

# Realization of Speech Auxiliary Function in Interactive Geometric Modeling

Fan Wang<sup>1, a</sup>, Yanping Hu<sup>2, b</sup>

<sup>1</sup>Dalian University of Technology, Dalian, 116024, China.

<sup>2</sup>Dalian University of Technology, Dalian, 116024, China.

<sup>a</sup>CodingSnail@163.com, <sup>b</sup>hypok@dlut.edu.cn

**Keywords:** Geometric modeling, speech recognition, Interactive mode.

**Abstract.** At present, the traditional interactive mode is mainly used for geometric modeling software. In order to improve the efficiency of modeling, a voice-assisted geometric modeling system is built. The technologies used in the system are speech recognition technology, Extensible Markup Language (XML) technology and object-oriented programming. After receiving the voice command, the system uses the speech recognition engine to identify the voice commands according to the XML syntax file. And then the voice commands identified are parsed and processed to generate the SolidWorks command set that accomplish the geometric modeling. The system provides a kind of interactive mode that combines mouse and voice for users and improves the efficiency of modeling.

## Introduction

Geometric modeling software is based on parametric technique and restriction technique [1], so modeling method is more direct and accord with the designer's thinking. However, due to the complexity of the parametric technique and restriction technique, geometric modeling software users still have to design the model through menus, toolbars, dialog boxes, etc. The interactive mode has some restrictions. Sometimes, in order to construct a geometry, the user needs to repeat the multi-step operation and the keyboard input is also slower. Voice is the most common, the most convenient and quickest way to transfer information [2]. In the process of geometric modeling, the model can be quickly adjusted to the appropriate perspective through the control of voice command about rotation. The user can use the voice command directly to adjust perspective without having to exit the current command state. In addition to the rotating view command, the system defines a number of other commands such as chamfer, shell, etc. Therefore it can help users complete the geometric modeling conveniently.

## System Overall Design

The system is based on SolidWorks secondary development technology. Firstly, the system establishes the running environment for the speech application programming interface (SAPI) by activating the XML syntax file prepared, setting the event interested and other operations. Secondly, the system changes the recognition state to SPRST\_ACTIVE to start recognizing the voice that user input. Thirdly, the voice driver module filters and organizes the information correctly identified. Finally, according to the current state that the SolidWorks program feed back to, the system sends a series of command sets to the SolidWorks program for controlling the completion of geometric modeling, through the SolidWorks secondary development interface. The whole framework of the system is shown in Figure 1:

