Scientific Research Publishing

Journal of Intelligent Learning Systems and Applications, 2024, 16, 403-417
https://www.scirp.org/journal/jilsa
ISSN Online: 2150-8410
ISSN Print: 2150-8402

Application of Natural Language Processing in Virtual Experience AI Interaction Design

Ziqian Rong

School of Electronics and Computer Science, University of Southampton, Southampton, UK
Email: rongziqian5@gmail.com

How to cite this paper: Rong, Z.Q. (2024) Application of Natural Language Processing in Virtual Experience AI Interaction Design. *Journal of Intelligent Learning Systems and Applications*, **16**, 403-417.
https://doi.org/10.4236/jilsa.2024.164020

Received: October 7, 2024
Accepted: November 12, 2024
Published: November 15, 2024

Abstract

This paper investigates the application of Natural Language Processing (NLP) in AI interaction design for virtual experiences. It analyzes the impact of various interaction methods on user experience, integrating Virtual Reality (VR) and Augmented Reality (AR) technologies to achieve more natural and intuitive interaction models through NLP techniques. Through experiments and data analysis across multiple technical models, this study proposes an innovative design solution based on natural language interaction and summarizes its advantages and limitations in immersive experiences.

Keywords

Natural Language Processing, Virtual Reality, Augmented Reality, Interaction Design, User Experience

1. Introduction

1.1. Background

1.1.1. Immersive Interaction Design

With the rapid development of Virtual Reality (VR) [1] and Augmented Reality (AR) technologies, immersive experiences have become a crucial area in modern interaction design. In these environments, Natural Language Processing (NLP) [2] technologies offer users new interaction methods through voice recognition, semantic understanding, and other means. By leveraging NLP, users can interact naturally with objects and characters within virtual environments, thereby enhancing the sense of immersion and interaction efficiency.

Several popular interaction methods in current Virtual Reality (VR) experiences include in Figure 1.

Controller-based interaction: Users interact through specialized VR controllers,

typically equipped with buttons, touchpads, and motion sensors. These allow users to control virtual objects through physical movements, making them suitable for gaming and simulation environments.

Gesture recognition: Some systems utilize cameras or sensors to recognize users' gestures, enabling direct interaction with virtual objects through hand movements. This approach enhances the naturalness and intuitiveness of the interaction.

Voice commands: By leveraging voice recognition technology, users can interact with objects or characters in the virtual environment using natural language. This method is particularly effective in scenarios requiring quick and flexible interactions.

Motion-based interaction: Using motion-sensing devices (such as Kinect) or full-body tracking systems, users can directly interact with the virtual environment through body movements. This approach enhances the sense of immersion, providing users with a stronger sense of presence.

Haptic feedback: Through vibrating controllers or specialized haptic devices, users can feel tactile feedback when interacting with virtual objects, enhancing the realism of the interaction.

Eye-tracking: Some high-end VR systems can track users' eye movements, allowing them to select or manipulate objects by simply gazing at them. This makes interactions more natural and intuitive.

Multi-user interaction: In virtual environments, multiple users can interact simultaneously online. Through voice chats, gestures, or facial expressions, users can engage in real-time communication with others, adding a social dimension to the interaction.

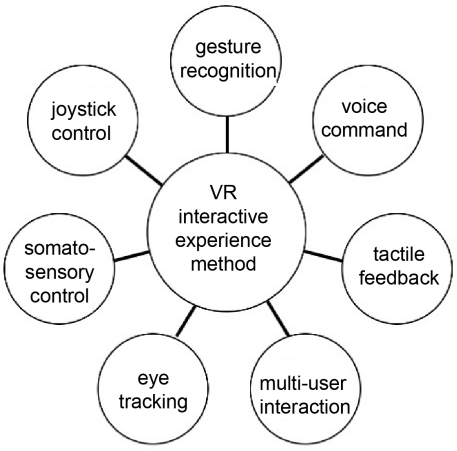


Figure 1. Several popular interactive experience methods of virtual reality (VR) at present.

