), pp. 2331-2352.

de Barcelos Silva, A., Gomes, M.M., da Costa, C.A., da Rosa Righi, R., Barbosa, J.L.V., Pessin, G., De Doncker, G. and Federizzi, G. (2020) ‘Intelligent personal assistants: a systematic literature review’, Expert Systems with Applications, 147, p. 113193.

de Borst, A.W. and de Gelder, B. (2015) ‘Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective’, Frontiers in Psychology, 6, p. 576.

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F. and Parasuraman, R. (2016) ‘Almost human: Anthropomorphism increases trust resilience in cognitive agents.’, Journal of Experimental Psychology: Applied, vol. 22, no. 3, pp. 331–349.

Dell’Aquila, E., Vallone, F., Zurlo, M.C. and Marocco, D. (2022) ‘SG-ACCORD: designing virtual agents for soft skills training in the school context’, Education Sciences, 12(3), p. 174.

Demeure, V., Niewiadomski, R. and Pelachaud, C. (2011) ‘How is believability of a virtual agent related to warmth, competence, personification, and embodiment?’, Presence, 20(5), pp. 431-448.

Di Battista, S., Pivetti, M. and Berti, C. (2020) ‘Competence and benevolence as dimensions of trust: lecturers’ trustworthiness in the words of Italian students’, Behavioral Sciences, 10(9), p. 143.

Dietz, G. (2011) ‘Going back to the source: why do people trust each other?’, Journal of Trust Research, 1(2), pp. 215-222.

Dineshshankar, J., Ganapathi, N., Yoithapprabhunath, T.R., Maheswaran, T., Kumar, M.S. and Aravindhan, R. (2013) ‘Lip prints: role in forensic odontology’, Journal of Pharmacy & Bioallied Sciences, 5(Suppl 1), pp. S95-S97.

Dobel, J.P. (1999) Public integrity. Baltimore: Johns Hopkins University Press.

Dryer, D.C. (1999) ‘Getting personal with computers: how to design personalities for agents’, Applied Artificial Intelligence, 13(3), pp. 273-295.

Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., Pierce, L.G. and Beck, H.P. (2003) ‘The role of trust in automation reliance. Trust and Technology’, International Journal of Human-Computer Studies, 58(6), pp.697–718.

Elmahal, D.M., Ahmad, A.S., Alomaier, A.T., Abdlfatah, R.F. and Hussein, D.M. (2020) ‘Comparative study between hologram technology and augmented reality’, Journal of Information Technology Management, 12(2), pp. 90-106.

Elster, J. (2007) Explaining social behavior: more nuts and bolts for the social sciences. Cambridge: Cambridge University Press.

Ferstl, Y., Thomas, S., Guiard, C., Ennis, C. and McDonnell, R. (2021) ‘Human or robot? Investigating voice, appearance and gesture motion realism of conversational social agents’, Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents. Virtual Event, 14-17 September. New York, NY: Association for Computing Machinery, pp. 76-83.

Fettke, P. (2009) ‘How Conceptual Modeling Is Used’, Communications of the Association for Information Systems, vol. 25 [Online]. DOI: https://doi.org/10.17705/1cais.02543 (Accessed: 15 March 2023).

Fiske, S.T., Cuddy, A.J.C. and Glick, P. (2007) ‘Universal dimensions of social cognition: warmth and competence’, Trends in Cognitive Sciences, 11(2), pp. 77-83.

Flament, F., Francois, G., Qiu, H., Ye, C., Hanaya, T., Batisse, D., Cointereau-Chardon, S., Seixas, M.D.G., Dal Belo, S.E. and Bazin, R. (2015) 'Facial skin pores: a multiethnic study', Clinical, Cosmetic and Investigational Dermatology, 8, pp. 85-93.

Fogg, B.J. and Tseng, H. (1999) 'The elements of computer credibility', Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Pittsburgh, PA, 15-20. New York, NY: Association for Computing Machinery, pp. 80-87.

Fuks, O. (2014) 'Gradient and categorically: Handshape's two semiotic dimensions in Israeli Sign Language discourse', Journal of Pragmatics, 60, pp. 207-225.

García-Pérez, M. A. (2012) 'Statistical Conclusion Validity: Some Common Threats and Simple Remedies', Frontiers in Psychology, vol. 3.

Gebru, B., Zeleke, L., Blankson, D., Nabil, M., Nateghi, S., Homaifar, A. and Tunstel, E. (2022) 'A Review on Human–Machine Trust Evaluation: Human-Centric and Machine-Centric Perspectives', IEEE Transactions on Human-Machine Systems, 52(5), pp. 952 – 962.

Geiskkovitch, D.Y., Thiessen, R., Young, J.E. and Glenwright, M.R. (2019) 'What? that's not a chair!: how robot informational errors affect children's trust towards robots', Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Daegu, 11-14 March. Piscataway, NJ: IEEE, pp. 48-56.

Ghasemi, Y., Jeong, H., Choi, S. H., Park, K.-B. and Lee, J. Y. (2022) 'Deep learning-based object detection in augmented reality: A systematic review', Computers in Industry, vol. 139, p. 103661.

Ghorbani, F., Taghavi, M.F. and Delrobaei, M. (2022) 'Towards an intelligent assistive system based on augmented reality and serious games', Entertainment Computing, 40, p. 100458.

Gkinko, L. and Elbanna, A. (2023) 'Designing trust: the formation of employees' trust in conversational AI in the digital workplace', Journal of Business Research, 158, p. 113707.

Goldin-Meadow, S. (2011) 'Learning through gesture', WIREs Cognitive Science, 2(6), pp. 595-607.

Goldin-Meadow, S. and Alibali, M.W. (2013) 'Gesture's role in speaking, learning, and creating language', Annual Review of Psychology, 64(1), pp. 257-283.

Goldin-Meadow, S. and Brentari, D. (2017) 'Gesture, sign, and language: the coming of age of sign language and gesture studies', Behavioral and Brain Sciences, 40, p. e46.

Golembiewski, R.T. and McConkie, M. (1975) 'The centrality of interpersonal trust in group processes', in Cooper, C.L. (ed.) Theories of group processes. New York, NY: John Wiley & Sons, pp. 131-185.

Graziano, M. and Gullberg, M. (2018) ‘When speech stops, gesture stops: evidence from developmental and crosslinguistic comparisons’, Frontiers in Psychology, 9, pp. 879.

Grivokostopoulou, F., Kovas, K. and Perikos, I. (2020) ‘The effectiveness of embodied pedagogical agents and their impact on students learning in virtual worlds’, Applied Sciences, 10(5), p. 1739.

Guazzaroni, G. (2022) ‘Virtual and augmented reality in art during the pandemic’, in Pillai, A.S. and Guazzaroni, G. (eds.) Extended reality usage during COVID 19 pandemic. Cham: Springer, pp. 89-93.

Guerra-Carrillo, B., Katovich, K. and Bunge, S. A. (2017) ‘Does higher education hone cognitive functioning and learning efficacy? Findings from a large and diverse sample’, Dalby, A. R. (ed), PLOS ONE, vol. 12, no. 8, p. e0182276.

Guneysu, A. and Arnrich, B. (2017) ‘Socially assistive child-robot interaction in physical exercise coaching’, Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). Lisbon, 28 August-1 September. Piscataway, NJ: IEEE, pp. 670-675.

Gunter, T.C., Weinbrenner, J.E.D. and Holle, H. (2015) ‘Inconsistent use of gesture space during abstract pointing impairs language comprehension’, Frontiers in Psychology, 6, p. 80.

Gupta, K., Hajika, R., Pai, Y.S., Duenser, A., Lochner, M. and Billinghurst, M. (2019) ‘In AI we trust: investigating the relationship between biosignals, trust and cognitive load in VR’, Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology. Parramatta, 12-15 November. New York, NY: Association for Computing Machinery, pp. 1-10.

Hachimura, K., Kato, H. and Tamura, H. (2004) ‘A prototype dance training support system with motion capture and mixed reality technologies’, Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication. Okayama, 22 September. Piscataway, NJ: IEEE, pp. 217-222.

Hadar, U., Dar, R. and Teitelman, A. (2001) ‘Gesture during speech in first and second language: implications for lexical retrieval’, Gesture, 1(2), pp. 151-165.

Hager, P. and Gonczi, A. (1996) ‘What is competence?’, Medical Teacher, 18(1), pp. 15-18.

Hanna, N. and Richards, D. (2019) ‘Speech Act Theory as an evaluation tool for human–agent communication’, Algorithms, 12(4), p. 79.

Hardin, R. (2006) Trust. Cambridge: Polity Press.

Hartholt, A., Mozgai, S., Fast, E., Liewer, M., Reilly, A., Whitcup, W. and Rizzo, A.S. (2019) ‘Virtual humans in augmented reality: a first step towards real-world embedded virtual roleplayers’, Proceedings of the 7th International Conference on Human-Agent Interaction. Kyoto, 6-10 October. New York, NY: Association for Computing Machinery, pp. 205-207.

Hartholt, A., Traum, D., Marsella, S.C., Shapiro, A., Stratou, G., Leuski, A., Morency, L.P. and Gratch, J. (2013) ‘All together now’, in Aylett, R., Krenn, B., Pelachaud, C. and Shimodaira, H. (eds.), Proceedings of the 13th International Conference on Intelligent Virtual Agents. Edinburgh, 29-31 August. Berlin: Springer, pp. 368-381.

Hartson, H.R. and Pyla, P. S. (2019) ‘The UX book: Agile UX design for a quality user experience’, 2nd edn, Cambridge, Ma, Morgan Kaufmann.

Hassemer, J. and McCleary, L. (2018) ‘The multidimensionality of pointing’, Gesture, 17(3), pp. 417-463.

Helm, B.W. (2014) ‘Trust as a reactive attitude’, in Shoemaker, D. and Tognazzini, N. (eds.) Oxford studies in agency and responsibility. Oxford: Oxford University Press, pp. 187-215.

