# Application of Speech Recognition Technology to Virtual Reality System

## Zhu Zhongxiang[a], Li Shanshan, You Jiali, Song Zhenghe

Beijing Key Laboratory of Optimal Design of Modern Agricultural Equipment, China Agricultural University, Beijing, 100083, China

[a]zhuzhonxiang@cau.edu.cn

**Abstract.** Appling the speech recognition technology into the virtual reality system can not only expand the application of speech recognition in scene roaming, but also make up the shortage of the interaction of virtual reality software, and improve the efficiency of interaction between users and the virtual environment. This paper aims to combine the speech recognition technology with the VR technology, and control the users' viewpoint in the VR system by speech. Using Microsoft Speech SDK5.1, the speech recognition program is developed and the interface is designed to connect the speech recognition module with the VR software. Based on the EON SDK, the EON nodes of speech recognition and scene roaming are programmed. The above EON nodes and the built models are imported into EON. Running the speech recognition program and then the scene roaming controlled by speech is realized eventually. The experiment on walking mazement shows that the speech commands can exactly control the users' motion to avoid obstacles and successfully go through the mazement.

## Introduction

As a multidisciplinary and emerging technology, virtual reality (VR) possesses the characteristic of immersive, interactivity and imagination [1]. Users can actively interact with virtual objects in a virtual environment via interactive devices such as 3D mouse, data gloves, head mounted display and position tracking device. Among them a data gloves is commonly used together with a position tracking device for the virtual assembly of mechanical parts and virtual driver training [2-5]. Machine vision sensor and Kinect inductor are also usually applied to virtual reality system for man-machine interaction, which can recognize the posture of hand or body during control operation [6] [7]. In addition, some researchers applied the speech command to drive the spatial scene roam in the virtual environment [8] [9]. Lin Mingxiao applied the speech recognition technology based on Speech SDK5.1 to the 3D simulation platform and proposed a method for dynamic word recognition [10]. Overall, the existent interactive devices used in virtual reality system are mainly depending on the motion of human body, and users rarely communicate with the virtual environment by speech, which, in some extent, reduced the actor's immersion in the virtual environment.

In this paper, Microsoft Speech SDK is used for the software development of speech recognition. Meanwhile, a secondary development of EON nodes is implemented to realize the roaming control by speech.

## Speech Recognition Technology

Speech recognition technology is a comprehensive discipline involving acoustics, physiology, digital signal processing, pattern recognition, artificial intelligence etc., which can convert the human words into the computer-readable input, and give computer the ability to understand the human spoken language.

The speech recognition technology can be divided two types: one is based on a specific person and the other is non-specific. The former is restricted by speaker, namely, the speaker's voice is learned firstly and identified later. Therefore, the speech recognition process (Figure 1) contains the

training stage and the recognition stage. The aim of the first stage is to establish the acoustics model of a speaker [11].
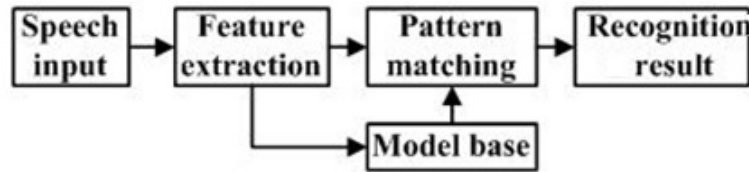


Fig.1. Flow of speech recognition for specific person

### Introduction to microsoft speech SDK

The Microsoft Speech SDK is a perfect tool for developing speech software, which is composed of Speech Application Programming Interface (SAPI), Microsoft Continuous Speech Recognition (MCSR), TTS, etc. The SAPI's APIs (Application Program Interface) take the COM as the interface standard, thus, application developers can easily use speech recognition technology without concerning about the complex speech technology, which greatly reduces the programming code and make the program more robust.

The functions of speech recognition are implemented together by four interfaces of the SAPI, namely, IspRecognizer, IspRecoContext, IspRecoGrammar and IspPhrase. The workflow of SAPI is very normative. For the most application, according to the results of the needs analysis, the speech application program can be assembled by the interfaces in the workflow order.

In this case, the Microsoft Speech SDK is used to identify the commands of a roamer, in which the speech recognizer is trained with the specific speaker's voice first before being used to recognize his or her speech.

### Development of speech recognition program

Before programing the speech recognizer, it is necessary to take working procedure of the speech recognition into the operation mechanism of Windows application. Finally, the speech recognizer can run in parallel with the Windows application. The detail programing steps are listed below.

(1) The grammar file (XML) which defines the commands that needed to recognize is compiled.