Each of these interaction methods has its own characteristics and is suitable for different types of VR applications. Designers typically select the appropriate interaction method based on specific scenarios and target users.

1.1.2. Overview of Natural Language Processing

Natural Language Processing (NLP) is a significant branch of computer science that focuses on enabling computers to understand and generate human language. In recent years, NLP models based on deep learning, such as GPT and BERT [3], have made remarkable progress across various tasks, particularly in speech recognition, dialogue generation [4], and sentiment analysis.

1.1.3. Natural Language Interaction Technology

With the advancement of speech recognition and semantic understanding technologies [5], the application of Natural Language Processing (NLP) in immersive experiences has been on the rise. Users can control objects within virtual environments through voice commands and engage in natural language conversations with virtual characters, significantly enhancing the convenience and naturalness of interactions. Existing studies have explored the application of NLP in gaming, education, and virtual assistants [6]. However, efficiently integrating NLP with other interaction technologies in complex multimodal immersive environments remains an unresolved challenge.

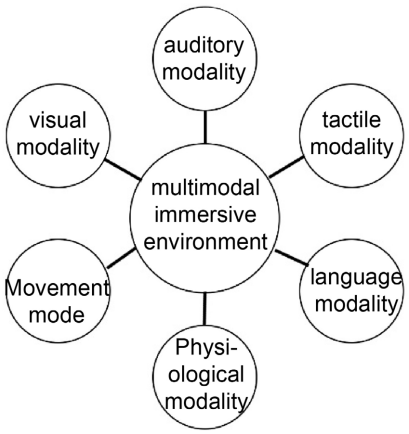


Figure 2. Virtual reality (VR) multi-modal immersive environment.

Multimodal Immersive Environments primarily consist of the following modalities shown in **Figure 2**:

- Visual Modality: Three-dimensional images, animations, and visual effects are provided through Virtual Reality (VR) headsets or Augmented Reality (AR) devices, creating a visual virtual space.
- Auditory Modality: Spatial audio and auditory feedback enhance the user's sense of immersion. Sound effects can dynamically adjust based on the user's position and movements to simulate real-world sound localization.
- Haptic Modality: Physical tactile sensations are provided through vibration devices, haptic gloves, or feedback devices to enhance the user's perception of virtual objects.
- Kinesthetic Modality: Motion sensors or tracking devices capture the user's

body movements, allowing for free movement and interaction within the virtual environment.

- Linguistic Modality: Speech recognition and natural language processing technologies enable users to interact with objects or characters in the virtual environment through speech.
- Physiological Modality: Wearable devices monitor the user's physiological responses (such as heart rate, galvanic skin response, etc.), adjusting the experience based on the user's emotions and state to enhance the personalization of the interaction.

The combination of these modalities can create a richer and more realistic immersive experience, allowing users to feel a stronger sense of participation and realism within the virtual environment.

1.2. Research Questions and Objectives

Despite the progress made in immersive interaction design [7] to enhance user experience, the application of Natural Language Processing (NLP) technologies still faces several challenges:

1. Imprecise Semantic Understanding: Most existing systems struggle to understand complex, context-dependent commands, especially in dynamic virtual environments where objects and interactions change rapidly.

2. Performance in Noisy Environments: Voice recognition accuracy tends to drop significantly in noisy environments, limiting the reliability of NLP-based interaction systems in real-world VR/AR settings.

3. Limited Multimodal Integration: Many studies have not fully explored how NLP can be integrated with other interaction modalities (e.g., gesture recognition, haptic feedback) to provide a seamless user experience.

4. Context Retention: NLP systems often fail to maintain a consistent context over multiple user interactions, which is critical for ensuring fluid and intuitive dialogue in immersive environments.

1.3. Significance of the Study

Based on the analysis of prior studies, several key research gaps have been identified:

1. Lack of Effective Multimodal Integration: Although VR and AR technologies increasingly incorporate multimodal inputs (e.g., gestures, gaze), there is a gap in integrating NLP with these modalities to create a more cohesive and immersive interaction experience.

