Social Believability of Human-Driven Embodied Conversational Avatars in Shared Virtual Worlds:

The Impact of the Adaptation Gap on Users' Experience of Interacting with VR Avatars

by

Andrey Goncharov

B.Sc. (Interactive Arts and Technology), Simon Fraser University, 2020

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

SIMON FRASER UNIVERSITY
Spring 2024

Declaration of Committee

Name:	**Andrey Goncharov**
Degree:	**Master of Science**
Title:	**Social Believability of Human-Driven Embodied Conversational Avatars in Shared Virtual Worlds:** ***The Impact of the Adaptation Gap on Users' Experience of Interacting with VR Avatars***
Committee:	**Chair: Jon Corbett** Instructor, Interactive Arts and Technology
	Ozge Nilay Yalçın Supervisor Assistant Professor, Interactive Arts and Technology
	Steve DiPaola Committee Member Professor, Interactive Arts and Technology
	Jim Bizzocchi Examiner Professor Emeritus, Interactive Arts and Technology

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

The research investigates socially believable interactions of human-driven Embodied Conversational Avatars visiting shared 3D virtual communities. Many social VR platforms allow users to use two kinds of avatar control mechanisms: motion-tracked or automated-gesturing. This potentially creates a discrepancy in the user's experience and impacts immersion when exposed to different avatars. This thesis adapts Komatsu et al.'s Adaptation Gap concept to measure the difference in participants' expectations and perceptions of both avatars, testing the assumption that the difference in avatar control type has an impact on user experience. The results support the argument that avatar control type matters ($p < .05$) and so does the context (or scenario) in which the interaction takes place ($p < .05$). The difference between user expectations and their perceptions is correlated ($p < .05$) with the believability of the avatars, which is known to have an impact on their interaction.

Keywords: Virtual Reality; Avatar; Motion Tracking; Automated Gesture; Adaptation Gap; User Perception

Dedication

To my family and friends, and to my cat, Pushok.

Acknowledgements

Thank you to my supervisors Dr. Steve DiPaola and Dr. Nilay Yalçin for their encouragement and support. Especially a big thank you to Nilay for helping me through the hardest parts of my research.

I also want to thank my SIAT cohorts Servet Ulas, Hanieh Shakeri, and Chelsea Mills for their support and help, including answering my constant influx of questions.

Thank you also to my friends and family who kept me sane all these months/years Alex, Olga, Olga Jr., Sasha, Maria, Nick, Sydney, Nick, and Yang.

Must also be said, thank you to all my participants who took time out of their day to help me with my study by participating.

Thank you to the Tivoli Cloud VR community, especially thank you to Caitlyn and Maki for helping with technical issues and supporting my work in Tivoli, and also thank you to Raz and XaosPrincess for volunteering and helping with the videos for the study. In addition, thank you to Jeremy for touring me through various VR platforms.

Finally, thank you to Eric Yang, Kevin Schut, and Trinity Western University for allowing me to work during these hard and stressful times and giving me an opportunity to practice and express my passion for Game Design and Development.

Table of Contents

Declaration of Committee ii
Ethics Statement iii
Abstract iv
Dedication vi
Acknowledgements vii
Table of Contents viii
List of Tables x
List of Figures xi

Chapter 1. Introduction 1
1.1. Background Context 1
1.2. The Research 4
1.3. The Study 7
1.4. The Contribution 8
1.5. Thesis Structure 9

Chapter 2. Concepts and Related Work 10
2.1. Embodied Conversational Agents/Avatars 11
2.1.1. Behaviour Generation for ECAs 14
2.1.2. ECAs and Virtual Worlds 16
2.2. Social Believability 19
2.2.1. Enhancing Believability of Avatars and Virtual Humans 20
2.2.2. Believability and Games 22
2.2.3. Evaluating Believability 24
2.3. Adaptation Gap 26

Chapter 3. Difference Between the Adaptation Gap and the Current Study 29

Chapter 4. Social VR Platform Taxonomy and the Study Environment 32
4.1. Automated Gesture Generation Systems in Social VR Platforms 33
4.1.1. Method of Selection 34
4.1.2. Choosing a Platform 36
4.2. The Two Avatar Control Systems in Tivoli Cloud VR 49

Chapter 5. The Study and the Results 52
5.1. Study Overview 54
5.1.1. The Participants 55
5.1.2. Creating the Avatar Videos for the Study Survey 56
5.1.3. Creating the Survey 61
5.1.4. Study Procedure 63
5.2. Results 67
5.2.1. Social Fidelity Expectations before Meeting the Avatars 68
5.2.2. Observer's Reception to their First Avatar 69

5.2.3. Adaptation Gap Scores Regardless of Exposure Order 73
5.2.4. Adaptation Gap versus Believability 76
5.2.5. Avatar Preference and Qualitative Results 77

Chapter 6. Discussion 81
6.1. Observations of the Open-Ended Questions 86
6.2. Potential Factors Explaining Study Results 89
6.3. Limitations 91
6.4. Recommendations on Multiple Conversational Avatar Exposure 93
6.4.1. Conversational Avatar Adaptation Gap and Gaming 97

Chapter 7. Future Work 100
7.1. Addressing the Limitations 101
7.2. Expanding the Research 103

Chapter 8. Conclusion 108

References 110

Appendix A. Sample of Social Believability Planning Document: Survey Structure and Questions 118

Appendix B. Avatar Exposure Qualitative Responses 127

Appendix C. End of Survey Qualitative Responses 156

List of Tables

Table 1. Tanenbaum et al.'s (2020) NVC High Level Categories with their inner Categories and Sub-Categories 17

Table 2. Categories of our Taxonomy and their explanations as seen in Tables 3 - 5. Categories include citations of the work where the concepts were taken from. 35

Table 3. Social VR Platform Taxonomy. Part 1: VR and Desktop Support. 37

Table 4. Social VR Platform Taxonomy. Part 2: Gonzalez-Franco & Peck's (2018) Avatar Embodiment categories. 38

Table 5. Social VR Platform Taxonomy Part 3: Tanenbaum et al.'s (2020) Believability categories and Other Categories. 43

Table 6. The items in the Social Fidelity questionnaire. 62

Table 7. Believability questionnaire used in the study. 62

Table 8. 2x2 within-subject design with counterbalanced orders (Desktop = automated-gesturing; VR = motion-tracked). 67

Table 9. Results of the Pairwise Comparison for Adaptation Gap Scores including t-values, significance levels (p) and effect sizes (d) (Desktop = automated-gesturing; VR = motion-tracked). 75

List of Figures

Figure 1. Tivoli Cloud VR, an immersive 3D social platform, website landing page (Tivoli Cloud VR, N.D.) 2

Figure 2. A group photo of users using motion-tracked and automated-gesturing avatars in Tivoli Cloud VR (Meeks, 2020) 5

Figure 3. Examples of different avatar visual and behavioural realism. Mozilla Hubs with disembodied avatar (left), Engage VR with fully embodied and face scanned avatar (right) 12

Figure 4. Diagram showing relationship between the Adaptation Gap and Expectation/Fidelity variables. Diagram modified from Komatsu et al. (2012). 27

Figure 5. Screenshots from visits to the social VR platforms. VR Chat (left), Engage VR (right). 36

Figure 6. Additional screenshots from visits to the social VR platforms. From left to right, top to bottom: Neos VR, Mozilla Hubs, Alt Space VR, Tivoli Cloud VR. 37

Figure 7. Screenshot of Tivoli's Automated-Gesturing Avatar in Pitch Scenario shown closer to the screen, this version internally referred to as the Desktop Avatar. 49

Figure 8. Screenshot of Tivoli's Motion-Tracked Avatar in Disco Scenario shown closer to the screen. 50

Figure 9. Screenshots capturing a sequence of conversational gestures of the Motion-Tracked avatar in Disco Scenario from Tivoli Cloud VR. 51

Figure 10. Screenshots capturing a sequence of conversational gestures of the Automated-Gesturing avatar in Pitch Scenario from Tivoli Cloud VR. 51

Figure 11. Diagram showing the study structure. The study blocks have their own structure inside. 54

Figure 12. Diagram showing the detailed structure of the Study Blocks shown in Figure 11. 55

Figure 13. Screenshot of a meeting room with Automated-Gesturing avatar in Pitch scenario. 56

Figure 14. Screenshot of Automated-Gesturing avatar in Pitch scenario pointing at an art board. 57

Figure 15. Screenshot of the disco pub with Motion-Tracked avatar (left) in Disco scenario. 58

Figure 16. Screenshot of Motion-Tracked avatar in Disco scenario pointing at a photo of a dog. 58

Figure 17. Snapshot of survey editor in Survey Monkey editing the landing page of the Desktop-Pitch variant of the survey. 63

Figure 18. Editor screenshots of Desktop-Pitch variant survey preview showing test questions. Left: video check test (private video does not represent what participants see). Right: segment of post video check questions. 64

Figure 19. Editor screenshot of VR-Pitch variant survey preview showing avatar photo (Motion-Tracked avatar in Pitch scenario) and initial impressions questions. 65
Figure 20. Editor screenshot of Desktop-Disco variant survey preview showing the first avatar scenario video. 66
Figure 21. Editor screenshot of Desktop-Disco variant survey preview showing part of the post video 1 questionnaires. 66
Figure 22. Boxplot graphs showing the Pre-Social Fidelity values for the groups based on the initially exposed Avatar Type (Desktop = automated-gesturing; VR = motion-tracked). There were no significant differences between avatar control types. 69
Figure 23. Descriptive Statistics showing Mean and Std. Deviation for Adaptation Gap between Scenario and Avatar Type (Desktop = automated-gesturing; VR = motion-tracked). 70
Figure 24. Test of Between Subject Effects for Adaptation Gap between Scenario and Avatar Type, including the interaction effect. There is a statistically significant interaction between Avatar Type and Scenario. 71
Figure 25. Estimated Marginal Means of Adaptation Gap with Avatar Type (Desktop = automated-gesturing; VR = motion-tracked). 71
Figure 26. Estimated Marginal Means of Adaptation Gap with Scenario (Desktop = automated-gesturing; VR = motion-tracked). 72
Figure 27. Boxplot graphs showing the Adaptation Gap values for the interaction groups based on Avatar Type and Scenario (Desktop = automated-gesturing; VR = motion-tracked). 74
Figure 28. Scatterplot showing the linear regression of Adaptation Gap and Believability values with a line of best fit, confidence and prediction intervals. 76
Figure 29. Screenshot of Autimated-Gesturing avatar from Disco scenario pointing at a picture of a dog with a red box indicator. 89

Chapter 1.

Introduction

1.1. Background Context

Virtual Reality and 3D Metaverses have increasingly been areas of interest in the contexts of remote work and collaboration, particularly in response to 2019's onset of the COVID-19 virus pandemic where the social distancing requirements especially exposed the need for these technologies. Virtual Reality technology and devices are a growing industry, with Fortune Business Insights stating in their report on the market share of Virtual Reality applications that Virtual Reality (or VR) is a developing and growing industry (Fortune Business Insights, 2020). Its applications range from health to entertainment and design, with a market size of 3.1 billion USD in 2019 and "projected" to grow even higher. The topic of the Metaverse was adopted by Facebook (now Meta) who have been talking about their growing plans for its universal Metaverse - an interconnected set of digital worlds where both people and virtual agents coincide in an immersive and engaging way of work, socialization, and play (Newton, 2021; Ravenscraft, 2021). There has been increasing interest in the concept of Metaverses, where Businesswire (2023) stated the Global Metaverse Market is projected to have a compound annual growth rate of 40% and a projected revenue of 700 Billion US Dollars, with the article stating the growth will be due to the increasing popularity of Virtual Reality, Augmented Reality, and Mixed Reality technologies that will help "the metaverse [in] gaining momentum as it helps in connecting the physical world with the virtual environment".

This thesis focuses on immersive 3D social environments or platforms (including virtual worlds, metaverses, social VR platforms, immersive environments, and 3D virtual communities) and the user embodiment in these environments. Albeit some of their differences, these terminologies are often used interchangeably both by its users and by academic or media publications.

- *Virtual worlds* are considered as online environments that are accessed by multiple users, using 3D digital rendering to represent the environment and the

users in the form of avatars, where they can interact with each other and the environment (Dass et al., 2011).

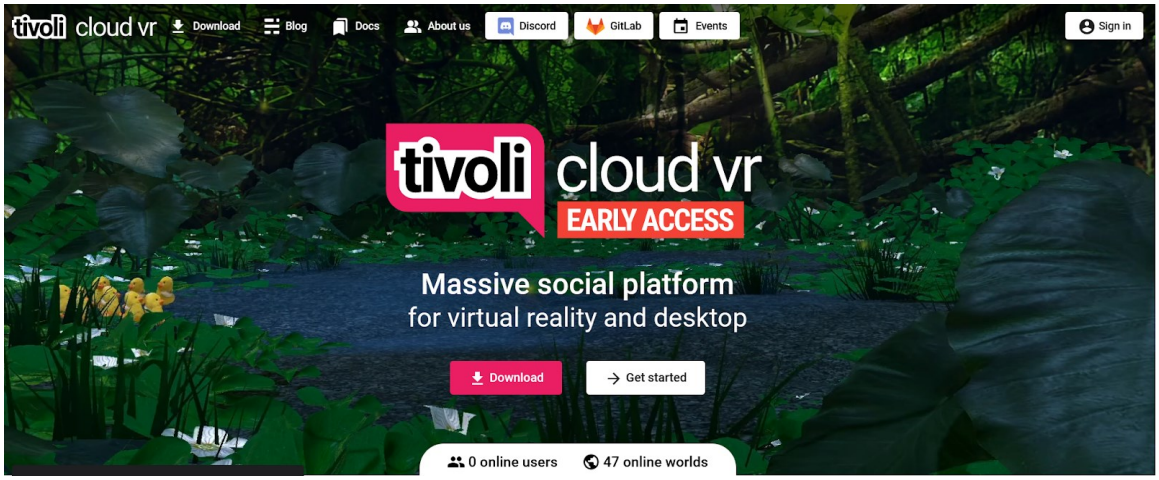

Figure 1. Tivoli Cloud VR, an immersive 3D social platform, website landing page (Tivoli Cloud VR, N.D.)

- Metaverses are a recently popular term for virtual worlds that reflect the physical world but also can dynamically change and connect - allowing for the physical and virtual worlds to interact together (Wang et al., 2022).

- *Social VR Platforms* are digital platforms/communities that host virtual worlds for users to interact and engage in social activities, mostly by using VR equipment. The users of social VR platforms use the technology for formal and informal purposes - from casual meetings, recreation, get togethers, parties; to business meetings, brainstorming sessions, and cooperative work sessions.

- *Immersive Environments* are defined as 3D virtual environments that focus on delivering a high degree of presence and engagement with sensory inputs for the user (Rubio-Tamayo et al., 2017).

- *3D Virtual Communities* is another word for Social VR Platforms but with a focus on 3D environments and building/supporting communities. These communities found favor due to their online nature that allows members across the globe to inhabit spaces together and interact, collaborate, socialize, and play.

With the improvements in Virtual Reality (VR) technology (including VR headsets and real-time movement tracking systems), there has been a rapid increase in social VR and multiplayer VR games and platforms, such as VRChat Inc.'s VR Chat (2014) and Tivoli Cloud VR (Meeks et al., 2020), among others. These social VR platforms allow users to interact in a multi-player setting with the VR environment and other users using avatars, which are digital representations or embodiment of users in the virtual environment. Due to the social, multi-user, and online nature of these platforms, avatars that are capable of showing socio-emotional behaviors (i.e., gestures, movement and expressions of the user) can be crucial for natural interaction, along with the need to increase their fidelity into immersive and believable agents/avatars. Some platforms such as the aforementioned Tivoli Cloud VR (Figure 1), a fork of High Fidelity (2013 - 2019), allow for users to control their avatar behavior, including these socio-emotional behaviors, using different methods: *motion-tracked* or *automated-gesturing*.

- *Motion-tracked* avatars can be driven via real-time standing body movement of the user using VR Technology such as Oculus Rift (Oculus, 2016) or HTC Vive (HTC, 2016) head-mounted displays and head plus hands tracking hardware.

- *Automated-gesturing* avatars are controlled without VR technology, from a user sitting in front of a computer, where typically a software algorithm based on the user's voice automates the gestural behaviour of avatars including lip-sync, facial gestures, head and upper body movement, and is sometimes augmented via keyboard and mouse control.

The former ("Motion-Tracked avatars") transfers actual individual movements from the user to the avatar, and the latter ("Automated-Gesturing avatars") automates via software-programmed movements of the avatar augmented by voice or mouse controls. Note that there is a spectrum of different motion-tracked and automated-gesturing avatars, from quite limited to quite advanced systems, where one avatar can incorporate both mechanisms for different types of behaviour control. The coupling between the human user and the avatar behaviors in the virtual environment highly depends on the quality of the behaviors these different avatar control mechanisms can create. Believability of these systems in terms of their social behaviors can therefore also be affected by these control mechanisms.

Although both the popularity of social VR platforms in terms of active users and the techniques to generate avatar behavior are steadily increasing, the impact of using different avatar control mechanisms on user experience in these immersive 3D social environments has not been previously examined. Our research is intended as a first step to evaluate the effect of using different avatar control mechanisms in social VR platforms (i.e., automated gesturing and motion tracking) through the lens of Social Believability. Our study does not involve fully automated computer agents, sometimes called non player characters (NPCs) in gaming, that do not involve a human controller. We use the term “conversational avatars“ or just “avatars” which are defined as embodied characters that are specifically driven by humans. This avatar definition involves the notion of 'binding the pair': the unification of a remote user with their online corresponding (virtual embodied) avatar (DiPaola & Collins, 1999; DiPaola et al., 2011). However, our “avatar” research touches upon important topics concerning all digital characters that can inhabit these 3D social immersive environments, including automated gesture generation and believability studies.

1.2. The Research

Against this backdrop, our research focuses on the Social Believability of human-driven 3D Conversational Avatars in Social VR Platforms and the differences of social believability between different avatar control mechanisms. Social Believability is the user’s reception of an entity or virtual human, including the level of immersion, ease of suspension of disbelief, and an entity’s ability to behave convincingly in social contexts (Afonso & Prada, 2009; Nixon, 2009, p. 9 - 13). The better an entity’s social fidelity or level of depth and complexity of social behavioural properties and features (see Chapter 3 for more details), the more socially believable it is.

Believability in this thesis is used as a term associated with natural interaction behaviour, experience, and performance of an entity. It can be applied to virtual characters, animated characters, humans playing as characters, etc. In traditional forms of media and performance arts - believability is defined as how well a character allows a viewer to suspend their disbelief and allows the portrayal of convincing personalities expected by the viewer (Loyall, 1997, p.1). In other words, it provides the character with the “illusion of life” (Gomes et al., 2013, p.3). Furthermore, character believability can also be influenced by the context of the interaction or the narrative (Bizzocchi et al., 2013;

Tanenbaum & Bizzocchi, 2009). Social Fidelity (see definition in Chapter 3) is a component of Believability, which is especially important in characters who engage in social interaction. If a character's social and behavioural features are of high level and quality, and thus having a high level of Social Fidelity, then that will produce a highly believable character. In the context of our research, one of the motivations for our research is the work being made toward the goal of making believable and immersive conversational avatars.

To examine the social believability of avatars, we conduct a study that evaluates the disparity in Social Fidelity of different avatar control mechanisms using a well-known social VR platform. Prior research focused on various implementations of avatars, virtual agents and robots (Ali et al., 2020; Cassell et al., 1994; Cassell, 2000; Greenwald et al., 2017; Lee & Marsella, 2006; Morie et al., 2012; Yalçın, 2018) and their evaluation in terms of believability (Bevacqua et al., 2017; Gomes et al., 2013; Gonzales-Franco & Peck, 2018). In this research, we specifically evaluate the effect of user expectations on the social fidelity and believability assessment of the avatars, using the concept of the Adaptation Gap (Komatsu et al., 2012), which was not examined before.

Figure 2. A group photo of users using motion-tracked and automated-gesturing avatars in Tivoli Cloud VR (Meeks, 2020)

The initial challenge of our research was to find a suitable virtual world platform for the study. When we started there were many virtual world platforms on the internet, with some focusing on visiting the worlds using Virtual Reality and a select few offered Desktop accessibility without VR hardware. While there are many variants of how a user can interact in virtual worlds, we will mainly be discussing *Desktop* versus *Virtual Reality* as entry points to a social VR platform. With Desktop as an entry point - the user is interacting in a 3D world with others as avatars but viewing it on a 2D monitor and controlling their avatar (typically via automatic methods) via the traditional keyboard, mouse and microphone. When we refer to Virtual Reality as an entry point, a user is immersed in the same 3D world but uses a full VR Head Mounted Display (in Stereo), hand controllers and voice to see the world in stereo as well as control their movements. We also refer to the avatars typically provided by the Virtual Reality as entry point setup as *motion-tracked* and select avatars (depending on the platform) provided by the Desktop as entry point setup as *automated-gesturing*. In order to guide our evaluation, we developed a taxonomy encompassing platform features as well as social and behavioural aspects of the avatars. Our taxonomy is adapted from Tanenbaum et al.'s work (2020) which includes categories such as: Movement & Proxemic Spacing, Facial Control, Gesture & Posture, and Virtual Environment Specific NVC. The categories were then selected and/or expanded to include avatar control features, including: Level of Motion Capture/Tracking, Use of Lip Sync, Use of Automated-Gesturing, Lip Sync Quality, Environmental Interaction in VR, and others. Additionally, we included Gonzalez-Franco & Peck's (2018) Avatar Embodiment categories, considering our focus on avatars. The categories were adapted to our taxonomy includes the following: Body Ownership, Agency and Motor Control, Tactile Sensations, Location of Body, External Appearance, and Response to External Stimuli. Chapter 4 describes this process in detail.

After careful examination of multiple social VR platforms under this taxonomy, Tivoli Cloud VR was chosen as the platform to use for the study for the main reasons being: 1) enabling different methods of avatar control mechanisms, 2) inclusion of automated gesture generation ability when conversing.

Platforms like the aforementioned Tivoli Cloud VR (2020), Vircadia (2023), and Overte (2022) allow users to use two kinds of avatar control mechanisms, namely the motion-tracked avatar and automated-gesturing avatar (see Figure 2). The different avatar control mechanisms allow users to control the facial and bodily gestures of the avatars,

locomotion, and emotional expressions. This enables users to engage in social interactions and collaborations in virtual spaces. The two mechanisms create different user entry point setups for a user visiting a social virtual world: one involves entering with VR technology and so is paired with a motion-tracked avatar, and the other entering without it and so is paired with an automated-gesturing avatar (a Non-VR avatar). The ability to use motion-tracked or automated-gesturing avatars in a shared networked world allows more users to inhabit these worlds but also creates a discrepancy in the behavior and performance of the avatars and can impact user expectations. This impact on expectations in turn can affect the believability and immersion of the avatars and the environment, affecting user experience (Loyall, 1997). In this context, this thesis is aimed to examine whether avatar control mechanisms affect believability by changing user expectations.

1.3. The Study

In order to examine the effect of different avatar control mechanisms on user perception and the believability of the avatars in social VR platforms, we focused on the notion of Social Fidelity gap which refers to the difference between the expected and the perceived social fidelity of an avatar. Social Fidelity gap is a modified version of the concept of Adaptation Gap introduced by Komatsu and colleagues (2012), which is the difference in the expected functionality versus the perceived functionality of an entity. Adaptation gap is found to affect user reception and continued interaction with an entity. As modification, we use Social Fidelity Gap that focuses on social behaviors of the avatars that are related to the concept of social believability. It is from the idea of the difference in social fidelity that we conducted a study to evaluate whether participants can notice a Social Fidelity gap among the different avatar control mechanisms (motion-tracked vs. automated-gesturing) and involves the utilization of the Adaptation Gap concept by Komatsu and colleagues.

The study seeks to answer the following research questions:

- Does an observer in a shared virtual environment notice a Social Fidelity gap among human-driven conversational avatars with different avatar control mechanisms?

- Are the Social Fidelity Gap scores for the conversational avatars correlated by the perceived believability?

In order to answer these questions, we recorded interactions with avatars with two different behaviour control mechanisms (motion-tracked vs. automated-gesturing) in a suitable social VR platform, previously referred to as avatar control type, and compared believability and Social Fidelity scores. Specifically, the study involved the final total of 88 participants watching two videos of the different avatar control types and then answering Likert scale-based questions based on the avatars. Participants answered pre-exposure questions before watching a video of each avatar and then answered post-exposure questions after each video, with no avatar control type or performance being repeated. The responses were analyzed in order to obtain believability scores and Adaptation Gap scores (calculated from Social Fidelity scores, see Section 5.1.3) for each participant. Our results showed that different avatar control methods have a significant difference in terms of Adaptation Gap scores, and this difference also changes depending on the interaction scenarios. The results further showed a significant correlation between believability and Adaptation Gap scores.

1.4. The Contribution

In this research, we created a taxonomy suitable to evaluate the social believability of avatars in social VR platforms by adapting Tanenbaum and colleagues' taxonomy (2020) and Gonzalez-Franco & Peck's (2018) work. Tanenbaum and colleagues' work focuses on nonverbal communication in social VR platforms and Gonzalez-Franco & Peck's (2018) work on Avatar Embodiment categories. To the best of our knowledge, our work is the first to apply these categories in evaluation of VR avatar control mechanisms for social believability. The closest work to resemble ours being Liu & Steed's (2021) work on comparing and evaluating social VR platforms, though our work specifically evaluates automated-gesturing and motion-tracked avatars, and includes Social Believability categories. Our research is also the first example of using the Adaptation Gap theory in the context of avatars in social VR platforms by measuring the Social Fidelity gap between two avatars. In addition, it showcases the importance of standardizing avatar visual presentation, and that social VR platforms need to cater avatar exposure more carefully to visitors in order to increase the visitor's acceptance rate and improve their virtual world experience. This is supplemented with recommendations about exposure order and the

necessity of reduction of the believability gap among different avatar control types. We also intend to provide the research community and practitioners with recommendations for how to bridge the Social Fidelity gap.

1.5. Thesis Structure

This introductory chapter provides an overview of the research and the thesis. It presents the context of the research and some details about social VR platforms and avatars, including the motivation for utilizing the Adaptation Gap theory to measure the Social Fidelity of conversational avatars. Chapter 2 provides some details of related work and theories that have been developed for virtual agents and avatars. Chapter 3 will discuss some differences between the original Adaptation Gap research and our approach to utilizing the Adaptation Gap with conversational avatars. Exploring and selecting the study environment for the research and the various criteria and taxonomies utilized are described in Chapter 4. The study details and their results are described in Chapter 5, with a discussion of the results followed in Chapter 6. Future work for conversational avatars and Social Fidelity based on the results are discussed in Chapter 7. Finally, Chapter 8 concludes the thesis by revisiting the contributions and summarizing the final takeaways.

Chapter 2.

Concepts and Related Work

In Virtual Reality (VR) environments, the users are represented by an “avatar” in which they can engage in social interactions by acting in the immersive VR world (Meadows, 2007). The actions each avatar can perform in the VR environments are often designed to resemble the face-to-face interactions between humans and increase body-ownership illusion where users can control the avatar’s fine-grained behaviours synchronously such as their gestures, posture, or eye gaze (DiPaola & Collins, 1999; Kokkinara & Slater, 2014; Wei et al., 2022). The extent of this control can be closely tied to the actual behaviour of the user through motion tracking or can have different degrees of automation through gesture generation models. Some avatars in VR environments can also be virtual agents, also referred to as Embodied Conversational Agents or ECAs (Cassell, 2000), that can have automated behaviours. Although virtual agents can sometimes be human-controlled in what is called Wizard-of-Oz (WoZ) studies where the users are told they are completely automated.

There are three important concepts that this research utilizes: Embodied Conversational Agents/Avatars, Social Believability, and the Adaptation Gap. *Embodied Conversational Agents* or ECAs are computer-generated virtual agents that visually resemble a human (e.g., using an avatar), are able to communicate both verbally and nonverbally, and can be controlled automatically by computers or humans. ECAs can have automated social behaviors including: animating lip movements, gaze, emotional expressions, face and body gestures, locomotion, and engage in dialogue using speech (Cassell, 2000). *Social Believability* of an avatar or a virtual agent refers to the extent to which they are perceived as socially competent, natural and immersive in interaction contexts (DiPaola & Collins, 1999; Alfonso & Prada, 2009; Nixon, 2009; Li et al., 2014; Hashemain et al., 2018). It can be determined by how immersive and how convincing an entity or virtual human is in social contexts. It is evaluated on two general levels: 1) Social – complexity of personality and social behaviour (Afonso & Prada, 2009); 2) Believability - complexity of behaviour and interaction, and range of movement (Nixon, 2009). The concept of the *Adaptation Gap* was introduced by Komatsu and colleagues (2012) as the difference between a user’s expectations and their actual perceptions on

the functions of an interactive robot. A positive Adaptation Gap is therefore defined as an interactive robot exceeding user's expectation, whereas a negative Adaptation Gap shows an interactive robot not meeting the user's expectations.

This chapter will provide detailed analysis of the various research and theories that fall into the categories of ECAs, Social believability, and the Adaptation Gap theory, which our research adapts and utilizes. It is important to note that VR and social platform development, including research into ECAs and Social Believability, is an ongoing and ever-changing endeavor. By the time this research is published, other research and technologies would have surely come out. In order to avoid an almost infinite loop of constant revisiting of platforms, updating research and study details - a conscious choice was made to commit words to paper. As such, this thesis serves not only as the result of the research and work that went on up to 2023 but also serves as a time capsule of the state of consumer VR, social platforms, avatars and ECAs of that period. Since the research on ECAs and Social Believability is ongoing, each selected work is provided with an overview and our justification for its inclusion in our research.

2.1. Embodied Conversational Agents/Avatars

Embodied Conversational Agents (or ECAs) are digital characters that can engage in interactions that resemble typical human face-to-face conversations through a virtual (e.g., 2D/3D computer models) or physical (e.g., robots) visual representation (Cassel et. al., 2000). ECAs can be used to represent the computer (or the agent) or represent the human users in a computational environment (as avatars). ECAs can emulate verbal and non-verbal multimodal behaviors (e.g., speech, facial displays, hand gestures, locomotion and body stance), to achieve natural conversation with the humans or other agents in its environment. The visual representation and behaviors are aimed to achieve the sense of embodiment in the agent's environment and ECAs can have differing levels of visual and behavioral realism (see Figure 3).

Figure 3. Examples of different avatar visual and behavioural realism. Mozilla Hubs with disembodied avatar (left), Engage VR with fully embodied and face scanned avatar (right)

These differences in both avatar embodiment as well as their behavioral capabilities are known to significantly affect the believability (Poggi et al., 2005), engagement (Loveys et al., 2020), and trustworthiness (Ruttkay & Pelachaud, 2004) of these ECAs. However, human-like avatars tend to increase the expectations on their behaviour and capabilities to be more human-like (Go & Sundar, 2019). Research into ECAs, including the design, implementation, and evaluation of creating the avatars and the techniques of sustaining social believability, are an extensive topic and come from various fields including psychology, linguistics, computer science, arts, and HCI research (Cassell et. al., 2000). Some cover the technical aspects of automating ECA behavior such as text-to-gesture generation systems (Ali et al., 2020), where others cover theory or best practices such as improving the visual believability of video game characters (Afonso & Prada, 2009) or ways to evaluate ECAs including evaluating believability and co-presence (Bevacqua et al., 2017).

Video games also use digital characters controlled by the computer that engage in interactions and conversations, which are called non-player characters (NPCs). This means the research and lessons on ECAs and Social Believability, including those coming from this thesis on conversational agents, are not just applicable to ECAs and social VR platforms but also to video games and game characters. We believe the variety of believability research for games alone (Prada & Paiva, 2005; Zammitto et al., 2008; Alfonso & Prada, 2009; Verhagen et al., 2013; Morgan & Papangelis, 2015) shows that game characters and non-player characters could benefit from applying ECA research

and techniques to enhance their believability. This is especially important when multiplayer VR enabled games are concerned. Especially for games where multiple users using avatars play with or against each other online using VR. Examples include Rec Room (Rec Room Inc., 2016), Star Trek: Bridge Crew (Red Storm Entertainment, 2017) and Space Team VR (Cooperative Innovations, 2020), among others.

The field involves research and development into creating more realistic and believable ECAs including improving aspects regarding their embodiment and conversational capabilities. ECA behaviors can be generated by artificial intelligence models or can be driven by humans who represent themselves in virtual environments. It is essential to clarify that our study does not involve ECAs that are representing computer agents as our avatars are mainly embodied and driven by humans, and have very little automation and behavioural systems behind them to be considered as agents. Instead, we refer to them in our study as *"conversational avatars"* to more accurately reflect the depth of their conversational and behavioural implementations. That being said, our research touches upon important topics concerning ECAs, including automated gesture generation and believability studies. We also believe the results and conclusions from our work on conversational avatars can be transferred to the ECA research. Work in this field includes, but is not limited to, research such as introducing empathy to ECAs (Yalçın, 2019), adding personality via nonverbal behavior (Saberi, et. al., 2021), the creation of ECAs and various approaches/concepts (Cassell, 2000), and evaluating ECA social presence in Virtual Reality (Greenwald et al., 2017), among others. In particular, we will take a look at the closest related research to our study, research on behaviour generation for ECAs, and research on ECA techniques and improvements.

The closest related research project to our research is one by Greenwald et al. (2017) which looks at the topic of Social Presence and Embodied Avatars in VR. The paper served as our starting point on what to look out for and what to keep in mind when thinking about improving the implementation of Non-VR versions of conversational avatars. It investigates the effectiveness of communication and interaction when embodying avatars in Virtual Reality and compares the effectiveness to alternate forms (face-to-face, etc.). The research outcomes show the effectiveness of communication in VR and a strong sense of social presence. Some shortcomings were noticed by the researchers, including limitations of VR and face-to-face interaction and the "limited anthropomorphic" avatars (p. 1). A space that was not explored is Embodied Avatars in

the same virtual environment using VR and Non-VR technology. Our research differs from this paper by introducing the concept of the Adaptation Gap and comparing the use of motion-tracked and automated-gesturing avatars. The topic of conversational avatars and the Adaptation Gap in gaming is discussed further in Section 6.4.1. We believe the Adaptation Gap and Social Fidelity gap research for conversational avatars can help improve immersion, interaction, and communication for players, by assisting in the future development of video game characters.

2.1.1. Behaviour Generation for ECAs

When talking about ECAs and techniques for their creation it is worth mentioning the resource that is most referenced, which is Cassell's book *Embodied Conversational Agents* (2000). In their book, Cassell and colleagues describe a compilation of different specialist approaches to design, implementation, and evaluation of ECAs from many researchers, plus existing systems being used to create ECAs. This is done by utilizing several authors and experts from different fields that contribute their input and research to the ECA topic. This book serves as a good starting point and supportive material for many components of ECA research and can help with research on human-driven ECAs. Due to its age, there are opportunities to build upon what is included in the book with more details. For example: details into the design and implementation of human-driven ECAs, evaluating Social Believability of ECAs, and an updated ECA "Turing" test focusing on believability. What is also worth noting is that the book does not provide a unified approach to creating ECAs nor a unified approach to evaluating them. This can be an opportunity for researchers to create a unified evaluative system for ECAs. What's also worthy of note is Cassell's prior work on improving ECAs (1994) which served as the initial inspiration for our research topic and the idea of automated generation of facial expressions and gestures of ECAs and their potential as improvements to Non-VR based avatar social believability. This serves as good contextual material when dealing with automated-gesturing avatars like those seen in High-Fidelity forks (copies of a code repository) such as Tivoli Cloud VR (Meeks et al., 2020), Vircadia (vircadia.com, 2023), and Overte (overte.org, 2022). The only drawback of the work is that the ECA systems described in the book are not meant for real-time applications, where the previously mentioned platforms all use real-time systems.

ECAs can be driven by either Artificial Intelligence (AI), rule-based systems, or can be driven by humans also acting as agents. Ali et al. (2020) and Lee & Marsella (2006) provide two approaches to generating behaviours for ECAs. Ali et al. in their paper present an approach for automating “rule-based co-speech” gesture generation mapping without human intervention. The automation of gesture and rule generation systems which contribute to creating models for ECAs helps the process of creating believable ECAs, and the models can make the rules generated be contextually believable. What was produced was a trained model generating the rules for text-to-gesture triggers/behaviours that were then evaluated and proven successful. Along the same vein, Lee & Marsella provides support for how important nonverbal communication is and what goes into creating behaviour rules for ECAs. In their work, they describe their framework creation process by first analyzing video data and annotating different nonverbal behaviours, then creating mappings for behaviours and key utterances, after which they moved on to creating behavioural rules using the mappings (Lee & Marsella, 2006, p. 246 - 253). The rules allow to specify which nonverbal behaviours should be generated based on utterance features (syntactic, semantic, affective). Finally, the rules were implemented into a behaviour generator that was then implemented into the SmartBody project from the University of Southern California. Lee & Marsella found that the system successfully generates nonverbal behaviours such as head movements, facial expressions, and body gestures in real-time, improving the interaction between users and ECAs. In addition, the system is user-extensible, allowing users to easily modify or extend behaviour generation rules, which allows for customization of behaviours according to specific user requirements. These works are also useful in other applications. For example, using machine learning and frameworks for text-to-speech behaviour generators applied not only towards computer-driven ECAs but human-driven ones as well.

A more recent example of work towards improving ECAs is Yalçın’s work (2018) which covers some topics about AI-driven ECAs and how they perform and gesture during conversations, but most importantly explores models to support the simulation of empathetic emotions. ECAs allow for “natural and effortless” (p. iv) interactions that encompass similar conversational properties and patterns that we as humans employ. Yalçın’s work expands on prior ECA work by introducing empathy to ECAs and highlights recent research focusing on emulating various emotions/behaviours, with Yalçın focusing on empathy with the goal of enhancing conversational interactions (p. iv). The journal goes

into detail on the modelling and implementation of empathetic models for ECAs but also addresses issues of utilizing complex behaviours in real-time and automated systems. This gives some insights into important conversational gestures and automation that can be used for gesturing during speech and also serves as another inspiration for our research.

More recent work uses generative AI models to learn behaviors from motion-capture or video data including human behavior (see Nyatsanga et al., 2023 for a review). Other AI models also can capture human behavior, including facial gestures, from video, to directly map to avatar expressions with realistic rendering (Zheng et al., 2022). However, these novel systems often require too much computational power for online VR environments to be run in real-time, which makes it unfeasible to use at the time of writing this thesis.

For our research, there's potential to talk about an approach to Social Believability of human-driven avatars to bridge the gap between motion-tracked and automated-gesturing avatars by utilizing models to generate text-to-gesture rules to automate the gesturing of human-driven avatars, such as the Automated-Gesturing Avatars used in Tivoli Cloud VR (discussed more in chapters 6 and 7). In addition, bridging the difference gap can also potentially be done using Lee & Marsella's described framework combined with motion tracking or motion recording. Combining this with Ali et al.'s research, it serves as a good supplemental resource for gesture generation as part of the Automated-Gesturing Avatar implementation of Tivoli, which involves automated gesturing during speech.

2.1.2. ECAs and Virtual Worlds

A paper on ECAs in Virtual Worlds by Morie et al. (2012) offers some limited supportive contributions and evidence on ECA technology and techniques for integrating them into virtual worlds. However, it brings up an important point related to our research which also involves different ECAs inhabiting a shared networked world. Morie et al.'s paper discusses the importance of networked worlds and underlines the need to focus on making ECAs and the environment more responsive. The paper highlights the need being especially important for networked worlds with a particular focus on the performance of multiple agents in said networked worlds. The gap here is further evidence to support the

need to enhance agents or the environment to be more responsive to users. This provides an opportunity to apply the approach to human-driven Conversational Avatars and how to make them more responsive to the environment and each other. We found this to be a good starting point for such research where researchers investigate making ECAs that are driven by humans using various inputs indistinguishable yet equally responsive.

