A Speech Interaction Method in Immersive Virtual Maintenance

Yuxue Jin[1,2], Chuan Lv[1,2] and Jie Geng[1,2*] and Di Peng[2]

[1]State Key Laboratory of Virtual Reality Technology and System, Beijing, 100191, PR China

[2]School of Reliability and Systems Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, PR China

[*]Corresponding author

Abstract—**Human-computer interaction, which involves interactive information processing, directly affects the vividness and immersion of virtual operation. Therefore, this paper aims to achieve the naturalization and intelligence of human-computer interaction. We present two language understanding models constructed by knowledge database and case database. Firstly, architecture of voice interactive system is generally introduced. And then, expression of operation object's feature information and fuzzy recognition of operator's intention are at a discussion. At end, experiments have been carried out to verify the significance. Based on the steps above, we develop a methodology of immersive virtual maintenance interaction based on speech interaction technology which can be more accurate and effective.**

Keywords-human-computer interaction; speech recognition; immersion virtual maintenance

I. INTRODUCTION

At present, immersive virtual maintenance inter-operability is an important issue of concern. The existing human-computer interaction methods [1] are mostly traditional mouse-keyboard. Therefore, when repairing in accordance with the maintenance procedures, operators often need auxiliary help owing to hands occupation [2]. Therefore, it is necessary to propose an optimized interactive way to make the maintenance personnel more fluid and efficient.

Rapid development of speech recognition technology in many areas, such as mobile phones and automotive and others, can lead to a more efficient interaction. In 3D immersion virtual maintenance environment, the main task of people is to complete the assembly and maintenance operations[3]. Therefore, both hands are occupied. Therefore, conveying the intention and feedback information through the voice command, to a large extent, makes human-computer interaction more direct and efficient.

II. DEMAND ANALYSIS IN IMMERSIVE VIRTUAL MAINTENANCE

Immersive virtual maintenance interactions [4] can be divided into two parts: interaction between user and system interaction as well as operation simulations in maintenance process, figure 1 shows the details. This paper analyzes the requirements of each type of voice interaction and lays the foundation for the constructing voice interactive system of immersive virtual maintenance.
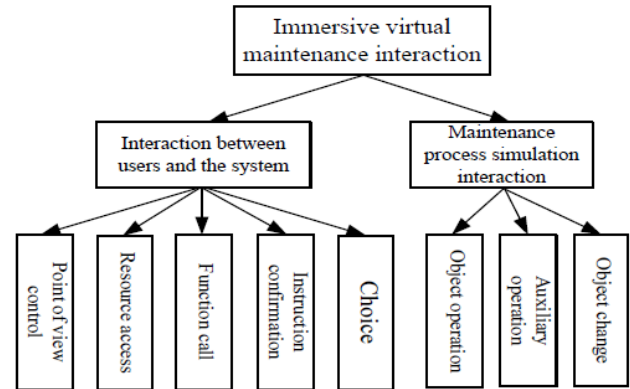

FIGURE I. INTERACTIVE REQUIREMENTS CLASSIFICATION IN IMMERSIVE VIRTUAL MAINTENANCE

1) Viewpoint control. Viewpoint rotate or translate in accordance with the specified direction to a certain speed. To implement control function, system needs to obtain three aspects of information including object, predicate and adverbial.

2) Resource access. Users access and read the various types of resources to assist in operations completion of virtual human maintenance. Resources include electronic maintenance manuals, maintenance tools, maintenance, and other hand tools. The voice commands for resource access include two parts including perform actions and manipulate objects. Interactive voice instructions are structured by "verb predicate + noun object" constitute.

3) Function call. A function call is a user's call to a function in a immersive virtual maintenance system. Corresponding features include 3D maps, visibility analysis, reachability analysis, fast upper limb (RULA) analysis, and operational spatial analysis. Function call voice control command structure and resource access are similar, can be designed as "verb predicate + noun object" structure.