Herpich, F., Guarese, R.L.M. and Tarouco, L.M.R. (2017) ‘A comparative analysis of augmented reality s aimed at the development of educational applications’, Creative Education, 8(9), pp. 1433-1451.

Hinnell, J. and Parrill, F. (2020) ‘Gesture influences resolution of ambiguous statements of neutral and moral preferences’, Frontiers in Psychology, 11, p. 587129.

Hoffman, R.R., Johnson, M., Bradshaw, J.M. and Underbrink, A. (2013) ‘Trust in Automation’. IEEE Intelligent Systems, 28(1), pp.84–88.

Holz, T., Campbell, A.G., O’Hare, G.M.P., Stafford, J.W., Martin, A. and Dragone, M. (2011) ‘MiRA—mixed reality agents’, International Journal of Human-Computer Studies, 69(4), pp. 251-268.

Hornbæk, K. and Oulasvirta, A. (2017) ‘What Is Interaction?’, Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. Denver, CO, 6-11 May. New York, NY: Association for Computing Machinery, pp. 5040-5052.

Hostetter, A.B. (2011) ‘When do gestures communicate? A meta-analysis’, Psychological Bulletin, 137(2), pp. 297-315.

Hu, P., Kaashki, N.N., Dadarlat, V. and Munteanu, A. (2021) ‘Learning to estimate the body shape under clothing from a single 3-D scan’, IEEE Transactions on Industrial Informatics, 17(6), pp. 3793-3802.

Huang, A., Knierim, P., Chiossi, F., Chuang, L.L. and Welsch, R. (2022) ‘Proxemics for human-agent interaction in augmented reality’, Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. New Orleans, LA, 29 April-5 May. New York, NY: Association for Computing Machinery, pp. 1-13.

Huang, X. and Wild, F. (2021) ‘A new metric scale for measuring trust towards holographic intelligent agents’, Proceedings of the 34th British HCI Workshop and Doctoral Consortium 2021. Virtual Event, 20-21 July. Swindon: BCS Learning and Development Ltd, pp. 1-8.

Huang, X., Wild, F. and Whitelock, D. (2021) ‘Design dimensions for holographic intelligent agents: a comparative analysis’, Proceedings of the 1st International Workshop on Multimodal Artificial Intelligence in- Education (MAIED 2021). Utrecht, 10-14 June. Aachen: CEUR, pp. 17-26.

IBM (2023) IBM Watson is AI for business. Available at: https://www.ibm.com/uk-en/watson (Accessed: 15 March 2023).

Imber, S. (1973) 'Relationship of trust to academic performance', Journal of Personality and Social Psychology, 28(1), pp. 145-150.

Iqbal, M.Z., Mangina, E. and Campbell, A.G. (2019) 'Exploring the use of augmented reality in a kinesthetic learning application integrated with an intelligent virtual embodied agent', Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). Beijing, 10-18 October. Piscataway, NJ: IEEE, pp. 12-16.

IshowoOloko, F., Bonnefon, J., Soroye, Z., Crandall, J., Rahwan, I. and Rahwan, T. (2019) 'Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation', Nature Machine Intelligence, [Online] 1(11), pp.517–521. DOI: https://doi.org/10.1038/s4225601901135 (Accessed: 18 December 2023).

Issa, T. (2007) Development and evaluation of a methodology for developing websites. PhD thesis. Curtin University. Available at: https://espace.curtin.edu.au/handle/20.500.11937/2013/ (Accessed: October 2007).

Issa, T. and Isaias, P. (2022) 'Usability and human–computer interaction (hci)', In Sustainable Design: HCI, Usability and Environmental Concerns. London, pp. 23-40.

Iverson, J.M. and Goldin-Meadow, S. (2005) 'Gesture paves the way for language development', Psychological Science, 16(5), pp. 367-371.

James Jr, H.S. (2002) 'The trust paradox: a survey of economic inquiries into the nature of trust and trustworthiness', Journal of Economic Behavior & Organization, 47(3), pp. 291-307.

Jiménez, S., Juárez-Ramírez, R., Castillo, V.H., Licea, G., Ramírez-Noriega, A. and Inzunza, S. (2018) 'A feedback system to provide affective support to students', Computer Applications in Engineering Education, 26(3), pp. 473-483.

Jones, G.R. and George, J.M. (1998) 'The experience and evolution of trust: implications for cooperation and teamwork', Academy of Management Review, 23(3), pp. 531-546.

Jones, K. (1996) 'Trust as an affective attitude', Ethics, 107(1), pp. 4-25.

Jones, L. (2019) 'Compassion and trust: A qualitative investigation into fostering "compassionate trust" in society'.

Joshi, M.L. and Kanoongo, N. (2022) 'Depression detection using emotional artificial intelligence and machine learning: a closer review', Materials Today: Proceedings, 58, pp. 217-226.

Jung, C. (2016) Psychological types. London: Routledge.

Kajopoulos, J., Cheng, G., Kise, K., Müller, H. J. and Wykowska, A. (2020) 'Focusing on the face or getting distracted by social signals? The effect of distracting gestures on attentional focus in natural interaction', Psychological

Research, vol. 85, no. 2, pp. 491–502 [Online]. DOI: https://doi.org/10.1007/s00426-020-01383-4 (Accessed: 27 March 2023).

Kandana Arachchige, K.G., Simoes Loureiro, I., Blekic, W., Rossignol, M. and Lefebvre, L. (2021) ‘The role of iconic gestures in speech comprehension: an overview of various methodologies’, Frontiers in Psychology, 12, p. 634074.

Kapp, S., Barz, M., Mukhametov, S., Sonntag, D. and Kuhn, J. (2021) ‘ARETT: augmented reality eye tracking toolkit for head mounted displays’, Sensors, 21(6), p. 2234.

Kappagantula, S.R.K., Adamo-Villani, N., Wu, M.L. and Popescu, V. (2020) ‘Automatic deictic gestures for animated pedagogical agents’, IEEE Transactions on Learning Technologies, 13(1), pp. 1-13.

Karpus, J., Krüger, A., Verba, J.T., Bahrami, B. and Deroy, O. (2021) ‘Algorithm exploitation: Humans are keen to exploit benevolent AI’, iScience, [Online] 24(6), p.102679. DOI:https://doi.org/10.1016/j.isci.2021.102679 (Accessed: 18 December 2023).

Kasap, Z. and Magnenat-Thalmann, N. (2007) ‘Intelligent virtual humans with autonomy and personality: state-of-the-art’, Intelligent Decision Technologies, 1, pp. 3-15.

Katz, L. (2019) VR lets you fire a virtual employee to practice doing it for real. Available at: https://www.cnet.com/culture/vr-lets-you-fire-a-virtual-employee-to-practice-doing-it-for-real/ (Accessed: 15 March 2023).

Kelly, S., Kaye, S. and OviedoTrespalacios, O. (2023) ‘What factors contribute to the acceptance of artificial intelligence? A systematic review’, Telematics and Informatics, vol. 77.

Kendon, A. (1980) ‘Gesticulation and speech: two aspects of the process of utterance’, in Key, M.R. (ed.) The relationship of verbal and nonverbal communication. Berlin: De Gruyter Mouton, pp. 207-228.

Kendon, A. (1995) ‘Gestures as illocutionary and discourse structure markers in Southern Italian conversation’, Journal of Pragmatics, 23(3), pp. 247-279.

Kendon, A. (2004) Gesture: visible action as utterance. Cambridge: Cambridge University Press.

Kim, H., Ali, G., Pastor, A., Lee, M., Kim, G.J. and Hwang, J.I. (2021a) ‘Silhouettes from real objects enable realistic interactions with a virtual human in mobile augmented reality’, Applied Sciences, 11(6), pp. 2763.

Kim, H., Lee, M., Kim, G.J. and Hwang, J.I. (2021b) ‘The impacts of visual effects on user perception with a virtual human in augmented reality conflict situations’, IEEE Access, 9, pp. 35300-35312.

Kim, J. and Im, I. (2023) ‘Anthropomorphic response: Understanding interactions between humans and artificial intelligence agents’, Computers in Human Behavior’, 139(11), p. 107512.

Kim, J.C., Laine, T.H. and Åhlund, C. (2021) 'Multimodal interaction systems based on internet of things and augmented reality: a systematic literature review', Applied Sciences, 11(4), p. 1738.

Kim, K. (2018a) Environmental physical-virtual interaction to improve social presence with a virtual human in mixed reality. PhD thesis. University of Central Florida. Available at: https://stars.library.ucf.edu/etd/6158/ (Accessed: 1 March 2023).

Kim, K. (2018b) 'Improving social presence with a virtual human via multimodal physical -- virtual interactivity in AR', Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal, 21-26 April. New York, NY: Association for Computing Machinery, pp. 1-6.

Kim, K., Boelling, L., Haesler, S., Bailenson, J., Bruder, G. and Welch, G.F. (2018) 'Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR', Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Munich, 16-20 October. Piscataway, NJ: IEEE, pp. 105-114.

Kim, K., Bruder, G. and Welch, G. (2017) 'Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality', Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. Gothenburg, 8-10 November. New York, NY: Association for Computing Machinery, pp. 1-7.

Kim, K., Bruder, G., Maloney, D. and Welch, G. (2016) 'The influence of real human personality on social presence with a virtual human in augmented reality', in Little Rock, A. (ed.) Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments. Little Rock, 7-9 October. Goslar: Eurographics Association, pp. 115-122.

Kim, K., de Melo, C.M., Norouzi, N., Bruder, G. and Welch, G.F. (2020) 'Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making', Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Atlanta, GA, 22-26 March. Piscataway, NJ: IEEE, pp. 529-538.

Kim, K., Schubert, R., Hochreiter, J., Bruder, G. and Welch, G. (2019) 'Blowing in the wind: increasing social presence with a virtual human via environmental airflow interaction in mixed reality', Computers & Graphics, 83, pp. 23-32.

Kim, K.J., Park, B.S., Kim, J.K., Kim, D.W. and Seo, Y.H. (2020) 'Holographic augmented reality based on three-dimensional volumetric imaging for a photorealistic scene', Optics Express, 28(24), pp. 35972-35985.