Starting with the research that helped with virtual world platform selection, something that we also discuss in Chapter 4. The researchers, Tanenbaum et al. (2020), provide a breakdown of Non-Verbal Communication (or NVC) of conversational avatars from ten popular social VR platforms with the goal to highlight the most prominent VR features and techniques and underline areas for further improvement (see Table 1 for all categories). Specifically, the research highlights commonly featured NVC behaviours by looking at features such as: proxemic spacing, facial expression control, gesture, posture, gaze fixation, and others (p. 1). This was done by researchers visiting and analyzing the VR experience of their chosen social VR platforms, thus producing a table of NVC functionality in these platforms that allows for comparison. This research provided us with a good cross reference for checking our own selection of social VR platforms but also provided an example of taxonomy or feature breakdown/analysis. Table 1 was used specifically to select the social VR platform to carry out the study in and is discussed in Chapter 4.

Table 1. Tanenbaum et al.'s (2020) NVC High Level Categories with their inner Categories and Sub-Categories

High-Level Category	Category (Sub-Category)
Movement and Proxemic Spacing	Direct Teleportation (Facing Selection, Destination Validation, Vertical Movement)
	Analog Stick Movement (Smooth/Continuous, Snap/Step)
	1:1 Player Movement
	3rd Person Movement
	"Hot Spot" Selection
Facial Control	Expression Preset (Independent/Direct Selection, Dependent/Indirect Selection)
	"Puppeteered" Expressions (Lip Sync)
	Gaze and Eye Fixation (Object Tracking)
Gesture and Posture	Poseable Appendages (Hands/Arms, Head, Torso, Legs)
	Dependent/Indirect Selection
	Mood, Posture, and Status (Dependent/Indirect Selection, Mood Preset)

High-Level Category	Category (Sub-Category)
Virtual Environment Specific NVC	Multi-Avatar Interactions (Consent)
	Collisions (With other Avatars, With Environment)
	Emotes
	POV Shifts

Tanenbaum et al. compiled their taxonomy using four high-level categories: Movement and Proxemic Spacing, Facial Control, Gesture and Posture, and Virtual Environment Specific NVC (2020, p. 5). Three out of the four of these high-level categories were based on key nonverbal communication features appropriated from prior work on VR locomotion, facial expression in VR, and gesturing in VR. The final category is of their own creation based on the unique features of Social VR platforms identified by them. They used the taxonomy to evaluate ten popular social VR platforms by identifying the top utilized features and designs from their categories. They then contributed future opportunities and challenges for the future development of nonverbal communication in VR. Tanenbaum and colleagues also mention "automation" in NVC, specifically concerning the inclusion or improvement of automation towards blinking, looking or eye gazing, posing/posture, lip syncing, and facial expressions for VR avatars only. The work does not include automation for avatars using Desktop as an entry point setup or automation using conversational gestures, which is the focus of our work.

Using Tanenbaum and colleagues' taxonomy as a guide, we used the general high-level categories related to avatar behavior control, including locomotion, facial control, and gesturing from their taxonomy. We then adapted those categories to include the following categories for our own taxonomy and exploration of top social VR platforms: Level of Motion Capture/Tracking (VR only), Level of Desktop Puppeteering, Use of Gestures and/or Emotes, Use of Lip Sync, Use of Automated-Gesturing, Lip Sync Quality, Environmental Interaction in VR, Environmental Interaction on Desktop, Extra Social Features (if any), Coding/Scripting capabilities of the platform, and Navigation Control Scheme used for Desktop and VR. The only high-level category that did not end up being utilized was the fourth category, Virtual Environment Specific NVC, as it was not relevant to our research focus. Instead, we extended our taxonomy with additional categories from Gonzalez-Franco & Peck's (2018), that relates to Avatar Embodiment categories. For more details, see Section 4.1.1.

Tanenbaum et al.'s work and Gonzalez-Franco & Peck's work also appear in Liu & Steed's work on Social VR Platform comparison and evaluation (2021). This work is in some ways similar to our approach in that it aims to utilize evaluative categories to evaluate top Social VR Platforms. The big difference with our work is that our research also evaluates automated-gesturing avatars and not just motion-tracked, and includes categories involving Social Believability. Liu & Steed's work, while influenced by Tanenbaum et al. and Gonzalez-Franco & Peck, utilize their own approach to evaluation while focusing more on the usability of the platforms themselves. They utilize categories in the form of tasks or task groups to evaluate how well platforms support or implement features in these categories, such as: Identification (users identifying objects or people), Communication (verbal and non-verbal communication), Navigation (planning and navigating virtual spaces), Manipulation (interacting with objects), and Coordination (or cooperating with other users).

Since we are covering conversational avatars in social virtual worlds, which includes being able to visit and control avatars using VR technology, it means that one would also be dealing with VR avatars. There are many sources and research on VR avatar representations and technologies, for example: Valkov et al. (2016), Young et al. (2015), Argelaguet et al. (2016), and Aseeri & Interrante (2021). However, we recommend examining Weidner et al.'s (2023) literature review on avatar visualizations for Augmented Reality and Virtual Reality. Their article reviews and compiles various rendering styles and uses of body representations in Augmented Reality and Virtual Reality applications.

2.2. Social Believability

The Social Believability field researches various aspects of avatar or agent Believability that include factors such as Immersion and Social Expression (Afonso & Prada, 2009; Nixon, 2009). Social expression has to do with the range and flexibility of expression with regards to verbal and non-verbal communication in a social context. Immersion is more concerned with features and factors that support a user's suspension of disbelief and involves concepts like the visual look and mechanical behaviour of the entity/agent (Bates, 1994; Yalçın, 2018; Greenwald, 2017). Social Believability assists in the development of more realistic and believable ECAs by improving aspects regarding their embodied, conversational and socio-emotional behaviours and gestures. In addition, the Performance versus Believability trade-off (Morgan & Papangelis, 2015) is also an

important factor, being an area of research that is concerned with both Immersion and Social Expression and their balance when it comes to having a wide array of functionality and a strong focus in only a few areas of believability. Work in this field includes, but is not limited to, research of Social Models for believable video game characters (Afonso & Prada, 2009), aesthetic and expression systems for virtual human behaviour/movement (Nixon, 2009), and how an intelligent agent's performance effects believability and vice versa (Morgan & Papangelis, 2015).

2.2.1. Enhancing Believability of Avatars and Virtual Humans

One of the preliminary umbrella topics that we are interested in that lies parallel to our research is how to improve or enhance ECAs' believability. An example is Nixon's (2009) covered topic on designing virtual humans and principles that allow for those virtual humans to be immersive and socially believable. In addition, other researchers also point out the importance of narrative in character believability (Bizzocchi et al., 2013; Tanenbaum & Bizzocchi, 2009). Virtual humans are digital computer models that act as substitutes that can be used for evaluation or for representation in virtual environments (Nixon, 2009, p. 9). Delsarte's aesthetic system approach delves into the performance and poses of believable characters, which can assist in improving avatars' social believability. Explorations are conducted with animators implementing animations using standard methods and then using Delsarte's Aesthetic System, which is then converted into a model, after which videos are produced of the performances. The videos are then shown to participants who evaluate their personality. The study was done within a limited application of animators animating a character and then recording the output, with results and analysis showing that the model improved "the presentation of personality traits" of the characters (p. 146). We believe additional support can be provided for Delsarte's system for enhancing the believability of characters by further applying and studying the system with conversational avatars being experienced by participants in real time. Nixon's research also provides contextual background and supporting materials when it comes to character movement and expression, which can assist in our research when it comes to evaluating the performance of conversational avatars.

Nixon references an earlier related work by Bates (1994) that takes a similar artistic approach to enhancing believability. It involves making believable characters by focusing on "artistic theory" used by animators and cartoonists like those from Disney animation (p.

1), with a particular focus on the expression of emotion. Bates emphasizes the artistic expertise of “appropriately timed and clearly expressed emotion” as a vital component of making believable characters. The paper describes various properties and approaches to applying and expressing emotions in synthetic characters and how that allows it to communicate the illusion of life. By combining artistic approaches with interactive characters, one can then be able to achieve “believable agents”. This is important research to the field of believable agents as it focuses on emotions specifically, and something we noticed lacking in quite a few virtual world conversational avatars.

Loyall in their PhD dissertation provides a valuable definition of believability: “a character is considered believable if it allows the audience to suspend their disbelief and if it provides a convincing portrayal of the personality they expect or come to expect” (Loyall, 1997, p.1). However, this is not the only contribution from Loyall. The dissertation proposes a set of properties that agents must have in order to be perceived as socially believable. These are: Personality, Emotion, Self-Motivation, Change, Social Relationships, Consistency, and the Illusion of Life. The Illusion of Life is further expanded by Loyall to include: Appearance of Goals, Concurrent pursuit of Goals, Parallel Action, Reactive and Responsive, Situated, Resource Bounded (in body and mind), Exists in a Social Context, Broadly Capable, and Well Integrated (capabilities and behaviours) (p. 15-26). Loyall’s work served as a foundation for believability and can be seen utilized in other works, such as by Gomes and colleagues (2013), where they both provide vital categories and metrics for evaluating character believability. Loyall’s dissertation is additionally divided into four parts, where the work displayed exemplifies an applied attempt towards the creation of believable agents: part 1 analyzes and provides the problems necessary to solve for believable characters, part 2 shows an architecture called “Hap” created to support the need for believable agents, part 3 provides an approach to natural language processing and generation, with the final part showing an agent built utilizing the described frameworks and its demonstration that resulted in it achieving a good level of believability (Loyall, 1997, p. iii).

Musick (2021) ended up being a good supplemental paper when paired with Komatsu et al.’s Adaptation Gap paper (2012). While not using the Adaptation Gap approach, quantitative analysis, or a single agent, it does look at how humans react to teammates (multiple agents) when they perceive them as autonomous AI agents. Like Komatsu et al., they found that human perception of the agents has an effect on how they

respond and treat these agents when working on team-based tasks and challenges. This is important for us as we care about the Social Believability perception of users towards a conversational avatar when exposed to multiple avatars. Factors that Musick et al. found contributing to the negative perception of the agents were related to low reliability, transparency issues, and lack of independence and agency (p. 1), resulting in the researchers concluding that “the perception of team composition did affect sentiments toward teammates, team processes, cognitive states and the emergence of a system of team cognition” (p. 4).

2.2.2. Believability and Games

There is plenty of material covering believability within the context of video games, looking at various aspects from social, personality, and group models to techniques in improving character believability. Afonso and Prada (2009), for one, focus on the topic of the Social Believability of virtual agents, specifically in Role-Playing Games (or RPGs). The goal of their research is to increase the immersion for players by improving the believability of in-game Non-Player characters by introducing a “social relationship” model. This social model allows virtual agents to exemplify “social deepness” (p. 35), which involves binding agents together in an awareness and social-based relationship. The work provides an interesting approach to believability with the introduction of the social model and utilizes the context of video game characters. While not necessarily what our research aims to study, it can certainly be useful for future research and directions for believability improvements.

Similarly to Alfonso and Prada (2009), Morgan and Papangelis (2015) in their paper focus on a limited-scope agent in a video game, but the approach and analysis detailed are important as it talks about concepts such as the performance and believability of an agent and how too much or too little of one affects the other. Results were in the form of statistics and qualitative interview data from participants watching videos of recorded gameplay and showed that performance does indeed impact the believability of an agent. This is related to our research as the goal is to have conversational avatars, driven through different inputs, perform in a way that is believable and indistinguishable. This can expand and fill the gap in research into interactive/social agents, with an opportunity to test the performance versus believability of conversational avatars, both computer and human-driven, with a focus on social believability.

In one excerpt from *Advances in Computer Entertainment* (2013), Verhagen and colleagues introduce the topic of Social Believability in video games in the form of a workshop in order to capture a wide net of expertise and ideas to contribute to improving the social believability of characters in video games. The end goal of the workshop was to compile different approaches to modelling and implementing intelligent behaviours with the inclusion of emotional and social behaviours. Most importantly this piece contributes a compilation of various believability models and constructs. Some of the entries our research is interested in are ones such as Bates' breakdown of believable characters as those that are "lifelike, whose actions make sense, who allows you to suspend disbelief" (1994). In addition, a set of requirements for good believability from the Oz group of Carnegie Mellon University (1999) which includes: personality, emotion, self-motivation, change, social relationships, consistency of expression, illusion of life, and well-integrated capabilities and behaviours. This resource served as a good starting point for our research as working with social VR platforms, as one could argue they have many similarities to video games in their implementation and use many video game technologies and techniques. The requirements for good believability especially serve as a good list to contextualize the believability of conversational agents and can also serve as a resource for future improvement of their believability.

In another paper on believable synthetic characters (Prada & Paiva, 2005), the researchers aimed to develop a model that supported group dynamics of synthetic agents in order to increase their believability. The developed model involves group dynamics based on similar theories in human social psychology, with the model being implemented and then evaluated on autonomous agents in a video game. The authors describe the model as being based on principles of group and user awareness of members to the point of supporting efficient building and reasoning of the social model within the group (p. 38). Most importantly they contribute to the topic of social believability by expanding the idea by explaining that agents' believability depends on the depth of actions and interactions, expressions, and "on how well they lead the user to the suspension of disbelief" (p. 1). The study done with the SGD model involves one user and a few limited agents, so there is an opportunity for a wider application done through tests on multiple users with a group of SGDs or ECAs, and observing if believability is or is not maintained.

Finally, Zammitto et al.'s 2008 paper focuses on improving player engagement with video game non-player characters by improving their natural behaviour and

believability through the introduction of personality modelling. The research gives an overview of a multidimensional hierarchical personality model that allows supporting of character systems, and showcases an implementation with a facial character system using XML for scripting character behaviour. The system is called the iFACE system and is divided into 3 parallel parts (Knowledge, Personality, Mood) plus the renderer or "geometry" part for displaying the faces (pg. 2). This research introduces an interesting gap in our exploration of character believability and implementation that is not always considered in today's social platforms and conversational avatars, and that being the use of personality and moods to influence expressions. This Influence on expressions is an important aspect for our research to keep an eye out for when looking at and evaluating conversational avatars of social VR platforms. It serves as another avenue for discussion on improving conversational avatar believability for social VR platforms, something that we touch upon in Chapter 7.

2.2.3. Evaluating Believability

One of the challenges of our research was creating survey questions that would allow us to evaluate the believability of the conversational avatars. To do so, we utilized a number of research papers and questionnaires developed by leading researchers in believability and ECA research.

One major resource was Gomes et al.'s (2013) paper which discusses the concept of "believability" in virtual characters in interactive narratives. They propose metrics for evaluating the perceived believability of virtual characters, which include the following dimensions: behavior coherence, change with experience, awareness, behavior understandability, personality, emotional expressiveness, social, visual impact, and predictability. Since our research and study involve the evaluation of an avatar's social believability, this research proved valuable in providing categories from the aforementioned dimensions and was adapted into our study as evaluative metrics, which then assisted in the creation of specific survey questions. Our study ended up utilizing the following metrics from the research: awareness, behaviour understandability, personality, visual impact, predictability, and behaviour coherence (see Section 5.1.3 for details).

Next, we utilize avatar embodiment, which is related to believability as the better a user's avatar embodiment is - the easier it is for the user to suspend their disbelief of the

avatar body being their body. Gonzalez-Franco and Peck's (2018) research proposes a standardized questionnaire for measuring a user's sense of embodiment when using VR avatars. Embodiment in VR refers to the extent to which a user feels that their virtual body matches their real body. This sense of embodiment can be influenced by factors such as the visual appearance and movement of the avatar, as well as the level of control the user has over the virtual body. The proposed questionnaire consists of questions designed to assess various aspects of embodiment, such as the sense of ownership over the virtual body, the level of immersion in the virtual environment, and the emotional response to the avatar's appearance and movements. The aforementioned questionnaires and the categories are useful for our research as a basis for evaluating the level of embodiment of an observed entity. Our research makes the assumption that the same questions that are used to evaluate one's own embodiment can be used to a certain extent to also evaluate the level of embodiment of an observed user, specifically evaluating how well an entity fits or is integrated into their environment and character. This served as our basis for evaluating various virtual world platforms in order to select one for our study.

In conjunction with using Gonzalez-Franco and Peck's work for creating the survey questions and evaluative categories for our conversational avatars, Bevacqua et al.'s (2017) work was also used and adapted to the study survey. Specifically, the evaluative categories: avatar control, realism and behavior believability. These were adapted to our work by utilizing the questions associated with those categories in their own survey, and deriving our own survey questions using their questions as examples and guides. Table 7 shows the Believability questionnaire categories used in our study, where "Control" and "Believability" categories are the resulting questions and categories inspired by the three Bevacqua et al.'s categories. While the former mentioned work focuses on embodiment, the latter focuses on agent interaction and co-presence. The goal of their work was to provide an evaluation for the believability and co-presence model of an agent interacting with a human. Interestingly, Bevacqua et al.'s conclusions and discussions on the model also see that agents and their role/behaviour have "an important impact on the human perception of the agent itself" (p. 1), on similar veins as Komatsu and colleagues (2012) saying that a user's expectation of an agent impacts user's reception to that agent (see Section 2.3).

When dealing with the input and performance of ECAs, we had to deal with concepts that involve the transfer of movement, behaviour, and personality from the user

to the avatar. One such avenue was the previously mentioned Gonzalez-Franco and Peck's (2018) work that looks at how closely the user embodies themselves with their virtual avatar. Exploring further leads us to the case study by DiPaola and Turner (2008) that looks at the Traveler virtual world and community and how users interact with each other in that world. Of particular focus is the use of oral transmission and intimate modes of communication that are utilized in combination with avatars that are used to express oneself. The paper covers the Traveler platform itself, a user named "Purple Tears", and their "Uninvited" virtual space. An important aspect that our research found useful is the term "binding the pair", which they describe as "the unification of the remote user and the corresponding avatar in the mind of the local viewer" (p. 5). "Binding the pair" is an important immersion factor. It affords the user's ability to not only comfortably and accurately express themselves but also supports the observer's immersion of experiencing another user as close to their real personality or persona as possible.

2.3. Adaptation Gap

The Adaptation Gap is a concept introduced by Komatsu and colleagues (2012) where they focus on user perception of interactive robots. They define a term called the "Adaptation Gap", which is the difference between a user's expectations regarding the functions of an entity and the function that they actually perceive. Positive Adaptation Gaps are entities exceeding the expectations of the users. Negative Adaptation Gaps are entities defying the expectations of the users. The goal was to define and investigate how the "Adaptation Gap" affects the acceptance rate of users toward a robotic agent. The idea is that agents invoke an initial expectation from the user and the agents can either exceed or defy such expectations during an interaction. It is the shifting of expectations that can either positively or negatively affect a user's perception and thus their behaviour towards the agent. To acquire the Adaptation Gap value (AG) one would need to administer pre- and post-exposure questionnaires assessing expectations and/or perceptions of the participants. Using an example of participants answering an appropriate questionnaire in Likert scale: one first totals the values from the pre-exposure test (F_{before}) and the post-exposure test (F_{after}), then subtracts the value from the post-exposure test total with the value from the pre-exposure test total. Komatsu et al. denote this as: $AG = F_{after} - F_{before}$ (p. 5). In return, one gets an Adaptation Gap value that is either positive or negative. Users

with positive Adaptation Gap values will have higher acceptance rates than those with negative values. Figure 4 shows a diagram that details this calculation.

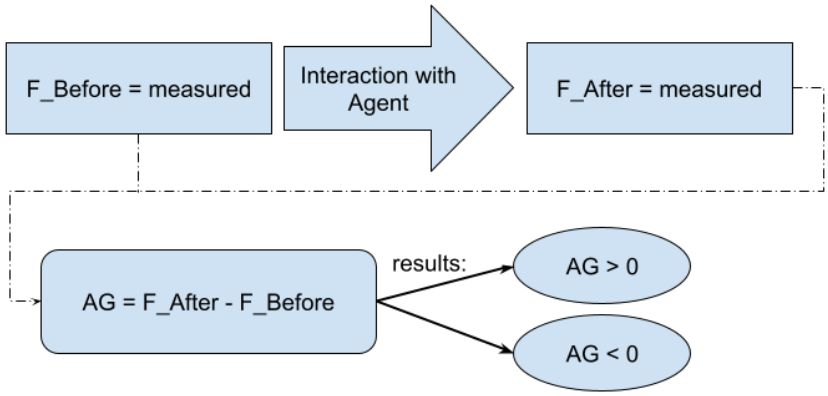

Figure 4. Diagram showing relationship between the Adaptation Gap and Expectation/Fidelity variables. Diagram modified from Komatsu et al. (2012).

Komatsu & Yamada defined the Adaptation Gap concept in 2010 where they started with a theory and then tested the effect of the Adaptation Gap on user impressions of a robotic agent. In that paper they found that the positive or negative signs of the Adaptation Gap plus the subjective impressions of the user before exposure (their study utilized 2 different-looking robotic agents) resulted in significant differences in the user's final impressions of the agents (Komatsu & Yamada, 2010, p. 1). Their study, which they clarified in their revised 2011 paper, included both Adaptation Gap signs (independent variable) and participant subjective impressions (dependent variable). Subjective impressions ended up being different due to the visual difference of the robotic agent used in the study. They go on to say that the use of subjective impressions, due to providing visually different robotic agents, complicated the study and so they aimed to simplify the study to understand the Adaptation Gap effect better. In a follow-up study (Komatsu et al., 2012), they built upon their Adaptation Gap work by studying the effect of the Adaptation Gap but now using only one robot and looking at expectations versus the perceived function of an agent. Their results showed that a positive Adaptation Gap produced a higher acceptance rate for users than a negative one. Ultimately, this allowed them to validate their theory about the Adaptation Gap, which led to their conclusion that "agents that evoke higher expectations than their actual functions should not be used for interaction with users" (Komatsu et al., 2012, p. 114).

Adaptation Gap is one of our main chosen concepts for our research and is deemed important as it can be directly used when evaluating agents. A limitation that was noticed from the prior work is that they did not study ECAs or avatars, only robotic agents, and did not study the visual look or social behaviors of the agent and how it plays a role in the adaptation gap. This is of particular use to ECAs as believable agents strive for believability and should not produce a negative adaptation gap. An important point the paper states that also lies as the basis of our research is that "agents that evoke higher expectations than their actual functions should not be used for interaction with users" (Komatsu et al., 2012, p. 114). We discuss more about the differences between Komatsu et al.'s work and our research in Chapter 3. Later in Chapter 5, we describe how the social fidelity score is used in calculating the Adaptation Gap for conversational avatars.

Chapter 3.

Difference Between the Adaptation Gap and the Current Study

In this chapter, we will briefly discuss the main differences between Komatsu and colleagues' (2012) research on the Adaptation Gap and our approach to using this concept for our research.

In Komatsu and colleagues' research, the main focus was on evaluating robotic agents' expected and perceived practical functionality. Their study used a Lego Mindstorm[1] robot that could correctly guess where a treasure was hidden in a game the user was playing. Users were expected to evaluate the robot after observing their behavior in the game. The evaluation questions they chose mainly focused on task completion, performance, popularity and commercial preference. Our research, inspired by this approach, utilizes the Adaptation Gap concept beyond robotic agents and task completion - focusing more on social and behavioural aspects. Embodied Conversational Agents (ECAs) and avatars have some overlap with the robotic agents described and used by Komatsu et al., but with a key difference being their use in social interaction settings. This includes the fact that agents are not necessarily driven by AI and can be human-driven as well. By human-driven we include the idea of user's intentionally driving agents or avatars through various inputs, however, we do not include the idea of a human driving and playing the role of a robotic or AI agent behind the scenes in a study with a "Wizard of Oz" setup. See for example, Komatsu et al.'s Lego Mindstorm robot (2012, p. 3) or Chaves & Gerosa's chatbots (2018, p. 1). While our research studies human driven avatar systems, it is also applicable to other avatar systems (e.g., Player characters in VR enabled multiplayer games) and AI-driven agents (e.g., ECAs and NPCs). We believe the Adaptation Gap concept is applicable to ECAs and Conversational avatars in accordance with the idea that if a robotic agent can have its expected and perceived functionality evaluated, with negative adaptation gaps signifying agents that should not be used for interaction with users (Komatsu et al., 2012, p. 6), then similarly an ECAs' expected and perceived believability can also be evaluated, with their negative adaptation gaps also

[1] https://www.lego.com/en-ca/themes/mindstorms/about

signifying agents that are not believable and so should not be used for interactions with users.

Another aspect Komatsu et al.'s research did not focus on more was the visual look of the agent and how it plays a role in the Adaptation Gap. The Lego Mindstorm bot was assembled to be serviceable, and the visual aspects of the robot were not part of the evaluation and the questions asked by the researchers. We saw an opportunity to research human-driven agents and include aspects of the visual look of the agent in addition to focusing on conversational and behavioural properties of the agent rather than practicality or functionality of the agent. Hence our research investigates how a conversational avatar's Social Fidelity and Social Believability in terms of the visual looks, interactions, and behavioural capabilities affect the Adaptation Gap for avatars on user reception to the avatars.

Social Fidelity, in the context of Conversational Avatars, is the level of depth and complexity of an entity's social-behavioural properties and features, including the extent to which they allow a virtual character to emulate social features of real-world interactions (Sinatra et al., 2021, p.3; Alexander et al., 2005, p. 4; Garau, 2003, p. 34). Social Fidelity supports not only the quality of the interaction of a conversational avatar but also factors such as an avatar's believability, a user's immersion and their suspension of disbelief during prolonged interactions. Simply, the better an entity's social fidelity is - the more socially believable it is.

Social Fidelity has two components or categories: Physical and Functional. As described by Alexander et al. (2005) and further elaborated by Sinatra et al. (2021): Functional social fidelity is concerned with how the interaction and the content of the interaction match real-world interactions, reinforced through the depth and complexity of those interactions and behaviours. In the context of Conversational Avatars, this can be thought of as the functional behavioural features of the avatar (being able to smile, gesture, etc.). Physical social fidelity is concerned with auditory and methods of communication, and how those are similar to real-world interactions. In the context of Conversational Avatars, this can be seen as the technical implementation of how the user receives or perceives the avatar's interaction. An example would be how accurate the speech synthesis of a conversational avatar is to how a real person would speak. Garau (2003) provides similar categories for "minimal fidelity" for avatars: Visual Fidelity (related

to appearance) and Behavioural Fidelity (related to behaviour and responsiveness). Our research primarily focuses on the Functional social fidelity of our chosen Conversational Avatars.

All in all, our research utilizes the Adaptation Gap concept and adapts it into a Social Believability Gap for evaluating the Social Believability of multiple ECAs by focusing on their social fidelity. But before one can evaluate the Adaptation Gap and Social Believability of a conversational avatar, one would need to have an avatar and an environment for them to be in so that one can capture and study their behaviour in that environment. This proved to be a challenge as we needed to pick one from a list of many social VR platforms and the chosen platform needed to allow users to control motion-tracked avatars using VR hardware and automated-gesturing avatars without using VR hardware. Addressing these challenges and the selection of the study environment is discussed next in Chapter 4.

Chapter 4.

Social VR Platform Taxonomy and the Study Environment

In this chapter, we discuss our approach to selecting a social VR platform for our study environment, our criteria for that selection using suitable taxonomy, analysis of a variety of platforms using that taxonomy and the final platform that was selected. The selected study environment was used in the process of recording videos of avatar performances in that environment. Before we begin, however, we would like to explain some key points to contextualize our search for a suitable social VR platform for our study.

Our main focus for the research is the social believability of human-driven avatars in social VR platforms. There can be many factors that might affect the believability of avatars in online social VR platforms, similar to non-VR counterparts, such as: responsiveness (Morie et al., 2012), task performance (Morgan & Papangelis, 2015) as well as behavioural control (e.g., control through keyboard versus VR controllers). Currently, there are two popular approaches to avatar and behaviour control taken in social VR systems (e.g., Meeks et al.'s Tivoli Cloud VR, 2020; Vircadia, 2023), which can be grouped as the "motion-tracking" and "automated-gesturing" avatar control approaches. Motion-Tracked Avatar control uses motion tracking to detect and map human behaviour into the avatar in real-time, including gestures, body movement as well as voice input. It requires the user to be using VR equipment, often standing while being fully immersed in the environment. Automated-Gesturing Avatar controls, on the other hand, allow users to control the avatar in front of a computer without the need for VR technology. The gestures of the avatar can be rule-based, keyboard-controlled or automated in some cases. Apart from the differences in the quality of gestures each technique provides, there are also usability, performance, and convenience factors that might affect the use of one versus the other (Tanenbaum et al., 2020). The implementations also have technical and optimization reasons that fall under believability versus performance tradeoff (Morgan & Papangelis, 2015) when such systems are used for real time 3D interaction and motion over a network. This means developers must keep in mind the limits of technology and data transfer when striving for a dynamic and immersive experience at a high frame rate. However, to the best of our knowledge, the

potential effect of different avatar controls on believability has not been evaluated in social VR platforms.

4.1. Automated Gesture Generation Systems in Social VR Platforms

One of the challenges of this research was finding a suitable virtual world platform to use for the study that can accommodate the various methods of avatar control in real-time. During our research exploration period between 2021-2022, there were many 3D social immersive platforms on the internet, with some focusing on visiting the worlds in Virtual Reality, and a select few of those that allowed users the option to visit the virtual world platform on Desktop without VR hardware.

Since our main research focus is on the social fidelity and believability of both motion-tracked and automated-gesturing avatars, we needed criteria to help us select the most suitable platform. This brings us to the *first main criterion*: the social VR platform must be able to provide two or more entry point setups that have the potential of providing automated-gesturing and motion-tracked solutions for using and visiting their platform. Meaning users can either enter the virtual worlds while using Virtual Reality hardware (e.g., using an HTC Vive headset and controller) or can enter with just their keyboard, a mouse and a flat monitor.

Social fidelity and believability being our main area of interest for the avatars means our *second main criterion* is that the users using an avatar without VR hardware need to have some social/behavioural features already in place for them to engage in social interaction. This is chosen to exclude some exceptions where a platform can have a Desktop as an entry point setup that provides an avatar version where the avatar is considered “static”, meaning it has minimal or almost non-existent behavioural and conversational features needed for social believability. One can imagine such avatars as those that stand still during interactions or conversations. Examples of some of the important conversational features include but are not limited to: gestures, emotes, eye gaze, lip sync, and environmental interaction, that are considered as essential for believable social interactions with ECAs (Cassell, 1994; Gonzalez-Franco & Peck, 2018; Tanenbaum et al., 2020; Loyall, 1997).

4.1.1. Method of Selection

A *3 step* approach was used to determine the suitability of a virtual platform for our study, which include:

1. Does the platform support Desktop and VR entry point setups with automated-gesturing and motion-tracked 3D avatar versions;
2. Levels of Embodiment for the avatars; and
3. Other Believability Features.

The first step involves fulfilling the requirement of having Desktop and VR entry point setups for the user on the platform that provides automated-gesturing and motion-tracked avatar behaviour controls. To determine the suitability of the avatars, we included additional criteria for our *second step*, which focuses on the level of 3D embodiment. In this step, Gonzalez-Franco & Peck's (2018) Avatar Embodiment categories were adapted to our platform selection criteria, which include: Body Ownership, Agency and Motor Control, Tactile Sensations, Location of Body, External Appearance, and Response to External Stimuli. The goal was to select the platforms that had the highest levels of embodiment for all the avatars provided. *The third step* focused on Believability features that our research was interested in, with some expanding the categories from steps 1 and 2. These extra categories, partially inspired by Tanenbaum et al.'s work (2020), were created with the intention of focusing on specific social platform features and also served as our "wish list" features that depend on their depth of implementation on the platform. The categories include: Level of Motion Capture/Tracking (VR only), Level of Desktop Puppeteering, Use of Gestures and/or Emotes, Use of Lip Sync, Use of Automated-Gesturing, Lip Sync Quality, Environmental Interaction in VR, Environmental Interaction on Desktop, Extra Social Features (if any), Coding/Scripting capabilities of the platform, and Navigation Control Scheme used for Desktop and VR. All the above categories were organized into a taxonomy document that was filled out by the researcher during their investigation into the social VR platforms and can be seen in Tables 3 - 5, Table 2 shows the taxonomy categories and their explanations. The notation "(Desktop/VR)" shows categories that apply to both Desktop and VR versions of the platform access.

Table 2. Categories of our Taxonomy and their explanations as seen in Tables 3 - 5. Categories include citations of the work where the concepts were taken from.

Category	Explanation
VR Support	Does the platform support visits using VR technology?
Desktop Support	Does the platform support visits without VR technology using standard Desktop technology with a mouse and keyboard?
Level of Embodiment (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	How well is the avatar represented as a human body being embodied by the user? Can range from disembodied with head and hands to fully embodied with full body.
Body Ownership (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	Subcategory of Level of Embodiment. How well does one feel they "own" the avatar body.
Agency and Motor Control (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	Subcategory of Level of Embodiment. How much agency and control does one have over the avatar body behaviour.
Tactile Sensations (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	Subcategory of Level of Embodiment. Any tactile or vibration feedback from interactions.
Location of Body (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	Subcategory of Level of Embodiment. The location of the avatar body in relation to one's real body location.
External Appearance (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	Subcategory of Level of Embodiment. The appearance of the avatar.
Response to External Stimuli (Desktop/VR) (Gonzalez-Franco & Peck, 2018, p. 3)	Subcategory of Level of Embodiment. Avatar's response to external influence/forces. (e.g., can the avatar body be pushed?)
Level of Motion Capture/Tracking (VR) (Tanenbaum et al., 2020)	How well is the movement of the user captured and then represented by the VR avatar controlled with VR hardware.
Navigation (Desktop/VR) (Tanenbaum et al., 2020)	What navigation or control scheme is used to move around.
Gestures/Emotes (Desktop/VR) (Tanenbaum et al., 2020)	Are there any systems to trigger gestures and emotes for Desktop avatars. Motion-tracked VR avatars could potentially have a system for triggering gestures and emotes but usually some can be performed by the user through motion tracking.
Lip Sync (Desktop/VR) (Tanenbaum et al., 2020)	Are there any systems for controlling and showing lip sync for VR hardware controlled and non-VR hardware controlled, motion-tracked VR and desktop avatars during speech.
Lip Sync Quality (Desktop/VR) (Tanenbaum et al., 2020)	The quality of the resulting lip sync ranging from speaking/not speaking movement to volume, intonation and phoneme.
Automated-Gesturing (Desktop)	Is there any system for triggering and controlling gestures automatically during speech.
Desktop Puppeteering (Desktop)	Is there any system to allow the puppeteering of Desktop avatars without VR hardware that allows us to mimic the movements or behaviour of motion-tracked VR avatars.
Environmental Interaction (Desktop/VR)	Can the avatar interact with the environment and how is it done.
Extra Social Features	Are there any social features that make the platform stand out from the rest.

Category	Explanation
Coding/Scripting	Any coding or scripting features to allow for custom features and logic to be added by the user.
Notes	Additional notes or comments from observations

4.1.2. Choosing a Platform

Figure 5. Screenshots from visits to the social VR platforms. VR Chat (left), Engage VR (right).

With the 3 step approach and taxonomy categories selected as described above, we moved to visiting and logging the details and features of six popular VR social platforms during the summer period of 2021. The explored platforms are: Neos VR (Solirax, 2018), VR Chat (2012), Engage VR (Immersive VR Education Ltd., 2016), Mozilla Hubs (Mozilla, 2018), Alt Space VR (Microsoft, 2015), and Tivoli Cloud VR (Meeks et al., 2020). Figure 5 and Figure 6 show screenshots documenting the visit to each platform. Each platform has a different level of avatar appearance and behavioral realism, as well as differences in the levels of control of their social behaviors. During our investigations of these platforms, we rated the capabilities and qualities of each platform and their avatar control mechanisms in terms of the selected categories.

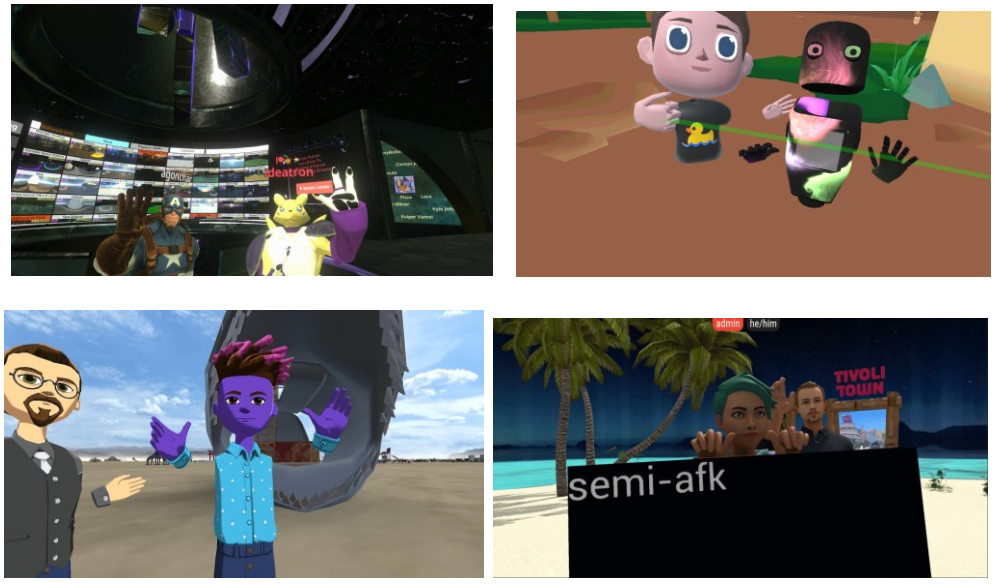

Figure 6. **Additional screenshots from visits to the social VR platforms. From left to right, top to bottom: Neos VR, Mozilla Hubs, Alt Space VR, Tivoli Cloud VR.**

The breakdown of the outcome of the visit and notes on the observed features and behaviours can be viewed in Tables 3 – 5. These tables serve as a taxonomy of the most prominent social VR platforms (mentioned above) and their features and performance observed during our visits. The tables were then used to assist in selecting a suitable social VR platform for use with the research study. Table 2 found previously provides some brief further details of the categories that are seen in Tables 3 - 5, where "VR" represents outcomes from visits using VR hardware and a motion-tracked Avatar, and "PC" represents outcomes from visits done without VR hardware. All final platforms selected had both VR and Desktop support (see Table 3).

Table 3. **Social VR Platform Taxonomy. Part 1: VR and Desktop Support.**

PLATFORM/ CATEGORIES	Neos VR	VR Chat	Engage VR[2]	Mozilla Hubs	Alt Space	Tivoli Cloud VR
VR Support	Yes	Yes	Yes	Yes	Yes	Yes
Desktop Support	Yes	Yes	Yes	Yes	Yes	Yes

[2] Some Engage VR content is locked behind a paid Plus account, which can be considered as a limitation.

Table 4. Social VR Platform Taxonomy. Part 2: Gonzalez-Franco & Peck's (2018) Avatar Embodiment categories.

