Advances in Machinery, Materials Science and Engineering Application
M. Chen et al. (Eds.)

doi:10.3233/ATDE220460

Multimodal Modeling System in Virtual Reality Scene

Chunxi LI [a,1] and Wenjun HOU [a] and Yiting CHENG [a]
[a] *Human Computer Interaction Intelligent Laboratory, BUPT, 10 Xitucheng Road, Beijing, China*

Abstract. In recent years, upgrading manufacturing informatization has put forward higher requirements for the designer. The traditional design method requires designers to switch their thinking on the platform of multiple dimensions, which increases the difficulty. With the development of virtual reality technology, the design process in a 3D environment can be realized. Aiming at the conceptual design stage in the design process, this paper carries out multi-modal interactive design for different tasks in the full three-dimensional environment. A multi-modal modeling method based on gesture interaction and supplemented by voice interaction and eye movement interaction is established through relevant experiments and analysis. The relevant platform is developed based on the Unity 3D engine, and the modeling process is completed. This paper proposes a solution to the modeling process in a 3D environment and the Enlightenment of further research.

Keywords. Multimodal interaction, modeling, virtual reality

1. Introduction

The design stage is a procedure of a mechanical product, and its accuracy and efficiency directly affect the service performance of the product. As a key product design stage, conceptual design needs to fully reflect the designer's design intention and product function [1]. At present, the existing design methods require designers to constantly change their thinking on multi-dimensional platforms, resulting in the invariance of information exchange, as shown in figure 1 [2-4]. Solving this problem, the model can be designed in a whole three-dimensional environment, and the development of virtual reality (VR) technology provides a good platform for this need. Still, the research on the interaction of the modeling scene is not deep enough and lacks systematic interaction [5]. Therefore, aiming at the modeling process in the virtual reality environment, this paper mainly studies the modeling method of the combination of gesture, eye movement, and voice.

[1] Chunxi Li, Corresponding author, Human computer interaction Intelligent Laboratory, BUPT, 10 Xitucheng Road, Beijing, 100876, P. R. China, E-mail: lichuxni@bupt.edu.cn.

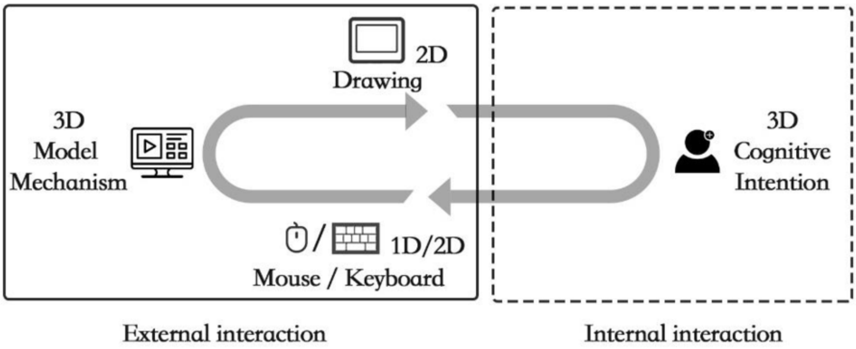


Figure 1. Changes in thinking dimensions of existing modeling methods.

Existing designers mainly carry out conceptual design with the help of computer-aided design (CAD), including AutoCAD, Sketch-Up, 3ds Max, and SolidWorks. The modeling software uses keys and mice as input, creates geometry based on parametric modeling mechanism, and forms complex design products by boolean operation basic geometry. Some researchers proposed highly structured [5] and deep-seated pages provide strong modeling ability by secondary development CAD software. Meanwhile, the software lacks flexibility and expressiveness and affects the designer creation process [2]. Moreover, the complex interface brings great learning difficulty, which requires professionals to master it skillfully, thus increasing the product development cost [3]. More importantly, the current CAD system has limitations on thinking, and the inherent tools affect the design thinking, which determines the design quality [6]. To sum up, 3D conceptual design needs to change the current interaction mode. The existing interaction methods require designers to suppress the design intention into low-dimensional interaction instructions and reorganize the two-dimensional information of the screen into three-dimensional information through cross-spatial perception in the brain. This process has experienced two-dimensional collapses in the reconstruction process, as shown in figure 1. This process leads to the problem of low naturalness and intuition in the modeling process. The development of virtual reality technology [7], natural interaction [8], and multimodal interaction [5] puts forward solutions to this problem.

