AALBORG UNIVERSITET

Title:
TruePresence: Affecting Social Presence and Uncanniness with Animation Variety in LLM-Based Virtual Humans

Theme:
Master's Thesis

Project Period:
10th SEMESTER

Project Group:
1002

Participant(s):
Julius Ebenau Winther
Krestian Holm bendixen
Lukas Lund Nysted

Supervisor(s):
Ali Adjorlu

Page Numbers: 16

Date of Completion:
May 26, 2025

Abstract:

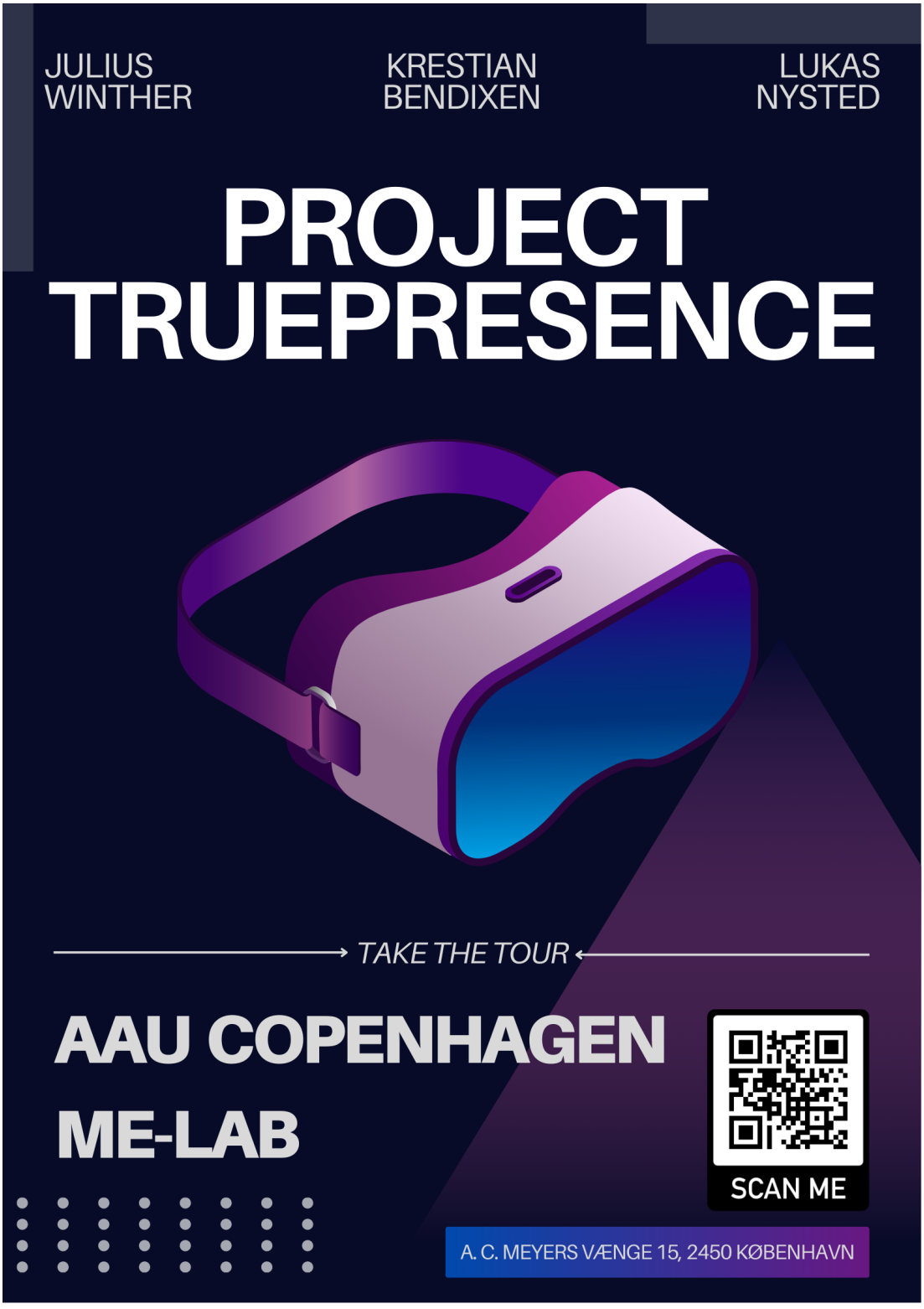
In this paper, we investigate the perceived Social Presence and Uncanny Valley Effect of a virtual reality application experienced by users engaging with a character built on a large language model (LLM). The investigation features two variations of the same VR experience with different degrees of Animation Variety. Animation variety is studied to investigate its effect on users' perceived Social Presence and Uncanniness from an LLM-Based character built on the Convai framework. The two variants were subsequently tested by using a 7-point Likert-type multiple-choice self-report questionnaire to find the perceived Social Presence and Uncanniness, as well as following qualitative questions to identify the most relevant features of the character have the most impact on the user perception.
We found that there is statistically significant evidence to support that increased animation variety of LLM-based Non-Player Characters (NPCs) in VR increases social presence and decreases perceived uncanniness.

TruePresence: Affecting Social Presence and Uncanniness with Animation Variety in LLM-Based Virtual Humans

Julius Ebenau Winther* MSc Medialogy

Krestian Holm Bendixen† MSc Medialogy

Lukas Lund Nysted‡ MSc Medialogy



Abstract

In this paper, we investigate the perceived Social Presence and Uncanny Valley Effect of a virtual reality application experienced by users engaging with a character built on a large language model (LLM). The investigation features two variations of the same VR experience with different degrees of Animation Variety. Animation variety is studied to investigate its effect on users' perceived Social Presence and Uncanniness from an LLM-Based character built on the Convai framework. The two variants were subsequently tested by using a 7-point Likert-type multiple-choice self-report questionnaire to find the perceived Social Presence and Uncanniness, as well as following qualitative questions to identify the most relevant features of the character have the most impact on the user perception.
We found that there is statistically significant evidence to support that increased animation variety of LLM-based Non-Player Characters (NPCs) in VR increases social presence and decreases perceived uncanniness.

Index Terms: Social Presence, Uncanniness, Animation, Virtual Human, Artificial Intelligence, Large Language Models, Gen-AI, Convai, Unity, Virtual Reality, Digital twins

1 Introduction

Large language models (LLMs) have seen a lot of development in recent years, gaining popularity in products such as Chat-GPT, DeepSeek, and Google Gemini, serving as assistants for productivity, but have recently started being deployed for use in virtual characters [8]. This technology allows the characters to adapt to unscripted player interactions, providing dynamic real-time conversations in virtual environments, and we believe that a sizable number of applications will feature these characters in the future. VR has also seen uses for teaching competencies, including soft skills such as preparing for job interviews, as seen in the VR-ACE project co-funded by the European Union [57]. The project features only scripted non-player characters, but adding LLM functionalities to use cases like it would perhaps prove a much more versatile and engaging experience for users. To explore the conjunction of these emerging technologies this paper, supported by our external collaborators Virsabi, intends to investigate the perceived co-presence of users and the uncanny valley effect in a virtual reality application featuring one such character, acting as a guide in a digital twin of the Multisensory Experience Lab (ME-Lab) at Aalborg University Copenhagen (AAU CPH).

*e-mail: jwinth20@student.aau.dk
†e-mail: hbendi20@student.aau.dk
‡e-mail: lnyste20@student.aau.dk

2 Related Work

This section addresses useful findings from previous studies and established literature regarding topics related to our research.

2.1 Why Apply LLM-Based Virtual Humans

> *"Building a humanoid agent, an artificial entity as an equivalent to a human has been one of the core motivations for the research in the field of Artificial Intelligence since its inception [11]."* **[62]**

In 2005 [21] found that co-presence can be elicited from conversational agents in virtual environments even with shallow knowledge, which according to [7] is enough to elicit believability. Yet in modern times [50] found that participants experienced uncanniness and discomfort when text was perceived as - or actually was AI-generated. Luckily participant AI detection was low (at 42% in their study). But discomfort was still rated high when text was actually AI generated.

This aligns with [13], who showed that a personalized view is more important for perceived helpfulness and that suspecting AI origin negatively affects perceived helpfulness. This means that character responses should aim to be believable and not too generic.

Considering that the use case of the experience is a Digital Twin of a laboratory with the purpose of educating visitors, it seems relevant that [2] found that non-generative AI can significantly increase Continuation Desire, which might be worth considering if the goal of the use case is to make participants explore all that the virtual lab has to offer.

However more importantly [35] found that interactive LLMs increased engagement compared to non-interactive systems and that context relevance, conversation memory and rule-based response is more important than knowledge retrieval for a personalized learning experience. The attempt for striving towards contextual responses. can result in breaking believability according to [60] who said that users might disregard the agent if the user is delivered inaccurate, missing, or outdated content, when expecting the agent affords interactive and contextual responses. Meaning that maintaining accuracy is still important. This risk however can be negated using RAG, or retrieval-augmented generation, which involves having separate or specific knowledge bases with increased priority helps accuracy and reduces hallucinations[27].

A generated personalized experience increased engagement, Agency, interactivity, narrative reflection, usability, and possibly Immersion in VR[32]. Additionally [62] found that Contextual understanding and emotional appropriateness is important for plausible, realistic responses.

Believing events are real is essential to Plausibility Illusion, which increases when events in the virtual environment refer directly to the user or are out of their direct control [55]. This could be indicators that generative content is also beneficial for overall Presence, since it encompasses Plausibility Illusion, which is related to Social Presence.

Interestingly, [34] found that precision and context relevance of LLMs can be improved by emotional stimuli, and among tested

LLM models ChatGPT-4o had most improvement with emotional stimuli and increased temperature.

[58] Evaluated information accuracy and found that even personality traits of an LLM model can affect its accuracy. The trait of 'Meticulousness' in the 'Personality Style' tool of Convai increased its accuracy because it affected the dependency of internal sources of information compared to external sources. Having an adjustable factor to affect the ratio between memory and context based responses seems like an invaluable tool when having to accommodate a satisfactory compromise of contextual dialogue and accurate replies. Aside from this accuracy optimization, it was also found that format and superfluous information in the knowledge base also affected accuracy.
Besides these usable features, Convai's Knowledge Bank is conveniently a RAG system, which can further help us find the golden ratio between Contextual dialogue and avoid inaccuracies.

Luckily LLMs are better at generating contextual responses due to their autoregressive generation, making each new response factor in all previous ones, especially since GPT-4o presents a novel opportunity by analyzing text-based data to infer emotional states and engagement levels[22]. This is particularly useful when wanting both accurate and context based answers [58].

2.1.1 Embodiment

According to [21], a conversational agent in a virtual environment significantly increased social presence even when participants did not communicate with it.

This correlates with [56] who said interpersonal distances and proxemics apply in VR too.

Similarly, [4] found that participants elicited social behavior like interpersonal distance and reacted to non-verbal cues such as eye contact and interpersonal distance even with an agent in non-verbal situations compared to a cylinder, indicating that people behave differently when around an embodied humanoid representation. Interpersonal distance might therefore be a more reliable measure of Social Presence than subjective survey ratings[4]. Interestingly, another indicator of social behavior being exhibited by users were their inhibited performance of memory tasks when surveilled, which also could be interpreted as an argument for social presence being elicited by the agent. Which is in line with [6] who also argues that self-report questionnaires might be effective and the best tool for measuring how people perceive an embodied agent on a conceptual level, but not necessarily how they will react to that embodied agent. This is probably due to self-report measures being adequate for measuring the appearance's effect on elicited co-presence yet not for behavioral realism in their study.

> *"Overall, our findings highlight embodiment as a crucial factor of AI's influence on human behavior"*
> **[36]**

[36] found that VR-Embodiment conversation agents increase spatial presence, co-presence, and social presence compared to those who interacted with text-based versions.

Similarly, [51] showed Embodied conversational VR-agents significantly enhancing social presence compared to a non-embodied (voice-only) agents, but as a consequence, agents now elicited uncanniness.
Evidently, an embodied character should be designed to be believable yet desirably not uncanny.

2.2 Virtual Reality

To include VR in the project, it's relevant to research what it specifically offers in improvements to the quality of the experience. While not essential it is still beneficial for the use case of the experience that first-person storytelling in VR increases immersion and VR increases engagement, interactivity, Agency and narrative reflection.[32]

2.2.1 Social- and Co-Presence

> *"Creating high levels of co-presence is a complicated process involving an appropriate fit between levels of each type of realism"* **[6]**

When measuring VR's effect on Social Presence and believability in a Social Skills Training scenario with virtual characters [23] found significantly increased perceived Social Presence when using VR compared to conventional computer screens, and believability, while not significantly higher, was also partially increased in VR.

> *"This study should also encourage further work on measurement of effects of social presence and its impact on the design of VR experiences.""* **[23]**

Presence in VR is a rich and diverse concept, often measured using subjective questionnaires [56].

According to [6] measured Co-Presence is affected by how much an embodied agent resembles a human being in appearance and behavior, making it potential candidate for evaluation in this project.

s

2.2.2 Performance

Hardware performance in VR varies from headset to headset, but it is generally very restricting. This is partly because VR headsets have to render every frame twice, once for each eye, but also because many VR headsets have moved away from being wired up to a computer, requiring headsets to do onboard computing, while still fitting in a somewhat small and light frame.
Using a cutting-edge VR headset such as the Meta Quest 3 as an example, it is recommended not to have more than 1.3 million to 1.8 million triangles rendered at once, which is a big leap over the previous generation (Meta Quest 2) with a recommended 750k-1m triangles [41]. This is not a big triangle budget, and while triangles are not the only thing to keep in mind when it comes to performance, it is indicative of the general performance limits.