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K. and Green, A. (2009) 'Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures', Neuropsychologia, 47(1), pp. 169-179.

Kita, S., Alibali, M.W. and Chu, M. (2017) 'How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis', Psychological Review, 124(3), pp. 245-266.

Kita, S., van Gijn, I. and van der Hulst, H. (1998) 'Movement phases in signs and co-speech gestures, and their transcription by human coders', in Wachsmuth, I. and Fröhlich, M. (eds.), Proceedings of the International Workshop on Gesture and Sign Language in Human-Computer Interaction. Bielefeld, 17-19 September 1997. Berlin: Springer, pp. 23-35.

Knote, R., Janson, A., Söllner, M. and Leimeister, J. M. (2019) Classifying Smart Personal Assistants: An Empirical Cluster Analysis, scholarspace.manoa.hawaii.edu, Proceedings of the 52rd Annual Hawaii International Conference on System Sciences [Online]. Available at https://hdl.handle.net/10125/59642 (Accessed: 27 March 2023).

Kocaballi, A.B., Laranjo, L. and Coiera, E. (2019) 'Understanding and measuring user experience in conversational interfaces', Interacting with Computers, 31(2), pp. 192-207.

Kopp, S. and Wachsmuth, I. (2004) 'Synthesizing multimodal utterances for conversational agents', Computer Animation and Virtual Worlds, 15(1), pp. 39-52.

Kopp, S., Tepper, P. and Cassell, J. (2004) 'Towards integrated microplanning of language and iconic gesture for multimodal output', Proceedings of the 6th International Conference on Multimodal Interfaces. State College, PA, 13-15 October. New York, NY: Association for Computing Machinery, pp. 97-104.

Kory-Westlund, J.M. (2023) 'Implications of children's social, emotional, and relational interactions with robots for human–robot empathy', in Mezzenzana, F. and Peluso, D. (eds.) Conversations on empathy: interdisciplinary perspectives on empathy, imagination and othering. New York, NY: Routledge, pp. 256-279.

KRAFTON (2022) KRAFTON teases the company's virtual human, ANA. Available at: https://krafton.com/en/news/press/krafton-teases-the-companys-virtual-human-ana/ (Accessed: 13 March 2023).

Krämer, N.C., Rosenthal-von der Pütten, A.M. and Hoffmann, L. (2015) 'Social effects of virtual and robot companions', in Sundar, S.S. (ed.) The handbook of the psychology of communication technology. Hoboken, NJ: John Wiley & Sons, pp. 137-159.

Kramer, R.M. (1999) 'Trust and distrust in organizations: emerging perspectives, enduring questions', Annual Review of Psychology, 50(1), pp. 569-598.

Krauss, R.M., Chen, Y. and Gottesman, R.F. (2000) 'Lexical gestures and lexical access: a process model', in McNeill, D. (ed.) Language and gesture. Cambridge: Cambridge University Press, pp. 261-283.

Kulms, P. and Kopp, S. (2018) 'A social cognition perspective on human–computer trust: the effect of perceived warmth and competence on trust in decision-making with computers', Frontiers in Digital Humanities, 5, p. 14.

Laamarti, F., Eid, M. and El Saddik, Abdulmotaleb (2014) ‘An Overview of Serious Games’, International Journal of Computer Games Technology, Hindawi Publishing Corporation, vol. 2014, p. 358152.

Lai, M.L., Tsai, M.J., Yang, F.Y., Hsu, C.Y., Liu, T.C., Lee, S.W.Y., Lee, M.H., Chiou, G.L., Liang, J.C. and Tsai, C.-C. (2013) ‘A review of using eye-tracking technology in exploring learning from 2000 to 2012’, Educational Research Review, 10, pp. 90-115.

Lakoff, G. and Johnson, M. (1980) Metaphors we live by. Chicago: University of Chicago Press.

Lakoff, G. and Johnson, M. (2008) Metaphors we live by. Chicago: University of Chicago Press.

Lampen, E., Lehwald, J. and Pfeiffer, T. (2020) ‘Virtual humans in AR: evaluation of presentation concepts in an industrial assistance use case’, Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology. Virtual Event, 1-4 November. New York, NY: Association for Computing Machinery, pp. 1-5.

Landrum, A.R., Mills, C.M. and Johnston, A.M. (2013) ‘When do children trust the expert? Benevolence information influences children’s trust more than expertise’, Developmental Science, 16(4), pp. 622-638.

Lang, P.J. (1980) ‘Behavioural treatment and bio-behavioural assessment: Computer Applications’, in Sidowski, J.B., Johnson, J.H. and Williams, T.A. (eds.) Technology in mental health care delivery systems. Norwood, NJ: Ablex, pp. 119-137.

Lang, Y., Liang, W. and Yu, L.F. (2019) ‘Virtual agent positioning driven by scene semantics in mixed reality’, Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Osaka, 23-27 March. Piscataway, NJ: IEEE, pp. 767-775.

Larson, K. (2020) ‘Serious games and gamification in the corporate training environment: A literature review’, TechTrends 64, 319–328.

Leary, M.R. and Allen, A.B. (2011) ‘Personality and persona: personality processes in self-presentation’, Journal of Personality, 79(6), pp. 1191-1218.

Lee, J.D. and See, K.A. (2004) ‘Trust in automation: designing for appropriate reliance’, Human Factors, 46(1), pp. 50-80.

Lee, J.J., Knox, B., Baumann, J., Breazeal, C. and DeSteno, D. (2013) ‘Computationally modeling interpersonal trust’, Frontiers in Psychology, 4, p. 893.

Lee, M., Bruder, G., Höllerer, T. and Welch, G. (2018) ‘Effects of unaugmented periphery and vibrotactile feedback on proxemics with virtual humans in AR’, IEEE Transactions on Visualization and Computer Graphics, 24(4), pp. 1525-1534.

Lee, M., Norouzi, N., Bruder, G., Wisniewski, P.J. and Welch, G.F. (2021) ‘Mixed reality tabletop gameplay: social interaction with a virtual human capable of physical influence’, IEEE Transactions on Visualization and Computer Graphics, 27(8), pp. 3534-3545.

Lepetit, V. and Fua, P. (2006) ‘Keypoint recognition using randomized trees’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(9), pp. 1465-1479.

Lester, J.C., Voerman, J.L., Towns, S.G. and Callaway, C.B. (1999) ‘Deictic believability: coordinated gesture, locomotion, and speech in lifelike pedagogical agents’, Applied Artificial Intelligence, 13(4-5), pp. 383-414.

Lhommet, M. and Marsella, S. (2016) ‘From embodied metaphors to metaphoric gestures’, Proceedings of the 38th Annual Conference of the Cognitive Science Society. Philadelphia, PA, 10-13 August. Austin, TX: Cognitive Science Society, pp. 788-793.

Li, C., Androulakaki, T., Gao, A.Y., Yang, F., Saikia, H., Peters, C. and Skantze, G. (2018) ‘Effects of posture and embodiment on social distance in human-agent interaction in mixed reality’, Proceedings of the 18th International Conference on Intelligent Virtual Agents. Sydney, Australia, 5-8 November. New York, NY: Association for Computing Machinery, pp. 191-196.

Li, J., Galley, M., Brockett, C., Spithourakis, G.P., Gao, J. and Dolan, B. (2016) ‘A persona-based neural conversation model’, arXiv. Available at: https://arxiv.org/abs/1603.06155 (Accessed: 10 March 2023).

Li, J., Zheng, Z., Wei, X. and Wang, G. (2021) ‘FaceMe: an augmented reality social agent game for facilitating children’s learning about emotional expressions’, Adjunct Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology. Virtual Event, 10-14 October. New York, NY: Association for Computing Machinery, pp. 17-19.

Li, J., Zhou, M.X., Yang, H. and Mark, G. (2017) ‘Confiding in and listening to virtual agents: the effect of personality’, Proceedings of the 22nd International Conference on Intelligent User Interfaces. Limassol, 13-16 March. New York, NY: Association for Computing Machinery, pp. 275-286.

Lin, Y.L. (2017) ‘Co-occurrence of speech and gestures: a multimodal corpus linguistic approach to intercultural interaction’, Journal of Pragmatics, 117, pp. 155-167.

Lindenberg, R., Uhlig, M., Scherfeld, D., Schlaug, G. and Seitz, R.J. (2012) ‘Communication with emblematic gestures: shared and distinct neural correlates of expression and reception’, Human Brain Mapping, 33(4), pp. 812-823.

LipSyncPro (2023). Available at: https://lipsync.rogodigital.com/SALSA (Accessed: 15 March 2023).

Lithner, J. (2007) ‘A research framework for creative and imitative reasoning’, Educational Studies in Mathematics, vol. 67, no. 3, pp. 255–276.

Liu, Z., He, S. and Xiong, W. (2008) ‘A fuzzy logic based emotion model for virtual human’, Proceedings of the 2008 International Conference on Cyberworlds. Hanzhou, 22-24 September. Piscataway, NJ: IEEE, pp. 284-288.

Llorach, G., Evans, A., Blat, J., Grimm, G. and Hohmann, V. (2016) ‘Web-based live speech-driven lip-sync’, Proceedings of the 2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES). Barcelona, 7-9 September. Piscataway, NJ: IEEE, pp. 1-4.

Lyons, J. (1977) Semantics: Vol. 1. Cambridge: Cambridge University Press.

Matsumoto, D. and Hwang, H.S. (2013) 'Emblematic gestures (emblems)', in Keith, K.D. (ed.) The encyclopedia of cross-cultural psychology. New York, NY: John Wiley & Sons, pp. 464-466.

Matsumoto, Y. and Dobs, A.M. (2017) 'Pedagogical gestures as interactional resources for teaching and learning tense and aspect in the ESL grammar classroom', Language Learning, 67(1), pp. 7-42.

Mayer, R.C., Davis, J.H. and Schoorman, F.D. (1995) 'An integrative model of organizational trust', Academy of Management Review, 20(3), pp. 709-734.

McAllister, D.J. (1995) 'Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations', Academy of Management Journal, 38(1), pp. 24-59.

