,

Human-machine interaction:

A robot avatar that looks just like you

Supervised by: Ioanna Giorgi.

University of Kent - School of Computing

MSc Artificial Intelligence

Word count: - 10,667

UNIVERSITY OF KENT

ACCESS TO A MASTER'S DEGREE OR POSTGRADUATE DIPLOMA DISSERTATION

In accordance with the Regulations, I hereby confirm that I shall permit general access to my dissertation at the discretion of the University Librarian. I agree that copies of my dissertation may be made for Libraries and research workers on the understanding that no publication in any form is made of the contents without my permission.

Notes for Candidates: by submitting your dissertation, you agree with the following:

1 Where the examiners consider the dissertation to be of distinction standard, one copy may be deposited in the University Library and/or uploaded into Moodle as an example of good dissertation for future students.

2 If a copy is sent to the Library, it becomes the property of the University Library. The copyright in its contents remains with the candidate. A duplicated sheet is pasted into the front of every thesis or dissertation deposited in the Library. The wording on the sheet is:

> *"I undertake not to use any matter contained in this thesis for publication in any form without the prior knowledge of the author."*

Every reader of the dissertation must sign and date this sheet.

3 The University has the right to publish the title of the dissertation and the abstract and to authorise others to do so.

....
SIGNATURE: Robert Whiting **DATE: 09/09/2024**

....
FULL NAME: Robert Whiting

CERTIFICATE ON SUBMISSION OF DISSERTATION

I certify that:

1 I have read the University Degree Regulations under which this submission is made;

2 In so far as the dissertation involves any collaborative research, the extent of this collaboration has been clearly indicated; and that any material which has been previously presented and accepted for the award of an academic qualification at this University or elsewhere has also been clearly identified in the dissertation.

....
SIGNATURE: Robert Whiting **DATE: 09/09/2024**

Declaration

I certify that I have read and understood the entry in the School Student handbook on Plagiarism and Duplication of Material, and that all material in this assignment is my own work, except where I have indicated with appropriate references.

Signed: Robert Whiting

Date: 09/09/2024

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Ioanna Giorgi, for her invaluable guidance, support, and encouragement throughout this project. Her expertise and insights have been instrumental in shaping the direction of my research, and I am immensely grateful for her patience and mentorship.

I would also like to extend my heartfelt thanks to my friends, who have been a constant source of motivation and support during the more challenging moments of this journey. Your encouragement and understanding have meant the world to me.

Finally, I am eternally grateful to my family. Without your unwavering belief in me and your continued emotional and moral support, this dissertation would not have been possible. Thank you for standing by me every step of the way and for being my pillar of strength.

Abstract

The rapid advancements in computer vision and machine learning technologies have significantly impacted human-machine interaction (HMI), leading to the creation of more realistic and customizable virtual avatars. This dissertation investigates the development of a personalized robot avatar that closely mirrors my own facial features and expressions using Unity and the Furhat Robotics platform. The primary objective of this research is to design a 3D model of my face, integrate it into a virtual avatar framework, and ensure real-time synchronization of facial movements and speech using advanced computer vision techniques.

Photogrammetry, 3D scanning, and blendshapes were employed to achieve high fidelity in facial texture and geometry, while machine learning algorithms were integrated for real-time animation and interaction. The project also explores the application of ChatGPT to enable dynamic, context-aware conversations, allowing for more natural and engaging interactions. Testing revealed that the avatar achieved a high degree of realism and responsiveness, though challenges remain in low-light environments and with certain complex facial expressions.

This research contributes to the growing field of human-machine interaction by enhancing the realism and customization of virtual avatars. It also sets the foundation for future developments in areas such as education, healthcare, and virtual collaboration, where lifelike avatars can facilitate more intuitive and immersive user experiences.

Contents

Abstract **4**

List of Figures **8**

1. Introduction **11**

1.1 Overview.. 11

1.2 Aims 13

1.3 Chapter Structure 14

2. Description of the Topic Tackled **16**

2.1 Introduction to Human-Machine Interaction and the Role of Humanoid Avatars 16

2.2 Problem Definition 17

2.2.1 Overview of the Research Problem 17

2.2.2 Challenges in Avatar Realism 17

2.3 Significance of the Research 18

2.3.1 Applications of Humanoid Avatars 18

2.3.2 Enhancing User Experience 19

2.3.3 Impact on Human-Machine Interaction 19

2.4 Key Terminology 20

2.5 Conclusion 22

3. Literature Review **23**

3.1 Background 23

3.2 Integration of Multiple Technologies 24

3.3 Software Use in Avatar Development 26

3.4 Accessibility and Democratization 28

3.5 Avatar Realism.. 29

3.5 Conclusion 30

4. Possible Approaches and Reasons for Approaches Taken........................ 31

4.1 Possible Methodologies 31

4.1.1 Traditional 3D Modelling and Animation 31

4.1.2 Photogrammetry and 3D Scanning 33

4.1.3 Real-Time Animation with Blendshapes 34

4.1.4 Machine Learning and AI-Driven Animation 36

4.2 ARKit and Unity's Animation and Rendering Capabilities 36

4.3 Integration of AI and Machine Learning 37

4.4 Virtual Avatar texture Application 38

4.5 Conclusion 38

5. Description of the Work Carried Out 39

5.1 Initial Setup and Tool Selection 39

5.1.1 Research and Development Approach 39

5.2 Workflow and Implementation Process 41

5.2.1 3D Model Creation and Customization 41

5.2.2 Integration and testing into Unity 45

5.2.3 Real-Time Expression Mimicry with Furhat Robotics API 45

5.2.4 Speech Recognition, Lip Sync, and Control Logic Using VSCode and Blockly 47

5.2.4.1 Voice Control & Selection 47

5.2.4.2 Control Logic Using Blockly 48

5.2.4.3 Speech Recognition and Machine Learning 50

5.3 Conclusion 51

6. Results, Testing, Verification 52

6.1 The Final Model and Texture 52

6.2 Application of the 3D Model and Textures 55

6.2.1 Texture Mapping and Alignment 55

6.2.1.1 JSON Code Structure 56

6.2.1.2 Adjusting the Texture Mapping in JSON 57

6.2.2 Visual Accuracy and Realism 59

6.3 Advancing the Virtual Avatar 61

6.3.1 Voice Interaction and Gesture Implementation 62
6.3.2 Integrating Blockly for Avatar Interaction 63
6.3.3 Integration of ChatGPT for Enhanced Conversational Ability 65
6.4 Conclusion 66

7. Evaluation 68

7.1 Key Research Questions Addressed 68

7.2 Strengths of the Project 69

7.2.1 High-Fidelity 3D Facial Modelling 69
7.2.2 Real-Time Facial Expression Synchronization 70
7.2.3 Integration of ChatGPT to Enhance Human-Like Robots 70
7.2.4 Summary of Strength 70

7.3 Weaknesses of the Project 71

7.3.1 Limited Expression Range 71
7.3.2 Performance Issues in Unfavorable Conditions 71
7.3.3 Summary of Weaknesses 72
7.3 Overall Evaluation 72

8. Conclusions 74

9. Overview of Future Directions 75

10. References 76

List of Figures

Figure 1: Furhat Robotics showcasing diverse avatars 11

Figure 2: Lineup of the current humanoid robots in development 18

Figure 3: Three-dimensional (3D) research themes evolution over time 31

Figure 4: Demonstrating how Photogrammetry is the practice of stitching 2D images from multiple angles into a 3D object 32

Figure 5: The Uncanny Valley: Plot of Perceived Familiarity vs. Human-Likeness 34

Figure 6: Software Architecture for Speech Recognition 36

Figure 7: Windows 3D Library “Give this person a face” model 39

Figure 8: Avatar Maker avatar created from Reference Picture 40

Figure 9: Reference Images for Full Head 3D Model Creation: Multi-Angle Views 41

Figure 10: 3D Mesh Overlay and Facial Landmark Detection for 4/6 Reference Pictures 42

Figure 11: Eye Texture Rendering Issue in 3D Model: Misalignment 43

Figure 12: Eye Texture Rendering in 3D Model: Alignment 43

Figure 13: Directory Structure for Facial Recognition Data Files in Furhat SDK . 45

Figure 14: List of available Furhat Voices within the SDK 47

Figure 15: Interaction Flow Diagram for Avatar's Conversational States Using Blockly 48

Figure 16: Emotion-Gesture Mapping for Avatar Responses Based on Keyword Detection 50
Figure 17: My 3D-Model with and without Textures applied 51
Figure 18: The Texture .PNG 52
Figure 19: Assortment of Facial Expressions for my Model With no Texture 53
Figure 20: Assortment of Facial Expressions for my Model with Texture 53
Figure 21: Furhat Model with Original .PNG File 54
Figure 22: JSON Code Structure 55
Figure 23: Furhat Model with Width to Width .PNG File 56
Figure 24: Different Textured files to be Tested 56
Figure 25: Furhat Model with Near Perfect .PNG File 57
Figure 26: Furhat Model with Perfect .PNG File 58
Figure 27: Selection of Gesture within the Furhat SDK 58
Figure 28: Model expressing "BigSmile" Gesture 59
Figure 29: Model expressing "ExpressAnger" Gesture 59
Figure 30: Model expressing "ExpressSad" Gesture 60
Figure 31: Voice Interaction Code Snippet 61
Figure 32: Furhat Conversation with Gestures Code Snippet 62
Figure 33: Blockly Visual Conversation Blocks 63
Figure 34: Communication with Furhat Integrated ChatGPT Via Text Box 65

Figure 35: Communication with Furhat Integrated ChatGPT Via Microphone .. 65

Figure 36: Random Furhat model emotional response via ChatGPT questions 65

Figure 37: Expression Limitations Due to Absence of Inner Mouth Detail in Avatar Models 70

Figure 38: Effects of Lighting Conditions on Facial Texture and Eye Detection Accuracy 71

Chapter 1: Introduction

1.1 Overview

In the world of modern human-machine interaction (HMI), the emergence of humanoid avatars as proxies for users marks a significant evolution in virtual communication and interaction as these avatars, capable of replicating human-like gestures, facial expressions, and behaviours, are transforming how we engage with technology in virtual environments [1][2][3], as imagine the possibility of interacting with a digital representation of a loved one or acquaintance through a robot avatar that mirrors their appearance and mannerisms, bridging physical distances and enhancing personal connections.

The historical reliance on verbal communication between humans is evolving as technology enables nonverbal exchanges through virtual avatars. These avatars are designed to convey emotions and intentions akin to face-to-face interactions, leveraging advancements in Human-Computer Interaction (HCI) to create immersive and empathetic virtual experiences. From educational simulations [5] to cultural training [4], humanoid avatars are reshaping how we perceive and engage with virtual environments and those who inhabit them.

Humanoid avatars are therefore not just static representations but dynamic entities capable of interaction and adaptation. In interactive settings such as virtual classrooms or collaborative workspaces, avatars facilitate natural and intuitive communication by mirroring users' facial expressions and gestures. This means it extends beyond the visual fidelity to include synchronized responses and sensory feedback, enhancing the sense of presence and engagement in virtual interactions.

The HCI is then crucial for creating a sense of authenticity and emotional connection as avatars that accurately reflect human behaviour, and emotions enable more meaningful communication, particularly in scenarios where physical presence is limited or impractical as supported through research which underscores the importance of avatar realism in influencing user engagement and interaction satisfaction in VR environments [7]. So the design and development of interactive avatar faces are crucial when attempting to optimize user experience and acceptance in virtual environments, as by studying user perceptions and responses to avatars versus physical robots, avatar capabilities can align with user expectations and preferences.

This iterative process not only improves avatar realism but also explores new avenues for enhancing HMI across diverse applications. In practical terms, the flexibility and adaptability enables customization tailored to an individual needs and preferences as they can be personalized to reflect diverse identities, preferences, and cultural contexts, fostering inclusivity and accessibility.

Figure 1: Furhat Robotics showcasing diverse avatars [8]

In conclusion, the development and integration of humanoid avatars in HMI represent a transformative leap towards enhancing HMIs and as technology continues to evolve, the concept of interacting with an avatar that mirrors human appearance and behaviour becomes increasingly feasible, bringing in a new era of immersive and empathetic virtual interactions.

1.2 Aims

The objective of this research is to develop a customized virtual robot avatar that accurately replicates my own facial features and expressions. By leveraging the Unity platform, a detailed 3-D model of my face will be created and seamlessly integrated into an existing virtual avatar framework, such as the Furhat robot. In addition, the virtual avatar will be programmed using Java and Furhat Robotics' API as the mouth, eyes, and eyebrows, will be precisely calibrated within the avatar's configuration file to ensure natural movements during interactions. This programming shall facilitate responsive interactions, enhancing the immersive quality of HMI. To reach these aims, the following critical research questions have been formulated to ensure they're achieved:

Can advanced computer vision techniques be utilized to create a realistic and customizable 3D model of a human face for use in virtual robot avatars?

This question investigates the potential of cutting-edge computer vision methodologies to generate highly detailed and customizable 3D models of human faces. The goal is to enhance the realism in virtual HMI.

How can the synchronization of facial movements and speech be optimized in a virtual robot avatar using real-time computer vision algorithms, and how does this optimization affect the accuracy of facial texture mapping?

This question focuses on optimizing the real-time synchronization of facial movements with speech during interactions. The aim is to achieve natural synchronization of facial expressions and gestures with speech, thereby improving the accuracy of facial texture mapping and overall visual realism.

1.3 Chapter Structure

Chapter 2: Description of the Problem/Topic Tackled

Focuses on the context and significance of the research topic as well as the challenges being addressed. Additionally, a brief explanation about the key terminology used to ensure clarity and understanding.

Chapter 3: Literature and Technology Review

Summarizes research related to virtual avatars, while discussing the technologies and tools (e.g., Unity, JAVA, Furhat API) used in the project. In addition, gaps in existing literature will be identified and examine.

Chapter 4: Possible Approaches and Reasons for Approach(es) Taken

Explores different methodologies and techniques that could be used to achieve the research objectives while also explaining the rationale behind selecting specific methods and tools.

Chapter 5: Description of the Work Carried Out

Details the process of creating the virtual avatar to describe how computer vision techniques and programming are used to develop the 3D model and integrate it into the Furhat.

Chapter 6: Results, Testing, and Verification

Displays the outcomes of the development and implementation process.

Chapter 7: Evaluation

Reflects on the strengths and weaknesses of the project as well as answering the 2 Research Question mentioned previously.

Chapter 8: Conclusions Overview of Future Directions

Summarizes the key findings and contributions of the research.

Chapter 9: Overview of Future Directions

Summarizes the future of the research.

Chapter 2: Description of the Topic Tackled

This chapter aims to define the specific problem addressed by this research, discuss its significance, and outline the challenges that must be overcome to achieve realistic and responsive virtual avatars. Key terminology relevant to this study will also be clarified to ensure a solid foundation.

2.1 Introduction to Human-Machine Interaction and the Role of Humanoid Avatars

HMI involves the study and design of systems where humans and machines communicate, ranging from simple interface manipulations to complex interactions with advanced AI systems. The development of effective humanoid robots, capable of natural and intuitive interactions, requires addressing significant challenges. These include the need for multi-modal communication that emulates human social behaviors and responses, as well as the technical complexities involved in animating virtual humans to achieve lifelike characteristics.

Humanoid avatars play a crucial role in enhancing HMI by serving as virtual representations that mimic human physical attributes and behaviors. These avatars create more personalized and engaging interfaces, allowing users to interact with virtual environments in a realistic and interactive manner. The integration of humanoid avatars into HMI systems not only advances the visual and functional aspects of these systems but also ensures that they meet user expectations for natural interaction. A key challenge in this integration is developing avatars that are not only visually convincing but also capable of handling multi-modal interactions, such as gestures and facial expressions, which are critical for authentic communication and engagement [26].

2.2 Problem Definition

2.2.1. Overview of the Research Problem

The primary challenge of this research is to create virtual avatars that can accurately replicate human facial features and behaviours in real-time interactions. This involves addressing several technical hurdles, such as the creation of detailed 3D models, the implementation of real-time facial recognition, and the synchronization of facial movements.