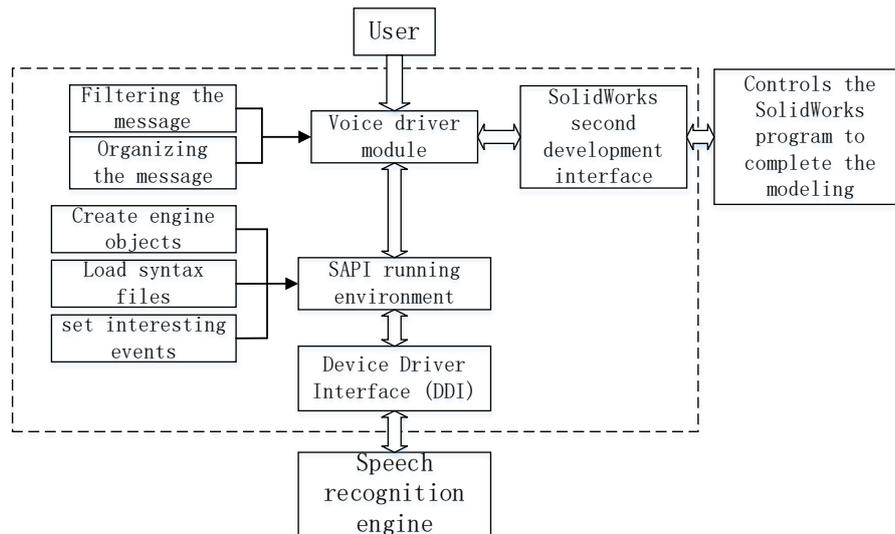


Fig. 1 Voice - assisted geometric modeling system

### Set the SAPI Running Environment

**Building Complete Grammar File.** Words, phrases and sentences expected to be identified are defined in Grammar rules [3]. It is a prerequisite for recognizing voice commands successfully, so it is important to write a complete grammar file. Usually, there are two kinds of grammar rules: Dictation Grammar, Command and Control Grammar [4]. Dictation Grammar can identify a large number of vocabulary in the engine dictionary and is suitable for continuous speech recognition [5]. Command and Control Grammar is mainly used to identify some specific command vocabularies and sentences defined in the grammar file. In order to assist the geometric modeling, the voice commands are divided into five categories: Sketch Rendering and Editing, Entity Modeling and Editing, Operating View, Digital Driving and Text Inputting.

According to the voice commands designed, the structure of XML syntax file used to achieve the voice interaction is as follows:

```

<?xml version="4.0" encoding="utf-8"?>
<GRAMMAR LANGID="801">
  <DEFINE>
    <ID NAME="VID_MODEL_EDIT" VAL="50"/>
    ...
    <ID NAME=" VID_NUMBERLIST " VAL ="800"/>
  </DEFINE>
  <RULE ID=" VID_MODEL_EDIT " >
    <L PROPID=" VID_MODEL_EDIT " >
      <P VAL="VID_Stretch"> chamfer </P>
      ...
    </L>
  </RULE>
  ...
  <RULE ID="VID_OrderList" >
    <L>
      <RULEREFF REFID="VID_NUMBERLIST" PROPID="VID_NUMBERLIST_PRO"/>
      ...
      <RULEREFF REFID="VID_MODEL_Edit" PROPID="VID_MODEL_EDIT_PRO"/>
    </L>
  </RULE>
</GRAMMAR>

```

The above is the overall framework of the grammar file in the geometric modeling system of speech interaction. Basically digital drive is the core of the geometric design operation, so digital recognition occupies a very important role in the entire system. In the design of grammar files, because the digital driving is the most complex, it needs to be constantly modified in practice to achieve the purpose for using.

**The Process of Initializing Recognition Engine.** Firstly, the system initializes the COM library and creates a speech recognition engine object through the function `CoCreateInstance()` [6]. There are two types of recognition engine objects: `InProctorRecognizer` and `SharedRecognizer`. The former is only used by the application that creates speech recognition engine object, and the latter can be used by multiple applications. Secondly, a voice recognition context (`RecoContext`) object is created for sending and receiving message notifications related to speech recognition. Thirdly, the message notification mechanism is set and the function `SetNotifyWindowMessage()` associates the `RecoContext` object with relational function that handles message. Finally, the system defines the relevant content that the engine needs to recognize and activate the recognition syntax rules that is loaded.

## **Realization and Processing of Voice Commands**

**The Recognition of Voice Commands.** The Windows Operating System maintains a message queue for each running application, so this mechanism can be used to implement real-time recognition of incoming speech. The speech recognition engine sends a message (`WM_RECOEVENT`) to the subject function of the message loop when a speech recognition event is triggered. After receiving the message, the body function determines the type of message. According to different needs, the system deals with different types of messages correspondingly.

Sentences that are successfully recognized cannot be used to achieve follow-up message-driven directly. It is necessary to properly process results identified, such as the interception and conversion of digital information, the handling and control of noise, etc. After the recognition is successful, the corresponding control message is found and placed in the message queue according to the mapping relationship between the message and the voice command. If the control message is not found, it is judged as a noise message and discarded directly.

