

How can virtual agent's modalities impact player immersion in role-playing games?

Group 22

October 2024

First Name	Last Name	Student Number
Mateusz	Wilk	4003748
Anna Michelle	van der Spek	4077644
Adrien Joon-Ha	Im	3984389

Table 1: Group 22 Members

Abstract

Virtual agents are capable of both sophisticated verbal and non-verbal interaction, which can be used to enhance user experience, in entertainment scenarios. The aim of this research is to assess the effect on user immersion of a virtual agent in a role-playing game context. Through three agents in one scene with varying interaction types we seek to test how improved speech modulation and gesture generation will impact the user immersion, namely engagement and perception. This will be measured both by qualitative observations and post-interaction questionnaire. We find in the pilot studies that improved modalities can have an impact on user perception, however the association seems to be lower than originally hypothesized. We conclude the study by suggesting further research that would test the impact of modalities on user perception with statistical significance.

1 Introduction

Storytelling and collaboration are fundamental aspects of how humans interact within society, and role playing games like Dungeons & Dragons (D&D) purely exhibit these traits. There is growing potential to enhance these experiences with virtual embodied agents (IVAs) that are capable of sophisticated verbal and non-verbal interactions. With recent advancements, in virtual agent technology [3], it is also possible that they can be used to substitute physical robots, since they are not limited by embodiment constraints of physical alternatives. Humanoid IVAs can more easily replicate human-like gestures and expressions, than a physical robot. Isaac Wang and Jaime Ruiz [6] explored the capabilities of non-verbal communication in IVAs and concluded that they are successful in communicating conversational intent and attention.

This paper explores the use of IVAs in a role-playing game context, specifically D&D, where cooperation and improvisation are based on the narrative choices by the player. Through an

exploratory study, we aim to assess the benefits and challenges of integrating virtual agents into collaborative storytelling, focusing on how they enhance user perception and immersion.

2 Background and related work

Conducting a study that involves a virtual agent for entertainment was inspired by the work of Jia, Li and Zhenjiang, Miao [2], who developed an interactive virtual environment with intelligent agents that are capable of learning and interacting with each other, which can be used for gaming and simulations. They did not however focus on the gaming context specifically, so the aim of our study is to explore this particular aspect of interaction with virtual agents for entertainment.

Cooperation between humans and agents was introduced in [4] where it was mentioned that both humans and agents can operate cohesively in a hybrid team to accomplish common tasks.

Using a game together with virtual agents was done by [5], where in a part of their study that was promoting awareness on the HPV vaccine they used an NPC character in a fantasy game to guide the study participants.

3 Research Question

The focus of the study will be to investigate the effects of virtual embodied agents, exploring their use of both verbal and non-verbal cues to better understand how they affect the user's level of immersion, namely perception of the agent, through the following research question:

How does the use of different modalities in a virtual embodied agent influence the level of immersion experienced by players in a role-playing game, such as Dungeons and Dragons?

3.1 Hypotheses

Based on the research question we can hypothesize that players interacting with a virtual embodied agent that utilizes a richer set of non-verbal cues such as varied gestures, facial expressions, and gaze patterns, will show higher levels of immersion compared to those interacting with agents with only basic non-verbal cues. Furthermore it is expected that virtual agents with more gesture generation and more elaborate speech production will correlate with higher levels of engagement during gameplay, as they will be perceived to be more realistic by the player. The quality of the interaction and immersion will be studied through a post-interaction questionnaire and qualitative observations.

4 Method

In a game, like Dungeons and Dragons where the role of the Dungeon Master/narrator is to adapt to all the changes made in the story by the player, improvisation is key. An exploratory study will be conducted to test the extent of this cooperation between the player and the virtual agent. We seek to test this through both verbal and non-verbal interaction. There will be 3 levels of interaction types for a given scenario demonstrated by different characters. These interactions differ in terms of the voice's realism and how many gestures are generated.