2.2.2 Challenges in Avatar Realism

When considering realistic avatars, various technical aspects have to be considered such as representation, motion control and etc. [13]. Some challenges that can briefly be acknowledged are:

3D Modelling

Creating realistic avatars involves developing detailed 3D models of human faces, capturing fine-grained features and textures. Traditional techniques often fall short, requiring advanced methods like photogrammetry and scanning for high-resolution accuracy [14]. Therefore, the complexity of human faces, with their unique features and variations, demands technical precision and a deep understanding of anatomy.

Real-Time Animation:

Achieving natural synchronization of facial movements with speech and gestures in real-time animation is a formidable challenge, demanding precise alignment of lip movements with spoken words and appropriate facial expressions across diverse languages and dialects. Low latency is critical for maintaining seamless interactions, necessitating optimization across data acquisition, processing, rendering, and display [15].

2.3 Significance of the Research

2.3.1. Applications of Humanoid Avatars

Normal everyday life has already been influenced by AI, but humanoid avatars amplify this impact by enabling remote operators to experience physical presence through the avatar. This breakthrough technology alleviates labor shortages handling mundane, hazardous, or strenuous tasks, as exemplified by Elon Musk's vision for the Tesla Bot, Optimus, slated to revolutionize industries. These avatars are poised to integrate seamlessly into sectors such as Education, Healthcare, and Entertainment, fundamentally reshaping societal dynamics and productivity.

As described by Elon Musk in October 2022 about the Tesla Bot, named Optimus:

> **"It will upend our idea of what the economy is... it will be able to do basically anything humans don't want to do... it's going to bring *an age of abundance*." -- Elon Musk**

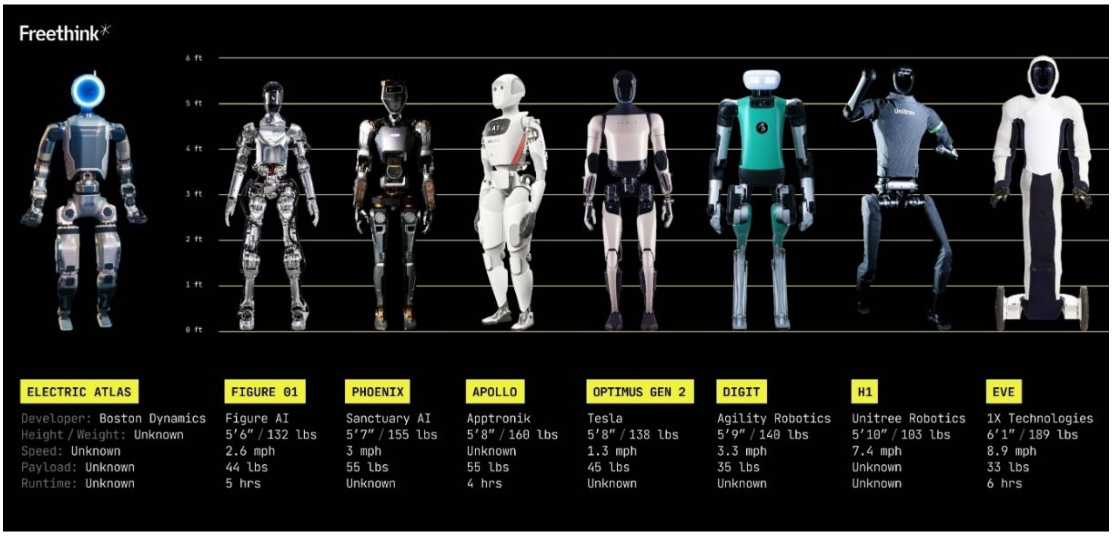

Figure 2 - Lineup of the current humanoid robots in development [27]

2.3.2 Enhancing User Experience

Realistic avatars enrich HMI by accurately simulating human behaviors and expressions, thereby creating immersive virtual interactions. These avatars effectively bridge the gap between physical and digital realms. Whether facilitating remote teamwork or enhancing social interactions in virtual environments, realistic avatars enhance communication by responding intuitively and authentically to users' interactions. Their ability to convey emotions and intentions through nuanced facial expressions and gestures makes interactions more natural and engaging.

2.3.3 Impact on Human-Machine Interaction

The development of realistic avatars represents a transformative leap in HMI, revolutionizing digital interface usability and accessibility. By integrating advanced technologies such as facial recognition and natural language processing, these avatars personalize interactions and adapt dynamically to users' preferences and needs. This capability not only enhances user satisfaction but also facilitates seamless communication across diverse applications, reshaping how humans interact with intelligent systems, promoting more empathetic, efficient, and productive communication channels.

2.4 Key Terminology

Terminology relevant to the study will be defined and explained. These terms are fundamental for understanding the complexities and advancements discussed.

Human-Machine Interaction (HMI):

The study and design of interfaces and systems where humans and machines interact.

Humanoid Avatars

Are digital representations of humans that replicate physical appearance and behaviour in virtual environments.

Unity:

Is a cross-platform game engine that is widely used for developing interactive 3D and 2D content. It provides tools for creating, importing, and manipulating 3D models, textures, and animations.

Furhat Robotics API:

Programming interface that allows control over a Furhat robots' behaviours, including speech and facial expressions.

Photogrammetry:

3D modelling technique that uses photographs to create highly accurate and detailed digital models of real-world objects such as a human face.

Texture Mapping:

Technique of applying images (textures) to 3D models to create realistic surface details.

JSON Configuration File:

A lightweight data-interchange format used to configure and control the behavior of software applications. In this case, it is used to customize the movements and expressions of virtual avatars.

Blendshapes:

Method of facial animation where different facial expressions are represented as distinct shapes or meshes that can be morphed between in real-time.

Facial Recognition:

Technology used to identify or verify individuals by analyzing and comparing unique facial features from images or video

2.5 Conclusion

This chapter has outlined the specific problem of creating virtual avatars, delving into the challenges associated with achieving this project's goal as well as a focus on developing humanoid avatars that accurately replicate human facial features in real-time. The discussion underscored the importance of detailed 3D modelling and the synchronization of facial movements with speech to overcome the technical challenges.

Furthermore, the chapter explored the broader impact of humanoid avatars on everyday life and industry, underscoring their potential to revolutionize fields such as manufacturing, remote operations, and virtual collaboration. Realistic avatars, therefore, represent a transformative leap in HMI by enhancing user experience and facilitating more natural interactions. In conclusion, this chapter sets the foundation for understanding the complexities and advancements involved in creating realistic virtual avatars. The subsequent chapters will build upon this groundwork, detailing the methodologies and technologies employed to address the outlined challenges and achieve the research objectives.

Chapter 3: Literature and Technology Review - 1922

3.1 Background

The extant literature indicates that the progress in human-machine interaction is greatly affected by the creation of realistic and compelling humanoid avatars. This narrative review explores the current research on avatar creation, leveraging software tools like Unity, and the significance of avatar realism [29]. Additionally, this section lays the groundwork for the methodologies and approaches utilized in this dissertation by highlighting key aspects of the development of humanoid avatars. The development of such avatars encompasses numerous intricate processes, including 3D modeling, facial recognition, and real-time animation [30].

Existing research also highlights the complexities of creating lifelike avatars for meaningful user engagement. The advent of advanced technology has revolutionized the field of human-machine interaction, and the development of realistic robot avatars has become a captivating area of research. One particularly intriguing concept is the idea of a robot avatar that physically resembles the user, allowing for a more immersive and personalized interaction experience that can enhance engagement and a sense of presence [30][31].

One significant advancement in this area is the development of real-time facial animation technology, which enables the seamless transfer of an actor's facial performance to a digital character with "extremely high fidelity". This breakthrough allows for more dynamic and engaging interactions, as the avatar can mimic the user's facial expressions and gestures in real-time [23].

Arcao et al. (2019) highlight the use of holographic projection and Kinect sensors to create realistic 3D avatars that accurately mimic human body motion, providing a more immersive and natural interaction. The significance of developing realistic and relatable virtual humans is critical [32]. When an avatar's appearance and behaviour closely resemble those of a real human, it elicits more natural and engaging reactions from observers, enhancing the overall interaction experience [33]. However, the "uncanny valley" phenomenon indicates that at a certain level of similarity, an observer’s reaction may shift from positive to discomfort, only becoming more favourable as the avatar further diverges from a real human [32][34][35].

Building on these advancements, the concept of a robot avatar resembling the user presents opportunities for remote participation in meetings, enabling individuals to attend without being physically present [19][36]. The creation of detailed, precise 3D models of human faces, capturing their nuances and subtleties, is essential for the development of realistic avatars [37]. This technology is applicable to virtual assistants, telepresence systems, and other contexts where personalized, engaging interaction is desired.

3.2 Integration of Multiple Technologies:

integrating multiple technologies is crucial for creating avatars that not only appear realistic but also exhibit lifelike movement and reactions, going beyond mere modelling [30][38][17]. Arcao et al. (2019) highlight the value of combining motion capture systems with computer vision techniques, leveraging the strengths of both to produce highly realistic and responsive avatars [17].While motion capture systems are vital for capturing human movement by using body-worn sensors to translate motion into digital data, they alone are insufficient for creating truly lifelike avatars. Motion capture often misses the nuanced facial expressions and subtle gestures crucial for achieving high realism in avatar animation [17][39].

Recent studies by Derpanis et al. (2019) and Roble et al. (2019) demonstrate significant advances in avatar realism through the integration of motion capture with computer vision techniques. By utilizing depth sensors, these researchers captured detailed 3D data, which neural networks then processed to replicate intricate patterns of human movement and expression. This integration allows avatars to mimic human expressions and gestures with remarkable fidelity. Depth sensors provide granular data on facial contours and movements, while neural networks generate real-time, lifelike animations, resulting in avatars that visually resemble humans and exhibit realistic behaviours, enhancing user immersion and engagement [23][40].

By capturing detailed facial features and body movements and using advanced machine learning algorithms to translate this data into realistic animations, researchers have developed highly customized and responsive virtual representations that closely resemble the original human form [23]. This integration of cutting-edge technologies has enabled avatars to achieve both visual and behavioural fidelity, paving the way for more natural and engaging human-machine interactions [41].Berthault et al. (2023) further illustrate the potential of this hybrid approach, combining depth sensors and convolutional neural networks to enhance avatar realism. Their research, focused on virtual reality settings, demonstrated how integrating these technologies resulted in avatars that responded to user interactions naturally and intuitively, greatly improving the overall user experience.

Combining different technologies offers a practical solution to the limitations of individual methods. Photogrammetry is excellent for creating high-quality static models but falls short in capturing dynamic facial expressions. On the other hand, motion capture excels at recording movement but struggles with intricate facial details [25]. By integrating these techniques, researchers can now create avatars that are not only visually impressive but also capable of lifelike, dynamic interactions. However, this integration isn't without its challenges—coordinating data from multiple sources can cause delays, and processing large volumes of data in real time demands significant computational power [42]. Despite these challenges, this hybrid approach holds great potential for advancing avatar realism and improving human-machine interactions..

3.3 Software Use in Avatar Development

The software tools used in developing avatars are indispensable, with Unity standing out for its versatility in animation, real-time rendering, and 3D modelling, enabling the creation of avatars that not only look realistic but also exhibit lifelike movement and reactions [21][17]. Unity's built-in features, such as its physics engine, allow for precise simulation of avatar movements and interactions, while its lighting and rendering tools enable the creation of detailed, photorealistic character models. Furthermore, Unity's support for various motion capture formats and integration with computer vision techniques streamline the incorporation of these technologies into avatar development [43][23].

Additionally, Unity's compatibility with machine learning libraries like TensorFlow and PyTorch allows researchers to harness neural networks for tasks such as facial expression recognition and emotion synthesis [44][23]. By combining Unity's comprehensive game engine features with advances in machine learning, developers can create avatars that not only achieve visual realism but also demonstrate natural and responsive behaviours, enhancing user experience. While Unity is a leading tool in avatar development, other software solutions, such as Unreal Engine and Blender, are also crucial. Unreal Engine's advanced rendering capabilities and real-time ray tracing make it a popular choice for high-fidelity avatars, especially in virtual and augmented reality applications [45][46]. Meanwhile, Blender, an open-source 3D modelling and animation tool, is widely used for its robust toolset for sculpting, rigging, and animating, supported by an active developer community, making it a key resource for avatar-related projects [45].

Unity supports the seamless creation and integration of avatars into various applications. As Brown et al. note, Unity's Mechanism animation system is particularly effective for rigging and animating humanoid characters. This system allows for complex animations and enables developers to create avatars capable of performing a wide range of human-like actions. For example, Berthault et al. (2023) used Unity to develop an educational simulation featuring avatars that interacted with students in a virtual classroom, demonstrating its effectiveness in creating engaging educational tools.

Beyond modelling, the integration of various technologies is essential for creating avatars that not only appear realistic but also move and react in lifelike ways. Unity's real-time rendering capabilities and support for high-fidelity graphics are vital in achieving this. Wilson et al. highlight these tools' effectiveness in creating visually stunning avatars that operate in real-time environments. Their research, involving avatars for a healthcare application, demonstrated how the avatars provided real-time feedback and interaction with patients, showcasing Unity's potential in sensitive and interactive scenarios.

In summary, the strategic use of software tools like Unity, Unreal Engine, and Blender has been instrumental in advancing avatar realism and interactivity. These tools enable the integration of technologies such as motion capture, computer vision, and machine learning, allowing for the creation of avatars that are both visually impressive and behaviourally realistic, thereby enhancing the user experience in virtual and augmented reality applications [29][47][23][6].

3.4 Accessibility and Democratization

Once the domain of large corporations and specialized teams, the development of realistic avatars has become more accessible due to advancements in AI and machine learning. The emergence of large-scale pre-trained models and advanced human representation systems, such as the CLIP framework, has enabled smaller studios and individuals to create high-quality avatars with reduced time and cost [22][42]. Additionally, the increasing availability of open-source software and cloud-based platforms has lowered barriers to entry, allowing more researchers and developers to experiment with avatar creation.

This democratization of avatar technology has wide-ranging implications, from social interaction and communication to education and healthcare. Jourabloo et al. (2021) emphasize the potential of accessible avatar technology to enhance remote collaboration and telework, thereby reducing travel and its environmental impact. Furthermore, the ability to create personalized avatars can improve users' sense of presence and engagement in virtual environments. Waltemate et al. (2018) found that customized avatars led to heightened emotional responses and behaviour changes in users.

In addition to animation, Unity’s real-time rendering capabilities and support for high-fidelity graphics are critical. According to Borkman et al. (2021), these tools help create visually impressive avatars that function in real-time environments. Their research, focused on healthcare applications, demonstrated how avatars could provide real-time feedback and interaction with patients, showcasing Unity’s potential in sensitive and interactive scenarios.

Beyond mere modelling, integrating various technologies is crucial for creating avatars that not only look realistic but also move and react in lifelike ways.

3.5 Avatar Realism

The realism of avatars is not merely about visual accuracy; it's about creating a connection that feels authentic and engaging. This authenticity is critical for user satisfaction and the overall success of HMI applications. Studies consistently highlight that avatar realism significantly impacts user engagement. According to Latoschik, E, M. et al. (2017), realistic avatars enhance the immersive experience, making interactions feel more genuine and emotionally engaging. Their research showed that users interacting with highly realistic avatars reported higher levels of satisfaction and engagement compared to those interacting with less realistic avatars.

This finding underscores the importance of investing in high-quality avatar-creation processes. To achieve this level of realism, various techniques are employed. For example, using blend shapes and morph targets allows for the creation of natural facial expressions, as highlighted by Hong et al. (2022). These techniques ensure that avatars can exhibit a wide range of emotions, enhancing the user's sense of connection. Additionally, motion capture technology, combined with machine learning algorithms, can be used to replicate human gestures accurately. This was demonstrated in a study by Fortini et al. (2023)., where avatars created using these techniques were able to perform complex gestures that closely mimicked human movements.

3.6 Conclusions

This literature review has examined the development of realistic and compelling humanoid avatars, highlighting their crucial role in enhancing human-machine interaction

(HMI). Research consistently emphasizes that realistic avatars significantly improve user satisfaction and immersion, laying the groundwork for the methodologies discussed in this dissertation. Technological advancements have made it possible to create highly realistic avatars. Techniques like blend shapes, morph targets, and motion capture, combined with machine learning algorithms, allow avatars to exhibit natural facial expressions and accurate human gestures. These technologies result in avatars that not only look realistic but also behave lifelike, thereby enhancing user engagement.