CATEGORIES		PLATFORM					
		Neos VR	**VR Chat**	**Engage VR**	**Mozilla Hubs**	**Alt Space**	**Tivoli Cloud VR**
Level of Embodiment (Parent category)	PC	Moderate	Moderate	Weak	Weak-Moderate	Weak-Moderate	Moderate
	VR	Strong	Strong	Moderate	Weak-Moderate	Moderate	Strong
Body Ownership	PC	Moderate	Moderate	Weak	Weak	Moderate	Moderate-Strong
	VR	Strong	Strong	Strong	Moderate	Moderate-Strong	Strong
Agency & Motor Control	PC	Weak	Moderate	Moderate	Moderate	Moderate	Moderate
	VR	Strong	Strong	Moderate	Moderate	Moderate	Strong
Tactile Sensations	PC	-	-	-	-	-	-
	VR	Yes	Yes	-	-	Yes	Yes
Location of Body	PC	Strong	Strong	-	-	Moderate	Strong
	VR	Strong	Moderate	Moderate	Weak	Moderate	Strong
External Appearance	PC & VR	Drastic	Drastic	Realistic	Simplified	Simplified	Drastic & Realistic
Response to External Stimuli	PC & VR	Safe or Drastic	Safe or Drastic	Safe & Realistic	Weak	Weak	Safe & Realistic

When discussing the ***Level of Embodiment*** for avatars with VR and Desktop entry point setup, we evaluated how well is the avatar represented as a human body being embodied by the user. This category encompasses multiple sub-categories according to Gonzalez-Franco & Peck's work (2018, p. 3), including body ownership, agency & motor control, tactile sensations, location of body, external appearance, and response to external stimuli. What we have observed from each platform, excluding Tivoli Cloud VR, is that platforms have a variety of embodiment levels for their avatars in these sub-categories. To be more specific, starting with embodiment levels with Neos VR, this platform has a strong level of embodiment for the VR setup (e.g., strong body ownership and agency where the system provides intuitive controls and accurate positions for the arms and body), but the avatar provided by the Desktop entry point setup was emulating VR inputs and interactions - where the controls and movement mapping was not intuitive and required getting used to. Specifically, the user could control gestures, object interactions in the world, UI interactions, and hands at rare points using the mouse and the movement keys with some difficulty, requiring the user time to get used to the controls. VR Chat exhibited a strong level of embodiment for VR (e.g., strong body ownership where the system provided good and easy to learn controls, tactile feedback, and accurate positions for the arms and body), where the Desktop version could not have control over hands or could not see them. Engage VR has a weak sense of embodiment compared to previous platforms (e.g., limited interactivity and certain actions, like sitting, disembodies the user by removing the camera or user's view from your body making it so that the user's real body is no longer aligned with their avatar body). Mozilla Hubs has the weakest sense of embodiment out of all platforms, where only the user's hand location is tracked and represented in the real world accurately to other users, meaning the user experiences a disembodied experience (floating hands, head, and body). Alt Space VR has a weak sense of embodiment (e.g., disembodied where the body is represented as a floating hand and torso with no legs, but has some tactile feedback and location of body parts is more accurate), though not as weak as Mozilla Hubs.

Going through specific sub-categories, ***Body Ownership*** is concerned with how well one feels they "own" the avatar body. We see the Desktop and VR versions of all platforms having different levels of body ownership. For example, in VR Chat's Desktop control, user can see body and hands but can't use their hands; where the VR version allows for hand control where the user can feel like their body is represented in the environment. Mozilla Hubs provides a weak level of body ownership in both Desktop entry

points, mostly due to missing representation of the whole body and hands, and in the VR version where it only shows hands from the controlling user's point of view. Engage VR, similarly does not provide any body representation in the Desktop version and hence a "weak" rating for body ownership, whereas the VR entry point allows for the user to see their full body when looking down and realistically represents their full arm movement (including elbow bending) when moving their hands. Alt Space provides a representation of body and hands, however, the Desktop version cannot utilize hand movements, where the VR version can allow for hand movement control. However, both Desktop and VR versions of Alt Space has no visual representation of the lower half of the body and therefore not fully embodied. Similarly, in Neos VR Desktop control, even though body is represented realistically, the lack of motor control gives a moderate body ownership, whereas the VR version provides a stronger sense of ownership. Finally, Tivoli Cloud has full body representation in both Desktop and VR versions, however, Desktop version body control can be unnatural as user doesn't have control of automated gesturing functionality and can't do some actions such as crouching.

Location of Body, similarly, could differ between Desktop and VR versions of the platforms. For example, there are no avatar bodies in first person in the Desktop versions of Mozilla Hubs and Engage VR, where users can only see their floating hands and their avatars are only visible by others. The VR version of Mozilla Hubs shows only hands through UI interaction where the avatar is a floating character (see Figure 6 top right), hence it received a "weak" rating. In Engage VR, VR version, avatar is forced to assume the sitting position when attempting to sit on a chair or on the floor but the camera of the user does not follow the character's movement during this time – creating a moment of disembodiment for the user. Both VR Chat and Alt Space VR provides a moderate level of fit for the location of body. For VR Chat, this is due to their wide selection of various avatar bodies, including drastic and fantasy ones, which may lead to many avatar bodies not matching the proportions of the user. This leads to, for example, their arms, legs, height positions, among others – not matching the proportions of their user. The experience can also differ dependent on the chosen type of avatar, where body and hands can be seen but there are no feet and therefore no full embodiment. Alt Space VR provides body and arms in correct locations but have a disembodied experience due to the body, arms, and head not being connected, in addition to the absence of legs. In rare instances, during movement, the torso model and some hand models do not always match the body and hand position of the user exactly. Finally, both Tivoli Cloud VR and Neos VR have a

strong fit for the location of body compared to user body, where they can see their full body including all body parts, with accurate tracking during movement.

In ***Agency & Motor Control***, this focuses on how much control the users have over the movements of the avatar body. Motor control in Desktop access was moderate in most platforms, while controlled with a keyboard and mouse may not be as intuitive, especially for first time users. Neos VR has no hand or body control in the Desktop version, hence receiving a "weak" score, except for some hand gestures that can happen through the UI menu. Interactivity and actions in Engage VR is very limited in both Desktop and VR versions, where walking action can end up unrealistic while avatars are floating in the environment with very limited gesturing and emote capability. Tivoli Cloud Desktop version additionally allows for strafing (sliding side to side in one direction to allow player to move with increasing speed), although it is debatable how much this action increases embodiment or realism in avatars. VR versions of Engage VR, Mozilla Hubs and Alt Space provides similar amount of motor control with their Desktop counterparts. However, VR controls of Neos VR, Tivoli Cloud VR and VR Chat allows for more gesturing and ease of motor control as it is more closely mapped with user body.

In terms of ***Tactile Sensations***, none of the chosen platforms provide tactile feedback when accessed via the Desktop entry point. While accessed with VR equipment, Neos VR and VR Chat provides hand controller vibration, albeit very rare and situational. Alt Space, similarly, provides hand controller vibration that are situational during interactions with environmental items. Tivoli Cloud VR allows for haptic feedback with hand controller implemented through certain scripts if written by developers. This can allow for extended applications and interaction; however, the functionality relies on the ability for the developer to include these scripts and are not automated. Engage VR and Mozilla Hubs VR do not have tactile feedback.

External Appearance of Neos VR and VR Chat were drastic, meaning, they involved exaggerated characters (see Figure 6 top left for Neos VR and Figure 5 left for VR chat), that are not necessarily realistic looking. Mozilla Hubs and Alt Space involved simplified cartoonish characters (see Figure 6 top right for Mozilla Hubs and bottom left for Alt Space characters), that had floating hands that are not attached to the avatar bodies. Mozilla Hubs also did not have legs attached, resulting in Weak "Location of Body" compared to other platforms. Tivoli Cloud can provide both realistic and drastic looking avatars, depending on the environment and selection of the user. Some environments in

Tivoli Cloud can force a particular avatar aesthetic on the user (e.g., user can use only cartoon avatars, else one is assigned by default) via an environmental policy, with their original avatar being restored when the user leaves that particular environment.

Response to External Stimuli involves avatar's response to external influence/forces, whether avatars can be affected from external objects or can interact with them and in what manner. Both Mozilla Hubs and Alt Space platforms have weak responses, as avatars lack collisions on some objects, can end up going through the world, and lacking collisions on most objects and interactions. Although this provides a safe interaction, where users don't have to worry about collisions with other objects and avatars, it can also feel unrealistic due to lack of responses to stimuli. Neos VR and VR Chat allow for different responses dependent on the world or environment parameters the avatars are situated in. Avatars in both platforms can either allow for a safe interaction, where there is a protective barrier surrounding the avatar that does not allow collisions or could use "drastic" responses where the avatars can be impacted by the collisions or even could be picked up or thrown away by other users in the environment. This drastic interaction setting, when activated, could result in unrealistic interactions but also an unsafe environment where the users need to be aware of their surroundings and keep their distances to certain objects in the environment. Finally, Engage VR and Tivoli Cloud VR both allow for safe interaction where users have the protective barrier from any collisions, while also allowing for realistic responses while interacting with the objects in the environment. For example, if a ball is thrown and it hits an avatar, it would bounce back while not applying any force or unsafe consequences to the avatar itself. Tivoli VR also allows for turning off external collisions altogether.

When all these observations from the sub-categories were combined, only the Level of Embodiment for Neos VR, VR Chat and Tivoli Cloud VR platforms were the strongest among all platforms. Even though Desktop versions are generally lower in terms of their levels of embodiment, regardless Tivoli Cloud VR and VR Chat consistently got high ratings on both Desktop and VR versions.

Some of these sub-categories share similarities with Tanenbaum et. al.'s (2020) believability categories or additional categories we included that are shown in Table 5 below. For example, Agency and Motor Control category in Table 4 is related to Navigation, Gestures & Emotes and Lip Synch categories in Table 5. Responses to External stimuli from Table 4, can be related to Environmental Interaction in Table 5.

Table 5. Social VR Platform Taxonomy Part 3: Tanenbaum et al.'s (2020) Believability categories and Other Categories.

CATEGORIES		PLATFORM					
		Neos VR	**VR Chat**	**Engage VR**	**Mozilla Hubs**	**Alt Space**	**Tivoli Cloud VR**
Level of Motion Capture/ Tracking	VR	One to one - IK	One to one - IK	One to one - IK	One to One - only hand location	One to One - hand and head location	One to one - IK
Navigation	PC	FPS Style (WASD+Mouse)	FPS Style (WASD+Mouse)	FPS Style (WASD+Mouse)	FPS drag mouse to look (WASD+Mouse)	FPS Style (WASD+Mouse)	FPS drag mouse to look (WASD+Mouse)
	VR	Joystick & controllers	Holoport(third person teleport), Joystick & controllers	Teleport, Joystick & controllers	Teleport, Joystick & controllers	Teleport, Joystick & controllers	Teleport, Joystick & controllers
Gestures & Emotes	PC	-	Finger Gestures, Emote Wheel, Emoji Wheel	Limited	3D Emojis	Speech bubble and Emojis	Emojis in chat and Emotes
	VR	Tracking Based	Finger Gestures, Emote Wheel, Emoji Wheel	Tracking Based + Limited	Limited (Controller basic gestures)	Limited (Controller basic gestures and emojis)	Emotes + Tracking based
Lip Sync	PC&VR	Yes	Yes	Yes	Yes	Yes	Yes
Lip Sync Quality	PC&VR	Phoneme	Phoneme	Volume based blendshapes	Volume based (levels of mouth open/close)	Volume based (mouth open/close)	Volume based blendshapes

CATEGORIES		PLATFORM					
		Neos VR	**VR Chat**	**Engage VR**	**Mozilla Hubs**	**Alt Space**	**Tivoli Cloud VR**
Automated-Gesturing	PC	-	-	-	-	-	Yes
Desktop Puppeteering	PC	Hand Gesturing via UI menu	Finger Gestures	-	Hand gestures via virtual pen	-	-
Environmental Interaction	PC	Yes-Point and grab (VR Emulation)	Yes - Point and grab	-	-	Yes - Point and grab	Yes - Point and grab
	VR	Yes - Point and grab	Yes - Point and grab	-	-	Yes - Point and grab	Yes - Point and grab
Extra Social Features	PC&VR	-	Emoji Wheel	3D VR Lecture recordings; Face scan for avatar creation; Shared media viewer	3D Emojis; Spawning 3D models/objects; Pen 3D drawing; Dropping in videos;	Emojis; Speech bubble; Eye gaze	Spawning Objects; Eye gaze; Positional and directional audio; Sign holding system; Script “store” to run social scripts

Most functionality on Level of Motion Capture and Navigation were similar among all platforms. All platforms allowed for one to one tracking, although Mozilla Hubs and Engage VR only were tracking the user's hands. Rest of the platforms use Inverse Kinematics (IK) in their tracking system. All Desktop navigation were inspired from First Person Shooter (FPS) style controls, where keyboard keys "W-A-S-D" were used for movement and mouse for pointing towards the desired location. All VR platforms also use joystick for movement and mostly allow for point-click teleportation to desired location. Only VR Chat was using Holoport functionality instead of teleportation, which is similar to point and click teleportation but with a third person twist. With Holoport, the user points to where they want to go and a representation or copy of their avatar moves to that location in real time while the user stays in their original spot and observes (other users only see the copy move/walk to that new location). Once the copy of the avatar arrives, then the user is teleported to that new location, with the user and the avatar now occupying the same spot but in that new location.

Most platforms except Tivoli Cloud VR had no automated gesturing capability, where some platforms also lack fidelity when it comes to automated lip sync and eye gaze. Neos VR, Mozilla Hubs and VR Chat platforms did not have any auto-gesturing features, with the avatar supplied by the Desktop entry point setup coming across as static during conversations. However, VR Chat has lots of gesture and emoji features accessible through its UI, allowing users to pick the gestures they want, including some finger gestures and an emoji wheel that seems to be heavily utilized by its users. Similarly, Engage VR contains limited gestures and emotes that are accessible by the UI such as raising hand or clapping, and no automated gesturing system. Mozilla Hubs has no automated-gesturing system and in terms of UI it is limited, where it only had a emoji wheel that spawned 3D emoji icons. Finally, Alt Space has a limited set of emojis and gestures compared to other platforms and has no automated-gesturing system.

Compared to Embodied Conversational Agents or the latest motion-capture technology, lip synch quality is generally low in both Desktop and VR versions of the platforms. Neos VR and VR Chat provides phoneme-based lip movements, where Engage VR and Tivoli Cloud provides volume based blendshapes. The lip synch in Alt Space changes the mouth image to a different picture (mouth open vs. mouth closed) when user is talking, similar to a binary toggle. Mozilla Hubs provide volume based scaling for lip

movements for speech, where the mouth opens and closes gradually based on different levels of volume.

In Engage VR there is no avatar head or body rotation, specifically any rotation is represented by the entire body rotating. However, one noteworthy feature is the ability to scan (take a picture using a mobile phone) the user's face during user account setup and use that face as an image on an avatar (the face picture is superimposed onto the avatar face as a texture, see Figure 5). Engage VR also allows users to visit it on their phones, like an iPhone or Android phone. While the experience is inferior, the phone experience in terms of controls and interactions is similar to the one experienced by users who go through the Desktop entry point setup on the same platform. Mozilla Hubs has a benefit in that it runs on the web browser for both users going through VR or Desktop entry point setups. Inside the platform, avatars suffer from collision issues where users can't interact with the majority of objects and can end up going through world elements or terrain. Alt Space has a good eye gaze system where the eyes follow the nearest speaker. But the bodies of the users are disembodied - with floating hands, heads, and bodies, with no legs. The immersion in Alt Space is further broken with the mouth animations, where they are just static images swapping and do not use blend shapes or phonemes like the other platforms. Both Alt Space and Tivoli Cloud's gaze system in the Desktop version automatically follows the nearest speaker/listener, although these systems can occasionally fail to capture the actual focus of attention.

In terms of environmental interactions with the avatar, two platforms (i.e., Mozilla Hubs and Engage VR) did not allow for any as there were no collision physics or interaction with the environment, as we explained in the "Response to External Stimuli" section. However, Mozilla Hubs provides limited UI interaction with some objects that can be spawned, which are not seen in relation to avatar body. Similarly, Engage VR only allows for sitting down and drawing gestures, which are controlled via the UI and cannot be seen in relation to the avatar's body from the user's perspective. All other platforms allowed for a point and grab interaction for both Desktop and VR versions. Neos VR's Desktop version tried to emulate input as VR input and so a lot of VR based interactions work off the get go in desktop mode, although can be non-intuitive to control. In Tivoli Cloud VR, while in a Game environment, some objects aren't configured to interact with the mouse.

We found that none of the platforms allow for the puppeteering of Desktop avatars without VR hardware, except for moving in the environment using keyboard or mouse. Mimicking the movements or behaviour of motion-tracked VR avatars through Desktop was mostly missing. The main type of puppeteering is the head movement in relation to the mouse pointer as mouse look movement, although Neos VR does not allow for just head movement alone but instead moves the whole body altogether. Apart from this Neos VR allows for hand gesturing through the UI menu with the help of Inverse Kinematics. Mozilla Hubs allows limited puppeteering using a virtual pen via a separate interaction menu and VR Chat only has puppeteering through some finger gestures/posing using some key buttons and some UI elements. Finger gestures in VR Chat allows users to pose their fingers in various positions, like making a pointing sign or a peace sign, using the keyboard and mouse. Tivoli Cloud does not allow for additional puppeteering by default and any additions need to be scripted. Tivoli VR uses automated gesturing capability instead of desktop puppeteering, to help with providing more natural interactions with the environment, while using a Desktop entry point.

During our exploratory investigations to Tivoli Cloud VR, first-time users expressed having to do a mental pause when exposed to an Automated-Gesturing avatar first before a Motion-Tracked avatar - as supposedly the Automated-Gesturing avatars had convincing conversational behaviours with similarities to Motion-Tracked avatars, and so users paused when having to recognize and understand the two types of possible avatars in the virtual world platform. Albeit anecdotal, it is due to these observations that the research was inspired to investigate the Social Believability Gap between the two avatars.

In the end, *Tivoli Cloud VR* was chosen as the social VR platform to use for the study. The main reasons for this selection being:

- A generally stronger level of embodiment for both avatar control types compared to other platforms.
    - Intuitive and easy controls of the body.
    - Ability to sprint and walk.
    - Haptic feedback delivered through certain objects with the additional ability to be able to add haptic feedback through scripting.

    - Accurate location of body parts with the body being fully represented.
    - Good mix of realistic and drastic (e.g., fantasy, cartoon, etc.) avatar models.
    - Most objects interacting and colliding with the avatar body though the avatar does not react, plus an option to turn off collisions entirely.
- Inclusion of automated-gesturing features for avatars that are not controlled through VR hardware when conversing, something that other platforms did not contain.
- Volume based blend shape lip sync, including automated eye gaze and eye contact when conversing.

It is important to note that VR and social platform development is an ongoing and changing endeavour. For example, the platform Tivoli Cloud VR while it was active during the year 2021, sadly was shut down in the year 2022. As of the time of writing the popularity and/or features of these virtual platforms have most certainly changed since 2021. For the purposes of this study, only the selected platform and its features captured from the Summer 2021 investigation were used for the creation of our presented research.

4.2. The Two Avatar Control Systems in Tivoli Cloud VR

Figure 7. Screenshot of Tivoli's Automated-Gesturing Avatar in Pitch Scenario shown closer to the screen, this version internally referred to as the Desktop Avatar.

Tivoli Cloud VR (as well as other social VR platforms like Vircadia and Overte) uses two avatar control systems depending on whether users are visiting the virtual world with VR hardware or without it. The research refers to these two avatars as Motion-Tracked avatars and Automated-Gesturing avatars. The platform and the two avatars were chosen because of the automated-gesturing features of the avatar that is controlled without VR hardware having convincing social believability potential. Note that, both avatar control types can be used with all 3D models and environments in Tivoli Cloud VR, and the only visual difference is the types of gestures they support and control they have over the conversational behaviors including gestures, body posture, locomotion, gaze and facial expressions. For example, Figure 7 shows a screenshot while using Automated-Gesturing avatar in a meeting room environment for the Pitch scenario and Figure 8 shows the Motion-Tracking avatar in a dance platform environment for the Disco scenario, both avatar models and scenarios can be used interchangeably for two avatar control mechanisms. The platform includes many environments including, but not limited to, a cafe in a rainforest, a small town, and a beach island. However, for the purposes of our study, we only used the two that fit our intended scenarios - the meeting room and the disco-pub environment.

Figure 8. Screenshot of Tivoli's Motion-Tracked Avatar in Disco Scenario shown closer to the screen.

The difference between the two types of avatars is that the Motion-Tracked avatar uses motion-tracking technology afforded by the VR technology (head tracking using the headset and hand tracking using the hand controllers) while the Automated-Gesturing avatar uses automated behaviours and features to enhance its performance/presentation during conversations. Motion-Tracked avatars, like the one in Figure 8, are driven by VR technology such as Oculus Rift (Oculus, 2016) or HTC Vive (HTC, 2016) head-mounted displays and tracking hardware. They allow the affordance to more directly replicate the behaviour and quirks of an individual's conversational habits due to the motion-tracking nature of the hardware. An example of this can be seen in Figure 9, which shows a captured sequence of a conversational moment of the motion-tracked avatar from Figure 8, giving an example of what a motion-tracked avatar driven by motion-capture technology looks like gesturing in Tivoli Cloud VR. Automated-Gesturing avatars, like the one in Figure 7, are driven using standard computer technology (keyboard and mouse, etc.) without VR technology to control movement. This affords the user the ability to control the avatar's head using the mouse and to control the body (where the user goes and where they are facing) using the keyboard. The user can make the avatar gesture and show emotions using keyboard shortcuts and UI buttons. During conversations, the Automated-Gesturing avatar uses automated gestures to animate conversational gestures at run-time and is typically based on the voice to automate the gestural behaviour and to animate the lip sync. An example of this can be seen in Figure 10, which shows a captured sequence

of a conversational moment of the Automated-Gesturing avatar from Figure 7.

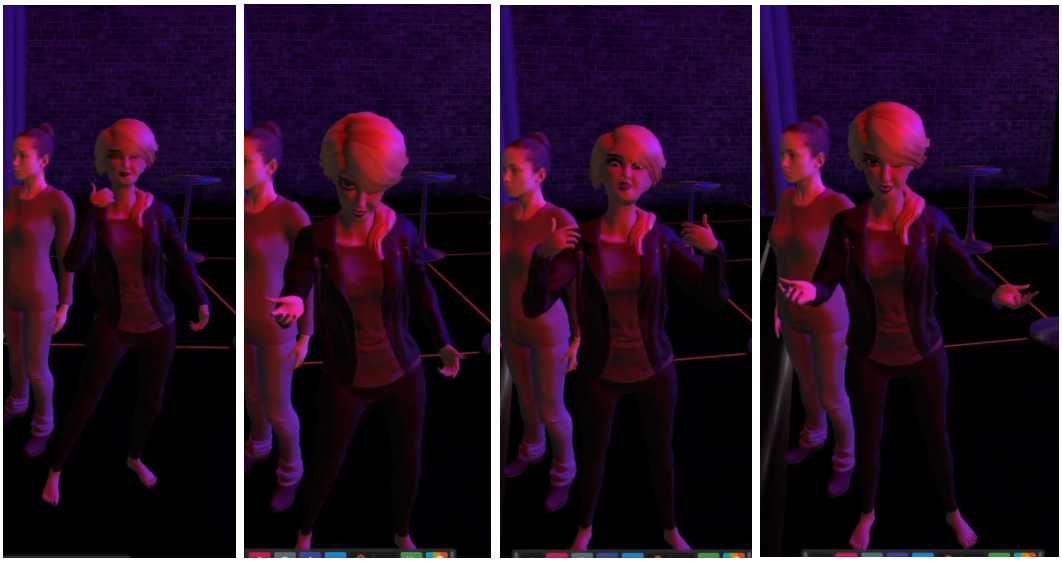

Figure 9. Screenshots capturing a sequence of conversational gestures of the Motion-Tracked avatar in Disco Scenario from Tivoli Cloud VR.

Figure 10. Screenshots capturing a sequence of conversational gestures of the Automated-Gesturing avatar in Pitch Scenario from Tivoli Cloud VR.

Both of these avatar control types are used in our study, where the Pitch and Disco scenario is selected to record videos of avatar interactions in Tivoli Cloud VR. The next chapter will explain the process of gathering materials and details of the study.

Chapter 5.

The Study and the Results

This chapter provides details of our study and its results. Specifically, it will cover the research questions, participants, study details, and the results from our analysis.

To restate our main research questions, we were interested in the following questions when looking at the difference in social fidelity among human-driven Conversational Avatars with different control mechanisms:

- Does an observer in a shared virtual environment notice a Social Fidelity gap among human-driven conversational avatars with different avatar control mechanisms?

- Are the Social Fidelity Gap scores for the conversational avatars correlated with the perceived believability?

For the study and for our analysis we broke down the general questions into more specific questions, with four research questions in total, which we will also label here for convenience. The following is the list of the four research questions and our hypotheses associated with them. We refer to Avatar Control Mechanisms or Avatar Control Types (automated-gesturing vs. motion tracking) as *Avatar Types* from now on. The different scenario types the avatar interactions are taking place is referred as *Scenario*.

- RQ1 - Does the introductory photo for the first conversational avatar to be experienced have an effect on user expectations in terms of Social Fidelity before meeting that avatar?

    - Hypothesis: Introductory photos **will not affect user expectations significantly** before meeting the conversational avatar.

    - Explanation: We needed to ensure any changes in avatar perception is not stemming from the visual differences, but rather from behaviors of the avatars. To further eliminate the visual and scenario confounds, we also counterbalance each 3D model and scenario in our study.

- RQ2 - Keeping in mind the avatar control mechanisms (Avatar Types), what is the observer's reception to the first conversational avatar after meeting them, specifically did the different Avatar Types exceed or unmeet user expectations?
    - Hypothesis: Observer's first exposure to a conversational avatar, regardless of avatar type and scenario, will exhibit a **non-zero Adaptation Gap value when meeting their first avatar**.
    - Explanation: Regardless of prior user expectations, we hypothesize that observing avatar behavior during interaction will cause a change in the observer's understanding of avatar capabilities, which will cause a positive or negative Adaptation Gap, using Social Fidelity scores. Similar to Komatsu et. al.'s work (2012), this question is intended to understand the extent in which the avatar control is affecting the Social Fidelity scores and the main focus of our study.
- RQ3 - Does Avatar Control Type (automated-gesturing vs. motion-tracked) and Scenario (Pitch vs. Disco) have an effect on Adaptation Gap scores, regardless of the avatar exposure order?
    - Hypothesis: motion-tracked avatars will have the observers exhibit **more positive Adaptation Gap scores** than automated-gesturing avatars, with the scenario **not affecting the scores significantly**.
    - Explanation: Similar to the previous research question, this is intended to understand the effect of avatar control mechanism and scenario on the user expectations. However, we aim to get a more holistic view of the Adaptation Gap concept using Social Fidelity scores, by incorporating within-subject effects that might arise due to avatar exposure order.
- RQ4 - Is there a relation between the Adaptation Gap scores and Believability scores of the conversational avatars, specifically can Adaptation Gap be an approximation of the Believability scores?

- Hypothesis: Based on Komatsu et al.'s (2012) paper claiming "participants with positive adaptation gap signs had a significantly *higher acceptance rate* than those with negative ones" (p. 1), we hypothesize there **will be a relationship between Believability scores of a conversational avatar and the observer's Adaptation Gap scores**.

- Explanation: This question further investigates the relationship between the proposed Adaptation Gap concept using Social Fidelity scores and the well-studied Believability concept in avatars in Social VR environments.

5.1. Study Overview

In order to answer the aforementioned research questions, a research study is designed for the evaluation of users' expectations and perceptions of a conversational avatars in terms of different Avatar Types (motion-tracked vs. automated-gesturing). We recorded two avatar performances (based on Avatar Types) in two different scenarios using the Tivoli Cloud VR platform (total of 4 recordings). We used a 2x2 within-subject design where each participant sees recordings of two avatar performances in different scenarios with counterbalanced orders, and evaluate their behavior in terms of believability and social fidelity. The overall study structure can be seen in Figure 11.

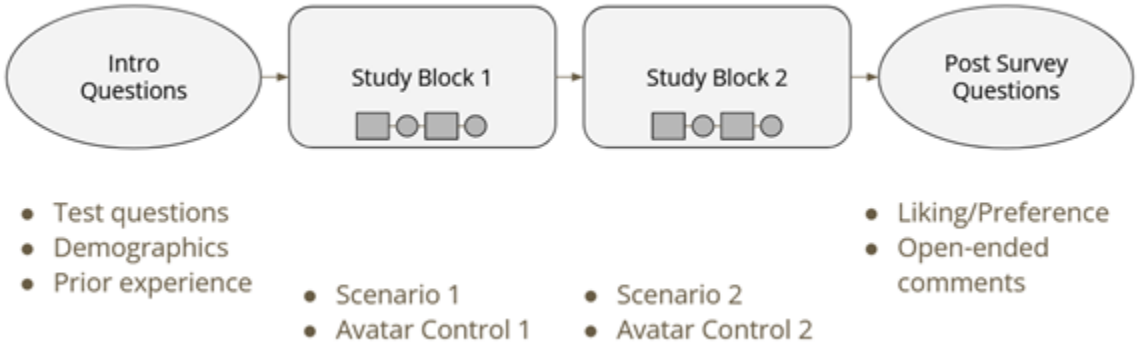

Figure 11. Diagram showing the study structure. The study blocks have their own structure inside.

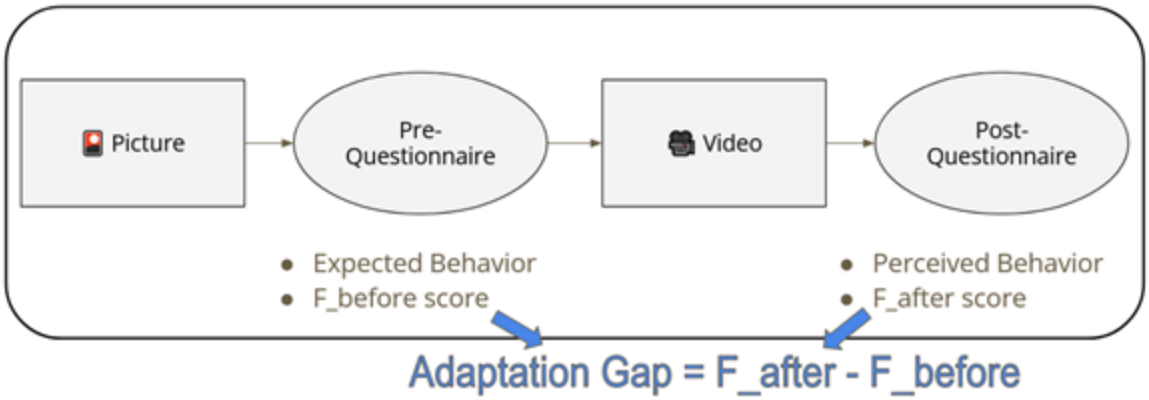

Figure 12. Diagram showing the detailed structure of the Study Blocks shown in Figure 11.

In order to evaluate Adaptation Gap for avatars, each study block starts with a picture representation of the avatar, followed by a questionnaire that measures expected Social Fidelity before meeting the avatar (F_before). This is followed by the recorded video of the avatar in the interaction setting, and a post-questionnaire that asking the perceived Social Fidelity of the avatar after seeing their behavior (F_after). The Adaptation Gap for each avatar is therefore calculated by taking the difference of these two values, as shown in Figure 12, and mentioned in Chapter 3. Each participant sees two of such study blocks, each with a different avatar control type (see Figure 11). Participants completed the study through Survey Monkey[3] platform which contained pictures of the avatars, videos of the avatar performances, and a set of survey questions related to the pictures and videos (order as shown in Figure 11).

5.1.1. The Participants

Participants are recruited from SFU's internal participant recruitment platform (SONA[4]) where graduate and undergraduate student participants can join different studies for course credit. Our study was approved by the university's ethics office[5], and we gave 1% course credits to each student who completed the study.

We recruited a total of 116 students. Out of the 116 students, 88 students passed the validation and attention tests, with the remaining 28 students' responses removed from

[3] https://www.surveymonkey.com/
[4] https://sfu-siat.sona-systems.com/
[5] Ethics approval code 30001194.

the analysis. Attention checks ensured the participants were able to hear and see the videos presented on their screens, and included one test video where participants were required to report what they hear and see in the video. Of the 88 students, 22 students were assigned to each Avatar Type and Scenario combination (2x2 = 4 combinations in total), and each participant was assigned to a dedicated time slot to complete the study. 88 participants successfully completed the survey, of which 29 (33%) identified as males and 57 (65%) as females, 1 (1%) identified as Other (“agender”) and 1 (1%) opted to not identify themselves. 83 (94.3%) of the participants declared their age to fall within 19-24 years of age, with 2 (2.3%) identifying their age falling within 25-34 years, and 3 (3.4%) identifying their age falling within 35-44. When asked if participants had any experience with video games: 87 (99%) responded as having experience with games, with 44 (51%) of those stating as having extensive experience. When also asked if participants had any experience with avatars, 76 (86%) responded as having experience with avatars, with 66 (87%) of those stating as having extensive experience. We defined avatars as 3D humanoid interactive game characters or 3D digital interactive representations of people online.

Figure 13. Screenshot of a meeting room with Automated-Gesturing avatar in Pitch scenario.

5.1.2. Creating the Avatar Videos for the Study Survey

There were four versions of videos, where each avatar control type, or sometimes referred to as just “Avatar Type” in this thesis (Automated-Gesturing and Motion-Tracked, see Chapter 4.2 for details), was seen performing in two different scenario conditions

(Pitch and Disco). In the Pitch scenario, the avatar is located in a meeting room (Figure 13) and pitches to the viewer a fictitious game. The scenario involves some mildly technical conversations related to video games and genres, and includes the avatar referencing and interacting with the room's board/poster (Figure 14). The Pitch scenario represents a formal use case of virtual world meetups. In the Disco scenario, the viewer is invited to a fictitious meet-up with an avatar at a disco pub (Figure 15). The scenario involves meeting a dancing avatar and having the avatar introduce themselves and their interests/hobbies. Included are interactions with a fictitious photo (Figure 16) and a bar-like area. The disco scenario represents an informal use case of virtual world meetups. Each video takes about 2.3 minutes (M=2min20s, SD=4s). The informal and formal nature of the scenarios were based on conversations with the developers of Tivoli Cloud VR on the frequent use cases of social VR platforms expressing users utilizing the platform often for formal meetings and informal get togethers.

Figure 14. Screenshot of Automated-Gesturing avatar in Pitch scenario pointing at an art board.

Concerning the creation of the four videos for the survey, the goal was to capture the performances of both avatar control types in certain contexts, knowing that the driver or performer behind the avatar would also be a factor of presentation. Considering we used a 2x2 within-subject design, the videos were designed with the intention that participants would be exposed to both avatar control types in different scenarios or contexts, though in different orders and combinations. This led to the creation of 4 participant groups that participants could be assigned to with each group having a different

video they watch as their first avatar video and a counterbalanced video (opposite avatar control type and scenario) for their second avatar video.

Figure 15. Screenshot of the disco pub with Motion-Tracked avatar (left) in Disco scenario.

Figure 16. Screenshot of Motion-Tracked avatar in Disco scenario pointing at a photo of a dog.

The first thing our readers will notice is that the study opted for videos for participants to watch, instead of the study involving them experiencing the scenarios and avatars in real-time in VR using Virtual Reality headsets and technology. Here are a few reasons why the study went the video route. *Firstly*, it is important to clarify that the video is not a static video showing the performance of an avatar standing in place. We aimed for the video to emulate a viewer in VR observing the avatar as one would experience if

they were to visit the platform themselves. The video captures various important aspects to convey the dynamic and "observer" nature of the experience by providing a point of view or "POV" of the observer as the main driving camera for the video. To accurately "prime" the user that the video conveys a real VR experience, the video starts with the observer looking down at their own body, moving around rooms, and standing in various places to view the avatar - thus providing an accurate POV of what a user would see when immersed in the platform and participating in the social interaction experience. *Secondly*, the video approach was taken over participants visiting in VR to mitigate some confounding variables that might impact the study results. These include: participants getting motion-sick in VR, the variance in the time it takes to acclimatize to using VR hardware, COVID-19 being prominent at the time of the study and so was a big health concern for participants (issues like sharing headsets, cleaning, sanitizing, ethics approval for an in-person study), and finally the time it takes for participants to get set up for a VR session versus accessing an online survey and viewing a video. *Lastly*, it was very important for us that the experience remained consistent across all sessions. This led to the video approach that allowed us to record videos and facilitate consistency as opposed to attempting to facilitate consistency with participants visiting the virtual worlds in VR. The videos allow us to provide a controlled and consistent point of view for the viewers, as opposed to dealing with individual participant issues. Each participant might not be used to moving in VR, understanding the controls of various VR platforms and hardware, and either interrupting performers or being interrupted/distracted while observing. Ultimately, recording videos of avatar performances was the best solution for the study considering the circumstances.

Next, a pair of scenarios was designed and written in order for the volunteers and avatars to perform within, as a constraint for their performance. The goal of the scenarios, as mentioned previously, was to provide one formal context and one informal context for the avatars to perform in, and so the Pitch and Disco scenario was created. To reiterate, the Pitch scenario involves an avatar pitching to the viewer a fictitious game. The Disco scenario involves the viewer being invited to a fictitious meet-up with an avatar at a disco pub (see beginning of Section 5.1.3 for more details). The purpose of any meeting or get-together in the real world can either be formal or informal. We based the conceptualization of our scenarios on these two categories because we observed frequent use of social VR platforms where users can meet for formal (for example, business meetings) or informal

(for example, social hangouts) purposes. The Disco and Pitch scenarios are fictitious, where the content and stories presented in the scenarios are not taken from real interactions. Despite this, the creation of the scenarios was inspired by secondhand conversations and personal experiences. The scenarios were designed in such a way that they could still be plausible and can represent examples of the use case for social VR platforms when it comes to social activities. The Disco and Pitch scenarios are not the extent or limit of possible scenarios or events that can happen in social VR platforms. The reason we landed on the two specific scenarios was based on a brainstorming session where these two scenarios ended up being the most fleshed out and plausible as a real-world scenario in a social VR platform. We acknowledge that Disco and Pitch can be replaced with any other example of a formal and informal scenario, and hence it was important for our study to counterbalance the scenarios per avatar control type. We further included scenarios as an independent variable of some of our evaluations.

Our videos include an avatar that would gesture, talk, and perform while an observer watches the avatar. To do this we required a volunteer to drive and perform the avatar of interest while a researcher (acting as the observer's point of view) recorded the scenario being acted out. We required volunteers exceptionally familiar with Tivoli Cloud VR and its avatars to control the avatars and act out the performances in a pre-selected environment while being recorded by a screen capture program. With the help of the Tivoli Cloud VR community, a pair of volunteer community members recommended by the developers were selected to help with the video recording by playing the role of the character of interest in the scenario, which included driving the avatars for the recording. These two volunteers needed to reprise their role/scenario twice as each volunteer needed to record their verbal and physical/virtual performance for both the motion-capture avatar and automated-gesturing avatar as closely as possible while working within the avatar control constraints. The researcher recorded the performances using a program called OBS from obsproject.org (Bailey, 2017) while using a first-person camera view inside Tivoli Cloud VR in an avatar body in the same environment as the performers. The video was recorded from the point of view of an "observer" in the scenarios, which involved the researcher rehearsing a set of body and camera movements that were performed in parallel with the volunteer's avatar performance.

Each volunteer was assigned to perform only one of the scenarios. A script was created for each scenario and given to the volunteers to practice and act out. In parallel

an avenue was selected as the set/backdrop for the scenario, which was provided by the community and developers of Tivoli Cloud VR which included a lecture hall-like environment and a disco/pub-like environment for the respective scenarios. The video capture process involved dry runs in the environment with the volunteers acting out their roles and then two or three recording runs of the final performance using one Avatar Type. The volunteers would then repeat their performance again but using the remaining Avatar Type. A challenge for the volunteers was to make sure the verbal performance between the Avatar Types remained consistent, with a bigger challenge being to make sure the physical/virtual performance also remained consistent to a certain degree. Considering the two avatar control types behave differently in terms of controls and their output behaviour, volunteers and the researcher worked together to plan the use of equivalent suitable gestures and behaviours of the automated-gesturing avatar to match as close as possible to the motion-tracked avatar's performance. The intention was to reduce the effect of the performances on the study results being due to unintended variations in avatar performances that we are not interested in, like variations in verbal performances. With all the videos recorded, the best videos in terms of consistency of performance between the Motion-Tracked and Automated-Gesturing avatars were selected for use in the study survey.