2.3 Creating a Believable Character

> *"The pursuit of the perfect virtual human has been the holy grail for graphics researchers for the last few decades[12][63]"* [43]

Creating a realistic simulation of a virtual character requires that the virtual character feels believable [38]. To develop a believable character, there are many factors to consider, some of which counteract each other.
Some of the facets of believable characters, such as appearance, do not rely on behavior and can therefore be achieved through conventional AI-based characters. However, other facets are more complicated to achieve and lend themselves well to being explored through LLM-based characters. Examples include: Being aware of- and reactive to the environment, memory, language understanding and -generation [38].
It could be argued that a good indicator of a believable character is the level of Social Presence experienced by the user, as it refers to the feeling of really being there with the character [10].

Therefore, striving to increase co-presence should be an important goal when designing a believable character. However, achieving a high level of Social Presence is difficult, as Social Presence is a complicated concept that requires multiple counteracting modalities of realism to work in sync [6]. For example, higher appearance- and animation realism lead to a stronger feeling of Social Presence as well as making a character be perceived as more attractive [1], however, if too large of an incongruence between appearance and behavioral realism occurs, it can diminish Social Presence [6]. This is supported by[56] who said characters should move appropriately depending on context, meaning high-quality anthropomorphism should be met with very natural animation, while cartoon characters do not have to meet the same expectations. which is in accordance with [18] who found that "machine-like" motion was rated better than more natural motion when applied to a low anthropomorphic robot with synthetic text-to-speech audio.

Similarly [26] stated that animation does not have to be anthropomorphic to convey intent and emotion.

[26] However, it has also been found that making users more attentive to the character improves Social Presence [9].

When evaluating overall acceptance, trust, and appropriateness in virtual humans, [49] found it to be significantly affected by natural animations and a human-sounding voice and interestingly, less animated agents were more persuasive.

Additionally a study by[18] suggested that voice and body motion with higher anthropomorphism elicited more likability and realism.

2.3.1 Visual Fidelity

Visual fidelity of the character refers to the "life-likeness" of the 3D character model and texturing. It is important to consider this when creating a believable character designed to induce social presence, since a more realistic/life-like character increases social presence when compared to stylized- or lower fidelity characters [39][1][19]. However, this does not apply to the user's own avatar, as having a realistic user avatar seemingly does not boost social presence [39].
Creating a visually high fidelity character also impacts users' trust in the character as well as perceived realism, social presence, and empathy [19]. Visual high fidelity was also found to impact the perceived realism of a character's voice, in the sense that a realistic character is perceived to have a less realistic voice than a stylized character [19]. However, perceived voice realism was not found to affect Social Presence in a large way [19].
Furthermore, a study by *M. Mustafa, et al.* [43] showed a correlation between NPCs that scored high realism on a Likert scale and measured uncanniness. This indicates that increasing social presence through visual fidelity inconveniently also increases uncanniness similar to [39]. However, following that logic, it could also indicate that uncanniness achieved through visual fidelity does not affect social presence.
Finally, it should also be considered that often, a more realistic character costs a larger part of the overall performance budget, due to being more densely detailed both in terms of 3D geometry and texture size, and as was mentioned earlier, VR is restrictive when it comes to performance.

2.3.2 Character Aesthetics and Personality

In addition to how realistic the character looks and acts, the aesthetics and personality of the character also matter. Character aesthetics are generally the most distinctive feature of a character; however, if the gender of the character is taken into account, voice takes precedence over character aesthetics [46]. A way to increase humanness is through attractiveness [25], which can be increased by a combination of body movement and character aesthetics[46]. Aesthetic appeal is also mentioned by [32] as significantly increasing engagement and continuous learning. A study by *Raul B Paradeda, et al.* [47] also found an indication that having a well dressed and kind character caused higher levels of felt trustworthiness and competency. However, they also discovered that if a character was acting mean and being dishonest it led to more immersion [47]. This means that if a use case requires an NPC to be trustworthy and reliable, it risks making the experience feel less immersive. It could indicate that in these cases, creating a believable character would have to rely more on other believability-inducing features, such as visual fidelity mentioned earlier. However, there is evidence to suggest that animation outweighs the influence of visual fidelity [48], and interestingly, disinclination to approach a character is influenced by behavioral realism and has nothing to do with character aesthetics [6], which points to the importance of well-crafted behavioral realism when it comes to user and character interactions.

2.3.3 Behavioral Realism

> *"Movement is fundamental to animals—including human beings—and thus to robots as well. Its presence changes the shape of the uncanny valley."* **[42]**

Behavioral Realism refers to a character's body movement and how they react to the world around them. Like visual fidelity, behavioral realism has also been found to increase Social Presence [64][5][51][52]. However, creating realistic behavior is difficult as incongruences between how a character should act and how they actually act, can elicit unwanted feelings or responses from a user [54][61]. For example if the animation speed of a smile is halved it

is considered creepy instead of happy [42], or if a character lacks facial expressions, it feels more uncanny [59]. If the character has facial expressions, and there is an incongruence between the emoted facial expressions and the expected facial expressions, the character might seem psychopathic [59]. Nevertheless, facial expressions performed at the right time in the right context can increase focus, immersion, involvement, and make communication feel easier [30][20]. Furthermore, complex gesture sequences have been found to make characters feel more natural, friendly, and trustworthy [44]. If those gestures are then slightly altered between sequences without changing the meaning of those gestures, the behavioral congruence is improved [17]. However, users are more lenient regarding temporal incongruence of voices and gestures in inter-character conversations when compared to body movement [16], meaning body animations have to be produced in a more precise manner.

Conversely, [40] claimed that visual cues aren't as perceptually important when motion cues are strong. Interestingly, what determines well perceived behavioral realism differs between male and female users [30]. Men prefer functional expressions and being in close proximity with the character, whereas women prefer emotional expressions and being at a farther distance from the character [30]. However, women also feel a larger difference in Social Presence based on distance; a character being physically close to a woman decreases comfort, perceived realism, and Social Presence [30]. This means that when designing proxemics concerning behavioral realism for a character, the gender of the user should, in an optimal situation, be taken into account.

2.4 Action Paralysis

Where LLM-based NPCs differ from conventional NPCs is their ability to have open conversations with the user. This is an important distinction to consider when developing an LLM-based NPC because in a 2025 study by *Yu Nong, Hai-Tao Zhang and Jia-Qiang Sun* [45] it was found that:

> *"Some participants, especially those less experienced with AI systems, initially found the open-ended nature of conversations challenging, suggesting a need for better onboarding or tutorials"* [45].

This is known as Action paralysis, also referred to as "Decision Paralysis" or "Choice Overload", and it is a concept that describes people's significantly decreased motivation to take action when presented with an overwhelming amount of options [28].
A way to alleviate potential action paralysis is to simply do the opposite of what causes action paralysis in the first place, namely, reducing the number of choices presented to the user [28].
While this does not directly influence our research question, it is something to be aware of when designing the application to minimize conversational friction between the user and the AI when conducting tests.

2.5 State of the Art

It was relevant to look upon contemporary solutions, as having LLM-powered characters is a fairly new concept, and the concept has seen a lot of development over the past few years. An early inspiration for our project was user generated content, also known as a modification or "mod", for the game Elder Scrolls V: Skyrim. This mod turned NPCs with conventional AI into LLM-based AI NPCs that are context- as well as world aware, and can even perform actions, all based on prompts created from user input in the form of speech to text [3]. This mod also allows the user to either run the LLM model in the cloud or locally on the user's PC [3], likely in an effort to make it cheaper. This is because LLM models running in the cloud restrict users' ability to send prompts to the model, generally by either setting a daily limit of tokens or by adding a paywall. Running the model locally does however come at an added performance cost. This means that having the model run locally is most likely not suitable for wireless VR headsets due to their limited onboard computing power.

Inspired by the Skyrim mod, we sought to find a way to replicate some of the demonstrated capabilities in a way that could be integrated into our VR experience, which would be developed in Unity. This is where we encountered Convai, a powerful LLM NPC API with a large suite of tools allowing for both integration in Unity and Unreal Engine, as well as easy adjustments of the LLM model parameters [15]. Conveniently, Convai already contained many of the features found in the Skyrim mod, or at least the ability to create them within its own framework.

At the time of writing, recent developments in the industry have occurred. Epic Games, a powerhouse of the video game industry, developed and released an LLM-based NPC into their flagship game Fortnite. The NPC is a representation of the villain Darth Vader from the Star Wars universe, -is capable of answering voice prompts and somewhat aware of the world and its entities. According to an article by Polygon [53], the base model of the LLM is Gemini 2.0 Flash and utilizes a voice clone of the original Darth Vader voice actor, James Earl Jones. We believe this should be considered a sign of which direction the industry is headed, in regards to LLM-based NPCs being more prevalent in video games.

2.6 Research Question

Based on the points from the analysis, we decided on the following research question:

How does adding realism, specifically; Visual Fidelity, facial expressions, animation, Embodiment, and non-verbal cues, to LLM-based Virtual Humans influence Social Presence, and Uncanniness for users in VR?

3 METHODS

3.1 Design and Prototype implementation

Based on our research we would start the development of a VR prototype with the intent of testing our hypotheses.

3.1.1 Use-Case

To test our hypotheses we decided on creating a scenario where a non-player character (NPC) guides the user around the ME-Lab at AAU Copenhagen, providing information about the research, staff, and experiments in the lab itself, but also the tools, gadgets, and objects within.
To provide the character with LLM functionalities, we chose the Convai API to power our NPC as it seemingly offered the best and most expansive solution for implementing LLM models into Unity.

3.1.2 Visuals

Designing the NPC was a task that covered multiple topics, such as aesthetics and presented personality, but also the underlying artificial intelligence and all its aspects.
Our goal was to design the optimal NPC guide for the digital tour we were creating. To do this, we started with the NPC's visual design.

Our NPC's visual design needed to fulfill three requirements. First, it should manifest a high feeling of Social Presence to fit in line with our research. Secondly, the visual design of the NCP should not elicit uncanniness. Thirdly, to ensure that people trust that the NPC presents them with correct information about the ME-Lab.
To create a more present feeling NPC, we needed for it to have a certain degree of visual fidelity as visually realistic NPCs create Social Presence. Since our game had to function in VR we had to strike a balance between realism and system performance. We therefore chose to go with a character from Reallusion's Actorcore group, which features photo-scanned characters with optimized topology, meaning that the character models are based on real 3D scans of human beings, while still optimized enough to be able to run on standalone VR headsets. The Actorcore model also worked well with our internal animation pipeline, as well as another technology we made use of, namely "Simple Automated Lip Sync Approximation" or SALSA for short.
As mentioned before, we also wanted the NPC to seem trustworthy. To visually communicate this, two factors are important: being attractive and well-dressed. Due to time and budget constraints we did not have the resources to conduct a test to properly choose a character design to fit these requirements. We therefore had some internal deliberations about what we thought would fit those three categories, while also looking like they would work in the ME-Lab. After some deliberation, we went with the character seen in **Figure 1**.

Figure 1: Our NPC, Emilio, standing in the middle of the ME-Lab explaining research to a user.

3.1.3 Animation

The animations for the character were made in Reallusion's animation tool Iclone, using a traditional keyframing method. The aim of the animations were to keep them as realistic as possible and not to keep them over-exaggerated. The animations were also set up to easily interpolate between each other, having a shared idle, as well as start- and end frames.

3.1.4 Lip sync

Considering [31] showed that partial face animation is worse than none and animation of the lower face is more uncanny than animation of the upper face, therefore, it is important that lip synchronization is implemented properly. ConvAI's included lipsync, visemes, and facial animation tool seemed more experimental but less customizable and -established than SALSA's lip sync and visemes. Which is why SALSA was chosen as the primary facial animation tool over what was included in the ConvAI SDK.

3.1.5 Facial Expression

SALSA also includes Emote customization, which is partially randomized facial expressions.

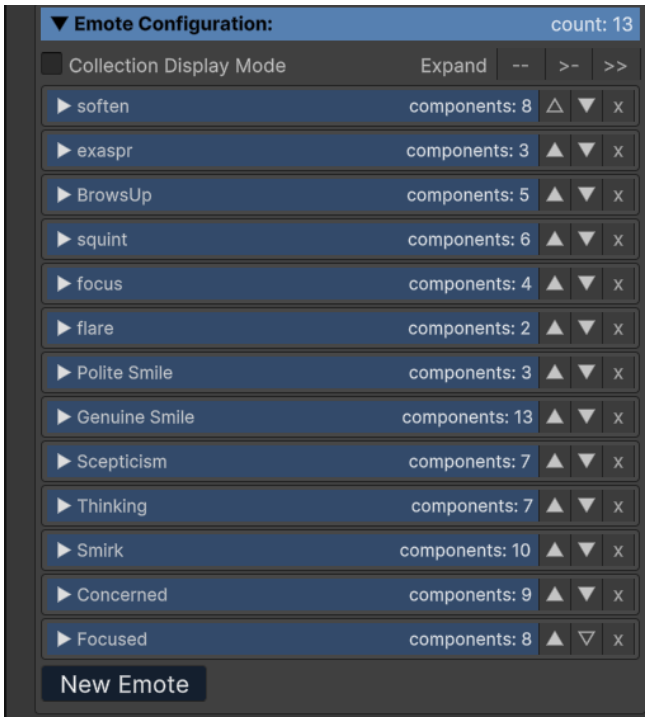


Figure 2: List of facial Emotes in Condition A of which; soften, BrowsUp, squint focus and flare we also included in condition B.