Software tools like Unity are essential in this process, offering robust capabilities in animation, real-time rendering, and 3D modelling. Studies have demonstrated Unity's effectiveness across various applications, such as educational simulations and healthcare interactions. The integration has further advanced avatar realism, allowing for precise replication of human expressions and movements. However, challenges remain in achieving real-time interaction and ensuring user acceptance.

Chapter 4: Possible Approaches and Reasons for Approaches Taken

This chapter examines various methodologies and techniques for developing realistic and responsive virtual avatars, focusing on addressing the challenges outlined in previous chapters. These challenges include creating detailed 3D models, achieving real-time synchronization of facial movements, and lifelike interactions. The chapter assesses different approaches based on their effectiveness in tackling these issues and provides a rationale for selecting particular approaches. It also discusses potential implementation challenges and proposes strategies for overcoming them, ensuring the project stays aligned with its objectives.

4.1 Possible Methodologies

To achieve the goal of developing realistic and responsive humanoid avatars, several methodologies were considered. These approaches range from traditional 3D modelling techniques to more advanced methods that integrate machine. The following sections provide a detailed exploration of these methodologies.

4.1.1 Traditional 3D Modelling and Animation

Traditional 3D modelling techniques offer significant control and precision over the visual details of a model, making them a tried-and-true approach. However, when it comes to developing highly realistic avatars, this approach may not be the most efficient. The process required for traditional modelling is time-consuming and demands substantial expertise, posing a challenge for projects that require rapid iteration and real-time performance. One of the primary limitations of traditional 3D modelling is the extensive manual effort needed to capture the subtle nuances of human facial expressions. This level of detail is crucial for achieving realism but places a considerable burden on the modelling process.

So while traditional 3D modelling remains a viable option for certain aspects of avatar development, recent advancements in machine learning offer promising alternatives. These advancements enhance both the efficiency and realism of avatar creation, enabling real-time animation with significantly less manual input and expertise especially as there's constant progress and breakthroughs being made every decade **(See Figure 3)** .

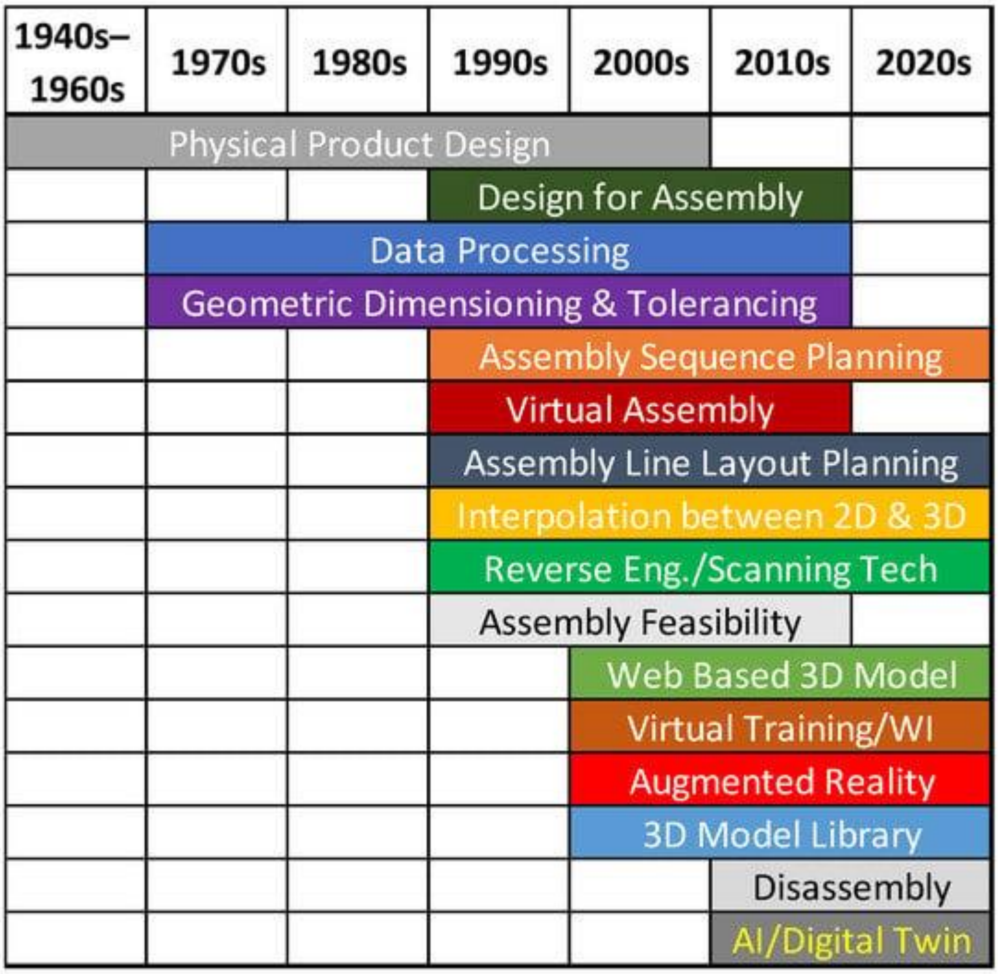

Figure 3 - Three-dimensional (3D) research themes evolution over time [57]

4.1.2 Photogrammetry and 3D Scanning

Photogrammetry and 3D scanning are advanced techniques used to capture highly detailed 3D models from multiple images or scans of real objects, often focusing on human faces. These methods are crucial for achieving high realism in virtual avatars by producing accurate textures and geometries. By stitching together 2D images from multiple angles, photogrammetry creates a 3D object that captures the intricate details of the human face, providing a strong foundation for lifelike virtual representations **(see Figure 4)**.

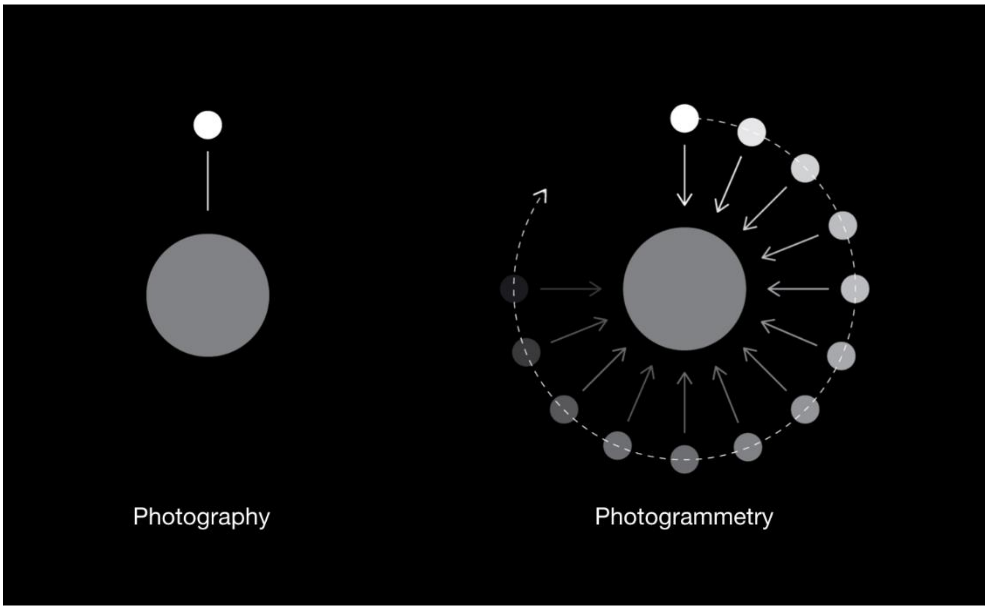

Figure 4 – Demonstrating how Photogrammetry is the practice of stitching 2D images from multiple angles into a 3D object

Given the need for precise facial features, photogrammetry and 3D scanning offer significant advantages. They reduce the manual effort required in traditional 3D modelling, while delivering a higher degree of accuracy—essential for making avatars appear lifelike and authentic in their interactions. However, these techniques come with challenges, such as the high cost of scanning equipment, complexities in integrating the scanned data into real-time engines like Unity. To mitigate these obstacles, machine learning-based tools can be integrated to automate the processing of scanned data, reducing hardware costs while maintaining high fidelity in the resulting avatar models.

Therefore a hybrid approach by integrating 3D scanning and photogrammetry to create the initial avatar model, followed by the use of blendshapes for testing facial expressions and movements could be potential solution. The 3D scanning process captures the avatar’s facial features with precision, providing the structural realism needed for accurate representation. Photogrammetry enhances this with detailed textures, further improving visual authenticity. Once the detailed model is created, blendshapes are applied for real-time animation in Unity, enabling the avatar to perform facial movements and engage in natural interactions without excessive computational demands to it feasibility to be applied to the Furhat.

This combined methodology ensures both realism and responsiveness. Moreover, this approach is scalable, making it accessible to smaller studios that may not have the resources for traditional high-end methods, but still need to achieve a high level of realism and interactive performance.

4.1.3 Real-Time Animation with Blendshapes

Blendshape animation involves creating various facial expressions and morphing between them in real-time, while rigging refers to building a skeleton for a 3D model to control its movements. These techniques are often combined to enable real-time facial animation, which is crucial for interactive avatars. Blendshapes offer a flexible approach to dynamically animating facial features, making them ideal for responsive avatars that engage naturally with users. Unity’s support for these techniques enhances their practicality and efficiency, allowing developers to test facial models and textures to see if they retain their lifelike quality during blendshape animations.

However, achieving realistic expressions requires a well-constructed set of blendshapes, which can be complex and time-consuming. Smooth transitions between expressions are vital to avoid the "uncanny valley" effect, where avatars appear almost human but cause discomfort due to slight imperfections. To address these challenges, a hybrid approach that combines blendshape models with physics-based simulation techniques can be employed, leveraging the strengths of both methods to create detailed facial animations while maintaining a high degree of realism and interaction.

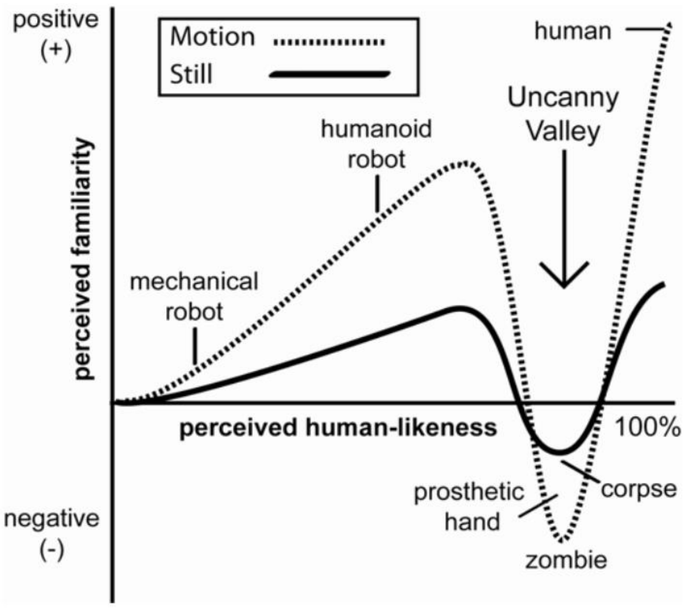

Figure 5: The Uncanny Valley: Plot of Perceived Familiarity vs. Human-Likeness, Adapted from Mori's Model (MacDorman & Minato Translation) [58]

4.1.4 Machine Learning and AI-Driven Animation

Machine learning techniques have revolutionized the field of animation by enabling the analysis and replication of human facial expressions, movements, and speech patterns.

AI-driven animation systems can generate real-time responses based on user input designs, resulting in more dynamic and personalized interactions. As by training models on vast datasets of human facial expressions and speech, avatars can learn to replicate these features with high accuracy and adapt to different user interactions. This approach aligns with the research objectives of achieving lifelike avatars capable of natural, multimodal communication. However, this approach also presents challenges. Integrating machine learning models with Unity's engine for real-time rendering can be complex. To mitigate these challenges, starting with simpler, pre-trained models and gradually introducing more complex AI systems allows for testing and optimization at each stage, ensuring a smoother integration process.

4.2 ARKit and Unity's Animation and Rendering Capabilities

Unity would be a primary development platform for this project due to its extensive tools for 3D modelling, animation, and real-time rendering. Its versatility, particularly in handling a variety of animation techniques such as blendshapes and rigging, made it ideal for this project's focus on creating a lifelike avatar. Unity's powerful rendering engine ensured that the avatar's facial expressions, texture mapping, and animations could be displayed accurately in real-time, delivering a visually realistic experience.

A significant enhancement to this system would be the incorporation of ARKit. ARKit provided high-quality facial tracking data that directly influenced the avatar's ability to reflect natural expressions. By leveraging ARKit, facial recognition data would map onto the 3D model in Unity, helping to test and refine blendshapes with greater accuracy. This integration enabled real-time tracking of facial movements, ensuring that the avatar responded in sync with the user's expressions and further improving the realism of interactions.

4.3 Integration of AI and Machine Learning

Integrating AI-driven animation models with Unity's real-time engine is a complex task, particularly the real-time processing demands of interactive avatars. To mitigate these, beginning with a simpler, pre-trained models, which are then integrated and tested within the Unity environment would be a potential solution. This method ensures that potential issues are identified and addressed early in the development process, resulting in a smoother and more efficient integration. Additionally, cloud-based processing can be considered for offloading some of the AI computations. This reduces the burden on local hardware and helps maintain performance efficiency, particularly in resource-intensive scenarios.

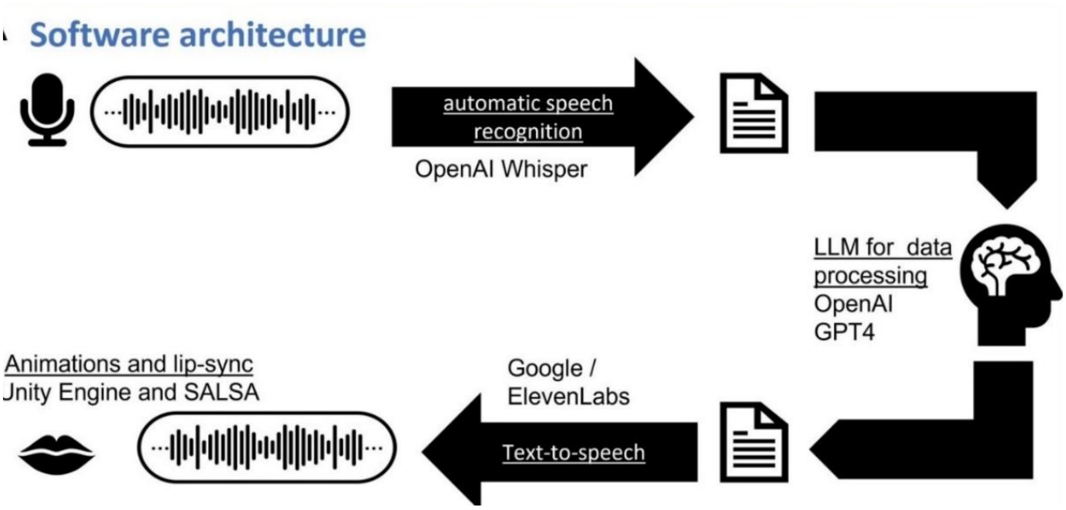

Figure 6: Software Architecture for Speech Recognition [59]

4.4 Virtual Avatar texture Application

For this project, a key aspect of customizing the Furhat robot avatar to resemble myself would involve applying a texture created from the photography onto the 3D model. This texture would include all key facial features such as skin tone, freckles, and fine details to closely resemble my appearance.

Once the model is created, the texture could be extracted to then be applied to the virtual avatar using Furhat's SDK, which relies on JSON configuration files to precisely map the texture to the avatar's facial structure. These configurations allowed for detailed calibration of the mouth, eyes, eyebrows, and other facial features to ensure they moved naturally during interactions, such as speaking or expressing emotions. The positioning and alignment of the texture are critical in achieving a realistic representation, requiring iterative testing and adjustment of the texture extracted do that it'd align perfectly with the Furhat model. This process would ensure that the avatar's expressions mirrored my own, enhancing the overall realism of the virtual robot.