5.1.3. Creating the Survey

Each participant is asked to fill the Social Fidelity questionnaires before and after meeting the avatar, where the Social Fidelity scores (F) are calculated as a sum of the Social Fidelity questionnaires (F = Q1 + Q2 + …) which asks about the socio-emotional behavioural range of the agent. The Adaptation Gap scores (AG) were calculated by subtracting the post Social Fidelity Scores (F_{after}) from the pre Social Fidelity scores (F_{before}) that are calculated after and before seeing the avatar video respectively ($AG = F_{after} - F_{before}$). This is similar to the procedure described by Komatsu et al. (2012). The original study and paper from Komatsu et al. focus on questions about task completion. In this study, we focus on the social and emotional behaviour capabilities of the agent. Note that the wording of the questions between the two versions slightly differs to point out their expectations in the pre Social Fidelity questionnaire, with their perceptions in the post Social Fidelity questionnaire. The set of question items used for Social Fidelity calculation

can be seen in Table 6, where the questions are prompted with “How would you rate your expectation/perceptions of the avatar’s...”, and are answered using a 5-point Likert scale.

Table 6. The items in the Social Fidelity questionnaire.

Measurement	Item
Social Skills	social cues, gestures, body language
Conversational Skills	ability to immerse you in a conversation
Movement Skills	range of movement
Emotional Skills	range of emotion
Non-verbal Skills	ability to communicate non-verbally
Visual Realism	feeling like a real person

The Believability of agents were evaluated by using the metrics for Character Believability by Gomes et al. (2013), while removing the social interaction and change with experience measures as neither of our scenarios included examples of these types of behaviours. In addition, we added two items from Bevacqua et al.’s work (2017): avatar control, realism and behaviour believability. The complete set of believability questions can be seen in Table 7, which were rated by using a 5-point Likert scale of participant’s agreement on the statements ranging from “Strongly Disagree” to “Strongly Agree”.

Table 7. Believability questionnaire used in the study.

Measurement	Item
Awareness	Agent perceives the world around it
Understandability	It is easy to understand what the agent was thinking about
Personality	Agent has a personality
Visual Impact	Agent’s behaviour draws my attention
Predictability	Agent’s behaviour is predictable
Coherence	Agent’s behaviour is coherent
Control	Agent was controlled by someone else
Believability	Agent’s behaviour was believable

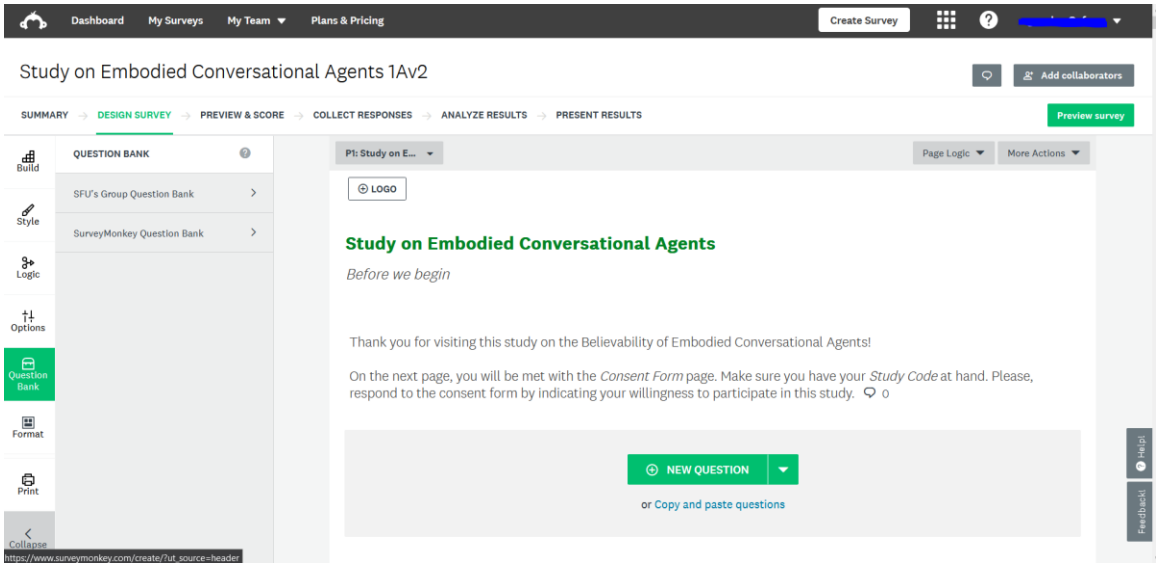

Figure 17. Snapshot of survey editor in Survey Monkey editing the landing page of the Desktop-Pitch variant of the survey.

The survey was created using a survey creation and deployment service called Survey Monkey (see Figure 17) with a university-supplied license. The survey was created using a survey template designed by the research team using Survey Monkey tools and consisted of pre-study test questions, two study blocks (with questions, videos, and pictures), and post-study questions (see Figures 18 - 21). Using the template, four variants of the survey were created to match the combinations of avatar control types and scenarios that the participants would be randomly assigned to, as seen in Table 8. The survey was then deployed within Survey Monkey, with each variant getting a shareable survey link. Each participant would be randomly assigned to a survey variant with no indicators of the variant number or any indication of other variants, with a survey link sent to the participants after signing up. Upon visiting the survey, participants are greeted with a landing page and another page containing the study's consent form and further details about the study. In order to continue with the survey, participants would need to indicate their willingness and consent to participate in the survey by selecting the appropriate button.

5.1.4. Study Procedure

Upon accepting the study participation, participants are asked to complete some test questions, where they watch a video and are asked to report what they saw and heard

(example: see Figure 18). If this test step fails, the participants are not allowed to continue with the study. Participants who pass the test questions are further asked about some demographic questions asking their age, gender and their familiarity with video games and virtual avatars. Then they are presented with a short description of the study procedure before starting the study blocks.

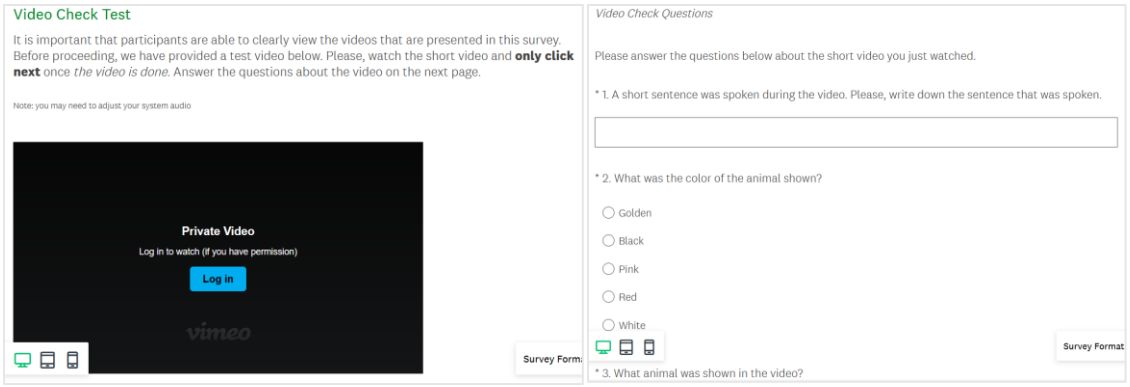

Figure 18. Editor screenshots of Desktop-Pitch variant survey preview showing test questions. Left: video check test (private video does not represent what participants see). Right: segment of post video check questions.

Each participant is then presented with two sets of study blocks. In each block, participants first see a picture of an avatar (example: see Figure 19) and are asked about their expectations of the pictured avatar's behaviour in terms of the Social Fidelity questionnaire.

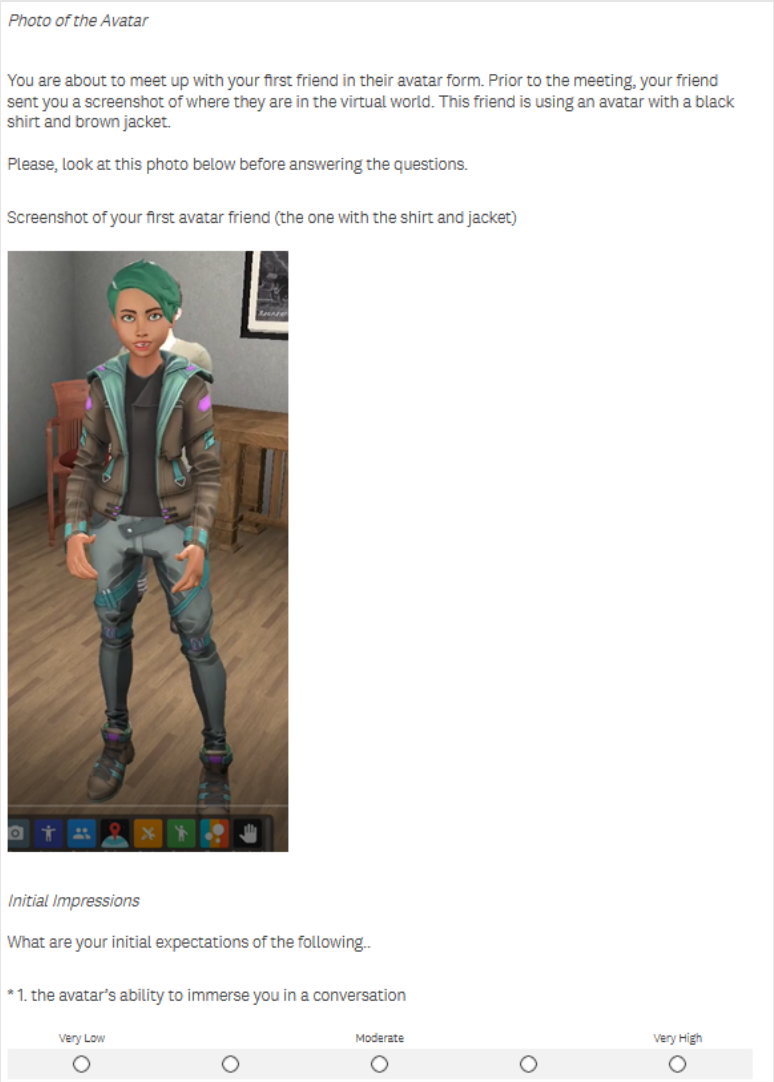

Figure 19. Editor screenshot of VR-Pitch variant survey preview showing avatar photo (Motion-Tracked avatar in Pitch scenario) and initial impressions questions.

Participants are then presented with a video of that avatar (example: see Figure 20) in an interaction scenario and are asked two attention-check questions related to the video to ensure they watched and heard the interaction in the video successfully. Then, they are presented with a post-questionnaire (example: see Figure 21) including their perceived scoring for the avatar in terms of its Social Fidelity, followed by a set of Believability questions.

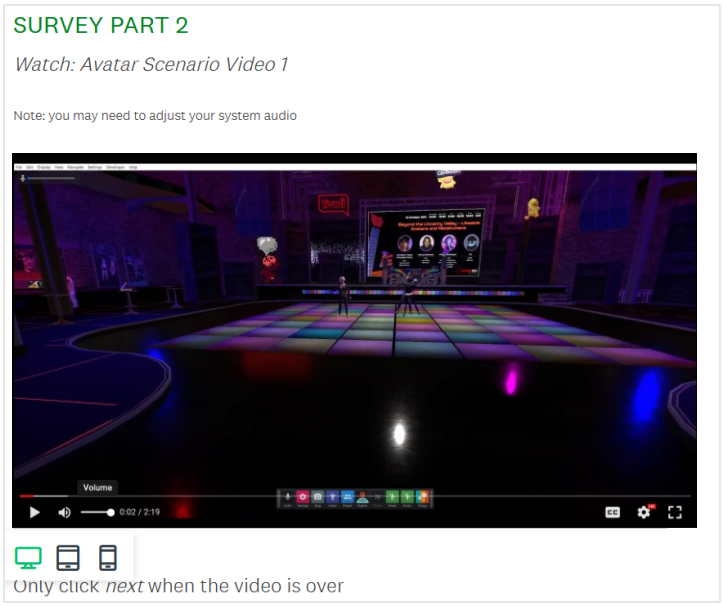

Figure 20. Editor screenshot of Desktop-Disco variant survey preview showing the first avatar scenario video.

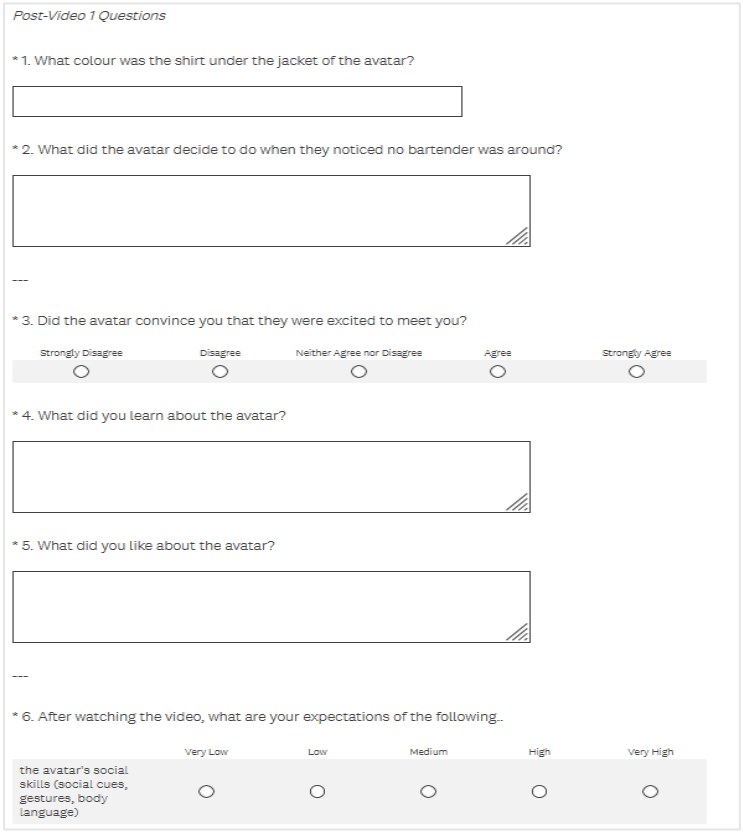

Figure 21. Editor screenshot of Desktop-Disco variant survey preview showing part of the post video 1 questionnaires.

Each participant saw two different videos with two different avatar control types (automated-gesturing and motion-tracked), in two different scenario conditions (Pitch and Disco environment). Participants were counterbalanced in the order in which they saw the motion-tracked or automated-gesturing avatar versions. The number of participants assigned to different combinations of avatar control types with the different scenarios was randomized but also partially controlled to make sure one variant was not being predominantly selected by the randomization over another. This helped to make sure participants were evenly spread between the 4 study variants as evenly as possible by the end of the research study. Table 8 provides an overview of the four groups with their coding that participants were assigned to in the study and the avatars/scenarios they were exposed to and in which order.

Table 8. 2x2 within-subject design with counterbalanced orders (Desktop = automated-gesturing; VR = motion-tracked).

Study Group Number	Study Group Coded	First Exposed Avatar Type and Scenario	Second Exposed Avatar Type and Scenario
1	1A	Desktop-Pitch	VR-Disco
2	1B	VR-Pitch	Desktop-Disco
3	2A	Desktop-Disco	VR-Pitch
4	2B	VR-Disco	Desktop-Pitch

The avatar control types, motion-tracked and automated-gesturing, internally were denoted as Desktop and VR avatars, then coded internally as A and B (A = Desktop, B = VR). The scenarios internally were denoted as Pitch and Disco and also coded internally as 1 and 2 (1 = Pitch, 2 = Disco). Following this, each of the four videos was given a code to represent the assigned avatar control type and scenario: 1A, 1B, 2A, 2B.

After going through both blocks, the participants were also asked to fill in a post-study questionnaire about which avatar they liked the most and asked to write about why they made that choice in a free-form text box. The whole study took about 20 minutes to complete. For samples of the study questions and structure, see Appendix A.

5.2. Results

We first removed the participants who failed to answer any of the attention checks during the study. This left us with a total of 88 participants who were included in the final

analysis, after the removal of 28 participants. We used IBM's SPSS software (IBM Corp., 2019) and R Software (R Core Team, 2022) for our analysis, with the lme4 package for R (Bates et al., 2015) for Linear Mixed Method Analyses. Effect sizes are calculated using Cohen's d, where .2 indicates small, .5 medium and .8 indicates large effect (Cohen, 1988).

5.2.1. Social Fidelity Expectations before Meeting the Avatars

For *research question 1* we wanted to know if the introductory photo for the first conversational avatar to be experienced has an effect on user expectations before meeting that avatar. Specifically, we analyzed whether different participants had different expectations on the avatars' social fidelity before encountering any of them. This allows us to control if the participants had different expectations about avatars' behaviour, before interacting with them.

To achieve this, we compared the pre Social Fidelity scores for the first avatar each participant encountered using a t-test, with Social Fidelity pre-scores of the first avatar as the dependent variable and Avatar Type (Automated-Gesturing vs. Motion-Tracked) as the independent variable. There were 44 participants that had an automated-gesturing avatar as their initial avatar exposure and 44 participants with a motion-tracked avatar as their initial avatar exposure. An independent-samples t-test was run to determine if there were differences in the Automated-Gesturing and Motion-Tracked groups in their Social Fidelity scores before meeting their first avatar. There were no outliers in the data, as assessed by inspection of a boxplot. Social Fidelity scores for each level of Avatar Type were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .653$).

Results showed that the Automated-Gesturing avatar group's initial social fidelity scores ($M = 17.57$, $SD = 4.25$) were slightly higher than the Motion-Tracked avatar group's initial social fidelity scores ($M = 17.32$, $SD = 4.53$), a non-statistically significant difference, $M = 0.25$, 95% CI [-1.61, 2.11], $t(86) = 0.27$, $p = .653$. As expected, our results showed there was no statistically significant difference between means ($p > .05$), and therefore, we fail to reject the null hypothesis and reject the alternative hypothesis. Figure 22 shows a visual representation. This means the introductory photos do not influence the

participant's pre-exposure scores, confirming that the different pictures of the avatars did not have an effect on the expectations of the participants on agent capabilities.

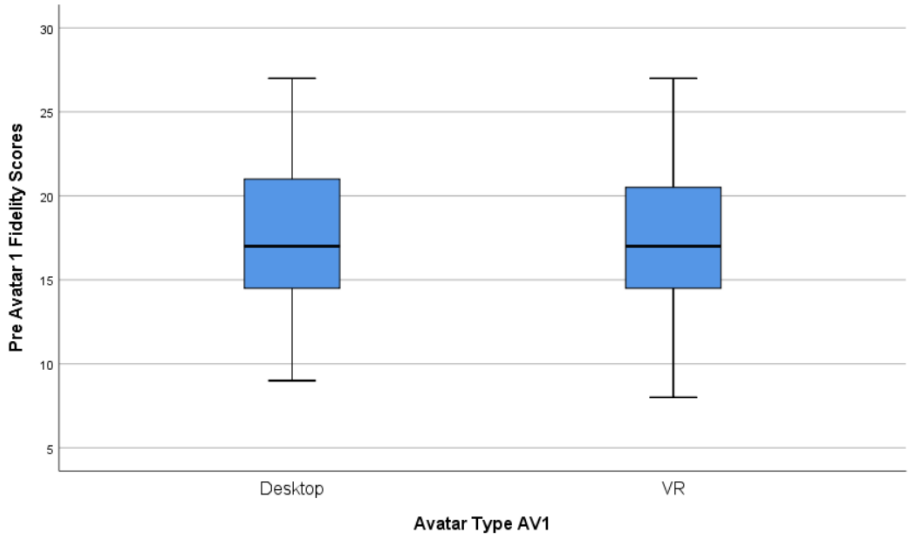

Figure 22. Boxplot graphs showing the Pre-Social Fidelity values for the groups based on the initially exposed Avatar Type (Desktop = automated-gesturing; VR = motion-tracked). There were no significant differences between avatar control types.

5.2.2. Observer's Reception to their First Avatar

For *research question 2* we wanted to know what the observer's reception to their first conversational avatar was after meeting them. Specifically, we were interested if the different Avatar Types exceeded or unmet observer expectations. This allows us to understand the extent of the first Adaptation Gap observers would experience.

A two-way ANOVA was conducted to examine the effects of Avatar Types (Automated-Gesturing vs. Motion-Tracked) and Scenario (Pitch vs. Disco) as independent variables on Adaptation Gap score as the dependent variable for first avatar exposure. Residual analysis was performed to test for the assumptions of the two-way ANOVA. Outliers were assessed by inspection of a boxplot, normality was assessed using Shapiro-Wilk's normality test for each cell of the design, and homogeneity of variances was assessed by Levene's test. There were no outliers, and there was homogeneity of variances (p = .561). Data was normally distributed except for the Disco-Desktop cell (p = .018). However, upon inspection of the QQ plot, Disco-Desktop looked to be acceptably normally distributed. Therefore, we continued with the two-way ANOVA analysis.

Our results showed a significant main effect for Scenario ($F(1, 84) = 11.027$, $p = .001$, partial $\eta 2 = .116$), with the Adaptation Gap for the Disco scenario (M= 3.95, SD= 4.281) was higher than the Pitch scenario (M= 0.45, SD=5.716), which showed the Pitch scenario was often not able to meet the expectations of the participants. However, we did not find a significant effect on the main effect of Avatar Type alone ($F(1, 84) = 1.071$, $p = .304$, partial $\eta 2 = .013$), where motion-tracked avatars (M=2.75 , SD=5.327), had slightly higher Adaptation Gap than the automated-gesturing avatars (M=1.66, SD=5.318).

Descriptive Statistics

Dependent Variable: Adaptation Gap 1

Scenario AV1	Avatar Type AV1	Mean	Std. Deviation	N
Disco	Desktop	4.55	3.985	22
	VR	3.36	4.573	22
	Total	3.95	4.281	44
Pitch	Desktop	-1.23	4.956	22
	VR	2.14	6.034	22
	Total	.45	5.716	44
Total	Desktop	1.66	5.318	44
	VR	2.75	5.327	44
	Total	2.20	5.320	88

Figure 23. Descriptive Statistics showing Mean and Std. Deviation for Adaptation Gap between Scenario and Avatar Type (Desktop = automated-gesturing; VR = motion-tracked).

There was also a statistically significant interaction between Avatar Type and Scenario for Adaptation Gap score, $F(1, 84) = 4.65$, $p = .034$, partial $\eta 2 = .052$. Therefore, an analysis of simple main effects for Scenario and Avatar Type was performed with statistical significance receiving a Bonferroni adjustment. All pairwise comparisons were run for each simple main effect with reported 95% confidence intervals and *p*-values Bonferroni-adjusted within each simple main effect.

Tests of Between-Subjects Effects

Dependent Variable: Adaptation Gap 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	409.318[a]	3	136.439	5.583	.002	.166
Intercept	427.682	1	427.682	17.499	.000	.172
Starting_Scenario	269.500	1	269.500	11.027	.001	.116
Starting_Avatar	26.182	1	26.182	1.071	.304	.013
Starting_Scenario * Starting_Avatar	113.636	1	113.636	4.650	.034	.052
Error	2053.000	84	24.440			
Total	2890.000	88				
Corrected Total	2462.318	87				

Figure 24. **Test of Between Subject Effects for Adaptation Gap between Scenario and Avatar Type, including the interaction effect. There is a statistically significant interaction between Avatar Type and Scenario.**

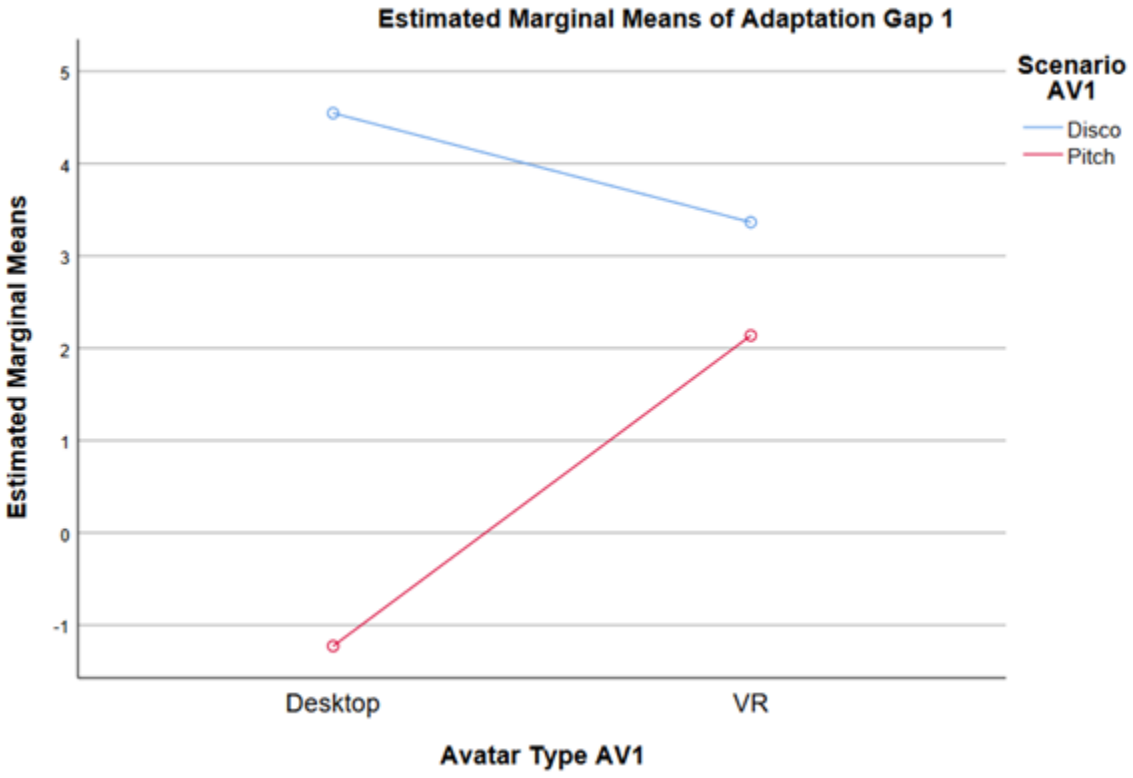

Figure 25. **Estimated Marginal Means of Adaptation Gap with Avatar Type (Desktop = automated-gesturing; VR = motion-tracked).**

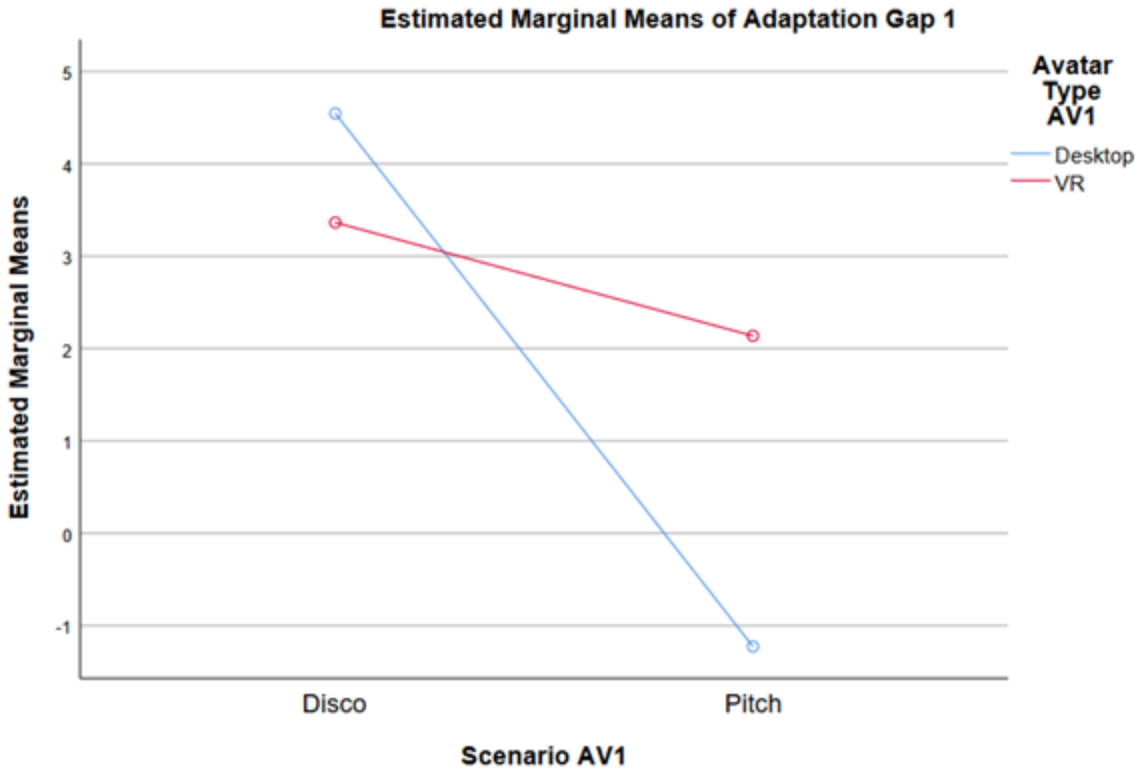

Figure 26. Estimated Marginal Means of Adaptation Gap with Scenario (Desktop = automated-gesturing; VR = motion-tracked).

There was a statistically significant difference in mean Adaptation Gap scores for Pitch and Disco scenarios who had an Automated-Gesturing avatar as their first avatar, $F(1, 84) = 15.00$, $p < .001$, partial η2 = .152, unlike for Motion-Tracked avatar, $F(1, 84) = 0.68$, $p = .413$, partial η2 = .008.

For an Automated-Gesturing as the initially exposed avatar, mean Adaptation Gap score for Disco was 4.55 (*SD* = 3.99) and -1.23 (*SD* = 4.96) for Pitch scenario, a statistically significant mean difference of 5.77, 95% CI [2.81, 8.74], $F(1, 84) = 15.00$, $p < .001$, partial η2 = .152. For a Motion-Tracked as the initially exposed avatar, mean Adaptation Gap score for Disco was 3.36 (*SD* = 4.57) and 2.14 (*SD* = 6.03) for Pitch scenario, a mean difference of 1.23, 95% CI [-1.74, 4.19], $F(1, 84) = 0.68$, $p = .413$, partial η2 = .008, which was not statistically significant.

There was a statistically significant difference in mean Adaptation Gap score between Automated-Gesturing and Motion-Tracked avatars in the Pitch scenario, $F(1, 84)$

= 5.09, p = .03, partial η2 = .057, unlike for Disco scenario, F(1, 84) = 0.63, p = .43, partial η2 = .007.

For the Disco scenario, mean Adaptation Gap score for Automated-Gesturing avatar was 4.55 (*SD* = 3.99) and 3.36 (*SD* = 4.57) for Motion-Tracked avatar, a mean difference of 1.18, 95% CI [-1.78, 4.15], F(1, 84) = 0.63, p = .43, partial η2 = .007, which was not statistically significant. For the Pitch scenario, mean Adaptation Gap score for Automated-Gesturing avatar was -1.23 (*SD* = 4.96) and 2.14 (*SD* = 6.03) for Motion-Tracked avatar, a mean difference of 3.36, 95% CI [0.4, 6.33], F(1, 84) = 5.09, p = .03, partial η2 = .057, which was statistically significant.

5.2.3. Adaptation Gap Scores Regardless of Exposure Order

We next evaluated, for *research question 3*, whether Avatar Type or Scenario have an effect on the Adaptation Gap scores regardless of the avatar exposure order. Note that, this research question is different from the prior question, where we evaluated the Adaptation Gap for the first avatar, making it a between-subjects analysis. This research question and its analysis, on the other hand, will give us a more holistic view of the Adaptation Gap scores and the effect of the avatar control type and scenario in a mixed-methods evaluation that takes into account both within and between subjects effects. To do so, we fitted a linear mixed model with Adaptation Gap as a dependent variable, Avatar Type (Automated-Gesturing vs. Motion-Tracked) and Scenario (Pitch vs. Disco) as fixed effects while including the interaction effect between the two, and finally included participant ID as random factor as each participant sees two different Avatar Type×Scenario combinations.

To test the assumption of conditional normality, a Shapiro-Wilk test was run on the response Adaptation Gap for each combination of levels of factors Avatar Type and Scenario. All combinations were found to be statistically non-significant except condition (Desktop-Disco), which showed a statistically significant deviation from normality (W = .941, p = .034). Due to the exception in one condition failing the normality test, this provided the reason for continuing the analysis through the use of linear mixed models. The normality assumption on the residuals was tested with a Shapiro-Wilk test on the full model. The test was statistically non-significant (W = .991, p = .45), indicating compliance with the normality assumption. A Q-Q plot of residuals visually confirmed the same.

We used the lmerTest (Kuznetsova et al., 2017) library to test the significance of each variable on the Adaptation Gap scores. Figure 27 shows an interaction plot with ±1 standard deviation error bars for Avatar Type and Scenario. A linear mixed model analysis of variance indicated a statistically significant effect on Adaptation Gap of Avatar Type $F(1, 86) = 3.77$, $p = .05538$; Scenario $F(1, 86) = 68.42$, $p < 0.001$; and of the Avatar Type×Scenario interaction $F(1, 86) = 6.26$, $p = 0.0143$. The boundary significance of the effect on Adaptation Gap of Avatar Type can be explained by the small effect size from the Automated-Gesturing and Motion-Tracked avatars ($d = -.29$), which is reported next.

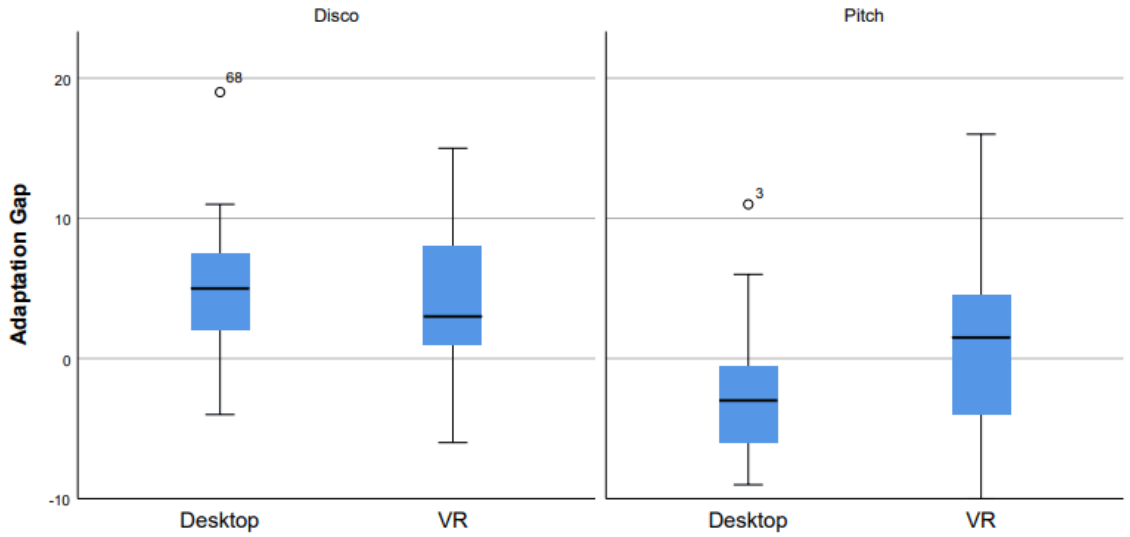

Figure 27. Boxplot graphs showing the Adaptation Gap values for the interaction groups based on Avatar Type and Scenario (Desktop = automated-gesturing; VR = motion-tracked).

Participants seem to experience significantly smaller Adaptation Gaps for the automated-gesturing avatar ($M = 1.08$, $SD = 5.71$), compared to the motion-tracked avatar version ($M = 2.40$, $SD = 5.65$) with a small effect size ($d = -.29$); and for the Pitch scenario ($M = -1.07$, $SD = 5.51$) compared to the Disco scenario ($M = 4.55$, $SD = 4.36$) with a large effect size ($d = 1.25$).

Pairwise comparisons using paired-sample t-tests, corrected with Holm's sequential Bonferroni procedure, indicated that the Disco scenario in both Avatar Types was not significantly different, however, it always resulted in a significantly larger Adaptation Gap regardless of the Avatar Type given it is compared to the Pitch scenario. Within the Pitch scenario, the Motion-Tracked avatar showed a significantly larger Adaptation Gap score compared to the Automated-Gesturing version. When comparing

Automated-Gesturing Avatar Type from both scenarios, there was a significant difference with the Automated-Gesturing avatar from the Disco scenario scoring higher than the one from the Pitch scenario, likely due to the context and nature of the scenario. Now comparing the Motion-Tracked avatar of both scenarios, similarly, the Motion-Tracked avatar from the Disco scenario scored higher than the Motion-Tracked avatar from the Pitch scenario, showing also the context and the scenario playing a role in the difference. Interestingly, despite the Automated-Gesturing avatar being generally weaker when compared against the Motion-Tracked avatar, there is a significant difference when comparing the Automated-Gesturing avatar from the Disco scenario and the Motion-Tracked avatar from the Pitch scenario - with the Automated-Gesturing avatar scoring higher than the Motion-Tracked avatar, further showing that the scenario seems to be boosting the score for Automated-Gesturing avatars. However, the Pitch scenario based on the data seems to be a less influential scenario when it comes to influencing the Adaptation Gap scores and so the Motion-Tracked avatar from the Disco scenario naturally ends up with a significant difference with a larger Adaptation Gap score when compared to the Automated-Gesturing avatar from the Pitch scenario. Figure 27 shows a visual representation, and Table 9 shows the statistics results for the pairwise comparisons.

Table 9. Results of the Pairwise Comparison for Adaptation Gap Scores including t-values, significance levels (p) and effect sizes (d) (Desktop = automated-gesturing; VR = motion-tracked).

Pairwise Comparison	t-value	p	d
Desktop Disco – VR Disco	0.615	.54	.14
Desktop Disco – Desktop Pitch	7.313	<.001	1.68
Desktop Disco – VR Pitch	4.475	<.001	.095
VR Disco – Desktop Pitch	7.222	<.001	1.54
VR Disco – VR Pitch	3.536	.002	.81
Desktop Pitch –VR Pitch	-3.162	.004	-.73

5.2.4. Adaptation Gap versus Believability

For *research question 4*, we further investigated the possible relation between the Adaptation Gap scores and Believability scores of our agents, to test if Adaptation Gap scores can be seen as an approximation to the Believability scores. A linear regression was run to understand the effect of Adaptation Gap scores on Believability scores, with Believability scores set as a dependent variable, Adaptation Gap scores as an independent variable. To assess linearity a scatterplot of Adaptation Gap scores against Believability scores with a superimposed regression line was plotted. Visual inspection of these two plots indicated a linear relationship between the variables. There was homoscedasticity and normality of the residuals. There were 3 outliers detected that deviated more than ±3 standard deviations. Upon inspection, the values are for the Believability score and they had values under the maximum possible. They were not deemed serious outliers and were kept in the analysis.

