

Intelligent Voice Augmented Reality Interactive

Xinyi Song, Junjie Zhang

University of Shanghai for Science and Technology, Shanghai, China, 200090

Abstract: Voice enables people to transmit information better and more quickly, and people can control all kinds of machines to communicate and work by intelligent voice. This paper intends to use intelligent voice to achieve new cloud classroom teaching. The effect that the teacher can move the picture in real time through voice control and reply accordingly can be achieved by speech synthesis, speech recognition and voice interaction technology. The efficiency of the classroom is improved while the interest of the classroom has been enhanced.

Keywords: Intelligent voice; Cloud classroom; Voice recognition ; Voice interaction;

1. Introduction

With the development of information technology, intelligent voice products are gradually integrated into all aspects of people's life, such as intelligent speakers, intelligent watches, intelligent learning machines and so on. Intelligent voice technology has become the most convenient and effective means for people to obtain and communicate information. At present, intelligent speech technology mainly includes four key links. VPS signal processing, ASR pattern recognition, NLP natural language processing and TTS speech synthesis.^[1] Based on this, this paper puts forward the concrete scheme of applying intelligent speech in cloud classroom teaching.

2. Status of the study

Language is a unique communication tool between people at the beginning. With the continuous development of science and technology, human can communicate with machines through language and convey their own instructions, and machines will do the corresponding work according to the instructions. In recent years, intelligent machines have developed rapidly, and more and more exchanges with machines are made in daily life. Voice recognition technology is an important part of intelligent machines, which enables people and machines to communicate better. Intelligent voice interaction is a new type of interaction mode based on voice input, which can receive feedback results by communicating with machines. The traditional speech interaction system mainly focuses on speech recognition and speech synthesis, which can be found from the current products. In most recognition tasks, the audio used is often sampled in strictly restricted environment, which eliminates the influence of external environment on audio. However, there are some problems such as noise interference, low dialect recognition rate and poor semantic understanding fault tolerance in real audio environment, which affect the accuracy of speech interaction system to varying degrees.^[2]

3. Design scheme of voice recognition system

Cloud classroom teaching camera based on speech recognition system needs to be combined with specific

application scenarios. Combined with the current online teaching background, the design and optimization of two different operation links of teacher-side voice control and student-side voice control are carried out, and two design schemes that can be used in voice recognition system are discussed and analyzed.

3.1 The first Scheme , student voice control system

The student-side voice control system is mainly the teacher-side camera uploads the shooting content to the web page in the form of live broadcast or video recording. The student-side can realize the function of scaling and moving through the picture panel of the voice control web page, as shown in Figure 2-1.

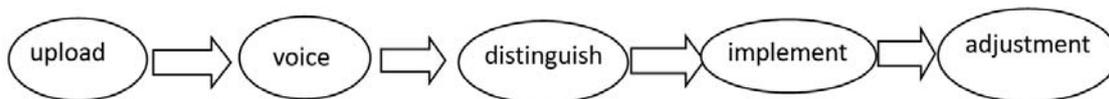


Figure 2-1 Student-side voice control system

By transmitting classroom information to the web page in real time through the cloud, students can adjust the panel content by themselves, obtain the key content information more flexible and free, and can adjust according to their own learning needs.

3.2 The second Scheme ,Teacher-side voice control system

The teacher-side voice control system mainly controls the mobile behavior of the camera by the teacher’s real-time voice. Its main functions are shown in Figure 2-2.



Figure 2-2 Teacher-side voice control system

The teacher-side voice control system can be directly controlled by the teacher in real time according to the content. By sending the corresponding voice instructions, the camera can realize the functions of scaling and moving, which can further clarify the key content of the explanation, and transmit the page more intuitively to the student side. Finally, the classroom content will be sent to the cloud for teachers or students to watch again. Thus improving the quality of teaching and enhancing students’ learning efficiency. The advantage is that the overall function is easy to realize. The use experience is optimized, and the teacher-side voice control system has a wide range of applications.

The second scheme is selected for design and implementation by Considering synthetically.

4.The key technology and main features

4.1 Human-computer interaction technology

The various functions of the camera are waked up and controlled by the voice module. The voice module has the highest priority.

After power on, the camera screen defaults to the center of the blackboard, and gives voice prompt. ‘Hello, I am Xiao Zhi, welcome to Intelligent Cloud class.’ In view of the problem of classroom discipline, instructions except for power on are “answered” by flashing lights. The instructions are as follows.

