# Research on Speech Recognition System Based on Virtual Reality Technology

Jing Luo[(✉)]

Guangzhou Nanyang Polytechnic College, Guangzhou 510925, China
`85810733@qq.com`

**Abstract.** The rapid development of cloud things (cloud computing, Internet of Things, big data, artificial intelligence), especially the innovative development of virtual reality technology, has had a profound impact on school enrollment and employment promotion. Based on this, this paper briefly introduces the content of virtual reality technology, analyzes the application advantages of virtual reality technology in the panoramic campus, focuses on the application process of speech interaction in the virtual scene, and focuses on speech recognition module and semantic recognition module, hoping to provide useful reference for the followers in this field.

**Keywords:** Virtual reality technology · Panoramic campus · Speech recognition · semantics

## 1 Introduction

With the rapid development of cloud things (cloud computing, Internet of Things, big data, artificial intelligence) and the improvement of people's independent choice. In various industries in China, the application of virtual reality technology is gradually deepening in breadth and depth. The expansion of all kinds of information, as well as the expansion of the intensity and mode of campus recruitment publicity, has gradually integrated into the life of the public. However, the content and mode of publicity are not different from each other, which makes it difficult for the target students to distinguish the situation of the school and the design of the major, especially in the remote schools. As a result, the students will feel dissatisfied after arriving at the school and drop out, resulting in the loss of the school and students. Therefore, AR technology is integrated into the panoramic campus. Through interactive and immersive viewing mode, students can understand the school, major, accommodation and life information, optimize the audience's feelings on the digital campus, and improve the enrollment rate of students.

In order to solve the problem of students browsing campus, this paper proposes a speech recognition system based on virtual reality technology. The voice recognition system can identify the semantics of users, open corresponding virtual scenes according to the semantics, and browse the campus, which is conducive to improving the efficiency of users using the virtual campus. Voice guidance can help users quickly complete the experience by avoiding complicated operation buttons.

## 2  Related Work

### 2.1  The Nature of Virtual Reality Technology

Virtual reality technology is a computer-generated simulation environment, which can integrate a variety of information, at the same time can be human-interactive virtual multidimensional dynamic scenes, but also can be integrated into the simulation system of personal behavior, so that the audience can be integrated into the environment [1].

Virtual reality (VR) mainly includes simulated environment, perception, natural skills and sensing equipment. The simulation environment is a real - time dynamic campus 3D image generated by computer. Perception refers to the perception of hearing, touch, force and motion in addition to the visual perception generated by computer graphics technology. Natural skills refer to the human head rotation, gestures or other human behavioral actions, which are processed by the computer and the data corresponding to the participants' movements, and the real-time response to the user's input, and respectively fed back to the user's facial features. Sensing equipment refers to three-dimensional interactive equipment [2].

### 2.2  Advantages of Using Virtual Reality Technology in Panoramic Campus

- Cost saving

Students need to know about the school and learn about the campus with the help of virtual reality technology. They don't need to spend any energy. They only need to prepare virtual reality equipment to learn about the campus and major on the official website. Schools can also set up virtual reality experiences at enrollment points, so that students can personally understand the school, and improve the popularity and recognition rate of the school.

- Integration of media and people

With the help of virtual reality technology, the computer and human neural network will be connected together, and the development of the man-machine interface from the non-sensory communication to the perception, smell and touch knowledge will eventually be incorporated into the human ontology. The humanization of virtual reality technology will eventually reflect the perfect state of "unity of nature and man" [3].

- Rich forms

In the traditional enrollment design scheme, parents and students have a relatively simple understanding of the school information, usually through the website.

The information obtained in this way is relatively one-sided and incomplete. Consumers cannot analyze and judge the information source, and it is difficult to fully understand all the contents of the information source. With the support of virtual reality technology, parents can break through the limitations of traditional vision, and the sensory experience extends from the original hearing and vision to touch and even smell. Through the feedback and collation of various sensory systems, we can get a comprehensive cognition of the campus.

### 2.3  Selection of Panoramic Campus Areas

Guangzhou Nanyang Polytechnic Vocational College is selected as the campus. First, all-round three-dimensional display of campus environment style, improve the visibility and visibility of the school. The second is to solve the campus planning, in the virtual environment, feel the reasonable space and the applicability of the design, reduce the modeling cost and shorten the design time, so that the building auditor immersive all-round review, more accurate.

### 2.4  Speech Recognition Technology

Voice recognition is used to control the virtual scene terminal. This speech recognition uses a speech recognition Engine based on HUAWEI HiAI Engine to provide AI application layer API to developers. This technology can convert speech files and real-time speech data streams into Chinese character sequences with an accuracy of more than 90% (95% for local recognition) [4].