The conceptual design process is a visualization of designer ideas without the natural world and authentic environment. Therefore, the completely virtual environment in virtual reality is suitable for the conceptual design process. In the virtual environment, the user interface is adopted. It does not rely on the command line and graphical interface but on an invisible page in which users interact continuously and incrementally. The environment comes from the real world, and the operation process is natural and intuitive. As a typical feature of genuine interaction, non-touch screen modal interaction provides a flexible and efficient form of interaction. The modeling scene usually includes gesture, voice, eye movement, and handle. In order to provide a more realistic way of interaction, this paper focuses on gesture, voice, and eye movement without additional devices to explore the process of three-dimensional environment modeling.

2. 3D Modeling Environment

In the virtual reality environment, the modeling form combining different interactive methods is adopted for various geometric features to task for the following four levels : Creating, Editing, Operation, Boolean. Finally, complete the establishment of the whole model.

2.1. Development Environment

In order to realize multimodal interaction, it is necessary to collect gesture, voice and eye movement data. Integrate corresponding data through unity 3D engine and provide corresponding visual environment. The specific system configuration is shown in the table 1 below.

Table 1. Virtual modeling multimodal interactive basic development environment information.

System Composition	Content
Product Positioning	Establishing a conceptual design platform of the multi-modal interactive 3D model in a virtual environment based on natural intelligence
Application scenario	Carrying out three-dimensional design freely in a certain area by using a standing posture
Software Environment	Unity-2018.3.1c1(Developed with Visual C#)
Hardware Configuration	PC (Win 10)
Gesture	Leapmotion
Eye Movement	HTC vive Pro Eye
Voice	HTC vive Pro

As an excellent real-time 3D interactive content creation and operation platform, unity 3D can provide multi-platform design and optimization. Therefore, most devices offer the software development kit (SDK) for the unity environment. All external devices involved in this paper provide SDK, which reduces the difficulty of development and improves data accuracy. Furthermore, the system needs to provide the three-dimensional environment and the medium of different interactive modes, as shown in figure 2.

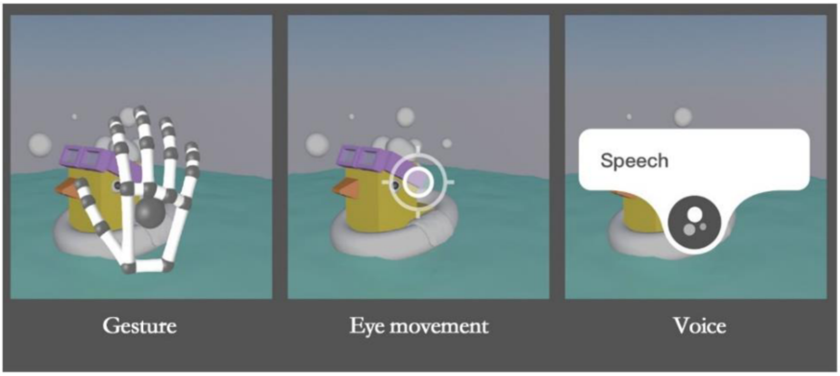


Figure 2. The avatar of interactive mode in three-dimensional environment.

3. Interactive Method Design

In the virtual reality environment, three-dimensional interaction needs to be adopted. Compared with two-dimensional interaction, three-dimensional interaction provides more operation freedom. Its interaction scene and intention are richer, bringing more design space and requiring a new interaction metaphor. Interactive metaphor abstracts the mechanism existing in the real world so that users can clearly understand the existing process and self intention to complete the corresponding task. The tasks of modeling in the virtual reality environment can be divided into model creation, model editing, and model manipulation. The creation of a model is the basis of most operations, which generates complex geometry by creating simple geometry and then manipulating and editing simple geometry. The task modules included in this article include as shown in figure 3:

1) Creation of geometry: As the central module of 3D modeling, call the corresponding geometric features in the scene the basis for subsequent editing.

2) Editing of geometry: According to the existing geometric features, modify their corresponding parameters according to their different types to meet the requirements of the division involved.

3) Operation of geometry: When there are multiple geometries in the scene, you can select the corresponding objects and complete basic operations such as rotation, movement, and deletion.

4) Boolean operation of multiple geometric features: For multiple geometries with contact, it is integrated into a complex geometry through boolean operations such as difference, intersection, and union. This paper only deals with the Boolean operation of two geometries.