**Processing the Recognition Result.** The message processing of voice information is a key step in the process of geometric modeling voice command driven. There are a few steps to implement this process: 1) The system reads the contents of the message mapping file and generates a list of class objects based on the serialization of the information read; 2) According to the object(`IspPhrase`) that exists in the recognition engine and its structure(`SPPHRASE`), the system can get the current working rules(`RuleID`)[7]; 3) According to the `RuleID` acquired and the information stored in the object, the system traverses the objects in the list and get the type of this rule. These types include sketch rendering and editing, solid modeling and editing, operational views, digital driving and text input commands; 4) For digital drive commands, the Chinese text information recognized is needed to convert into Arabic numbers. After the conversion is complete, the Arabic numerals are sent to the application as the part of the message parameters; 5) The text input command will be sent directly without any processing. 6) For other types of rules, the system obtains the parameters that the sending message function needed by traversing the class object list and matching the associated attributes.

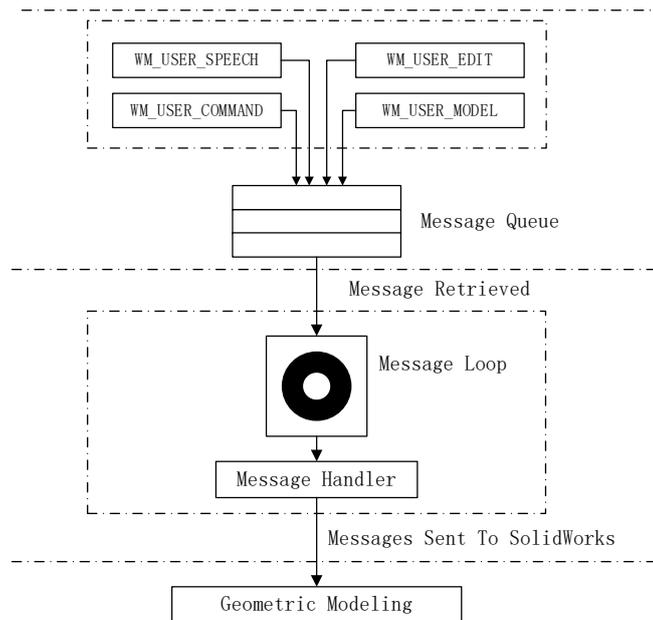


Fig. 2 The schematic diagram of message mechanism

When the speech recognition module intercepts the user's voice stream and processes it with SAPI, command that the user inputs is converted into Windows messages. The module send these messages predefined to Windows message queue. The plug-in program developed with the SolidWorks API SDK references message header files compiled to capture messages from the speech recognition module. The function that handles messages in the program interprets and processes messages intercepted. Finally, the plug-in program sends commands to SolidWorks to control the completion of geometric modeling. The function of voice-assisted geometric modeling can be achieved, through the above process.

**Improve the Matching Rate of Voice Command.** First of all, users read a specific training file. The system generates the corresponding configuration file by collecting the user's pronunciation features that includes the noise environment, the frequency of pronunciation, etc. The system can be trained to generate multiple configuration files for multiple users. The recognition engine can perform voice recognition based on the profile that the user selects. There are multiple grammar rules in the same grammar file. Depending on the current operating state, the system qualifies the syntax rules that should currently be activated by feedback. Thereby, the range of the identified content is narrowed and the purpose of improving the voice command matching rate is achieved.

The system filters the noise by using the method that controls the threshold. The user adjusts the noise threshold based on its own volume, and the system only recognizes the speech within this threshold range. If the volume is greater than the maximum threshold or less than the minimum threshold, it is treated as noise without any further processing. The average volume, maximum volume and minimum volume is recorded for a period of time by defining a struct. According to the value recorded, the system filters the noise through one way that extreme value filtering and mean filtering. Extreme filtering is mainly used to filter sudden and instantaneous noise, while the mean filter is used for non-intermittent noise.

### The Analysis of Application Example

The view occupies a very important position in the process of geometrical modeling, and the geometry can be viewed from any angle by manipulating views such as rotating view, dynamic zooming, zooming out views, and so on. In the actual modeling process, sometimes the user wants to adjust to the appropriate perspective, but it needs a series of operations through the mouse and keyboard. In some cases the user even needs to withdraw from the current state of the order, so it is cumbersome. For these questions, the voice commands for manipulating views are defined, including standard views, perspectives, rotating views, section views, etc. The cross-sectional view can be used for situations where the internal structure of the cabinet needs to be observed frequently.