Voice input requirements are as follows:

a) *use wav or PCM audio format.*
b) *the current client support identification of mandarin.*
c) *speech length less than 20 s.*
d) *client hardware devices using: the sampling rate of 16000 Hz, mono.*

## 3  Application Problems and Countermeasures of Virtual Reality Technology

### 3.1  Application of Virtual Reality Technology

- Contradiction between virtuality and reality

    The virtual scene is consistent with the campus scene, but the virtual reality technology will be affected by the personnel, with personal subjective emotions, resulting in different judgments. Irrational use will lead to false publicity and bring bad negative effects to the school. The audience has personal subjective feelings, which leads to the error of judgment and the possibility of feeling cheated. So, in these cases, you need to create regulatory authorities, supervision and control, to prevent a crisis of confidence.

- Combination of panoramic content and virtual reality technology

    By combining aerial photography with panoramic technology and virtual reality (VR) technology, and connecting with the 3D intelligent cloud platform of panoramic campus, the campus features that the school needs to display (such as student living area, business district, canteen, leisure and entertainment area and characteristic teaching area, etc.) are copied on the Internet by means of panoramic shooting, 3D multi-dimensional display and building virtual exhibition hall. Realize the 720° online rotation of the real scene display, 720° scene without visual dead Angle for virtual tour, at the same time, personalized asteroid opening way and custom navigation buttons can improve the visual

experience. All-round three-dimensional display of campus environment style, improve the visibility and visibility of the school [5].

- Current Technical Problems

Virtual reality input and output devices are mainly used to realize the interaction between users and virtual scenes, including stereoscopic display, three-dimensional sound restoration, action mapping between real people and virtual people, haptic feedback, etc. Interaction is a way of communication between users and virtual environment. Users' choice and operation of objects in virtual scene and control of simulation process must be realized through human-computer interaction. The Razer Hydra and STEM systems currently in use work well in many ways, but ultimately do not mimic the user's hands. Currently, no fixed place is provided to complete the VR device experience.

### 3.2 Countermeasures of Virtual Reality Technology

- Display of proprietary platforms

The virtual reality platform is mainly composed of three interactive systems, namely individual social interaction, virtual reality equipment and virtual reality environment. The coordinated management of the three systems and the internal data is conducive to improving the application effect of virtual reality technology in the design of preset scenes and information data transmission [6].

Functions and interfaces provided by Vega are used for secondary development to realize interactive operation in virtual environment and experience immersive feeling. In practical applications, a reasonable virtual reality platform is built, in which individuals are transformed into a series of information data that can be analyzed and tracked, and the needs and characteristics of the audience are understood through data analysis, providing reliable data support for the subsequent platform optimization [7].

- Legal Constraints

The school provides specifications for the use of virtual reality equipment, formulates user manuals, and provides scientific and standardized guidance for the reasonable application of this technology.

## 4  Speech Recognition System Based on Virtual Reality Technology

After the completion of the campus virtual scene, professional personnel are required to cooperate with the visit. In order to reduce the operator's work burden during the visit, voice recognition system is used to control the scene selection. Through voice control, non-contact scene selection can be adopted, which is conducive to the non-touch access at the present stage.

### 4.1  Architecture of Speech Recognition System

The voice recognition interaction system is set on the virtual scene selection screen. First, input the speech, detect the length of the speech, and record the effective speech.
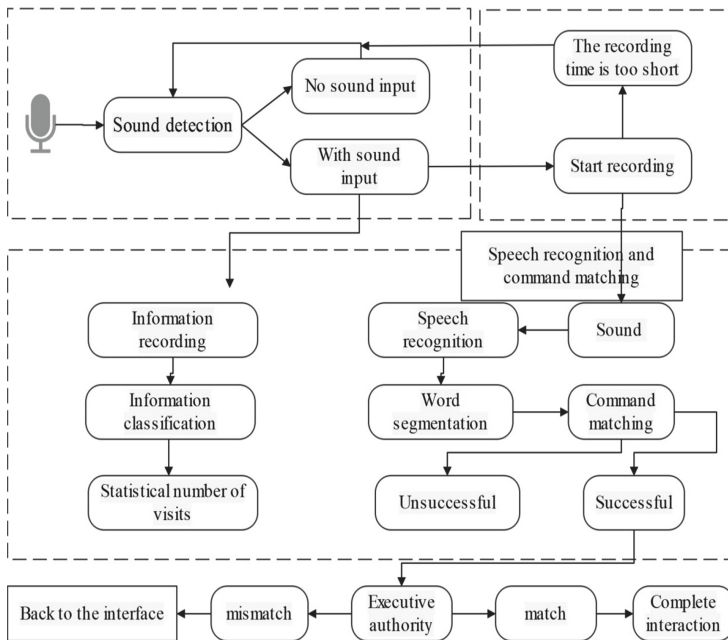
**Fig. 1.** Overall architecture of speech recognition

Secondly, the speech recognition software of iflytek is used for speech recognition. The recognition is carried out by word segmentation. Keywords are selected for command matching. If no, return to the screen and start a new voice input. This is shown in Fig. 1.

## 4.2 Analysis of Main Modules of Speech Recognition Interactive System

Speech recognition interaction is an important entry point of virtual scene. Such intelligent recognition interaction can avoid touch operation and improve operation efficiency [8].