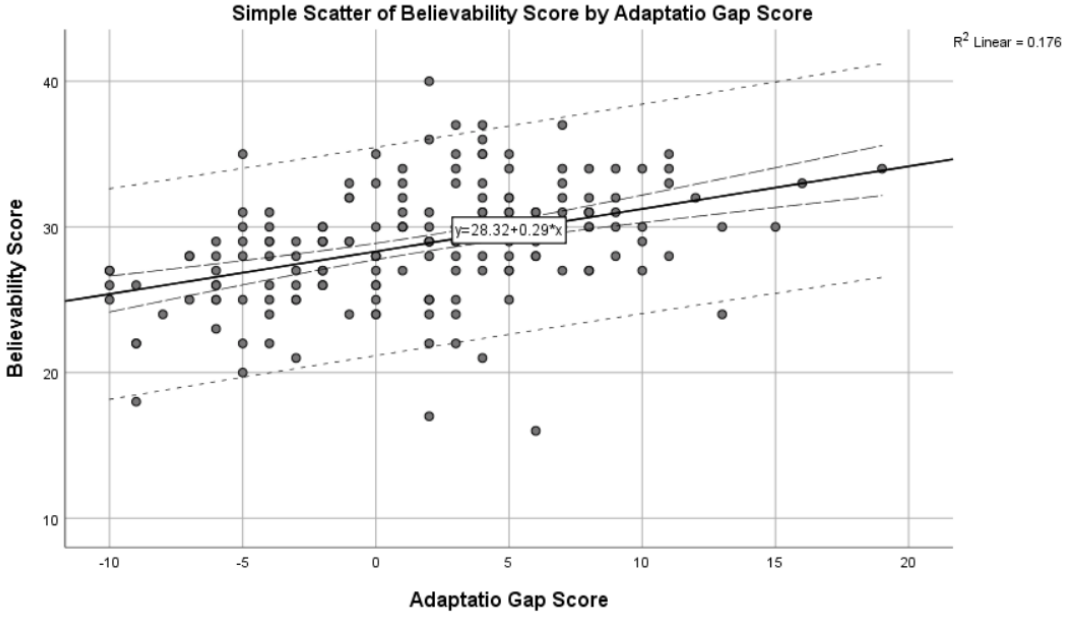

Figure 28. Scatterplot showing the linear regression of Adaptation Gap and Believability values with a line of best fit, confidence and prediction intervals.

The prediction equation was: Believability = 28.316 + (0.292 x Adaptation Gap). Adaptation Gap scores statistically significantly predicted Believability scores, $F(1, 174) = 37.23$, $p < .001$, accounting for 17.6% of the variation in Believability scores with adjusted

R^2 = 17.2%, a small size effect according to Cohen (1988). Predictions were made to determine mean Believability scores for those people who got average positive Adaptation Gap scores of 5, 10 and 15. For AG = 5, the mean Believability score was predicted as 29.78, 95% CI [29.16, 30.4]; for AG = 10 it was predicted as 31.24, 95% CI [30.29, 32.19]; and for AG = 15, it was predicted as 32.7, 95% CI [31.34, 34.06]. Figure 28 shows a visual representation of the relationship between Adaptation Gap and Believability.

5.2.5. Avatar Preference and Qualitative Results

At the end of the survey, participants were asked to vote for their favourite avatar between the two that were shown to them (see Table 8 for which groups had which avatars). We specifically asked the following question to do so: “You've just watched TWO videos of different avatars. Regardless of their personalities and the specific scenarios they were in and focusing strictly on their ability to communicate and collaborate, which avatar would you choose as the best one, again in terms of its ability to communicate and collaborate?” From the total of 88 participants, 52.25% of them had a preference for the motion-tracked avatar, and 68% of them chose the Disco scenario. Looking into the automated-gesturing avatar versus the motion tracked avatar within each scenario, we will start off with the Pitch scenario. From the total of 88 participants:

- 28 preferred the performance of the Pitch scenario avatar.
    - 15 of the votes were in favour of motion-tracked avatar.
    - 13 were in favour of the automated-gesturing avatar.
- 60 preferred the performance of the Disco scenario avatar.
    - 31 of the votes were in favour of motion-tracked avatar.
    - 29 were in favour of the automated-gesturing avatar.

Although the wording of the question was intended to minimize the effect of scenario and other effects, it cannot be guaranteed that this was achieved. To better understand the reasons for this preference, one has to look further into the detailed qualitative comments the participants provided.

At the end of the survey, participants provided justification for their votes and were allowed to optionally provide additional comments. From these comments, we observed and compiled a few repeating themes and items that are of interest. We will report them below for all the study groups (see Table 8 for details). Full list of participant comments can be seen in Appendix B and C.

1A GROUP - DESKTOP PITCH vs. VR DISCO

Viewers preferred the VR-Disco avatar over the Desktop-Pitch avatar (second exposed avatar), with a 77.27% preference ratio.

Desktop-Pitch: Viewers found this avatar's movement robotic, however, some were surprised by the amount and quality of hand gestures. Some focused on the Masculine voice and how it differed from the body/outfit that was presented. Eye contact and gaze were noticed and how good it was. Some complimented the outfit and how it matched the game aesthetic. Interactions with the wall and referencing the art wall were mentioned. Compared to the VR Disco one, some say this avatar was calmer and easier to understand.

VR-Disco: Viewers liked the casual nature of this avatar and found it more outgoing. The avatar in general was more expressive, had a wider range of vocals and gestures, and had more interactions with the environment and the viewer. People generally liked this avatar's personality more and noticed that the disco environment was more believable and preferred over the pitch environment.

1B GROUP - VR PITCH vs. DESKTOP DISCO

Viewers preferred the Desktop-Disco avatar over the VR-Pitch avatar (second exposed avatar), an 81.82% preference ratio.

VR-Pitch: Viewers initially notice the male voice over the gender-ambiguous body. Avatar has limited facial expressions but good conversational gestures. Some noticed the amount of eye contact the avatar makes. Some compliments on the interesting avatar design and outfit. Avatar displayed good body language though most movements were jittery, especially positional and leg movements. Many remembered the avatar referencing the artboard and pointing to it. In general, viewers say the avatar moved around a lot and the body language and gestures fit what was being said.

Desktop-Disco: Viewers really notice the avatar's love for dogs. This avatar had better hand and finger gestures, was more expressive, and had a better personality by being enthusiastic and outgoing. They were also less jittery with their feet and movement than the other avatar. Viewers liked their voice and the context of the meet up. Most complement the outfit by being more believable. Many remember the avatar being able to point at things and a red block appearing showing what the avatar is pointing to. This avatar also lacks facial expressions. Most also reference the dancing in the beginning as being attention-grabbing and a memorable moment.

2A GROUP - DESKTOP DISCO vs. VR PITCH

There was no apparent preference from the viewers preferring the Desktop Disco or VR-Pitch (second exposure) avatars as scores were tied, with a 50% preference ratio.

*Desktop-Disco***:** Viewers liked the personality of the avatar and particularly found the avatar talking about its dog to be memorable. Viewers complained about the lack of facial emotions and expressions. The dancing portion at the beginning done by the avatar made good impressions on the viewers, and particularly the viewers enjoyed the enthusiasm of the avatar as their voice had more emotions, which coupled well with the more natural outfit the avatar had. The avatar also displayed clear body movements and gestures, with smoother animations, and was more interactive with the viewer and environment.

*VR-Pitch***:** The first thing viewers mention is the cool stylized outfit of this avatar, which is not natural but grabs attention. This avatar had a calmer voice, which some preferred, and compliment on their coherent speech and better speech mannerisms - meaning it was easier to understand this avatar. However, avatar had jittery movement, particularly around the feet. Also was lacking in facial expressions and emotions. The avatar displayed more natural gestures and body movement fitting with the context and the spoken words, though jittery at times, which caused some unnatural poses in addition to lots of shifting of the feet.

2B GROUP - VR DISCO vs. DESKTOP PITCH

Viewers preferred the VR Disco avatar (first exposed avatar) over the Desktop-Pitch avatar, with a 63.64% preference ratio.

VR-Disco: Viewers most remember the avatar talking about their dog. The first thing they notice is that the avatar is lacking in facial expressions, but also that the avatar is outgoing, expressive, and has a good attitude. The high expressiveness of the avatar seemed to annoy some viewers. The more realistic clothing was also preferred by some but not by others. The avatar generally had smooth movements and had authentic gestures though rather exaggerated. Voice and their accent were also something that stood out for viewers along with the occasional weird leg movement of the avatar doing physical movement. Interaction with the viewer was the top reason for some viewers as to why they liked this avatar more.

Desktop-Pitch: First thing viewers point out is that this avatar has a cool outfit. The avatar had smooth hand movements but lacked facial emotions. The upper body movement was natural but everything else did not work so well. It was especially pointed out that the avatar repeated gestures a lot, and some pointed out that the voice did not match the avatar's visual look. What helped viewers remember this avatar was the references to the image board the avatar made. Viewers did not like the monotone, script reading, nature of the avatar though some preferred how clearly this avatar spoke.

Chapter 6.

Discussion

This chapter discusses the results from the previous chapter, some of the observations taken from the qualitative questions of the survey, and ends with a discussion on some of the limitations of the research.

Participants' prior expectations of avatar behavior

For *research question 1* we wanted to know if the introductory photo for the first conversational avatar to be experienced has an effect on user expectations before meeting that avatar. Komatsu et. al.'s research (2012) suggests agents with significant differences in appearance may in fact affect the user expectations. This was an important question to answer first because if there was any evidence that the images alone cause a significant difference in initial impressions (and hence cause a difference in initial Social Fidelity scores), then we would have needed to account for that difference in expectations and also comparisons of the Adaptation Gap. Our results showed that, as expected, there was no significant difference in the initial Social Fidelity scores, which validated our expectation that the introductory photos of different 3D avatar models we used do not have a major influence on the initial Social Fidelity scores, and as such do not create a disparity among these avatars.

The effect of avatar control type on social fidelity

For *research question 2* we wanted to know what the observer's reception to their first conversational avatar's social fidelity was after meeting them and watching their behavior, which is judged based on the Adaptation Gap score, and whether it changes between avatar control types. Our results showed that, as expected, the Motion-Tracked avatar resulted in a larger and positive Adaptation Gap than the other avatars. Given that the initial Social Fidelity scores (the scores prior to watching the videos) were similar between the avatars, this indicates that the Motion-Tracked avatars exceeded participants' expectations of Social Fidelity more so than the Automated-Gesturing avatars. This confirms our hypothesis that VR-controlled motion tracking systems can have greater social fidelity with the technology's tight coupling with human behaviour. Since Motion-

Tracked avatars scored higher than Automated-Gesturing avatars, this could create a disparity between the two avatars.

What was interesting is that the results also showed an interaction effect between avatar control type and scenario. Specifically, the effect of Automated-Gesturing (Desktop) avatar control type showed a drastic difference in Adaptation Gap scores between the two scenarios, where the Motion-Tracked (VR) avatar control consistently had high Adaptation Gap scores regardless of the scenario type. This shows that user expectations can significantly change in automated-gesture control depending on the scenario the avatars are being used at. Our results even showed a negative Adaptation gap score for the Automated avatar control in the Pitch scenario, suggesting users' expectations for this avatar control and scenario combination were unmet. Compared to this, average Adaptation Gap scores for the Motion-Tracked avatar control was consistently high and positive, showing the avatars exceeded expectations in both scenarios. However, we also see that in the Pitch scenario, some participants still rated Motion-Tracked avatar as not meeting expectations (negative Adaptation gap score), and Disco scenario was consistently rated higher, likely due to the scenario effect which we will cover in the next subsection.

Moving on to our results that correspond to the *research question 3*, where we evaluated whether the Avatar Type or the Scenario have an effect on the Adaptation Gap scores, regardless of the avatar exposure order. This allowed us to incorporate any within-subjects effects into our model, to examine whether the participants were changing their expectations when they encountered the second avatar with a different avatar control mechanism. Albeit significant, our results show a small effect size (d=-.33) for Avatar Type, where Automated-Gesturing avatars had lower Adaptation Gap scores than the Motion-Tracked avatars for the participant. This means that Automated-Gesturing avatars perform worse in terms of Social Fidelity than Motion-Tracked avatars. However, due to the small effect size, this means that Avatar Type alone does not explain the effect on the Adaptation Gap score sufficiently and other factors such as scenario have a lot of effect on the user's expectations.

These quantitative results, combined with the qualitative feedback we received from the participants showed us that Automated-Gesturing avatar control types could meet or exceed user expectations given the correct context, such as an informal interaction

scenario like Disco environment, where the expectations likely are low to begin with. However, Automated-Gesturing avatars are not on par with Motion-Tracked avatars and (ignoring the scenario) will not be favoured over the Motion-Tracked variant.

The Motion-Tracked variant consistently performs better than the Automated-Gesturing avatar, with participants voting it as their most preferred avatar when comparing all Avatar Types in all scenarios. Based on the open-ended answers discussed in Section 5.2.5, the Motion-Tracked avatar is favourable due to a couple of factors: better expression and more variety of movement and gestures due to the motion-tracked nature of the technology, the translation of unique personality quirks of the performers from their body movement to their avatar performance, and more frequent user/viewer interaction. However, Automated-Gesturing also had note-worthy features and mentions that allowed it to receive positive scores for its participants and even receive votes from participants who preferred this system. These factors include: natural and smoother (automated) gestures, good eye gaze and eye contact, consistent harmonized movement between the lower and upper body, and additional behavioural features like the finger-pointing gesture. Automated-Gesturing avatars and Motion-Tracked avatars show one kind of discrepancy due to the difference in gesture quality. But once the scenario the avatar performs in is taken into account, another kind of discrepancy emerges, which is discussed next looking at the effect of scenario on social fidelity.

The significant difference between the avatar control versions shows there still is some benefit in using VR systems to control these agents in social interaction platforms. However, the cost of these systems as well as the physical effort it requires from the human controlling the avatar might pose constraints on its use. A potential avenue for exploration is motion-captured data, which can be fed into the “text and affect behaviour” rule generator in order to allow user-driven avatars to trigger gestures based on speech-to-text, discussed further in the next chapter.

The effect of scenario on social fidelity

Our results further showed a significant effect of the scenario on Adaptation Gap scores, both in the initial avatar exposure and within-subjects analyses, where the Pitch scenario consistently had the worst scores than the Disco scenario. It is possible that due to the more serious and formal nature of the Pitch scenario, users were expecting more

from the avatar in terms of realistic behaviour, gestures, and emotions, in comparison to the Disco scenario which was more informal. When comparing the effect of scenario on the Adaptation Gap score, the Pitch and Disco scenario generated a large effect size (d=1.25), with the Disco scenario (*M* = 4.55, *SD* = 4.36) scoring higher than the Pitch scenario (*M* = -1.07, *SD* = 5.51). Interestingly, within the Pitch scenario the Motion-Tracked version still seemed to perform better than the Automated-Gesturing version for which the Adaptation Gap scores were almost always negatively rated. Compared to Automated-Gesturing, the Motion-Tracked version had a higher and often positive Adaptation Gap rating, suggesting that unintentional body movements and gestures in the Motion-Tracked avatar might have helped the positive perception of the avatars. However, there were no significant differences within the Disco scenario when comparing the Motion-Tracked version to the Automated-Gesturing version, where they both seem to be performing equally well.

The Disco scenario scoring higher potentially has to do with the more informal nature of the scenario where there are no significant expectancy on how avatars need to be behaving or gesturing. However, the score can also be influenced by the fact that each scenario was performed by different performers, so there was an effect of the performer on the scenario which we did not control for.

Moreover, we also found a significant effect of the interaction between Avatar Type and Scenario on Adaptation Gap scores - this means that depending on scenario, the avatar control types have a significant effect on user expectations. Going into more detail, we found that the scenario the avatars are performing in does have an effect, where the Disco scenario always had higher Adaptation Gap scores compared to the Pitch scenario. Similar to our results from the first avatar exposure, this indicates that in a more social scenario, both automated-gesturing and motion-tracked avatars were exceeding expectations. This could be due to the monotone delivery of the Pitch scenario's avatar, in which the person performing primarily memorized their lines, but had to reference their script. Whereas in the Disco scenario, the performer was allowed to improvise their lines. Despite the performer improvising their lines, measures were taken to make sure the improvised moments remained consistent across the different Avatar Type performances. This difference might have resulted in the avatars not meeting the expectations of the participants in the Pitch scenario, and often having negative Adaptation Gap scores. Moreover, the Pitch scenario includes less vocal emotional expression, resulting in less

facial emotional gesture generation by both Motion-Tracked and Automated-Gesturing avatars.

Another explanation could be the perception of the personality or gender of the performers affecting the results. Since the different scenarios had different performers controlling the avatars, this means the performances also had different personalities, vocal performances, and perceived gender of the voices affecting the performance. Particularly, the use of male and female voices from the performers only relegated to one particular scenario, where a male performer was used for the Pitch scenario and a female performer was used for the Disco scenario. Moreover, the performer in the Disco scenario was perceived as more extroverted and more expressive during the interaction, which was commented on by many participants. Compared to that, the performer in the Pitch scenario was calmer. Our qualitative results also showed that the preference for the expressiveness of the interaction partner could differ within some participants. This might affect the significant effect of the scenario on the Adaptation Gap score and could be a sign of more than just the effect of the scenario on the score, but also the specific performer. Future work can address this by having both performers participate in both scenarios (discussed further in Section 7.1). That being said, we do not believe this would affect the results we had on avatar control type as that was counterbalanced in our study.

Relation between Adaptation Gap and Believability

For *research question 4*, we have found that the Adaptation Gap scores were significantly correlated with the Believability scores of the avatars, suggesting that not meeting user expectations on Social Fidelity might result in low believability of the avatars, confirming our assumptions. We hypothesized similarly to Komatsu et al.'s (2012) hypothesis, that the more believable an agent is, the more likely users would continue to interact with it. Similarly to Komatsu et al., we found that the higher and more positive the Adaptation Gap score is, the higher the acceptance rate of the user will be. From this we further hypothesized that there could be a relationship between Believability and acceptance of an agent, and so there would be a relationship between the Believability scores and Adaptation Gap scores. Since our results show there is a relationship, Komatsu et al.'s statement that positive gaps garner a higher acceptance rate ends up being true for our conversational avatars. This potentially justifies the Adaptation Gap

approach's usefulness and applicability for evaluating the Social Believability of conversational avatars and ECAs.

However, the correlation was not particularly strong, indicating that there might be other factors that were not fully captured by the Adaptation Gap which affect believability. This would be expected considering believability also involves other factors including the physical appearance and narrative (Gomes et al., 2013; Gonzalez-Franco & Peck, 2018; Bizzocchi et al., 2013; Tanenbaum & Bizzocchi, 2009) whereas the Social Fidelity questions mainly focused on the social gestural behaviour of the avatar. This means that further evaluation is needed to understand the effect of these possible additional factors and their relationship with the Adaptation Gap. The small effect size can potentially be explained by the small sample size of our study, and so this could potentially be rectified with a larger sample size. However, the significant results show that indeed there is an important relation between Adaptation Gap and Believability of the avatars in Social VR environments.

6.1. Observations of the Open-Ended Questions

Looking at the qualitative answers from the surveys in the form of open-ended questions, we observe and potentially explain some of the results seen in the quantitative answers from our analysis in this section.

Firstly, we mention what was observed as common for both avatars (Motion-Tracked and Automated-Gesturing) in both scenarios (Pitch and Disco), which is that participants expressed that both avatars *lack facial expressions*. This coincides with Tanenbaum et al.'s (2020) conclusions where they also point out the lack of facial expression interfaces and features in top social VR platforms (p. 9). We believe this should be one of the first areas to be improved, as facial expressions are important in conveying emotional information and personality (Ekman & Oster, 1979), which are key factors for social interaction and believability (Loyall, 1997).

Moving on to the comments that were different between the two avatar control types. For the Automated-Gesturing avatar implementation participants expressed the avatar having more robotic movements in comparison to the Motion-Tracked avatar implementation. However, the Automated-Gesturing avatar implementation had better

hand gestures than the Motion-Tracked avatar implementation, despite the gestures repeating quite frequently. This might mean that repetition itself might break the illusion of life, where Automated-Gesturing systems might need to overcome this issue by including more variety in their gestures. Lastly, for the Automated-Gesturing avatar implementation participants noticed natural upper body movement for the avatar, but the lower half of the body was not up to the same quality and naturalness as the upper body. By this we mean that the behaviour and motion of the lower half of the avatar body was not as smooth or dynamic as the upper body. The lower body of the avatar does not animate or "update" as often as the upper body, and occasionally jitters when moving. Potential avenues for improvement can be to introduce postures and stances based on spoken word or gestures, introducing body weight shifting (shifting center of mass through the hips and the legs), and other motion captured movement. On the Motion-Tracked avatar implementation side, participants frequently expressed that this avatar had more natural gestures and body movement than the Automated-Gesturing avatar implementation. However, participants found the jittery movements of the avatar (due to the tracking quality of the VR hardware) to be noticeable and distracting. This shows there is also room for improvement on the motion-tracked avatar control to make the movements smoother.

Looking at the avatar's general performance (regardless of Avatar Type) in the scenarios, we start with the Pitch scenario. Participants noticed a disparity between the visual gender representation of the avatar and the voice of the avatar (i.e., masculine voice with a feminine-looking avatar) in the Pitch scenario. This difference might have affected the expectations in the performance or perception of the avatar. Prior research showed that voice can be an important cue for social capabilities and the perception of personality (Lee et al., 2005). The gender of the voice could also have an effect on user preference, depending on the similarity of the participant or attraction (Ozogul et al., 2013). However, we did not control these parameters per scenario, and did not examine the perception of personality of the different users acting in the two different scenarios. Further work might need to address this and its effect in the Adaptation Gap based on each scenario.

Apart from gender, prior studies also showed that ethnicity of the avatars can have an effect on user perception and preferences, following the similarity-attraction principle (Moreno & Flowerday, 2006). The performer having a hard accent with their verbal delivery, as the user who was controlling the avatar in the Disco scenario was from Germany and had a German accent. Few participants commented on this and mentioned

that it was difficult to understand the Disco avatar due to their accent, however this did not seem to affect the preference or the higher scores of social fidelity of the Disco scenario. Moreover, despite this avatar being easier to understand on the Pitch scenario, compared to the accent of the Disco scenario - participants commented that the monotone delivery of the avatar's verbal performance in the Pitch scenario made it hard to concentrate for long periods of time. Disco scenario was also, overall, more preferred. It is important to note that our study did not capture the ethnicity of our participants and did not examine the perception of ethnicity and its effect on participant scores. Future work should address the effect of participant's and avatar's ethnicity and evaluate it further.

Moreover, a few participants mentioned that the avatar performance in the Pitch scenario was calmer and much easier to understand in comparison to the avatar performance in the Disco scenario. This is potentially explained by the different people driving the avatars, with the one driving the Disco avatar being more expressive and extroverted person. For the Motion-Tracked avatar of the Disco scenario, participants mentioned that this avatar was more expressive and used a wide range of gestures, but certain participants commented on the gestures being too exaggerated and that the avatar was gesturing a little more than they expected, especially when compared to the Motion-Tracked avatar's performance of the Pitch scenario. This can be attributed to the personality of the avatar and the performer driving it, as we believe they are more outgoing and extroverted than the performer driving the Automated-Gesturing avatar in the Pitch scenario. Indeed, regardless of the Avatar Type in the Disco scenario, participants commented on the personality of the Disco avatar: they were more outgoing and enthusiastic, especially when it came to the dog photo the avatar would reference. Certain participants also mentioned liking the Disco avatar's wide range of vocals in their performance, though others preferred the more calm nature of the Pitch avatar's performance.

With the Disco scenario, participants noticed some interesting differences between the Automated-Gesturing avatar and the Motion-Tracked avatar used in the scenario. For the Automated-Gesturing avatar in the Disco scenario, participants referenced the avatar pointing and referencing the photo of their dog - as in the Desktop entry point setup mode the avatar pointing at the photo produces a red cube that assists viewers in seeing exactly what the avatar is pointing at, similar to a laser pointer (see Figure 29). Participants also found the avatar referencing the whiteboard drawings in the Pitch scenario to be a

memorable part of the scenario, especially when they interacted with it by facing it and pointing to specific parts of it. Participants particularly liked the dance moves of the Disco avatar's performance displayed at the beginning of the video, and mentioned the avatar performance interacts more with the viewer and the environment. For example, the avatar attempted to touch the viewer's shoulder and give the viewer a hug. This suggests that interaction with the environment as well as other participants might be improving the believability of the avatars. Participants also mentioned that the Disco avatar, wearing a leather jacket and a casual outfit, had a more believable outfit than the Pitch avatar who was wearing a cyberpunk-style outfit. This might also signal the importance of the avatars' visual appearance being consistent with the scenario or the environment it is in, although some participants also expressed preferring and liking the fantasy outfit of the Pitch avatar (see Figures 7 and 8).

Figure 29. Screenshot of Autimated-Gesturing avatar from Disco scenario pointing at a picture of a dog with a red box indicator.

6.2. Potential Factors Explaining Study Results

To summarize, the majority of the participants reported preferring the Motion-Tracked avatar from the Disco scenario as opposed to the others, which coincides with our results from Research Question 3 (see Table 9). There are many potential factors that could explain the preference of the participants, and thus the results from Research Question 3. Generally, the results can be explained due to the difference between the Automated-Gesturing and the Motion-Tracked avatar control types and their performance in terms of their Social Fidelity. Automated-Gesturing avatars have good automated

general gestures that are triggered when talking but tend to repeat themselves after a certain time. Some ways to improve the Automated-Gesturing avatar include, but not limited to, varying the gestures used during speech, a system to contextually tie gestures to the spoken word, and personalization of gestures and behaviours. Motion-Tracked avatars are the most expressive and varied due to their VR hardware and tracking nature, with the personality and quirks of the users being more accurately represented with this avatar. However, the avatars are not perfect and come with a few issues: jittery foot movement, and disassociated behaviour/movement between the upper and lower body. The Motion-Tracking avatar can be improved by additional systems that include, but not limited to, those that smoothen or match movement/behaviour between the upper and lower body, in addition to systems that can help with smoothening arm gestures in case of poor hand tracking.

In terms of the effects we've seen in the scenarios, we see some additional factors that further explain the results. We will summarize the top few we have noticed and previously mentioned:

1. *Personality and Expressiveness* - The individuals who volunteered to control the avatars in the videos for the study have different personalities and ways of expressing themselves. For example, the Disco avatar is more expressive and exhibits a more extroverted personality, while the Pitch avatar is calmer although not as expressive and exhibiting a more introverted personality. This might be causing differences in the preference of the users in terms of the Scenario effect. Prior work shows people's preferences might change in terms of personality, according to personality adaptation-convergence (similarity-attraction) (Lee et al., 2005) and divergence (complementary-attraction) (Isbister & Nass, 2000; Gurtman, 2009; Liew & Tan, 2016), which we did not control for.

2. *Voice* - the difference in voice delivery and range was observed in the comments as having an effect, particularly as most preferred the more lively delivery of the Disco scenario's Motion-Tracked avatar. This is further explained by participants calling the Pitch scenario avatar's verbal performance "robotic". We believe this is because the performer for the Pitch scenario's avatar memorized their lines, where the performer behind the Disco

avatar improvised their performance in certain places, thus providing a livelier delivery. The research team made sure that each performer's verbal delivery for both the Automated-Gesturing and Motion-Tracked avatar performances were as close as possible.

3. *Disco avatar's movement* - the Disco scenario's Motion-Tracked avatar was observed in the comments as having movement that is both varied and exaggerated. This was preferred by most participants over the more stationary and subdued movements of the Pitch scenario's Motion-Tracked avatar. The Automated-Gesturing version of both scenarios' avatars - while doing well in certain areas - generally came out with similar performances. One of our concerns looking back at the avatar performances in the videos was that the Pitch avatar in VR generally kept their arms and their gestures near their sides or around the waist level more often than the Disco avatar, which could explain the participant preferences mentioned above.

4. *Avatar setup* - we believe another factor, despite not being mentioned in the participant's comments, is the avatar setup inside the social VR platform. It is possible due to the use of the different avatars (outfit and other features), that the avatars could be technically set up differently. For example, the avatar used by the Disco scenario for both Automated-Gesturing and Motion-Tracked implementations could have better blendshapes set up for the mouth movement than the avatar used for the Pitch scenario, resulting in more expressive mouth movements.

6.3. Limitations

Despite our success with some of the results, there are some limitations to our study that can be addressed in future work and research. The main limitations are:

- Not gathering feedback from the actual people controlling the avatars (first person feedback) and how do they feel.

- The personalities of the avatars and the performers driving them potentially influencing the Adaptation Gap scores in each scenario type.

- The perceived ethnicity, personality and gender of the avatars potentially influencing the Adaptation Gap scores (due to having different people drive the Disco and Pitch avatars).
- Only including two kinds of scenarios as testing environments.
- The participant's exposure to one avatar at a time in the videos instead of exposing the participants with multiple simultaneous avatars.

One of our limitations is our study not controlling and accounting for the factors of personality and expression of the avatars and the performers. When seeing that participants showed preference for the Disco scenario as opposed to the Pitch scenario, this can be attributed to the individuals who volunteered to control the avatars in these different scenarios having different personalities and expressions. Participants commented on the performer in the Disco scenario as being more outgoing and extroverted than the performer driving the avatar in the Pitch scenario. The difference in personality and expressions might be causing differences in the preference of the users in terms of the Scenario effect, where prior work shows these factors being important and influential (Lee et al., 2005, Isbister & Nass, 2000; Gurtman, 2009; Liew & Tan, 2016). However, this would not affect our results in the effect of avatar control types, as our study was counterbalanced in terms of scenarios.

Our results show that the scenario has a significant effect on the Adaptation Gap score. Having only 2 scenarios on two extremes (informal and formal) means we do not know the extent of the influence scenario has on the score and/or how the scenario effects/interacts with Avatar Type, regardless of the performer, when it comes to the Adaptation Gap scores. To address this, we would need to design and implement more varied scenarios where some can cover categories in between the informal and formal extremes, for example: casual, semi-casual, business casual. In addition, due to the disparity of the avatar performances - it is worth attempting to have the avatar performers recorded for each Avatar Control Type for each scenario. Meaning, for example, the more extraverted performers and the introverted performers should be recorded performing in both informal and formal scenarios as an additional counter balance to accommodate both the effect of scenario and the effect of the performance (vocal and gestural). This would also allow us to track the effect of gender, personality, and expression.

In order to properly ascertain one's reception and reaction to multiple avatar exposures, and hence the impact of avatar control type order when it comes to exposure, participants would need to ideally be exposed to the avatars in the same session/experience and ideally in person. In our videos, we expose the participant to one avatar control type and then have the participant later watch another video of another avatar control type. The videos limit the impact of participants experiencing both avatars at the same time and the validity of that experience when it comes to their genuine response and reception to both avatars. Our research did not examine simultaneous avatar exposure as we needed to examine interactions with a single avatar before examining multiple simultaneous avatar exposure in a social VR platform. Our research is the first study to focus on the effect of the Adaptation Gap on different avatar control type exposure and the use of Social Fidelity scores to capture user reception. In the light of the results of our current work, simultaneous avatar exposure can be addressed in future work (see Section 7.1).

6.4. Recommendations on Multiple Conversational Avatar Exposure

We would like to end our discussion chapter with our recommendation to developers concerning the use of multiple conversational avatars (also applicable to embodied conversational agents) of different types in social VR platforms. While our research did not study multiple simultaneous exposures to different avatar control types, it did cover exposure to different avatar control types with counterbalancing. The results can help us predict what would happen if users were to be exposed to these avatars simultaneously within the same environment or context, knowing that the avatar's user's personality, behaviour, and social habits are also factors impacting immersion. Seeing as the results show that the Automated-Gesturing avatar version is generally weaker in terms of the Adaptation Gap than the Motion-Tracked version, we confirm Komatsu et al.'s (2012) conclusion and state as our *first policy*:

It is recommended to not use conversational avatars that have an Adaptation Gap that is *negative generating,* that is, avatars that cannot meet the user expectations.

Our results indicate that the scenario or the context within which an interaction takes place matters and affects the user's adaptation gap. So much so that even the

Automated-Gesturing avatar version can generate a positive adaptation gap - though generally weaker than the one generated by the Motion-Tracked avatar - when encountered within a fitting and engaging scenario as seen in the Disco scenario. The automated-gesturing features (and other believability features) of the Automated-Gesturing avatar have the potential to create a viable avatar for positive experiences within social VR platforms - especially when paired with an engaging scenario.

That being said, *our second policy* is that developers are recommended to improve the experience of non-VR or non-motion-tracked avatar implementations by potentially doing either or both of the following:

1. Improve the believability of non-VR or non-motion-tracked avatar implementation (e.g., Desktop or automated-gesturing avatar versions) with more and better believability features; and/or

2. Cater the "new user" experience by showing the avatar with the relatively weakest Adaptation Gap first before the strongest.

Explaining our second policy's first point further, we understand it is extremely difficult to have a non-VR (non-motion-tracked) implementation of an avatar match the adaptation gap scores that can be generated by a Motion-Tracked avatar, especially given the motion-tracked nature of VR technology. However, the intent here is not to match but *to reduce* the believability gap (difference in adaptation gap scores) between the two avatars. The goal is to have the adaptation gap score for non-VR avatars improve in such a way that it reduces the adaptation gap difference between the two avatars. We believe such improvements can eventually remove the need for platforms to worry about avatar exposure order (related to policy 2 point 2) assuming both avatars can produce similarly high adaptation gap scores for the user.

Moving on to the second point of our second policy. We have seen from the Adaptation Gap concept (see Section 2.3) utilized in our research that an initial exposure producing a high social fidelity score followed by a lower score after exposure would result in a negative Adaptation Gap, which is not desirable. Conversely, an initial low social fidelity score followed by a high score would result in a positive Adaptation Gap score. This approach can also be applied to the general adaptation gap experience of a user exposed to multiple avatars, assuming both avatars generate positive adaptation gap

scores. Using a similar approach, we substitute social fidelity scores with adaptation gap scores of the individual avatars (AG_{av2}, AG_{av1}) in order to get the general or overall adaptation gap score (AG_{final}) for the user, denoted as: $AG_{final} = AG_{av2} - AG_{av1}$. This represents the user's experience of being exposed to both avatars at the same time. With this in mind, the same rules carry over if a user is exposed to a weaker avatar (low AG_{av1} score) and then meets a stronger avatar (high AG_{av2} score). This would produce a better outcome for the user, with an overall positive adaptation gap score ($AG_{final} = AG_{av2} - AG_{av1}$; producing a positive AG_{final}). This is why we recommend exposing users to the weakest avatar first, as exposing a user to a strong avatar first (high AG_{av1} score) followed by a weak avatar (low AG_{av2} score) will produce an overall negative adaptation gap for the user ($AG_{final} = AG_{av2} - AG_{av1}$; producing a negative AG_{final}), impacting their reception and experience with the platform.

When gazing forward at the potential future for ECAs, conversational avatars, virtual world platforms, and metaverses, a few key points arise that coincide with the results of our research. These have been compiled into recommendations for developers. *First and foremost*, there is an opportunity to work on conversational avatar exposure order for new users - whether as an onboarding tool (designed to streamline the process of integrating new users into the platform) or by catering first introductions with other conversational avatars through filters, smart matchmaking (server selection), visibility, etc. Similar to how a user would get a positive or negative Adaptation Gap score based on an avatar's believability - a user would similarly get an overall Adaptation Gap score based on the believability of two avatars and in which order a user would experience them. We recommend developers focus on exposing to new users the relatively weakest avatar in terms of social believability first (see policy 2.2 above) and order the exposure from weakest to strongest. Thus, increasing the chances of an overall positive Adaptation Gap for users, increasing the acceptance and reception of the different avatar control types of a social VR platform, in addition to improving the experience.

Our *next recommendation for developers*, and probably one that requires the most effort, is improving the social believability of non-VR or non-motion-tracked conversational avatars. This can be done by focusing on avatars, like the automated-gesturing avatar, by adding new features and improvements. A select few critical improvements we recommend are: 1) introducing personality and mood to conversational avatars, 2) emotion and facial expressions during conversations, and 3) improving the variety and

appropriateness of automated gestures. Introducing personality and mood in the form of personality models comes from Zammitto et al.'s (2008) work. For our recommendation, we foresee that this can be in the form of stances, postures, tailored gestures and facial expressions. One way to achieve this is possibly through a system that can allow users to select various poses and postures, and even select specific reactions (gestural or emotional) based on spoken words that the system can detect. This can be done using the system already in place to trigger gestures during speech. Such an improvement would help differentiate behaviours of conversational avatars and potentially allow to also introduce more emotions into avatar facial expressions, which ties into our next recommendation. Our research observed frequent comments on the lack of emotional and facial expressions in conversational avatars during conversations. This matches Tanenbaum et al.'s (2020) conclusions pointing out the lack of facial expression interfaces and features in top social VR platforms (p. 9). Some include the platforms our researchers explored in Chapter 4. This underlines the importance of it being one of the first features that need to be added to conversational avatars. Bates himself stresses that for "believable agents" to be convincing, they need to have accurately timed and explicit emotions expressed within a given social context. Bates continues that if agents do not "react emotionally to events", then humans will not be convinced of their genuineness (Bates, 1994, pg. 2). Finally, some of the comments from our study highlight the amount of repeated gestures, which we observed to also occasionally include gestures that do not quite fit with the spoken context of the avatar. A good improvement would be to introduce more variety of gestures to the automated-gesturing system and then introduce a system to allow the matching of a select amount of gestures to frequently repeated words or phrases that the avatar system can keep track of through voice. Since the virtual world platforms that use automated-gesturing avatars already have audio analysis systems to trigger gestures and lip sync, it is beneficial to take advantage of such systems to also keep track of words and phrases with the help of tools such as "speech-to-text". Examples of this can be seen in Ali et al.'s (2020) and Lee & Marsella's (2006) work. However, we believe there is a more interesting approach inspired by Ali et al. and Lee & Marsella for improving ECA and conversational avatar believability by also providing some individuality, which we will discuss a little further below.

Our results show that the Adaptation Gap effect is present on both avatar control types as there are no adaptation gaps at a value of zero. As discussed earlier in this

section, there is potential for the effect to multiply and worsen as more varieties of avatar control types are used on a platform. Currently, with platforms like Tivoli Cloud VR, there are only two avatar control types. But if we were to move closer to the realization of a global metaverse like the one proposed by Meta (Newton, 2021; Ravenscraft, 2021), it is possible to imagine such platforms will have multitudes of entry point setups for such metaverses. If there are multiple entry point setups for metaverses then we can imagine there will be multiple options to control or embody the avatars. As a result, this not only multiplies the variety of avatar implementations but will also multiply the occurrence of users needing to adapt to each avatar control type a user encounters. We believe users can be led to fatigue due to the mental strain of having to adapt to each avatar's performance and behaviours, especially if: worst-case scenario the users experience a constant switching of adaptation gap values (from positive to negative, negative to positive), or best-case scenario they experience a frequent adaptation to each avatar control type. Part of this is explained by Komatsu et al. as a user's mental adaptation having to do with the mental models people create when meeting an agent or avatar (Komatsu et al., 2012, pg. 109), which they specify involves various factors like the agent's appearance, the agent's behaviour, and the user's own preferences. In addition, when it comes to believability there are additional factors like those from Verhagen et al.'s (2013) requirements for good believability. When any factors end up changing too frequently or too drastically, that frequent adaptation of the mental model would be the source of fatigue for a user. Stemming from this is the *final recommendation for developers,* which is to reduce the amount of ECA or conversational avatar implementations in order to unify them into a single front-end (or user-facing) avatar implementation. Specifically, users can still opt to use VR technology, their keyboard and mouse, or their mobile phones, but the users on the receiving/observing end should receive a consistent and predictable performance of the avatars regardless of the entry point setup the initial user has chosen. For this, we propose a unified front-end ECA performance system, which is also a potential future research direction that we will discuss in the next section.