4.1 Materials

In this study, we will be using the Unity Game Engine to construct a controlled environment, along with the convai.com API plugin to support the non-verbal interaction of the IVA. This plugin will facilitate gestures, gazes, and facial expressions that align with scenario-specific needs. The humanoid agent allows for more greater expressiveness, along with a voice that supports varied intonation. The virtual world also offers the potential for greater immersion, by incorporating assets that are thematically consistent with D&D narrative.

4.1.1 Interaction description

We developed a virtual agent for a specific D&D scenario, where the user will interact with three characters, making use of the Convai narrative design feature (essentially a finite-state machine), which supports a predefined flow of the interaction. The user can speak to one character at a time, and may switch between them as many times as they please. At first they receive the initial introduction from the narrator (the Dungeon Master), who explains the context of the scenario and the capacity for interaction to the user. The player is in a tavern with three characters who have found a secret treasure, but one of them stole the treasure. The objective is for the player to find out which of the three characters stole the treasure. Once the context of the interaction is given by the DM, the player must first make a choice for the next character to speak with. The user can then choose one of the three characters: Bob the pirate, Bjorn the Viking, and Brad the merchant. If the player manages to get an answer from the character that they have the treasure (who is Brad), the interaction ends. The flow of this interaction is illustrated in Figure 1.

4.1.2 Affordances and modalities

The virtual agent(s) will have a digital model resembling a human and will be capable of facial expressions, producing speech, and gestures. The user will be able to interact with the agent through speech. The agent can also react to a user's responses, which can be either expressed by speech, and then be processed by a speech-to-text algorithm or through a text box, similar to conventional chatbots. Different variations of voices and gestures will be used. For example, one version of the agent will have a more robotic voice and another version will have a more realistic voice.

4.1.3 Design Method

We are using the engineering design method, as we do not have the resources to involve the end users more in the design process. However, the robustness of the interaction will be user tested following an iterative development cycle, through testing a prototype during the pilot study. Based on the results of this study, necessary changes to the interaction will be made.