In the process of geometric modeling, rotating view command is the most widely used. Voice commands can rotate the model to the appropriate angle of view quickly through the rotation of any angle and the angle of the model can also be adjust easily in the process of modeling. These voice commands can be interspersed throughout the geometric modeling process and be used in combination with each other, so they can assist users to complete the geometric modeling better.

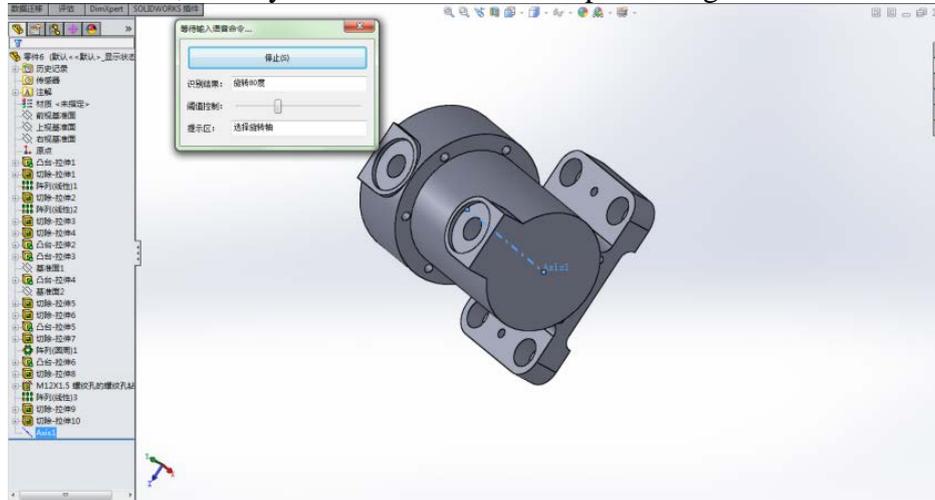


Fig. 3 The example of rotation view application

## Conclusion

According to the design process of the speech-assisted geometric modeling system, this paper defines the modeling commands that are suitable for voice description and conform to human habits. According to these voice commands, the construction of the grammar file is completed. The voice commands and Windows messages are mapped to generate MSM files. On the basis of the grammar file, the function of speech recognition is realized by encapsulating the speech recognition interface and initializing the recognition engine. After recognizing the correct command, according to the mapping relationship in the MSM file, the system gets the corresponding message. After the different types of messages are processed accordingly, the system uses the Windows message loop mechanism to complete the communication between the speech recognition module and the geometric modeling module to control the completion of geometric modeling. Interactive mode combined with the voice and mouse is more suitable for human natural habits and ways of thinking. Therefore, the speech-assisted geometric modeling is feasible and has a certain improvement in modeling efficiency.

## References

- [1] Yang Yu. The parametric design and intelligent assembly system based on the secondary development of solidworks: IEEE, 2010
- [2] Mustafa Nazmi Kaynak, Qi Zhi, Adrian David Cheok et al. Analysis of lip geometric features for audio-visual speech recognition. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2004, 34 (4) : 564-570
- [3] Lawrence R. Rabiner, Biing-Hwang Juang. Fundamentals of speech recognition., 1993
- [4] Sadaoki Furui. Speaker-independent isolated word recognition using dynamic features of speech spectrum. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1986, 34 (1) : 52-59
- [5] Todd F. Mozer, Pieter J. Verneulen. Background speech recognition assistant: Google Patents, 2015
- [6] Vishakha Garg. Speech signal processing. Journal of VLSI Design and Signal Processing, 2015, 1 (2)

[7] Ian Bennett, Bandi Ramesh Babu, Kishor Morkhandikaret al. Multi-language speech recognition system: Google Patents, 2015