This was beneficial considering full facial animation means the higher perceived attractiveness and -positivity, but inconveniently also higher uncanniness [31].

In conclusion, either full or zero facial animation should be applied, considering partial animation is the most uncanny, least realistic, and more negatively perceived. Coherence was of importance, since implementing both also increases uncanniness. which is why emote detail was as important as visemes were for the prototype. Regarding coherence, since more dynamic facial animations increase perception of virtual humans positively [31], and emotions are more easily interpreted when exaggerated [29]. This meant that, in contrast to our approach in animation, we were inclined to strive for emotional intelligibility and expressiveness when designing Emotes. However, feedback from our internal pilot testing at Virsabi quickly revealed that they were overexaggerated and elicited uncanniness. Therefore, emotes were adjusted accordingly to a more moderate degree, but still trying to communicate probable emotional expression.

3.1.6 Behaviour

After having settled on the visuals of the NPC named Emilio, we had a somewhat more complicated task to complete: developing his behavior. As mentioned in **??**, the reason for this being more complex, is that some aspects of what makes a character socially present also make them feel less trustworthy, like, being dishonest, acting mean, complex/varied behavior, and high behavioral realism all contribute to Social Presence.

We chose to present the NPC as well-mannered, truthful, and helpful. While that should make the NPC feel less present, the alternative would be counter-intuitive to the use case and feel out of place in the overall experience and purpose of the NPC.

How this ”personality” was implemented is deeply entangled with the AI aspect of the NPC, and as such, will be covered in the following section.

Finally, we opted to involve some behavioral realism by having the NPC automatically turn towards the user and maintain eye contact while conversing.

3.2 AI Design

3.2.1 Model

The Convai API has a functionality that allows the base model to be selected from a range of providers: Claude, Google, OpenAI, Llama, and others. We chose to use OpenAI’s GPT-4o model.

Convai also has a content moderation filter, which can be turned off for some of the models. This was not an option for OpenAI’s GPT-4o model, but it did not affect us since we wanted to keep the experience free of any inappropriate content.

The last parameter that can be altered to adjust the base model is the ”temperature”, which essentially dictates how closely the AI should stick to its knowledge base or how imaginative it should be, on a scale from zero to one, where zero is extremely limited and one is extremely imaginative. For this parameter, we chose a value of 0.7 as we wanted the AI to prioritize context relevance, making conversations feel more realistic. Taking this choice meant that we made the model more prone to hallucinations, therefore, we had to implement some form of guardrails, which will be expanded on later in this section.

3.2.2 Knowledge Base

The AI’s knowledge base is divided into three parts: A backstory or ”Character Description”, information about the ME-Lab or a ”knowledge bank”, and virtual environment knowledge. The character description tool also works in tandem with another system called ”Personality”, which adjusts how the AI uses its knowledge base to respond to queries, simulating emotions in its verbal responses. How we set up ours can be seen in **Figure 3**.

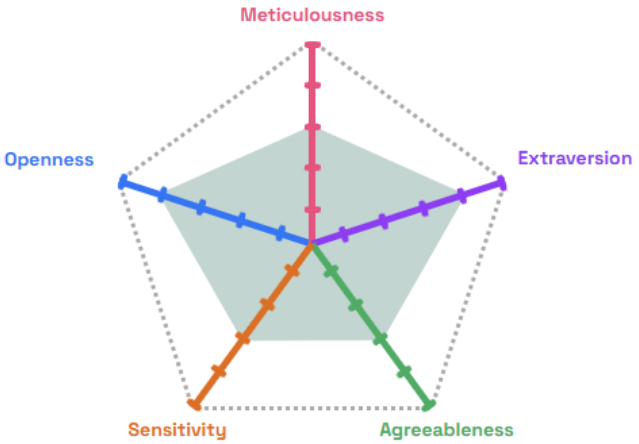


Figure 3: The AI’s Personality Style configuration

As for the more generalized knowledge base, we provided the AI with domain-specific papers about general research conducted in the ME-Lab, as well as information specific to objects and locations, such as information about the lab's anechoic chamber. We also provided it with a list of the staff working in the lab.
And as seen in **Figure 4** the virtual environment knowledge consisted of the name of an object or location, simple variations of that name, a brief description of the object, and a location transform in the virtual environment.

Figure 4: Description of the anechoic chamber as seen in Unity.

3.2.3 Capabilities

Because this AI exists in a virtual environment and has to function as a tour guide, It is required to exhibit behavior and exert actions to reflect that role. To facilitate this, we made sure that the AI had the following capabilities:

- The AI is able to understand users through text-to-speech and reply using a high-quality voice model.
- The AI can identify and navigate to specific marked locations in the lab based on varying levels of verbal descriptions provided by the user.
- The AI knows what objects users are holding in their hands and can provide descriptions and conversations about them.
- The AI can animate based partially on the "emotions" that the AI model produces.
- The AI has "memory" of what has happened in a single session, for example, it can keep track of how much of the lab the user has explored or which objects have been previously picked up.
- Because of the high temperature of the model and its extensive knowledge of the ME-lab, the AI can relate Me-Lab matters to topics relevant to the guidee.

Some of these capabilities came prepackaged with Convai, while others had to be developed. The feature that was the most complex to develop was having the character animate based on emotion.
Convai's implementation sends a list of emotions "felt" by the character to Unity, based on user input and the personality parameters of the LLM model. The possible list of emotions that the AI can "feel" is quite long, so we grouped the emotions into seven categories. The group with the most matching emotions in the list would then correspond to a number used as a blend parameter for our animation setup. Furthermore, to add animation variety, an additional random value was added to the blend parameter. Finally, it is important to mention that, as can be seen in **Figure 5**, some of the emotions that the AI could experience were quite extreme and would likely never be relevant in our use case, such as fear or despair. We therefore did not create animations for all emotion pools.

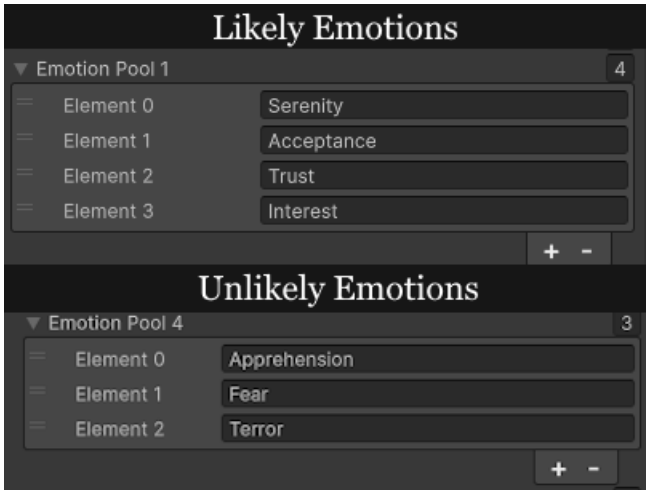


Figure 5: This figure shows two out of seven emotion pools, where one contains emotions likely to be experienced by the AI, and the other pool contains emotions unlikely to be experienced by the AI

3.2.4 Tour

We also needed to design how the AI would handle the aforementioned tour. Our first iteration used Convai's built-in narrative design structure, as seen in **Figure 6**. This structure allowed the designer to set up trigger zones in the environment and utilize them to deliver the AI precise information about where it currently was in the narrative, which in turn could be used to give the AI direct instructions on what to say or do next. This structure turned out to be too rigid and restrictive, while it was also prone to breaking out of the flow. We therefore decided to scale down the use of narrative design to only be used as an introduction to the experience, as an effort to try and eliminate action paralysis by giving users context and options to choose from.
Instead, we focused on creating an extensive knowledge bank, a descriptive backstory, and fine-tuning location descriptions and positions. This allowed for a more flexible tour of the ME-lab, where the user could choose in which order and how much they wanted to see and experience. While this approach was less prone to breaking

the experience, it did present other challenges due to its nature of being more open-ended. One such challenge was the AI's tendency to hallucinate explanations if it did not have the information ready in its database. A good way to fix this was a combination of two simple solutions. The first solution was to implement in the AI's backstory that it should act as a person who did not answer questions without proper evidence, as well as not stray to subject matters unrelated to the ME-Lab. Another challenge would be that the tour would be presented mostly on the initiative of the user and their own curiosity, possibly resulting in awkward first interactions instead of having the NPC dictate the direction of the tour.

> *"Emilio is very dedicated to his work and therefore does not talk about anything unrelated to the "Me-Lab". Emilio is academic and therefore does not make uneducated guesses on practical matters. He only answers a question when he knows the answer from his knowledge base."*
> **- Snippet from Emilio's character description**

Secondly, to limit hallucinations, we simply increased the amount of information available in the AI's knowledge base, as well as assigning common keywords and synonyms that users would use to describe objects and locations in the environment.

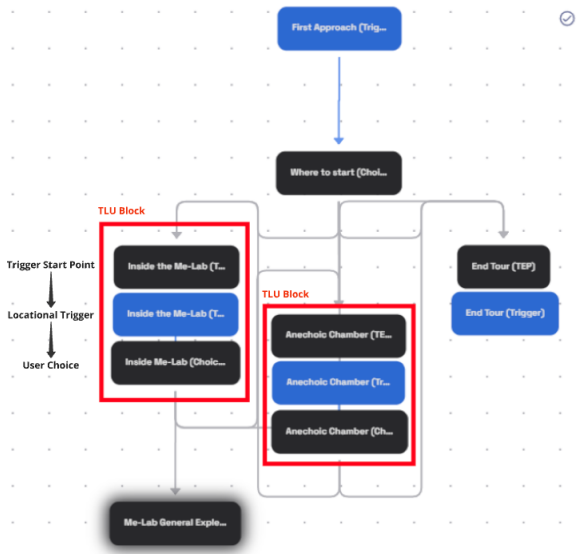


Figure 6: An initial attempt at using Convai's Narrative Design Structure

3.3 Virtual environment

As can be seen in **Figure 7**, the environment was designed to be a realistic and accurate representation of the laboratory ME-lab at AAU CPH, featuring many recognizable features of the lab that Emilio could explain and describe for users. The lab space also features multiple interactable objects for users to interact with, such as the ones seen in **Figure 8**.

Figure 7: An overview of the virtual ME-lab with an insert view of the anechoic chamber.

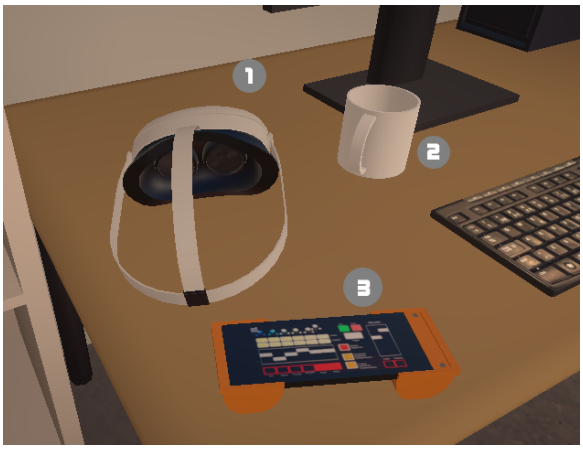


Figure 8: Objects 1 through 3 shown in the figure are some of the objects that can be interacted with in the virtual lab.

3.4 Experimental Design

In this section the approach for designing and conducting the tests is addressed and the tools for evaluating the research question and hypotheses are presented.

3.4.1 Measures

To investigate our hypotheses and answer our research question, a between-subjects experiment was designed with one independent variable: Animation variety. In condition A, the character has more displayable animations; 3 idle animations, 6 talking animations, 13 facial emotes, and the visemes have been adjusted for more visual acuity and expression, whereas in condition B the character only has 1 idle, 1 talking and 5 facial emotes and the visemes are more rudimentary.

It was decided that a minimum of 32 participants per condition group, resulting in 64 total participants, were required to meet the Central Limit Theorem 'threshold' for normality, allowing for t-tests, ANOVA etc. and reaching the statistical power desired for detecting medium effects[14].

To avoid subjecting our test participants to excessive cognitive load and fatigue, we only used select items, modules and sub-scales

from validated frameworks rather than the full questionnaires. For measuring Social Presence, parts of three validated scales were used, and two additional ad-hoc items. Eight items from the subscale for Co-presesence from[10], four items from [33], and two items from [37]. For measuring Uncanniness eeriness, five items from [24] were used. On top of these, we also added four open-ended questions for qualitative data to explore user sentiment in more detail. For providing the quantitative data for evaluating hypotheses 7-point Likert-type items were used instead of 6-point, because we wanted to give participants the option to rate indifferently and not try to skew our data towards more polar ratings.

3.4.2 Procedure and Apparatus

The VR playthrough of the experiment is a single-player activity where presence and immersion are essential, meaning participants were intended to experience it individually and undisturbed from external sources. For this reason, relatively remote and quiet locations at AAU CPH were selected to host the tests, and virtual playspaces on Oculus Quest 2's were set up so that the users could move freely in a relatively large space. Each of the participants were given 10 minutes to experience the prototype, unless cyber-sickness or discomfort inclined the user to quit the experiment ahead of time.

Figure 9: Two participants in separate testing locations, with ample room to move around in the VR playspace. At least one test conductor, as well as a PC with the questionnaire, were present during the whole procedure.

Based on our previously conducted pilot tests, the average duration of the entire experiment was estimated at 14 minutes per participant, which turned out to be a convincing factor when persuading people to voluntarily participate in our experiment.