4.5 Conclusion

This chapter has provided a comprehensive overview of the various methodologies and tools considered for developing realistic and responsive virtual avatars. By combining 3D scanning with real-time animation techniques in Unity, this project aims to show a high level of realism and responsiveness in the avatar, addressing the challenges identified in earlier chapters. The selected methodologies have been critically evaluated, with the rationale for their selection clearly articulated. These approaches not only align with the project's objectives but also lay the groundwork for future developments in the field of virtual avatar creation.

Chapter 5: Description of the Work Carried Out

The primary goal was to develop a highly realistic virtual avatar that mirrors myself and my facial expressions while synchronizes them with speech. This avatar was to be integrated through a pipeline utilizing the Unity platform enabling it to replace the skin texture of an existing avatar like the Furhat robot. The work involved several steps, including model creation, facial recognition integration, synchronization testing through the use of

Blendshapes through the ARKit, and calibrating the position of the facial feature in the characters .Json configuration so that they move "naturally" during the interaction.

5.1 Initial Setup and Tool Selection

In previous chapters, concepts were evaluated to meet the project's objectives of creation and realism. Unity was chosen as the primary platform, supported by Blender for advanced 3D modelling. Blender's detailed mesh creation and seamless compatibility with Unity's asset pipeline ensured smooth integration between modelling and animation. Unity managed real-time interaction and rendering, while Blender provided high-fidelity models, especially for facial features. For facial recognition integration a physical Furhat robot was unavailable, so instead, the Virtual Furhat via the SDK Launcher's web interface would facilitate avatar interaction. This integrated approach using Unity, Blender, and the Furhat API would enable the creation of the Model and its lifelike nature.

5.1.1 Research and Development Approach

To establish a strong foundation for the project, the initial step involved exploring simpler, pre-trained models, which would then be integrated and tested within the Unity environment. Some simpler pre-built 3D models included those from the Windows 3D Library. One such model, the "Give this person a face" model **(see Figure 7)**, provided a starting point for experimentation with facial structures and geometry. However, the model lacked the realism needed for the project due to inconsistencies in texture application and the inability to finely control facial features, leading to an artificial appearance.

Figure 7: Windows 3D Library “Give this person a face” model

Some Pre-trained models would include Avatar Maker and Ready Player Me. Avatar Maker, for example, is a Unity plugin capable of generating 3D avatars from a single selfie. Although, despite its ease of use, Avatar Maker presented significant limitations in terms of realism. The avatars created using a single image lacked crucial details, particularly when facial accessories like glasses or facial hair were involved **(see Figure 8)**. While the tool was effective for rapid prototyping, it could not deliver the level of detail and texture quality necessary for this project which highlighted the benefits of photography.

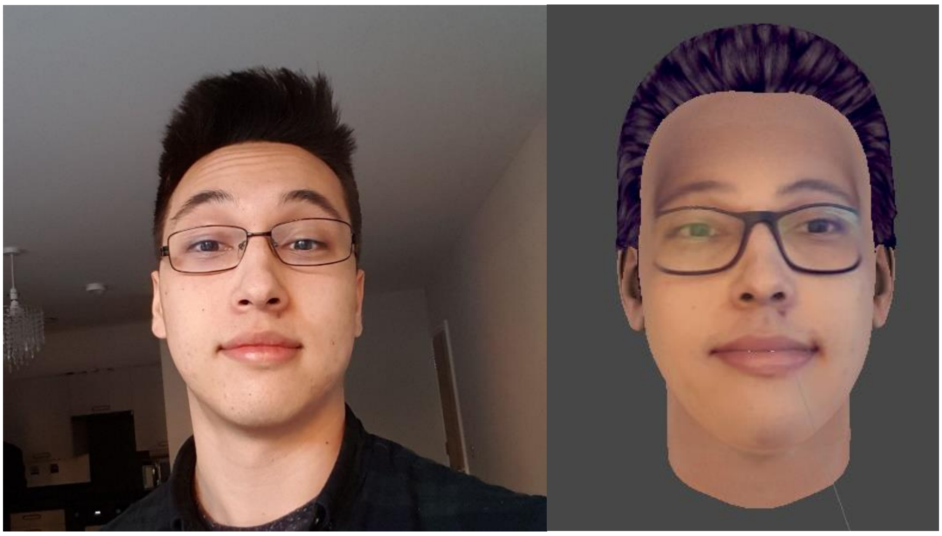

Figure 8: Avatar Maker avatar created from Reference Picture

5.2 Workflow and Implementation Process

5.2.1 3D Model Creation and Customization

With the limitations of pre-trained models and avatar creation tools established, the next step in the process was the manual creation and customization using Blender. This process would make use of the Photogrammetry and 3D Scanning techniques mention in the previous chapter to produce accurate textures and geometries. The construction process went as follows:

- **Initial Mesh Construction**:

The creation process began by importing six reference images **(See Figure 9)** of my face into Blender. These images were taken from multiple angles, providing full head coverage rather than relying on a single image. By capturing the front, sides, back, and top of the user's head, these images ensured that the 3D model would be as accurate as possible, covering every detail of my facial features and skull structure.

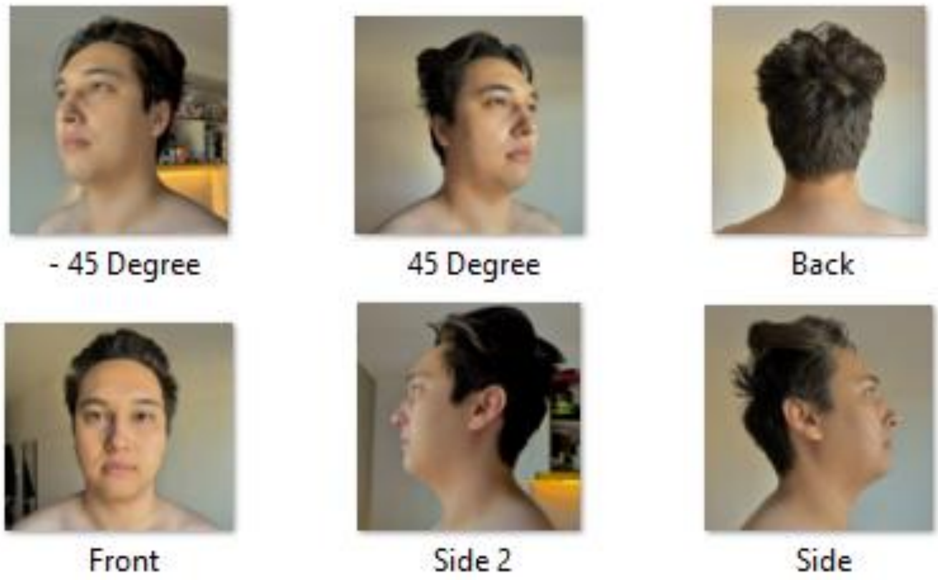

Figure 9 – Reference Images for Full Head 3D Model Creation: Multi-Angle Views

Using Blender’s polygonal modelling tools, the face shape would be adjusted based on the six reference images as adjusting one image would in-turn affect others **(See Figure 10)**. This step was vital as the use of multiple images provided a significant advantage over using just one image, as it allowed for greater precision in the modelling process for particular features and ensured that the avatar’s head had a consistent geometry from all angles. This process resulted in a more lifelike and detailed 3D mesh.

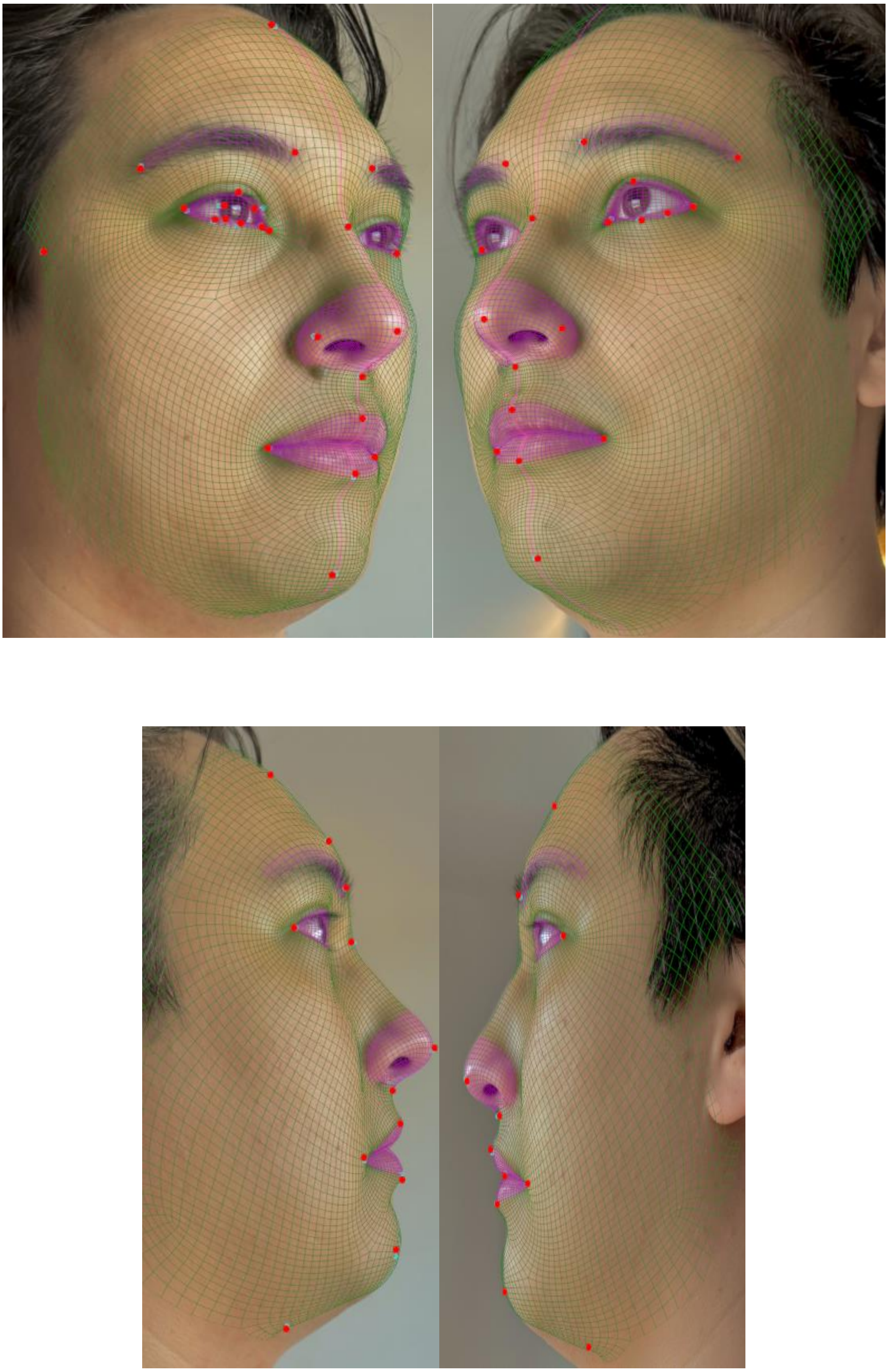

Figure 10 - 3D Mesh Overlay and Facial Landmark Detection for 4/6 Reference Pictures

- **Mesh Refinement and Texture Mapping**:

Once the base mesh was created, further refinements would be made using Blender's sculpting tools. These tools allowed for adjustment on finer details, such as the lips, eyes, nose and mouth so that any issues that arise from the photogrammetry could be resolved **(See Figure 11 & 12)**. Those critical areas which were highlighted in purple received particular attention, with increased polygon counts (**number of polygons make up a 3D model, with higher polygon counts resulting in more detailed and smoother models**). This was crucial for achieving realism as it allow the possibility of generating realistic skin textures, capturing fine details such as pores, freckles, and subtle variations in skin tone. **(See Figure 12)** which shows the inclusion of my mole below my nose as an example of this.

Figure 11: Eye Texture Rendering Issue in 3D Model: Misalignment

Figure 12: Eye Texture Rendering in 3D Model: Alignment

- **Exporting to Unity**:

The model would be exported from Blender in FBX format, which is compatible with Unity alongside a texture .PNG file. The export process ensured that all the geometry, textures, and animation-ready elements such as Blendshapes were preserved. This seamless export-import pipeline between Blender and Unity allowed for efficient workflow, reducing the chances of data loss or misalignment during the transition.

5.2.2 Integration and testing into Unity

The next step involved utilizing Unity's animation and rendering capabilities to test for interaction and expression mimicry. To achieve real-time facial expression changes, Blendshapes were utilised. These Blendshapes defined various facial expressions, such as smiling, frowning, and raising eyebrows, which the avatar needed during interactions. Additionally, ARKit's facial tracking capabilities could further test this through iOS devices and allowed detailed testing of facial expression tracking, ensuring accuracy and fluidity in the avatar's responses.

5.2.3 Real-Time Expression Mimicry with Furhat Robotics API

The next phase involved integrating the model into the Furhat Robotics API through the SDK, which allowed the virtual Furhat robot to respond while closely resembling myself.

The work required would involve:

- **Connection to the Furhat SDK**

The API allowed for real-time control of the avatar's expressions based on input from the SDK Web Interface.

- **Facial Integration and calibrating**

Through the Furhat API, the system accessed real-time facial data from files stored within the SDK Launcher (see Figure 13). These files included key elements such as Eyes and Skin texture. The access to these files as well as the relevant JSON configurations, allowed me to insert my custom model with its specific facial features into the system as well as any relevant JSON code to ensure the configuration pulled from the right files. Once the facial texture was mapped onto the avatar, further calibration would be required to ensure the texture was correctly aligned. To calibrate the texture file, I'd use Canvas **(online graphic design tool that allows users to create and edit images)**.

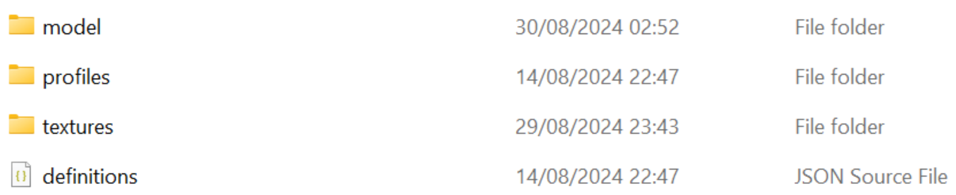

Figure 13: Directory Structure for Facial Recognition Data Files in Furhat SDK

5.2.4 Speech Recognition, Lip Sync, and Control Logic Using VSCode and Blockly

After application of the facial texture so that the Furhat resembles me and its ability to operate affectively, a key part of advancing an engaging virtual avatar would be to achieve synchronize speech with facial expressions in real time. This would be accomplished by integrating a speech recognition and lip sync system, enhanced by control logic implemented using VSCode and Blockly.

5.2.4.1 Voice Control & Selection

VSCode was used to develop scripts for managing the voice selection and interaction behaviours of the avatar. This allowed for real-time control and customization of the avatar's speech-related behaviours.

- **Voice Selection and Customization**

This was to allow myself to customize the avatar’s voice from a list of available options, ensuring that the avatar could offer personalized and regionally appropriate voices. The code filtered and selected voices from Furhat’s 119 available options **(See Figure 14)**, allowing myself to select a preferred voice, which was then stored for future use.

```
Available voices:
1: {'language': 'fr-BE', 'name': 'Isabelle-Neural'}
2: {'language': 'en-US', 'name': 'Danielle-Neural'}
3: {'language': 'en-US', 'name': 'Gregory-Neural'}
4: {'language': 'tr-TR', 'name': 'Burcu-Neural'}
5: {'language': 'en-US', 'name': 'Kevin-Neural'}
6: {'language': 'tr-TR', 'name': 'Filiz'}
7: {'language': 'sv-SE', 'name': 'Elin-Neural'}
8: {'language': 'sv-SE', 'name': 'Astrid'}
9: {'language': 'ru-RU', 'name': 'Tatyana'}
10: {'language': 'ru-RU', 'name': 'Maxim'}
11: {'language': 'ro-RO', 'name': 'Carmen'}
12: {'language': 'pt-PT', 'name': 'Ines-Neural'}
13: {'language': 'pt-PT', 'name': 'Ines'}
14: {'language': 'pt-PT', 'name': 'Cristiano'}
15: {'language': 'pt-BR', 'name': 'Vitoria-Neural'}
16: {'language': 'pt-BR', 'name': 'Vitoria'}
17: {'language': 'pt-BR', 'name': 'Ricardo'}
18: {'language': 'pt-BR', 'name': 'Camila-Neural'}
19: {'language': 'pt-BR', 'name': 'Camila'}
20: {'language': 'pl-PL', 'name': 'Maja'}
21: {'language': 'pl-PL', 'name': 'Jan'}
22: {'language': 'pl-PL', 'name': 'Jacek'}
23: {'language': 'pl-PL', 'name': 'Ewa'}
24: {'language': 'pl-PL', 'name': 'Ola-Neural'}
...
116: {'language': 'fr-FR', 'name': 'Remi-Neural'}
117: {'language': 'it-IT', 'name': 'Adriano-Neural'}
118: {'language': 'pt-BR', 'name': 'Thiago-Neural'}
119: {'language': 'ar-AE', 'name': 'Zayd-Neural'}
```

Figure 14: List of available Furhat Voices within the SDK

5.2.4.2 Control Logic Using Blockly

Utilizing the Furhat Remote API is just one of 3 ways to program the Robot, another option is through the use of Blockly, a visual programming tool that allows users to create code by dragging and dropping blocks, making it easier to build complex logic and behaviors without needing to write traditional text-based code.