McCafferty, S.G. (2002) 'Gesture and creating zones of proximal development for second language learning', The Modern Language Journal, 86(2), pp. 192-203.

McDonnell, R., Breidt, M. and Bülthoff, H.H. (2012) 'Render me real? Investigating the effect of render style on the perception of animated virtual humans', ACM Transactions on Graphics, 31(4), pp. 1-11.

McEvily, B. and Tortoriello, M. (2011) 'Measuring trust in organisational research: review and recommendations', Journal of Trust Research, 1(1), pp. 23-63.

Mcknight, D.H., Carter, M., Thatcher, J.B. and Clay, P.F. (2011) 'Trust in a specific technology: an investigation of its components and measures', ACM Transactions on Management Information Systems, 2(2), pp. 1-25.

McNeill, D. (1995) 'Hand and mind', in Sebeok, T.A. and Umiker-Sebeok, J. (eds.) Advances in visual semiotics. Boston: De Gruyter Mouton, pp. 351-374.

McNeill, D. (2005) Gesture and thought. Chicago: University of Chicago Press.

Mellor, D. and Moore, K.A. (2014) 'The use of Likert scales with children', Journal of Pediatric Psychology, 39(3), pp. 369-379.

Mierowsky, R., Marcus, N. and Ayres, P. (2020) 'Using mimicking gestures to improve observational learning from instructional videos', Educational Psychology, 40(5), pp. 550-569.

Miller, M.R., Jun, H., Herrera, F., Yu Villa, J., Welch, G. and Bailenson, J.N. (2019) 'Social interaction in augmented reality', PLoS One, 14(5), p. e0216290.

Mixamo (2023). Available at: https://www.mixamo.com/#/ (Accessed: 15 March 2023).

Miyake, S. and Ito, A. (2012) 'A spoken dialogue system using virtual conversational agent with augmented reality', Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference. Hollywood, CA, 3-6 December. Piscataway, NJ: IEEE, pp. 1-4.

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A. and PRISMA-P Group (2015) 'Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement', Systematic Reviews, 4(1), pp. 1-9.

Mol, L., and Kita, S. (2012). 'Gesture structure affects syntactic structure in speech', in Proceedings of the 34th Annual Conference of the Cognitive Science Society, eds N. Miyake, D. Peebles, and R. P. Cooper (Sapporo: Cognitive Science Society), pp. 761–766.

Moorman, C., Zaltman, G. and Deshpande, R. (1992) 'Relationships between providers and users of market research: the dynamics of trust within and between organizations', Journal of Marketing Research, 29(3), pp. 314-328.

Mori, M. (1970) 'The uncanny valley', Energy, 7(4), pp. 33-35.

Mostajeran, F., Burke, N., Ertugrul, N., Hildebrandt, K., Matov, J., Tapie, N., Zittel, W.G., Reisewitz, P. and Steinicke, F. (2022) 'Anthropomorphism of virtual agents and human cognitive performance in augmented reality', Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, 12-16 March. Piscataway, NJ: IEEE, pp. 329-332.

Mostajeran, F., Reisewitz, P. and Steinicke, F. (2022) 'Social facilitation and inhibition in augmented reality: performing motor and cognitive tasks in the presence of a virtual agent', Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). Christchurch, 12-16 March. Piscataway, NJ: IEEE, pp. 323-328.

Naffziger, D. W., Hornsby, J. S. and Kuratko, D. F. (1994) 'A Proposed Research Model of Entrepreneurial Motivation', Entrepreneurship Theory and Practice, vol. 18, no. 3, pp. 29–42.

Nass, C., Steuer, J. and Tauber, E.R. (1994) 'Computers are social actors', Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Boston, 24-28 April. New York, NY: Association for Computing Machinery, pp. 72-78.

Nasution, A., Rizki, Y., Nasution, S. and Muhammad, R. (2020) 'An augmented reality machine translation agent', Proceedings of 2nd International Conference on Science, Engineering and Technology (ICoSET 2019). Pekanbaru, 5-7 September 2019. Setúbal: SciTePress, pp. 163-168.

Navas Medrano, S., Pfeiffer, M. and Kray, C. (2020) 'Remote deictic communication: simulating deictic pointing gestures across distances using electro muscle stimulation', International Journal of Human–Computer Interaction, 36(19), pp. 1867-1882.

NEON (2023). Available at: https://neonlife.ai/ (Accessed: 15 March 2023).

Nikolaev, A.R., Pannasch, S., Ito, J. and Belopolsky, A.V. (2014) 'Eye movement-related brain activity during perceptual and cognitive processing', Frontiers in Systems Neuroscience, 8, p. 62.

Nirme, J., Haake, M., Gulz, A. and Gullberg, M. (2020) 'Motion capture-based animated characters for the study of speech–gesture integration', Behavior Research Methods, 52(3), pp. 1339-1354.

Nizam, S.S.M., Abidin, R.Z., Hashim, N.C., Lam, M.C., Arshad, H. and Majid, N.A.A. (2018) ‘A review of multimodal interaction technique in augmented reality environment’, International Journal on Advanced Science, Engineering and Information Technology, 8(4-2), pp. 1460-1469.

Norouzi, N., Kim, K., Bruder, G., Bailenson, J.N., Wisniewski, P. and Welch, G.F. (2022) ‘The advantages of virtual dogs over virtual people: using augmented reality to provide social support in stressful situations’, International Journal of Human-Computer Studies, 165, p. 102838.

Norouzi, N., Kim, K., Bruder, G., Erickson, A., Choudhary, Z., Li, Y. and Welch, G. (2020) ‘A systematic literature review of embodied augmented reality agents in head-mounted display environments’, in Argelaguet, F., Sugimoto, M. and McMahan, R. (eds.), Proceedings of the 30th International Conference on Artificial Reality and Telexistence and the 25th Eurographics Symposium on Virtual Environments. Virtual Event, 2-4 December. Goslar: Eurographics Association, pp. 101-111.

Norouzi, N., Kim, K., Lee, M., Schubert, R., Erickson, A., Bailenson, J., Bruder, G. and Welch, G. (2019) ‘Walking your virtual dog: analysis of awareness and proxemics with simulated support animals in augmented reality’, Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Beijing, 14-18 October. Piscataway, NJ: IEEE, pp. 157-168.

Novack, M.A. and Goldin-Meadow, S. (2017) ‘Gesture as representational action: a paper about function’, Psychonomic Bulletin & Review, 24(3), pp. 652-665.

Novack, M.A., Congdon, E.L., Hemani-Lopez, N. and Goldin-Meadow, S. (2014) ‘From action to abstraction: using the hands to learn math’, Psychological Science, 25(4), pp. 903-910.

O’Hare, G.M.P., Stafford, J.W. and Campbell, A.G. (2005) ‘NeXuS: delivering perceptions to situated embodied agents’, in Panayiotopoulos, T., Gratch, J., Aylett, R., Ballin, D., Olivier, P. and Rist, T. (eds.), Proceedings of the 5th International Conference on Intelligent Virtual Agents. Kos, Greece, 12-14 September. Berlin: Springer, pp. 500-500.

Obaid, M., Damian, I., Kistler, F., Endrass, B., Wagner, J. and André, E. (2012) ‘Cultural behaviors of virtual agents in an augmented reality environment’, in Nakano, Y., Neff, M., Paiva, A. and Walker, M. (eds.), Proceedings of the 12th International Conference on Intelligent Virtual Agents (IVA 2012). Santa Cruz, CA, 12-14 September. Berlin: Springer, pp. 412-418.

Oh, S. and Byun, Y.C. (2012) ‘The design and implementation of augmented reality learning systems’, Proceedings of the 2012 IEEE/ACIS 11th International Conference on Computer and Information Science. Shanghai, 30 May-1 June. Piscataway, NJ: IEEE, pp. 651-654.

OpenAI (2022) ChatGPT: Optimizing Language Models for Dialogue [Online]. Available at https://openai.com/blog/chatgpt.

Oufqir, Z., El Abderrahmani, A. and Satori, K. (2020) ‘ARKit and ARCore in serve to augmented reality’, Proceedings of the 2020 International Conference on

Intelligent Systems and Computer Vision (ISCV). Fez, 9-11 June. Piscataway, NJ: IEEE, pp. 1-7.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. and Moher, D. (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', BMJ, 372, p. n71.

Papadopoulos, T., Evangelidis, K., Kaskalis, T.H., Evangelidis, G. and Sylaiou, S. (2021) 'Interactions in augmented and mixed reality: an overview', Applied Sciences, 11(18), p. 8752.

Parenti, L., Lukomski, A.W., De Tommaso, D., Belkaid, M. and Wykowska, A. (2022) 'Human-likeness of feedback gestures affects decision processes and subjective trust'. To be published in International Journal of Social Robotics. Available at: https://doi.org/10.1007/s12369-022-00927-5 (Accessed: 10 March 2023).

Park, K.B., Choi, S.H., Kim, M. and Lee, J.Y. (2020) 'Deep learning-based mobile augmented reality for task assistance using 3D spatial mapping and snapshot-based RGB-D data', Computers & Industrial Engineering, 146, p. 106585.

Park, K.H. and Jeong, Y.S. (2019) 'Indoor dialog agent in mixed reality (video)', Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services. Seoul, 17-21 June. New York, NY: Association for Computing Machinery, pp. 708-709.

Pearce, W.B. (1974) 'Trust in interpersonal communication', Speech Monographs, 41(3), pp. 236-244.

Peng, H., Shang, X., Guo, C., Xiong, G., Nyberg, T.R., Fan, D. and Wang, Y. (2016) 'A survey on big data for human body shape', Proceedings of the 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI). Beijing, 10-12 July. Piscataway, NJ: IEEE, pp. 145-150.

Peters, C., Li, C., Yang, F., Avramova, V. and Skantze, G. (2018) 'Investigating social distances between humans, virtual humans and virtual robots in mixed reality', Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. Stockholm, 10-15 July. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, pp. 2247-2249.