- “Xiaozhi, move to the left.”
- “Xiaozhi, move to the right.”
- “Xiao Zhi, enlarge the picture .”
- “Xiao Zhi, narrow down the picture .”
- “Xiao Zhi, switch the picture to the blackboard .”
- “Xiao Zhi, switch to PPT.”

When the power of microphone is lower than 15%, voice prompt will be given. ‘The battery is low, please charge it in time.’ After class, the teacher can turn off Xiaozhi by voice command “turn off Xiaozhi camera” or key.

4.2 Voice recognition technology

Speech recognition technology is referred to as ASR technology. It is to transform voice information into text information that can be recognized by the machine. The construction of voice recognition system includes two stages. The one is Data training stage, the other is model matching recognition stage. In the data training stage, the single chip microcomputer analyzes and processes the collected voice samples, extracts the voice feature information from them, and establishes a feature model offline. In the voice recognition stage, the user’s voice data will be automatically matched and recognized, which is usually completed online.

4.3 voice synthesis technology

The essence of voice synthesis technology is to transform text information into voice information. This design uses Alibaba cloud speech synthesis platform and cb5654 development board to extract the corresponding voice parameters from the recorded sound library. In the synthesis process, we make the sequence decision of context-related HMM training by analyzing the input text, then send the generated voice into the parameter synthesizer, and finally output the synthesized voice.