- **Defining Interaction States and Triggers**:

With Blockly's drag-and-drop interface, different states were created to represent the various stages of an interaction (e.g., greeting, responding, listening) **(See Figure 15)**. These states were linked to triggers that dictated how the avatar should transition between behaviours based on user input.

- **Gestures and Facial Expressions**:

This'd allow the avatar could perform gestures (e.g., nodding, smiling) in response to specific user inputs or during pauses in the conversation. These gestures, combined with facial expressions, enhanced the naturalness of the interaction.

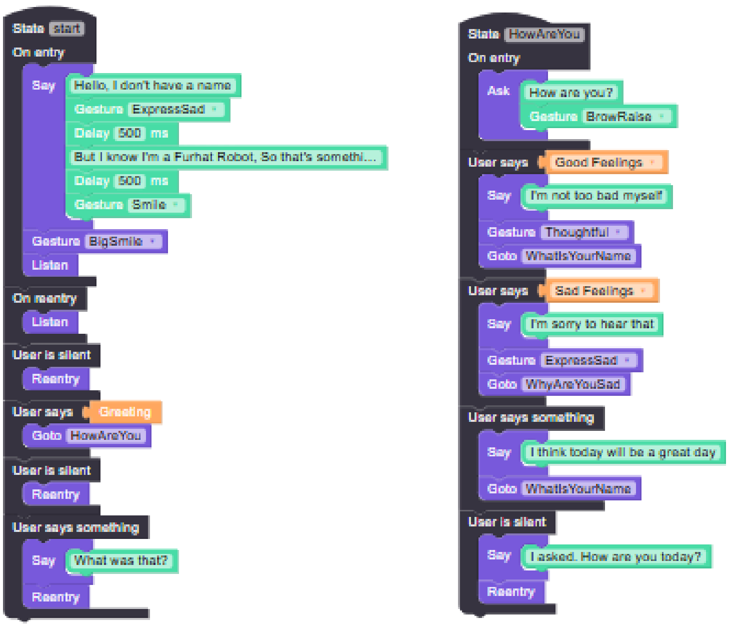

Figure 15: Interaction Flow Diagram for Avatar's Conversational States Using Blockly

5.2.4.3 Speech Recognition and Machine Learning

To further enhance the avatar's interaction capabilities, speech recognition was combined with machine learning using ChatGPT for real-time conversational responses. This integration allowed the avatar to engage in dynamic, contextually relevant dialogues, making interactions feel more natural and personalized.

- **ChatGPT for Conversational Response**:

ChatGPT enabled the avatar to generate responses based on user input without relying on predefined scripts. This allowed for contextual awareness, where the avatar could maintain the flow of conversation and adjust its responses according to the user's emotional tone. The system could personalize interactions over time, tailoring its responses to the user's preferences.

- **Real-Time Speech Recognition and Phoneme Detection**:

In addition to the typing version of interaction with the Furhat through GPT, the speech recognition system offered an alternative approach by analyzing user speech in real time, specifically keywords which would trigger the system to correspond gestures and facial expressions, enabling the avatar to visually express emotions and respond more naturally to user interactions. This mapping of keywords to gestures is illustrated in **Figure 16**.

```python
# Define mappings of emotions/gestures to certain keywords
emotion_gesture_map = {
    "happy": "BigSmile",
    "good": "Smile",
    "sad": "ExpressSad",
    "angry": "ExpressAnger",
    "fear": "ExpressFear",
    "surprised": "Surprise",
    "confused": "Thoughtful",
    "hello": "Nod",
    "bye": "Wink",
    "thanks": "Nod",
    "sorry": "BrowFrown"
}
```

Figure 16: Emotion-Gesture Mapping for Avatar Responses Based on Keyword Detection

5.3 Conclusion

The development of a highly realistic virtual avatar involved multiple stages of experimentation and refinement, beginning with the selection of tools like Unity, Blender, and the Furhat Robotics SDK. Initial attempts with pre-built models highlighted the need for full customization, leading to the use of Blender for precise 3D modeling and texture mapping. Photogrammetry enabled detailed control over facial features, while Blendshapes provided realistic expressions for real-time interaction. Integrating the avatar with Unity’s animation system for ARKit testing and synchronizing it with Furhat Robotics API allowed for advanced features such as real-time speech recognition and dynamic expression changes. The use of Blockly further streamlined interaction design, making the system accessible for users with different technical backgrounds, setting the stage for testing and evaluation against project goals.

Chapter 6: Results, Testing, Verification

This chapter discusses the results of testing and evaluation conducted on the virtual avatar, focusing on its accuracy, performance, and interaction quality. The primary areas of testing included the avatar's 3D model, texture mapping, real-time expressions, speech synchronization, and the overall system in the attempt to create an avatar that resembles myself.

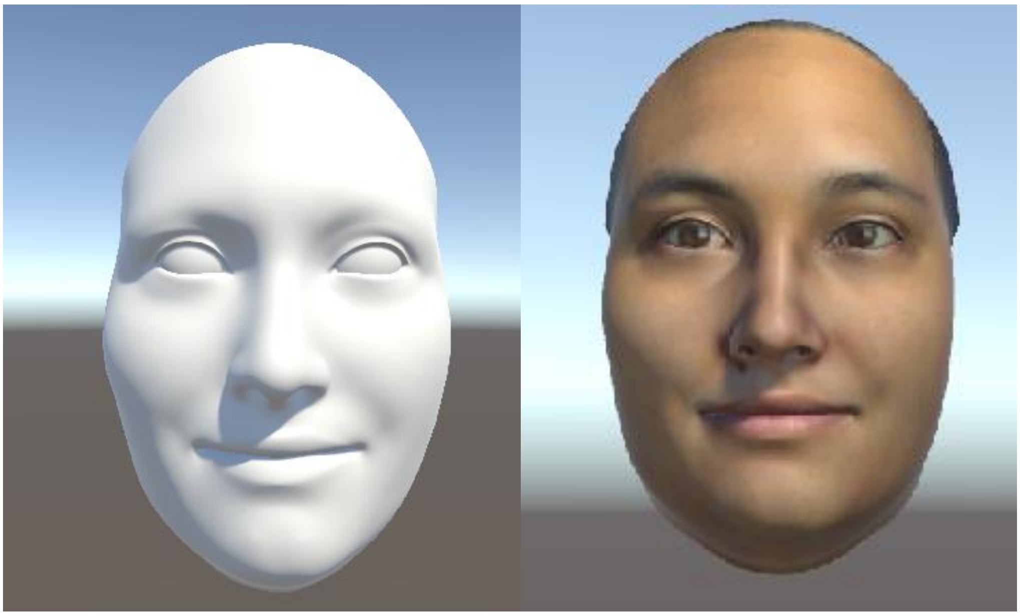

Figure 17: My 3D-Model with and without Textures applied

6.1 The Final Model and Texture

The final 3D face model produced **(See Figure 17)** was the product of the development process outlined in Chapter 5. Throughout this phase, the focus was on ensuring that all critical facial features would be best represented. Once the 3D geometry was finalized, the next step required obtaining a high-resolution texture. This texture includes detailed features such as skin tone variations and facial marks, providing a lifelike representation of my face. The final texture attempted to excluded unnecessary areas, such as hair and neck, to concentrate solely on the essential parts of the face that would be used the

Furhat robot **(See Figure 18)**, but there would be some hair included which due to how the Furhat would be presented with the wrapping of the texture, wouldn't be visible.

Figure 18: The Texture .PNG

With both the model and texture in place, blendshapes were applied with the use of the ARKit to test a range of facial expressions. These pre-configured shapes allowed the model to simulate key expressions, in a controlled environment. The purpose of these tests was to evaluate how well the model and texture responded to dynamic movements and how the avatar would react when different facial muscles were engaged **(See Figure 19 & 20)**.

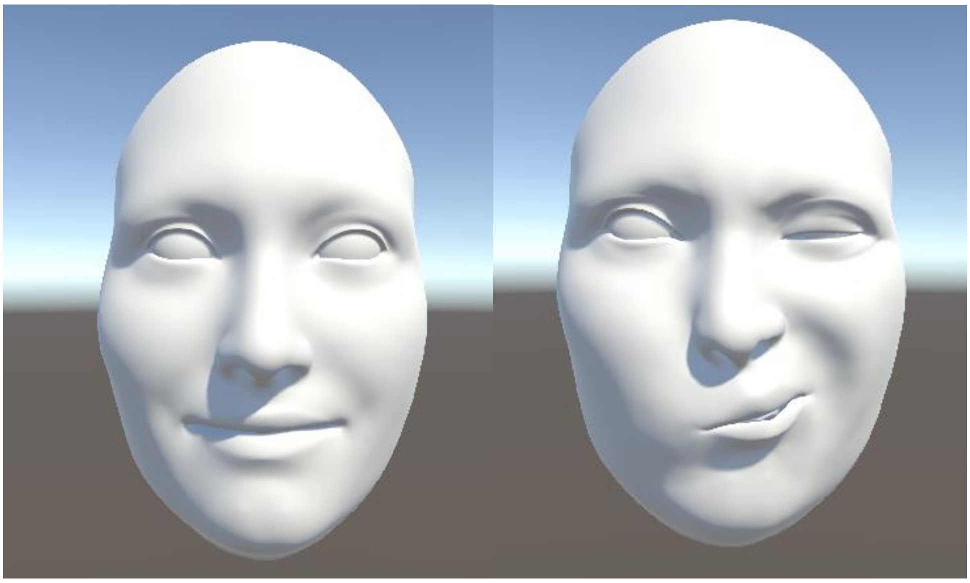

Figure 19: Assortment of Facial Expressions for my Model With no Texture

Figure 20: Assortment of Facial Expressions for my Model With Texture

The resulting model successfully captured my facial features and responded fluidly to various expressions. The Blendshapes revealed that the model could transition smoothly between different emotional states, from subtle movements like eyebrow raises to more exaggerated expressions. This final model and texture combination represented a high-fidelity version of the avatar, ready for integration into the Furhat SDK for real-time testing and interaction.

6.2 Application of the 3D Model and Textures

The 3D model and texture mapping of the avatar were critical elements of the project, aiming to replicate my facial features. To assess out these areas, focus was placed on two separate areas, Visual Accuracy and Realism, and Texture Mapping and Alignment.

6.2.1 Texture Mapping and Alignment

To ensure the Furhat avatar resembled me accurately, the project required not only careful texture creation but also the configuration of JSON code to define how the textures were applied to the Furhat model. Once the texture was created in Blender, it was exported as a PNG file and had to be referenced in the Furhat SDK using JSON configuration files. Initial placement of the texture without adjustment resulted in **Figure 21.**

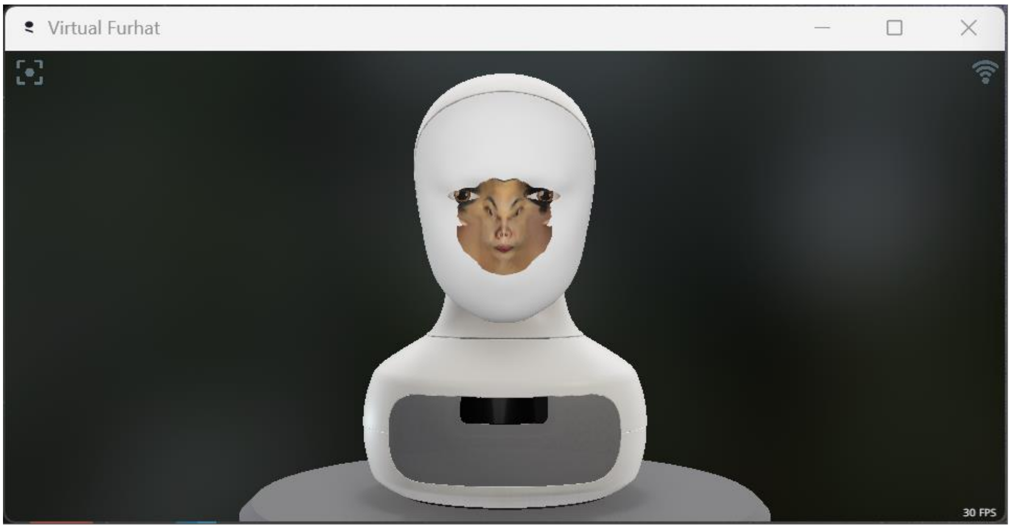

Figure 21: Furhat Model with Original .PNG File

6.2.1.1 JSON Code Structure

The texture application required generating a custom JSON file that directed the SDK on how to apply the PNG texture to the virtual Furhat model. JSON code snippet looked like this:

```
{"Parameters":[],
"TextureController":{
    "FaceSkinColor":"#FFFFFFFF",
    "TextureSets": {
        "eyes":"ted",
        "skins":"robert"
        },
        "Overlays":[]
        },
        "MaskID":"adult [legacy]",
        "RigID":"adult [legacy]"
}
```

Figure 22: JSON Code Structure

In this code, the "TextureSets" section specifies the PNG files for different facial components. For example, "robert" refers to the primary texture file for the face, and "ted" applies to the eye textures. These values were updated with each PNG, particularly after resizing the texture to 1024x1024 pixels, which matched Furhat’s requirements **(See Figure 23)**. The "FaceSkinColor" parameter was set to a neutral value (white) to avoid any conflicts with the texture colours, allowing the PNG file to determine the appearance fully.

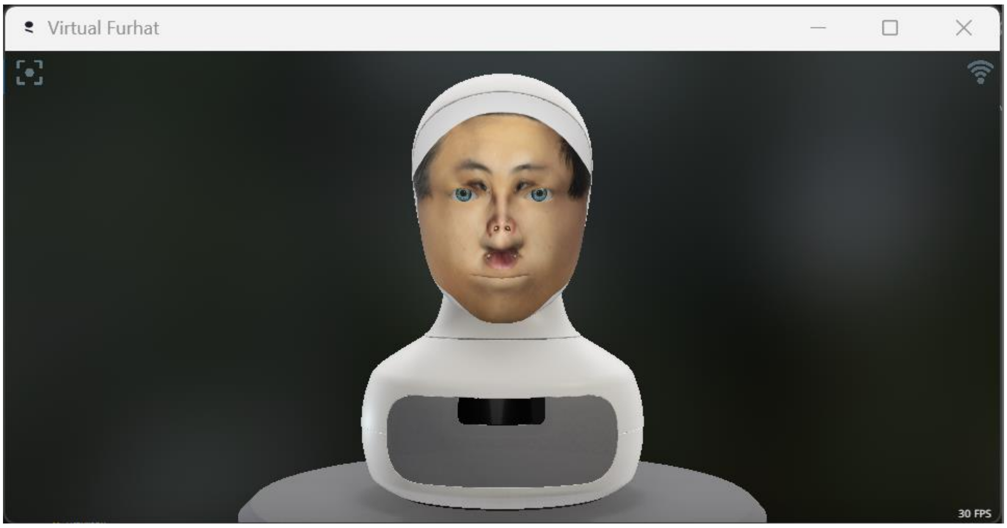

Figure 23: Furhat Model with Width to Width .PNG File

6.2.1.2 Adjusting the Texture Mapping in JSON

Testing the alignment of the texture required iterative adjustments to the JSON file and the PNG itself. Each update would be loaded into the Furhat SDK, and real-time expressions would be tested to identify any misalignment or texture stretching (see **Figure 24** for different texture variations tested on Furhat). As initial testing revealed that while the lower regions of the face, such as the jawline, aligned well, the upper portions, including the eyes and forehead, required further adjustment.