3.4.3 Hypotheses

From our research question 2.6 and the findings from related work the following hypotheses were formulated:

- **H1 (Alternative Hypothesis):** Participants exposed to increased animation variety from VR LLM NPCs will report significantly higher Social Presence than participants exposed to less animation variety.
- **$H1_0$ (Null Hypothesis):** There will be no significant difference in reported Social Presence between participants exposed to high- and low animation variety VR LLM NPCs.
- **H2 (Alternative Hypothesis):** Participants exposed to increased animation variety from VR LLM NPCs will report significantly lower uncanniness ratings than those exposed to less animation variety.
- **$H2_0$ (Null Hypothesis):** There will be no significant difference in uncanniness ratings between high- and low animation variety VR LLM NPCs.

4 Results

In this section, data from the questionnaires, the processing, and the results thereof are presented. 65 total tests were conducted, but 1 result (participant 47) was discarded due to a glitched avatar position, resulting in the participant experiencing the environment and character at an unnatural height and distance. This was not discovered until immediately after the participant had already provided data, which is why an additional test was conducted, in order to reach our goal of 64 participants.

4.1 Quantitative results

Descriptive statistics were calculated, followed by a frequency analysis to visualize the response distribution of the provided ratings.

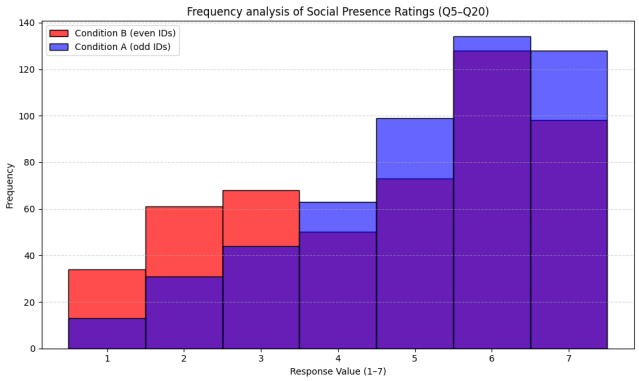


Figure 10: Frequency Analysis of Social Presence Ratings for condition A(Blue) and B(Red)

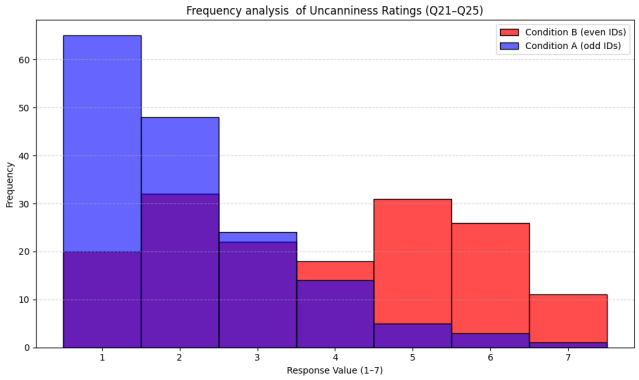


Figure 11: Frequency Analysis of Uncanniness Ratings for condition A(Blue) and B(Red)

Condition A Social: Mean = 5.18, Std = 1.63
Condition B Social: Mean = 4.65, Std = 1.91
Condition A Uncanny: Mean = 2.12, Std = 1.28
Condition B Uncanny: Mean = 3.81, Std = 1.87

As expected the Likert-type ratings were not normally distributed, therefore Mann–Whitney U tests were applied to evaluate significant differences:

Mann–Whitney U Test results for Social Presence:
Two-tailed: U = 150605.000, p = 0.0000263
One-tailed: U = 150605.000, p = 0.0000131

Mann–Whitney U Test results for Uncanniness:
Two-tailed: U = 6152.000, p = 0.000000000000000238
One-tailed: U = 6152.000, p = 0.000000000000000119

From a frequency analysis perspective, the two null hypotheses can be rejected (3.4.3 and 3.4.3).

However, individual participants' mean scores were expected to be normally distributed for both measures and conditions, therefore the distributions of means were analyzed for normality using Shapiro-Wilks tests:. Social Presence A: Shapiro-Wilk p-value = 0.375 → Normally distributed Uncanniness A: Shapiro-Wilk p-value = 0.050 → Normally distributed Social Presence B: Shapiro-Wilk p-value = 0.305 → Normally distributed Uncanniness B: Shapiro-Wilk p-value = 0.246 → Normally distribute

These means were calculated along with standard deviations and ranges.

Condition A social: Mean = 5.18, Std = 0.89, Range = 3.31
Condition B social: Mean = 4.65, Std = 0.98, Range = 3.44
Condition A uncanny: Mean = 2.12, Std = 0.81, Range = 3.40
Condition B uncanny: Mean = 3.81, Std = 1.37, Range = 5.00

A box plot for each condition and measurement was generated for visualization.

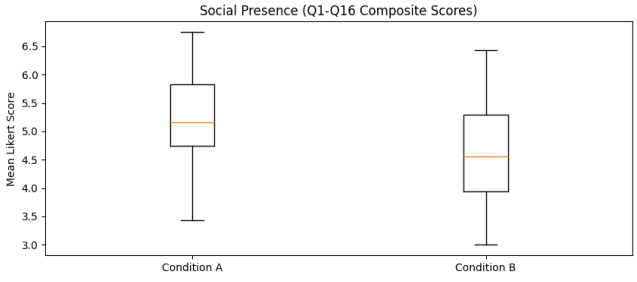


Figure 12: Social Presence Box-Plot for condition A(Left) and B(Right)

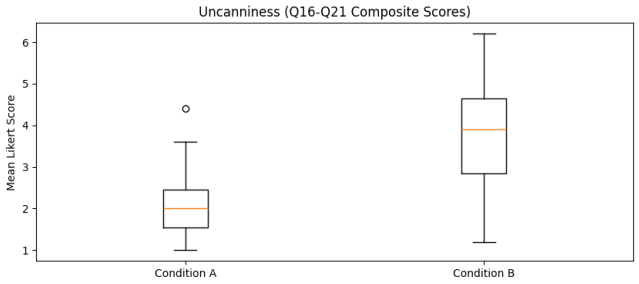


Figure 13: Uncanniness Box-Plot for condition A(Left) and B(Right)

For both conditions means of social presence and uncanniness were normal distributions, which led to a Levene's test for variance and homogeneity. Homogeneity of variance for social (p=0.442) Homogeneity of variance for uncanny (p=0.002)

The Social presence means were Homogeneous, therefore a Student's t-test was applied: Student's t-test for Social: t = 2.260, p = 0.027 p was lower than 0.05 meaning the null hypothesis 3.4.3 was rejected. The Uncanniness means however was not within acceptable range of homogeneity and variance, meaning a Welch's T test had to be applied. Welch's t-test for Uncanniness: t = 5.938, p = 0.0000003 p was lower than 0.05 meaning the null hypothesis 3.4.3 was rejected.

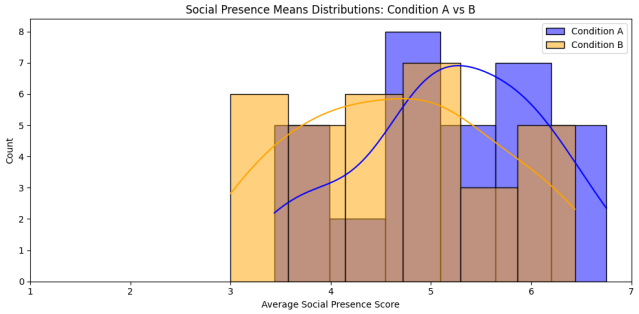


Figure 14: Social Presence Means Distributions for condition A(Blue) and B(Yellow)

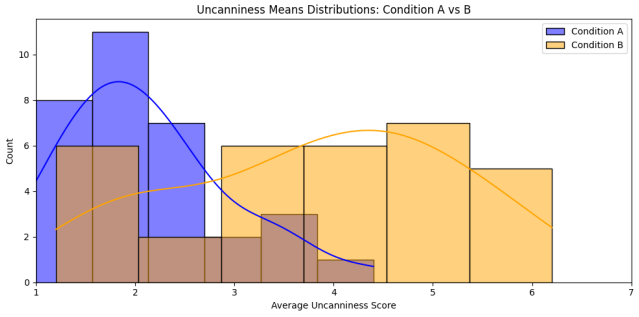


Figure 15: Uncanniness Means Distributions for condition A(Blue) and B(Yellow)

Cohens' d was used to measure effect size between the conditions with the followings results: Cohen's d (social presence): 0.57 Cohen's d (uncanniness): 1.48

4.2 Qualitative results

The figures in this subsection are made based on the coded categories of the qualitative responses from the participants in the user test. The graphs show the distribution of these codes in the individual questions across testing conditions.

These are sorted by most- to least frequent positive, followed by most- to least frequent negative code. Positives are always green, teal and blue, and negatives are always red, yellow, magenta, brown and Orange.
The qualitative answers acquired from the test indicate a popular skew in multiple factors towards condition A could be interpreted as A being interpreted as perceived the most ”humanlike” of the two.

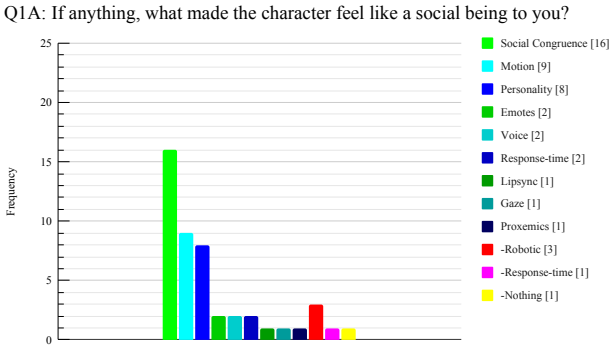


Figure 16: Many users would describe the social congruence of condition A in a positive way, followed by motion and personality, and only few would remark it negatively as robotic.

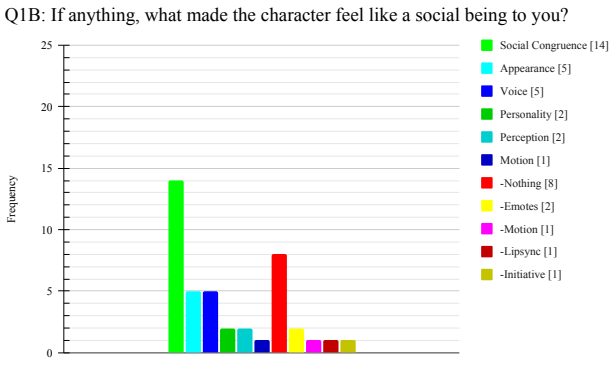


Figure 17: Social congruence, was also rated highly for condition B, followed by visuals and voice, though many would refrain from naming anything that made the NPC feel like a social being. There were also observed far more negative responses than in **figure 16**.

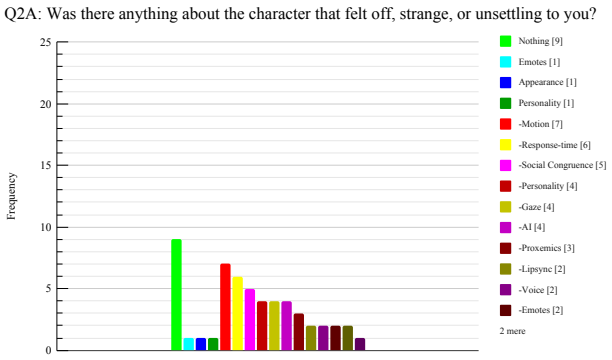


Figure 18: It is observed that the largest portion of the users in condition A have nothing of note to add on what felt strange, though a large portion reported that the motion and response-time of the character made him feel off.

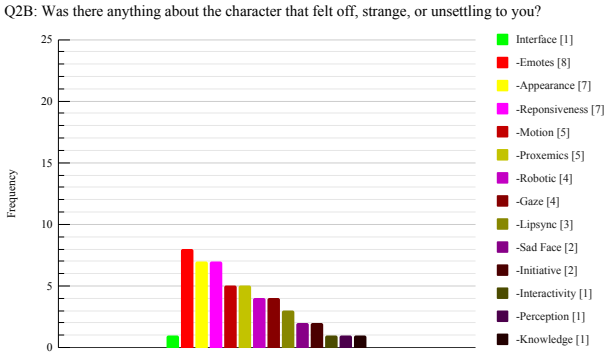


Figure 19: In the answers it was clear that unlike **figure 18**, users felt that much of the visual representation, along with the responsiveness, made the character feel off.

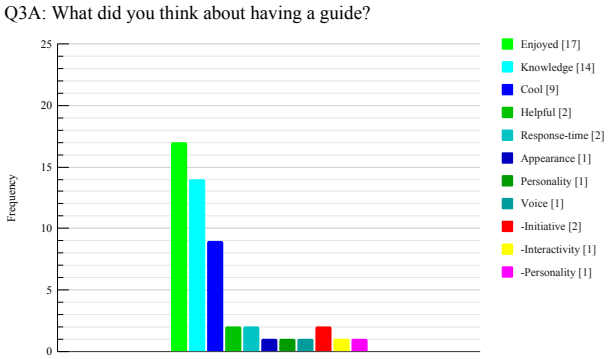


Figure 20: The answers indicated a clear positive reaction to users having a guide show them around the premises of the lab, with only a few commenting on a lack of initiative on behalf of the guide.

