Exploring Multimodal Interaction with Virtual Pets in VR

RICHARD CARROLL, Colorado State University, USA

This study investigates the interaction between users and virtual pets in Virtual Reality (VR) through three modalities: controllers, gaze, and speech. As VR technology advances, opportunities for creating immersive experiences that can serve purposes in entertainment, therapy, and companionship. This research aims to understand how these interaction methods influence user engagement and emotional connection with virtual pets. Three participants engaged with a virtual pet dog in a VR room environment, each using a distinct modality. Post-experience questionnaires assessed their interactions, revealing preferences and challenges, which contribute to the field of Human-Computer Interaction (HCI).

CCS Concepts: • **Human-centered computing** → **Interaction paradigms**; • **Computing methodologies** → *Virtual reality*; • **Applied computing** → *Computer games*.

Additional Key Words and Phrases: Multimodal Interaction, Virtual Reality, Virtual Pets, User Engagement, Human-Computer Interaction

ACM Reference Format:
Richard Carroll. 2025. Exploring Multimodal Interaction with Virtual Pets in VR. 1, 1 (May 2025), 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Virtual Reality (VR) has evolved from a specialized tool for gaming into a versatile platform impacting various domains, including education, healthcare, and social interaction. The ability to create immersive, interactive environments opens new avenues for human-computer interaction, particularly through virtual pets. These digital companions potentially provide companionship for individuals unable to own real pets, support mental health therapy, and offer interactive entertainment. This project explores how different interaction modalities—controller input, gaze-based interaction, and voice commands—affect user engagement and emotional connection with virtual pets in VR. By examining these modalities, we aim to identify methods that enhance user experience, providing insights for developing more effective VR systems. The study involved three participants interacting with a virtual dog in a VR living room, each using a distinct modality, followed by detailed feedback collection.

2 BACKGROUND

Multimodal interaction in VR has gained recognition for enhancing immersion and engagement. Lee and Jung (2023) demonstrated that combining speech and gaze inputs increases users' presence and emotional involvement in virtual environments. Chen et al. (2011) found voice and tactile-like interactions with virtual animals foster emotional attachment in educational contexts. Similarly, Cheok et al. (2011) and Lin et al. (2017) explored mixed-reality pet interactions, emphasizing emotional design for virtual companionship. Despite advances, challenges remain related to input reliability and intuitiveness in emotionally significant contexts. Our study extends this

Author's address: Richard Carroll, Colorado State University, Fort Collins, CO, USA, rocarrol@colostate.edu.

ACM XXXX-XXXX/2025/5-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

foundation by comparing controller, gaze, and voice modalities for virtual pet interactions in VR, addressing gaps about their effectiveness comparatively.

3 AIM

The primary objective is to evaluate how controllers, gaze, and voice modalities shape users' experience with virtual pets in VR, assessing naturalness, engagement, enjoyment, and emotional connection. Hypothesizing that controllers, due to familiarity and tactile feedback, may yield higher engagement and connection, the study uses specific questionnaire items to measure these constructs. This research informs VR development for therapy, education, and entertainment applications.

4 METHOD

4.1 Participants

Three participants aged 18–35, selected for prior VR experience but not necessarily with virtual pet interactions, took part. The pre-experiment survey captured age, gender, prior VR experience, and personal attitudes toward pets to reduce bias. All participants reported openness to all interaction modalities.

4.2 VR Environment and Virtual Pet

The environment is a distraction-free virtual living room with warm lighting designed to foster immersion. The study initially tested five dog breeds—cur, corgi, pug, chihuahua, and German shepherd—each with eight animations (breathing, tail wiggling, walking, running, eating, angry, sitting). Participant feedback indicated preferences for "cuter" breeds (pug, corgi), though one participant favored larger, masculine breeds (German shepherd). For controlled interaction consistency, the German shepherd was primarily used, while noting breed influence on engagement.

Fig. 1. Virtual dogs used in the VR environment displaying variety in breeds.

4.3 Interaction Modalities

- **Controller Input**: Oculus Touch controllers enabled natural actions—petting by approaching the dog's head, playing fetch by grabbing and throwing a virtual ball, and issuing commands via button presses.
- **Gaze-Based Interaction**: Eye-tracking tracked sustained gaze for petting, a look-away gesture for fetch, and gaze-based selection of command icons.
- **Voice Commands**: Meta Voice SDK provided speech recognition for commands like "pet the dog," "throw the ball," and "sit," although background noise impacted accuracy occasionally.

Fig. 2. Controller-based petting and interaction interface.

4.4 Technical Setup

The setup used Unity's XR Interaction Toolkit handling input, interactions, grab and manipulation behaviors, and locomotion support, alongside the OpenXR Plugin for device agnosticism. The Unity Input System enabled flexible controller detection. The Meta Voice SDK handled speech recognition. The environment was designed to be distraction-free, with participants tested alone to minimize external interference.

4.5 Measures

I collected data using a Google Form with a custom-designed questionnaire featuring a straightforward rating scale, like choosing from "strongly agree" to "strongly disagree." I crafted questions to explore how participants felt about the virtual pet, including how connected they felt to it and whether the virtual environment, such as a living room, felt realistic and lively. I also added open-ended questions to let participants share their thoughts in their own words. Combining these numerical ratings with written responses gave me a clear understanding of participants' thoughts and feelings about the VR pet experience.