Figure 24: Different Textured files to be Tested

The most challenging part of this alignment was the mouth. The original texture showed the mouth in a partially open state, which was incompatible with the Furhat robot's closed-mouth default expression. Additionally, the eye textures were refined to ensure accurate placement within the Furhat model's eye socket PNG files with the texture Sub-Folder, as any misalignment would cause noticeable distortions during blinking and eye movement. This is seen in **(Figure 25)** where you can see my textured eyes above the eye lids as well as my noses nostrils.

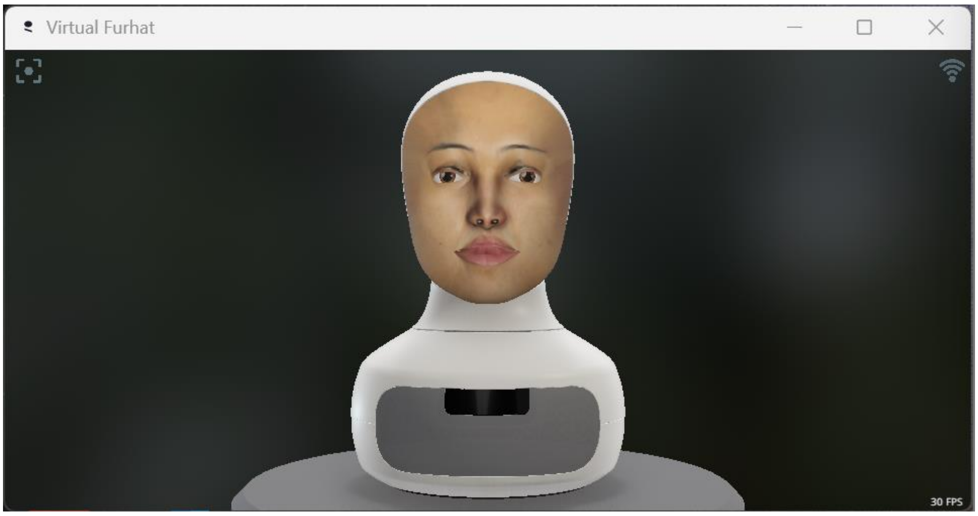

Figure 25: Furhat Model with Near Perfect .PNG File

These real-time tests helped identify and correct issues with stretching or misalignment, particularly during dynamic movements like speaking or blinking. The final result **(See Figure 26)** was a well-aligned texture that worked effectively with the Furhat robot, producing realistic facial expressions while maintaining visual consistency across different movements and interactions.

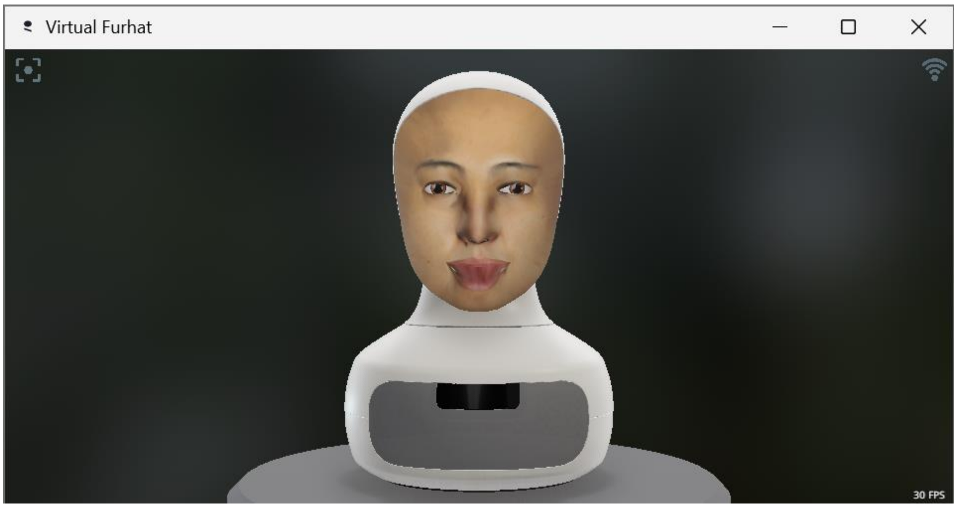

Figure 26: Furhat Model with Perfect .PNG File

6.2.2 Visual Accuracy and Realism

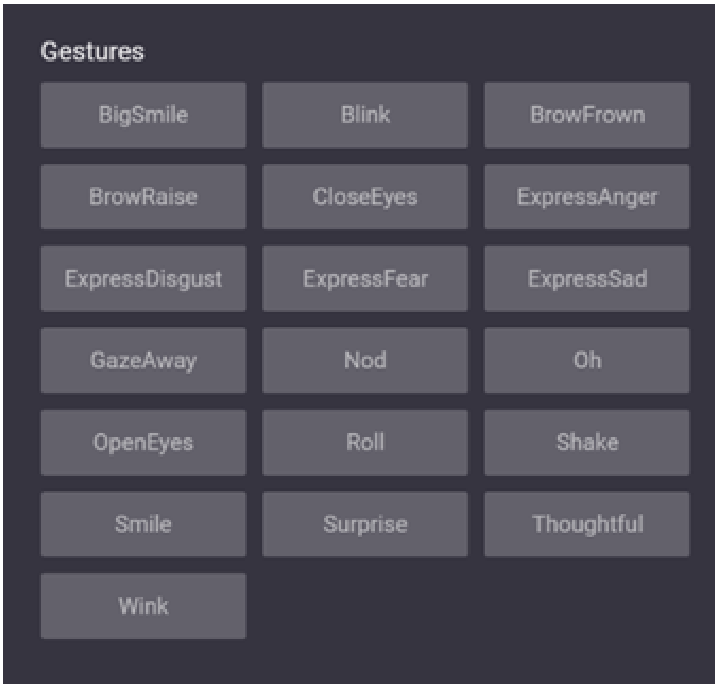

Figure 27: Selection of Gesture within the Furhat SDK

Once the 3D model was constructed using photogrammetry, testing focused on how accurately the avatar captured the user's facial features and how well it responded in terms of realism during dynamic interactions. To test expressions, the avatar was put through a series of predefined facial movements which were available through the Furhat SDK **(See Figure 27)**. The goal was to observe whether the avatar could smoothly transition between various emotional states—such as “BigSmile” **(Figure 28)**, “ExpressAnger”**(Figure 29)**, or “ExpressFear**(Figure 30)**”—while maintaining a natural appearance.

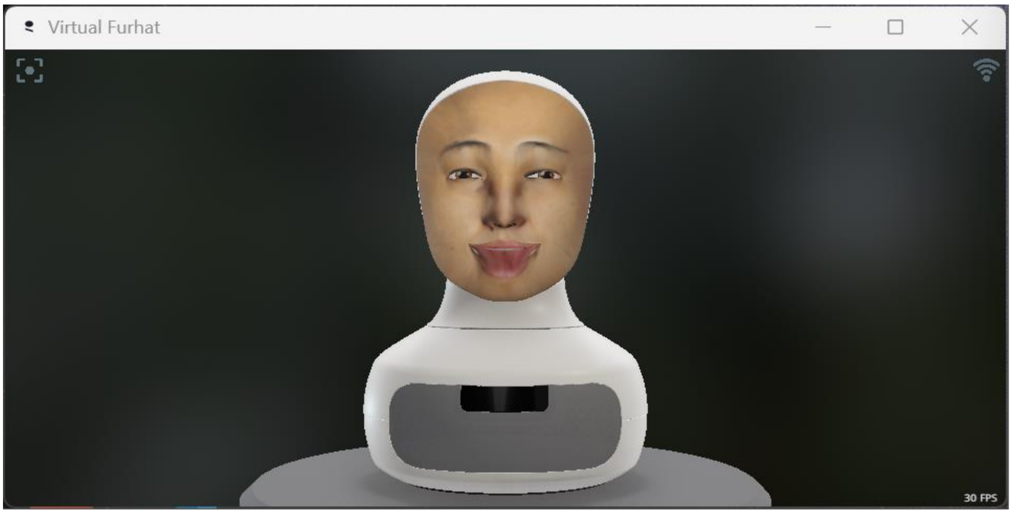

Figure 28: Model expressing “BigSmile” Gesture

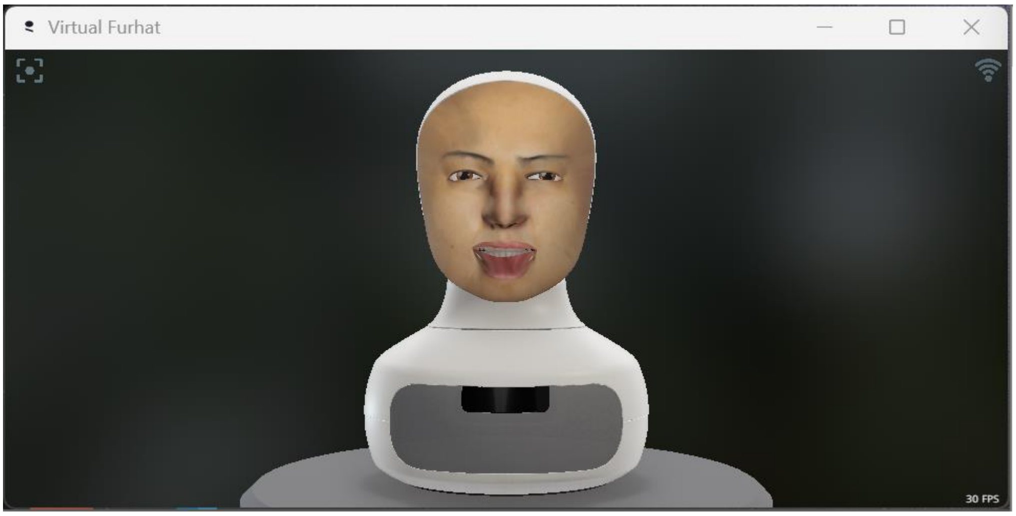

Figure 29: Model expressing “ExpressAnger” Gesture

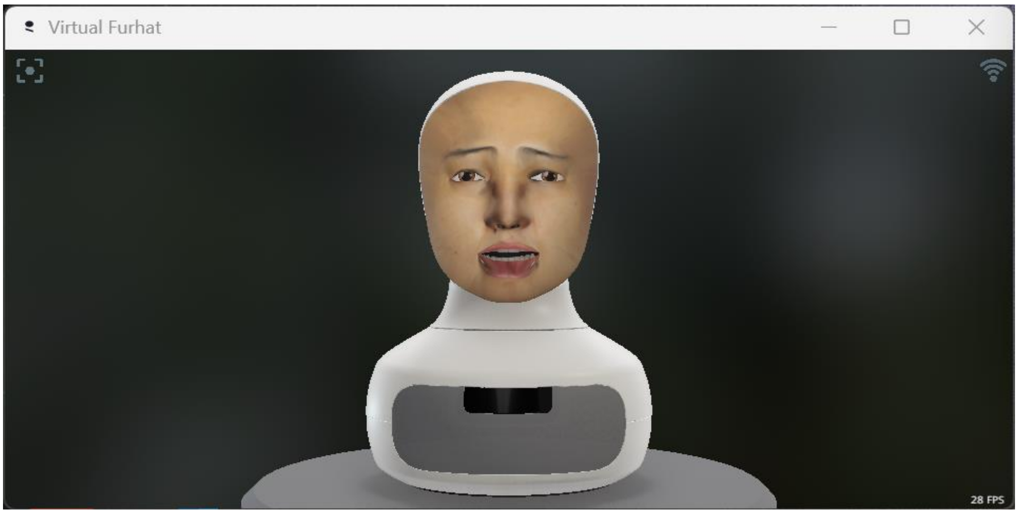

Figure 30: Model expressing "ExpressSad" Gesture

Testing revealed that the 3D model responded fluidly, with minimal distortion or stretching of the texture during more exaggerated expressions. The model's skin texture adapted well to different expressions, preserving important details that added to the realism. In conclusion, the visual accuracy and realism of the avatar were validated through a series of tests that focused not just on static representation but on dynamic expressions.

6.3 Advancing the Virtual Avatar

Following the development of the 3D model and texture mapping, the next phase of advancement involved writing code to enable real-time interactions and dynamic expressions. To achieve this, several key components were developed and integrated, including conversational flows, gesture mapping, and voice interaction. This section outlines the technical challenges, solutions, and specific code implementations used to achieve real-time interaction with the Furhat robot, expanding on the work described in Chapter 5.

6.3.1 Voice Interaction and Gesture Implementation

The implementation of real-time voice and gesture-driven interactions significantly advanced the capabilities of the virtual avatar, enhancing its ability to communicate naturally and engage users. By integrating voice selection and gesture-based responses, the avatar would provide more personalized and immersive interactions.

Voice Interaction and Customization

The voice interaction system allows selection from a range of available voices through the Furhat Robotics API. After the user selects a voice, the avatar immediately switches and confirms the selection by speaking with the newly chosen voice. This real-time switch ensures smooth, uninterrupted conversation and adds a personalized touch to the interaction.

```
if selected_voice_name in voice_names:
    # Set the selected voice on Furhat
    response = furhat.set_voice(name=selected_voice_name)
    print(f"Voice switch response: {response}")

    # Confirm the new voice with a delay to ensure the change is applied
    furhat.say(text=f"Please wait a moment while I switch to the {selected_voice_name} voice.", blocking=True)

    # Delay to allow the voice change to take effect
    import time
    time.sleep(2)  # Adjust the time as necessary

    # Now speak with the new voice
    furhat.say(text=f"Hello! I'm now speaking with the {selected_voice_name} voice.", blocking=True)
```

Figure 31: Voice Interaction Code Snippet

Gesture-Based Interaction

Gesture-based responses, such as nodding, smiling, or displaying thoughtfulness, were integrated to enhance the avatar's non-verbal communication. These gestures are triggered in response to specific user inputs, making interactions more intuitive and engaging.

```python
# Start the conversation with a smile gesture
furhat.gesture(name="Smile")
furhat.say(text="Hello! How are you today?")

# Listen to the user's response with a 10-second timeout
response = listen_with_timeout(timeout=10)

# Check the response and respond accordingly with gestures
if response.message and ("good" in response.message.lower() or "great" in response.message.lower()):
    furhat.gesture(name="BigSmile")
    furhat.say(text="I'm glad to hear that! What can I help you with today?")
elif response.message and ("not good" in response.message.lower() or "bad" in response.message.lower()):
    furhat.gesture(name="ExpressSad")
    furhat.say(text="I'm sorry to hear that. Is there anything I can do to make your day better?")
elif response.message:
    furhat.gesture(name="Thoughtful")
    furhat.say(text="I see. How can I assist you today?")
else:  # No response or unrecognized input
    furhat.gesture(name="BrowFrown")
    furhat.say(text="I didn't quite catch that. How are you feeling today?")

# Listen to the user's request with a 10-second timeout
request = listen_with_timeout(timeout=10)
```

Figure 32: Furhat Conversation with Gestures Code Snippet

Synchronization and Real-Time Interaction

The system was designed to execute both speech and gestures in real time, preventing any disjointed or out-of-sync movements. This synchronization helped create a fluid and engaging user experience. By combining voice selection and gesture-based responses, the avatar was transformed into a more responsive and interactive entity. These enhancements significantly improve user engagement, aligning with the research goal of creating a lifelike, personalized avatar.

6.3.2 Integrating Blockly for Avatar Interaction

An advancement was the integration of Blockly, a visual programming tool, to design and manage conversational flows. This integration led to significant improvements in the development process, making it more efficient and accessible, while enhancing the avatar’s interactivity. The visual interface allowed for quicker creation and modification of behaviors by arranging blocks representing actions, conditions, and logic. As a result, this streamlined the workflow and made managing complex interactions much simpler.

Figure 33: Blockly Visual Conversation Blocks

However, Blockly's conversational flow is still limited in its ability to manage more complex, open-ended interactions due to its automated state-based structure. To overcome this limitation, the integration of ChatGPT offers a significant enhancement. By incorporating ChatGPT, the avatar can manage dynamic, context-aware conversations, responding to a wider range of dialogues. This integration complements Blockly’s structured flows by adding depth and flexibility, allowing the avatar to manage complex queries and maintain richer interactions.