6.4.1. Conversational Avatar Adaptation Gap and Gaming

We believe our research can guide not only the future development of Conversational Avatars and ECAs but can also be used in the development and improvement of video game characters or NPCs. Video games use digital characters

controlled by the computer that engage in interactions and conversations. These characters that are controlled by the computer or game AI are called Non Playable Characters or NPCs. Oppositely, video game characters that are instead controlled by the player would be named Playable Character or just Player Character. These characters can benefit from the research and lessons on ECAs and Social Believability, including the Adaptation Gap and Social Fidelity gap covered by this thesis. We see video games benefitting from the utilization of evaluative categories to evaluate video game characters, or using the Adaptation Gap to judge or predict the acceptance rate of new character designs, or using Social Fidelity or Believability research to enhance conversational features of video game characters and improve player immersion.

Game Characters use similar rendering techniques (e.g., using Global Illumination, OpenGL libraries), technology for creating 3D models of humanoids (e.g., Zbrush, Blender and Maya software), technology for animating characters (e.g., Blender and Maya software), and control schemes (e.g., VR controller, game controller, keyboard and mouse controls) similar to social VR platforms - with the biggest difference being that Game Characters are primarily made for gameplay and game related interactivity. But even with this difference - due to the many similarities, we believe they can benefit and improve just as much as ECAs or Conversational Avatars utilizing Social Fidelity gap and Believability research. Just looking at Social Believability research alone shows us the variety of research and interest in Social Believability for video games and game characters (Prada & Paiva, 2005; Zammitto et al., 2008; Alfonso & Prada, 2009; Verhagen et al., 2013; Morgan & Papangelis, 2015). This is not just for NPCs but can also be utilized for player characters. Especially for games that are multiplayer VR enabled games where multiple users using avatars play with or against each other online using VR. Examples include Rec Room (Rec Room Inc., 2016), Star Trek: Bridge Crew (Red Storm Entertainment, 2017) and Space Team VR (Cooperative Innovations, 2020), among others. These games and more utilize avatars and can benefit from having their believability and social fidelity improve. Even in these games - being able to see each other and communicate in the game is still part of the game package despite the reduced importance of social behavioural and conversational features when compared to the importance of fun gameplay. Nevertheless, developers can benefit from improving the immersion and believability of not only NPCs but Player Characters as well - which can then help improve enjoyment or the user acceptance rate of these avatars. All in all, we see the Adaptation

Gap and Social Fidelity gap approach to conversational avatars being beneficial for the future development of video game characters.

Chapter 7.

Future Work

Our research results show promise in the uses of the Adaptation Gap to gauge the believability of multiple conversational avatars. It also provides opportunities for future research regarding the Adaptation Gap, believability, and conversational avatars. For example, researching believability features and improvements that can bridge the believability gap (or difference in Adaptation Gap scores) between two avatars. That being said, we also acknowledge our study and research are constrained by the limitations expressed in Section 6.3. This means there is also future work to be done in further refining, clarifying, and perfecting some areas of the research. For example, measuring Adaptation Gap scores of two conversational avatars (motion-tracking and automated-gesturing) with the users experiencing the avatars in person in virtual reality instead of watching a video. These ideas and more are discussed in this section.

The video based study was taken over participants visiting in VR to mitigate some confounding variables that might impact the study results. These include: participants getting motion-sick in VR, the variance in the time it takes to acclimatize, participants differing in how long they can be comfortable in VR using VR hardware, COVID-19 being prominent at the time of the study and so was a big health concern for participants (issues like sharing headsets, cleaning, sanitizing, ethics approval for an in-person study), and finally the time it takes for participants to get set up for a VR session versus accessing an online survey and viewing a video. Lastly, it was very important for us that the experience remained consistent across all sessions. This led to the video approach that allowed us to record videos and facilitate consistency as opposed to attempting to facilitate consistency with participants visiting the social VR platforms in VR. The videos allow us to provide a professional and consistent point of view for the viewers, as opposed to dealing with individual participant issues. Each participant might not be used to moving in VR, understanding the controls of various VR platforms and hardware, and either interrupting performers or being interrupted/distracted while observing. Ultimately, recording videos of avatar performances was the best solution for the study considering the circumstances. Despite these positives, we recognize that video recordings of an avatar's performance and interaction are not the best medium with which to have a participant experience a

conversational avatar, albeit allowing us to have control over the participants' experience. Some shortcomings include the passive nature of the participants experiencing the interactions, and the “observer” in the video not being a direct one-to-one relation to the participant who is supposed to be the true observer. This made the evaluation from an interaction and behavioural perspective difficult, in large part due to the passive nature of the video form. One way to address this would be to repeat the study but have the participants visit the social VR platform (and the avatars) in person using VR technology and an assigned avatar so they can experience it in real-time.

This is due to the main argument being that a stronger connection to the performances and interactions could have been made, with results impacting the Adaptation Gap scores, if the participants were to experience the avatars in situation - meaning, in this case, wearing a VR headset and visiting Tivoli Cloud VR in those same environments with the aforementioned avatars, experiencing it in real-time. Since the time of the completion of our study Covid 19 protocols have loosened but also Tivoli Cloud VR has closed, meaning future studies would have to be in other High-Fidelity forks (copies of a code repository), which Tivoli was a fork of. This includes platforms like Vircadia (vircadia.com, 2023) and Overte (overte.org, 2022). This brings us to the *first future research direction*, which is to recreate the Adaptation Gap study but have the participants experience the avatars in VR and inside a social VR platform environment instead of a video. However, this would come with its own challenges of executing the study in person, with some of the challenges being the technological overhead required for setup and deployment of the study, organization of the performing avatars, and the likelihood of a lower participant turnaround.

7.1. Addressing the Limitations

One of the limitations from Section 6.3 was the fact that our study did not factor in the personality and expressions of the performers and the avatars, which potentially could impact the Adaptation Gap scores. Our *Second proposed future work* and *research direction* would be the need to include questions and study designs that allow to take into account the different user’s personality and behavioural expressions, in addition to potentially counter balancing these factors in order to see their effect on the Adaptation Gap score.

Additional areas that can be included in our second proposed future work are those that were not covered by the study but were brought up during the open ended questions sections were the topics of participant's and avatar's gender and ethnicity. Additionally, the study did not capture the ethnicity, personality and gender of our participants and did not examine the perception of ethnicity, personality and gender, and its effect on participant scores. As this could be a potential factor affecting Adaptation Gap scores and the correlation between Believability and Adaptation Gap, *future work and research directions* should address the effect of participant's and avatar's ethnicity and gender and evaluate its effect. The effect of ethnicity and gender can be combined with our previously discussed limitation of the effect of personality and expression of the performers and avatars, potentially by counterbalancing the performers and avatars, in addition to making sure each performer/avatar performs in every scenario designed for the study. Expanding this further, to further understand the effect of scenario on the Adaptation Gap score with the additional features (ethnicity, gender, etc.), *future work and research* should include more than two scenarios and ideally should encompass more variety in terms of formal and informal scenarios, with some being somewhere in between.

Finally, our research did not gather feedback from the actual performers (those driving the avatars) as first-person feedback on their experience and how they feel about their avatar's social fidelity. Future work should include additional open ended or interview questions just for the avatar performers in order to capture the Social Fidelity Gap not just for the observers but also for those performing and controlling the avatars. This can allow us to examine the extent in which the users of the avatar control mechanisms on how well they think it captures their behaviors, and whether it affects the presence they feel in the environment.

Expanding on the direction of having participants be in VR inside the social VR platform, we have discussed in Section 6.4 the possibility of the adaptation gap for a user changing either severely and/or frequently with multiple avatar exposures. While it was our recommendation to reduce the frequency of exposure to different avatar control types and to start with the weakest implementation, this recommendation was based on our current findings. To strengthen the argument, our *third proposed future research direction* is a study that should be carried out to address and capture the change in a user's Adaptation Gap score when exposed to multiple conversational avatars within the same session. This approach would help to more accurately capture the real-world experience

of users (contrasting the study involving watching videos) and the reality of users meeting multiple avatars within the same visit or session, and so hypothetically would produce more accurate Adaptation Gap results. In addition, we are also curious if the medium and technology through which users are experiencing and interacting with the virtual worlds (e.g., through a VR headset or a desktop monitor) also have an impact on Adaptation Gap scores. Our *fourth proposed future research direction* is to do a follow-up study to investigate the effect of the observer's entry point setup while interacting with these avatars on the Adaptation Gap. For example, do users in VR using a VR headset observing a conversational avatar have a better connection with the observed avatar than those observing through a monitor? Viewing through VR can potentially provide better easing of suspension of disbelief, thus potentially allowing users to more easily accept conversational avatars and as a result could produce higher positive Adaptation Gap scores.

7.2. Expanding the Research

First and foremost, we see from the results that there is a Believability Gap, or a difference in Adaptation Gap scores, between the Automated-Gesturing and Motion-Tracked avatars. In order to unify the experience for users, especially when it comes to metaverses, developers would need to bridge the Believability Gap and have that difference be as little as possible. The easiest approach would be to increase the believability of the automated-gesturing implementations in social VR platforms by improving and adding more behavioural and believability features. To start, what are the best or most important features for developers to implement first that can provide the most significant difference when it comes to bridging the believability gap? This would be our *fifth proposed research direction*, which involves researching, implementing, and testing new behavioural and believability features for the Automated-Gesturing avatar. This would require researching and utilizing new AI techniques, for example, those that observe correlated movements to spoken words, which would allow to trigger the gestures again based on the spoken word. It also entails testing the Adaptation Gap outcome from the implemented features against the Motion-Tracked avatar. Some of the top features to try first are the ones that were recommended at the beginning of Chapter 7. These include introducing personality and mood to conversational avatars, emotional and facial

expressions during conversations, and improving the variety and appropriateness of automated-gesturing gestures.

Another one of our recommendations to developers from the beginning of this chapter was about reducing the number of avatar control types users are exposed to in order to reduce mental strain. Other potential avatar control types and mechanisms can include using the screen and touch input of a mobile phone, a tablet, a screen in an electric car, or a large interactive screen or wall. Our recommendation is influenced by our results and Komatsu et al.'s (2012) research, that suggests different control mechanisms can lead to differences in the social behaviour of these avatars, which could lead to low believability and mental strain. We hypothesize that multiple changes of expectations and getting used to different avatar behaviours can lead to fatigue and mental strain. While we included this idea as part of our final developer recommendation, we believe this concept is worth confirming and researching further. This would be our *sixth proposed research direction*, where a study would need to be carried out to measure fatigue and mental straining for two kinds of possible situations users can be exposed to. Part of the challenge is to figure out what the minimum (currently 2 avatar control types) and the maximum amount of avatar control types a user can be exposed to before feeling mental fatigue and straining. The two kinds of situations are: 1) having multiple avatar control types - where each avatar control type has radically different social fidelity levels and so produces radically different Adaptation Gap results; 2) a situation where there are multiple avatar control types exposed to the user in the same sessions but do not necessarily have radically different Adaptation Gap results. The study would require the utilization of sensors and have participants follow up with activities and/or interview questions that measure the fatigue and mental strain of the user. This potential future research would allow researchers and developers to understand better the effects of user exposure to multiple conversational avatars. It should also help support the argument for the need to reduce the number of discernable avatar control types that a user can visually distinguish during the avatar's performance or conversations.

Expanding upon our final developer recommendation from earlier in the chapter is the idea of a unified front-end ECA performance system. This is our *seventh proposed research direction* and probably the most exciting one. To give a brief explanation, it is a system that unifies the output performance regardless of the input technology or entry point setup to deliver a consistent performance that does not inherit the drawbacks of the

input technology used by the user. The foundation of this idea lies in the concept of "binding the pair" termed by DiPaola & Turner (2008) in their work. The idea here is to match or "bind" the behaviour and movement intricacies of a user to their virtual representation as closely as possible. One way to do this is with better tracking technology to accurately track the movement and behaviour of a user. However, for our research, we suggest focusing on the Automated-Gesturing avatar as that avatar, from our results, has the weakest Adaptation Gap value. To improve the "binding" of the Automated-Gesturing avatar - more technology and systems would need to be implemented. Part of the reason to focus on the Automated-Gesturing avatar is both the promise and potential of the automated-gesturing system based on how it impacted users' opinions and scores in our study. This is despite the frequent participant comments about the system having the issue of repeated gestures (see our second developer recommendation from this chapter).

To improve the Automated-Gesturing avatar and bridge the believability gap while working towards a unified system would require some research and development. Inspired by Lee & Marsella's work (2006), this is where we propose a potential avenue for exploration to be in motion-captured data which can be fed into a "text and affect behaviour" rule generator. This is done in order to allow user-driven avatars to trigger gestures based on speech-to-text when using an avatar that is not tracking the motion of the user (Non-VR avatars). One of the motivations for this approach is to have avatars like the Automated-Gesturing avatar perform with gestures unique to the user, captured from their VR sessions, and have users be able to visit social VR platforms without relying on cumbersome VR equipment. Using machine learning, the system can extract behavioural properties and patterns from a user's performance. The extraction occurs when users are in VR using a Motion-Tracked avatar, with the data then integrated into a dynamic and fast animation system tied to social and verbal inputs. The social and verbal inputs also include data from the user's social and verbal patterns and behaviour that is fed along with the motion-capture data into the machine learning system. The ideal result here would be a recreation of a user's Motion-Tracked avatar behaviour and performance without the user driving it in VR or even necessarily driving the avatar at all. This ultimately results in observers not being able to tell which avatar control type or input technology the avatar is using. This is especially useful for both non-VR avatar implementations but also use cases that involve the user being away from their computer or "away-from-keyboard". There's potential for users to use their avatars, for example, at virtual kiosks or as virtual assistants

at digital conventions that can provide help and guidance while the actual guide is away. Another example is avatars would be able to behave with a high degree of accuracy and immersion when the user no longer wants to be in VR hardware for their session. It is vital then that If that user decides to switch back to being in VR that the transition on the avatar side should be unnoticeable to any observers.

Finally, our *last recommended research direction* is one that aims to delve further into explaining our results for research question 4. Our results showed that Believability showed statistical significance in being able to predict Adaptation Gap scores. However, it had a small effect size, which means there are likely other factors also explaining the relationship between Believability scores and Adaptation Gap scores. Our research can be further extended by looking at various levels of avatar embodiment and how that boosts or hinders social believability and social presence. Taking a cue from Gonzalez-Franco & Peck's (2018) work, we believe the level of embodiment or the perceived level of embodiment of the avatar is probably one of the other potential factors impacting Believability, the Adaptation Gap, and the relationship between the two. We believe the level of embodiment also in tandem goes well with Weidner et al.'s (2023) work that discusses different avatars and avatar representations (e.g., disembodied and fully embodied representations). We propose a future research study that investigates how different levels of avatar embodiment and representations affect Believability scores and Adaptation Gap scores, and if the factor could further explain the relationship between Believability and Adaptation Gap scores. To explain further, there are two studies that could be carried out: one, for example, that looks at disembodied avatars (floating heads and hands, etc.), fully embodied avatars (like the Tivoli Cloud VR avatars), and any other level of embodiment or representation in between, and see how those affect the scores. Another study could also look at the potential factor of the level of accuracy of matched versus unmatched movement between the user and the avatar, and how it is perceived by observers in the form of Believability and Adaptation Gap scores. This was inspired by the performance of the Automated-Gesturing and Motion-Tracked avatars of Tivoli Cloud VR, but we believe this study should focus more on the Motion-Tracked avatar implementations instead. Motion-Tracked avatars could either perform with fully matched movement to what the user is performing or could be approximating, smoothing, and/or transforming user movement to better suit the avatar's capabilities or "character". An example of a user's avatar that would require the user's movement to be transformed to

fit the avatar would be if a user were to embody a horse avatar instead of a humanoid. We believe studying how the different approaches to matching or translating user movement and how that affects Believability and Adaptation Gap scores could provide further insight into its effect on the relationship between the two scores. That being said, we acknowledge that there are potentially more factors that could better explain or predict the relationship between Believability and Adaptation Gap. However, we believe the factors related to the level of avatar representation and embodiment plus the level of matched movement are good starting points.

Chapter 8.

Conclusion

To summarize, our research aimed to evaluate different avatars controlled by humans in social VR environments through the lens of Social Believability. This was done with the help of the Adaptation Gap concept, where a believability gap can be formed between the motion capture-driven (using VR) and non-motion capture-driven (automated-gesturing) avatars. The study sought to answer two overarching research questions: 1) Does an observer in a shared virtual environment notice a Social Fidelity gap among human-driven conversational avatars with different avatar control mechanisms? 2) Are the Social Fidelity Gap scores for the conversational avatars correlated with the perceived believability? To answer this, we first came up with a taxonomy that would help us better analyze and compare the social believability of human driven conversational avatars using different avatar control mechanisms. We used this taxonomy to pick the most suitable to run a controlled quantitative study.

The study preparations involved finding a suitable virtual social world platform and avatars, which would be used to record videos for the study. This involved exploring six VR platforms and their avatars and selecting the most suitable one for our study. A document of embodiment categories and feature wish lists was created to evaluate the platforms and their avatars. In the end, Tivoli Cloud VR was chosen due to a) a generally stronger level of embodiment for both avatar control types compared to other platforms, b) inclusion of Automated-Gesturing features for Desktop avatars when conversing, c) volume-based blend shape lip sync, including automated eye gaze and eye contact when conversing. Four videos were then recorded with each avatar control type and scenario combination (2x2) using Tivoli's avatars and were later used inside our survey. The survey was structured with an initial video test plus test questions. After which, if qualified, participants would move on to two blocks of the study and finish with some final questions. Each block involved looking at an avatar photo of one avatar control type and scenario and answering a few questions, then watching a video and answering some questions afterward. The second block repeated the process but showed an alternate avatar and scenario.

Our results showed that the Adaptation Gap was correlated with the Believability scores and the motion-tracked avatar control exceeded observers' expectations with their performance. When compared to the automated-gesturing avatar, the motion-tracked avatar performed better in comparison. We also found a significant effect of scenario, which might influence agent behaviour, where a social scenario (Disco) created a larger positive Adaptation Gap compared to its formal counterpart (Pitch).

For future work, we plan to further test the effect of the Adaptation Gap with users while immersed in the social VR platform instead of watching a video capture of the VR event. We also plan for the test to include a static avatar for comparison. Comparing the differences between the different controlled avatars, when both agents are collaboratively or socially interacting together, might also help us examine any changes in the expectations and perceptions while seeing both agent types together. Moreover, understanding the Adaptation Gap of these systems from the users' perspective, in addition to the observers' perspective which this paper focused on, would be important to examine.

The results allowed us to show the importance of ECA or conversational avatar behaviour control in social VR platforms. In addition, it also allowed us to show the need for carefully catering conversational avatar exposure to visitors in order to increase the visitor's acceptance rate, meet their expectations and improve their virtual world experience. All in all, the paper also serves as an encouragement for VR Social Platform developers to understand the importance of gesture quality and scenarios to guide their endeavours to improve accessibility and immersion for their users. Studying the experience of an Adaptation Gap among multiple human-driven avatars has allowed us to inform developers on how to create better believable, accessible, and more fulfilling avatars/agents and experiences. One key recommendation stemming from the study is to research machine learning of user gestures and behaviours in VR and combine it with speech-to-gesture rule generation for triggering user-recorded gestures when using automated-gesturing avatars. Considering how all the "magic" happens over the internet or on servers - there should be no reason why the industry and its developers could not strive towards a unified ECA or conversational avatar presentation. Especially to the level where end users as observers should not be able to easily tell the difference between different avatar control schemes.

References

Afonso, N., & Prada, R. (2009). Agents That Relate: Improving the Social Believability of Non-Player Characters in Role-Playing Games. *Lecture Notes in Computer Science Entertainment Computing* - ICEC 2008, 34-45. doi:10.1007/978-3-540-89222-9_5

Alexander, A. L., Brunye´, T., Sidman, J., & Weil, S. A. (2005). From gaming to training: A review of studies on fidelity, immersion, presence, and buy-in and their effects on transfer in pc-based simulations and games. *DARWARS Training Impact Group*, 5, 1–14.

Ali, G., Lee, M., & Hwang, J.-I. (2020). Automatic text-to-gesture rule generation for embodied conversational agents. *Computer Animation and Virtual Worlds*, 31(4–5), e1944. https://doi.org/10.1002/cav.1944

Argelaguet, F., Hoyet, L., Trico, M., & Lecuyer, A. (2016). The role of interaction in virtual embodiment: Effects of the virtual hand representation. *2016 IEEE Virtual Reality (VR)*, 3–10. https://doi.org/10.1109/VR.2016.7504682

Aseeri, S., & Interrante, V. (2021). The Influence of Avatar Representation on Interpersonal Communication in Virtual Social Environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5), 2608–2617. https://doi.org/10.1109/TVCG.2021.3067783

Bailey, H. (2017). the OBS Project Contributors. *Open Broadcasting Software*. Retrieved from https://www.obsproject.org/

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01.

Bates, J. (1994). The Role of Emotion in Believable Agents. *Communications of the ACM*, 122–125.

Bevacqua, E., Richard, R., & De Loor, P. (2017). Believability and Co-presence in Human-Virtual Character Interaction. *IEEE Computer Graphics and Applications*, 37(4), 17–29. https://doi.org/10.1109/MCG.2017.3271470

Bizzocchi, J., Nixon, M., DiPaola, S., & Funk, N. (2013). The Role of Micronarrative in the Design and Experience of Digital Games. *DiGRA Conference*. http://www.digra.org/wp-content/uploads/digital-library/paper_181.pdf

Business Wire. (2023, January 17). *Global metaverse market analysis report 2022: Expected revenue of $700 billion by 2030 - researchandmarkets.com*. Retrieved July 10, 2023, from https://www.businesswire.com/news/home/20230117005761/en/Global-Metaverse-Market-Analysis-Report-2022-Expected-Revenue-of-700-Billion-by-2030---ResearchAndMarkets.com

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., . . . Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* - SIGGRAPH 94. doi:10.1145/192161.192272

Cassell, J. (2000). *Embodied conversational agents* / edited by Justine Cassell [and others]. MIT Press.

Chaves, A. P., & Gerosa, M. A. (2018). Single or Multiple Conversational Agents?: An Interactional Coherence Comparison. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3173574.3173765

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic [Google Scholar]

Dass, S., Dabbagh, N., & Clark, K. (2011) Using virtual worlds: what the research says. *Quaterly Review of Distance Education,* 12(2), 95.

DiPaola, S., Collins, C. (1999). A 3D Natural Emulation Design to Virtual Communities. *ACM SIGGRAPH 99 Conference Abstracts and Applications*, SIGGRAPH '99, Los Angeles, 208-208

DiPaola, S., & Turner, J. (2008). Authoring the Intimate Self: Identity, Expression and Role-playing within a Pioneering Virtual Community. *Loading...*, 2(3), Article 3. https://journals.sfu.ca/loading/index.php/loading/article/view/40

DiPaola, S., Turner, J., & Browne, P. (2011). Binding the pair: Making a historical case for avicentric self-expression within 3D virtual communities. *International Journal of Web-Based Communities*. Vol 7. 157-173. 10.1504/IJWBC.2011.039508.

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology*, 30(1), 527-554.

Fortune Business Insights. (2020, September). *Virtual Reality Market Share, Growth: VR Industry Trends [2020-2027].* Retrieved December 9, 2020, from https://www.fortunebusinessinsights.com/industry-reports/virtual-reality-market-101378

Garau, M. (2003). *The impact of avatar fidelity on social interaction in virtual environments*. [Doctoral, UCL (University College London)]. https://discovery.ucl.ac.uk/id/eprint/10103871/

Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, 304–316. https://doi.org/10.1016/j.chb.2019.01.020

Gomes, P., Martinho, C., Paiva, A., & Jhala, A. (2013). *Metrics for Character Believability in Interactive Narrative*. 8230. https://doi.org/10.1007/978-3-319-02756-2_27

Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI*, 5. https://www.frontiersin.org/articles/10.3389/frobt.2018.00074

Greenwald, S. W., Wang, Z., Funk, M., & Maes, P. (2017). Investigating Social Presence and Communication with Embodied Avatars in Room-Scale Virtual Reality. *Communications in Computer and Information Science Immersive Learning Research Network*, 75-90. doi:10.1007/978-3-319-60633-0_7

Gurtman, M. B. (2009). Exploring Personality with the Interpersonal Circumplex. *Social and Personality Psychology Compass*, 3(4), 601–619. https://doi.org/10.1111/j.1751-9004.2009.00172.x

Hashemian, M., Prada, R., Santos, P. A., & Mascarenhas, S. (2018). Enhancing Social Believability of Virtual Agents using Social Power Dynamics. *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 147–152. https://doi.org/10.1145/3267851.3267902

Isbister, K., & Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, 53(2), 251–267. https://doi.org/10.1006/ijhc.2000.0368

IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp. URL https://www.ibm.com/spss

Immersive VR Education Ltd. (2016). *Engage VR* [Computer software]. Retrieved November 14, 2023, from https://engagevr.io/

Kokkinara, E., & Slater, M. (2014). Measuring the effects through time of the influence of visuomotor and visuotactile synchronous stimulation on a virtual body ownership illusion. *Perception*, 43(1), 43-58.

Komatsu, T., Kurosawa, R., & Yamada, S. (2012). How Does the Difference Between Users' Expectations and Perceptions About a Robotic Agent Affect Their Behavior? *International Journal of Social Robotics*, 4(2), 109–116. https://doi.org/10.1007/s12369-011-0122-y

Komatsu, T., & Yamada, S. (2010). *Effects of adaptation gap on user's variation of impressions of artificial agents*. Proc. WMSCI 2010.

Komatsu, T., & Yamada, S. (2011). Adaptation gap hypothesis: How differences between users' expected and perceived agent functions affect their subjective impression. *Journal of Systemics, Cybernetics and Informatics*, 9(1), 67-74.

Kuznetsova, A., Brockhoff, PB., Christensen, RHB. (2017). "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software*, 82(13), 1–26. doi:10.18637/jss.v082.i13.

Lee, J., & Marsella, S. (2006). Nonverbal Behavior Generator for Embodied Conversational Agents. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), *Intelligent Virtual Agents* (pp. 243–255). Springer. https://doi.org/10.1007/11821830_20

Lee, K., Lee, C., & Nass, C. (2005). Social-Psychological Origins of Feelings of Presence: Creating Social Presence With Machine-Generated Voices. *Media Psychology - MEDIA PSYCHOL*, 7. https://doi.org/10.1207/S1532785XMEP0701_2

Li, B.A., Thakkar, M., Wang, Y., & Riedl, M.O. (2014). *Data-Driven Alibi Story Telling for Social Believability.*

Liew, T. W., & Tan, S.-M. (2016). Virtual agents with personality: Adaptation of learner-agent personality in a virtual learning environment. *2016 Eleventh International Conference on Digital Information Management* (ICDIM), 157–162. https://doi.org/10.1109/ICDIM.2016.7829758

Liu, Q., & Steed, A. (2021). Social Virtual Reality Platform Comparison and Evaluation Using a Guided Group Walkthrough Method. *Frontiers in Virtual Reality*, 2. https://www.frontiersin.org/articles/10.3389/frvir.2021.668181

Loveys, K., Sebaratnam, G., Sagar, M., & Broadbent, E. (2020). The effect of design features on relationship quality with embodied conversational agents: a systematic review. *International Journal of Social Robotics*, 12(6), 1293-1312.

Loyall, A. B. (1997). *Believable agents: Building interactive personalities* [Phd]. Carnegie Mellon University.

Meadows, M. S. (2007). *I, avatar: The culture and consequences of having a second life*. New Riders.

Meeks, C. (2020, August 12). Becoming a B-corp, transparent web entities, Meward, parties at the Nut, and more... Retrieved April 18, 2021, from https://blog.tivolicloud.com/weekly-update-8-12-2020/

Meeks, C., Deprez, M., et al. (2020). *Tivoli Cloud VR* (Version 0.17.3) [Computer software]. https://tivolicloud.com/

Microsoft. (2015). *Alt Space VR* [Computer software]. Retrieved November 14, 2023, from https://web.archive.org/web/20221231034447/https://altvr.com/

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology*, 31, 186–207. https://doi.org/10.1016/j.cedpsych.2005.05.002

Morgan, E., & Papangelis, K. (2015). Comparing the Trade-off of Believability and Performance of Abstract Intelligent Agents and Humans Playing Super Mario Bros. In C. Stephanidis (Ed.), *HCI International 2015*—Posters' Extended Abstracts (pp. 759–763). Springer International Publishing. https://doi.org/10.1007/978-3-319-21380-4_128

Morie, J. F., Chance, E., Haynes, K., & Rajpurohit, D. (2012). Embodied Conversational Agent Avatars in Virtual Worlds: Making Today's Immersive Environments More Responsive to Participants. In Philip Hingston (Ed.), *Believable Bots: Can Computers Play Like People? (pp. 99–118)*. Springer. https://doi.org/10.1007/978-3-642-32323-2_4

Mozilla. (2018). *Mozilla Hubs* [Computer software]. Retrieved November 14, 2023, from https://hubs.mozilla.com/

Musick, G., O'Neill, T. A., Schelble, B. G., McNeese, N. J., & Henke, J. B. (2021). What Happens When Humans Believe Their Teammate is an AI? An Investigation into Humans Teaming with Autonomy. *Computers in Human Behavior*, 122, 106852. https://doi.org/10.1016/j.chb.2021.106852

Newton, C. (2021, July 22). *Mark in the metaverse*. The Verge. Retrieved February 25, 2022, from https://www.theverge.com/22588022/mark-zuckerberg-facebook-ceo-metaverse-interview

Nixon, M. (2009). *Enhancing believability: Evaluating the application of Delsarte's aesthetic system to the design of virtual humans* [Thesis, School of Interactive Arts & Technology - Simon Fraser University]. http://summit.sfu.ca/item/9766

Nyatsanga, S., Kucherenko, T., Ahuja, C., Henter, G. E., & Neff, M. (2023, May). A Comprehensive Review of Data-Driven Co-Speech Gesture Generation. In *Computer Graphics Forum* (Vol. 42, No. 2, pp. 569-596).

Overte. (n.d.). *Overte* [Computer software]. Retrieved April 17, 2023, from https://overte.org/

Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education*, 67(C), 36–50.

Poggi, I., Pelachaud, C., de Rosis, F., Carofiglio, V., & De Carolis, B. (2005). Greta. A believable embodied conversational agent. *In Multimodal intelligent information presentation* (pp. 3-25). Dordrecht: Springer Netherlands.

Prada, R., & Paiva, A. (2005). Believable groups of synthetic characters. *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, 37–43.* https://doi.org/10.1145/1082473.1082479

Picard, R.W. (1997). *Affective computing*. MIT Press, Cambridge Google Scholar

R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

Ravenscraft, E. (2021, November 25). *What is the metaverse, exactly?* Wired. Retrieved April 12, 2022, from https://www.wired.com/story/what-is-the-metaverse/

Rubio-Tamayo, J. L., Gertrudix Barrio, M., & García García, F. (2017). Immersive Environments and Virtual Reality: Systematic Review and Advances in Communication, Interaction and Simulation. *Multimodal Technologies and Interaction*, 1(4), Article 4. https://doi.org/10.3390/mti1040021

Ruttkay, Z., & Pelachaud, C. (Eds.). (2004). From brows to trust: Evaluating embodied conversational agents (Vol. 7). *Springer Science & Business Media*.

Saberi, M., DiPaola, S., Bernardet, U. (2021) Expressing Personality Through Non-verbal Behavior in Real-Time Interaction. *Frontiers in Psychology*, Vol 12, 5474. https://doi.org/10.3389/fpsyg.2021.660895

Sinatra, A. M., Pollard, K. A., Files, B. T., Oiknine, A. H., Ericson, M., & Khooshabeh, P. (2021). Social fidelity in virtual agents: Impacts on presence and learning. *Computers in Human Behavior*, 114, 106562. https://doi.org/10.1016/j.chb.2020.106562

Solirax. (2018). *Neos VR* [Computer software]. Retrieved November 14, 2023, from https://neos.com/

Tanenbaum, T. J., & Bizzocchi, J. (2009). Close Reading Oblivion: Character Believability and Intelligent Personalization in Games. *Loading...*, 3(4), Article 4. https://journals.sfu.ca/loading/index.php/loading/article/view/42

Tanenbaum, T. J., Hartoonian, N., & Bryan, J. (2020). “How do I make this thing smile?”: An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376606

Valkov, D., Martens, J., & Hinrichs, K. (2016). Evaluation of the effect of a virtual avatar's representation on distance perception in immersive virtual environments. 2016 *IEEE Virtual Reality (VR)*, 305–306. https://doi.org/10.1109/VR.2016.7504775

Vanian, J., & Levy, A. (2023, February 1). *Meta lost $13.7 billion on reality labs in 2022 as Zuckerberg's metaverse bet gets pricier*. CNBC. Retrieved January 29, 2024, from https://www.cnbc.com/2023/02/01/meta-lost-13point7-billion-on-reality-labs-in-2022-after-metaverse-pivot.html

Verhagen, H., Eladhari, M. P., Johansson, M., & McCoy, J. (2013). Social Believability in Games. In D. Reidsma, H. Katayose, & A. Nijholt (Eds.), *Advances in Computer Entertainment* (pp. 649–652). Springer International Publishing. https://doi.org/10.1007/978-3-319-03161-3_74

Vircadia. (n.d.). *Vircadia* [Computer software]. Retrieved April 17, 2023, from https://vircadia.com/

VRChat Inc. (2012). *VRChat* [Computer software]. Retrieved April 28, 2023, from https://hello.vrchat.com/

Wang, X., Yang, J., Han, J., Wang, W., & Wang, F.-Y. (2022). Metaverses and DeMetaverses: From Digital Twins in CPS to Parallel Intelligence in CPSS. *IEEE Intelligent Systems*, 37(4), 97–102. https://doi.org/10.1109/MIS.2022.3196592

Wei, X., Jin, X. & Fan, M. (2022). Communication in Immersive Social Virtual Reality: A Systematic Review of 10 Years' Studies. *In Chinese CHI '22: The Tenth International Symposium of Chinese CHI,* October 22–23, 2022, Guangzhou, China. ACM, New York, NY, USA, 11 pages.

Weidner, F., Boettcher, G., Arévalo Arboleda, S., Diao, C., Sinani, L., Kunert, C., Gerhardt, C., Broll, W., & Raake, A. (2023). A Systematic Review on the Visualization of Avatars and Agents in AR & VR displayed using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, PP, 1–11. https://doi.org/10.1109/TVCG.2023.3247072

Yalçin, Ö N. (2018). Modeling Empathy in Embodied Conversational Agents. *Proceedings of the 2018 on International Conference on Multimodal Interaction - ICMI 18*. doi:10.1145/3242969.3264977

Young, M. K., Rieser, J. J., & Bodenheimer, B. (2015). Dyadic interactions with avatars in immersive virtual environments: High fiving. *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, 119–126. https://doi.org/10.1145/2804408.2804410

Zammitto, V., DiPaola, S., & Arya, A. (2008). A Methodology for Incorporating Personality Modeling in Believable Game Characters. *Pre-Print Proceeding of the 4th International Conference on Game Research and Development*. Beijing, China.

Zheng, Y., Abrevaya, V. F., Bühler, M. C., Chen, X., Black, M. J., & Hilliges, O. (2022). Im avatar: Implicit morphable head avatars from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13545-13555).

Appendix A.

Sample of Social Believability Planning Document: Survey Structure and Questions

Note: slight changes were made to the wording between this planning doc and the actual survey on Survey Monkey

[SOCIAL BELIEVABILITY STUDY SURVEY PLAN V2]

Before we begin
Thank you for visiting this study on the Believability of Embodied Conversational Agents!

On the next page, you will be met with the Consent Form page. Please make sure you have Study Code at hand. Please, respond to the consent form by indicating your willingness to participate in this study. Please make sure you have your Study Code ready at hand.

CONSENT FORM FOR ONLINE STUDY

- [Agreement and Consent Form here]
- I give my informed consent and wish to continue with this study [button]
- I do not wish to participate in this study at the present time [button]

INTRODUCTION

Thank you for choosing to participate in this study!

The study will involve watching some videos and answering some questions.

The study should take somewhere between 20 to 30 minutes.

Participants may close the survey and exit at any time if they decide at any point that they don't feel comfortable or well and need to stop participating in the survey.

What will the participants do
In this study survey, the participants will start off with a photo of the first avatar they will be introduced to. Then they will answer some initial impression questions. Next, they will watch a video of that avatar in action and then answer some follow-up questions. After, they will do another round of watching a video and answering questions, but this time with a different avatar. The survey will end off with some final questions.

Instructions before we begin
This survey is a *unique instance* just for the participant. Please, **do not share** the survey contents or the survey link with anyone.

It is very important to make sure the survey responses are as *accurate as possible*. It is expected that participants will do the survey in **one sitting without pausing or leaving** their computer/survey.

Participants should prepare themselves before continuing by making sure of the following:

- they have allotted **20-30 minutes** to do the survey
- there are **no distractions** or interruptions nearby that might interfere with the survey
    - please, remove any potential distractions and anything that can interrupt the survey. Including but not limited to: phones, other computers, pagers, etc.
- they are in an area where there's a **good and stable internet connection**
- If they are in a busy, noisy area, we suggest using **headphones** for the duration of the survey, especially for the video portions
- the survey **does not allow revisiting questions**, it is also **not** possible to **redo submissions** or to **submit multiple times**.
    - make sure to not click *back* on the survey or hit the *back arrow* on this browser.
- there should be no need to hit the *refresh button* on this browser during the survey

The participant has prepared for the survey and done the above and is ready to do the survey in its entirety

- Yes, No

Please enter your Study Code
[Study Code here]

VIDEO CHECK
It is important that participants are able to clearly view the videos that are presented in this survey. Before proceeding, we have provided a test video below. Please, watch the short video and **only click next** once *the video is done*. Answer the questions about the video on the next page.

[test video here]

Video Check Questions
Please answer the questions below about the short video you just watched.

- What animal was shown in the video?
    - Dog
    - Cat
    - Fish
    - Hamster
    - Rabbit
- What was the colour of the animal shown?
    - Black
    - White
    - Golden
    - Red

    - Pink
- A short sentence was spoken during the video. Please, write down the sentence that was spoken.

[The first two (before randomization) question needs to be answered correctly in order to be validated to continue]

PRE SURVEY QUESTIONS

The survey should be done entirely inside the survey system (SurveyMonkey). There is no need to visit external sites or click on external links. Please, *avoid clicking away* from the survey or *clicking on any links/icons* that might have the participant leave the survey window.

Let us start with some preliminary questions. Please, answer the following:

- What is your age?
    - 19-24
    - 25-34
    - 35-44
    - 45-54
    - 55-64
    - 65+
- What gender do you identify as?
    - Male
    - Female
    - Other(enter text)
    - Prefer not to say
- Do you have any experience with video games?
- Please describe the extent of your experience with video games
    - [dropdown]
- Do you have any experience with avatars? (Avatars are 3d humanoid-like game characters you can interact with or people online coming through as 3d humanoid game characters you can interact with)
    - Yes, No
- Please describe the extent of your experience with avatars by choosing one or more of the following
    - I play a lot of video games involving using and creating avatars
    - I use VR for games and such that involve using and/or creating avatars
    - I visit virtual social worlds that use and/or create avatars
    - I'm aware of the use of avatars in one or all of the following but never used or worked with them: video games, VR, social platforms, film, animation
    - I only know about the film Avatar or the cartoon Avatar
    - I'm not 100% sure what an avatar is and/or what they look like
    - I've never encountered any virtual avatars
    - Other (please specify)

SURVEY PART 1

For the purpose of this survey, we present a fictitious scenario for you to immerse yourself in. We ask you to pretend that you are visiting a Virtual World platform on your computer (such as VRChat) populated with various avatars. You are here to see your friends/acquaintances who are also visiting this virtual world in their avatar form.