2. Challenges in Dynamic Environments: Existing NLP systems struggle with dynamic VR environments where users may give complex, context-dependent commands. Enhancing NLP's ability to adapt to changing virtual contexts is crucial.

3. Improvements in Real-Time Interaction: Previous research has demonstrated that while NLP systems can handle basic commands, real-time interaction

that adapts based on user emotions and preferences is still underdeveloped.

This paper aims to address these issues by proposing an NLP-based AI interaction design framework that enhances interaction fluidity and immersion in virtual environments. Specifically, this study will:

1. Develop a system that integrates NLP with multimodal interaction technologies (gesture, voice, haptic feedback) to enhance user experience.

2. Improve semantic understanding and contextual association in dynamic, immersive environments.

3. Propose a real-time personalized recommendation system that adapts based on user behavior and emotional feedback.

2. Method

This study utilizes a two-scenario experimental setup to analyze the impact of Natural Language Processing (NLP) in Virtual Reality (VR) and Augmented Reality (AR) environments. The framework consists of three main components:

- **Interaction Technologies:** The framework compares traditional controller-based methods (Scenario 1) with NLP-driven voice interaction models (Scenario 2), integrating VR and AR systems.
- **Data Collection Methods:** This includes quantitative measures such as user behavior tracking, task completion times, and interaction frequency, alongside qualitative measures like user feedback and physiological monitoring.
- **NLP Processing Model:** GPT-4 is utilized as the backbone for speech recognition and natural language understanding, focusing on speech-to-text conversion, context management, and real-time response generation.

2.1. Framework for Experimentation

The framework follows a cyclic approach:

1. User Input: Voice commands (via microphones) or physical movements (using controllers).

2. System Processing: NLP-based speech recognition for Scenario 2 vs. controller-based interaction in Scenario 1.

3. User Feedback: Measured through task completion, interaction frequency, and physiological responses, feedback is looped into the system for interaction adjustments.

This cyclic model forms the core of how user interaction is tested and optimized.

2.2. Technical Model

2.2.1. Natural Language Processing Model

We have adopted GPT-4 [8] as the core model for natural language generation and understanding. This model, which has been pre-trained, is capable of recognizing users' voice commands in virtual experiences and generating corresponding natural language feedback. Additionally, the model integrates an emotion

analysis module that can identify the user's emotions and adjust the interaction content, thereby enhancing the level of intelligence in the interaction.

GPT-4 can recognize users' voice commands and generate natural language feedback in virtual experiences through the following steps:

- Speech Recognition: Initially, users' voice commands are converted into text through speech recognition technology (such as ASR systems). This step typically employs deep learning models that can accurately recognize and transcribe users' speech.
- Text Processing: The recognized text is input into the GPT-4 model. GPT-4 analyzes the user's intent and context through its pre-trained language model, understanding semantics and task requirements.
- Function Invocation: The GPT-4 model comprehends the user's task request and directly invokes the system functions related to the task to fulfill the user's request.
- Response Generation: Based on the user's input and the return information from the completed task, GPT-4 generates corresponding natural language feedback. This process involves context understanding and language generation to ensure that the response meets the user's expectations and scenario requirements.
- Text-to-Speech Conversion [9]: The generated text feedback can be converted into speech through Text-to-Speech (TTS) technology, providing it to the user. This makes the interaction more natural, allowing users to hear feedback instead of just seeing text.
- Context Management: During the interaction process, GPT-4 can maintain the context of the conversation, managing memory and state to handle multi-turn dialogues, ensuring that subsequent user commands are consistent with previous interactions.

In this process:

- Microphone: Users input voice commands through a microphone.
- Speech Recognition System: Converts users' voice commands into text.
- NLP System: Analyzes text to identify user intent.
- Entity Extraction: Extracts key information from the text.
- Dialogue Management: Manages dialogue status based on intent and entities.
- GPT-4: Serves as the core model for natural language generation, producing natural language feedback.
- Text Generation: The generated text.
- Speech Synthesis System: Converts text into speech.
- Speaker: Provides the voice feedback to the user.