6.3.3 Integration of ChatGPT for Enhanced Conversational Ability

One of the most significant advancements in the virtual avatar's development was the integration of OpenAI's ChatGPT. As unlike pre-scripted responses that rely heavily on defined intents, ChatGPT provides the avatar with the ability to handle open-ended questions and complex user queries, resulting in more natural, engaging conversations.

Enhancing Flexibility and Depth in Conversations

The integration allowed the avatar to respond to a wider variety of user inputs, moving beyond basic question-answer interactions. For instance, while the avatar’s initial setup handled predefined topics like greetings or emotion-based responses, ChatGPT expanded its capacity to handle diverse subjects, such as providing detailed information on a given topic, engaging in discussions, or asking follow-up questions to keep the conversation going, giving more depth and making the avatar less reliant on rigid state transitions and pre-built dialogues, enhancing its ability to adapt to the user's input.

Technical Integration

The technical integration of ChatGPT into the avatar's existing infrastructure was a seamless process. By utilizing API calls, inputs could be sent to ChatGPT either through a text box **(See Figure 34)** or via code that transposed voice inputs into text for processing **(See Figure 35)**. This allowed for natural voice-based interactions where the integration was structured to ensure that real-time responses were delivered without noticeable delays, keeping the interaction smooth and engaging.

Figure 34: Communication with Furhat Integrated ChatGPT Via Text Box

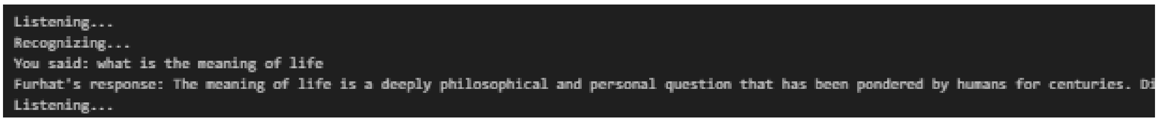

Figure 35: Communication with Furhat Integrated ChatGPT Via Microphone

Improving User Engagement

Integrating ChatGPT significantly improved user engagement by enabling the avatar to hold longer, more meaningful conversations that felt tailored to the user's needs. ChatGPT's ability to understand context and generate relevant responses kept interactions natural and reduced repetitive replies. Additionally, ChatGPT allowed the avatar to ask follow-up questions and provide deeper insights, creating a more interactive and human-like experience. Overall, this integration added flexibility and intelligence, enhancing the avatar's conversational depth and adaptability while opening new possibilities for responsive, dynamic interactions.

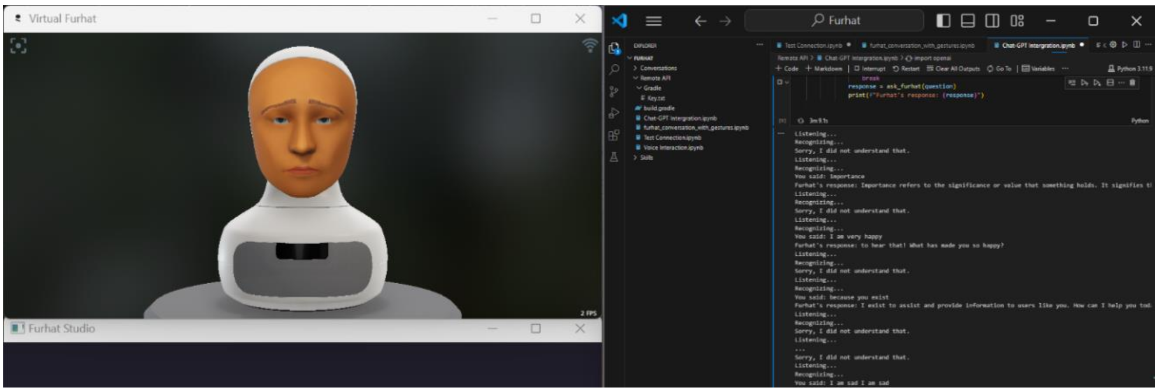

Figure 36: Random Furhat model emotional response via ChatGPT questions

6.4 Conclusion

In this chapter, the results of the testing and evaluation phases for the virtual avatar, focusing on its accuracy, performance, and interaction qualities were explored. The 3D model and texture mapping were tested thoroughly, ensuring that the avatar resembled the user while maintaining realistic expressions during dynamic interactions and the use of blendshapes demonstrated the fluidity of the model, confirming its ability to transition smoothly between various facial expressions.

In addition to the above, the integration of voice interaction and gesture-based responses made the avatar more engaging and lifelike and the addition of Blockly improved the development workflow. Finally, the integration of ChatGPT significantly enhanced the avatar’s conversational abilities, enabling it to handle complex, open-ended interactions with flexibility and depth. Overall, the results demonstrate the successful creation of a responsive and realistic avatar, meeting the project’s goals of both visual and interactive realism.

Chapter 7: Evaluation

The aim of this chapter is to critically evaluate the development of the virtual robot avatar. This assessment addresses the overall strengths and weaknesses, performance evaluation, and the impact of the research goals. Moreover, this chapter seeks to answer the key research questions from Chapter 1 and offers a detailed exploration of how advanced computer vision techniques and real-time algorithms contributed to the success of this project.

7.1 Key Research Questions Addressed

1. **Can advanced computer vision techniques be utilized to create a realistic and customizable 3D model of a human face for use in virtual robot avatars?**

One of the most significant strengths of this project lies in the successful use of advanced computer vision techniques to create a highly detailed and customizable 3D model of my face. Techniques such as photogrammetry and 3D scanning were instrumental in generating accurate textures and geometries. By stitching together 2D images from multiple angles, photogrammetry produced a high-fidelity 3D representation of my face, ensuring a strong sense of presence and immersion during virtual interactions.

However, the customization process was not without challenges. Early attempts with pre-trained models highlighted limitations in texture quality and avatar realism. This led to the decision to map the mesh of the avatar using Blender, resulting in a more lifelike model. The outcome was a detailed and realistic 3D model capable of capturing intricate facial expressions, providing a foundation for the second key question.

2. **How can the synchronization of facial movements and speech be optimized in a virtual robot avatar using real-time computer vision algorithms, and how does this optimization affect the accuracy of facial texture mapping?**

Real-time synchronization of facial movements with speech was achieved through the use of blendshapes in combination with ARKit’s facial tracking capabilities. These algorithms ensured that the avatar could mirror human expressions in real-time while maintaining accurate facial texture mapping . This synchronization allowed for fluid transitions between different facial expressions, ranging from subtle movements to more exaggerated gestures.

The project revealed that accurate synchronization directly impacted the texture mapping process, with textures adjusting dynamically based on the avatar's expression. This was especially important for maintaining realism during interactions. However, environmental factors like lighting and background noise introduced challenges. In low-light settings, the facial features might not be as visible causing irregular colour distortions and shading in particular areas of the face. Therefore location and lighting will impact low-budget attempts of recreation.

7.2 Strengths of the Project

7.2.1 High-Fidelity 3D Facial Modelling

The project successfully created high-fidelity 3D facial models using photogrammetry and 3D scanning. These methods captured intricate facial details and textures, resulting in a realistic, personalized avatar that enhanced the user's sense of presence and immersion during interactions.

7.2.2 Real-Time Facial Expression Synchronization

Real-time facial recognition and expression mimicry through blendshapes and ARKit allowed the avatar to dynamically adapt to the user's facial movements. This seamless synchronization between expressions improved user engagement, making the avatar more lifelike and immersive.

7.2.3 Integration of ChatGPT to Enhance Human-Like Robots

The integration of ChatGPT significantly boosted the avatar's conversational abilities, enabling it to engage in dynamic, context-aware conversations beyond scripted responses. This made interactions more natural, personalized, and responsive, enhancing the overall user experience and emotional connection with the avatar.

7.2.4 Summary Of Strength

In summary, the strengths of this project are rooted in the successful integration of advanced technologies for creating realistic, interactive, and dynamic avatars. The combination of high-fidelity facial modelling, real-time synchronization, and advanced conversational abilities through ChatGPT demonstrates a comprehensive approach to enhancing HMIs. These strengths provide a solid foundation for future developments in virtual avatar technology, highlighting the project's versatility, scalability, and potential for application across multiple domains.

7.3 Weaknesses of the Project

7.3.1 Limited Expression Range involving the mouth

Although the avatar effectively replicates facial expressions, it encounters challenges in conveying nuanced emotions and micro-expressions, particularly in relation to large mouth movements. To resolve this issue, adjusting the PNG texture file to depict the mouth in a closed position rather than open could offer a practical solution. The primary goal prior to this adjustment was to demonstrate proof of concept.

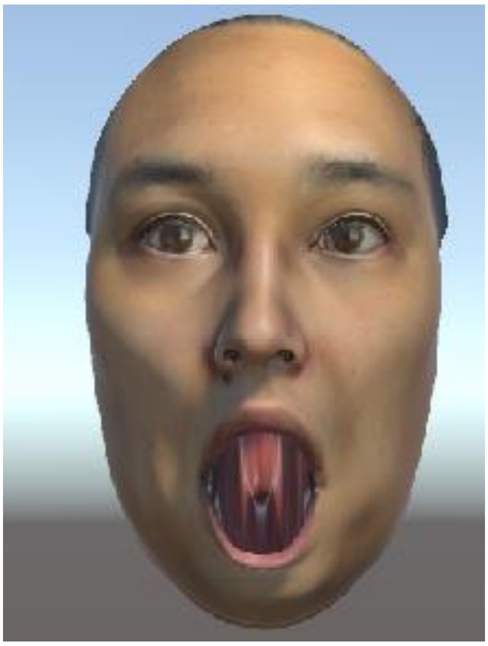

Figure : Expression Limitations Due to Absence of Inner Mouth Detail in Avatar Models

7.3.2 Performance Issues in Unfavourable Conditions

Environmental factors, such as lighting and background noise, present significant challenges. In low-light settings, facial features may become less visible, leading to irregular color distortions and uneven shading in specific areas of the face. As a result, the location and lighting conditions play a crucial role, especially in low-budget recreations. The avatar's performance diminishes in environments with poor lighting or excessive background noise.

Figure: Effects of Lighting Conditions on Facial Texture and Eye Detection Accuracy

7.3.3 Summary Of Weaknesses

Though the project made significant strides in avatar realism and interaction, the limitations in expression range, environmental adaptability, and computational efficiency present areas for improvement. Addressing these weaknesses would enhance the avatar’s usability and performance in diverse and practical applications.

7.4 Overall Evaluation

This project successfully demonstrated the potential of using advanced computer vision techniques and real-time algorithms to create a realistic, responsive, and customizable virtual avatar. Through high-fidelity 3D modelling, real-time expression synchronization, and the integration of ChatGPT, the avatar was able to deliver a highly immersive and interactive experience. These strengths show that the avatar can effectively replicate human-like features, expressions, and conversational abilities, paving the way for enhanced human-machine interaction in various fields.

Overall, the project has laid a solid foundation for future developments in virtual avatar technology. It showcases the ability to create engaging, lifelike avatars that can be personalized and adapted to a wide range of applications. Addressing the identified weaknesses would further enhance the avatar's versatility, making it more accessible and efficient across diverse use cases, such as education, healthcare, and entertainment.

Chapter 8: Conclusions

In conclusion, this dissertation has examined the complex and evolving field of HMIs, focusing on the development of realistic and responsive humanoid avatars. By applying advanced techniques in computer vision, 3D modelling, and artificial intelligence, this research has demonstrated the potential to create virtual avatars that closely replicate human facial features and expressions in real-time. These avatars act as a bridge between the physical and digital worlds, enabling more natural and immersive interactions within virtual environments.

The key contributions of this research are evident in the integration of cutting-edge technologies, such as Unity's animation and rendering capabilities, with real-time facial recognition and expression synchronization. The creation of a customizable 3D model that mirrors a user's facial movements and gestures in real time represents a significant advancement in HMI, enhancing both authenticity and emotional engagement in virtual settings.

This research also addresses the inherent challenges of creating lifelike avatars, including achieving high-fidelity 3D modelling, ensuring precise synchronization of facial expressions with speech, and overcoming computational and hardware limitations. While significant progress has been made, some limitations remain, particularly in the performance under varying conditions.

Ultimately, this dissertation has advanced the field of human-machine interaction by focusing on the creation of a virtual avatar that closely resembles the author, showcasing the future potential of realistic, responsive avatars.

Chapter 9: Overview of Future Directions – 288 Words

Looking ahead, this dissertation identifies promising avenues for future research and development, and as technology evolves, the concept of interacting with avatars that not only resemble humans but also exhibit human-like intelligence presents an exciting frontier for exploration. Developing personalized 3D avatar models with real-time expression synchronization represents a significant advancement in HMIs, but further research could enhance the utility and realism of these systems. As by integrating more advanced deep learning algorithms, avatars could better interpret and emulate a broader range of human emotions, making them more effective in sensitive contexts like counselling and cross-cultural communication. Additionally, optimizing system performance under diverse conditions remains a key challenge.

Future research should focus on improving the robustness of facial recognition and expression synchronization algorithms in varying lighting and environmental conditions. Leveraging cloud or edge computing could also reduce computational demands, making advanced avatars more accessible across different hardware platforms. Expanding the cultural adaptability of avatars is also essential for broader global adoption. Incorporating a wider range of cultural expressions and behaviours into training data could create avatars that resonate with users from diverse backgrounds.

These advancements would not only enhance user satisfaction but also broaden the applicability of virtual avatars across fields such as education, healthcare, and customer service, fostering a more inclusive digital experience where all users feel represented and understood in their interactions with technology.

Chapter 10: References

1. Blanz, V., & Vetter, T. (2023). A morphable model for the synthesis of 3D faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (pp. 157-164). [Online][Accessed: 17th June 2024]
Available at: https://dl.acm.org/doi/10.1145/3596711.3596730

2. Li, S. Z., Jain, A. K., Romdhani, S., Blanz, V., Basso, C., & Vetter, T. (2005). Morphable models of faces. *Handbook of Face Recognition*, 217-245. [Online][Accessed: 17th June 2024]
Available at: https://link.springer.com/chapter/10.1007/0-387-27257-7_11

3. Schroeder, R. (Ed.). (2001). *The social life of avatars: Presence and interaction in shared virtual environments*. Springer Science & Business Media. Chapter 1 [Online][Accessed: 17th June 2024]
Available at: https://books.google.co.uk/books?hl=en&lr=&id=d-7vegnIgJsC&oi=fnd&pg=PR9&dq=avatars+engage+with+technology+in+virtual+environments+&ots=mjb42jGDBt&sig=YFXTI4v3uLfNPzA8YLdpAUQjB28#v=onepage&q&f=false

4. Brown DE, Moenning A, Guerlain S, Turnbull B, Abel D, Meyer C. Design and evaluation of an avatar-based cultural training system. The Journal of Defense Modeling and Simulation. 2019;16(2):159-174. doi:10.1177/1548512918807593 [Online][Accessed: 20th June 2024]
Available at: https://journals.sagepub.com/doi/full/10.1177/1548512918807593

5. Falloon, G. (2010), Using avatars and virtual environments in learning: What do they have to offer?. British Journal of Educational Technology, 41: 108-122. [Online][Accessed: 20th June 2024]
Available at: https://doi.org/10.1111/j.1467-8535.2009.00991.x

6. Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003, April). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 529-536). [Online][Accessed: 20th June 2024]
Available at: https://dl.acm.org/doi/abs/10.1145/642611.642703

7. Furhat Robotics Website for the next generation of conversational Robots [Online][Accessed: 20th June 2024]
Available at: https://furhatrobotics.com/furhat-robot/

8. Unity Technologies. (2022). Unity user manual (2022.3 LTS). Unity Technologies. [Online][Accessed: June 23rd 2024]
Available at: https://docs.unity3d.com/Manual/UnityManual.html#unity-user-manual-2022.3-(lts)

9. Chamola, V., Vineet, A., Nayyar, A., & Hossain, E. (2020). Brain-Computer Interface-Based Humanoid Control: A review. *Sensors*, *20*(13), 3620. [Online][Accessed: 26th June 2024]
Available at: https://doi.org/10.3390/s20133620

10. Villa-Uriol, M., Sainz, M., Kuester, F., & Bagherzadeh, N. (2003). Automatic creation of three-dimensional avatars. *Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE*. [Online][Accessed: 26th June 2024]

Available at: https://doi.org/10.1117/12.473109

11. Zhang, Q., Liu, Z., Guo, B., Terzopoulos, D., & Shum, H. (2006). Geometry-driven photorealistic facial expression synthesis. *IEEE Transactions on Visualization and Computer Graphics*, *12*(1), 48–60.
[Online][Accessed: 26th June 2024]
Available at: https://doi.org/10.1109/tvcg.2006.9

12. Raga, E. A., & Sumida, S. S. (2009). Anatomical considerations in facial motion capture. *Computer Graphics*, *43*(2), 1–9.
[Online][Accessed: 26th June 2024]
Available at: https://doi.org/10.1145/1629216.1629220

13. Thalmann, D. (2000). Challenges for the research in virtual humans. In *Proc. AGENTS 2000*.
[Online][Accessed: 28th June 2024]
Available at: https://ieeexplore.ieee.org/abstract/document/7504761/

14. Latoschik, M. E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., & Botsch, M. (2017, November). The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM symposium on virtual reality software and technology* (pp. 1-10).
[Online][Accessed: 28th June 2024]
Available at: https://dl.acm.org/doi/abs/10.1145/3139131.3139156

15. Aneja, D., McDuff, D., & Shah, S. (2019, October). A high-fidelity open embodied avatar with lip syncing and expression capabilities. In *2019 International conference on multimodal interaction* (pp. 69-73).
[Online][Accessed: 28th June 2024]
Available at: https://dl.acm.org/doi/abs/10.1145/3340555.3353744

16. Diamandis, P. (n.d.). Metatrend #5: Humanoid Robots & Avatars Are Arriving. Peter Diamandis - Innovation & Entrepreneurship Community.
[Online][Accessed: July 8, 2024]
Available at:
https://www.diamandis.com/blog/metatrend_5_humanoid_robots_avatars_are_arriving#:~:text=These%20avatars%20enable%20a%20remote,require%20during%20times%20of%20shortage.

