Art and Performance Letters (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/artpl.2025.060102
ISSN 2523-5877 Vol. 6 Num. 1

Intelligent Interactive Space Art Based on the Needs of Smart Cities under Internet Technology

Yuanlong Tian[1,a], Shengnan Wang[2,b,*]

[1]School of Arts, Weifang University of Science and Technology, Shouguang City, Shandong Province, China
[2]Department of Integrated Arts, Silla University, Busan Metropolitan City, South Korea
[a]19862603283@163.com, [b]w13255364736@163.com
[*]Corresponding author

Keywords: Intelligent Space Voice Interaction System, Internet Technology, Voice Interaction, Smart City

Abstract: Relying on the development of smart cities, the art of intelligent interactive space has received extensive attention, forcing traditional industries to start the road of intelligent transformation. How to improve the intelligence level of spatial interaction has attracted much attention. In view of this problem, it is of great significance to study the intelligent spatial interaction method. The application research of voice interaction technology is gradually expanding in spatial interaction, and its performance advantages are crucial for solving intelligent transformation problems. This article aims to study the art of intelligent interactive spaces based on the needs of smart cities under internet technology, and also analyzes the construction of voice signal processing, voice interaction technology, and interaction systems. The results indicate that: The interactive space art embedded in this system has a higher user experience satisfaction score than the traditional interactive art, with a difference of 27.57%. It can be seen that the system can meet the needs of intelligent interactive space art, and the level of intelligence and user satisfaction have been greatly improved.

1. Introduction

Voice interaction technology plays an important role in various fields of daily habits, with significant effects in solving intelligent interaction problems and a wide range of applications. However, in the field of human-computer interaction, the application and research of voice interaction technology are relatively limited, and there is still significant room for development. Therefore, the study of using voice interaction technology to enhance the level of intelligent interaction space art is of great significance.

At present, with the continuous advancement of intelligent transformation under the concept of smart city, more and more scholars have explored intelligent spatial interaction. Among them, to achieve remote control of machines, Chen T investigated the predictive remote operation of computer numerically controlled (CNC) machines virtually controlled by gestures [1]. To assess user emotions in real-time systems, Hibbeln M discussed the use of human-computer interaction input devices to infer emotions [2]. Rozado D proposed the open source accessibility software FaceSwitch and showed how to help subjects with movement disorders to interact effectively with a computer hands-free [3]. Michalakis K discussed deploying the Internet of Things (IoT) on existing Internet infrastructure to extend this interaction by providing customized applications and services that allow human-computer interaction to be integrated into the automation of everyday life [4]. For more efficient interaction, Correia N N proposed Audio Visual User Interface (AVUI), a new type of UI that connects interaction, sound and image [5]. However, the intelligent level of the methods used for space art interaction is not high.

Voice interaction technology can be used in intelligent space interactive art, and has a good performance in the accuracy of speech recognition. Among them, in order to study the environmental factors affecting speech recognition in interaction, Birch B used the speech activation system he developed to conduct a multi-environment human-computer interaction test [6]. To improve the accuracy of speech recognition for speech interaction, Zhang H proposed a cross-modal speech-text retrieval method using an interactively learned convolutional autoencoder (CAE) [7]. Motta I conducted an online questionnaire survey of smartphone users and interviewed users of voice assistants to understand the timing and problems of users using voice assistants [8]. To improve the effectiveness of voice interaction, Cho E examined the effects of mode, device, and task differences on perceived human similarity and attitudes toward voice-activated VA [9].

In order to solve the above-mentioned problem of low intelligence of interactive space art, this paper uses voice interaction technology to analyze the space art voice interaction system, and simulates the algorithm and system to achieve the effect of improving user interaction experience. The innovation of this paper is: Using Internet technology, it analyzes how speech acquisition, speech recognition and text output in interactive systems play a role in the research and research of intelligent interactive space art based on the needs of smart cities under Internet technology. The proposed intelligent voice interaction system is expounded. Through experiments, it is found that the system runs stably, has strong intelligent interaction, and greatly improves the user interaction experience.