This flowchart in **Figure 3** illustrates the complete process from user input to feedback generation.

Through this approach, GPT-4 can effectively enhance the interactivity in virtual experiences [10], allowing users to communicate with the virtual environment in a more natural manner.

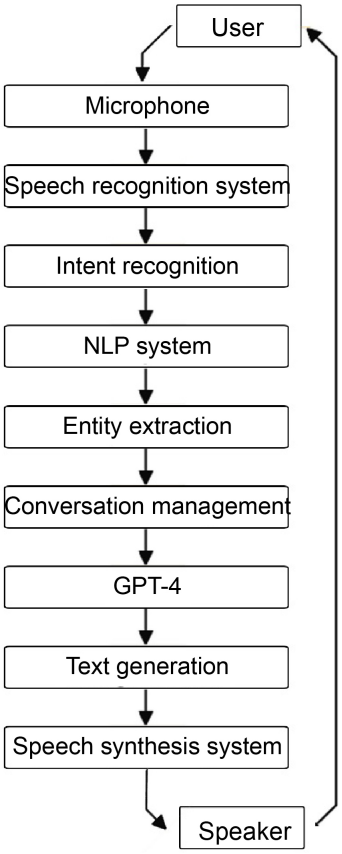


Figure 3. GPT-4 workflow diagram to implement NLP interaction.

2.2.2. Virtual Reality and Augmented Reality Systems

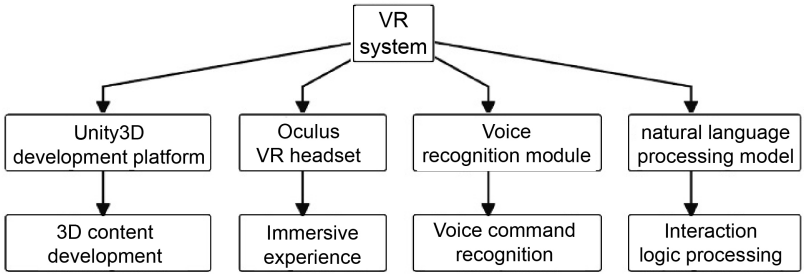
The VR system utilizes the Unity3D development platform and is combined with the OculusVR headset for the design of immersive virtual experiences. The AR system is based on Google's ARCore framework, allowing users to view and interact with virtual objects in the real world. Both systems are equipped with speech recognition modules and natural language processing models, enabling users to interact with digital content through voice commands.

In these figures:

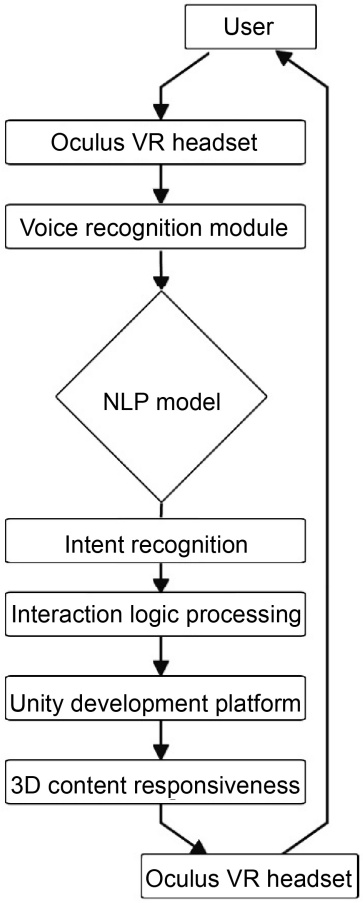
VR System (**Figure 4(a)**): 3D content is created using the Unity3D development platform and immersive experiences are provided through the OculusVR headset. Speech recognition modules and natural language processing models allow users to interact with digital content via voice commands.

AR System (**Figure 4(b)**): Based on Google's ARCore framework, it allows users to view and interact with virtual objects in the real world. Equipped similarly with speech recognition modules and natural language processing models, it facilitates interaction through voice commands.