4) Instruction information confirmation. When users send voice commands to system, the system is necessary to confirm the user's order to ensure the correctness of the implementation of the action. Feedback voice statement structure is "Do you want to say '**' ?" The users can select" confirm "or" cancel ".

5) Function selection. When operating, users need to make some selection of features, such as gesture library

selection, maintenance tool selection and so on. Implementation of selecting function can be done through the voice interactive. When system voice prompt "Please select the option you need", the users need to enter the corresponding voice options.

6) Maintenance object operation. The operation of the maintenance object includes three types of auxiliary operation, object operation and scene object change.

On the basis of things above, by analyzing the characteristics of immersive virtual maintenance interaction [5], interactive function is divided into two major categories 8 small classes. Stages of simulation preparation, maintenance simulation and evaluation are analyzed. Finally, interaction needs and corresponding behaviors are summarized, and the total interactive demand summary table is obtained.

III. VOICE INTERACTIVE SYSTEM ARCHITECTURE

The relationship between the functional modules of the voice interactive system about immersive virtual maintenance is shown in Figure 2.
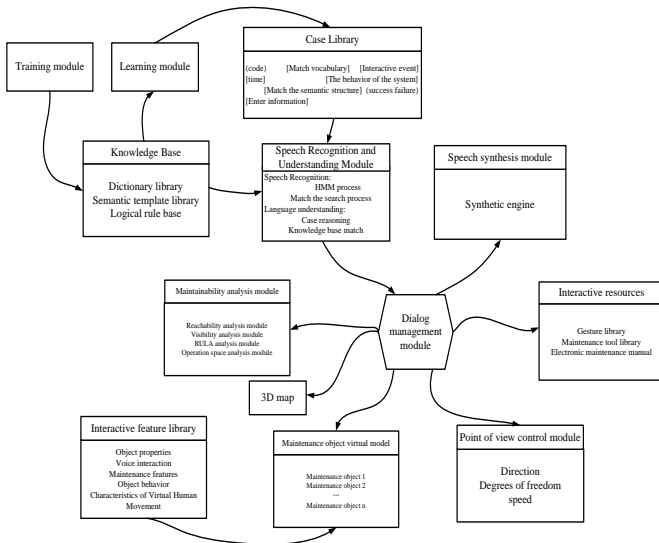

FIGURE II.   FUNCTION MODULE STRUCTURE DIAGRAM OF SYSTEM

IV. LANGUAGE UNDERSTANDING MODEL BASED ON KNOWLEDGE DATABASE AND CASE REASONING

The language understanding model is based on case database and knowledge base. The knowledge base contains the dictionary library, semantic template library and the logic rule library. Case library contains examples of success and failure recognition, including recognition input, output and evaluation. The understanding process can be shown in the following figure.
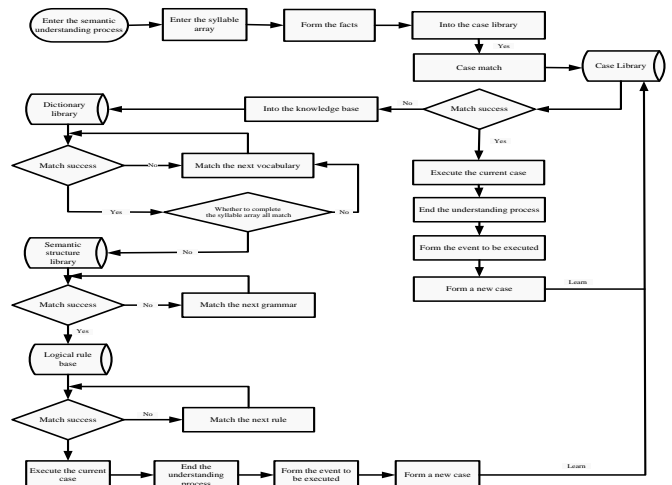

FIGURE III.   FLOW CHART OF LANGUAGE UNDERSTANDING PROCESS BASED ON KNOWLEDGE BASE AND CASE REASONING

A. Language Understanding Model Based on Case Reasoning

A very important part of the language understanding model based on case reasoning is the similarity calculation between syllable sequence and information input in the case database. The calculation method constructed in this paper takes character repetition and sequence consistency of syllable sequence as the parameters, then they are integrated by certain weights.