2. Method of Intelligent Interactive Space Art

2.1 Content and Organization of This Paper

The idea of smart city has been deeply rooted in the hearts of the people, and the traditional space art interaction method cannot meet the increasingly intelligent needs of users, so it is very important to improve the intelligence level of space art interaction [10-11].

This article proposes a study on the art of intelligent interactive space based on the demand for smart cities under internet technology [12]. This paper analyzes the performance of the speech acquisition and speech recognition algorithms and the system construction method, and improves the speech endpoint detection algorithm by combining multi-features to build an intelligent speech interaction system. The experimental results have shown that the interactive space art using the intelligent voice interaction system has a better experience on the interactive body than the ordinary interactive mode.

2.2 Construction of Intelligent Voice Interaction System

According to the collection of user interaction requirements, it can be obtained that since the establishment of the smart city concept, users' intelligent demand for space art interaction has been increasing day by day. They are not satisfied with the modal and simplified interaction experience of traditional interaction methods, and urgently need more intelligent interaction methods. As one of the most natural ways of interaction between humans and computers, voice has developed rapidly in recent years through artificial intelligence technology. Now, voice interaction technology is subtly changing people's living habits in various fields [13].

Aiming at the intelligent needs of users in interactive space art, this paper will build an intelligent voice interaction system using voice interaction related technologies [14]. It is applied to interactive space art in order to improve its intelligence level and promote the effect of user experience and satisfaction.

The intelligent voice interaction system goes through a series of processes such as voice signal preprocessing, voice recognition, semantic understanding, voice synthesis, and output from voice signal acquisition to voice output [15]. The process is shown in Figure 1:

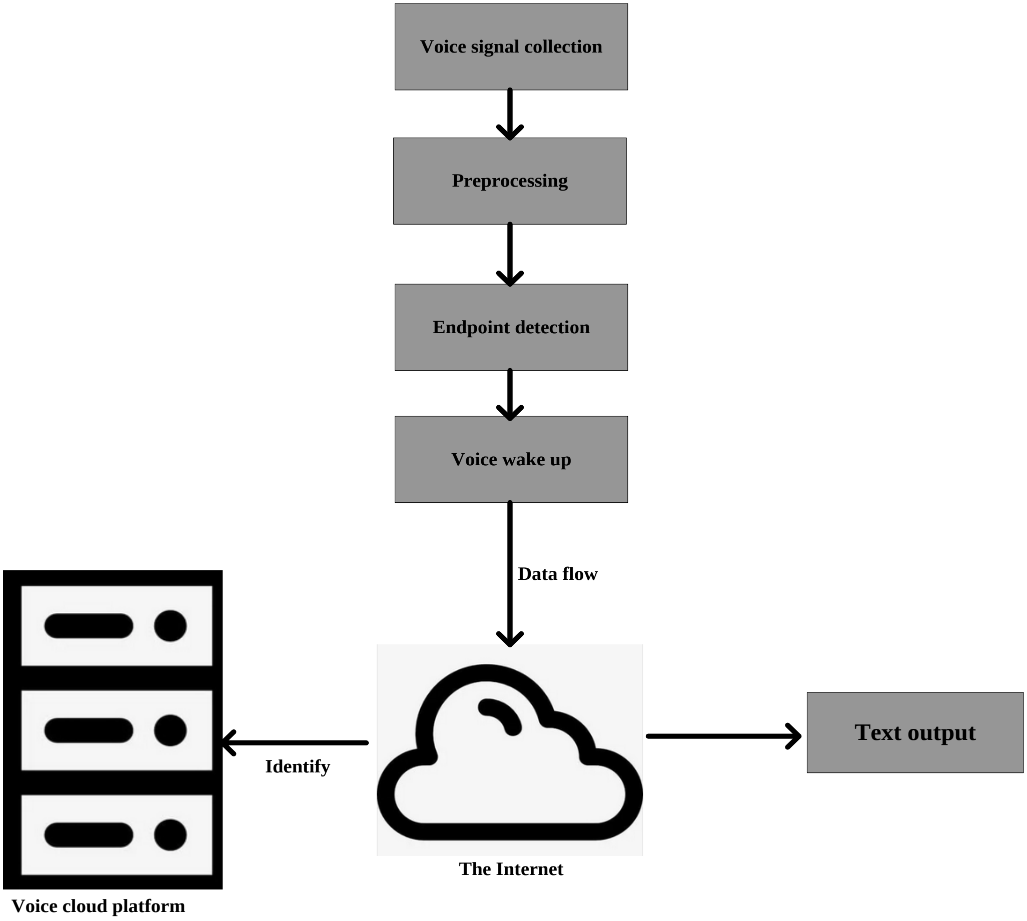


Figure 1: Speech recognition process

Preprocessing: In the processing of speech signals, good features are needed to obtain the expected results [16]. Operations such as pre-emphasis and noise reduction are necessary operations in the preprocessing stage. Its mathematical definition expression is shown in Formula (1):

$$Q(Z) = 1 - \alpha z^{-1} \tag{1}$$

Among them, the value of α is close to 1.

The sound is then segmented into frames and manipulated using the overlapping segmentation method. Then, each frame is weighted with an appropriate window function to obtain a windowed sound signal. The expression is shown in Formula (2):

$$e(w) = e(x) * c(x) \tag{2}$$

Among them, $c(x)$ is the window function; $e(x)$ is the speech signal of each frame. Choosing an appropriate function according to the actual situation can better reflect the effect.

Feature extraction: The short-term analysis method will be used for the characteristic analysis of the non-stationary signal such as speech [17]. The speech features are analyzed below.

The short-term average zero-crossing rate is the number of changes in the signal symbol of each frame sampling point of the signal, which changes with the frequency and can be used to roughly judge the spectral characteristics. Its definition is shown in Formula (3):

$$\mathrm{D}_n = \frac{1}{2}\sum_{m=1}^{N-1}\left|\operatorname{sgn}[e(m)] - \operatorname{sgn}[e(m-1)]\right| \tag{3}$$

Among them, $e(n)$ is the discrete signal; N is the effective length of the signal.

The symbolic function definition in the above formula is shown in Formula (4):

$$\operatorname{sgn}[e(n)] = \begin{cases} 1 & e(n) \geq 0 \\ -1 & e(n) < 0 \end{cases} \tag{4}$$

The short-term average energy can track the change trend of the speech signal, and the difference between the voiced frame and the unvoiced frame is obvious, so using this function can effectively identify the effective initial segment of the speech [18]. Its definition is shown in (5):

$$P_n = \sum_{m=-\infty}^{\infty}[e(m)c(n-m)]^2 = \sum_{m=n-N+1}^{n}[e(m)c(n-m)]^2 \tag{5}$$

Among them, $c(n)$ is the window function.

Mel frequency cepstral coefficient is also one of the audio signal characteristics, referred to as MFCC. The MFCC feature extraction process is shown in Figure 2:

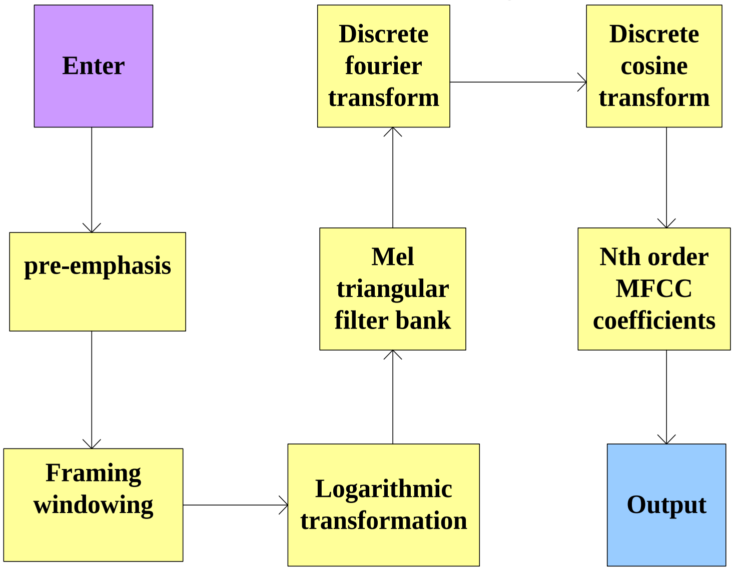

Enter
pre-emphasis
Framing windowing
Logarithmic transformation
Mel triangular filter bank
Discrete fourier transform
Discrete cosine transform
Nth order MFCC coefficients
Output

Figure 2: MFCC feature extraction

In order to average the contribution of each dimension and reduce the effects of audio signal distortion, further processing must be done using average normalization. After processing, it is shown in Formula (6):

$$S = (M - \overline{M}) * \frac{1}{C} \tag{6}$$

Among them, M is the feature of the current frame; $\overline{M}$ is the background noise feature; C is the standard deviation of the background noise feature.

In this paper, the L2 norm is used as the final judgment feature to calculate the L2 norm of each frame, and its definition is shown in Formula (7):

$$\|S\| = \sqrt{\sum_{i=0}^{n} S_i^2} \tag{7}$$

Among them, the order of features is n.

Endpoint detection: The function of endpoint detection is to detect the start and end points of the sound signal [19]. The performance of the algorithm is different for different characteristics. There are some problems in the traditional algorithm. Based on this, this paper improves the endpoint detection algorithm.

Since the change of noise in the actual living environment is impermanent, the buffer is used in this paper to track the change trend of the short-term zero-crossing rate, and the threshold update rule is shown in Formula (8):

$$Y_h = \overline{H} + \beta H_{std} \tag{8}$$

Among them, $H_h = \{H_0, H_1, ..., H_n\}$ is the buffer; $\overline{H}$ is the data mean; β is the update step; H_{std} is the standard deviation.

The threshold update rule of MFCC L2 norm is defined as Formula (9):

$$Y_m = \delta \overline{H} + \beta \tag{9}$$

Among them, δ, β is the adjustment factor; $\overline{H}$ is the mean value of the background noise norm.

Finally, the scanning idea is integrated in confirming that the speech is an effective segment, and the absolute difference method is used for comparison. The purpose is to solve the problem of various waveform shapes of short-term average zero-crossing under various noise conditions. This strategy can effectively reduce the false detection rate, reduce the number of system requests to the cloud, reduce system energy consumption, and improve system response speed.

This paper discusses the use of probabilistic and statistical modeling in speech recognition and text analysis for intelligent voice interaction systems. The speech recognition module uses Bayesian principles to predict sequential text outputs from speech input. Text analysis, using methods like TF-IDF and mutual information, extracts features to recognize user intent. The system integrates several modules through ROS, including audio preprocessing, cloud-based speech recognition, intent recognition, speech synthesis, and playback, enhancing human-computer interaction. The system's performance and stability are tested to optimize the user experience in interactive spaces [20].

3. Data Sources for Intelligent Interactive Space Art

The data in this paper is divided into two parts: questionnaire survey data and functional test data. Part of it is the current demand information of interactive space art users and the user experience evaluation information of interactive space art collected by issuing questionnaires to users.

In this paper, questionnaires were distributed to 300 users through online questionnaire surveys to collect information on the current interactive space art experience. The specific content of the questionnaire survey is shown in Table 1:

Table 1: Example of user experience survey table

Question1	How about the current interactive space art experience?	
Sequence	Answer options	
User 1	A	Very dissatisfied
User 2	D	Very satisfied
User 3	B	Average
User 4	C	Satisfied
Question2	What do you think is lacking?	
Sequence	Answer options	
User 1	B	Interactive experience
User 2	B	Interactive experience
User 3	D	Content
User 4	C	Interactive mode

Among them, question 1 collects information on the user's current experience satisfaction, and sets four answer options: very dissatisfied, average, satisfied, and very satisfied. Question 2 collects information for the shortcomings of current interactive space art, and sets four answer options: appearance; content, interaction method, and interactive experience.