Pfeuffer, K., Abdrabou, Y., Esteves, A., Rivu, R., Abdelrahman, Y., Meitner, S., Saadi, A. and Alt, F. (2021) 'ARtention: A design space for gaze-adaptive user interfaces in augmented reality', Computers & Graphics, vol. 95 [Online]. DOI: https://doi.org/10.1016/j.cag.2021.01.001 (Accessed: 14 January 2021).

Philip, P., Dupuy, L., Auriacombe, M., Serre, F., de Sevin, E., Sauteraud, A. and Micoulaud-Franchi, J.A. (2020) 'Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients', NPJ Digital Medicine, 3(1), pp. 1-7.

Pimentel, D. and Vinkers, C. (2021) ‘Copresence with virtual humans in mixed reality: the impact of contextual responsiveness on social perceptions’, Frontiers in Robotics and AI, 8, p. 634520.

Pimentel, J.L. (2010) ‘A note on the usage of Likert Scaling for research data analysis’, USM R&D Journal, 18(2), pp. 109-112.

Piumsomboon, T., Lee, G.A., Hart, J.D., Ens, B., Lindeman, R.W., Thomas, B.H. and Billinghurst, M. (2018) ‘Mini-me: an adaptive avatar for mixed reality remote collaboration’, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. Montreal, QC, 21-26 April. New York, NY: Association for Computing Machinery, pp. 1-13.

Piwek, L., McKay, L.S. and Pollick, F.E. (2014) ‘Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion’, Cognition, 130(3), pp. 271-277.

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P. and Vilà-Giménez, I. (2018) ‘Deconstructing beat gestures: a labelling proposal’, Proceedings of the 9th International Conference on Speech Prosody. Poznań, 13-16 June. Baixas: International Speech Communications Association, pp. 201-205.

Pujades, S., Mohler, B., Thaler, A., Tesch, J., Mahmood, N., Hesse, N., Bülthoff, H.H. and Black, M.J. (2019) ‘The virtual caliper: rapid creation of metrically accurate avatars from 3D measurements’, IEEE Transactions on Visualization and Computer Graphics, 25(5), pp. 1887-1897.

Qiu, L. and Benbasat, I. (2009) ‘Evaluating anthropomorphic product recommendation agents: a social relationship perspective to designing information systems’, Journal of Management Information Systems, 25(4), pp. 145-182.

Randhavane, T., Bera, A., Kapsaskis, K., Gray, K. and Manocha, D. (2019) ‘FVA: modeling perceived friendliness of virtual agents using movement characteristics’, IEEE Transactions on Visualization and Computer Graphics, 25(11), pp. 3135-3145.

Rebol, M., Güti, C. and Pietroszek, K. (2021) ‘Passing a non-verbal turing test: evaluating gesture animations generated from speech’, Proceedings of the 2021 IEEE Virtual Reality and 3D User Interfaces (VR). Lisboa, 27 March-1 April. Piscataway, NJ: IEEE, pp. 573-581.

Reeves, B. and Nass, C. (1996) The media equation: how people treat computers, television, and new media like real people and places. New York, NY: Cambridge University Press.

Reiersen, J. (2017) ‘Trust as belief or behavior?’, Review of Social Economy, 75(4), pp. 434-453.

Reinhardt, J., Hillen, L. and Wolf, K. (2020) ‘Embedding conversational agents into AR: invisible or with a realistic human body?’, Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction. Sydney, 9-12 February. New York, NY: Association for Computing Machinery, pp. 299-310.

Rempel, J.K., Holmes, J.G. and Zanna, M.P. (1985) ‘Trust in close relationships’, Journal of Personality and Social Psychology, 49(1), pp. 95-112.

Richards, M.N. and Calvert, S.L. (2016) ‘Parent versus child report of young children’s parasocial relationships in the United States’, Journal of Children and Media, 10(4), pp. 462-480.

Rickel, J. (2001) ‘Intelligent virtual agents for education and training: opportunities and challenges’, in de Antonio, A., Aylett, R. and Ballin, D. (eds.), Proceedings of the 3rd International Conference on Intelligent Virtual Agents. Madrid, 10-11 September. Berlin: Springer, pp. 15-22.

Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D. and Swartout, W. (2002) ‘Toward a new generation of virtual humans for interactive experiences’, IEEE Intelligent Systems, 17(4), pp. 32-38.

Rizzo, A., Shilling, R., Forbell, E., Scherer, S., Gratch, J. and Morency, L.-P. (2016) ‘Autonomous virtual human agents for healthcare information support and clinical interviewing’, in Luxton, D.D. (ed.) Artificial intelligence in behavioral and mental health care. San Diego, CA: Academic Press, pp. 53-79.

Roberts, J.S., Laughlin, J.E. and Wedell, D.H. (1999) ‘Validity issues in the Likert and Thurstone approaches to attitude measurement’, Educational and Psychological Measurement, 59(2), pp. 211-233.

Rokoko (2023). Available at: https://www.rokoko.com/ (Accessed: 15 March 2023).

Rolin, K.H. (2021) ‘Objectivity, trust and social responsibility’, Synthese, 199(1), pp. 513-533.

Rotter, J.B. (1967) ‘A new scale for the measurement of interpersonal trust’, Journal of Personality, 35(4), pp. 651-665.

Rousseau, D.M., Sitkin, S.B., Burt, R.S. and Camerer, C. (1998) ‘Not so different after all: a cross-discipline view of trust’, Academy of Management Review, 23(3), pp. 393-404.

Russian3DScanner (2023). Available at: https://www.russian3dscanner.com/ (Accessed: 15 March 2023).

Salanitri, D., Hare, C., Borsci, S., Lawson, G., Sharples, S. and Waterfield, B. (2015) ‘Relationship Between Trust and Usability in Virtual Environments: An Ongoing Study’, Kurosu, M. (ed), HumanComputer Interaction: Design and Evaluation, Cham, Springer International Publishing, pp. 49–59.

Sato, W., Hyniewska, S., Minemoto, K. and Yoshikawa, S. (2019) ‘Facial expressions of basic emotions in Japanese laypeople’, Frontiers in Psychology, 10, p. 259.

Scharowski, N., Sebastian, Svab, M., Opwis, K. and Brühlmann, F. (2023) ‘Exploring the effects of human-centered AI explanations on trust and reliance’, Frontiers in Computer Science, [online] 5. doi:https://doi.org/10.3389/fcomp.2023.1151150 (Accessed: 18 December 2023).

Schmidt, S., Ariza, O. and Steinicke, F. (2020) ‘Intelligent blended agents: reality–virtuality interaction with artificially intelligent embodied virtual humans’, Multimodal Technologies and Interaction, 4(4), p. 85.

Schmidt, S., Nunez, O.J.A. and Steinicke, F. (2019) ‘Blended agents: manipulation of physical objects within mixed reality environments and beyond’, Symposium on Spatial User Interaction. New Orleans, LA, 19-20 October. New York, NY: Association for Computing Machinery, pp. 1-10.

Sebo, S.S., Krishnamurthi, P. and Scassellati, B. (2019) ‘“I don’t believe you”: investigating the effects of robot trust violation and repair’, Proceedings of the 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). Daegu, 11-14 March. Piscataway, NJ: IEEE, pp. 57-65.

Seitz, L., Woronkow, J., Bekmeier-Feuerhahn, S. and Gohil, K. (2021) ‘The advance of diagnosis chatbots: should we first avoid distrust before we focus on trust?’, Proceedings of the 29th European Conference on Information Systems. Marrakech, 14-16 June. Atlanta: Association for Information Systems, pp. 1-11.

Sekine, K. and Kita, S. (2013) ‘Use of spatial information from cohesive gesture to comprehend subsequent sentences’, Presented at the Tilburg Gesture Research Meeting 2013. Tilburg, 19-21 June.

Sheppard, B.H. and Sherman, D.M. (1998) ‘The grammars of trust: a model and general implications’, Academy of Management Review, 23(3), pp. 422-437.

Sheth, A., Yip, H.Y., Iyengar, A. and Tepper, P. (2019) ‘Cognitive services and intelligent chatbots: current perspectives and special issue introduction’, IEEE Internet Computing, 23(2), pp. 6-12.

Shimobaba, T., Blinder, D., Birnbaum, T., Hoshi, I., Shiomi, H., Schelkens, P. and Ito, T. (2022) ‘Deep-learning computational holography: a review’, Frontiers in Photonics, 3, p. 854391.

Siau, K. and Wang, W. (2018) ‘Building trust in artificial intelligence, machine learning, and robotics’, Cutter Business Technology Journal, 31(2), pp. 47-53.

Skantze, G. (2021) ‘Turn-taking in conversational systems and human-robot interaction: a review’, Computer Speech & Language, 67, p. 101178.

So, W.C., Ching, T.H.-W., Lim, P.E., Cheng, X. and Ip, K.Y. (2014) ‘Producing gestures facilitates route learning’, PLoS One, 9(11), p. e112543.

Söllner, M., Hoffmann, A., Hoffmann, H., Wacker, A. and Leimeister, J.M. (2012) ‘Understanding the formation of trust in IT artifacts’, Proceedings of the 33rd International Conference on Information Systems (ICIS 2012). Orlando, FL, 16-19 December. Atlanta: Association for Information Systems, pp. 1038-1055.

Solomon, R. and Flores, F. (2003) ‘Building trust: In business, politics, relationships, and life’, Building Trust: In Business, Politics, Relationships, and Life, pp. 1–178.

Sottilare, R.A. (2018) ‘A review of intelligent tutoring system authoring tools and methods’, Assessment of intelligent tutoring systems technologies and opportunities. Paris: STO/NATO, pp. 1-8.

Sousa, S., José Cravino, Martins, P.A.F. and Lamas, D. (2023) ‘Human-centered trust framework: An HCI perspective’. arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.2305.03306 (Accessed: 18 December 2023)

Sousa, S., Lamas, D. and Dias, P. (2014) ‘A model for human-computer trust’, in Zaphiris, P. and Ioannou, A. (eds.), Proceedings of the First International Conference on Learning and Collaboration Technologies (LCT 2014). Heraklion, 22-27 June. Cham: Springer, pp. 128-137.