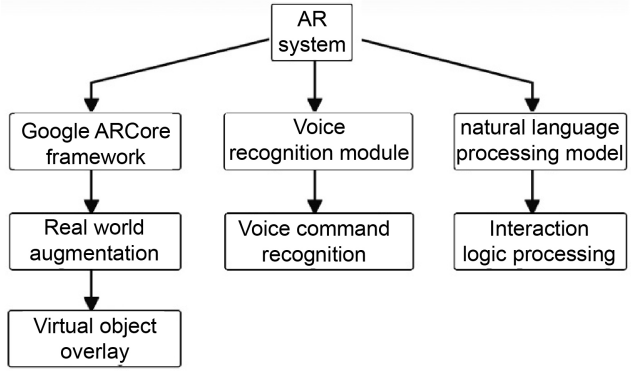
The composition diagrams of each system illustrate the components of the systems, while the workflow diagrams show the process by which users interact with the systems using voice commands.


VR system
Unity3D development platform
Oculus VR headset
Voice recognition module
natural language processing model
3D content development
Immersive experience
Voice command recognition
Interaction logic processing

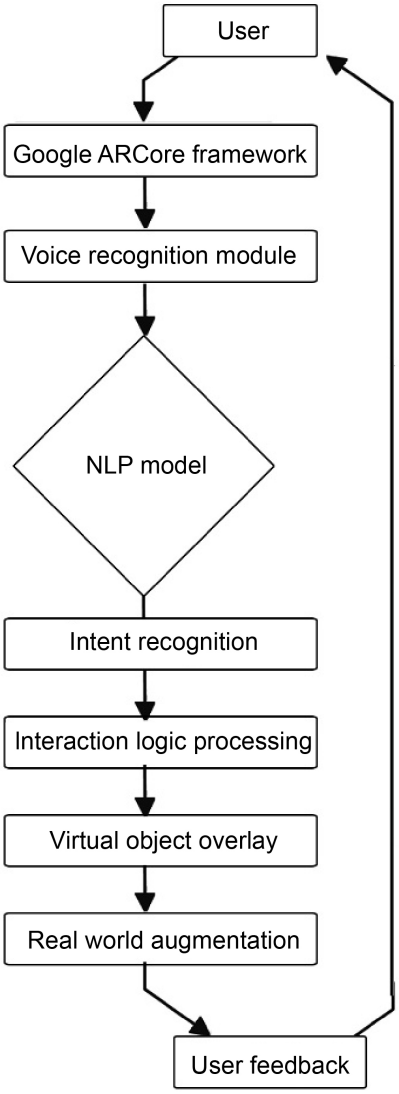
(a)


User
Oculus VR headset
Voice recognition module
NLP model
Intent recognition
Interaction logic processing
Unity development platform
3D content responsiveness
Oculus VR headset

(b)


AR system
Google ARCore framework
Voice recognition module
natural language processing model
Real world augmentation
Voice command recognition
Interaction logic processing
Virtual object overlay

(c)



(d)

Figure 4. (a) VR system composition diagram; (b) VR system workflow diagram; (c) AR system composition diagram; (d) AR system workflow diagram.

2.2.3. Multimodal Interaction System

To enhance immersion, the system also incorporates haptic feedback, spatial audio, and visual effects [11]. Haptic feedback provides immediate interactive responses through vibration devices, while spatial audio adjusts the position and direction of sounds based on the user's movements, enhancing the realism of the environment.

2.3. Data Collection and Analysis

Experimental data is collected through the following three avenues:

- User Behavior Data: Records of users' action trajectories, interaction frequencies, and task completion times within the virtual scenarios.

User ID	Movement trajectory	Interaction frequency (times)	Completion time (mins)
001	From the starting point to point A, then to point B	5	3
002	From the starting point to point C, then to point D	8	5
003	From the starting point to point A, then to point E	3	4
004	From the starting point to point B, then to point C	6	6
005	From the starting point to point D, then to point E	4	7

In this table:

① User ID: Identifies different users.