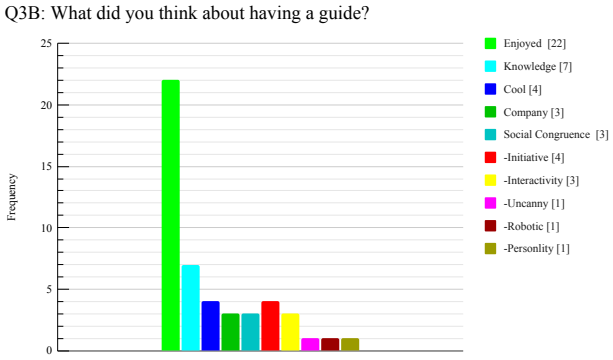


Figure 21: The answers indicate a positive reaction to the guide, again showing that users wishing for more initiative and interactivity.

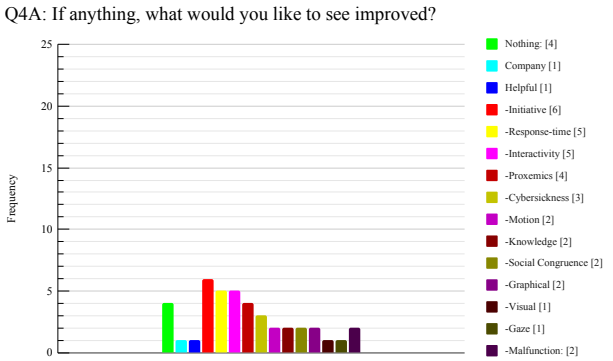


Figure 22: When asked about what they wanted to see improved, a sizeable amount of users noted that the NPC could show more initiative and response-time could be better. Though it is worth remarking that some users had no improvements they would like to see.

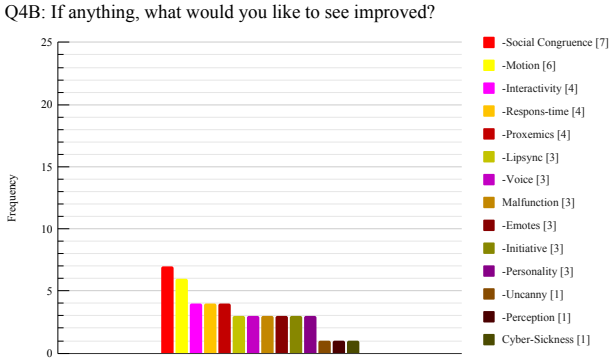


Figure 23: In comparison to **figure 22** it is remarkable that no users were completely content with the NPC and several thought congruence as well as motions were in need of an improvement.

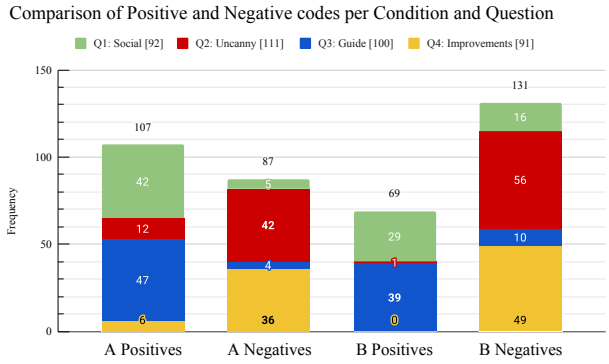


Figure 24: Comparing the amount of positive and negative codes of the qualitative answers showed a clear trend that the B condition was described with a higher amount of negative terms than condition A, both in sheer amount of statements, but also in relation to the ratio of statements.

5 Discussion

5.1 Data findings

H1: Participants exposed to increased animation variety from VR LLM NPCs will report significantly higher Social Presence than participants exposed to less animation

H1 Answer: Was supported with statistical significance by both parametric and non-parametric tests. Both total ratings and personal mean scores between conditions were statistically different, indicating that increased animation variety also increases how socially present a VR LLM NPC is perceived.

H2 Participants exposed to increased animation variety from VR LLM NPCs will report significantly lower uncanniness ratings than those exposed to less animation variety..

H2 Answer: Was supported with statistical significance by both parametric and non-parametric tests, indicating that increased animation variety lowers perceived uncanniness. This is interesting and contradictory to findings of [43], which indicate that increased realism increases uncanniness, and animation is behavioral realism according to [6], who, interestingly, were unable to detect a significant correlation between behavioral realism and social presence of VR NPCs.

5.2 Discussing the qualitative data

The qualitative data gave a clear insight into which facets of the NPC that were interpreted as having the most impact on the users. The answers collectively show that condition A was described more positively than condition B across all the questions, which ties in well with the results from the social presence and uncanniness tests, further supporting the claim that animation variety strengthens the two concepts. It is also relevant to observe that social congruence was one of the highest rated codes to be improved in condition B compared to A, despite the AI model being identical in both conditions. Both conditions proved to be joyful for both groups of test participants, despite the disparity in ratings between them.

5.3 Test Observations

During test procedures, we observed our test participants trying the VR experience. We did so to obtain additional qualitative data that could potentially expose certain aspects of our VR Experience we may not have anticipated, be it issues with usability or unexpected AI behavior.

- **Divided Attention**: Being able to observe unaware participants revealed that a considerable amount of time was spent not directly conversing with Emilio or looking at him. What distracted them from doing so was often exploring the lab or reading the chat log. In some cases, the use of Emilio's replies was reduced to a digital assistant similar to Siri, where only Voice or Text was perceived. This is a phenomenon that can have decreased the influence of the independent variable Animation Variety. A good way to have a metric for the time not spent looking at the character would be to include gaze-tracking data in VR. This could potentially also have given more revealing and interesting data.

- **Hallucinating due to high Temperature**: Emilio was able to briefly convince participants about functionality not present in the prototype or even physically possible. Among other things, this encompassed persuading participants that they were perceiving audio sensations differently while in the anechoic chamber, were able to turn on virtual VR-Headsets and play games, select and play music on Spotify on lab computers and test the wave field synthesizer.
  Being able to convince people about inaccurate statements is not inherently a bad thing; however, once discovered to be wrong, it might have impacted believability. The High temperature at 0.7, knowledge base, Personality Style and Character Description are all parameters that potentially have influenced this. Having a high chance of hallucinating, as well as being instructed to help participants learn and being knowledgeable could have caused these convincing yet false statements. As mentioned, this was intended, as we wanted the AI to be able to adapt to a variety of situations. This worked well in a situation where one test participant kept asking the AI to relate the tour to that test participant's study, and the AI seemingly did a satisfactory job of this. However, in some instances, the AI would hallucinate the correct answer, but then retract its answer when challenged. There was a situation where one test participant asked the AI if it knew the color of the coffee cup that the test participant was holding. The AI does not know this, but still guessed correctly. The test participant then asked if the AI was sure, to which the AI apologized for not knowing.
  Finally, there were also situations where the AI hallucinated being able to interact with the environment. This probably occurred because the AI has knowledge on how to operate the equipment in the lab. Therefore, the AI would sometimes ask if it should give a demonstration of how to use certain equipment, even though it does not have the capability to do that.

- **Communication issues**: Communicating with the AI seemingly had multiple points of friction. For example, some test participants would bring up the controller with the "talk button" to their mouth when speaking, presumably because they thought that the microphone was located there. This is wrong as it is located on the headset. While this was a non-essential problem that seemed not to interfere with the test participants' overall experience, there were some more drastic friction points. For example, the response time of the AI was varied. In most cases, it seemed that the time that the AI took to respond was just on the line of being too slow in the sense that people seemed almost to get confused or try to re-prompt him, but then he answered just in time. This was mostly prevalent in the first few interactions as the test participants would get used to the response time.
  There were also some instances where the test participants accidentally cut off the introduction that the AI was giving by trying to talk to it right before or after it started to speak. This sometimes led to action paralysis as they did not know how to naturally start the conversation. In other instances, test participants sometimes tried to correct themselves by pressing the talk button right after releasing it, leading to confusion, as that action deletes the initial prompt.
  While these problems are an indicator of usability issues, they could also affect the user's sense of presence as they might break immersion.

- **Initiative**: As alluded to in the previous paragraph, some test participants also experienced action paralysis, because the AI, for one reason or another, did not initiate the conversation. Some participants also mentioned that they wanted the AI to take initiative and portray a more active role. For example, some test participants found it confusing when the AI said a variant of the line: "Let me take you on a tour of the lab and I will explain", then only to proceed to walk to a location and stop and wait to be prompted again. Instead, the test participants seemingly wanted the AI to act more naturally by first walking to the location after having finished saying follow. Some test participants also wanted the AI to automatically speak to them when it had asked them to follow it to a location. Finally, some participants also wanted the AI to follow them around. These factors could indicate that if the character acts contrary to social behavioral and conversational norms, people tend to experience action paralysis. However, it also indicates that if these requirements are fulfilled, the chance of action paralysis occurring could be decreased.

5.4 Assessing the Research Question

Results and observations combined with findings from prior work provided the following answers to the research question 2.6:

> *How does adding realism, specifically; Visual Fidelity, facial expressions, animation, Embodiment, and non-*

verbal cues, to LLM-based Virtual Humans influence Social Presence, and Uncanniness for users in VR?

Visual fidelity, behavioral realism and animation quality and variety significantly increases the perceived social presence and lowers the uncanny valley effect.

5.5 Future Work

This section discusses the shortcomings of our research and how it could be expanded upon in the future.

5.5.1 Changes to the experiment

The attention of the user versus the attention put on the virtual environment. It would have been relevant to not just conduct a test based on subjective results, but also to implement objective results, such as a gaze-tracker, to actively measure how much attention was actually given to the character while playing the prototype. This would be relevant, since there's no clear metric in our experiment to indicate just how attentive our test participants were of our NPC during the playthrough of the experience.

5.5.2 Usability Testing

While we did conduct small, unofficial, and unstructured user tests during development as well as consult industry professionals, the goal of our research was not usability, so we never conducted a proper System Usability Scale (SUS) type test. However, it could be useful to allocate some resources to further refine the usability of our VR experience, before conducting more research into LLM-based NPCs. This would serve to make it a smoother experience for the users, which would in turn make it easier to conduct tests, but it would also be useful information for our external industry collaborator Virsabi, if they ever decide to make a product based on the VR experience we created.

5.5.3 Voice and Gender

Our initial analysis of the concept of creating a virtual character that could make a user feel a high degree of Social Presence indicated, along with multiple other factors, that gender and voice could also affect Social Presence. Sadly we did not have the resources to also research this, however, it would be interesting to test this in relation to our current NPC's elicited Social Presence, but it could also open up to potential biases, such as: if we change the gender of the character, we also have to change the voice model, meaning we will not know if it was the gender or the voice model change that would impact a potential result. This bias could maybe be circumvented by comparing multiple voice models of the same gender, but this would significantly increase research time, as that would be an entire extra test that would have to be conducted. We would also have to carefully consider two other factors between the male and female NPCs that could potentially be different if not curated correctly, namely, race and appearance, as those two factors might influence potential results.

5.5.4 Functionality

As the VR experience that we created is somewhat limited, expanding the functionalities and capabilities of the environment, user, and LLM-based NPC, could also lead to potentially interesting emergent discoveries. These new functionalities could come in the form of more and different ways for the user to interact with the environment, such as letting a user turn on the Razvan 3000 sound system and experiencing something new within the VR headsets, or giving the NPC more influence over the world around it, allowing it to interact with the objects in the environment, this would allow it to not just hallucinate the interactivity of the laboratory and thereby giving it more agency and letting it take on a more active role in the guided tour than what it currently has.

How does adding realism, specifically; Visual Fi- delity, facial expressions, animation, Embodiment, and non-verbal cues, to LLM-based Virtual Hu- mans influence Social Presence, and Uncanniness for users in VR?

6 Conclusion

In this study we set out to discover how adding realism, specifically; Visual Fidelity, facial expressions, animation, Embodiment, and non-verbal cues, to LLM-based Virtual Humans influence Social Presence, and Uncanniness for users in VR. We found that there is statistically significant evidence supporting that increased animation variety in VR LLM NPCs increases social presence. Furthermore, we found statistically significant evidence that more animation variety decreases perceived uncanniness. This is contradictory to findings of [43], which indicate that increased realism increases uncanniness, and animation is behavioral realism according to [6], who were unable to detect a significant correlation between behavioral realism and Social Presence of VR NPCs.
However, our hypotheses are further backed up by our qualitative data, which indicates that when a character lacks animation variety, test participants tend to refrain from naming any concrete features that made the character feel like a social being, even when explicitly asked to. Additionally, test participants would also start to find the character's visuals and emotes far more strange or unsettling than in the condition where the character had high animation variety. Finally a higher rate of overall positively coded statements and a smaller rate of overall negatively coded statements in our higher Animation Variety condition compared to the low Animation Variety condition were also observed in our qualitative data.

While our research was limited in some regards due to time constraints, in factors such as not having conducted a system usability test of the software we produced, or only having tested on a male character. This paper set out to look into a small subset of the effects that can cause an LLM-based character in a virtual world to feel more realistic and socially present, and in doing so, made discoveries that nudge our collective understanding of the relatively new field of LLM-based Characters/NPCs forward. Nevertheless, due to the novelty of this field, much more research has to be conducted, to truly grasp the opportunities and pitfalls that it presents.

Supplemental Materials

Acknowledgments

We wish to thank Ali Adjorlu for supervision, Erik Sikström for external supervision, Dina Ceraja Elmazi for providing assistance on creating the virtual environment and the Virsabi Crew for external collaboration and all their technical assistance during the implementation of the project.