17. Arcao, J A., Cadag, V C., Martinez, V., Roxas, E., Serrano, K J., & Tolentino, R E. (2019, April 1). Holographic Projection of 3D Realistic Avatar That Mimics Human Body Motion.
[Online][Accessed: July 8, 2024]

Available at: https://doi.org/10.1109/iciict1.2019.8741454

18. Jourabloo, A., Gecer, B., Saragih, J., Wei, S., Wang, T., Lombardi, S., Belko, D., Trimble, A., & Badino, H. (2021). Robust Egocentric Photo-realistic Facial Expression Transfer for Virtual Reality.
[Online][Accessed: July 10, 2024]
Available at: *https://arxiv.org/abs/2104.04794*

19. Barua, H B., Sarkar, C., Kumar, A A., Pal, A., & Balamuralidhar, P. (2020, January 1). I can attend a meeting too! Towards a human-like telepresence avatar robot to attend meeting on your behalf. Cornell University.
[Online][Accessed: July 10, 2024]
Available at: https://doi.org/10.48550/arxiv.2006.15647

20. Bombari, D., Mast, M S., Cañadas, E., & Bachmann, M. (2015, June 24). Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. Frontiers Media, 6.
[Online][Accessed: July 10, 2024]
Available at: https://doi.org/10.3389/fpsyg.2015.00869

21. Earnshaw, R., Magnenat-Thalmann, N., Terzopoulos, D., & Thalmann, D. (1998, September 1). Computer Animation for Virtual Humans. Institute of Electrical and Electronics Engineers, 18(5), 20-23.
[Online][Accessed: July 11, 2024]
Available at: https://doi.org/10.1109/mcg.1998.708557

22. Hong, F., Zhang, M., Peng, L., Cai, Z., Yang, L., & Liu, Z. (2022, January 1). AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars. Cornell University.
[Online][Accessed: July 10, 2024]
Available at: https://doi.org/10.48550/arxiv.2205.08535

23. Roble, D., Hendler, D., Buttell, J., Cell, M., Briggs, J., Reddick, C., Iannazzo, L., Li, D., Williams, M., Moser, L., Wong, C., Kachkovski, D., Huang, J., Zhang, K., McLean, D., Cloudsdale, R., Milling, D., Miller, R., Lawrence, J., & Chien, C. (2019, July 28). Real-time, single camera, digital human development.
[Online][Accessed: July 11, 2024]
Available at: https://doi.org/10.1145/3306305.3332366

24. Chatzitofis, A., Zarpalas, D., Kollias, S., & Daras, P. (2019). DeepMoCaP: Deep optical motion capture using multiple depth sensors and Retro-Reflectors. Sensors, 19(2), 282.
[Online][Accessed: July 11, 2024]
Available at: https://doi.org/10.3390/s19020282

25. Raga, E. A., & Sumida, S. S. (2009). Anatomical considerations in facial motion capture. Computer Graphics, 43(2), 1–9.
[Online][Accessed: July 11, 2024]
Available at: https://doi.org/10.1145/1629216.1629220

26. Gillies, M., Pan, X. and Slater, M. (2010) "Piavca: a framework for heterogeneous interactions with virtual characters," Springer Science+Business Media, 14(4),p. 221-228.
[Online][Accessed: July 13, 2024]
Available at: https://doi.org/10.1007/s10055-010-0167-5

27. Meet the humanoids: 8 robots ready to revolutionize work. (2024, April 27). Freethink. [Online][Accessed: July 13, 2024]
Available at: https://www.freethink.com/robots-ai/humanoids

28. Nowak, K L., & Fox, J. (2023, October 26). Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations.
[Online][Accessed: July 13, 2024]
Available at: https://rcommunicationr.org/index.php/rcr/article/view/39

29. Jiang, R., Wang, C., Zhang, J., Chai, M., He, M., Chen, D., & Liao, J. (2023, January 1). AvatarCraft: Transforming Text into Neural Human Avatars with Parameterized Shape and Pose Control. Cornell University.
[Online][Accessed: July 24, 2024]
Available at: https://doi.org/10.48550/arxiv.2303.17606

30. Dafarra, S., Pattacini, U., Romualdi, G., Rapetti, L., Grieco, R., Darvish, K., Milani, G., Valli, E., Sorrentino, I., Viceconte, P M., Scalzo, A., Traversaro, S., Sartore, C., Elobaid, M., Guedelha, N., Herron, C W., Leonessa, A., Draicchio, F., Metta, G., . . . Pucci, D. (2024, January 24). iCub3 avatar system: Enabling remote fully immersive embodiment of humanoid robots. American Association for the Advancement of Science, 9(86).
[Online][Accessed: July 24, 2024]
Available at: https://doi.org/10.1126/scirobotics.adh3834

31. Nagendran, A., Pillat, R., Hughes, C., & Welch, G. (2023, November 9). Continuum of virtual-human space: towards improved interaction strategies for physical-virtual avatars.
[Online][Accessed: July 24, 2024]
Available at: https://dl.acm.org/doi/10.1145/2407516.2407554

32. Bombari, D., Mast, M S., Cañadas, E., & Bachmann, M. (2015, June 24). Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. Frontiers Media, 6.
[Online][Accessed: July 24, 2024]
Available at: https://doi.org/10.3389/fpsyg.2015.00869

33. Garau, M., Slater, M., Pertaub, D., & Razzaque, S. (2005, February 1). The Responses of People to Virtual Humans in an Immersive Virtual Environment. The MIT Press, 14(1), 104-116.
[Online][Accessed: July 24, 2024]
Available at: https://doi.org/10.1162/1054746053890242

34. Cheetham, M., Suter, P., & Jäncke, L. (2011, January 1). The human likeness dimension of the “uncanny valley hypothesis”: behavioral and functional MRI findings. Frontiers Media, 5.
[Online][Accessed: July 25, 2024]
Available at: https://doi.org/10.3389/fnhum.2011.00126

35. MacDorman, K F., Green, R., Ho, C., & Koch, C T. (2009, May 1). Too real for comfort? Uncanny responses to computer generated faces. Elsevier BV, 25(3), 695-710.
[Online][Accessed: July 25, 2024]
Available at: https://doi.org/10.1016/j.chb.2008.12.026

36. Jalil, S A., Osburn, B., Huang, J., Barley, M., Markovich, M., & Amor, R. (2012, July 2). Avatars at a meeting.
[Online][Accessed: July 25, 2024]
Available at: https://doi.org/10.1145/2379256.2379270

37. Kabadayi, B., Zielonka, W., Bhatnagar, B L., Pons-Moll, G., & Thies, J. (2023, January 1). GAN-Avatar: Controllable Personalized GAN-based Human Head Avatar. Cornell University.
[Online][Accessed: July 25, 2024]
Available at: https://doi.org/10.48550/arxiv.2311.13655

38. Magnenat-Thalmann, N., Yumak, Z., & Beck, A. (2014, September 1). Autonomous virtual humans and social robots in telepresence.
[Online][Accessed: July 25, 2024]
Available at: https://doi.org/10.1109/mmsp.2014.6958836

39. Tanaka, K., Onoue, S., Nakanishi, H., & Ishiguro, H. (2013, May 1). Motion is enough: How real-time avatars improve distant communication.
[Online][Accessed: July 28, 2024]
Available at: https://doi.org/10.1109/cts.2013.6567270

40. Cao, C., Simon, T., Kim, J K., Schwartz, G., Zollhoefer, M., Saito, S., Lombardi, S., Wei, S., Belko, D., Yu, S., Sheikh, Y., & Saragih, J. (2022, July 1). Authentic volumetric avatars from a phone scan. Association for Computing Machinery, 41(4), 1-19.
[Online][Accessed: July 28, 2024]
Available at: https://doi.org/10.1145/3528223.3530143

41. Dafarra, S., Darvish, K., Grieco, R., Milani, G., Pattacini, U., Rapetti, L., Romualdi, G., Salvi, M., Scalzo, A., Sorrentino, I., Tomè, D., Traversaro, S., Valli, E., Viceconte, P M., Metta, G., Maggiali, M., & Pucci, D. (2022, January 1). iCub3 Avatar System. Cornell University.
[Online][Accessed: July 28, 2024]
Available at: https://doi.org/10.48550/arxiv.2203.06972

42. Habermann, M., Liu, L., Xu, W., Pons-Moll, G., Zollhoefer, M., & Theobalt, C. (2022, January 1). HDHumans: A Hybrid Approach for High-fidelity Digital Humans. Cornell University.
[Online][Accessed: July 28, 2024]
Available at: https://doi.org/10.48550/arXiv.2210.

43. Magnenat-Thalmann, N., & Egges, A. (2023, October 26). Interactive Virtual Humans in Real-Time Virtual Environments.
[Online][Accessed: July 29, 2024]

Available at: https://ijvr.eu/article/download/2682/8740

44. Wang, D., Dai, B., Deng, Y., & Wang, B. (2023, January 1). AgentAvatar: Disentangling Planning, Driving and Rendering for Photorealistic Avatar Agents. Cornell University.
[Online][Accessed: July 29, 2024]
Available at: https://doi.org/10.48550/arXiv.2311.

45. González-Franco, M., Ofek, E., Pan, Y., Antley, A., Steed, A., Spanlang, B., Maselli, A., Banakou, D., Pelechano, N., Orts-Escolano, S., Orvalho, V., Trutoiu, L., Wojcik, M., Sánchez-Vives, M V., Bailenson, J N., Slater, M., & Lanier, J. (2020, November 3). The Rocketbox Library and the Utility of Freely Available Rigged Avatars. Frontiers Media, 1.
[Online][Accessed: July 29, 2024]
Available at: https://doi.org/10.3389/frvir.2020.561558

46. Zakharov, E., Ivakhnenko, A., Shysheya, A., & Lempitsky, V. (2020, January 1). Fast Bi-layer Neural Synthesis of One-Shot Realistic Head Avatars. Cornell University.
[Online][Accessed: July 31, 2024]
Available at: https://doi.org/10.48550/arxiv.2008.10174

47. Higgins, D., Fribourg, R., & McDonnell, R. (2021, May 31). Remotely Perceived: Investigating the Influence of Valence on Self-Perception and Social Experience for Dyadic Video-Conferencing With Personalized Avatars. Frontiers Media, 2.
[Online]
[Accessed: July 31, 2024]
Available at: https://doi.org/10.3389/frvir.2021.668499

48. Jourabloo, A., Geçer, B., Torre, F D L., Saragih, J., Wei, S., Wang, T., Lombardi, S., Belko, D., Trimble, A., & Badino, H. (2021, January 1). Robust Egocentric Photo-realistic Facial Expression Transfer for Virtual Reality. Cornell University.
[Online][Accessed: July 31, 2024]
Available at: https://doi.org/10.48550/arXiv.2104

.

49. Waltemate, T., Gall, D., Roth, D., Botsch, M., & Latoschik, M E. (2018, April 1). The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. Institute of Electrical and Electronics Engineers, 24(4), 1643-1652.
[Online][Accessed: July 31, 2024]
Available at: https://doi.org/10.1109/tvcg.2018.2794629

50. Borkman, S., Crespi, A., Dhakad, S., Ganguly, S., Hogins, J., Jhang, Y., Kamalzadeh, M., Li, B., Leal, S., Parisi, P., Romero, C I E., Smith, W C., Thaman, A., Warren, S L., & Yadav, N. (2021, January 1). Unity Perception: Generate Synthetic Data for Computer Vision. Cornell University.
[Online][Accessed: July 31, 2024]

Available at: https://doi.org/10.48550/arxiv.2107.04259

51. Derpanis, K G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A A A., Tzionas, D., & Black, M J. (2019, January 1). Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. Cornell University.
[Online][Accessed: July 31, 2024]
Available at: https://doi.org/10.48550/arxiv.1904.05866

52. Roble, D., Hendler, D., Buttell, J., Cell, M., Briggs, J., Reddick, C., Iannazzo, L., Li, D., Williams, M., Moser, L., Wong, C., Kachkovski, D., Huang, J., Zhang, K., Hakim, S T., Cloudsdale, R., Milling, D., Miller, R., Lawrence, J., & Chien, C. (2019, July 28). Real-time, single camera, digital human development.
[Online][Accessed: July 31, 2024]
Available at: https://doi.org/10.1145/3306305.3332366

53. Berthault, A., Kato, T., & Shirai, A. (2023, January 1). Avatar Fusion Karaoke: Research and development on multi-user music play VR experience in the metaverse. Cornell University.
[Online][Accessed: August 1st, 2024]
Available at: https://doi.org/10.48550/arXiv.2308.

54. Hong, F., Zhang, M., Pan, L., Cai, Z., Yang, L., & Liu, Z. (2022, January 1). AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars. Cornell University.
[Online][Accessed: August 1st, 2024]
Available at: https://doi.org/10.48550/arXiv.2205.

55. Fortini, L., Leonori, M., Gandarías, J M., Momi, E D., & Ajoudani, A. (2023, January 1). Markerless 3D human pose tracking through multiple cameras and AI: Enabling high accuracy, robustness, and real-time performance. Cornell University.
[Online][Accessed: August 1st, 2024]
Available at: https://doi.org/10.48550/arXiv.2303.

56. Latoschik, E, M. et al. (2017) The effect of avatar realism in immersive social virtual realities.
[Online][Accessed: August 2nd, 2024]
Available at: https://doi.org/10.1145/3139131.3139156.

57. Kirpes C, Hu G, Sly D. The 3D Product Model Research Evolution and Future Trends: A Systematic Literature Review. *Applied System Innovation*. 2022; 5(2):29.
[Online][Accessed: August 22nd, 2024]
Available at: https://doi.org/10.3390/asi5020029

58. Edify. (2024, August 22). Realistic digital avatars and the uncanny valley. *Edify*.
[Online][Accessed: August 22nd, 2024]
Available at: https://www.edify.ac/blog/realistic-digital-avatars-and-the-uncanny-valley

59. Fink, M. C., Robinson, S. A., & Ertl, B. (2024). AI-based avatars are changing the way we learn and teach: Benefits and challenges. Frontiers in Education, 9, 1416307.
[Online][Accessed: August 22nd, 2024]

Available at: https://doi.org/10.3389/feduc.2024.1416307
60.