Actor Takeover of Animated Characters

Jingyi Zhang*
University College London

Abstract

With the rise in accessibility of consumer-grade virtual reality headsets, social virtual reality applications have now caught widespread attention. One's anticipation for the virtual world is for it to be lively with characters. However, current technology falls short of generating intelligent virtual agents, thus creating a densely-populated environment with which one could have a variety of interactions is out of reach. This position paper introduces a takeover system which enables a single actor to seamlessly take over control of multiple virtual characters in a social virtual reality environment. In our experiments, we expect our system to enhance the perceived social presence of the scene and to uncover the difference in perception regarding computer-driven agents and human controlled avatars when interacting with multiple characters in the same environment.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies

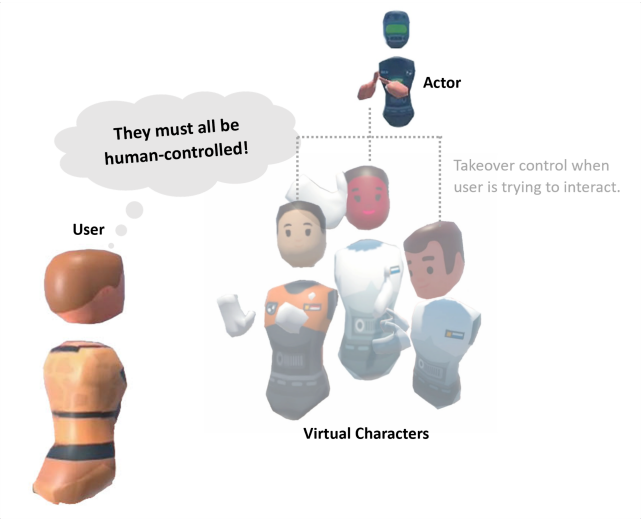

Figure 1: Concept sketch of an actor jumping among multiple characters and interacting with the user when needed

1 Introduction

The evolution of consumer-grade head mounted displays (HMDs) and high-speed internet services has facilitated access to social virtual reality (SVR) applications. To enhance the realism of the virtual world to bring it in line with the similar situations in reality, we expect the scene to populate with background characters. However, despite the infinite possibilities and freedoms that virtual reality interactions offer us, the technologies of autonomous intelligent agents are not at the required level. These agents would be required to possess the ability of natural verbal communication and realistic non-verbal behavioural generation such as logical interactions with objects and other characters, body languages, or facial expression.

Extensive research has been conducted in each of these fields, from generating human-like language with models like ChatGPT [9], to modeling non-verbal behavior using statistical model [8, 11], rule-based methods [1, 10] or machine learning model [2–5, 7]. However, such simulations still struggle with a decline in naturalness over time or a delayed reaction time which restrict their believability. Therefore, the reliance on human actors in controlling virtual characters remains.

Considering that background characters often outnumber users in a scene, and that during most of the time those characters are rarely in direct contact with the user, we see the need for an actor to switch between multiple characters and assume control to interact with users when required. We expect that an actor could take over an existing virtual character to steer plot lines, address specific queries, or introduce novelty into the scene.

*e-mail: jy.zhang@ucl.ac.uk

This position paper proposes research:

- to develop a system that allow a single actor to take over control of multiple virtual characters;
- to understand how users distinguish between human controlled avatars and computer-driven agents in multi-character scenarios;
- to investigate whether co-presence could be enhanced by having one actor controlling multiple characters

2 Prior Research

In a pilot study, we developed a takeover system which allowed a single actor to seamlessly transition among multiple virtual characters in virtual environment [12], as shown in Figure 1. The system enabled the actor to takeover control of one of the characters when the participant attempted interaction, and playing pre-recorded motions in loop during other times. The actor were able to view the participant's avatar, and several transparent avatars which represents the future key frames of available characters for takeover. The actor would be teleport to the position and rotation of the target character before takeover happened and a linear interpolation method would produce an animation to fill in the gap between the key frame avatar and the actor avatar's current position and posture. The system was built upon the Ubiq social VR platform [6].

To evaluate our system, we have created a cafe scenario with three virtual characters (a barista and two customers). An experiment was conducted to compare between our system and the one where responses to participants were made by playing recorded clips using "Wizard-of-Oz" setup. The experiment was designed as within-subject to populate the database, so the participants were asked to enter the same scene twice to experience both conditions. Both

participants and the actor could access the scene using Meta Quest 2[1]. During each of the trials, the participants were directed to complete tasks to interact with all three characters. After the whole experience, the participants were asked to answer co-presence questionnaire followed by an interview.

The results showed that the takeover system succeeded in making the participants unaware of the transfer of control and non of the participants realised the three characters were controlled by one actor. Our system improved the perception of social presence compared to the setup with recorded responses and invoked a stronger illusion that more characters were under human control rather than computer-driven. The interview also provided us insights on the criteria that participants used to determine human-controlled or computer-driven character. This included the character's facing direction, its ability to perform unexpected actions, the conventionality of the response, the size of movement area, the reaction speed and the naturalness of the character's motion.