Let A1 denote the syllable sequence in the source case, and A2 for the syllable sequence of the target case. The length of the syllable sequence can be expressed by Len( ). SameNum (A1, A2) represents the number of repetitive syllables of A1 and A2. Morphological similarity SimF (A1, A2) of A1 and A2 can be expressed as:

$$SimF(A1, A2) = 2 \times \frac{SameNum(A1, A2)}{Len(A1) + Len(A2)} \quad (1)$$

Let OneNum (A1, A2) indicate characters in A1 and A2 appearing only once. Lnv (A1, A2) represents the vector of the position number in the sequence A1.Snv (A1, A2) represents position sequence in A2 corresponding to A1. OrdNum (A1, A2) represents the number of reverse orders of the adjacent components of Snv (A1, A2).The word order similarity SimO (A1, A2) of the syllable sequences A1 and A2 can be expressed as:

$$SimO(A1, A2) = \begin{cases} 1 - \dfrac{OrdNum(A1, A2)}{OneNum(A1, A2) - 1}, & OneNum(A1, A2) > 1 \\ 1, & OneNum(A1, A2) = 1 \\ 0, & OneNum(A1, A2) = 0 \end{cases}$$

$$(2)$$

Finally, the similarity of syllable sequences A1 and A2 is weighted by word similarity and word order similarity. Let the

weight of the word similarity be α, the weight of the word order similarity be β, the similarity Sim (A1, A2) can be expressed as:

$$Sim\left(A1,\ A2\right)=\alpha SimF\left(A1,\ A2\right)+\beta SimO\left(A1,\ A2\right) \tag{3}$$

For syllable sequence similarity matching, the importance of the similarity of characters is much greater than the word order similarity. Generally, take α 0.9, β 0.1.

Whether the case can be reused need to judge according to the size of the similarity. Given the threshold for the similarity of the case, if the similarity between new case and old case is greater than the given threshold, the matching result of the new case can be directly called old one.

B. *Language Understanding Model Based on Knowledge Database*

Based on knowledge base, the semantic analysis is based on the matching search, which can be described in the model language.

The knowledge base consists of the dictionary library W, the semantic template library WS, and the logical rule base LR.

The model can be represented by a six-tuple array, as shown in the following equation:

$$\prod = \{A, P, PP, CC, CP, I\} \tag{4}$$

In this model, A represents syllable string processed by the previous step, which is the input information of the model based on the knowledge; P presents preprocessor, which preprocess string A; PP refers to the lexical meaning analyzer, which convert vocabulary description into text; CC is a semantic parser, and the user's intent is obtained by matching the new lexical sequence and the semantic template library one by on; CP is the interactive behavior processor.

V. EXPERIMENTAL ANALYSIS OF SPEECH INTERACTION

FIGURE IV.   VOICE INTERACTION EXPERIMENT

In order to verify real-time and accuracy, this paper analyzes the accuracy and response time of voice command recognition. According to different functional categories, five voice interactive commands were selected as the experimental objects. Twenty experiments were performed for each voice interactive command. The accuracy of each recognition result was calculated. The results are shown in Table 1.

1 represents recognition result is correct, 2 is not correct. Line graph analysis about results above is shown in Figure 4. Inaccurate parts are concentrated in the first 10 experiments. Therefore, study can make the identification results be accurate, which verifies accuracy of the method proposed in this paper.