Stefanini, S., Bello, A., Caselli, M.C., Iverson, J.M. and Volterra, V. (2009) ‘Co-speech gestures in a naming task: developmental data’, Language and Cognitive Processes, 24(2), pp. 168-189.

Techasarntikul, N., Ratsamee, P., Orlosky, J., Mashita, T., Uranishi, Y., Kiyokawa, K. and Takemura, H. (2020) ‘The effect of the presence of an embodied agent in an AR guiding system’, Transactions of the Virtual Reality Society of Japan, 25(1), pp. 68-77.

Thalmann, D., Musse, S.R. and Kallmann, M. (2000) ‘Virtual humans’ behaviour: individuals, groups, and crowds’, in Vince, J.A. and Earnshaw, R. (eds.) Digital media: the future. London: Springer, pp. 239-256.

Thomas, H. (1750) Human nature. London: J.M. Dent and Co.

Thomas, S., Ferstl, Y., McDonnell, R. and Ennis, C. (2022) ‘Investigating how speech and animation realism influence the perceived personality of virtual characters and agents’, Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Christchurch, 12-16 March. Piscataway, NJ: IEEE, pp. 11-20.

Thompson, J.C., Trafton, J.G. and McKnight, P. (2011) ‘The perception of humanness from the movements of synthetic agents’, Perception, 40(6), pp. 695-704.

Thuan, Nguyen Hoang, Drechsler, A. and Antunes, P. (2019) ‘Construction of Design Science Research Questions’, Communications of the Association for Information Systems, vol. 44, pp. 332–363.

Traum, D. (2009) ‘Models of culture for virtual human conversation’, in Stephanidis, C. (ed.) Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. San Diego, CA, 19-24 July. Berlin: Springer, pp. 434-440.

Trinh, H., Ring, L. and Bickmore, T. (2015) ‘DynamicDuo: co-presenting with virtual agents’, Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. Seoul, 18-23 April. New York, NY: Association for Computing Machinery, pp. 1739-1748.

Trochim, W.M.K. (2021) Likert scaling. Available at: https://conjointly.com/kb/likert-scaling/ (Accessed: 15 March 2023).

Tschannen-Moran, M. and Hoy, W.K. (2000) ‘A multidisciplinary analysis of the nature, meaning, and measurement of trust’, Review of Educational Research, 70(4), pp. 547-593.

Tsovaltzi, D., McLaren, B. M., Melis, E. and Meyer, A. K. (2012) ‘Erroneous examples: effects on learning fractions in a web-based setting’, International Journal of Technology Enhanced Learning, vol. 4, no. 3/4, p. 191 [Online]. DOI: https://doi.org/10.1504/ijtel.2012.051583 (Accessed: 29 May 2021).

Turk, M. (2014) ‘Multimodal interaction: a review’, Pattern Recognition Letters, 36, pp. 189-195.

Twyford, J. and Craig, S. (2013) ‘Virtual humans and gesturing during multimedia learning: an investigation of predictions from the temporal contiguity effect’, in Bastiaens, T. and Marks, G. (eds.), Proceedings of E-Learn 2013: World Conference on E-Learningin Corporate, Government, Healthcare, and Higher Education. Las Vegas, NV, 21 October. San Diego, CA: Association for the Advancement of Computing in Education, pp. 2145-2149.

Unity (2023). Available at: https://unity.com/ (Accessed: 15 March 2023).

Unreal Engine (2023). Available at: https://www.unrealengine.com/en-US/metahuman?utm_source=GoogleSearch&utm_medium=Performance&utm_campaign=%7Bcampaignname%7D&utm_id=19729682794&sub_campaign=&utm_content=&utm_term=unreal%20engine%20meta%20human/ (Accessed 27 March 2023).

Urbano, J., Rocha, A.P. and Oliveira, E. (2013) ‘The impact of benevolence in computational trust’, in Chesñevar, C.I., Onaindia, E., Ossowski, S. and Vouros, G. (eds.), Proceedings of the Second International Conference on Agreement Technologies. Beijing, 1-2 August. Berlin: Springer, pp. 210-224.

van Pinxteren, M.M.E., Wetzels, R.W.H., Rüger, J., Pluymaekers, M. and Wetzels, M. (2019) ‘Trust in humanoid robots: implications for services marketing’, Journal of Services Marketing, 33(4), pp. 507-518.

van Straten, C.L., Peter, J. and Kühne, R. (2020) ‘Child–robot relationship formation: a narrative review of empirical research’, International Journal of Social Robotics, 12(2), pp. 325-344.

Van Straten, C.L., Peter, J., Kühne, R., de Jong, C. and Barco, A. (2018) ‘Technological and interpersonal trust in child-robot interaction: an exploratory study’, Proceedings of the 6th International Conference on Human-Agent Interaction. Southampton, 15-18 December. New York, NY: Association for Computing Machinery, pp. 253-259.

Vanneste, B. S., Puranam, P. and Kretschmer, T. (2013) ‘Trust over time in exchange relationships: Meta-analysis and theory’, Strategic Management Journal, vol. 35, no. 12, pp. 1891–1902.

Vaughan, N., Gabrys, B. and Dubey, V.N. (2016) ‘An overview of self-adaptive technologies within virtual reality training’, Computer Science Review, 22, pp. 65-87.

Verma, V., Sharma, A., Jain, K.K. and Adlakha, S. (2021) ‘Digital assistant with augmented reality’, in Tuba, M., Akashe, S. and Joshi, A. (eds.), Proceedings of the 5th International Conference on ICT for Sustainable Development. Goa, 23-24 July. Singapore: Springer, pp. 395-404.

Vicon (2023). Available at: https://docs.vicon.com/ (Accessed: 15 March 2023).

Vlachopoulos, D. and Makri, A. (2017) ‘The effect of games and simulations on higher education: a systematic literature review’, International Journal of Educational Technology in Higher Education, vol. 14, no. 1.

Voulodimos, A., Doulamis, N., Doulamis, A. and Protopapadakis, E. (2018) ‘Deep learning for computer vision: a brief review’, Computational Intelligence and Neuroscience, 2018, p. 7068349.

Wadsley, T. and Ryan, M. (2021) ‘A belief-desire-intention model for narrative generation’, Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, 9(4), pp. 105-108.

Wagner, P., Malisz, Z. and Kopp, S. (2014) ‘Gesture and speech in interaction: an overview’, Speech Communication, 57, pp. 209-232.

Wakefield, E., Novack, M.A., Congdon, E.L., Franconeri, S. and Goldin-Meadow, S. (2018) ‘Gesture helps learners learn, but not merely by guiding their visual attention’, Developmental Science, 21(6), p. e12664.

Wand, Y. and Weber, R. (2002) ‘Research Commentary: Information Systems and Conceptual Modeling—A Research Agenda’, Information Systems Research, 13(4), pp.363–376.

Wang, I., Smith, J. and Ruiz, J. (2019) ‘Exploring virtual agents for augmented reality’, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Glasgow, 4-9 May. New York, NY: Association for Computing Machinery, pp. 1-12.

Wang, T., Qian, X., He, F., Hu, X., Cao, Y. and Ramani, K. (2021) ‘GesturAR: an authoring system for creating freehand interactive augmented reality applications’, Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology. Virtual Event, 10-14 October. New York, NY: Association for Computing Machinery, pp. 552–567.

Wang, T.J. (2022) ‘A study on utilizing 3D motion-capture analysis to assist in Chinese opera teaching’, Research in Dance Education, 23(2), pp. 194-222.

Wang, X. (2014) Personas in the user interface design. Available at: https://www.immagic.com/eLibrary/ARCHIVES/GENERAL/UCALG_CA/U071112W.pdf (Accessed: 13 March 2023).

Watts, R. (2020) An interactive, holographic 3D model using IBM Wats.

Wekit ECS (2022) MirageXR. Available at: https://github.com/WEKIT-ECS/MIRAGE-XR/ (Accessed: 15 March 2023).

Wesp, R., Hesse, J., Keutmann, D. and Wheaton, K. (2001) ‘Gestures maintain spatial imagery’, The American Journal of Psychology, 114(4), pp. 591-600.

Whittaker, L., Russell-Bennett, R. and Mulcahy, R. (2021) ‘Reward-based or meaningful gaming? A field study on game mechanics and serious games for sustainability’, Psychology & Marketing, vol. 38, no. 6.

Wild, F., Loesch, F., Huang, X. and Bouali, S. (2019) holoCARE. Available at: https://github.com/fwild/holocare (Accessed: 15 March 2023).

Wolf, E., Döllinger, N., Mal, D., Wienrich, C., Botsch, M. and Latoschik, M.E. (2020) ‘Body weight perception of females using photorealistic avatars in virtual and augmented reality’, Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). Porto de Galinhas, 9-13 November. Piscataway, NJ: IEEE, pp. 462-473.

Wolf, E., Fiedler, M.L., Döllinger, N., Wienrich, C. and Latoschik, M.E. (2022) ‘Exploring presence, avatar embodiment, and body perception with a holographic augmented reality mirror’, Proceedings of the 2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). Christchurch, 12-16 March. Piscataway, NJ: IEEE, pp. 350-359.

Wright, P. and Mccarthy, J. (2008) ‘Empathy and experience in HCI’, Proceedings of the 2008 Conference on Human Factors in Computing Systems. Florence, 5-10 April. New York, Association for Computing Machinery, pp. 637-646.

Yaghmale, F. (2009) ‘Content validity and its estimation’, J Med Educ, vol. 3.

Yeo, A., Ledesma, I., Nathan, M.J., Alibali, M.W. and Church, R.B. (2017) ‘Teachers’ gestures and students’ learning: sometimes “hands off” is better’, Cognitive Research: Principles and Implications, 2(1), pp. 1-11.

Yoo, S. and Tanaka, J. (2022) ‘Increasing motivation of walking exercise using 3D personalized avatar in augmented reality’, in Yamamoto, S. and Mori, H. (eds.), Proceedings of the Thematic Area Human Interface and the Management of Information, Part of HCI International 2022. Virtual Event, 26 June-1 July. Cham: Springer, pp. 172-191.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019) ‘Systematic review of research on artificial intelligence applications in higher education–where are the educators?’ International Journal of Educational Technology in Higher Education, 16(1), pp. 1–27.