References

[1] N. Amadou, K. I. Haque, and Z. Yumak. Effect of appearance and animation realism on the perception of emotionally expressive virtual humans. In *Proceedings of the 23rd ACM international conference on intelligent virtual agents*, pp. 1–8, 2023. 4

[2] N. L. B. Andreas Jin Sum, Rune Korsgaard Ludvigsen. Navigating ai in real-time for reactive and immersive video-game experiences. Unpublished student report, Department of Architecture, Design, and Media Technology, Aalborg University Copenhagen, 2024. 2

[3] ArtFromTheMachine. Mantella - bring npcs to life with ai, August 2023. 5

[4] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments*, 10(6):583–598, 2001. 3

[5] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis. Interpersonal distance in immersive virtual environments. *Personality and social psychology bulletin*, 29(7):819–833, 2003. 4

[6] J. N. Bailenson, K. Swinth, C. Hoyt, S. Persky, A. Dimov, and J. Blascovich. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence*, 14(4):379–393, 2005. 3, 4, 13, 15

[7] J. Bates, A. B. Loyall, and W. S. Reilly. Integrating reactivity, goals, and emotion in a broad agent. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 14, 1992. 2

[8] Y. Benkhoui. Spotlight: Convai reinvents non-playable character interactions, January 2024. 2

[9] E. Bevacqua, R. Richard, and P. De Loor. Believability and co-presence in human-virtual character interaction. *IEEE computer graphics and applications*, 37(4):17–29, 2017. 4

[10] F. Biocca, C. Harms, and J. K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, 12(5):456–480, 2003. 4, 10

[11] W. Bledsoe. I had a dream: Aaai presidential address. *AI Magazine*, 7(1):57–57, 1986. 2

[12] G. Borshukov and J. P. Lewis. Realistic human face rendering for" the matrix reloaded". In *ACM Siggraph 2005 Courses*, pp. 13–es. 2005. 3

[13] K. Candon, Z. Hsu, Y. Kim, J. Chen, N. Tsoi, and M. Vázquez. Perceptions of the helpfulness of unexpected agent assistance. In *Proceedings of the 10th International Conference on Human-Agent Interaction*, pp. 41–50, 2022. 2

[14] J. Cohen. A power primer. 2016. 9

[15] Convai. Introducing convai, August 2023. 5

[16] C. Ennis, R. McDonnell, and C. O'Sullivan. Seeing is believing: body motion dominates in multisensory conversations. *ACM Transactions on Graphics (TOG)*, 29(4):1–9, 2010. 5

[17] Y. Ferstl, M. Neff, and R. McDonnell. Understanding the predictability of gesture parameters from speech and their perceptual importance. In *Proceedings of the 20th ACM international conference on intelligent virtual agents*, pp. 1–8, 2020. 5

[18] Y. Ferstl, S. Thomas, C. Guiard, C. Ennis, and R. McDonnell. Human or robot? investigating voice, appearance and gesture motion realism of conversational social agents. In *Proceedings of the 21st ACM international conference on intelligent virtual agents*, pp. 76–83, 2021. 4

[19] Y. Gao, Y. Dai, G. Zhang, H. Guo, F. Mostajeran, B. Zheng, and T. Yu. Trust in virtual agents: Exploring the role of stylization and voice. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 4

[20] M. Garau. *The impact of avatar fidelity on social interaction in virtual environments*. University of London, University College London (United Kingdom), 2003. 5

[21] M. Gerhard, D. Moore, and D. Hobbs. Close encounters of the virtual kind: Agents simulating copresence. *Applied Artificial Intelligence*, 19(3-4):393–412, 2005. 2, 3

[22] E. Grassucci, G. Grassucci, A. Uncini, and D. Comminiello. Beyond answers: How llms can pursue strategic thinking in education. *arXiv preprint arXiv:2504.04815*, 2025. 3

[23] M. Guimarães, R. Prada, P. A. Santos, J. Dias, A. Jhala, and S. Mascarenhas. The impact of virtual reality in the social presence of a virtual agent. In *Proceedings of the 20th ACM international conference on intelligent virtual agents*, pp. 1–8, 2020. 3

[24] C.-C. Ho and K. F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, 2010. 10

[25] C.-C. Ho and K. F. MacDorman. Measuring the uncanny valley effect: Refinements to indices for perceived humanness, attractiveness, and eeriness, 2017. 4

[26] Y. Hu, P. Huang, M. Sivapurapu, and J. Zhang. Elegnt: Expressive and functional movement design for non-anthropomorphic robot, 2025. 4

[27] L. Huang, W. Yu, W. Ma, W. Zhong, Z. Feng, H. Wang, Q. Chen, W. Peng, X. Feng, B. Qin, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, 2025. 2

[28] S. S. Iyengar and M. R. Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995, 2000. 5

[29] T. Johansson and M. P. Eladhari. Emotional believability of non-playable game characters-animations of anger, sadness and happiness. In *International Conference on Interactive Digital Storytelling*, pp. 72–99. Springer, 2024. 7

[30] S. Kang, A. Nguyen, B. Yoon, K. Kim, and W. Woo. Gender differences in perceiving avatar face and interpersonal distance: Exploring realism and social presence in mixed reality. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 101–110. IEEE, 2024. 5

[31] P. Kullmann, T. Schell, T. Menzel, M. Botsch, and M. E. Latoschik. Coverage of facial expressions and its effects on avatar embodiment, self-identification, and uncanniness. *IEEE Transactions on Visualization and Computer Graphics*, 2025. 6, 7

[32] K. H. C. Lau, B. Yun, S. Saruba, E. Bozkir, and E. Kasneci. Wrapped in anansi's web: Unweaving the impacts of generative-ai personalization and vr immersion in oral storytelling. *arXiv preprint arXiv:2409.16894*, 2024. 2, 3, 4

[33] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff. A cross-media presence questionnaire: The itc-sense of presence inventory. *Presence: Teleoperators & Virtual Environments*, 10(3):282–297, 2001. 10

[34] C. Li, J. Wang, Y. Zhang, K. Zhu, W. Hou, J. Lian, F. Luo, Q. Yang, and X. Xie. Large language models understand and can be enhanced by emotional stimuli. *arXiv preprint arXiv:2307.11760*, 2023. 2

[35] C. Li, W. Zhu, W. Xing, and R. Guo. Analyzing student attention and acceptance of conversational ai for math learning: insights from a randomized controlled trial. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pp. 836–842, 2024. 2

[36] S. Lim, R. Schmälzle, and G. Bente. Artificial social influence via human-embodied ai agent interaction in immersive virtual reality (vr): Effects of similarity-matching during health conversations. *arXiv preprint arXiv:2406.05486*, 2024. 3

[37] M. Lombard, T. B. Ditton, and L. Weinstein. Measuring presence: the temple presence inventory. In *Proceedings of the 12th annual international workshop on presence*, pp. 1–15. University College London, 2009. 10

[38] N. Magnenat-Thalmann and D. Thalmann. Virtual humans: thirty years of research, what next? *The Visual Computer*, 21:997–1015, 2005. 4

[39] D. Mal, N. Döllinger, E. Wolf, S. Wenninger, M. Botsch, C. Wienrich, and M. E. Latoschik. Am i the odd one? exploring (in) congruencies in the realism of avatars and virtual others in virtual reality. *Frontiers in Virtual Reality*, 5:1417066, 2024. 4

[40] R. McDonnell. Appealing virtual humans. In *Motion in Games: 5th International Conference, MIG 2012, Rennes, France, November 15-17, 2012. Proceedings 5*, pp. 102–111. Springer, 2012. 5

[41] Meta. Testing and performance analysis, October 2024. 3

[42] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012. 4, 5

[43] M. Mustafa, S. Guthe, J.-P. Tauscher, M. Goesele, and M. Magnor. How human am i? eeg-based evaluation of virtual characters. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 5098–5108, 2017. 3, 4, 13, 15

[44] M. Neff. Hand gesture synthesis for conversational characters. *Handbook of Human Motion*, pp. 1–12, 2016. 5

[45] Y. Nong, H.-T. Zhang, and J.-Q. Sun. Natural language processing for immersive game interactions: Improving nlp models for more natural conversations with ai-driven npcs in ar/vr games. *Authorea Preprints*, 2025. 5

[46] J. Ondřej, C. Ennis, N. A. Merriman, and C. O'sullivan. Frankenfolk: Distinctiveness and attractiveness of voice and motion. *ACM Transactions on Applied Perception (TAP)*, 13(4):1–13, 2016. 4

[47] R. B. Paradeda, G. Freitas, S. E. Alves Filho, and A. A. Souza. The influence of well-dressed npcs on player perception, immersion and decision-making in gaming. In *2023 IEEE Conference on Games (CoG)*, pp. 1–8. IEEE, 2023. 4

[48] D. Parmar, S. Ólafsson, D. Utami, P. Murali, and T. Bickmore. Navigating the combinatorics of virtual agent design space to maximize persuasion. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1010–1018, 2020. 4

[49] D. Parmar, S. Olafsson, D. Utami, P. Murali, and T. Bickmore. Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Autonomous agents and multi-agent systems*, 36(1):17, 2022. 4

[50] K. Radivojevic, M. Chou, K. Badillo-Urquiola, and P. Brenner. Human perception of llm-generated text content in social media environments. *arXiv preprint arXiv:2409.06653*, 2024. 2

[51] M. Richter. *Embodied Conversational Agent for Urban Exploration of Historical Places in Virtual Reality*. PhD thesis, Technische Universität Wien, 2024. 3, 4

[52] R. Rickenberg and B. Reeves. The effects of animated characters on anxiety, task performance, and evaluations of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 49–56, 2000. 4

[53] I. Rouse. An ai james earl jones returns as darth vader in fortnite, with permission, May 2025. 5

[54] A. P. Saygin, T. Chaminade, H. Ishiguro, J. Driver, and C. Frith. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social cognitive and affective neuroscience*, 7(4):413–422, 2012. 4

[55] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009. 2

[56] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in virtual reality*, 3:914392, 2022. 3, 4

[57] StandoutEdu. Virtual reality for accelerated career entry for young job seekers, November 2024. 2

[58] M. V. Taylor, Z. Muwaffak, M. R. Penny, B. R. Szulc, S. Brown, A. Merritt, and S. T. Hilton. Optimising digital twin laboratories with conversational ais: enhancing immersive training and simulation through virtual reality. *Digital Discovery*, 2025. 3

[59] A. Tinwell, D. A. Nabi, and J. P. Charlton. Perception of psychopathy and the uncanny valley in virtual characters. *Computers in Human Behavior*, 29(4):1617–1625, 2013. 5

[60] H. C. van Vugt, J. F. Hoorn, and E. A. Konijn. Interactive engagement with embodied agents: An empirically validated framework. *Computer Animation and Virtual Worlds*, 20(2-3):195–204, 2009. 2

[61] A. M. Von der Pütten, N. C. Krämer, J. Gratch, and S.-H. Kang. "it doesn't matter what you are!" explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):1641–1650, 2010. 4

[62] H. Wan, J. Zhang, A. A. Suria, B. Yao, D. Wang, Y. Coady, and M. Prpa. Building llm-based ai agents in social virtual reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pp. 1–7, 2024. 2

[63] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, et al. Analysis of human faces using a measurement-based skin reflectance model. *ACM Transactions on Graphics (ToG)*, 25(3):1013–1024, 2006. 3

[64] K. Zibrek and R. McDonnell. Social presence and place illusion are affected by photorealism in embodied vr. In *Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games*, pp. 1–7, 2019. 4

APPENDIX

(A) PIE-CHARTS OVER PARTICIPANTS MENTIONED CODES

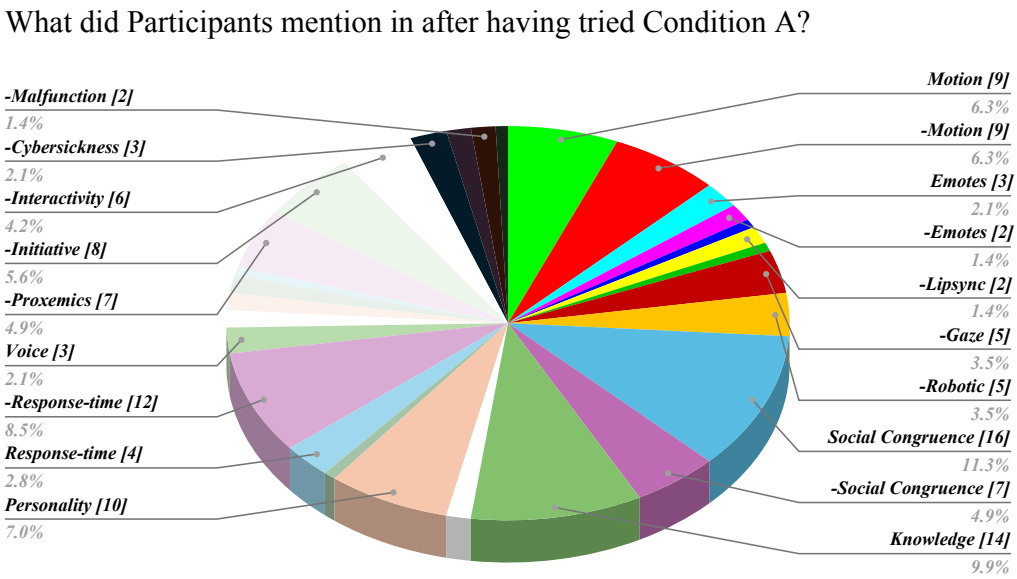


Figure 25: Pie-Chart over Mentioned Codes in Condition A.