In this paper, the collected 100 pieces of music and voice audio are processed and tested for the functions of each module of the system. The data parameters used in the test are as shown in Table 2:

Table 2: Test data parameters of module function part

Project	Data 1	Data 2	Data 3
Type of data	Music	Music	Chat
Length of time	58	16	5
File name	"My heart Will Go On"	"Bell"	"Greeting"
File format	Mp3	Mp3	Mp3

The data types that contain the test data are divided into two types: music and dialogue; the parameters also include the audio duration, file name and file format.

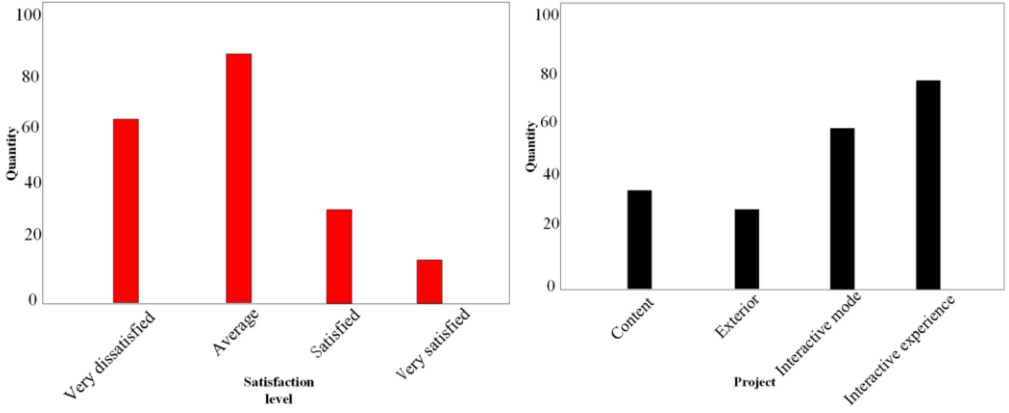
4. Results and Discussion of Intelligent Interactive Space Art

This paper firstly investigates the status quo of space art intelligent interaction through a questionnaire survey, and analyzes user needs to improve the space art interaction method and then improves the method. It builds an intelligent speech recognition system, tests the function of the system and then applies it to the interactive space art to test its effect. After the results are obtained, the results are analyzed and discussed below.

4.1 Intelligent User Interaction Requirements

This paper analyzes the demand and improvement direction of intelligent interaction after collecting information on the current interactive space art experience by issuing questionnaires to

300 users. The specific results are shown in Figure 3:



(a) Satisfaction situation

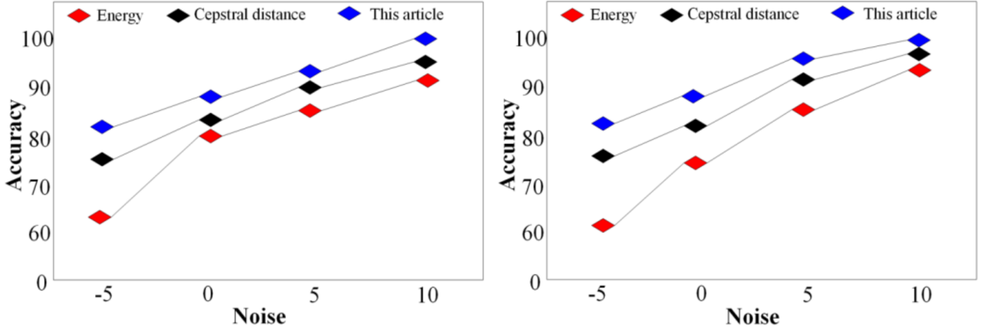
(b) Improvement needs

Figure 3: Current experience

As shown in Figure 3: In the survey of satisfaction with the current interactive space art experience, most users are not satisfied with the current experience, accounting for 63.4% of the total number of respondents; according to the survey data on the deficiencies in the current experience, most users believe that there is still room for improvement in the interaction method and interaction experience, accounting for 58.7%. It can be seen that the current needs of users cannot be met. The improvement of interactive space art should focus on improving the interactive mode and interactive experience, which determines the direction for the system construction.