- Administrator voiceprint recognition module

The administrator is responsible for virtual scene interaction. The administrator enables the virtual scene interaction platform to facilitate special management, follow-up information collection, and daily reception statistics.

Each person has different vocal cords, which produce different sound rating and different voicing print [9]. Voice print is a biological characteristic unique to each person, and voice recognition is a relatively reliable way of identity authentication. The specific recognition process is shown in Fig. 2.

First, the administrator's voice print is input to form a voice print library. The feature of voice print is extracted to form the feature database, and the model is trained to generate the model database. When the administrator's voice is input, the feature is extracted and
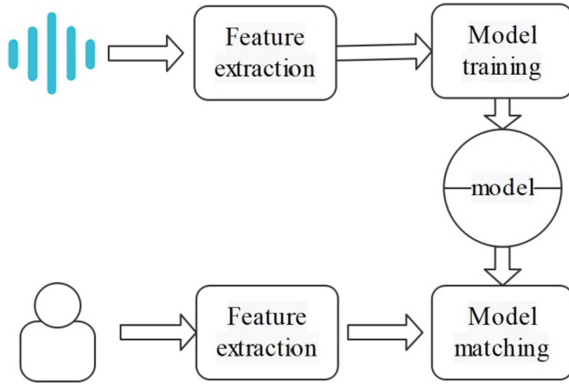
**Fig. 2.** Sound recognition process

**Table 1.** Semantic extraction process of words

| Operation procedure | Detailed procedure |
| --- | --- |
| voice | Get me a virtual scene in the cafeteria |
| Extract words | Action - open, place - canteen, target -virtual scene |
| command | Opening canteen scene |

matched with the model library. After matching, all permissions are unlocked and the scheduling of virtual scenes starts.

- Word command matching

The interaction of virtual scene is controlled by voice, which needs to extract semantic understanding. Firstly, the speech is confirmed, the semantics are mined, keywords are determined, and interactive commands are formed [10]. As shown in Table 1.

### 4.3 Simulation Experiment

This virtual scenario has been used in enrollment for two years, and the information obtained is shown in Table 2.

Data in Table 2, people are very concerned about the teaching environment, teaching staff and accommodation on campus. The overall number of interaction shows that people like this way of voice interaction and choosing virtual scenes to learn about the campus.

**Table 2.**  Frequency of voice interaction in virtual scenarios

|  | Teaching scene | Canteen scene | Library scene | Motion scene | Honor scene | Dormitory scene |
|---|---|---|---|---|---|---|
| 2022 summer enrollment | 8500 | 5800 | 2100 | 2000 | 9000 | 4500 |
| 2022 spring enrollment | 6000 | 6000 | 1500 | 1500 | 3000 | 6000 |
| 2021 Summer Enrollment | 6000 | 5000 | 1500 | 1000 | 3000 | 4000 |
| 2021 spring enrollment | 7500 | 6000 | 1500 | 1500 | 1500 | 3000 |

## 5   Conclusions

From the previous aspects of the research situation, the use of virtual reality technology in the panoramic campus is gradually emerging. From the theoretical point of view, the application of virtual reality technology in the panoramic campus needs more technical support and equipment support, as well as the support of various platforms.

Based on the analysis of speech recognition system, this paper introduces the work of speech interaction in virtual scene in detail from the whole architecture, administrator voice print recognition and word command matching. Therefore, speech recognition system based on virtual reality technology has strong applicability.

## References

1. Sun,Y,Z.(2020)Research on Cultural Communication of Chinese Traditional Ancient Architecture based on Virtual reality technology, Xi 'an Polytechnic University, Xi 'An.
2. Wang, J.(2022) Implementation of 3D Virtual Campus Scene Based on Sketchup and ArcGIS.Information and Computer (Theoretical Edition), 34(03):110–112.
3. Zhang, X.(2020)Evolution and Trend of intelligent Teaching Research in China in Recent Ten years. Distance Education in China. 09:123-124
4. Zhou,N.(2020)Research on UAV Speech Control Scheme and Simulation Based on HMM and RNN. Journal of System Simulation.03:88-85
5. Tian, G. (2020)Analysis on the Status quo and future prospects of Domestic media "VR+ News". 04:253–254
6. Shen,Y. (2020) Virtual Reality: A New Chapter in the Development of Educational Technology -- Interview with Professor Zhao Qinping, Academician of the Chinese Academy of Engineering. Research on Audio-Visual Education. 01:25-26
7. Nie,S,J.(2019) Application of VR technology in the construction of Sponge City in Nie Er Park. China Water Supply and Drainage. 12:214-214

8. Tao, J,H.(2022) Overview of Multi-mode Human-Computer Interaction. Journal of Image and Graphics. 06:146-148
9. Chen, X,Z.(2021)A Double-Threshold Speech Endpoint Detection Algorithm. Journal of Zhejiang Polytechnic of Industry and Trade. 02:198–199
10. Zhang,H,N(2021) Research on Speech Recognition Method in Strong Noise Environment Heilongjiang University,The Heilongjiang.