3 Planned Research

In subsequent research, we plan to improve the takeover system and conduct user case studies to evaluate its plausibility and effectiveness in improving perceived social presence of virtual environments inhabited by pre-recorded characters.

Regarding the takeover system, we aim to introduce more realism of interactions, including voice communication, additional tracking points for avatar control, and enhanced agent intelligence during non-interactions. By implementing audio features, users and characters can engage in verbal communication. A voice manipulation method will be employed to differentiate between the actor's voice while taking over different characters. By transitioning from three-point to six-point tracking and using a more realistic humanoid model, the virtual characters are allowed to express more intricate body language. This may also help to address concerns raised in prior studies about mechanical movements and lack of social presence even for human controlled characters, which may be caused by the cartoon like character appearance with only head and hands. Furthermore, we intend to integrate motion generation techniques to replace the recorded motion replayed in loop during non-interactions, and when giving responses in the comparison group in the experiment.

We aim to ease the takeover process for the actor by eliminating the need for mimicking key frame avatars. We expect an animation to seamlessly fill the gap in posture differences between the actor's avatar and the character being taken over. Additionally, we plan to develop a method that predicts the likelihood of user interaction with each character so that the actor can get prepared before interactions begins.

To gain a deeper understanding of whether the advanced takeover system with more interactions involved and easier takeover procedure could enhance the perceived social presence of the environment, we aim to use our system with different scenarios to conduct experiment that investigates:

- The influential factors that vary levels of perceived social presence and the number of perceived human-controlled avatars in multi-character environment.
- The key criteria that users use to distinguish computer driven agents and human controlled avatars when immersed in a virtual environment with multiple characters or even crowds.
- The feasibility of actors using our system to portray characters with different personalities. (This may aid in comprehending how users remember and differentiate various characters in the same virtual environment.)

[1]Meta Quest 2: https://www.meta.com/quest/products/quest-2/

I would appreciate feedback on the feasibility and practicability of this proposal during the Doctoral Consortium at the IEEE VR 2024 conference. Specifically, I am seeking insights on potential real-world applications of the system, appropriate testing scenarios and guidance on isolating individual factors in complex social VR environments to explore the above mentioned research questions.

References

[1] J. Cassell, H. H. Vilhjálmsson, and T. Bickmore. BEAT: The behavior expression animation toolkit. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 477–486. Association for Computing Machinery, New York, NY, USA, 2001. doi: 10.1145/383259.383315

[2] C. Curtis, S. O. Adalgeirsson, H. S. Ciurdar, P. McDermott, J. Velásquez, W. B. Knox, A. Martinez, D. Gaztelumendi, N. A. Goussies, T. Liu, and P. Nandy. Toward believable acting for autonomous animated characters. In *Proceedings of the 15th ACM SIGGRAPH Conference on Motion, Interaction and Games*. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3561975.3562941

[3] S. Dermouche and C. Pelachaud. Generative model of agent's behaviors in human-agent interaction. In *International Conference on Multimodal Interaction*, pp. 375–384. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3340555.3353758

[4] W. Feng, A. Kannan, G. Gkioxari, and C. L. Zitnick. Learn2Smile: Learning non-verbal interaction through observation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4131–4138, 2017. doi: 10.1109/IROS.2017.8206272

[5] Y. Ferstl and R. Mcdonnell. Investigating the use of recurrent motion modelling for speech gesture generation. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, pp. 93–98. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3267851.3267898

[6] S. J. Friston, B. J. Congdon, D. Swapp, L. Izzouzi, K. Brandstätter, D. Archer, O. Olkkonen, F. J. Thiel, and A. Steed. Ubiq: A system to build flexible social virtual reality experiences. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology*. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3489849.3489871

[7] G. Gomes, C. A. Vidal, J. B. Cavalcante Neto, and Y. L. B. Nogueira. An emotional virtual character: A deep learning approach with reinforcement learning. In *21st Symposium on Virtual and Augmented Reality (SVR)*, pp. 223–231, 2019. doi: 10.1109/SVR.2019.00047

[8] S. P. Lee, J. B. Badler, and N. I. Badler. Eyes alive. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 637–644. Association for Computing Machinery, New York, NY, USA, 2002. doi: 10.1145/566570.566629

[9] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, and B. Ge. Summary of ChatGPT/GPT-4 research and perspective towards the future of large language models. 2023. doi: 10.48550/arXiv.2304.01852

[10] S. Marsella, Y. Xu, M. Lhommet, A. Feng, S. Scherer, and A. Shapiro. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 25–35. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2485895.2485900

[11] M. Neff, M. Kipp, I. Albrecht, and H.-P. Seidel. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Trans. Graph.*, 27(1):1–24, mar 2008. doi: 10.1145/1330511.1330516

[12] J. Zhang, K. Brandstätter, and A. Steed. Supporting co-presence in populated virtual environments by actor takeover of animated characters. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 940–949, 2023. doi: 10.1109/ISMAR59233.2023.00110