4. System process

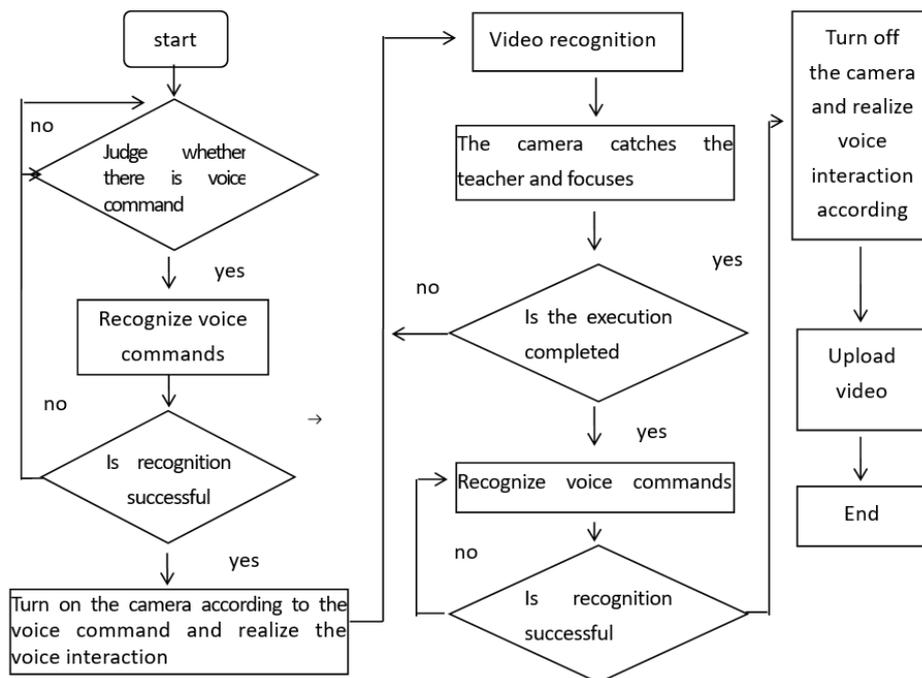


Figure 4-1 system flow chart

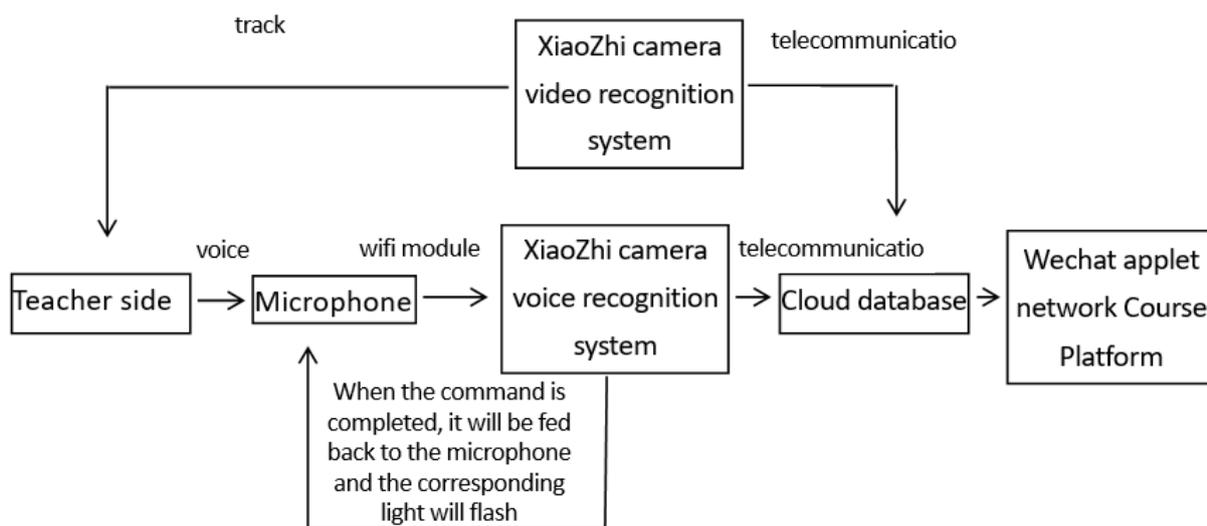


Figure 4-2 connection of each module

When the camera receives voice command (such as “XiaoZhi, move left”), it will execute the command in the way of moving. The teacher’s position is identified by target detection technology, and the blackboard’s position is identified by edge detection technology. After video recognition, it will rotate under the action of the steering gear until clearly captured the classroom picture. It will waiting for the next voice instruction to be issued after execution.

The voice instructions sent by the teacher side are received by CB5654 development board, and sent to the upper computer through the serial port to control the switch of the camera. The upper computer records the video, saves the video to the cloud database or local, and uploads it to the wechat applet.

5.voice recognition test

The teacher side sent out the voice instruction of “class begins” and CB5654 development board received and replied “class begins, please listen carefully”. The teacher side sent out the voice instruction of “class is over”, and cb5654 development board received and replied “class is over, goodbye teachers and students”.

Voice wake-up, voice capture and intelligent conversation are tested at different decibels. The environment below 30dB corresponds to classroom, the environment below 60dB corresponds to home, and the environment above 60dB corresponds to game room test. The results show that the noisier the environment is, the worse the recognition effect is. However, the probability of completing the whole intelligent conversation remains above 85%. The results are shown in the table below.

Table 1 voice test results

Decibels	Test items	Number of tests	Number of error	probability
≤30db	Voice wake up	300	0	100%
	Voice capture	300	2	99.33%
	Intelligent dialogue	300	2	99.33%
≤60db	Voice wake up	300	0	100%
	Voice capture	300	5	98.33%
	Intelligent dialogue	300	7	97.67%
60db≤	Voice wake up	300	23	92.33%
	Voice capture	300	27	91%
	Intelligent dialogue	300	36	88%

6. Conclusion

In this paper, intelligent voice technology is used to achieve new cloud classroom teaching, the teacher controls the movement of the picture changes by voice commands, which can accurately and quickly convey the classroom content to the students. Students can quickly distinguish the key content of the classroom, thus greatly improving the efficiency of learning. At the same time it can also carry out voice interaction. Active classroom atmosphere can also be invigorated by carrying out voice interaction. After that, through the information interaction of camera, database, web page or wechat applet, students can watch the classroom content again through wechat applet after class, consolidate knowledge points, and truly achieve intelligent and efficient teaching. Due to the information interaction involving different functional modules, in order to facilitate users to manage different classroom teaching information uniformly, it is necessary to establish a database in the cloud server for secure information interaction.^[3] However, The development language and environment of different functional modules are different, so it is difficult to realize the interaction between various modules, and the development cycle is long.

Reference:

1. Hao Ya Ya, Wu Xuan, Liu Rong Kai. Development status and application prospect of intelligent speech recognition technology [J]. Electroacoustic technology, 2020,44 (03): 24-26
2. Gao Xintao, Chen guaili. Development status and application prospect of speech recognition technology [J]. Gansu science and technology journal, 2007 (04): 13 + 76
3. Wang Bin, Wang Yujun, Cui Jianwei. Progress of intelligent voice interaction technology [J]. Artificial intelligence, 2020 (05): 14-28