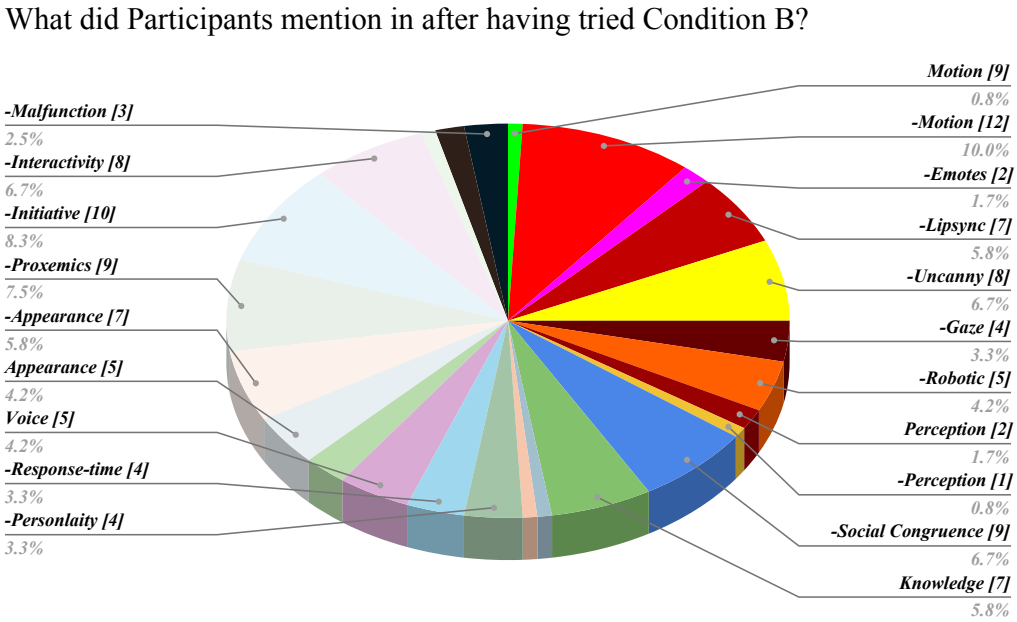


Figure 26: Pie-Chart over Mentioned Codes in Condition B.

(B) QUALITATIVE DATA AND CODING + APPLIED QUESTIONNAIRE

Condition A Responses

A_Q1: If anything, what made the character feel like a social being to you? Please describe.

Response:	Codes:
Hans "normale" udtryk, gjorde ham mere relaterbar, når han ligner en man kan se på gaden/studiet.	Emotes:
Quick response to questions	Response-time:
The pitch in his vocal was good	Voice:
actually walking when i asked him to show me around	Motion: Social Congruence:
Convincing voice and Motion	Voice: Motion:
The Motion felt fluid and the fact that it responded to my instructions pretty intuitively. It also wasn't giving me super generic responses like siri would, but like the extent of a dynamic - real interaction	Motion: Social Congruence: Personality:
Engaged in natural conversations, and understod my intentions even if I did not say it explicitly	Social Congruence: Personality:
relevant feed back too the questions	Social Congruence:
Proper responses	Social Congruence:
Responses were mostly fine, but a bit robotic, and took a while to respond	Social Congruence: -Robotic: -Response-time:
responding to his name, and responding directly to my questions - and the dancing	Personality: Response-time: Social Congruence: Motion:
He felt very much like robot, but there wasn't anything off-putting about him, almost like a robot who was nice to talk to. But it was clear for me that It was not another person I was talking to.	-Robotic: Personality: Social Congruence:
Talked about common interests. Talked to me casually instead of being unnecessarily informative. Had facial expressions and mouth Motion when he talked.	Personality: Social Congruence: Emotes Lipsync: Motion:
the clothing, natural hairstyle	Appearance:
That he kept an eye on me. Answered but also took the lead in the conversation sometimes.	Perception: Emotes: Social Congruence:
Den var god til at gå rundt, og så kunne den dance.	Proxemics: Motion:
Understanding me, all the times	Social Congruence:
When the character looked at me as if he is really addressing me, and the use of tonality.	Gaze: Voice:
Listened well and understood every question i asked	Social Congruence:
Han kunne danse	Motion:
The way he spoke in calm voice and the response seemed to clearly understand what I meant.	Voice: Social Congruence:
The way it responded according to what I said	Social Congruence:
The dancing	Motion:
That the responses fit almost exactly what I asked, they felt natural, more so than being entirely pre-programmed	Social Congruence: Personality:

I suppose it did move around like a person to some degree and used its hand when conversing, but its hard to overlook the robotic nature of it either way.	Motion: Personality: -Robotic
It didn't. It felt like a robot.	-No : -Robotic:
he could dance and talked to me	Motion: Social Congruence:
His responses	Social Congruence:

A_Q2: Was there anything about the character that felt off, strange, or unsettling to you? Please describe.

Response:	Codes:
more realistic facial expressions and body gestures.	-Emotes: -Motion:
Nope	No:
At one point he walked through me	-Motion: -Proxemics:
No	No:
his smiles,	-Emotes:
Nothing	No:
perhaps the awkward onset delay in response time, felt like I needed to fill in the gap of the silence. but other than that I was pretty satisfied with the interaction. OH and also the way he kept on leaving hahaha like I had to call him back in.	-Response-time: -Motion: -Proxemics:
Its too much "ChatGPT" for me to have a natural conversation with it. It answers too much, and does not read the room (or figure out what I want to know)	-Social Congruence: -Personality: -AI:
his voices is a bit AI like	-Voice: -AI:
That it just turned to look at me, and the mouth Motion was a bit weird	-Gaze: -lipsync:
He walked back to the starting position instead of closer to me when I told him to come to me. He kind of over explained a lot. I asked what a thing was and I got an autistic kind of fact list in addition to it. Nice, but more than I needed. The conversation also felt like a Q and A and not like a conversation. The questions of if I wanted to know more about something were good, but left it at me asking for more information, which is not a "normal" conversation. Good for the purpose of learning about the room though.	-Motion: -Social Congruence: -Personality:
Not really, it wasn't photoreal enough to register on the uncanny valley for me	No: Appearance:
the longer pauses after each question asked made the conversations less natural of course, and I felt like I could not just ask him any equation, but that it had to be relevant to why I was there and the Lab - but I didn't test this, which I could have	-Response-time: -Social Congruence: -Personality:
No i don't think so, as said before he just felt robotic, but pleasant enough.	No: -Robotic: Personality:

Some responses were a bit slow, making some awkward pauses sometimes.	-Response-time:
There was something about the eyes, and the walk and Motion were very stiff	-Gaze: -Motion:
I dont feel like, but some might feel his intensive eye-contact that way	-Gaze:
Nej synes jeg ikke	No:
He stood still a lot	-Motion:
Lack of empathy or human touch in the language which could still be improved.	-Social Congruence: -Personality:
The pauses of complete silence between questions is not very human	-Response-time:
Han tog lang tid om at svarer på spørgsmålene	-Response-time:
Only really the robotic voice	-Voice: -Robotc:
I was aware that it was an NPC but didn't feel uneasy	No:
The mouth	-Lipsync:
There was one point where the character walked straight through me so I saw inside their head model. Other than that nothing made them seem completely out of the ordinary, except for a few rigid responses	-Proxemics: -Social Congruence
Personally no, but it probably comes from seeing the character as entirely an artificial intelligence, rather than one trying to imitate a human.	No: -AI:
Though he made some funny expressions.	Emotes:
The robot-like manners, it felt like talking to an embodied AI.	-Robotic: -AI:
he just was a bit off but idk what exactly but not too bad	-Uncanny:
time lack. sometimes loss of eye contact. His skin looked sick	-Social Congruence: -Gaze: -Appearance:
His Motion and the delayed response	-Motion: -Response-time:
not really.	No:

A_Q3: What did you think about having a guide?

Response:	Codes:
God ide, men føler man skal have en ide om hvad det er for et sted eller hvad man skal være der for. For at kunne få optimalt ud af oplevelsen/rundvisningen.	Enjoyed:
great and learning experience	Enjoyed: Knowledge:
It worked well	Enjoyed:
A nice and novel experience	Enjoyed: Cool:
Very cool	Cool:
it was definitely a good visual and audio queue as opposed to text pop ups	Enjoyed: Visual: Voice:

Fine, but for me i think that having the kind of intractability made the tour kinda of boring compare to just run around on my own and explore and interact with stuff	Enjoyed: -Interactivity:
it was nice having a guide it felt like you got a tour of the real lab from someone that's knowlegable	Enjoyed: Helpful: Knowledge:
Difficult for me to answer	Pass:
It was very helpful in exploring the space	Helpful:
A good idea definitely, it gave good information about what the stuff I asked about was and it's purpose.	Enjoyed: Knowledge:
it was useful, as I wouldn't otherwise know what the lab was about and what to do there	Enjoyed: Helpful: Knowledge:
Worked very well in the context of needing to get to know what the me-lab space is and how the things there work.	Enjoyed: Knowledge:
I like to have a guide if I am in an area I do not know or if they are very informative on the current subject. In this case, it makes a lot of sense to have a guide for e.g. new students.	Enjoyed: Knowledge:
It was nice having a guide, but it was difficult to ask him questions, since it concerned an area, I know nothing about.	Enjoyed: Knowledge:
It was nice	Enjoyed:
Meget fedt når man så noget man ikke vidste hvad var	Enjoyed:
Very nice	Cool:
It was amazing! I really like this idea since it feels more personal rather than exploring it myself.	Enjoyed: Personality:
i made sense since i had no idea about the place.	Enjoyed: Knowledge:
godt	Enjoyed:
Helpful, to understand the uses of different areas.	Helpful: Knowledge:
It was nice. They were informative	Cool: Knowledge:
It was cool, but i expected him to lead me	Cool: -Initiative:
It was nice, it was cool how many extra questions you were able to ask them aside from the few key Me-Lab elements	Enjoyed: Cool: Knowledge:
It was handy, it answered all of my questions, although the answers did get a bit long winded at times.	Helpful: Knowledge: -Personality
It was good, he seemed very responsive to questions, and gave good answers.	Enjoyed: Response-time: Knowledge:
It was nice, but I think I wanted him to make more suggestions as or what to do and where to go.	Cool: -Initiative:
it was pretty cool, he could explain stuff to you	Cool: Knowledge:
Nice. worked well	Cool: Enjoyed:
It was great to be able to ask anything	Enjoyed: Knowledge:

A_Q4: If anything, what would you like to see improved?

Response:	Codes:
it's more helpful than being alone.	Helpful: Company:

Måske han kunne starte med at introducere stedet, hvis det er nogen der skal have rundvisning som ikke ved noget om rummet.	-Initiative:
Maybe if the use is uniteractive the NPC takes initiative and leads the way.	-Initiative:
Not really, but I do not have an experience to compare with	No:
det var lidt sløret, men ellers meget godt og nemt at bevæge sig	-Visual:
Nothing	No:
Maybe the delay in responses. i also dont want him to necessarily follow me around and tail me but maybe he shouldn't leave the room LMAO. Great job overall	-Response-time: -Proxemics:
The bot needs to have deeper knowledge of stuff, like a RAG system for receiving information from documents for context for its answer. det var for overfladisk information den gav	-Social Congruence: -Knowledge
The distance that the character showed was opposite sides of the room making the cyber sickness higher	-Cyber-sickness: -Proxemics:
I think it could be good if he followed me unless ordered not to	-Proxemics:
The following around. He could essentially follow when you go to another room and stand there for a little while instead of talking through the wall to you.	-Proxemics: -Perception:
It doesn't really set the user up to explore specific sections that don't already know about the lab. Maybe if it could suggest specific locations every now and then.	-Initiative:
maybe to be a part of an activity in the lab, so I didn't just visit it while it was empty - I still cant quite imagine what it looks like when experiments are taking place in the lab.	-Interactivity:
i don't know	No:
Just polish. The environment could be "warmer", more objects in the lab, more refined animations, etc. Other than that, not so much! Very cool	-Graphical: -Interactivity: -Motion:
He could have presented a short guide or there could have been signs separating areas of the lab	-Initiative:
Maybe some sidetask idk	-Interactivity:
Det var lidt sløret, men ellers meget godt og nemt at bevæge sig	-Graphical:
More eye contact	-Gaze:
Response to specific questions could be more detailed and logical.	-Knowledge:

lessen the feeling of sea sickness	-Cyber-sickness:
Tiden han tog på at svarer	-Response-time:
Maybe more examples of different room in use, like hearing sounds clip recorded inside vs outside the soundproof room.	-Interactivity:
Once when I asked him he didn't answer.	-Response-time: -Malfunction:
It was cool, but i expected him to lead me	-Initiative:
There was a few responses at the end where the responses had 2 apostrophes e.g. "I"d" our "can"t" This made the character pronounce things wrong which took you out of it a bit no major improvements though, only the possibility of asking about even more things	No: Malfunction:
Perhaps try to adjust the speech pattern of the character in real time to match that of the person in the experience and more natural and dynamic Motion transitions.	-Social Congruence: -Voice: -Motion:
He relies a lot on the player to respond to him and tell him to show you more, he doesn't take the initiative to show the room.	-Initiative:
pick more stuff up	-Interactivity:
time lack in interaction.	-Response-time:
The delay in the response	-Response-time:
i felt a little dizzy while moving	-Cyber-sickness:

Condition B Responses

B_Q1: If anything, what made the character feel like a social being to you? Please describe.