(2) During the initialization of Windows application, the functional code is added to initialize COM; each speech interface is instantiated in a particular order; recognition engine and recognition context are created; the recognition grammar and message are set up, and then the recognition grammar file is loaded.

(3) In the callback function of Windows application, the speech recognition message is added, and the result is obtained from the interface of IspPhrase. Then the recognition result is mapped to corresponding control command and the command is executed.

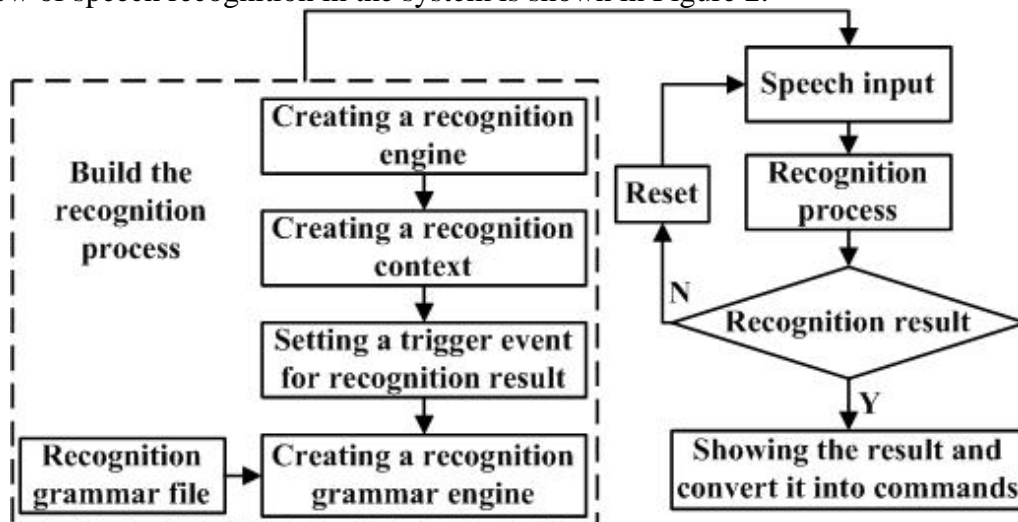The flow of speech recognition in the system is shown in Figure 2.



Fig.2. Algorithm flow of speech recognition

The corresponding relationship between the speech inputs, the recognition results and the controlling functions is defined as shown in Table 1. The speech recognition program is coded and the man-machine interface is shown in Figure 3.

Table 1 Corresponding relationship between the speech inputs, the recognition results and the functions

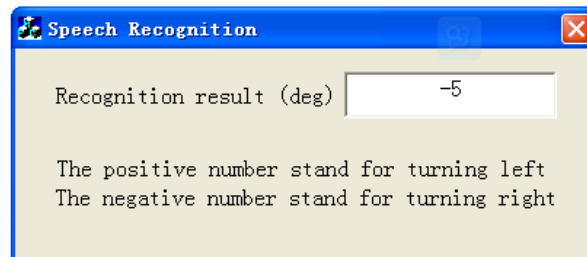| Speech inputs | Recognition results | Controlling functions |
|---|---|---|
| Turn bridle | -10 | Turn about |
| Turn right | -5 | 10 degrees to the right |
| Turn left | 5 | 10 degrees to the left |
| Go straight | 10 | Go forward |
| Stop here | 15 | Stop |



Fig.3. Display interface of speech recognition result

## Creation of Speech Node with EON SDK

EON Studio is a design tool totally based on graphical user interface and used to develop the real-time 3D multimedia applications of training and roaming in virtual reality.

### Interface between speech recognition program and EON

The method of shared memory is chosen as data communication tool between the speech recognition program and the speech node (EON node). The speech recognition program runs independently outside of EON, and writes the recognition result to a shared memory. On the other hand, the speech node created with EON SDK reads the recognition result in the shared memory and outputs the result to the roaming node.

### Development of interface program

In the design scheme of shared memory, the writing and reading of data in the shared memory are two different processes. The data is written into the shared memory by the speech recognition program and is read from by the speech node.

The basic of the shared memory is the memory-mapped-file (MVF) technology. Creation of a MVF is actually to make the file to associate with some areas of the virtual space in the process. To establish the association, a file-mapped object should be firstly created and the file-mapped-object view is obtained. Then the contents of the file can be checked and edited by the pointer of the above view. When the data is written into the file, the system is responsible for the cache, buffer, writing and call-in of data and the distribution and release of memory [12].