Pre-Video Instructions
Please, read through these instructions and state your agreement prior to proceeding. Later in the survey videos will be shown of an avatar's performance in a particular scenario. Please, focus on the avatar's behaviour, communication, and social skills.

For the sake of consistency and accuracy, the participant should make sure:

- they watch the video in its entirety and in one sitting
- They do not pause the video
- They do not click on video icons or links that make them leave the survey window
- They do not click back on the survey, back on the web browser, or the refresh button
- Only once the video is done do they hit the next button
- They do not click on any of the icons or images at the end of the video
- It is recommended to watch the video with *headphones*

The participant has read the above instructions and agrees that they have understood the instructions. They also agree that they are ready to watch the videos.

- Yes, No

[Separate page]
Photo of avatar

You are about to meet up with your first friend in their avatar form. Prior to the meeting, your friend sent you a screenshot of where they are in the virtual world.

Please, look at this photo before answering the questions
[Photo here]

Initial impressions

What are your initial expectations of the following..

- the avatar's ability to immerse you in a conversation
    - Very Low* * Moderate * *Very High
- the avatar feeling like a real person
    - Very Low* * Moderate * *Very High
- the avatar's social skills (social cues, gestures, body language)
    - Very Low* * Moderate * *Very High

How would you rate the avatar's..

- potential ability to communicate non-verbally
    - Not effective* * Moderately Effective * *Very effective
- potential range of emotion
    - No range* * Moderate Range * *Wide range
- potential range of movement

- No range* * Moderate Range * *Wide range

–

SURVEY PART 2

Pre-Video Instructions
A video is about to be shown of an avatar's performance in a particular scenario. Please, focus on the avatar's behaviour, communication, and social skills.

Watch Video
Now you will watch the video
[watch video]

Post-Video Questions

- [video question check]
  - Scenario 1(Video Game Pitch)
    - What colour was the hair of the avatar talking in the video?
    - Describe as best you can the game that was pitched in the video
  - Scenario 2(Social Meet Up)
    - What colour was the shirt under the jacket of the avatar?
    - What did the avatar decide to do when they noticed no bartender was around?
- [general questions]
  - Scenario 1
    - The avatar was really passionate about their product
      - Strongly disagree* * * * *Strongly agree
    - What did you learn about the avatar?
    - What did you like about the avatar?
  - Scenario 2
    - Did the avatar convince you that they were excited to meet you?
      - Strongly disagree* * * * *Strongly agree
    - What did you learn about the avatar?
    - What did you like about the avatar?
- [main questions]
  After watching the video, what are your expectations of the following..
  - the avatar's social skills (social cues, gestures, body language)
    - Very Low* * Moderate * *Very High
  - the avatar feeling like a real person
    - Very Low* * Moderate * *Very High
  - the avatar's ability to immerse you in a conversation
    - Very Low* * Moderate * *Very High

  Now that you've watched the video, how would you rate the avatar's..
  - potential range of movement
    - No range* * * * *Wide range
  - potential range of emotion
    - No range* * * * *Wide range
  - potential ability to communicate non-verbally
    - Not effective* * * * *Very effective
  - –

    - The avatar perceives the world around him/her (awareness)
        - Strongly disagree* * * * *Strongly agree
    - It is easy to understand what the avatar is thinking about (behaviour understandability)
        - Strongly disagree* * * * *Strongly agree
    - The avatar has a personality (personality)
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour draws my attention (visual impact)
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour is predictable (predictability)
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour is coherent (behaviour coherence)
        - Strongly disagree* * * * *Strongly agree
    - –
    - The avatar was controlled by someone else
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour was believable
        - Strongly disagree* * * * *Strongly agree
- [additional questions]
    - I really liked the avatar
        - Strongly disagree* * * * *Strongly agree
    - I would interact with the avatar again
        - Strongly disagree* * * * *Strongly agree
    - I felt very engaged with the avatar
        - Strongly disagree* * * * *Strongly agree
    - (Optional) How well did the avatar do in comparison to a real person?
    - (Optional) Was there anything specific about the avatar that helped you remember the conversation?
    - (Optional) Did you notice anything worth mentioning about the avatar's performance?

–

SURVEY PART 3

Photo of avatar

You are about to meet up with your second friend in their avatar form. Prior to the meeting, your friend sent you a screenshot of where they are in the virtual world.

Please, look at this photo before answering the questions
[Photo here]

Initial impressions

Please answer about your expectations of the following..

- the avatar's ability to immerse you in a conversation
    - Low* * * * *High
- the avatar feeling like a real person
    - Low* * * * *High
- the avatar's social skills(social cues, gestures, body language)
    - Low* * * * *High

Please state your rating for the following: the avatar's..

- potential ability to communicate non-verbally
    - Not effective* * Moderately Effective * *Very effective
- potential range of emotion
    - No range* * Moderate Range * *Wide range
- potential range of movement
    - No range* * Moderate Range * *Wide range

_

SURVEY PART 4

Pre-Video Instructions
A **reminder** to *please follow the video watching instructions* for the second avatar video.

Please make sure you: *watch the video in its entirety, do not pause, do not click away, do not click back, do not click on any icons, watch using **headphones**, and click **next** only once the video is done.*

Watch Video
[watch the video]

Post-Video Questions

- [video question check]
    - Scenario 1(Video Game Pitch)
        - What colour was the hair of the avatar talking in the video?
        - Describe as best you can the game that was pitched in the video
    - Scenario 2(Social Meet Up)
        - What colour was the shirt under the jacket of the avatar?
        - What did the avatar decide to do when they noticed no bartender was around?
- [general questions]
    - Scenario 1
        - Did the avatar convince you to buy/support their product?
            - Strongly disagree* * * * *Strongly agree
        - What did you learn about the avatar?
        - What did you like about the avatar?
    - Scenario 2
        - Did the avatar convince you that they were excited to meet you?
            - Strongly disagree* * * * *Strongly agree
        - What did you learn about the avatar?
        - What did you like about the avatar?
- [main questions]
  What are your expectations of the following after seeing the video..
    - the avatar feeling like a real person
        - Low* * * * *High
    - the avatar's social skills(social cues, gestures, body language)
        - Low* * * * *High
    - the avatar's ability to immerse you in a conversation
        - Low* * * * *High

  Having seen the video, how would you rate the avatar's..

    - potential range of emotion
        - No range* * * * *Wide range
    - potential range of movement
        - No range* * * * *Wide range
    - potential ability to communicate non-verbally
        - Not effective* * * * *Very effective
    - –
    - The avatar perceives the world around him/her (awareness)
        - Strongly disagree* * * * *Strongly agree
    - It is easy to understand what the avatar is thinking about (behaviour understandability)
        - Strongly disagree* * * * *Strongly agree
    - The avatar has a personality (personality)
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour draws my attention (visual impact)
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour is predictable (predictability)
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour is coherent (behaviour coherence)
        - Strongly disagree* * * * *Strongly agree
    - –
    - The avatar was controlled by someone else
        - Strongly disagree* * * * *Strongly agree
    - The avatar's behaviour was believable
        - Strongly disagree* * * * *Strongly agree
- [additional questions]
    - I really liked the avatar
        - Strongly disagree* * * * *Strongly agree
    - I would interact with the avatar again
        - Strongly disagree* * * * *Strongly agree
    - I felt very engaged with the avatar
        - Strongly disagree* * * * *Strongly agree
    - (Optional) How well did the avatar do in comparison to a real person?
    - (Optional) Was there anything specific about the avatar that helped you remember the conversation?
    - (Optional) Did you notice anything worth mentioning about the avatar's performance?

FINAL SURVEY QUESTIONS AND SUBMISSION

- You've just watched TWO videos of different avatars. Regardless of their personalities and the specific scenarios they were in and focusing strictly on their ability to communicate and collaborate, which avatar would you choose as the best one, again in terms of its ability to communicate and collaborate?
    - Avatar in video 1
    - Avatar in video 2
- Please explain the above choice
- Is there anything you want to mention?

–

Thank you for doing the survey

To complete the survey: you must hit the "submit" button to finalize the survey.

You may submit the survey by clicking the “submit” button or you may cancel your participation in the survey by closing this browser window.

We hope you found something interesting during this study and have possibly found an interest in Virtual Worlds (like VRChat) and potentially visiting them in the future in Virtual Reality.

About the content of the survey
A reminder to please not share the contents of the survey or the survey link with anyone

Contact Info and Further Inquiry
If you have any further interest/questions or if you know anyone that might be interested in doing this survey please contact us at:

[Submit] [Cancel]

Appendix B.

Avatar Exposure Qualitative Responses

Group: 1A	Scenario: Pitch	Avatar: Desktop	Exposure: 1

What did you learn about the avatar?

- The avatar was shy, though the voice and stance was confidant.
- The avatar was leading a pitch meeting for the robot video game, which implies they work for some kind of game studio.
- They speak exactly same as the human person.
- avatar can communicate with me face to face like a person, if it is good enough, I can see expressions and actions
- He had a slightly enthusiastic tone about the information he presented
- They are not as passionate as I thought they would be. From the previous screenshot I expected them to be more passionate and wider range of emotion/movement.
- It was a female avatar with a masculine voice. and the body movement was a bit odd. Like while talking the avatar turned the opposite direction and kept talking. This was before the screen was presentation.
- they seem to be a game developer or marketing director
- i learnt what they said about the video
- That the avatars face matches what is being said by the use
- It has green hair, a shirt and jacket.
- Their movement is very artificial and not coherent at all.
- They are developing a game and need the opinion of the player
- They are releasing a video game to switch, PlayStation, etc. and enjoy sci-fi
- The avatar can change the mouth shape according to the voice, and also has some hand movements. It can move freely in the virtual space.
- The avatar was knowledgeable about the game they were pitching, made little to moderate eye contact, and used hand gestures a moderate amount of time while speaking.
- Their voice sounded more masculine than I had imagined. They liked games, they seemed knowledgeable about video game technology, and they seemed excited to hear 'my' opinion about their pitch.
- Their movement and gestures were better than I thought they would be.
- The avatar is a he/she.
- he is trying to introduce their group's design idea of creating a game.
- It has a surprising range of movement with its arms
- I don't think any personal information was shared in the video.

What did you like about the avatar?

- Good eye contact. Expressive arm movements.
- I liked the hand gestures as they were talking. Although they were a bit clunky at times, this made the avatar seem a little bit more human-like.
- I liked how it had a really neat presentation, and asks my avatar if we would like to hear more, asking opinion.
- It is a new experience compared to traditional games and video conferencing
- Neutral, nothing special, just an average narrator
- I like the avatar's friendliness and body language, not too much but sufficient.
- The eyes either mimicked the real movement of the or used the new eye tracking movement to show realisum
- the hand motions
- that it sounded like a real human being
- I liked their facial expressions and body language
- It had a good aesthetic, such as hair, clothes and physical features. Also, it used hand gestures when talking.
- Mostly their jacket but not their movements.
- I like the style of the avatar and their facial expression
- They are friendly and easy to understand
- The avatar has a relatively neutral dress, and it can interact differently depending on the virtual world around it, such as pointing to the sketches on the wall while speaking.
- While speaking, the avatar sometimes spoke with their hands, something that people IRL do very commonly when presenting or pitching an idea.
- Their hair and fashion are cool. They enunciated their words well.
- The movement
- I like the hand gesture and outfit of the avatar
- he moves naturally when he speaks, such as some gesture. when he is not facing the camera, his voice is dimmer
- It has a surprising range of movement with its arms
- The hand movements and gestures were very realistic.

(Optional) How well did the avatar do in comparison to a real person?

- Avatar did fine. Was engaging enough, but I missed the deeper facial cues and fluid movements of a real person.
- Moderately well. The hand gestures added to the experience of speaking to a real person, but I did not like the lack of facial expressions on the avatar.
- When speaking, it felt like a real person was controlling the avatar from the back.
- It's actually pretty bad. Because the point of face-to-face communication is to be more efficient, and I can feel the other person's "temperature" (if the other person is happy, I can feel warm, if the other person is angry, I will be nervous).
- Not quite there yet
- close but need to work on leg movement
- The avatar seemed robotic and stiff as compared to a normal person. The model was distracting and felt as if it was harder to pay attention to the content being displayed

- Well, it was harder to understand the avatar because there was a lot of information about the other features of the game. Also, the movement was weirder in the later part of the video, but the hand gestures at the beginning were very human like.
- the movement of the avatar sometimes is not that coherent
- There were a few awkward movements when they were speaking that didn't feel natural
- Compared to a real person, the avatar's movements are lacking and a little stiff, but the perception and interaction of the surrounding space are good.
- I feel more compelled to listen to the avatar pitch the game than I would a real person, because I feel like the avatar feels like a video game character and they are personally inviting me to the game they are trying to pitch.
- While the avatar cannot express emotion as well as a real person, communicating the game concept was done well.
- The hand gesture and movements were a bit delayed
- Good effort, but does not compare
- I think the voice and the mouth movements we're not adequately synced in a realistic way

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- The interaction with the graphics on the wall â€“ as the avatar turned and looked at the images, this drew my attention there as well.
- How it s going to launch the video game, Nintendo switch.
- No, actually, I think avatars just make my experience better, but they're no different than video conferencing.
- N/A
- the jacket and anime eyes
- no
- I think the eyes of the avatar helped engage the conversation. The unique aesthetic of the avatar made it memorable.
- when they turned around and face the screen
- Some of the intenations of the voice, like when they talked about some of the exciting portions of the game.
- The avatar's interaction with the surrounding space. This allows for more references on the basis of purely oral descriptions.
- The avatar was dressed in a similar style to most sci-fi game protagonists typically dress, so I remember the conversation because I felt like I was speaking with a character from the game they were trying to present.
- When the avatar gestured with their hands, that helped me to regain focus on their pitch.
- Not really, the whiteboard helped me remember more.
- I was more distracted by the limits of the avatar honestly (gitters, awkward rotation and movement in space)
- no.

(Optional) Did you notice anything worth mentioning about the avatar’s performance?

- N/A
- It guides you to the slide, and having eye contact with out avatar.
- I think avatars need to be more like real people and have more expressions and movements. At the same time, I'm not optimistic about the development of avatars because they don't give us actual feelings such as hugs. But if it's combined with technology that allows our bodies to feel things, I believe it could become the most mainstream or even the only way to interact.
- awkward body language
- it would be inappropriate to say
- the avatar was calm and had decent presentation skills
- The avatar seemed to respond based on signals of the player. When the person nodded their head, the avatar continued speaking, which is what a real human would do.
- The avatar seemed tense, which made me feel slightly more tense, too. There was no relaxed or casual tone being evoked from the avatar's body language, so it was difficult to be immersed in the conversation because their body language (and in some instances, the way they spoke) conveyed "board meeting", while their clothing was more casual.
- I was impressed when they paused to wait for 'me' to respond, and then when they continued after only a nod.
- Lip-syncing was a little off.
- No

Group: 1A	Scenario: Disco	Avatar: VR	Exposure: 2

What did you learn about the avatar?

- Avatar felt a little cold or perhaps shy.
- She has an eight year old dog named Billa, who is a golden retriever. She is also the Chief Marketing Officer of a social VR platform, and is interested in all things "technical", such as NFTs.
- She speaks casually. She makes many gestures.
- avatar's voice is three-dimensional, as if there is a real person in front of you
- she didnt seem very normal
- She is very outgoing, and she seem to be a very good companion to be with, someone who would cheer you up.
- She work at the best VR soiclizing thing. and she is really into anyting technical
- their name was "keyus" they are a marketing director for a social media platform who loves tech and dogs.
- their hobbies, and that they have a dog
- They love anything technology related and work in a VR company (one of the best)

- Their name is Kass, they are chief marketing of some company, and their hobbies is anything technical, such as NFT. They have an NFT of their dog. They also had a sore throat last week.
- The way she talks is a bit unnatural and I have a slight feeling that her excitement was forced.
- The avatar is very outgoing, she is interested in anything technological. She has an 8 years old golden retriever.
- They like tech stuff, they might be a little drink, they are caring and very charismatic
- The avatar has very rich emotions and body gestures. It can not only perceive the virtual world around it but also interact with the objects around it.
- They are the CMO of a company, their name was somewhat incomprehensible but I believe it was Caius? Kaye? They like anything technical, such as AI and NFTs and proceeded to show me an NFT of their dog, Bilah? Milah?
- They're very talkative and enthusiastic, especially about their 8-year-old golden retriever. They are in a leadership position at a VR tech company. They had a cold last week, which affected their throat.
- They're into technology
- She is a marketer that loves technical things like NFT. She loves her dog.
- she likes ai, nft, and she shows a picture of her dog and like it so much
- They can control all of their limbs
- This avatar has a dog

What did you like about the avatar?

- I liked how the head tilts and eye contact felt engaging.
- She was very expressive, had a wide range of gestures, and speaking volume, and physically interacted with me (e.g., greeting with air kisses) as opposed to just talking at me.
- She sounded like a real person, a conversation which we can hear somewhere near us.
- voice
- nothing
- Her friendliness and her voice makes people feel safe.
- she can do a lot of body movement. and she loves her dog
- the range of motion and emotion was very high
- their range of movement was better than the last
- Very expressive, alot of open body language
- Very energetic, good body language, and energetic voice.
- That this one has more body movements than the one in the previous video
- I really like the emotion that the avatar expressed
- She was funny because she seemed a little loopy and friendly.
- The high spirits of this avatar touched me very much, the rich body language and the very lively and enthusiastic dialogue made me very happy.
- The avatar showed a level of enthusiasm and excitement to be conversating with me. The way they were dressed also made this interaction seem like a casual, believable meetup that would occur IRL.

- Their dog was cute. I also enjoyed their dancing at the beginning.
- I liked their dog
- I liked her fun personality and she tried to interact with me and other people.
- her movement is very natural, and she has different pitch when she talks. she also tries to interact with me such as kissing as greeting, and holding my hands to invite me to her place
- Their range of motion was better than the previous avatar
- this avatar had more human like behaviour

(Optional) How well did the avatar do in comparison to a real person?

- The non-vocal behaviour (dancing, waving, kissing) was quite similar to a real person, and I liked the range in her speaking volume (yelling for the bartender vs speaking at a normal volume)
- The voice and movement were great, but the expression was not
- fairly similar
- 80% real personanlity
- the avatar has a high range of expression. I could believe that it was a real person, although it still feels very scripted and stiff
- better movement, better expression of emotions
- Very well because it was able to express emotion with body language and tone. Also changed behaviour and actions based on change of situations.
- I think the avatar did very well in the tone of her voice and her interaction with the environment looks like a real person
- Movements were excellent however I found her words to be a bit mumbled at times
- The avatar's emotions and body movements are very close to real people, and the rich movement details such as the movement of the legs and the movement of the head make it closer to the movements of real people.
- Did alright.
- The avatar is comparable in terms of general full body control with their limbs and torso, but still lacks fine details and facial expressions
- I believe if the avatar had facial movements/expressions it would help in making a convincing avatar.

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- Her interactions with her environment (picking up and showing the NFT, waving her hands in front of the person next to her, dancing on the dance floor)
- She picked up the picture
- her questionable body language
- she was very forcing and gets what she wants
- not necessarily, their voice pitch was a bit uncomfortable

- The use of hands when talking and energetic body language. Also the aesthetic was engaging.
- How expressive she was and she showed emotion
- Holding the photo of the pet and pointing to the photo when the avatar introduces its pets I was very impressed with this part.
- The conversation wasn't very memorable.
- Nothing in particular

(Optional) Did you notice anything worth mentioning about the avatar's performance?

- voice
- looked like on drugs/drinks
- it feet wierd when she went right through my charater like Usain Bolt
- The performance was believable and i could understand if the performer was similar to the avatar. The performance was exaggerated and felt unreal
- Most of this avatar's movements are very close to real people, but the movement speed is a bit fast, like teleportation.
- I was impressed by the avatar being able to hold up the NFT of their dog.
- The dancing was surprisingly realistic
- The avatar lagged a lot or had choppy framerate

Group: 1B	Scenario: Pitch	Avatar: VR	Exposure: 1

What did you learn about the avatar?

- The avatar uses body gestures and facial impressions as it speaks but seemed a little unnatural. The avatar also has a voice of a male but the appearance looks gender-neutral.
- I learned that it can speak and move, however, it's facial expressions are limited
- They made a great pitch for the video game concept and are able to move around fairly realistically.
- The avatar has a range of expressions and movements with its hands and face.
- The avatar is a man with green hair, he knows the game well and was trying his best to introduce us to the game.
- Did not display any emotion. Also the legs don't really move, they kind of rotate awkwardly.
- I learned that the avatar seems to be knowledgeable about this up and coming game. They also seem to know a lot about the details of the game, and that they also have plans for marketing,
- i learned about the games platform, what the plot is, how they are planning to market and introduce other streams of profit through amibo
- The avatar was quite expressive with arm movements and the way it shifted, the mouth would move when it spoke.
- How will we play the game.

- Avatar had smooth tracking with its arms. Their audio came from the avatar's mouth. So for example when they were on my right I only heard them with my right ear.
- I learnt the main theme of the sci fi game, fighting in a chaotic enviroment when robots went free.
- He seems to be working on a game.
- The avatar is soft spoken and it tries to engage the viewer
- Avatar is a character that would closely look like humanoid to better replicate human's actions/emotions for better immersion in VR/online world
- They are much more physically expressive than I thought when it comes to capturing movements. The occasional jittering of the avatar's position in the space (not the intentional movements, but what appeared to be accidental shifting of their physical position) did make me feel a little sick and distracted me a bit from what was being said at times. But other than that, it was a pleasant experience.
- The eyes are dead, and there is no movement other than automatic blinking.The eyes are dead, and there is no movement other than automatic blinking. Although there are hand and foot movements, the key facial expressions are difficult to detect
- Not sure
- The avatar's body language and facial expressions matched everything they said. They were able to use hand gestures when talking to emphasis their message.
- that the mouth movement doesn't match the words the avatar was saying
- This avatar is capable of doing gestures, and interact with others.
- The avatar has very minimal movements, the mouth moves when speaking, and the arms moved when the player controlling them moved.

What did you like about the avatar?

- I liked how it mimicked the conversational gestures that humans might make, even it seemed unnatural, the gestures themselves were quite detailed. I liked how it blinked its eyes....
- I liked that it made eye contact with me so that it felt like I was actually there in the room. Having it speak and move makes it feel more believable.
- I liked that they used body language as they were talking. They weren't just standing still and talking like an NPC normally would. It made me more immersed, and I really felt like I was at a normal meeting.
- I liked the expressions it showed.
- Easy to recognize and remember, can somehow tell his personality from the video
- The model looks really nice and it was able to move its hands around.
- I liked the design of the avatar and the colours that were used. I also liked that there were subtle hand movements and gestures just as a human would use.
- i liked how expressive the avatar aims are
- It's design was interesting, it made I contact with you as it spoke and would also move its lips, making it seem more life like.
- The avatar has some behaviors that fits a real personâ€™s personality.

- The body and posture of this avatar felt very natural, especially compared to the ones around it.
- The avatar's motion movement was really smooth, and has good verbal communication.
- I liked its voice.
- It provides good eye contact and hand movements so it doesn't look stiff. I also like the fashion choice, makes the avatar seem tough but still able to communicate with.
- I like customization - it can look the way I want it to look
- I like the avatar's design. It is not overly complex and a nice balance of semi-realism; they felt like a character from a Pixar film. I also like how well the avatar was able to pick up movements, despite the transform jitter. The arm movements and non-verbal communication elements of the conversation were picked up very well in my opinion.
- Compared to the two avatars standing in the back, this avatar looks comfortable. There is some degree of physical movement.
- It made hand gestures
- Tthe body language of the avatar was laid back, so it felt very casual listening to them speak.
- the body movement feels somewhat natural even though still not as smooth.
- I like how the avatar talks with hand gestures, feels like a real person.
- I liked that it somewhat simulated the body language of a normal speaking human being.

(Optional) How well did the avatar do in comparison to a real person?

- Facial expression and the arm movements were pretty close to a real person, but the leg movements were way off.
- I think if I were in the VR environment for longer, I would be able to view it as a real person.
- I would give it an 8/10. It was almost spot on. Many times when explaining the concept, the avatar would speak with their hands, point and turn to the board when referencing something on it. With improved graphics, this could be a game changer for work or school.
- The avatar uses some body gestures while explaining the game, which looks like a real person, but the gestures are also somehow stiff and limited
- Avatar felt stiff, rotating towards the screen was awkward.
- I think that the avatar did fairly well compared to a real person. As the voice seemed to be from a real person, it felt like listening to someone talk just as normal. I do believe the hand movements are a good touch, however I am unsure about the movement of the whole avatar itself (eg. walking around). Additionally, the movements do feel a little stiff at times, making it a little funny to watch. Overall, the avatar did quite well at informing me about the game they were talking about.
- there was very little facial expressions and the pauses between when they were speaking seemed awkward, as i didn't know if they were done talking or not
- relatively well, its basic movements seemed similar but there was a lack in fine motor controls and its range of arm motion didn't seem to travel above the waist.

- The lack of emotion made it harder to feel engaged in the conversation.
- I think the Avatar is like a real person, except it has limited facial movement, I feel eye contact was an issue.
- He could maybe have more facial expressions.
- Well, the avatar still looked stiff. The voice through the avatar could be more engaging. But it had some qualities to a real person, like the avatar looks unique and in real life we are all different.
- I'm still actually unsure of whether or not this was a real person speaking and interacting with the player in a digital space. If that was a non-playable character, then I'm incredibly impressed. They definitely felt like a real person was controlling them.
- I couldn't see it as a real person because it dresses like a video game character and has limited movement
- Even though the avatar s interaction was almost like a real person, there were times when the avatar's language and movement came across as unnatural and awkward.
- the movement is still a bit rigid but the way the avatar moves did mimic how a real person move their body when they're talking
- Very similar body languages as a real person, but not very smooth.
- The hand movements felt very awkward, there was a point in the video where the avatar raised its arms and the hands seemed really oddly connected. The face doesn't move much other than the mouth which hinders the ability to convey emotion a lot.

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- nope..
- When it paused and gestured for me to view the board, it helped have a visualization and actions that helped me remember certain parts of the video. As opposed to it staying static the entire time, may lose focus over time. Also it pausing to ask question helped too.
- The person engaged me with their body language, the avatar seemed realistic because of it.
- I liked that the voice projection depended on which way the avatar was facing.
- No
- There wasn't anything specific that helped me remember, but I tried to remember the conversation by recalling what the avatar said. It is possible that the images in the background might have helped.
- Some of the bigger motions like the sweeping motion were good.
- The way the avatar could direct their body towards their presentation picture and gesture towards it.
- I think the Avatar's clothing fits very well with the game through the conversation.
- His eyes were staring at me too much to remember much about the conversation.

- The first time I watched the video I was too invested in looking around, to be honest, but on the second try I was able to follow up with Avatar and understand the message clearly
- Their design stood out much more significantly compared to the others in the scene. Their body language was friendly and felt natural; they appeared approachable and kind.
- Not really
- Hand movements
- the body movement
- Hand gestures, and body languages. I started paying more attention when the avatar is not still.
- No

(Optional) Did you notice anything worth mentioning about the avatar's performance?

- No facial expressions
- I didn't notice anything worth pointing out. The avatar just seemed like a friendly guy.
- It moved around a lot which was believable for someone giving a game pitch.
- When he wants to introduce the page on the wall, he will naturally lift up his left side arm to point it.
- The range of mouth movement didn't feel like it was matching up with its voice. Also it was interested how the hands seemed stuck in a certain position that felt unnatural and distracting (I'd imagine its the position that person has their hands holding onto controllers?)
- I really like the Avatar's gesture and the smoothness of motion, and verbal communication was excellent. The designer did a great job with modelling and rendering.
- Their legs in the first 10 seconds of the video were moving in a funny way (taking really small steps) But overall well done, especially with emotions and body movement - hand gestures, body rotation, and moving fingers for indication
- The only thing I could think of was the transform jitter. I have had some difficulty in the past with watching videos of VR experiences due to motion sickness, and the transform jitter of the avatar did make me feel a bit queasy at times and made it difficult for me to hear what they were saying. But otherwise, I really liked their design!
- Not really
- I noticed the feet movement/whole body movement glitched sometimes
- that the movement of the mouth doesn't match the words. it's hard to ignore as that was what I'm looking at the most when the avatar is talking
- The avatar seemed to somewhat float left to right as he was speaking. Almost as if the person controlling it was stepping left to right but the avatar didn't mimic the stepping

Group: 1B	Scenario: Disco	Avatar: Desktop	Exposure: 2

What did you learn about the avatar?

- The avatar is very outgoing and has an accent that is not North American. She love dogs and whisky. She has a chirpy voice and kind heart that cares for the sore throat I have. Very friendly.
- That they're the CMO of the company and loves their dog
- Their movements can be quite ecstatic
- That she is interested in technology and has minted a picture of her dog.
- She is a marketing officer who is very passionate about everything with technology. She is outgoing and friendly.
- Has a dog named Bella that's 8 years old.
- I learned that the avatar has a dog named Billa (?). They appeared to be very talkative and excited to talk to us.
- they were very passionate about their work and very friendly
- They had a very natural form of communication and their hand motions were more believably human
- She had a 8 years dog named Bella. She is interested in any technology like nft.
- They had pre-made animations that were mapped to their movements. Certain arm movements outwards and twisting of hands.
- The avatar was the marketing person for the tech company.
- She loves anything tech related
- Engaging personality, loves to dance and drink, talkative person.
- they had even more range of showing emotions. I was mesmerized by hand gestures.
- Again, they were much more expressive than I had thought. They did not have transform jitter, which made it easier for me to pay attention to what they were saying.
- better physical mobilityï¼Œ but the face modeling seems rougher than the last avatar
- eccentric personality and talking manner, makes the last one feel much more realistic now
- The avatar works in technology and VR and AI and owns a company. They also have a golden retriever dog and had a sore throat a while ago.
- the finger movement was delicate
- she has an 8-yea-old dog, and she is super outgoing.
- Her hobbies include anything technology related, and she has an 8 year old golden retriever

What did you like about the avatar?

- I liked how the tone of her voice was in sync with the gestures she made. Her movement was much much more natural than the green-hair avatar. I also liked how half of her face was covered with her hair because her facial expressions did not quite match with her tone of voice and body gestures, so it would have been weird to not see her smiling while her voice and actions are cheerful.
- The person using the avatar was very enthusiastic, there was more range of motion this time and their dancing was fun to watch
- The hand movements, gestures, and body language
- She seems very excited.

- Her personality makes people feel welcome and she is always energetic. Her body gestures and movement make me feel she is a real person.
- This avatar moved around a lot more and felt more expressive
- I liked that the avatar seemed energetic and was pretty warm in welcoming us. I also liked their dancing in the beginning.
- i liked how expressive her voice was
- They seemed very natural in the way they acted and the avatar was more believable in how it was dressed as well.
- She will ask me questions and interact with me.
- I like how the avatar could gesture towards far away objects with a cursor (the red block)
- She has sexy dance moves and seems to be fun to have a conversation with.
- Her clothes
- At first, by the screenshot, she looked scary. But she's friendly, engaging, good conversationalist, seems like a fun avatar.
- I really like that they can walk, not teleport from one place to another; also that their mouth move according to the speech. I also liked that when she was showing her dog - the red square appeared to indicate what she is referring to.
- I liked how physically expressive the avatar was. While I initially felt less impressed by the design than the previous avatar (based on personal preference for stylized design), I was very impressed by their movement capabilities.
- When she points at something, there will be a red square to show you where she is pointing
- funny because her speech and reactions were so dramatized
- They asked the user questions
- the body movement goes beyond just moving the arms but also including the fingers when pointing and/or moving
- She is nice, energetic and very friendly.
- It felt a lot more like I was talking to another person! Her hand gestures were very lifelike, so it felt as though I was talking to another person

(Optional) How well did the avatar do in comparison to a real person?

- The avatar seemed a little unrealistic. Although the context was set, it seemed like they were too free in their movements, and in person, not many people would be constantly using body language during their communication
- Her use of body language and how she interacts with the environment makes me feel she is a real person. She did a really good job.
- Fairly well, except the facial expression was still non-existent in the avatar
- The avatar seemed did pretty well. The gestures were nice, the movement was of a wide range. The appearance in this lighting was not the most visually appealing, primarily because of how the hair texture looks a bit slimy (not sure how to describe that).
- more believeable, tho facial expressions were a little limited
- Physically they moved around very similarly to a real person and was believable, however the way they spoke seemed unnatural as it would abruptly stop at certain points

- I donâ€™t think she is like a real person. Her movements are stiff.
- The avatar felt more like an AI following specific animations like in the Sims. Which made it feel automated and unnatural.
- I think she has various personalities from a dancer to a marketing personnel.
- Her movements were too much sometimes compared to a real person's movement in that situation.
- It was nice talking to her, it's like she's actually trying to talk to us. Pretty good range of movement, and has a good personality.
- Some of the behaviours felt a little sudden - like them calling for a barkeeper and then inviting the player over for tea - but other than that the expressiveness of the person voicing, and the avatar's range of movement were very impressive.
- It did not feel like a real person
- It was real but there were times when the avatar came across as fake
- the delicate movement of the fingers made me think of a real person talking
- She is aware of the world around her and can 'think' an alternative way. So she is really close to a real person.
- A lot better than the green haired avatar! The range of motion was really good, the hand and arm gestures were very lifelike. I even noticed as she walked away from me to the bar that the clothes on her back reacted correctly as though she was really walking.

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- not really
- Waved her hands around and pointed at things like the other lady at the bar
- Not the avatar itself but when they pointed to the dog picture it helped me remember the conversation was about a dog.
- Their pointing and hand motions made certain parts of the conversation more memorable
- She points the dog picture.
- Her gesture is really open, and it opens up the conversation really well, and sustains attention.
- Her exaggerated movements pointing to the other avatar helped me remember that part of the conversation (asking if she was a bartender).
- The objects around the room
- I believe it is a change of environment; each scene had it is own dialogue the disco - greetings the bar - introduction of Chaos (that is how I heard her name) cut off - inviting to the new scene; Chaos house (not shown though)
- Their voice really stood out to me as animated and lively. Some of the behaviours felt a bit sudden or unexpected, but nothing was jarring about the conversation. I remembered some of those sudden moments best, like the avatar pointing at things or addressing the player directly about not speaking.
- eccentric personality
- Body movement, like where they directed and pointed their body when talking.
- when she pointed out things while talking about it. example: pointing at the dog when talking about him

- I remembered when she introduced her dog and kept calling for bartenders.
- No, personally, I was very distracted by the dance floor in the background and the other dancing avatars, which is why I may have gotten her shirt colour wrong.

(Optional) Did you notice anything worth mentioning about the avatar's performance?

- seemed too unrealistic, too much movement
- Seemed more enthusiastic
- Outside of what I have already mentioned, no.
- The avatar's voice made it seem like she didn't actual want to talk to me it didn't feel genuine and she wanted me to buy her NFTs or something.
- I think the dance in the beginning really captured my attention, and drew my attention on her instantly.
- Could do some more facial expressions to be more "real"
- finger movements as stated above. Looked hyper-realistic
- Again, I was surprised by how physically animated the avatar was. While I do have a personal preference for stylized designs, I appreciated how lively this avatar felt!
- was much more unrealistic and odd than the last one but that almost helped my memory and engagement more
- the facial expression is more rigid so it's a little hard to understand or predict her emotion
- Hand/arm motion was really good!

Group: 2A	Scenario: Disco	Avatar: Desktop	Exposure: 1

What did you learn about the avatar?

- Their name, their interest in technology, their dog.
- she is joyful and positive
- Sheâ€™s the CFO of a company dealing with blockchain, and has a dog.
- She has a pet dog
- that she can speak and move around. she is into technology.
- She is interested about everything technology
- She has a dog and she dances at the beginning
- She works for a social VR company
- I learned that she loves technology and that she is some kind of chief officer I believe. I also learned her name, that she loves her golden retriever dog, and that she has an accent.
- The audio isn't really clear. In the beginning, the audio seems only appears on my right. So I didn't learn something from her.
- she likes tech stuff like AI, NFT,... she minted a photo of her dog
- she works in a VR company , loves everything to do with technology, and has a pet dog.

- that is into technology and has a dog
- She likes those high-tech things, such as AI and NFT.
- she can do various body motions but not much emotional expression in general.
- I found avatar can show very obvious body language and hand gestures, but it is still hard to see the facial expression. But I can learn their emotion through voice
- Yeah,I saw the building and place looks nice with photos and poster on the wall, and find the red point for the place character pointed.
- She likes to talk very much, but has a very gentle personality and is very understanding. She likes dogs very much. She's not a very polite person,
- She is very talkative and expressive. She is probably meeting someone(me) at the bar for blind date.
- The avatar talked a little about where they work, that they love all kind of technologies and have an eight-years old dog Billy.
- I learned that they are the Chief Marketing Officer of the club (top social VR club). They are interested in NFT and they love their dog.
- She likes technology, including AI, VR, crypto/NFTs. Also has a cute dog.

What did you like about the avatar?

- I like the expression of emotions through the voice. I liked the range of motion while dancing.
- her attitude
- She was very enthusiastic
- I liked how she understood that I was nodding and not much of a talker.
- her actions while she was speaking
- Visual and social cues, body language and delivery
- I like her voice.
- She was very communicative and tried to create conversation and even compensate at times
- I liked that the avatars voice was not robotic and sounded like a real human voice. I also like that she interacted with me as the character. For example, she wasn't just talking about herself she asked if I liked certain things, like her dog.
- I think the animated avatar is better than the screenshot one.
- her outfit
- she was very friendly and engaged in good conversation even though I was not talking at all.
- that is caring because recommended Pete to drink tea and also seems like she was interest in talking to Pete
- She has a lovely dog.
- the avatar is really talkative and has lots of stories to talk about.
- I like it when having a conversation with the avatar, it can include hand gestures. It feels more natural. And when we are talking about a person or thing, we can directly point to that.
- I like the topic(because I like dog), and CV sounds really nice, maybe we can see more examples in sims4? with more dramatic moves to show the action , even there is no words, but poeple all know their mood, and the action and sound look not together in the first(?maybe I remember wrong) and no music no colorful

light(there should be in some places like this), the place looks just four person, and the angle of me, can be more body in my eyes, I mean like if we are nromal person and walk, we can see our shoes and hand wave, and little nose, in story, we can change the dog's photo from the table to her phone(unless she is the boss of the bar), and directly point somewhere looks littl strange, what about use eyes or other actions to show the point they want?(but red point can see clear the place people want, I also like this)

- she likes dogs
- I like the tone and flow of her voice it's naturally engaging and exciting.
- I liked the general introduction of the avatar, she seemed nice and engaging especially when talking about having a cold and about the dog (I actually answered this one). Also, I noticed that the avatar had some nice hand gestures and body language in terms or rotating toward the responder or while looking for bartender.
- I like that they were very enthusiastic and they really enjoyed Pete's company.
- Full range of motion

(Optional) How well did the avatar do in comparison to a real person?

- Good, though the lack of facial expression meant you had to focus on the tone of voice to judge emotions.
- i think almost 60%
- The avatar was very straightforward, unlike what a real person would do at a nightclub.
- pretty we'll
- She did good just feel the voice over did not match her personality
- The avatar does not much like a real person. Especially the facial expression and hand, arm movement
- I feel she is looking good, and the movement is about to coherent. Probably need more work on the oral for example the mouth shape, and the voiceover.
- not too realistic since her movements were not natural, her mouth does not sync with her speech (maybe better if using mo-cap?)
- it did well, the only thing that separates the avatar form the real person is looking eye to eye with the other person
- the gesture and body movement was ok but the facial expression was not as same as a normal person
- It is very much like a real person. The avatar can read my body language and keep the conversation going.
- same answer with question 5
- The model's expressions and body language were uncomfortably stiff, and her emotions were expressed in her voice
- Pretty well, interaction with the avatar is more engaging than I expected.
- I did not expect that the avatar would be so engaging, once again I liked the body language, however, the avatar was really lacking the mimics/facial expressions and was not able to reflect the responder (well, there was no input), which is the thing real person do a lot.
- I think it lacked a bit of facial expressions which I find sometimes hard because I do not know what they felt.