Response:	Codes:
the body and face looked like a human	Appearance:
he could respond to questions	Social Congruence:
Polite, responded not just as a `fact giver' but as an informer	Social Congruence:
Pass	-No:
When it actually followed me.	:
The likeliness to an actual human, and the fact that the responses felt real. I felt like the character could interact realistically with me.	Appearance: Social Congruence:
talking to me, responding to me.	Voice: Social Congruence:
It has a natural voice, and was walking around while talking.	Voice:
not really	-No:
Not really, it was an NPC and reminded me of the hologram from Jurassic World	-No:
His appearance looked like a social being, but his mimicry was quite off.	Appearance: :Emotes -Lipsync:-Motion:

The way it knew where everything was when I asked (for example asking where the computer was, and it taking me towards it. Also the tinnitus room)	Social Congruence:
The facial movements	Emotes: Lip-sync:
He responded to my question in a way i would expect a normal person to respond	Social Congruence:
I could hear his voice move around	Voice:
He didn't	-No:
It did but i could also see it was not a human being but something close	Visual:
nothing	-No:
His language and choice of words, his voice and his looks. It seemed like I was interacting with some very introvert person, that sticked to his formula.	Social Congruence: Voice: Appearance:
He could hear me	Perception:
responding to my questions quickly and with answers that made sense.	Respons-time: Social Congruence:
It could understand general descriptions of the environment and make assumptions about what I was talking about. It knew what I was carrying and could talk about it despite me not being able to describe it, as could it describe the Razvan Benches when I just asked about the "area near the carpet"	Social Congruence:
Nothing.	-No:
Noticed I moved on	Perception:
As I am used to interact with "NPCs" in games and ai applications I was maybe too analytical in my approach and didn't as such feel a social interaction.	-No:
It responded quickly to my questions and the answers were relevant to my questions	Respons-time: Social Congruence:
The way he answered my questions	Social Congruence:
The adequate responses to even the smallest remarks.	Social Congruence:
It was nice that you gave him a name! You could have added to that by introducing him a bit more, given him more of a personality.	Personality: -Initiative:
Not really. He doesn't really interact about anything other than the lab.	-No: -Social Congruence: Personality:
understanding my questions, and thoroughly responding	Social Congruence:
I think the characters visual appearance was very human-like. Voice and responses where also quite good.	Appearance: Voice: Social Congruence:

B_Q2: Was there anything about the character that felt off, strange, or unsettling to you? Please describe.

Response:	Codes:
the details about his face	-Appearance: -Emotes: -Lipsync:
a bit uncanny valley - and didn't react timely to my questions	-Uncanny:

	-Respons-time:
the mouth particularly and the eye brows gave off a sort of depressed mood	-Appearance: -Sad face:
It doesn't look at me. -It didn't just feel smaller but feel like something between the portions wasn't correct. -It went through me like if I would be a ghost. -It didn't wait for me to follow. -Face expressions in overall is not bad but still off. -Was weird that it couldn't use the VR while demonstrating.	-Gaze: -Appearance: -Proxemics: -Emotes: -Uncanny: -Interactivity:
The character looked a bit sad, and it did not look "full" like an actual human (Something with the depth seemed off, it looked a bit more 2D than expected)	-Appearance: -Sad face: -Uncanny:
Stone cold look. Was there any blinking?	-Appearance: -Emotes: -Gaze:
he didn't really engage towards me, it felt very one-sided and robotic	-Initiative: -Robotic:
Mainly his facial expressions which were very stiff and unaligned with what he said	-Emotes: -Lipsync:
His mimicry and the way he talked was a bit off	-Emotes: -Lipsync: -Motion:
He was VERY monotone, so it was hard to concentrate on what he was saying. Was very nice to have a text box.	-Voice: Interface:
very stiff	-Motion:
Slow answering	-Respons-time:
It could be cool if his head turned towards to you	-Gaze:
The way he was walking. He walked very unnatural, like he was on a treadmill. He also walked right through me which felt unnatural and uncomfortable. He responded very slowly to me.	-Motion: -Proxemics: -Respons-time:
the time it took to answer and it did not really have any small movements	-Respons-time: -Movements:
the way it was standing	-Movements:
Yes. His mouth movements. His very strange movements around in the room. And he was always waiting for my answer and didn't start any irrelevant conversations.	-Lip-sync: -Movements: -Proxemics: -Initiative:
Repeated the same thing multiple times, and was very robotic in its answers. It felt like he didn't really listen to what I said but just took the short and sweet of it, with only really giving exposition and not some one who is interacting with me.	-Robotic: -Social Congruence: -Knowledge:
i could get very close to the character without it reacting, making it feel awkward at times.	-Perception: -Proxemics:
The model was a bit more towards Source Engine models, likely just the way lighting was rendered on it. It did not seem entirely real, but fake enough to not be uncanny valley	-Appearance: -Uncanny:
The character walked right through me at one point. It felt very weird.	-Proxemics:
Yes, too long pauses, and the way of speaking was akin to a computer/limited AI	-Respons-time: -Social Congruence:

Not particularly, maybe facial animations could help, but without it being very good it could make things uncanny.	-Emotes: -Uncanny:
It didn't stop in between sentences to allow me to interject in the conversation	-Social Congruence:
The delay in answering and the robotic voice	-Robotic: -Motion:
Sometimes the character rotating towards me felt very unnatural.	-Motion:
I caught the character staring directly at me a few times, lacked humane mannerisms, such as blinking or small movements. The guide was very still and robotic, there was a bit of lag between my audio input and the guide's response.	-Gaze: -Emotes: -Robotic: -Respons-time:
He seem like a low graphic doll, but he matches the environment.	-Appearance:
The smile was unsettling	-Emotes: -Uncanny:
The voice was monotone, but I think it was still good compared to other algorithms out there. Also, the long waiting time in the answer highlighted the human as a "robot"	-Voice: -Respons-time: -Robotic:

B_Q3: What did you think about having a guide?

Response:	Codes:
it was better than having none	Enjoyed: Company:
a good idea to make it more personally	-Personality:
fine, came with some nice pointers and examples as well as good responses.	Enjoyed: Social Congruence:
Nice	Enjoyed:
Wasn't intuitive. I would rather have me guide around first.	-Initiative:
It was fun, felt like i learned something new about the ME Lab.	Enjoyed: Knowledge:
fun idea - I didn't feel like having much of a guide though, because I had to prompt myself.	-Initiative:
it was comfortable to have a very knowledgeable person there with me, because I have never experienced the lab myself.	Enjoyed: Knowledge:
pretty cool	Cool:
It was cool, he had good answers.	Cool: Social Congruence: Knowledge:
I did not know what i could be guided in because I don't have an idea what is in the lab	-Initiative:
It was nice! Very cool	Enjoyed: Cool:
made it more interesting than just exploring alone	Company:
Felt nice, but i was unsure of what to ask about. He needed input before he responded	Enjoyed: -Initiative:
It was cool, so you could get information	Cool: Knowledge:
It was nice having a guide but he was a bit creepy and unreal also when he talked it was a bit hard to concentrate on his words so I didn't really pay too much attention. Maybe some drawings of what he was saying could help	Enjoyed: -Uncanny:
I liked that a lot. i did not need to read anything and it felt natural	Enjoyed:
it's nice but I'd prefer a human	Enjoyed: Human

	preference:
It was nice, because otherwise I had no idea what to do.	Enjoyed:
Very robotic	-Robotic:
i think it would be nice if there were people working on the computers felt empty compared to my experience there	-Interactivity:
It was a pleasant experience, quite unlike guidebooks. It was nice being able to just ask about whatever seemed interesting in the room and get info, as was the chat log a great addition, as I at one point forgot to listen, but could still catch up	Enjoyed: Knowledge: Interface:
It was nice having someone who was able to explain different things about the lab.	Enjoyed: Co-Presnece: Knowledge:
It's a good concept and works	Enjoyed:
It was good since I wouldn't know about the soundproof room or what the setup with the surround sound was without a guide.	Enjoyed: Knowledge:
I liked it, and I could imagine using it in a different setting	Enjoyed:
it was fun	Enjoyed:
It was nice and enhanced the experience in my opinion.	Enjoyed:
It was a nice way to make the tour interactive.	Enjoyed: Interactive
It seems like a good idea.	Enjoyed:
I think it was nice to be able to ask questions	Enjoyed: Interactivity: Social Congruence:
I liked it!	Enjoyed:

B_Q4: If anything, what would you like to see improved?

Response:	Codes:
time responses, delay	-Responsivenes:
the time of the responses - and i had a Malfunction where it didn't answer	-Responsivenes: -Malfunction:
the character particularly and the text from what the NPC said should be broken up more to make it easier to read	-Motion: -Emotes: -Lipsync: -Social Congruence:
If the guides speech could feel more natural	-Social Congruence: -Voice:
The creepy man.	-Uncanny:
Had some issues when i picked items up and wanted to talk to the guide, it did not register my speech every time.	-Respons-time: -Malfunction:
mimicry	-Motion: Emotes:: Lipsync:
It could gesticulate more, or maybe ask more questions back.	-Motion: -Initiative:
Responses time to questions, TTS recognition	-Respons-time: -Malfunction:
His facial expression and maybe his vocabulary, which was very stiff and a bit boring	-Emotes
I found it a bit difficult to read the text boxes, and he wanted to show me something, but he couldn't.	-Interactivity:
His tone of voice is very monotone. I would like to hear him sound more natural.	-Voice:

if the character followed me	-Perception: -Proxemics:
He could take more action when I was silent. I was unsure what to ask about.	-Initiative:
Maybe slower the answers	-Knowledge:
Him being more human like and the graphics to be better. The graphics made it feel unreal and my eyes could quite get used to the bad quality of the graphic. He was very stiff.	-Graphical: -Motion:
maybe some idle animations for the character while just standing ?	-Motion:
The reason of the "game". I was confused what to do and what to expect. Emilio was just there to guide me and not to join me in my experience	-Initiative:
More responses, less repeated sentences, perhapse making jokes or something	-Social Congruence: -Personality
a distance reaction	-Proxemics:
I want to be able to pat the guide on the head, he deserved it	-Interactivity:
The response time of the character. It could have been due to being on a hotspot, but the character took a while to answer. The character didn't recognize the cup when I held it in my hand and asked what it was. He thought I was referring to him.	-Respons-time: -Social Congruence:
Fewer pauses in speech. A little more "speed".	-Social Congruence:
If anything, just a clarification to whether or not there was a goal to it, or if it was just to casually experience the ME lab.	-Interactivity:
1. Better mouth movements to correspond to the speech. 2. Better pacing of the responses to allow for follow-up questions. 3. React and follow when I move elsewhere instead of staying stuck in the same spot	-Lip sync: -Social Congruence: -Proxemics:
not such a boring man :) fun clothes and more expressive	-Personality: -Appearance
Maybe character animation and having it interact with objects more.	-Interaction: -Motion:
I think the guide's tone of voice could have been made more personal -- does he have a vernacular, is a topic within his personal interest. To a certain degree he seemed excited about the topics, but it could have been more explicit.	-Voice: Personality:
Jeg har sagt en masse til Julle. Det må være nok.	???
It was a bit stressful that the character started answering while walking, it might be better if he waited to start answering the question when you were in position at the space related to the question	-Social Congruence: -Proxemics:
Shorter answering time, higher visual quality on the room (HMD quality), another way of moving maybe - I felt it was a bit uncomfortable to move while standing still.	-Respons-time: -Cybersickness:

Thank you for participating in the experiment!
You will now be asked a series of questions which will be used to investigate our study.

Signing this consent form, you agree to let the members of the group from Aalborg University Copenhagen, use the recorded answers of the questionnaire for study related purposes. You will remain completely anonymous, and you can withdraw your consent at any time by sending an e-mail to: hbendi20@student.aau.dk

All personal information obtained during the test will be kept confidential and the data you provide will only be used for research purposes, academic objectives and publications. You will be identified through identification numbers, and not your real name.

☐ I consent

What is your Age?

☐ 18-22
☐ 23-28
☐ 29-34
☐ 35-39
☐ 40+

What is your biological sex?

☐ Female
☐ Male
☐ Prefer not to say

Have you had any experience with NPCs (Non-Player Characters)?

☐ Never
☐ Rarely
☐ Occasionally
☐ Often
☐ Daily

Have you had any experience with VR (Virtual reality)?

☐ Never
☐ Rarely
☐ Occasionally
☐ Often
☐ Daily

Let the test conductors know that you are now ready for the VR experience.

Conductor denotes; Participant ID Number:

_____

Conductor denotes; Experiment Condition:

☐ A
☐ B

I was aware of the other individual.
(1 = Not At All, 7 = Very Much)

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The other individual was aware of me.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I felt like the other individual was watching me.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I felt like I could almost touch the other individual.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I paid close attention to the other individual.
(1 = Not At All, 7 = Very Much)

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The other individual paid close attention to me.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The other individual understood what I meant.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I understood what the other individual meant.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I felt I was interacting with another intelligence.
(1 = Not At All, 7 = Very Much)

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I felt that I was sharing the experience with another person.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I felt as if I was in the company of another person.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I had a sense of being with another person.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The character I interacted with seemed aware of me
(1 = Not At All, 7 = Very Much)

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

I responded to the character as if they were a real person.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The character's responses felt timely and relevant.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

It felt like the character could see me or react to me.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The character was creepy.
(1 = Strongly Disagree, 7 = Strongly Agree)

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The character was disturbing.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The character made me feel uneasy.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

There was something 'off' about this character.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

The character's movements/expressions were unnatural.

1 2 3 4 5 6 7

❑ ❑ ❑ ❑ ❑ ❑ ❑

If anything, what made the character feel like a social being to you?

Was there anything about the character that felt off, strange, or unsettling to you? Please describe.

What did you think about having a guide?

If anything, what would you like to see improved?