Zhang, Z., Wu, Y., Pan, Z., Li, W. and Su, Z. (2022) ‘A novel animation authoring for the virtual teacher performing experiment in mixed reality’, Computer Applications in Engineering Education, 30(2), pp. 550-563.

Zhou, S.Z.Y., Tan, D., Wei, D. and Nakayama, Y. (2009) ‘TangiCube: introducing bi-directional interactivity with a virtual agent in a tangible mixed reality environment’, Proceedings of the International Conference on Advances in Computer Entertainment Technology. Athens, 29-31 October. New York, NY: Association for Computing Machinery, pp. 413-414.

Zielke, M.A., Zakhidov, D., Hardee, G.M., Pradeep, J., Evans, L., Lodhi, Z., Zimmer, K. and Ward, E. (2018) ‘Exploring medical cyberlearning for work at the human/technology frontier with the mixed-reality emotive virtual human system platform’, Proceedings of the 2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH). Vienna, 16-18 May. Piscataway, NJ: IEEE, pp. 1-8.

APPENDICES

Extracting Teaching Gestures for Animation: An Experiment

Appendix 1: Participant information sheet

RESEARCH STUDY PARTICIPANT INFORMATION SHEET 'MOTION CAPTURE OF GESTURES USED IN TRAINING'

Conducted by
The Institute of Educational Technology (IET)
Prof Dr Fridolin Wild (f.wild@open.ac.uk, 01908-858 885)
and Xinyu Huang (xinyu.huang@open.ac.uk)
with the
Institute of Educational Technology (Prof Dr Denise Whitelock,
denise.whitelock@open.ac.uk, 01908-653777)

You are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully. Your participation is entirely voluntary and, should you work or study with the OU, it has no impact on your study success, evaluation, or progression at work.

With this study, we seek to explore whether there are specific gestures teachers and trainers use when they explain to learners or trainees how to do specific things – like building cardboard fort. Aim is to identify recurring gestures (such as particular pointing gestures) of effective instruction so that we can utilise these gestures in Augmented Reality teaching applications, equipping 3D character models with a repertoire for guiding and instructing. The research is partially financially supported by the from the European Union's Horizon 2020 research and innovation program under grant agreement No 856533.

For this, you will put on a motion capture suit, that helps us collect data about your limb and joint movement, orientation, and position (using 19 body-worn sensors). The motion capture suit is worn above your regular clothes. If you feel wearing the suit is uncomfortable for you, you have the right to withdraw from the study at any time. We will also provide you with a pair of Augmented Reality smart glasses. Other than Virtual Reality glasses, AR glasses do not suffer of any noticeable motion sickness. Should you, however, experience any discomfort wearing them, you have the right to withdraw from the study at any time. Additionally, we record a room video to facilitate studying your motions in the context in which where they were delivered. Basic demographic data will be collected (age bracket, gender) as well as measurements of limb lengths required to configure the motion capture.

You will instruct a trainee on three tasks:

- navigate the trainee to a specific location in the room and make them stand in a specific body pose
- assembly of a cardboard fort
- precision assembly of a specific Tangram shape (also known as "puzzle T")

We are interested in your natural behaviour, so we have designed the tasks to resemble commonplace tasks that do not require any training for you and that require only common sense in their execution.

The collected data will be anonymised (faces visible in the videos will be blurred, audio recording transcribed to subtitles). It is aimed for that the resulting data set be published as open research data (e.g. on our institutional open research data repository ORDO), likely pooled with data from other experiments. It will be referred to in publications in anonymised form only.

This project has been reviewed by, and received a favourable opinion from, The Open University Human Research Ethics Committee, reference HREC/4366/Wild

It is up to you to decide whether or not to take part. If you do decide to take part, you will be given this information sheet to keep and be asked to sign a consent form.

You have the right to withdraw from the study at any time during your participation, leaving the session and informing the researchers.

You have the right to ask for your data to be removed after your participation in the study by email to xinyu.huang@open.ac.uk by end of October 2022.

If you have a concern about any aspect of this study, please contact xinyu.huang@open.ac.uk and f.wild@open.ac.uk, and we will do our best to answer your query. I/we will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at The Open University who will seek to resolve the matter as soon as possible: Clair Hewson, Chair, Open University Human Research Ethics Committee, The Open University, Charles Pinfold Building, Level 3, Walton Hall, Milton Keynes, MK7 6AA, Email: Research-REC-review@open.ac.uk

1. THANK YOU

2. DATA PROTECTION

The Open University is the Data Controller for the personal data that you provide.

The lawful reason for processing your data will be that conducting academic research is part of The Open University's public task. (The consent we request from you relates to ethical considerations)

You have a number of rights as a data subject:

- To request a copy of the personal data we have about you
- To rectify any personal data which is inaccurate or incomplete
- To restrict the processing of your data
- To receive a copy of your data in an easily transferrable format (if relevant)
- To erase your data
- To object to us processing your data

If you are concerned about the way we have processed your personal information, you can contact the Information Commissioner's Office (ICO). Please visit the ICO's website for further details.

Appendix 2: Consent form

INFORMED CONSENT FOR 'MOTION CAPTURE OF GESTURES USED IN TRAINING'

Prof Dr Fridolin Wild, IET, WELS

Xinyu Huang, IET, WELS

Please highlight your choice by clicking inside the appropriate box

1. Taking part in the study

I have read and understood the information sheet for the 'teaching gestures' study, or it has been read to me.	YES ☐	NO ☐
I have been able to ask questions about my participation and my questions have been answered to my satisfaction.	YES ☐	NO ☐
I consent voluntarily to be a participant in this study and understand that I can refuse to do tasks I am not comfortable with, and I can withdraw from the study at any time by contacting Xinyu Huang (xinyu.huang@open.ac.uk) up until 31.10.2022, without having to give a reason.	YES ☐	NO ☐

I understand that taking part in the study involves wearing a motion capture suit for recording my joints movement and being filmed, while performing a given teaching task (navigation, assembly, and precision). Subsequently, I will fill in a simple questionnaire with basic demographic information.	YES ☐	NO ☐
I agree to photos and videos being taken during the observation sessions	YES ☐	NO ☐

2. Use of the information in the study

I understand that information I provide will be used for publications, but only in fully anonymised form.	YES ☐	NO ☐
I understand that personal information collected about me that can identify me will not be shared beyond the study team.	YES ☐	NO ☐

3. Future use and reuse of the information by others

I give permission for the motion capture and film data plus the additional simple demographic data that I provide to be deposited in a specialist data centre after it has been anonymised, so it can be used for future research and learning.	YES ☐	NO ☐

I would like to receive a copy of the summary of the findings of this study. *Please insert your email address in the space below if you answer 'yes'* Email address …………………………………………………………………..	YES ☐	NO ☐

4. Signatures

Name of participant [in CAPITALS] ________________________	**Signature** ____________________ **(electronic signatures may be accepted**	**Date** __________

For participants unable to sign their name, please mark the box instead of signing ☐

This project has been reviewed by, and received a favourable opinion from, The Open University Human Research Ethics Committee, reference HREC/4366/Wild

Appendix 3: NVivo coding data

Name	Files	Reference
Representational gestures	**132**	**2348**
Beat	16	41
Deictic	**125**	**1290**
Deictic_direction	86	251
Deictic_object	105	502
Deictic_position	100	537
Emblematic	**65**	**364**
Emblematic_No	23	52
Emblematic_counting fingers	18	40
Emblematic_remove	8	8
Emblematic_thumb up	46	238
Emblematic_wait/hold on	15	26
Iconic	**61**	**198**
Iconic_angle	9	17
Iconic_length	4	6
Iconic_shape	54	145
Iconic_size	3	6
Iconic_spatial position	15	24
Metaphorical	**7**	**8**
Mimicking	**44**	**155**
Transformational	**79**	**292**
Transformational _flip	20	31
Transformational _other side	4	5
Transformational_rotation	76	256

Trust towards Holographic AIs: An Experiment

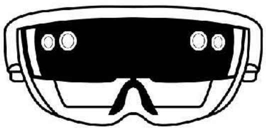
Appendix 4: Participant information sheet (Guardians)

RESEARCH STUDY PARTICIPANT INFORMATION SHEET 'DEVELOPMENT OF TRUST BETWEEN HUMAN AND HOLOGRAPHIC AIS' (PARENTS)

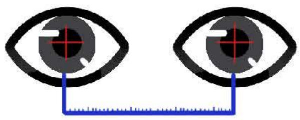
Conducted by
Xinyu Huang (xinyu.huang@open.ac.uk)
and Prof Dr Fridolin Wild (f.wild@open.ac.uk, 01908-858 885)
with the
Institute of Educational Technology (Prof Dr Denise Whitelock,
denise.whitelock@open.ac.uk, 01908-653777)

Your children are being invited to take part in a research study. Before you decide whether or not to take part, it is important for you to understand why the research is being done and what it will involve. This project that choosing to either take part or not take part in the study will have no impact on your marks, assessments or future studies and work. Please take time to read the following information carefully.

Trust is crucial in human-computer interaction (HCI), especially when working with artificial intelligence (AI) systems. This project aims to investigate how develop a sense of trust and acceptance towards holographic AIs, i.e., animated, smart, 3D character models. In this project, holographic AIs will help children identify different 3D shapes, such as cube, cuboid, cone, and sphere, by face-to-face communication.



We are inviting children between the age of 8 -11 years old to participate in our research. This project uses Microsoft HoloLens to project virtual objects into the real world. However, too young children possibly cannot observe whole view of virtual surrounding by using Microsoft HoloLens as interpupillary distance (IPD) of HoloLens is around 63mm, but the 8-11 years old kid's is 58mm-60mm. Therefore, we will try to adjust IPD by HoloLens's interpupillary calibration application at first, but if children cannot see whole virtual surrounding, we will replace the HoloLens with an iPad.