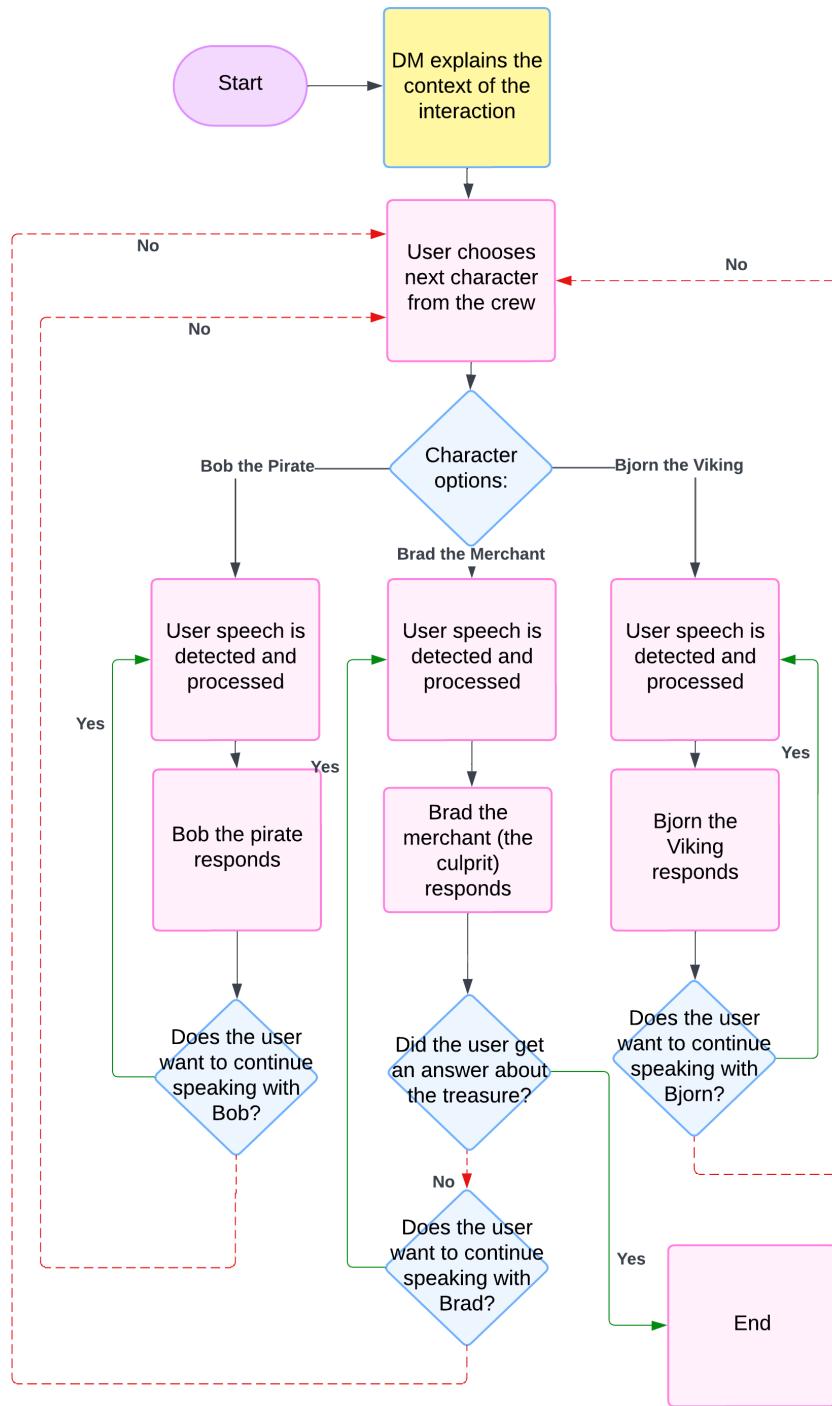


Figure 1: Interaction Diagram

4.2 Experimental setup / approach

The experiment will be conducted in two phases to appropriately address our research question.

- Phase 1: A pilot study to test robustness and usability of our interaction. We will test if the users are able to understand the context of the role-playing game, if they can communicate with the robot through the text-to-speech interface and if the virtual agents work correctly and robustly when using convai.com, especially the different modalities like speech and gesture generation.
- Phase 2: A pilot study focused on user perception. The main focus of this study is to understand whether the improved modalities can have an impact on more humanlike perception of the agent. The three agents will be set up in the following way with gradually increasing modalities: Brad the merchant will only display an idle animation, and will be using a standard TTS voice, Bjorn the Viking will be able to produce different gestures during speech and display a different idle animation, finally Bob the pirate will be able to move his lips, produce gestures during speech, have a separate idle animation and use a voice tailored to his character.

4.3 Measures

- In Phase 1 qualitative data will be collected through observations from each interaction as well as post interaction questions, capturing nuances in participant's responses and behaviors. These qualitative insights will provide contextual understanding, which will help us identify potential improvements for Phase 2.
- In the second phase, a version of the "Robotic Social Attributes Scale" (RoSAS)[1] will be used as a post-interaction questionnaire. This scale will help quantify participants' impressions of the IVA's warmth and competence, providing data to support our hypotheses about the impact of verbal and non-verbal cues.

5 Experimental results

5.0.1 Phase 1

During the first pilot study the interaction was still in development, therefore the agents were not capable of facial expressions. Two participants aged 19 and 18, respectively, were chosen to test the interaction, and their feedback as well as our notes from the experiment were collected.

In short, we observed that the users are able to correctly understand the purpose of the interaction. The affordance of speech of our agents seems to correctly correspond to the user's perception, they were able to communicate with the agents effectively through this modality.

The problems that the users identified were: the agents answering too vaguely, they found the faces unnatural due to the lack of movement, they didn't like the fact that the agents did not speak to each other, the individual stories did not feel connected or relevant at times.

Furthermore, we identified the following technical problems during the interaction: the pirate once did not respond to the question appropriately, as if he misunderstood it, the agents told conflicting stories, it could be perceived as if they were making the stories up.