To set up and access a file by using memory mapping, the programing steps are listed as follows. (1) Call the CreateFileForMapping function to open or create the memory-mapped file; (2) Use the CreateFileMapping function to create an object in memory and to tie the object to the file that is opened by the CreateFileForMapping function; (3) Use the MapViewOfFile function to create a view of the memory-mapped object; (4) Use the pointer that is returned by the MapViewOfFile function to gain direct access to the memory-mapped object.

## Roaming Control Based on Speech Recognition

### Mazement modeling

The 3D maze model is built by UG and exported into a PARASOLID file. Then the PARASOLID file is converted to a *.flt file using the self-designed conversion software. Importing the *.flt file into Creator, and surface merging and textures mapping are conducted to get the final model (Figure 4).

### Development of roaming node

A roaming node is developed based on EON SDK to move or rotate the camera, namely viewpoint. Then the roaming control by speech commands in the virtual environment is realized. The design flow is shown in Figure 5.
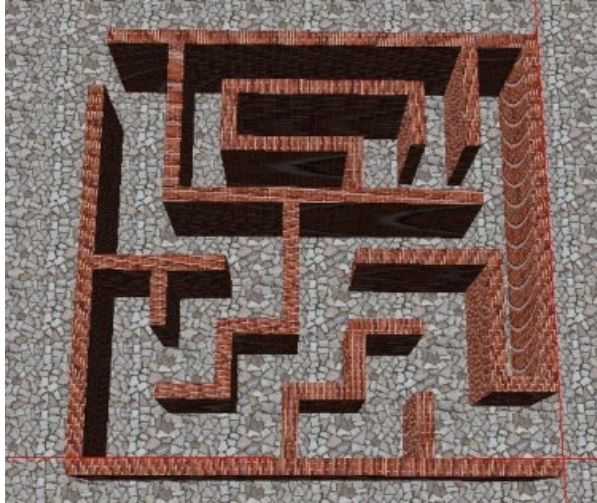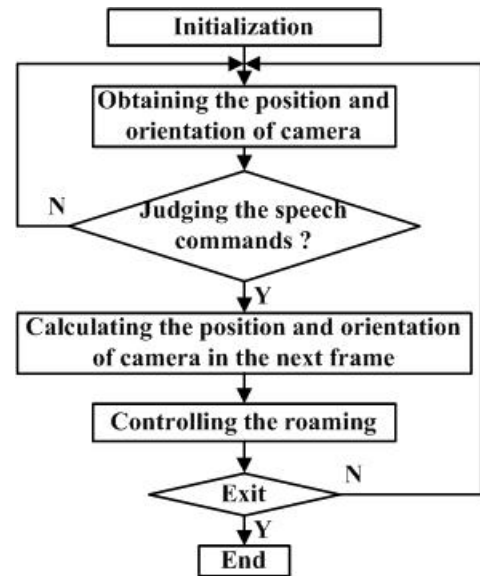


Fig.4. 3D mazement model



Fig.5. Algorithm of the roaming node

### Roaming control with speech recognition

The scene roaming system with speech control is designed, as shown in Figure 6. The three-dimensional mazement model with *.flt format is imported into EON, and the speech node and the roaming node are also added into and connected logically. Figure 7 shows the simulation tree of the EON program. When the EON program process starts and the speech recognition program process is also in running status, the observer can control the roam in the virtual scene with a microphone or other audio equipment. In order to validate the function of roaming controlled by speech, a path recording node is added into the EON program to record the roaming path. The recorded roaming path is then added into a new EON program to playback the roaming process as shown in Figure 8.
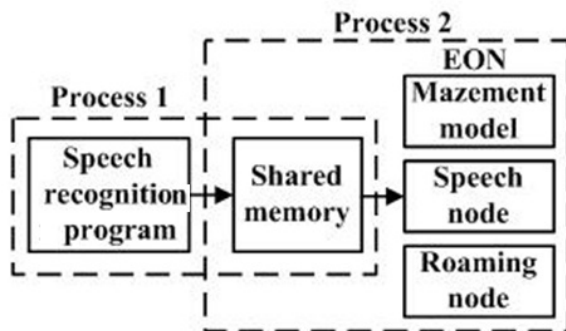