We respect the autonomy of every person, and you have the ultimate right to the decision of whether to participate in this study. If you decide to take part, you are still free to withdraw at any time and without giving a reason for such a move.

If you agree to participate in this study, you will need to confirm your agreement and sign your name in a consent form. We will introduce the application, and show how to use HoloLens and interact with holographic AIs. The experience will take up about 20 minutes of your time. After that, you need to fill in a questionnaire about your experience with the holographic AI, which takes about 5 minutes of your time. If, after you have started to take part, you change your mind, just let me know and I will not use any information you have given me in my writing.

This research contributes to the sense of trust towards holographic AIs and development of interaction between children and agents in order to improve its user experience. If you feel holographic AIs are not suitable for you. You can change to other holographic AIs, or you can stop the app immediately.

All information collected about the individual will be kept strictly confidential, and research data will be securely stored at all times. Your data is stored anonymously. The data generated in the course of the research will be kept securely in paper and electronic form for a maximum of ten years after the completion of a research project.

The result of the research will be published as a conference paper, dissertation, or presentation. A copy of the findings of the study will be offered to you, if you tick the box on the consent form.

You have the right to withdraw from the study at any time or contacting researchers. You have the right to ask for your data to be removed after the study by contacting researchers, up until all data have been aggregated for analysis.

If you want to take part in this study, please contact Xinyu Huang (xinyu.huang@open.ac.uk), the deadline for participation is the end of 11/2022.

The Open University is the Data Controller for the personal data that you provide.

The lawful reason for processing your data will be that conducting academic research is part of The Open University's public task. (The consent we request from you relates to ethical considerations.

DATA PROTECTION

The Open University is the Data Controller for the personal data that you provide.

The lawful reason for processing your data will be that conducting academic research is part of The Open University's public task. (The consent we request from you relates to ethical considerations)

You have a number of rights as a data subject:

- To request a copy of the personal data we have about you
- To rectify any personal data which is inaccurate or incomplete
- To restrict the processing of your data
- To receive a copy of your data in an easily transferrable format (if relevant)
- To erase your data
- To object to us processing your data

If you are concerned about the way we have processed your personal information, you can contact the Information Commissioner's Office (ICO). Please visit the ICO's website for further details.

Appendix 5: Participant information sheet (Children)

RESEARCH STUDY PARTICIPANT INFORMATION SHEET 'DEVELOPMENT OF TRUST BETWEEN HUMAN AND HOLOGRAPHIC AIS'

Conducted by
Xinyu Huang (xinyu.huang@open.ac.uk)
and Prof Dr Fridolin Wild (f.wild@open.ac.uk, 01908-858 885)
with the
Institute of Educational Technology (Prof Dr Denise Whitelock,
denise.whitelock@open.ac.uk, 01908-653777)

To be read and discussed with children aged 8-11 years old.

We would like to invite you to take part in a research study called Trust towards holographic AIs. Before you decide, it is important to understand what the study is and what will happen you if you take part. Please read the following information carefully and ask us about anything that you do not understand.

What is the study for?

The study is trying to improve feeling of trust between children and virtual humans. Sarah, a virtual humans, is an animated, smart, 3D female, she will help you to understand what 3D shapes and 2D shapes are. We will use HoloLens2, so that you will see Sarah in your real surrounding.

The result of all this work may help us to understand how to improve trust relationship with Sarah.

Who can take part?

You can take part in this study if you are between the ages of 8-11.

In order to take part, You must have permission from your parents or the person who looks after you if you want to take part.

Do I have to take part?

Not at all! It is up to you to decide if you do not want to take part. If you decide to take part, you will need to fill in the enclosed form.

You can leave the study at anytime without giving a reason.

What will happen to me if I take part?

- If you are able, you will be asked to fill out a consent form.

  The form is to say that you understand the study and what will happen. You will be given your own copy of the form to keep, as well as this leaflet.

- You will need to fill out a demographic form.

  This form will let us know your age and whether you had experience in playing with virtual humans.

- We will help you to understand the application, and show how to use HoloLens and interact with Sarah.

- After interaction, we will ask you to tell us how you feel by asking you to fill out a questionnaire. We will also send a little gift for you after finished the questionnaire.

What if the questions bother me or if I don't want to answer any questions?

You can stop at any time, and you do not have to give a reason, You can also contact the research team at any time, and say that you want your answers about certain questions to be destroyed, which we will do straight away.

What will happen to my data?

All of the information will be stored in a locked file cabinet and only researchers working on this study can look at this information.

Who has reviewed the study?

This study has received ethical approval from Open University

Research Ethics Committee (approval number:)

Who is responsible for this study?

Xinyu Huang from Open University is the main researcher for this study.

If you have any questions, you can contact us at any time. Email Xinyu Huang – xinyu.huang@open.ac.uk.

Thank you for reading this information leaflet.

Appendix 6: Consent form (Guardians)

CONSENT FORM (Guardians)

INFORMED CONSENT FOR DEVELOPMENT OF TRUST BETWEEN HUMAN AND HOLOGRAPHIC AIS

Xinyu Huang and Prof Fridolin Wild

Institute of Educational Technology

xinyu.huang@open.ac.uk

f.wild@open.ac.uk

Mark with √ in box	**Please initial box**
I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.	☐
I understand that my child's/ chirden's participation is voluntary and that I am free to withdraw at any time, without giving reason.	☐
I agree to the use of anonymised quotes in publications.	☐
I agree that an anonymised data set, gathered for this study may be stored in a specialist data centre/repository relevant to this subject area for future research.	☐
I agree to the interview.	☐
I would like to receive a copy of the summary of the findings of this study. Please insert your email address in the space below if you tick box Email address …………………………………………………………..	☐
I agree to take part in the above study.	☐

Name of participant [in CAPITALS]	Signature	Date
________________________	____________________	________
	(electronic signatures may be accepted	

If your project will be reviewed by HREC:

This research project has been reviewed by, and received a favourable opinion from, The Open University Human Research Ethics Committee – HREC reference number: HREC/4480/Huang

OR where the project does not need formal HREC review:

This research project conforms to and complies with the OU Human Research Ethics Committee's conditions for exemption from formal review.

Appendix 7: Consent form (Children)

CONSENT FORM (Children)

INFORMED CONSENT FOR DEVELOPMENT OF TRUST BETWEEN HUMAN AND HOLOGRAPHIC AIS

Xinyu Huang and Prof Fridolin Wild

Institute of Educational Technology

xinyu.huang@open.ac.uk

f.wild@open.ac.uk

You will be given a copy of this information sheet and consent form to keep. Taking part in this study is voluntary. This means you can refuse to be a part of this study. Also, you can decide to withdraw from this study at any point without any reasons. If you wish to stop at any time, just tell us.

	Mark with √ in box
Has somebody explained this study to you?	Yes / No
Do you understand what the study is about?	Yes / No
Have you asked all the questions you want?	Yes / No
Have you had your questions answered in a way you understand?	Yes / No
Do you agree to have a short interview?	Yes / No
Do you understand it's OK to stop taking part at any time?	Yes / No
Are you happy to take part?	Yes / No

If any answers are 'no' or you don't want to take part, don't sign your name!
If you do want to take part, please write your name and today's date

Name of participant [in CAPITALS]	**Signature**	**Date**
______________________	____________________	________
	(electronic signatures may be accepted	

If your project will be reviewed by HREC:

This research project has been reviewed by, and received a favourable opinion from, The Open University Human Research Ethics Committee – HREC reference number: HREC/4480/Huang

OR where the project does not need formal HREC review:

This research project conforms to and complies with the OU Human Research Ethics Committee's conditions for exemption from formal review.

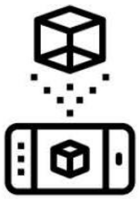
Appendix 8: Demographic Questions

Development of trust between human and holographic AIs

Demographic Questions

1. What is your gender? *(select one answer)*
   - ☐ Girl
   - ☐ Boy
2. What is your age? ✎ ______________ [numerical input]
3. Have you used Augmented Reality app before (such as Pok é mon Go)?

   (select one answer)
   - ☐ Yes, often
   - ☐ Yes, sometimes
   - ☐ No
   - ☐ I don' t know



4. Have you chatted with virtual cartoon/ human? *(select one answer)*
   - ☐ Yes, often
   - ☐ Yes, sometimes
   - ☐ No
   - ☐ I don' t know

Appendix 9: Questionnaire

Trust towards Sarah

The questionnaire evaluates how you feel about holographic AIs in order to analyse the relationship of trust between children and holographic AIs.

Section 1. Please tick ONE BOX ONLY in each row.

Strongly Disagree | Disagree | Undecided | Agree | Strongly Agree

1. Sarah is clever.
   - ☐ Strongly Disagree
   - ☐ Disagree
   - ☐ Undecided
   - ☐ Agree
   - ☐ Strongly Agree

2. Sarah knows what she is doing.
   - ☐ Strongly Disagree
   - ☐ Disagree
   - ☐ Undecided
   - ☐ Agree
   - ☐ Strongly Agree

3. I like Sarah.
   - ☐ Strongly Disagree
   - ☐ Disagree
   - ☐ Undecided
   - ☐ Agree
   - ☐ Strongly Agree

4. Sarah does not lie.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree
5. I think Sarah wants to do good.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree
6. Sarah is kind.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree
7. Sarah feels real to me.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree
8. Sarah looks out for me.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree

9. Sarah wants to help me.
   - ☐ Strongly Disagree
   - ☐ Disagree
   - ☐ Undecided
   - ☐ Agree
   - ☐ Strongly Agree

10. Sarah is caring.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree

11. Sarah is my friend.
    - ☐ Strongly Disagree
    - ☐ Disagree
    - ☐ Undecided
    - ☐ Agree
    - ☐ Strongly Agree

Section2. Please tell us more about your responses:

1. Would you like to share your story or life with Sarah? And why?

   ____

2. Do you think Sarah can keep your secrets?

   ____

3. What did Sarah do well?

4. What did Sarah not do so well?