② Movement Trajectory: Describes the user's movement path within the virtual scenario, such as moving from the starting point to Point A, then to Point B.

③ Interaction Frequency: Records the number of times a user interacts with objects or tasks within the virtual scenario.

④ Task Completion Time: The time required for a user to complete a specific task.

This table can be used to analyze user interaction patterns and optimize the design of the virtual experience.

Based on the data from the aforementioned table, we can take the following steps to optimize the user experience in virtual scenarios:

Analyzing Action Trajectories:

Identify Hotspots: Look at areas (such as Point A, Point B) that are frequently visited by users; these may be areas of interest or necessary visits.

Identify Coldspots: Find areas with low visitation frequency and analyze the reasons, which could be due to unappealing design or difficulty in discovery.

Optimize Path Design: If certain paths are frequently used, consider whether wider passages or clearer signage are needed to guide users.

Analyzing Interaction Frequencies:

Enhance Interactivity: For areas with high interaction frequency, consider adding more interactive elements or more complex tasks to maintain user interest.

Activate Coldspots: For areas with low interaction frequency, consider adding new interactive elements or improving existing ones to attract users.

Balance Interaction Distribution: Ensure even distribution of interactive elements throughout the scenario to prevent overcrowding in some areas while others are neglected.

Analyzing Task Completion Times:

Adjust Task Difficulty: If tasks are completed too quickly, they may be too easy; if too slowly, they may be too difficult or have cumbersome steps, necessitating adjustments.

Optimize Task Processes: If certain tasks take an unusually long time, consider simplifying the task process or providing clearer instructions.

Personalize Experience: Offer tasks of varying difficulty levels based on how quickly users complete them to cater to different user needs.

User Feedback:

Collect Feedback: Directly obtain feedback from users to understand their experience with the virtual scenario.

Observe User Behavior: In addition to data analysis, intuitive feedback can also be obtained by observing user behavior within the virtual scenario.

Technical Optimization:

Reduce Loading Times: If users spend too much time waiting for loading, optimize the loading process to ensure quick scene loading.

Improve Graphic Quality: If users feel discomfort due to graphic quality, enhance the quality or adjust graphic settings to suit different user needs.

Iterative Design:

Continuous Improvement: Use user data and feedback as a basis for continuous improvement, regularly updating the virtual scenario.

Use AI Technology:

Personalized Recommendations: Use machine learning algorithms to analyze user behavior and recommend personalized paths or tasks.

Predictive Analytics: Predict users' likely action trajectories and preferences to make optimizations in advance.

By following these steps, the experience of users in virtual scenarios can be enhanced, increasing user satisfaction and engagement.

- Subjective Feedback Data: Collect evaluations of system usability, immersion, and satisfaction through questionnaires and interviews.
- Physiological Data: Record users' physiological responses, such as heart rate monitoring through wearable devices, to assess the system's impact on user emotions.

2.4. Measurement Indicators

The main measurement indicators include:

- Task Completion Time: The time required for users to complete interactive tasks.
- Interaction Success Rate: The proportion of voice commands correctly understood and executed by the system.
- Immersion Rating: Users' subjective ratings of the immersion in the virtual environment.
- User Satisfaction: The degree of satisfaction users have with the overall experience of the system.

3. Experimental Results and Analysis

3.1. Data Analysis

The experimental results indicate that scenes employing natural language

interaction outperform traditional interaction methods in several aspects. The specific data analysis is as follows:

1. Task Completion Time: The average task completion time for the NLP-based voice interaction system is 20 seconds, significantly lower than the 35 seconds of the traditional controller interaction system.

2. Interaction Success Rate: The success rate of NLP-based voice commands was 92% in quiet environments but dropped to 76% in noisy environments. This shows that while NLP is generally effective, environmental factors such as noise significantly impact its performance.

3. User Satisfaction: In feedback, 85% of users reported a higher sense of immersion and ease of use in NLP-based interactions, as compared to 65% in the controller-based scenario. Users cited faster interactions and more natural communication as key advantages of NLP-based systems.