4.2 Speech Recognition Algorithm Performance Test

In this paper, the dual-threshold algorithm based on energy and cepstral distance and the improved algorithm in this paper are tested in different noise environments. The algorithm is tested for its accuracy by performing effective audio screening on 1000 prepared audio pairs, both voiced and unvoiced. The audio files are separately processed by adding noise and divided into four groups of -5, 0, 5, and 10 to test the algorithm. The specific results of the test are shown in Figure 4:



(a) white noise environment

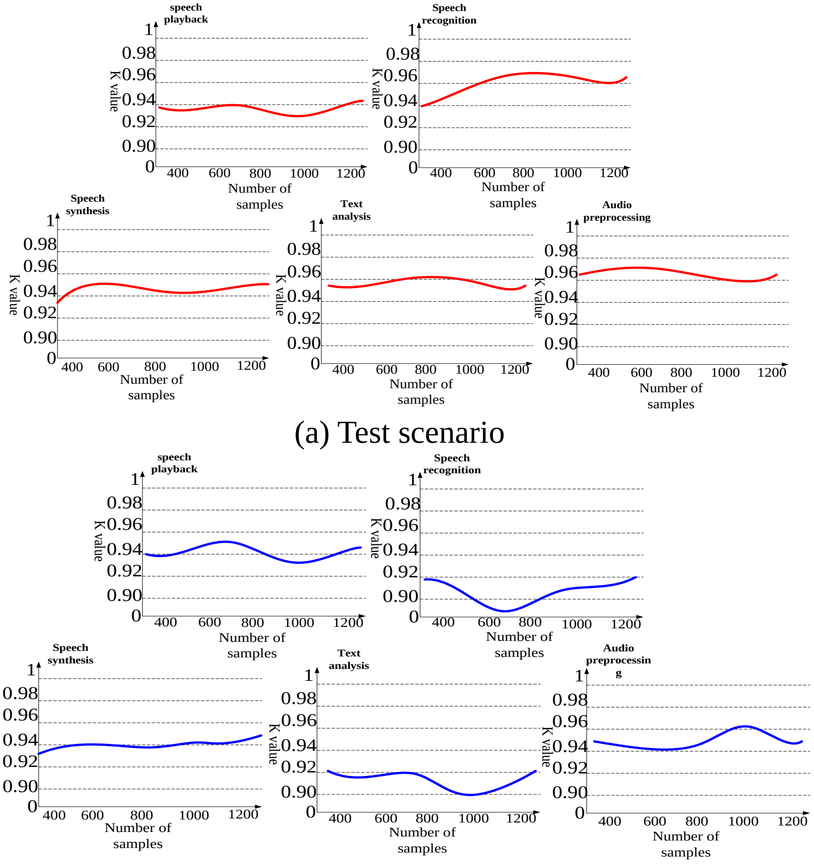
(b) pink noise environment

Figure 4: Accuracy of speech interception in different noise environments

As shown in Figure 4: In different test environments, the improved algorithm in this paper has improved accuracy in intercepting valid speech signals compared to the previous algorithm. Through the statistical calculation of the result data, the average interception accuracy of the energy-based dual-threshold algorithm is 80.3%; the average interception accuracy of the dual-threshold algorithm based on cepstral distance is 84.7%; the average interception accuracy of the improved algorithm in this paper is 91.6%. It can be seen that the improvement of the algorithm in this paper improves the accuracy of speech signal interception and obtains higher performance, which lays a good technical foundation for the construction of the subsequent system.