Based on these results we concluded that the users are able to understand the interaction reasonably well, the agent gestures in particular work as planned, but facial movements need improvements. Another problem is the lack of a concrete knowledge base for each agent, which can impact the robustness of the interaction as the agent sometimes might not respond correctly. We decided to improve agent knowledge for Phase 2. Since the study design requires gradual modalities for different agents, we also worked to incorporate lip movements for Bob the pirate.

5.0.2 Phase 2

For the second study, we adjusted the knowledge base to make it coherent between all agents, we added an introductory part which explains the story behind our D&D scenario and adjusted the gestures for the agents so that they gradually improve from Brad the merchant to Bob the pirate (as described in section 4.2). We chose three participants from the age group between 18 and 20 years old. We evaluated the interaction using a questionnaire at the end.

We asked the users the following questions regarding each agent (Bob, Bjorn, or Brad) individually:

- How natural and engaging do the gestures and expressions of the agent feel?
Two of the three users gave the highest score to Bob, the agent with the most elaborate gestures and lip movement.
- How warm is the agent's voice?
Three participants indicated that Bob's voice has the highest perceived warmth.
- How contextually appropriate are agent's gestures?
Interestingly, all agents seemed to have a similar level of appropriateness according to the users.
- How effectively does the agent respond to the user's decisions?
Again, all participants indicated that there is no difference between agents.
- How unnatural do the agent's gestures and expressions feel?
The users reported no difference between IVAs
- How mechanical or unfitting does the agent's voice sound?
Most participants replied that Bjorn's voice was the least natural, Brad and Bob both received similar notes.

The final questions we asked were about the interaction as a whole:

- Did the agent's gestures and facial expressions enhance the sense of realism in the scenario?
Two users replied that the gestures were quite natural, and the other one indicated that the faces of the agents were too still.
- Did the agents' voices add depth and believability to the scenario?
All participants replied that the voices were natural and fitting, one of them even suggesting that they were indistinguishable from the voice of a human actor.
- Did you feel like the IVAs were genuine characters within the narrative?
Three users agreed with this statement, with two of them attributing this fact to their robust backstories.

- Did IVAs presence in the scenario make the interaction less immersive?
All three participants disagreed.

Since our study does not have statistical significance we can only interpret the results of this case study in terms of making new assumptions or hypotheses. It can be argued whether improved voices and gestures positively impact player immersion, based on the provided user feedback.

The users seemed to perceive Bob the most positively, which could be due to his improved modalities. Interestingly, the difference in gestures did not seem to matter for the users, which could imply that the voice and the agent's response is more noticeable than the gestures it produces.

It is also worth noting that the users' overall impression of the IVAs and interaction was quite positive, they indicated that the conversation felt mostly natural and believable.

6 Discussion

The findings from our two pilot studies suggest that while improved modalities in the agents had an effect on the user's immersion, the effect of these modalities are less than initially hypothesized. Our hypothesis stated that agents with richer modalities, such as realistic gestures, natural facial expressions and modulation in the voice would lead to higher levels of engagement and immersion by the player. The study shows that improvement in gestures and facial expressions often but did not always link to a greater likability or potential for immersion by the user.

Of the three agents, Bob the pirate was consistently described by the participants of the study as the most engaging agent. This is likely due to this character having natural gestures, better facial expressions including lip movement, and a highly realistic voice. This is explained by Wang and Ruiz that virtual agents that are capable of non-verbal communication are successful in communicating intent and attention [6], making them more engaging. This aligns with our initial hypothesis that better modalities are linked with a more engaging and immersive experience.

However, this hypothesis was not supported for all agents. Bjorn the viking, despite having more advanced gestures and voice features than Brad, received mixed ratings from the participants. Although his voice was described to be realistic, it was also considered by participants to be not fitting with the character, and in turn caused for a lower immersion of the user. This shows that even with advanced modalities, user immersion can be decreased due to the mismatch between the agent's narrative role and personal features. In the case of Bjorn the viking, the mismatch between his voice and his narrative background caused participants to have a less realistic and immersive perception in spite of his improved gestures.