TABLE I. VOICE INTERACTIVE COMMAND RECOGNITION ACCURACY STATISTICS

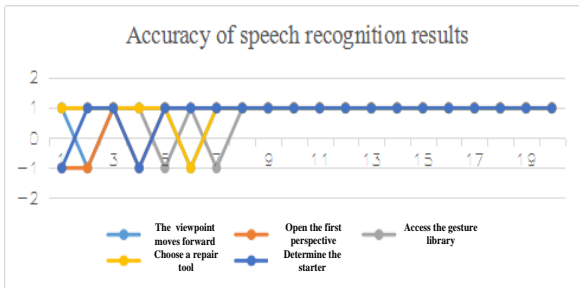| Interactive command	Whether the recognition result is correct (Y / N)																			
The viewpoint moves forward	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Open the first perspective	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Access the gesture library	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Choose a repair tool	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Determine the starter	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

FIGURE V.   ACCURACY ANALYSIS OF SPEECH RECOGNITION RESULTS

In addition, in order to verify the real-time nature of voice interactive system, the experiment [6] compares the working mode of the interactive mode with the traditional keyboard/mouse in the starter reachability analysis process.The virtual authentication environment of the two interactive modes and the operating rules are the same[7]. By recording interactive times, it is verified whether the real-time performance of the voice interactive system is optimized for the traditional keyboard/mouse interaction mode.

In the voice interaction experiment, through the voice interactive command "open/close the simulation tool" to control the accessibility analysis tool, in the process, with 6s. However, in traditional way, the operator with 5s, support staff with 3s. It can be seen, the voice interaction time is less than the mouse/keyboard interaction hours. Therefore, the method put forward is more convenient through the way of voice interaction, and can meet requirements of interactive real-time.

VI. CONCLUSION

To meet the requirements of interactive application, this paper studies the way to improve interactive efficiency of

immersive virtual maintenance by combining with engineering application. The advantages are summarized as follows:

1) Expressing the user's operational intent through the voice instruction, rather than other technical means which constantly guess the user's intention, saves computing rate;

2) Hearing and language are two important ways of interaction between man and world. Besides, the goal of human-computer interaction is to achieve the interaction of human and computer as much as possible. By introducing voice interactive technology into immersive virtual maintenance operation environment, a more natural human-computer interaction is achieved;

3) Owing to the complex interface and function of immersive virtual maintenance interaction, voice can obtain the information needed quickly and accurately from a large amount of information.

REFERENCES

[1] Jie Geng, Dong Zhou, Chuan Lv, Zili Wang, A Modeling Approach for Maintenance Safety Evaluation in Virtual Maintenance Environment, Compute-Aided Design, 2013,45(5): 937–949

[2] Jie Geng, Chuan Lv, Dong Zhou, Ying Li, Zili Wang, Compensation-based methodology for maintenance time prediction in a virtual environment, Simulation Modelling Practice and Theory,2014,47: 92–109

[3] Karutz, C.O. and J.N. Bailenson, Immersive Virtual Environments and the Classrooms of Tomorrow, The Handbook of the Psychology of Communication Technology, (2015)33:290.

[4] Marcos Paulo Alves de Sousaa, Manoel Ribeiro Filho, Marcus Vinícius Alves Nunes, Andrey da Costa Lopes. Maintenance and operation of a hydroelectric unit of energy in a power system using virtual reality. International Journal of Electrical Power & Energy Systems [J]. 2010,32(6): 599-606.

[5] Jie Geng, Ying Li, Ranran Wang, Zili Wang, Chuan Lv, Dong Zhou, A virtual maintenance-based approach for satellite assembling and troubleshooting assessment, Acta Astronautica, 2017,138:434-453

[6] Gan Lu • Lik-Kwan Shark • Geoff Hall • Ulrike Zeshan. Immersive manipulation of virtual objects through glove-based hand gesture interaction; Virtual Reality, 2012.16 (3): 243-252.

[7] Zhongxiang Zhu, Shanshan Li, Jiali You, Zhenghe Song. Application of Speech Recognition Technology to Virtual Reality System [A]. International Industrial Informatics and Computer Engineering Conference[C]. 2015:103-108.