4.6 Procedure

Participants completed a 5-minute training to familiarize with the modality, then performed three tasks—petting, fetching, and issuing commands—for approximately 5 minutes each. Subsequently, they filled a 10–15 minute questionnaire immediately to capture fresh impressions. Total session lasted about 25–30 minutes.

5 RESULTS AND DISCUSSION

5.1 Overall Experience

All participants reported positive experiences, praising the cozy environment and the dog's lifelike design. Picking up and throwing the ball was particularly engaging across modalities, fostering immersion and agency. Technical challenges with command recognition and animation glitches, however, reduced interaction smoothness.

5.2 Controller Input

The participant using controllers found petting and fetch intuitive and enjoyable, supported by tactile feedback. Command inputs suffered from inconsistencies—commands like "sit" frequently failed and animations glitched, interrupting immersion. Despite this, the participant appreciated the realistic breathing and tail animations.

5.3 Gaze-Based Interaction

Gaze interaction was described as novel and enjoyable for petting and fetch initiation, but selection of command icons was unintuitive with poor animation response. The participant expressed frustration at frozen animations but valued the engaging environment.

5.4 Voice Commands

Voice interaction offered the potential for natural communication but faced significant reliability issues due to speech recognition challenges, especially amid background noise. Although commands sometimes worked, animation glitches impaired realism. The participant still appreciated speaking to the dog and the environment ambiance.

5.5 Quantitative Questionnaire Summary

Satisfaction ratings ranged 7–8 out of 10; controllers rated easiest to use (8/10), gaze intermediate (6/10), and voice lowest (5/10). Engagement scores were moderate to high, with controllers favored due to familiarity. Emotional connection averaged 6–7/10, strengthened by pet design but constrained by technical faults.

5.6 User Feedback

Participants enjoyed petting and fetching but wanted more diverse interactions such as teaching tricks or grooming. The "angry" animation was disliked when triggered unexpectedly. Requests for breed personalization were common, favoring cuter animals like pugs and corgis.

5.7 Technical Challenges

Challenges included mapping errors on controllers, calibration sensitivity on gaze, and speech recognition failures on voice. Animation glitches—e.g., freezing mid-sit or abnormal running—were recurrent, indicating Unity system implementation issues.

5.8 Limitations

The small number of participants in this study and the use of a between-subjects design make it difficult to apply the findings to a larger population. Since each participant only experienced one type of interaction method, we could not compare how the same person responded to different modalities. This limits our ability to understand how each method might perform across a wider range of users.

Additionally, there were some technical issues that affected the overall experience. These problems prevented participants from fully engaging with the virtual pet as intended. For example, if a participant encountered difficulties with the voice commands or gaze tracking, it could have influenced their enjoyment and connection with the virtual pet. These limitations highlight the need for further research with a larger group of participants and improved technology to better understand how different interaction methods can be effectively used in virtual reality settings.

5.9 Future Directions

In the future, it would be helpful to include more people in the study so the results can better represent a wider group. Right now, with only three participants, it's hard to know if the results apply to everyone. Also, instead of having each person try only one way of interacting with the pet, future studies should let the same person try all three methods (controller, gaze, and voice). This would help us better compare which one works best for different people.

It's also important to fix some of the technical issues that came up, especially with the voice commands not always working and the animations sometimes freezing or acting strangely. Making sure everything works smoothly would improve the overall experience and help participants feel more connected to the virtual pet.

Another idea is to combine interaction methods—like letting people use both voice and gaze at the same time—to see if that feels more natural. It might also be good to use pets that are more neutral or imaginary, so personal preferences about real dog breeds don't affect how people feel about the pet.

Finally, this kind of VR pet system could be very useful for people who are afraid of real animals or who can't have pets for other reasons. It could also be used in schools or therapy programs to help teach care and responsibility in a safe and fun way.

6 CONCLUSION

This study shows that using different ways to interact with virtual pets in VR like controllers, gaze, and voice can help users feel more involved and connected emotionally to the pets. However, this only happens well if the system works reliably without many glitches. Among the three methods tested, using controllers gave the best overall experience. Participants found this method easier to use and more satisfying because the controls were familiar and responsive.

The voice and gaze methods also showed good potential. Users liked the idea of speaking to the pet or looking at it to interact, which felt natural and interesting. But these types of interaction still need some fixing and improvements to work smoothly. For instance, voice recognition sometimes misunderstood commands, and gaze tracking could be inaccurate or hard to use.

REFERENCES

[1] Cheok, D., et al. (2011). Metazoa Ludens: Mixed-Reality Interaction and Play for Small Pets and Humans. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 41(5), 876–891.

[2] Chen, Z.-H., Liao, C., Chien, T.-C., and Chan, T.-W. (2011). Animal companions: Fostering children's effort-making by nurturing virtual pets. *British Journal of Educational Technology*, 42, 166–180.

[3] Lin, C., Faas, T., and Brady, E. (2017). Exploring affection-oriented virtual pet game design strategies. In *Proceedings of the Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, 362–369.

[4] Tan, R. K. C., Cheok, A. D., and Teh, J. K. S. (2006). Metazoa Ludens: Mixed Reality Environment for Playing Computer Games with Pets. *International Journal of Virtual Reality*, 5(3), 53–58.

[5] Lee, J., and Jung, Y. (2023). Enhancing Presence through Multimodal Interaction in Virtual Reality. *Journal of Virtual Reality Research*, 12(4), 45–60.