4.3 System Performance Test Evaluation

This paper integrates each module after the system construction is completed, and evaluates the function of each module in the test environment and the actual use environment according to the evaluation method specified in the discussion. The system function is evaluated by the normalized index coefficient k. The specific results are shown in Figure 5:



(a) Test scenario

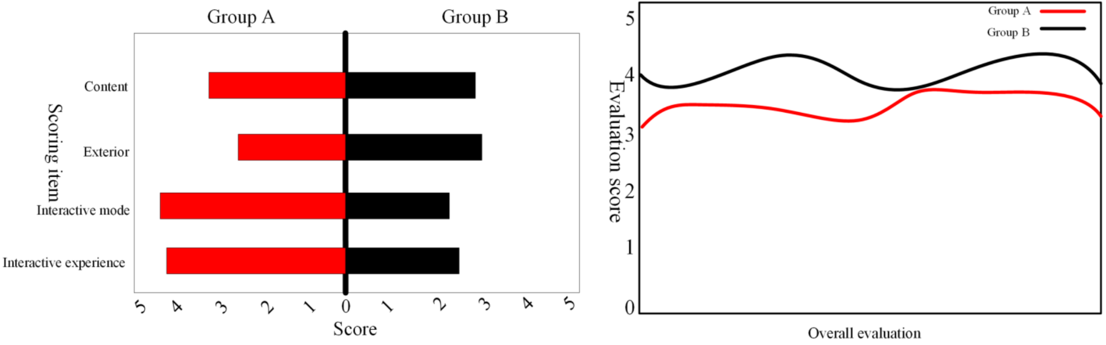
(b) Real scene

Figure 5: Performance evaluation of each module in system operation

As shown in Figure 5: the overall system runs smoothly and stably in the test and real scene use, but the evaluation index in the real scene is lower than the index in the test environment in terms of speech recognition and text recognition. This is because the actual environment in use is more complex than the test simulation environment, and there is a large gap between the data used for training and the input data in the real scene.

4.4 Satisfaction Evaluation

In this paper, 300 users are divided into 2 groups on average, in which group A is used for the interactive space art experience that integrates the intelligent voice interaction system constructed in this paper; group B is for the ordinary interactive space art experience, the time is 1 hour. After the experiment is over, the experience satisfaction of this interactive space art is evaluated, and the information is collected and counted for comparison. The specific results are shown in Figure 6:



(a) Comparison of the average satisfaction scores of each evaluation

(b) Overall satisfaction evaluation

Figure 6: Comparison of satisfaction ratings of different groups

It can be seen from Figure 6 that the users who have experienced the interactive space art integrated with the voice interaction system constructed in this paper have higher satisfaction feedback than the users in group B who have experienced ordinary space art. In contrast, the interactive space art that integrates the intelligent voice interactive system proposed in this paper is superior to the ordinary interactive space art in all aspects, especially in terms of interactive mode and interactive experience.

5. Conclusions

The development of interactive space art is inseparable from the contribution of the Internet and voice interaction technology. It can be seen that this intelligent voice interaction system integrated with Internet technology is superior to traditional interaction methods in all aspects. Through the test of each module of the system, the evaluation of each module is carried out, and the evaluation coefficient is above 0.9, which generally speaking, the system runs smoothly and has good performance. Through grouping experiments, the overall satisfaction score of group A using the system in this paper can reach 4.46 points, indicating that the system constructed in this paper can meet the user's intelligent interaction needs in the interactive space art experience.

References

[1] Chen T, Wang Y C, Lin Z. Predictive distant operation and virtual control of computer numerical control machines [J]. Journal of Intelligent Manufacturing, 2017, 28(5):1061-1077.

[2] Hibbeln M, Jenkins J L, Schneider C, Valacich JS, Weinmann M. How Is Your User Feeling? Inferring Emotion Through Human-Computer Interaction Devices[J]. MIS Quarterly, 2017, 41(1):1-21.