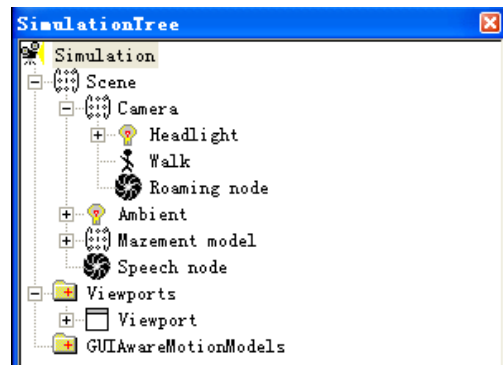
Fig.6. Scene roaming system with speech control
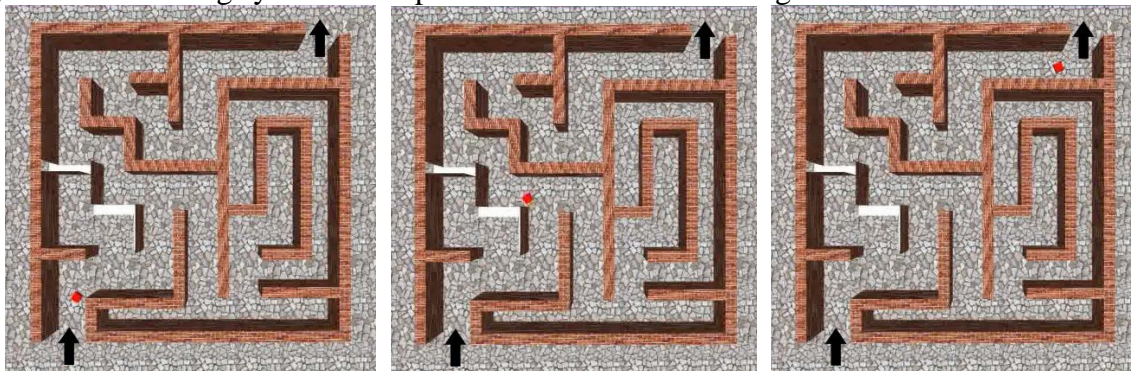


Fig.7. Simulation tree



Fig.8. Playback of the roaming in mazement

## Conclusion

Based on the Speech SDK 5.1 and EON SDK, the speech recognition program, speech node and roaming node are developed. Accordingly, the roaming control in virtual reality is realized. The experiment on walking the mazement indicates that the user can walk through the mazement by speech command. At the same time, the speech roaming system has a high accuracy and good real-time.

## Acknowledgement

## References

[1] Xu Wei. The domestic and foreign research and development on virtual reality technology. Modern Business Trade Industry, 2009, (2): 279-280.

[2] Xu Deyou. Hand gesture-based interaction in virtual reality training system. Journal of System Simulation, 2006, 18(S2): 386-389.

[3] Li Shanshan, Song Zhenghe, Chen Yueyang, Mao Enrong, Zhu Zhongxiang, Chen Li. Virtual assembly for mechanical gearbox of agricultural chassis. Transactions of the Chinese Society of Agricultural Machinery, 2012, 43(S1): 318-322.

[4] Zhu Zhongxiang, Chen Li, Li Shanshan, Liu Bo, Song Zhenghe, Mao Enrong. Critical techniques of virtual assembly for combine harvester chassis based on virtual reality. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(S2): 262-267.

[5] Wang Junfeng, Li Shiqi, Liu Jihong, Cao Pengbin. Application of multi-modal interaction technology in virtual disassembly. Computer Aided Engineering, 2004, 13(4): 16-19.

[6] Zhou Lai, Zheng Danli, Gu Hongbin, Wang Debao. Research of visual interaction for virtual reality flight training. Acta Aeronautica et Astronautica Sinica, 2013, 34(10): 2391-2401.

[7] Lu Ming, Wang Zhenshui, Tian Yuan, Li Lin. Posture interaction tool for virtual reality system based on Kinect. Journal of System Simulation, 2013, 25(9): 2124-2130.

[8] Wang Fei, Liu Huiyi, Wang Chunfeng. Vagile system based on speech interaction in virtual scene. Journal of System Simulation, 2009, 21(5): 1384-1386, 1389.

[9] Jin Zhu, Ma Xiaoping, Kan Hongwei. Scene navigation with speech interface based on Java3D and JSAPI. Microcomputer Information, 2007, 23(12-3): 178-179.

[10] Lin Mingxiao. Application of speech recognition technology in 3D simulation based on Speech SDK. Computer Technology and Development, 2011, 21(11): 160-162, 166.

[11] Zhu Runhua, Wang Peijun. The virtual assembly system research based on web and speech recognition. Machinery & Electronics, 2008, 04:66-67.

[12] Wan Ming, Zhang Fengming, Wang Xuefeng. Virtual instrument design based on MVF. Chinese Journal of Scientific Instrument, 2006, 11(27): 1542-1545.