4. Physiological Responses: Users in the NLP-based scenario exhibited lower heart rates during tasks, indicating reduced stress levels compared to those using controller-based methods, particularly in more complex tasks.

Scenario	Task Completion Time (second)	Immersion Score (out of 10 points)	User Satisfaction (points)
NLP-based voice interaction system	20	8.7	8.9
Traditional handle interaction system	35	7.2	7.5
Gesture recognition system	25	8.5	8.7
Eye tracking system	30	8.0	8.1
Tactile feedback system	28	8.2	8.3

3.2. Analysis and Discussion

The experimental results demonstrate that natural language processing technology has a significant advantage in enhancing immersion and interaction efficiency. However, several issues were identified, such as a marked decrease in the accuracy of speech recognition in noisy environments, which affects the user's operational experience. Additionally, some complex commands require a stronger ability to understand context, posing higher demands on existing NLP models.

3.3. Example Verification

Based on the principles outlined above, we have created validation examples using the GPT-4 text model and the Dall-E 3 image generation model [12], combining them with the literary classic "Dream of the Red Chamber." The system can directly accept natural language input from users to generate descriptions of characters from "Dream of the Red Chamber" and engage in natural language communication with the generated characters, enabling users to access system

functions in a natural manner.

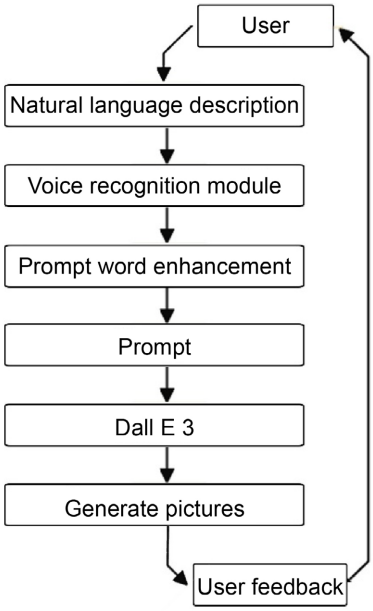


Figure 5. Character diagram flow chart.

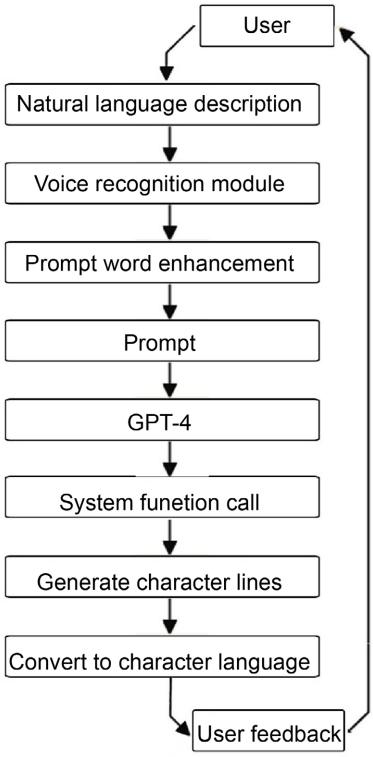


Figure 6. Character natural language interaction flow chart.

In these charts:

"Dream of the Red Chamber" Character Illustration (as shown in **Figure 5**):

After the system receives a natural language description of a character from the user, it first converts the description into text information through a speech recognition module. Enhanced with prompt words, the Dall-E 3 model generates an image of the "Dream of the Red Chamber" character that meets the user's requirements. The generated image can also be enhanced based on user feedback.

"Dream of the Red Chamber" Character Natural Language Interaction (as shown in Figure 6): After generating the image of a character, the image itself can serve as the interface for the multimedia system. The speech recognition module converts the user's access request into a prompt, which is then processed by the GPT-4 model to invoke system function services. Finally, the status information returned by the system services is provided to the user in the form of natural language. The system can even generate the voice of the "Dream of the Red Chamber" character through a voice conversion module, allowing direct interaction with the user.