[3] Rozado D, Niu J, Lochner M J. Fast Human-Computer Interaction by Combining Gaze Pointing and Face Gestures [J]. ACM Transactions on Accessible Computing, 2017, 10(3):1-18.

[4] Michalakis K, Aliprantis J, Caridakis G. Visualizing the Internet of Things: Naturalizing Human-Computer Interaction by Incorporating AR Features [J]. IEEE Consumer Electronics Magazine, 2018, 7(3):64-72.

[5] Correia N N, Tanaka A. From GUI to AVUI: Situating Audiovisual User Interfaces Within Human-Computer Interaction and Related Fields [J]. EAI Endorsed Transactions on Creative Technologies, 2021, 8(27):1-9.

[6] Birch B, Griffiths C A, Morgan A. Environmental effects on reliability and accuracy of MFCC based voice recognition for industrial human-robot-interaction[J]. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 2021, 235(12):1939-1948.

[7] Zhang H. Voice Keyword Retrieval Method Using Attention Mechanism and Multimodal Information Fusion[J]. Scientific Programming, 2021, 2021(8):1-11.

[8] Motta I, Quaresma M. Opportunities and Issues in the Adoption of Voice Assistants by Brazilian Smartphone Users [J]. Revista ErgodesignHCI, 2020, 7(1):138-149.

[9] Cho E, MD Molina, Wang J. The Effects of Modality, Device, and Task Differences on Perceived Human Likeness of Voice-Activated Virtual Assistants [J]. Cyberpsychology, Behavior, and Social Networking, 2019, 22(8): 515-520.

[10] Matteo, Ribet, Marco, Sabatini, Luca, Lampani, et al. Monitoring of a controlled space flexible multibody by means of embedded piezoelectric sensors and cameras synergy[J]. Journal of intelligent material systems and structures, 2018, 29(14):2966-2978.

[11] Greenberg S, Honbaek K, Quigley A, Reiterer H. Proxemics in Human-Computer Interaction[J]. Dagstuhl Reports, 2018, 3(11):29-57.

[12] Shneiderman, Ben. Revisiting the Astonishing Growth of Human–Computer Interaction Research[J]. Computer, 2017, 50(10):8-11.

[13] Sreekanth N S, Narayanan N K. Multimodal Human Computer Interactionwith Context Dependent Input Modality Suggestion and Dynamic Input Ambiguity Resolution[J]. International Journal of Engineering Trends and Technology, 2021, 69(5):152-165.

[14] Devi N, Easwarakumar K S. A Clinical Evaluation of Human Computer Interaction Using Multi Modal Fusion Techniques [J]. Journal of Medical Imaging & Health Informatics, 2017, 7(8):1759-1766.

[15] Yuan Q, Wang R, Pan Z, Xu S, Luo T. A Survey on Human-Computer Interaction in Spatial Augmented Reality[J]. Journal of Computer-Aided Design and Computer Graphics, 2021, 33(3):321-332.

[16] Pradip Kumar Sharm, Seo Yeon Moon, Jong Hyuk Par. Block-VN: A Distributed Blockchain Based Vehicular Network Architecture in Smart City [J]. Journal of Information Processing Systems, 2017, 13(1):184-195.

[17] Pereira G V, Macadar M A, Luciano E M, Testa, M G. Delivering public value through open government data initiatives in a Smart City context [J]. Information Systems Frontiers, 2017, 19(2):213-229.

[18] Anthopoulos L. Smart utopia VS smart reality: Learning by experience from 10 smart city cases [J]. Cities, 2017, 63(3):128–148.

[19] Shin J G, Jo I G, Wan S L, Sang HK. A Few Critical Design Parameters Affecting User's Satisfaction in Interaction with Voice User Interface of AI-Infused Systems [J]. Journal of the Ergonomics Society of Korea, 2020, 39(1):73-86.

[20] Faramita R, Lestari D P, Niwanputri G S. E-commerce Design Interaction with Voice User Interface using User-centered Design Approach[J]. International Journal of New Media Technology, 2020, 6(2):104-108.