- I never thought it was a real person. Just a person controlling an avatar.

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- The enthusiasm in the voice!
- Her outfit, like a fashion woman
- The avatar was excited when she pointed out a photo of her dog at a club...
- her movements and actions made me remember her
- The environment played a part
- She pointed her dog's picture by little red box
- She helped me remember things when she asked me if I like something. For example, when she was talking about her dog, she also asked me if I like the dog.
- Her movement, but just a little bit.
- she has blond hair, and she has a golden retreiver. she was really excited to meet and talk to me
- the tone of voice and gestures
- the dog picture
- body language such as pointing at the photo and "bar tender"
- yeah the action and the words are not hard to listen, this will help
- Pictures of dogs at the bar, her exaggerated body movements and yelling
- Her emphasize in tone during conversation and her body language.
- The things I remembered the best were the things I can relate to, so I vividly remember the part where the avatar discussed their dog, otherwise, there were not that many things that helped me remember something specific about the avatar.
- When they were looking for the bartender to give Pete tea. It was engaging and felt personal.
- Not really

(Optional) Did you notice anything worth mentioning about the avatar’s performance?

- I noticed she was dancing very awkwardly at the beginning of the video.
- how she introduced her dog and the way she was looking for barkeeper
- Tone of the avatar, was not genuine and felt forced
- arm movement is good
- The movements were not all the same, she points, dances, etc. This gives a better feel of the avatar's movement
- not yet
- it's good at looking at the person's face direction (yes/no)
- the dance
- the dancing looks really nice in the first, and beside person looks also dance nice, with different type of body

- The only thing that's kind of affecting the avatar's emotional expression is her unnatural eye movement.
- The part with going to the bar where it's quieter was very believable, however, I did not understand so sudden change from talking about hobbies, pets and general stuff to talking about cold, perhaps, I just haven't heard the voice of the responder correctly. After that the connection was somewhat lost but still believable.
- Looked like a full body motion capture was used for the avatar's movements.

Group: 2A	Scenario: Pitch	Avatar: VR	Exposure: 2

What did you learn about the avatar?

- Their interest in the game they are designing.
- He explained everything detailed.
- The avatar is pitching a video game
- I learned that the avatar was telling me a game using knuckleheads.
- avatar was introducing the game
- Goth persona, hair, and style. Not sure she's heterosexual but i could be wrong.
- He designed the game
- Not much, just heard him talk about the game
- They did not explain much about themselves but I learned that they are running the pitch meeting and wanted me to be apart of the game.
- What to prepare a game design pitch.
- the avatar has a cool fashion sense, with passion for the game they want to create
- the avatar is developing a game.
- that it's very passionate about the game
- He created a video game.
- more leg movement but not a lot of facial expressions.
- The avatar moves its mouth and blinks when speaks.
- learn about the idea and some precessing draw of this game.
- This model is much more human-like than the last one
- He is part of the game design team and is in charge of the pitch.
- The avatar was pitching the game they are making.
- I learned that they are part of creating the sci-fi action game.
- Nothing about the avatar itself other than that it developed games.

What did you like about the avatar?

- I liked that the body language and movements matched the tone of voice and the emotions that seemed to be present.
- His body movement act really well as human being.

- Talked coherently
- I liked how the avatar was very descriptive when he was pitching about the game.
- the hand movements while speaking
- Hair and style, voice over was better, emotions was good
- voice and outlook(outfit style and hair color)
- Very engaging in his voice, there was more flow in the way he talked rather than it being stagnent.
- I liked that the avatar was slightly moving when he was taking, he overall had good movement. Also when he first said hi, the avatar seemed very nice.
- His/Her style is really cool, and his movement is much more natural, especially the sight, it's similar to reality.
- body gestures are natural (hands, head)
- there wasn't much for me to notice about the avatar since he talked about the game only. his voice was very clear and easy to understand. with nice pauses in sentences.
- that used hand gestures to interact with the pitch description and to indicate what screen to look at
- I like the details of his outlook.
- the clothing style is younger than the first avatar being shown; seems to have a specific personality.
- I like when the avatar is talking, they look me in the eye, and the body is naturally moving with what they say.
- I like the model, this looks better than the first one, but the action doesn't look good, I know it's halloween, but people looks floating in the air and the background people look little horrible. I also like the drawing on the wall, this really help to understand what he is saying.
- Facial expressions and body movements are more natural in conversation
- He enunciates clearly during the pitch and there is sufficient pause.
- I liked their look, very unique and distinguishable both in artistic style and as something that reflects the real person. I also liked the presentational style and the way avatar describes the product.
- I like that the way they pitch their game.
- Generally a very clean and stylish model.

(Optional) How well did the avatar do in comparison to a real person?

- Still got something not very smoothy like human.
- The avatar spoke in a professional way, with no stuttering or mistakes whatsoever. Which might be a bit different compared to a real person.
- He does not much like real person
- The movements were not that awkward and resembled a real person.
- 80% like a real person
- body language feels natural (except for the legs, they are not standing still for some reasons), facial expressions can be improved more
- The emotional expressions seem to be limited and not sufficient enough to be a real person.

- looks better than the first one, but the action still no good enough, maybe eyes should pay more attention on me, the feeling I get, like, the avatar really doesn't like his job and me, but he should continue to do that, or he is just tired? and there are some electric sounds in the first
- Still stiff, but more convincing than the last one. It feels like a real person talking
- I don't think he did a good job. Not sure whether I'm inherently not interested in his pitch or this avatar is just not engaging as the last one.
- The avatar's movement was very unrealistic while rotating towards the presentation, also I was lacking both facial expressions and gesturing, so it was less engaging that talking to the real person.
- Not very well, its facial expression remained unchanged for the entire duration of the video so it didn't feel like the avatar was that engaged in what was happening. Again, just felt like a person controlling the avatar.

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- His green hair helped me remember the conversation.
- not really
- His/her movement and eye sight.
- Not really. In fact, the clothing style and color kind of draw my attention more rather than focusing on the conversation.
- eye contact
- no,,,
- His mouth opens and closes when he speaks
- No it was hard to pay attention to the conversation when the avatar is glitching.
- No

(Optional) Did you notice anything worth mentioning about the avatar’s performance?

- I noticed his gestures were moving unnaturally.
- not really
- The avatar would have different head moments, sometimes slightly looking down , etc. They held eye contact the whole time.
- eye sight, really impressive.
- the avatar sounded confident and passionate about their project
- there are more legs movement from this avatar compared to the first one.
- The avatar sometimes has some displacement without moving their feet.
- yeah, the model looks really nice and sentences are more coherent, rather than mechanistic instructions,I also feel more involved.
- The movement of his mouth does not really match his words.
- The avatar had a lot more robotic voice than that of the first one. While it did not feel bad for the presentation style, the actual simulation of the dialogue would have been awkward. Also, I did not understand how the avatar reflect the input given. Let's say that the responder nods when being ask if the avatar should

continue (based on camera movement), but the avatar does not respond and continues. On one side it reflected given input (nod) on other I was lacking actual communication.
- It kept sliding around on the floor which was kind of distracting.

Group: 2B	Scenario: Disco	Avatar: VR	Exposure: 1

What did you learn about the avatar?

- Shes blonde
- she loves technology and her dog
- She likes her golden retriever
- They're the manager of some kind of tech company (accent was hard to understand) and their hobbies include anything technical like AI and NFTs
- The avatar made gestures according to what they were saying and was talkative
- She likes NFTs and has a NFT of her dog, she also mentioned her dog's name. She wanted to drink whiskey. She mentioned her name, but I forgot it.
- It can move like humans. I first thought that there would be an option that users must choose to do certain gestures.
- They love technology and NFTs (showed me an NFT of her dog), and they are the chief operating officer of some company.
- She's a chief marketing officer
- the avatar had a full range of motion with its arms and some range with their legs, the avatar's facial expressions conveyed their speech well, but lacked some emotion. the hands also felt a little stiff
- She is outgoing, loud, and friendly.
- the avatar is chief of marketing at a software company. They enjoy anything involving tech such as nfts, etc.
- Very outgoing
- she is a CE in a company she is obviously interested in me!! she has a dog called Bella, 8 year-old, and her image was sold as NFT.
- she has a cute brown dog that age is 8 year old. she likes something like AI.. she also got terrible cold.
- avatar is like a virtual friend
- I learned she has a dog, I learned she is a marketing manager for the vr game I'm in I think. I learned she likes technical things.
- they have a dog
- She's the chief marketing officer of the VR platform we were using, loves anything technical, and has an NFT of her dog
- In marketing, loves tech, loves 8 year old dog, has NFT of their dog
- has a dog whom she loves very much and caught something that had caused her to have a sore throat
- she invests in NFT and cryptos

What did you like about the avatar?

- She kept talking and gesturing
- she was very expressive with her arms
- She has an expressive voice
- The dialogue felt fairly natural and their movements felt fairly organic
- When looking at the photo, it looked like the avatar's gestures were a little stiff but in the video, they seem fine and a little more natural as she were moving while talking.
- I liked that the movement was pretty smooth and expressive, and additionally, I liked how the voice matched the avatar's facial expressions pretty accurately. I liked how the avatar was representative of a real person in terms of looks and clothing.
- The gestures and body language are surprisingly realistic
- Seemed friendly
- nothing particular
- the design was well suited for the environment it was in. the design was cartoonish enough to not feel it hit the uncanny valley, but that is hard to completely avoid.
- That she is outgoing, loud, and friendly.
- I liked the expressive talking and the movements/gestures were smoothing than I thought theyâ€™d be
- Attentive towards people
- she was so engaging and gave me the feeling of importance.
- when we meet first time, she hug and kissed me. I felt that she was so friendly.
- her attitude
- I liked the range of movement of the avatar, it expressed emotion well. I also thought the movements were pretty natural looking, which is hard to do in vr sometimes.
- sounded excited
- Very animated, expressed emotions pretty well
- Friendly
- very friendly
- the way she interacts with me was intuitive

(Optional) How well did the avatar do in comparison to a real person?

- i think the avatar did a good job of seeming like a real person though some of the syncing of voice and animation felt somewhat off
- In comparison to real people, this avatar feels a lot less reactive. While this isn't unique to this avatar, video game avatars' conversations generally feel less natural due to their interactions. Independently the avatar seem more realistic, but in conversation, the feedback the avatar receives from me feels unsatisfying.
- It felt like the avatar was controlled by a real actor. I would probably not engage this avatar in a game after the first time interacting with them, but they felt real
- The gestures of the avatar seems awkward and not something I would see when talking to a real person. It seems almost a little too exaggerated.

- I think it felt like a real person, but it might be partly because of the expressive voice. Mostly I felt that the avatar matched the voice pretty well, which made it seem more real. Generally, the way the avatar behaves makes it more human-like than its appearance.
- It is surprisingly well because it does act like a real person instead of just standing still and move like robots
- Facial expressions were missing so I relied a lot more on the tone of voice. There was body movement and body language but it felt a bit forced and unnatural. The graphics were not super clean/crisp which made it feel more like animation than a real person.
- The avatar looked, sounds, and speaks creeply and robotically.
- there were a few instances where the range of motion failed the avatar, especially in the fingers and legs. VR technology will need to work those kinks out to feel like they are more fully fleshed out.
- Worse, due to the lack of facial emotion displayed.
- I think the gestures were relatively accurate however the quality of the graphics and how it looked always took me out of the conversation. It also was pretty noticeable the changes in volume didnâ€™t seem real to how it would be in person and the mouth movements didnâ€™t sync up. The flow of the conversation seemed on par with a real person.
- It was pretty similar to how people would act
- her behaviour, voice and verbal acts were perfect, but I guess her physical gestures need more work to be more realistic, it still does not gives me the sense of the real person.
- Her gestures and words are just like real people. I could immerse myself in her conversation.
- The interaction followed the same steps that an in person interaction would, however it felt much more scripted. The transition from dance floor to bar, and bar to stairs was like playing a video game quest, not like talking to a real person.
- They did okay, the lip syncing with the voice was often off which made the disconnect between a person speaking into a mic and an animated avatar quite apparent
- Hard to answer. I was playing a passive role in an exercise.
- It was engaging similarly to how a real person would greet you at a club. felt like I was meeting a long time friend I met online for the first time.
- the movement was a bit weird

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- the image of the dog really helped with remembering her likes
- Images
- Her accent was hard to understand but added credibility to her character. She included personal details that felt pulled from the real actor's life, but felt pre-canned - like the details you would pull out for an interview or meet-and-greet. The believability of these details made them easier to remember
- The avatar's voice

- I think the avatar's physical gestures helped, such as when she was introducing the NFTs and her dog.
- The body movements
- Holding up the NFT of her dog as a visual cue/reminder of what was discussed was helpful
- I wouldnâ€™t say anything in particular based on the avatar. I more so remembered parts of the conversation based on keywords that stuck out to me
- The loud voice of the avatar
- Her social and warm character, and her kisses :))
- body gestures.
- Tone of voice
- Her hair and facial expressions helped me understand the situation. Her body movements also helped a lot.
- The photo of the dog helped me remember information the avatar said about the dog
- When the avatar had mentioned that I seemed to be very quiet. Even though I was not controlling the avatar, or it was my avatar to begin with, I felt like I was a part of the conversation.

(Optional) Did you notice anything worth mentioning about the avatar’s performance?

- Voice
- The physical movements were surprisingly organic feeling, even compared to some mocap I've seen in games like Detroit Become Human. The hand gestures were believable but felt kind of exaggerated, like she felt she had to work hard to speak with her hands, which I'm not sure was needed.
- I think when the avatar was 'walking', it seemed more like it was floating to its destination and when the avatar invited the user to her place, it seemed almost like she teleported away to a far spot when the camera angle was pointed at her once again after the user turned towards her direction.
- I think that the highly expressive voice might possibly make the avatar more expressive/human-like rather than its physical appearance.
- It felt a bit scripted, and a little forced. Almost like the tone of voice was forcing them to be excited.
- not particularly.
- The avatarâ€™s interests are very likely to a real human
- she was so warm and the location was basically a bar, she was inviting me to her house for a drink and you know, based on these information I guess her clothing was not matched with her character, she needs more tempting clothing.
- she showed me the dog image. so I can easily get interest of her
- maybe improve her ability to hold objects
- I think the way her avatar moves is very unnatural. I noticed it a lot when she was walking away from my character, I don't like how it appears that she travels through my body at some points, it ruins the immersion.
- Some of the hand movements helped make the character a bit more realistic.
- perhaps worth mentioning the avatar's engagement to their surroundings, for example waving their hands in front of the avatar that was beside us at the bar.

Group: 2B	Scenario: Pitch	Avatar: Desktop	Exposure: 2

What did you learn about the avatar?

- The avatar was trying to talk about his product
- that he was making the game
- The avatar had a pitch in a video game concept that he was passionate about
- They want to pitch you a game
- The avatar seemed like he is having his game pitched to the user, hence the more formal and business-like gestures
- The avatar is probably a game developer type of character and likely the leader of the team. I assume we are working with the avatar in their team, and they are pitching the game to us to get our input. They hope to release their game on the switch, PC, etc and release amiibos to boost game sales.
- The avatar is there to explain the project on the wall
- They were game engineers building a video game. Sounded very professional and knew what they were doing
- did not recall
- the avatar has several emotes it can use to convey hand motions and expressions to emulate real hand gestures while speaking.
- They know their game, how to appeal to the audience, promotion, etc.
- not much from what I can remember
- The avatar is likely to work at a game developing company
- He was not that much passionate about the game, it seems that he is only there to introduce the game to me.
- he talked about the game, the robot, rocket power, something chaos, design thing.
- the avatar is professional but I cannot feel the friendliness
- I learned its pitching a game and that it is the developer of the game.
- they are pitching a game concept
- The avatar was pitching the aforementioned game and discussing how it could sell effectively on Nintendo Switch, PC, and Xbox
- Nothing personal or character driven. It was just describing the game
- the avatar is making a game, and is carefully trying to explain the game to us
- talks about the game development

What did you like about the avatar?

- The avatar was actively engaged between me and the product screen
- he seemed alot more human like
- Outfit was cool
- They looked cool
- The casual style of the avatar. I can imagine it in a game.

- I like how the avatar looks physically unique and memorable, and I feel that the physical body movements seem natural.
- Hand movements are smooth
- I really liked their outfit. Sounded more natural like they were giving a pitch. Using hand motions was also a good addition
- passionate, moderate body language and non awkward
- the design was appealing, the gestures helped it feel a bit more alive.
- I like how I could see their hand gestures, but still not really facial movements.
- how it looked. The voice would become a little less clear when turned away which made the avatar seem more realistic.
- The avatarâ€™s voice was really nice to hear
- His appearance
- hmmmm his character outfit?
- body language
- I liked how the voice sounded farther away when their head was turned away from me, it wasn't implemented perfectly, because the distortion was very immediate instead of gradual, but the effect was noticed.
- the hand gestures they used when talking
- Their jacket and pants were pretty cool-looking, almost cel-shaded look reminded me of Overwatch
- The aesthetic of the avatar
- I liked their upper body movement, but would like to see more movement in the lower body. Would think that walking and explaining through the storyboard would make the explanation clearer.
- the gestures

(Optional) How well did the avatar do in comparison to a real person?

- did well but lacked expressive motion in limbs
- Movements were erratic and shifty at times. Occasionally sliding left and right.
- It felt like a real voice actor and that that person was kind of controlling the avatar, but it didn't feel like a real person
- I think the gestures of this avatar was very stiff. I believe the avatar was also making the same gestures over and over again, which I guess for this context it was fine.
- The hand movements are the only thing that is similar to a real person, however it is too repetitive
- similar to a real person
- outside of VR, it's slightly harder to do nonverbal communication. The emotes the avatar was doing helped it feel more alive, but the animations were a little too stiff to be believable. It made it feel like a roleplay of a person, rather than a believable motion.
- gestures werenâ€™t anything crazy, but they were subtle enough to feel like a person would gesticulate like that while going on a long pitch.
- Similar to a person who is passionate about their work
- his verbal ability was low and the pronunciation of the words and his mouth motions were not the same.

- I think he looks like real person. but I did not interest about his talking topic. that is why I cannot understand what he want to present.
- I don't think this one did as well as the first one. It's movements were much more jagged and less human-like. I really didn't like when its head moved separately from its body, it took any sense of reality away, and looking jarring.
- The avatar was very static and not very life like
- did not feel as engaged in the conversation. For something that was very context heavy, it would help to. use the storyboard early on in the conversation, similarly to how a real person would in my opinion
- the appearance is not matching the voice for some reason, and the hands are overlapping sometimes so I feel the avatar is unreal

(Optional) Was there anything specific about the avatar that helped you remember the conversation?

- No, I had to listen to it twice. Listening to a pitch for something is pretty boring in comparison to listening to someone talk about themselves though
- When the avatar kept going towards the presentation's screen when presenting his game idea.
- When the avatar gestured more drastically/had a greater range of body movements such as pointing to the board, I had an easier time remembering the conversation about amiibos and sale boosts.
- The avatar pointing at the project on the wall
- the natural movement of the avatar, hair color
- not particularly.
- The avatarâ€™s voice was nice to hear
- I felt like its voice didn't match its appearance. Its features were very feminine but the voice sounded masculine. It's clothes also looked like they belonged in a sci-fi game, and didn't really match the setting that we were in, which was a mostly plain looking room.
- The diagram on the back was more helpful in remembering information than the avatar itself
- hand gestures were great and engaging, but could be used to point at something more relevant in the conversation
- gestures

(Optional) Did you notice anything worth mentioning about the avatar’s performance?

- The character moving but the legs not walking was distracting. It also felt like the arms lifting was coordinated with the actors movements, but the hand gestures didn't feel as natural
- I think the avatar's blinking and hand gesture was very distracting. The more I look at it, the more I feel that it is very robotic when it is moving. Also not sure if its the video quality or something else, but there seems to be a delay with the avatar's voice and its mouth movements.

- I felt that in this avatar there was a bit of a difference in the range of emotion of the face compared to the body. I felt that the body movements were greater and more natural compared to the face. The facial movements seemed to be more lacking as while the avatar spoke, there was little to no movement but the body had small but natural movements comparable to a human's when talking.
- It's quite poor compared to the first one
- not particularly
- I think if he want to do presentation, then use more image is help to understand.
- his emotion is not flexible
- I think the movements are what made it inhuman looking. Its stance was also very rigid.
- The avatar seemed pretty monotone and was pretty lifelessly selling their product. The range of motions and expressions didn't help with making them feel passionate.
- The avatar did not add anything to the situation. The voice was realistic and there was a diagram in the background. This presentation would have been just as effective as a slide deck presentation.
- the hands are overlapping sometimes

Appendix C.

End of Survey Qualitative Responses

Group: 1A

Fav. Avatar	Please explain the above choice	Is there anything you'd like to mention?
1A_Desktop	Avatar 1 felt more approachable and congenial. Avatar 1 was also slightly more expressive and had more varied and less "computer-like" robotic movements.	
2B_VR	The second avatar was much more expressive as she had a wider range of vocal ability, larger and more dynamic gestures, and more physical interaction with her environment.	The environment of the second avatar (bar/club) was also more believable and relatable than an empty room for the pitch meeting in the first avatar's situation.
1A_Desktop	I liked that there weren't a specific disadvantage of listening to the presentation in the virtual world. It got me to think that people can interact more in the virtual world in the future.	I hope that the development of the virtual world gives us advantages, rather than the disadvantages of cutting f communications in the real world.
2B_VR	Her voice is more three-dimensional, and she has more moving, and more functional	I think the modeling of the avatar needs to be more like the real person, the expression needs to be more obvious, and the scene also needs to be more realistic.
1A_Desktop	Video one has a much normal interaction rather than the second one makes you question why you're there	First video needs to sound a little less robotic/monotone
2B_VR	I personally would feel more engaged and more likely to communicate with the 2nd avatar. The 1st avatar seemed more calm, but it is hard for me to predict their actions and thus I don't feel safe around them.	Nothing.
2B_VR	enthusiastic and felt like the conversation was real.	nice dance moves on the dance floor by the supervisor of this research Unlike in the first video Zombie like behavior
1A_Desktop	avatar 1 was calm and collected. Their explanations were easy to understand and i	

	felt as though i could remember more when i was with them	
2B_VR	the one in the second video has a better range of movement, feels more real, and showed more emotions	
2B_VR	The second Avatar was much more expressive with their body language, making them feel more real. It also felt like I was talking more to a person than a robot, but that might be more with the avatar's personality	
2B_VR	Avatar 2 seemed to be very friendly and fun to be around. Also, they tried to come up with solutions when the player had a problem.	Avatar 1's voice was very robotic and monotone, so it didn't give as much human like interaction as Avatar 2.
2B_VR	This is because the second one has more of a personality and is less emotionless than the first one.	Their lip movements and the audio don't really match.
2B_VR	Compears to Avatar 1, avatar 2 looks less than a real person. However, her vivid emotion, movement, and interaction with her surrounding are way better than avatar 1. It convinced me that is is an avatar with human characteristics.	
2B_VR	Her body movements and gestures were much more natural even though her form was not the most pleasing. Her unique voice also helped with keeping me engaged.	
2B_VR	In comparison, the body movements of the second avatar are much richer and more detailed than the first avatar. The first avatar has only mouth shapes and simple arm movements, while the second avatar has more head movements as well as hand movements. When talking, the second avatar also has more speech-matching movements, such as when looking for the bartender it probes into the bar to find it. Compared to the first avatar, there was only a simple turn, and no further actions to cooperate with the dialogue.	The second avatar allowed me to see a wealth of action and detail that I hadn't been exposed to while experiencing VR on my own. The second avatar allowed me to see a wealth of action and detail that I hadn't touched while experiencing VR on my own. But I also noticed that the same problem I had when operating the avatar was the movement of the avatar. Because of the limitation of the field of reality, I could only use the joystick to move the avatar, and on the second avatar, I felt that I saw something similar way to move.
2B_VR	The avatar in the second video seemed more personable, as if someone who was similar to them in terms of mannerisms and	

	means of conversation exists in real life. They displayed a lot more personality in the short video clip compared to the first avatar, and the second avatar seemed like they were fun to talk to, very conversational, and seemed overall like someone pleasant to be around.	
2B_VR	The second one was more expressive, especially with physical closeness and vocal variation.	I think the virtual settings also had an effect on the communication for me â€" I was distracted by the drawings in the background for the first video, while the second avatar specifically changed the location to be easier to focus.
1A_Desktop	Spoke a bit more clearly.	Second video being a "date" setting made me cringe a little so I didn't listen to too much of the conversation
2B_VR	Avatar 2 feels more like a real person to me because she talked with more emotions. The way she tried to interact with me showed that she was more friendly than the first one. This helped me to be more engaged in the conversation. Avatar 2 was also funnier than the first one. I like her reactions.	
2B_VR	The second avatar feels like a real person instead of controlled by someone. the first avatar feels like a robot and just read out from the script	
2B_VR	Avatar 2 had tracked control of their legs while Avatar 1 did not. Avatar 2 also did a better job demonstrating their range of full body movement.	I did not like the NFTs part
2B_VR	The second one engaged a lot more with the user which shows ability to communicate. Also, due to them being more realistic as a person, it felt that they were better at communicating.	

Group: 1B

Fav. Avatar	Please explain the above choice	Is there anything you'd like to mention?
2A_Desktop	I would say that the second avatar was more	

	interactive and the topic of the conversation was more general and colloquial so it was easier to communicate even if I was not talking. So definitely, the avatar was more engaging, as if I was actually talking with her. The second avatar had more realistic dress code and the voice had much more variation in tone than the first one which makes it easier to tell the emotion of the avatar and hints the personality of the avatar was well. The avatar also asked questions which shows a good understanding of what "communication" is, as it is a talking in both ways instead of just one person talking which is what the first avatar did. Also the gestures and the voice were in sync which made her seem more real. Also, the avatar seemed to have a gender, ethnicity, age unlike the first character that seemed like a character from a cartoon or something.	
2A_Desktop	The second avatar seemed to have a wider range of motion i.e. pointing, turning, dancing, and walking around	N/A
1B_VR	The timing of their gestures and body language was key. Avatar 1 used limited hand gestures, and only turned when necessary. Sometimes, less is more, and for this reason, they felt more realistic than the second avatar, who was constantly moving while they talked and it was hard to follow (too distracting).	Enjoyed this study! Good luck team
1B_VR	I like the expression of the first avatar better than the second one, and the movements seemed to be more natural.	
2A_Desktop	The second avatar is more close to a real person in terms of the way she speaks, her body language and how she interacts with the things around her.	
2A_Desktop	The avatar was dancing, legs moving around and jumping. It felt more believable. The first avatar was stiff and even when turning towards the screen, its feet were awkwardly shuffling.	Facial expressions were lacking which is the main thing that lowers believability
2A_Desktop	I feel like the method of communication in the second video was more immersive for me. I felt more engaged in the conversation since it seemed like they were having a conversation to us rather than lecturing us (informing) in the first one. The second avatar seemed to respond to us a little more.	Nope!
2A_Desktop	her movement with here arms seemed more	

	wide range and the point that shows what she is pointing at helped	
2A_Desktop	The hand motions were more believable and the way it spoke had more personality to it.	For some reason having the avatar dressed more realistically made it more realistic to talk to.
2A_Desktop	She has some interaction with me and the context she said is much easier to remember.	In the second scenes, the dancing movements of two people on the stage are supper weird lol.
1B_VR	The model was more pleasing to look at aesthetically and its movements felt more realistic.	While I preferred the first video's avatar I did also prefer how the second video's was much clearer to hear (being heard in both ears clearly)
2A_Desktop	The second avatar had more body gestures as well as facial movement. She also had an interesting voice.	nevertheless, both caught my attention with good body language and verbal communication.
2A_Desktop	The second avatar seemed more like a normal person. I think I would be less likely to run into someone with green hair so it personally feels less real.	
2A_Desktop	Avatar 2 was more engaging, I definitely remembered the conversation more and saw a larger range of movement for avatar 2.	Avatar 1 is good too but compared to avatar 2 it could be more to engage the person. Visuals and movements helps greatly.
2A_Desktop	in the first video, avatar did not move much around the room so it is harder to tell about their degree of movement compared to the second one. Also second avatar had more human feature rather than first one that looked a bit cartoonish (soft face features)	However i believe first avatar is able to move their eyes more than the second avatar
2A_Desktop	This was a difficult choice for me. I would have liked to choose the avatar in video 1, however the transform jitter made it difficult for me to focus on what they were saying. I couldn't remember a lot of what they said compared to the second video because of that. I felt the first avatar had more potential for communication and collaboration, but the second avatar's ability to point (and have the red icon show up on their target) was helpful in giving me a direct visual to what they were speaking about. For this reason, I would choose the avatar in video 2 at this point in time.	I feel that if the slight shifting of the first avatar's physical transform was less noticeable, and the ability to point was enabled and/or demonstrated, I would actually have chosen the first avatar. They felt very friendly and already expressive and kind in their design. While I did like the second avatar, they reminded me of NPCs in past games I've watched people play, and felt a little less expressive in their design. I was surprised by how fluid some of their movements were, though!

1B_VR	The first avatar looks better and more detailed	Maybe two avatars should be operated by the same person, the emotional expression of the operator has a great impact on the experience
2A_Desktop	I feel like there was more range in movement and body language	no
2A_Desktop	Regardless of twos personality, it felt more like interacting with a real person because it felt like they had a mind of their own. They were unpredictable and their movement didn't seem like it was controlled by someone else. They could express thoughts and ask questions and initiate well when communicating.	Even though the avatar in video one was more laid back and came across natural, the avatar in the second video mimicked human interaction more. In video one I felt like I was interacting with an avatar but in video two (regardless of the outfit and look of the avatar) I felt like I was interacting with a human.
2A_Desktop	I really like how the movement of the arms is more detailed compared to the first avatar in terms of how it even includes the fingers.	I like how the first avatar's body movement is broader and more distinct when talking
2A_Desktop	The avatar in video 2 is more realistic and the topic is more engaging. Whereas in video 1, my thoughts start wandering off during his speeches.	
2A_Desktop	The range of motion was much more lifelike, the clothing was also much more like something people wear in real life. The first avatar's clothing seemed right out of a pokemon game with all the straps and belts.	

Group: 2A

Fav. Avatar	Please explain the above choice	Is there anything you'd like to mention?
1B_VR	I found the second avatar more believable as a real person as I think their emotions and body language was easier to see.	
1B_VR	Body movement is richer, much better	Eyes and mouth movement can be improve, i know they are hard to complete.
2A_Desktop	Despite the less detailed avatar, avatar 1 had more movement and the pointing gesture was delivered clearly. The avatar was also more emotive.	
1B_VR	The avatar in video 2 was able to discuss	I really like the projected

	more on the subject matter more in more detail than the avatar from video 1.	game drawings on the wall in video 2.
2A_Desktop	I think avatar in video one had better movements and emotions while she was talking	no
1B_VR	I prefer the voice over on 2, she caught my attention more and was a curious character. I did enjoy 1 but the voice over was not giving it for me, also I liked how the scene of 1 began where she carried the audience somewhere else but 2 was overall more put together.	N/A
2A_Desktop	First avatar has better body movement and language performance	The avatar in video 2 his feet are unstable, like skating. also his hands movement and eye contacts are little weird. overall his body is not natural.
1B_VR	There was a higher level of detail in the visuals of the second avatar making him more engaging and realistic. His voice in the way he would speak had more flow and didn't feel fake in that sense.	
2A_Desktop	I chose the first one because she asked more questions compared to the other avatar. I also think that she had more personality than the other avatar because she was a bubbly character.	I think the the overall movements on the second avatar where better than the first one.
1B_VR	The second one is more convincing for me.	Like the guys dancing in the first scene.
1B_VR	more natural and realistic body language, facial expressions are a little better	
2A_Desktop	The avatar in video 1 seemed to interact with me and would ask for my opinion. there was a conversation.	
2A_Desktop	it included interactive buttons for the avatar to go through and asked multiple questions throughout the discussion and seemed to care more for the person	I like how the avatar seem interested in the person
2A_Desktop	I think the Avatar in video 1 is the best one. The way she interacts with the woman near the bar makes me feel like she is a real person who is trying to find a bartender.	Both avatars in videos 1 and 2 are kind of lack of emotion on their face. It will contain more personalities if both avatars could have more facial expressions.
1B_VR	I like clothing styles and the first impression of the second avatar more. I think it's because the avatar's appearance seems to someone the same age as me.	I wish there were more facial expressions for each avatar so it would be much more engaging for me to remember the conversation with that avatar.

2A_Desktop	Although Avatar in video 2 feels more natural to me with its body movements and the way they looks at me. But Avatar in video 1 made me feel more enthusiastic. And the greater range of movement also draws my attention to her rather than other persons or objects in the scene.	
2A_Desktop	Obviously, although the first model is not as good as the second, the first character is more enthusiastic and the words are easier to understand, so I will choose the first one.	what about adding subtitles to let people understand more?
1B_VR	The second Avatar's communication skills were strong, his voice sounded more polite when speaking and his body language was more natural than the first avatar's	The appearance of the first Avatar was uncomfortable. The eyes were too big and the hair covered half of his face
2A_Desktop	Although there is not much difference of modelling and rendering quality between the two avatars, the first one is just naturally more engaging because of her variation in tones and accent.	
2A_Desktop	While I like the way the second avatar looks (I think it an represent real person better in terms of look), the first avatar feels like having a lot smoother animation for movement in general, more gestures and more vivid and engaging voice.	
1B_VR	I like that they are very engaging not just verbal but also with their use of body language and eye contact. They make sure that the person they are talking to are still part of the conversation and does not feel bored.	I think what is missing from both of the avatars are their facial expressions. While the tone of their voices gives it away, it might still be hard to tell how they feel. Without facial expressions, it somewhat takes away the realness of the surrounding and interaction, and may just look like a doll is starting at you.
1B_VR	Visually it's the most pleasing to look at, even though it doesn't have a wide range of emotions. It does have a more appealing style to it, which feels like the point of virtual worlds like this.	

Group: 2B

Fav. Avatar	Please explain the above choice	Is there anything you'd like to mention?

1A_Desktop	The avatar had more motion and awareness. The first avatar felt like she was reading a script and did not feel like like a normal interaction. The second avatar carefully tried to explain things.	There were times when the avatar's mouth would not be moving but there would be talking.
2B_VR	the first avatar did not look as realistic but had more range of movement to allow expressiveness and seemed like she was able to communicate better rather than just vocals.	expressive motion is very improtant, though the face of avatar one was a bit uneasy she still had the range of character to express what she was trying to get across.
2B_VR	The avatar in video 1 was more responsive to me in conversation. This may be due to the specific scenarios presented, but avatar 1's movement also felt more fluid.	
2B_VR	It felt more like I was talking to a real person inhabiting a digital character. The avatar felt more flexible. For example, when she leaned over the counter looking for the bartender, that felt real, while the second character couldn't even turn around without me noticing that his legs didn't move.	The first avatar was in a darker environment, which could obscure some of that avatar's less organic motions. It had a more boring but possibly more realistic visual design. The second avatar had really boring conversational material, which could have made me perceive it as less engaging. It had a more interesting visual design, which could have made me see it as more interesting, but it's voice didn't feel like it matched how it looked. It would be easier to compare the two if they were in a similar environment, had dialogue in the same vein, both had similar visual design, and were both voice acted by people who felt personable. That being said, even if I imagine these things as equal, I still think I would pick the first one just on the basis that the way it moved felt more aligned with someone's real movements.
1A_Desktop	I think its just because of the overall look of the avatar. Although avatar 1's overall look seems more realistic, in a VR chatroom, I rather interact with an avatar that is out of the ordinary. Avatar 2 gives off a more game-	Both movements of the avatar seemed stiff and robotic. It would be nice if they were able to move a little smoother and do actual

	like vibe and overall is more visually appealing to the eye as it is dressed more interesting and makes usage of different colours.	walks rather than floating to the next spot.
2B_VR	I think that the first avatar utilized greater physical movements in body and expression. I felt that it was easier to grasp the conversation of the first avatar as they felt more like a real person due to their move expressive behaviour, such as waving and leaning over the bar to call the bartender. Also when they greeted you, they behaved in a way that seemed genuinely excited, probably due to the combination of the expressive voice and more open body language.	I think that the second avatar probably seemed like it communicated worse to me as compared to the first avatar they weren't really moving as much and not as expressive, along with a more monotone voice compared to the first avatar.
2B_VR	The able to interact with "us" such as greeting and hugging plays a huge role in communicating non-verbally and also the movements are not that clunky and repetitive as the second one.	
1A_Desktop	They sounded and acted a lot more natural. Using hand motions was more intuitive and easy to follow.	
1A_Desktop	The first avatar moves and talks awkwardly and seems unnatural, obnoxious and "trying too hard". The second avatar seems genuine, human, caring and natural.	The first one seems to have a very different accent than the second one, not in terms of race/ethnicity but the inotations and pitch.
2B_VR	VR's hand motion and interaction lends itself more to human interaction, given its input methods.	
1A_Desktop	The avatar seemed more realistic, looking more like a human instead of a weird doll. I could perceive their gestures more.	Keep working hard!
2B_VR	although the graphics were worse. I think the first avatar had more fluid movements that really fit the environment	no
1A_Desktop	The first avatar had a rather thick accent which I was not really to be able to comprehend while the second avatar spoke more clearly which was more easy to comprehend	
1A_Desktop	The Avatar 2 resembles the real human more to me.	
2B_VR	the girl avatar is really friendly who is blonde hair and where the black jacket and dance well.	he use lots of body gesture, that make me more focus on the communication.
2B_VR	Avatar in video motivated me to engage more than the one in video 2. I can feel her	

	energy	
2B_VR	The movements of Avatar 1 were way better at showing lifelike movement. I think that the potential for conversation is much much higher in the first avatar, because the second one had many actions that looked robotic. When the second avatar gestured with its arms, it looked much more rigid, and like it had less range of motion than the first one. The first avatar also did things like kiss my character on the cheek, which requires moving its entire body in very particular ways. I think that the first avatar did a great job at emulating such a complex action, and I think the second avatar would really struggle to do something as complicated as that and show it successfully.	I think that if the voice of the second avatar had matched its appearance better, I may have been able to be immersed in the conversation a bit better, and I also think its clothes distracted me a bit. The second avatars movements were the things that I liked the least about our interaction though, because I feel like its movements were very limited and linear if that makes sense. It seemed as if it didn't have free range of motion as much as avatar 1, which makes me think that it would be worse to have a conversation with. The facial expressions of avatar 1 were also way better, avatar 2 showed no real emotion that I could see, and only gestured with their arms.
1A_Desktop	the speech was more clear and the hand gestures helped more with grabbing my attention.	
2B_VR	Way more animated and expressive. The second avatar hardly moved at all, the only benefit to the second avatar is that it seemed the lip-sync was better, and the character model was a bit cooler, but the first one was just way more engaging overall.	
2B_VR	Avatar 1 was more dynamic and we moved around in the space more which helped feel more engaged in the surroundings and conversation.	
2B_VR	more engaging and more aware of the environment along with my own avatar	it feels a bit drastic to compare the two together because the second one had more content, requiring more of na attention span from the listener.
2B_VR	the first one has a matching voice, more leading gestures and high pitch which can draw my attention.	