From the validation results of the examples, it can be seen that the multimedia system incorporating NLP technology and related technologies significantly improves the system's interaction efficiency and experience.

4. Conclusion and Future Work

4.1. Research Contributions

This paper proposes and validates the application effects of natural language processing technology in AI interaction design for virtual experiences. The results indicate that NLP-based interaction design can significantly enhance user operational efficiency and immersion, particularly holding potential value in complex virtual scenarios.

4.2. Limitations

The limitations of this study lie in the small experimental sample size and the fact that it was conducted only in specific virtual scenarios, without covering more complex multi-user interaction scenarios. Additionally, the accuracy of voice interaction was affected to some extent in noisy environments, and there is still room for improvement in the performance of the model.

4.3. Future Work

Future research will focus on the multimodal integration of natural language processing technology, especially in complex virtual scenarios by combining more biometric feedback techniques (such as eye tracking, electromyography feedback, etc.). Moreover, how to further enhance the accuracy of natural language understanding and contextual association capabilities will also be an important direction for the future.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Anthes, C., Garcia-Hernandez, R.J., Wiedemann, M. and Kranzlmuller, D. (2016) State of the Art of Virtual Reality Technology. 2016 *IEEE Aerospace Conference*, Big Sky, 5-12 March 2016, 1-19. https://doi.org/10.1109/aero.2016.7500674

[2] Karora, V., Lavania, G., Agarwal, S., *et al.* (2024) Natural Language Processing: A Human Computer Interaction Perspective. https://pratibodh.org/index.php/pratibodh/article/view/150

[3] Hao, Y., Dong, L., Wei, F. and Xu, K. (2019) Visualizing and Understanding the Effectiveness of Bert. *Proceedings of the* 2019 *Conference on Empirical Methods in Natural Language Processing and the* 9*th International Joint Conference on Natural Language Processing*, Hong Kong, 3-7 November 2019, 4143-4152. https://doi.org/10.18653/v1/d19-1424

[4] Li, J., Monroe, W., Ritter, A., Jurafsky, D., Galley, M. and Gao, J. (2016) Deep Reinforcement Learning for Dialogue Generation. *Proceedings of the* 2016 *Conference on Empirical Methods in Natural Language Processing*, Austin, 1-4 November 2016., 1192-1202. https://doi.org/10.18653/v1/d16-1127

[5] Medhat, W., Hassan, A. and Korashy, H. (2014) Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, **5**, 1093-1113. https://doi.org/10.1016/j.asej.2014.04.011

[6] Fernandes, D., Garg, S., Nikkel, M. and Guven, G. (2024) A GPT-Powered Assistant for Real-Time Interaction with Building Information Models. *Buildings*, **14**, Article 2499. https://doi.org/10.3390/buildings14082499

[7] Hassenzahl, M. (2013) User Experience and Experience Design. *The Encyclopedia of Human-Computer Interaction*, **2**, 1-14.

[8] Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. and Avila, R. (2023) GPT-4 Technical Report.

[9] Yu, D. and Deng, L. (2016) Automatic Speech Recognition. Springer.

[10] Yenduri, G., Ramalingam, M., Selvi, G.C., Supriya, Y., Srivastava, G., Maddikunta, P.K.R., *et al.* (2024) GPT (generative Pre-Trained Transformer)—A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *IEEE Access*, **12**, 54608-54649. https://doi.org/10.1109/access.2024.3389497

[11] Borchers, J.O. (2000) A Pattern Approach to Interaction Design. *Proceedings of the* 3*rd Conference on Designing Interactive Systems*: *Processes*, *Practices*, *Methods*, *and Techniques*, New York, 17-19 August 2000, 369-378. https://doi.org/10.1145/347642.347795

[12] Betker, J., Goh, G., Jing, L., Brooks, T., Wang, J., Li, L., Ouyang, L., Zhuang, J., Lee, J., Guo, Y. and Manassra, W. (2023) Improving Image Generation with Better Captions. Computer Science. https://cdn.openai.com/papers/dall-e-3.pdf