Brad the merchant, who had the least advanced modalities, also showed low ratings by the participants. He had minimal gestures, and the least realistic voice out of the three agents. One participant noted that as Brad had a mask on, the lack of lip movements or facial expressions did not lessen realism of the character. This shows that some design elements can be used to balance out the lack of modalities of some characters.

It has additionally been noted by participants that despite the realistic voice of some agents, the lack of facial expressions on them was noticeable and reduced realism of the IVA. This suggests that immersion is not only linked to a single modality, but likely due to a combination of multiple factors.

Overall, the response of the participants with regards to the virtual agents was positive, and it has been observed that the users were immersed when interacting with the agents. We should note

that the results of this pilot study come from a very small sample size, and that all participants were of a similar background in terms of age, and prior exposure to virtual agents. The limitations in this study show the need for further research in order to make conclusive and precise findings regarding how these modalities exactly contribute to better immersion.

7 Conclusion and further research

In conclusion, the preliminary study suggests that enhanced modalities of the IVAs may have a positive effect on player immersion in a role playing game. It was demonstrated that virtual agents with better voice, gestures, and facial expressions can be better perceived by the user and can allow them to be more immersed in the game. However, it was also observed that different modalities do not all have an equal effect on player immersion. Improved voice characteristics seem to have the highest effect on player immersion, while gestures have a more minimal effect and are sometimes not noticed by the player.

Further research regarding the effects of different modalities on player immersion could be done to test our findings with statistical significance for the general population. We did not prove the association in general, these are our findings for the small study we did, and it can be argued that gestures can affect the user perception of the interaction.

Additionally, a legitimate direction of future research could involve trying to understand which modalities seem to have the highest impact on the user. In our study we used a combination of several modalities for each agent, and it could be hard to determine the exact modalities that contributed to the agents being perceived by the users in a way that it was noted in the results.

Overall, this study demonstrated the importance of virtual agent modalities such as facial expressions, gestures, and voice in order to enrich the player experience in a role playing game. It can serve as a foundation to develop more engaging and immersive virtual agents not only in gaming but also for broader applications in entertainment.

References

- [1] Colleen M. Carpinella, Alisa B. Wyman, Michael A. Perez, and Steven J. Stroessner. The robotic social attributes scale (rosas): Development and validation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 254–262, 2017.
- [2] Li Jia and Miao Zhenjiang. Entertainment oriented intelligent virtual environment with agent and neural networks. pages 90–95, 11 2007.
- [3] Samuel Mascarenhas, Manuel Guimarães, Rui Prada, João Dias, Pedro A. Santos, Kam Star, Ben Hirsh, Ellis Spice, and Rob Kommeren. A virtual agent toolkit for serious games developers. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pages 1–7, 2018.
- [4] Tim Schwartz, Michael Feld, Christian Bürckert, Svilen Dimitrov, Joachim Folz, Dieter Hutter, Peter Hevesi, Bernd Kiefer, Hans-Ulrich Krieger, Christoph Lüth, Dennis Mronga, Gerald Pirk, Thomas Röfer, Torsten Spieldner, Malte Wirkus, Ingo Zinnikus, and Sirkо Straube. Hybrid teams of humans, robots, and virtual agents in a production setting. In *2016 12th International Conference on Intelligent Environments (IE)*, pages 234–237, 2016.

- [5] Ian Steenstra, Prasanth Murali, Rebecca B. Perkins, Natalie Joseph, Michael K Paasche-Orlow, and Timothy Bickmore. Engaging and entertaining adolescents in health education using llm-generated fantasy narrative games and virtual agents. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.
- [6] Ruiz J Wang, I. “examining the use of nonverbal communication in virtual agents.”. *International Journal of Human-Computer Interaction* 37 (17): 1648–73. doi:10.1080/10447318.2021.1898851., 2021.