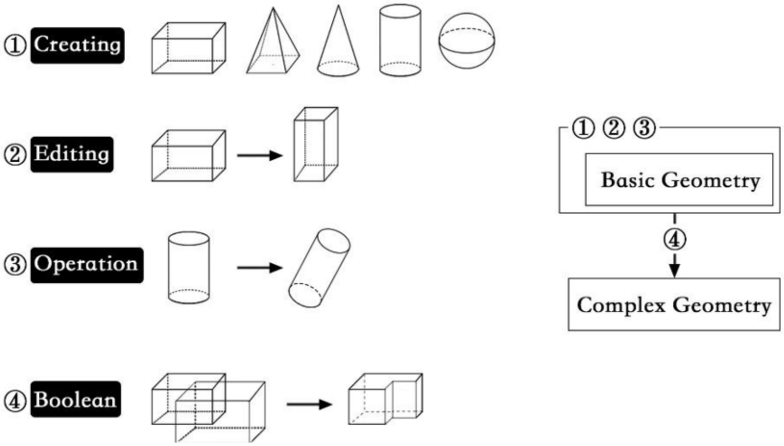


Figure 3. Operations required during modeling.

Through the recognition and analysis of multimodal data, including gesture, eye movement, and voice, to finish the modeling task. Gesture interaction is the most common kind of human posture, and it is also the earliest one used in virtual reality systems. However, human physiological reasons limit the use of gestures in the virtual environment. To expand the application scope of gesture interaction and approach the problem of multichannel interaction information fusion, this paper proposes a multichannel interaction technology with gesture interaction as the main body and other interaction modes as the auxiliary.

3.1. Gesture Interaction

In the three-dimensional environment, the semantics contained in gesture and the universality of gesture has significant advantages over other interaction methods. There are two types of gesture interaction: volley gesture interaction and collision gesture interaction. Volley gesture refers to the interaction form when the hand is in contact with the object in the scene, while collision gesture is an interactive way to start the task by contacting the hand with the object. Through heuristic experiments, users can customize the corresponding modeling gestures and calculate the consistency of gestures between different users for the same task through the following formula.

$$AR(r) = \frac{|P|}{|P|-1} \sum_{P_i \in P} \left(\frac{P_i}{P} \right)^2 - \frac{1}{|P|-1} \tag{1}$$

In the formula 1, P represents the frequency of all gestures, P_i indicates the frequency of a gesture. AR ≤ 0.1 indicates low consistency, $0.3 <$ ar ≤ 0.5 indicates high consistency, and AR > 0.5 indicates high consistency. By selecting the gesture with the highest frequency, the corresponding gesture set of the following tasks is established. The relevant experimental process and results are shown in the papers previously published by the same laboratory and will not be introduced here, as shown in figure 4 [10].

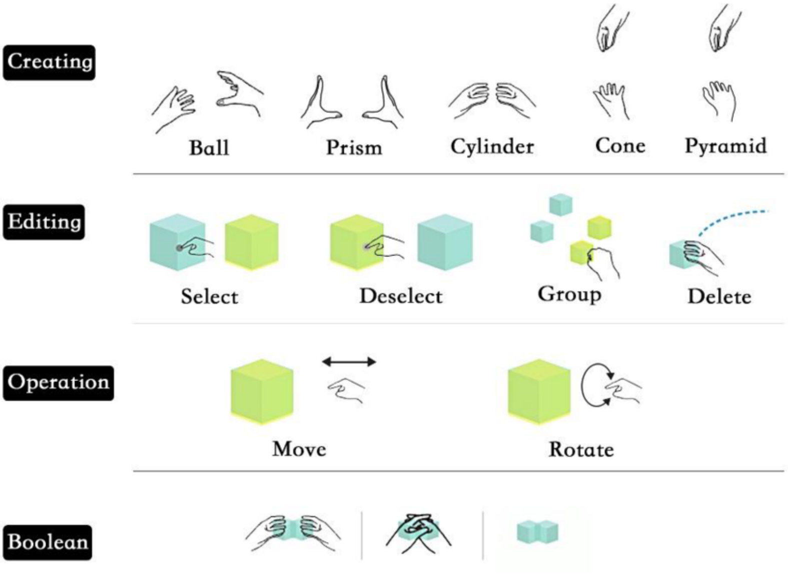


Figure 4. Definition of Volley gesture operation in modeling process in 3D environment.

3.2. Eye Movement Interaction

Compared with gesture interaction, eye movement interaction has a wider range of interactions. The input device used in this paper obtains the two-dimensional plane position coordinates of the annotation points of the eyes in the helmet. However, there are three interactive dimensions of X, Y, and Z in the three-dimensional environment, so it is necessary to complete the transformation of the two-dimensional coordinates of eye annotation points and the corresponding three-dimensional coordinates, as shown in figure 5. Establish the ray emitted from the midpoint of the binocular through the gaze point in the screen, and take the first object in contact with the ray as the three-dimensional coordinates of eye movement data reaction. The implementation method is shown in the figure. The forms of eye movement interaction include gaze, jump, and sweep. In this paper, the gaze is mainly used to complete eye movement interaction tasks.

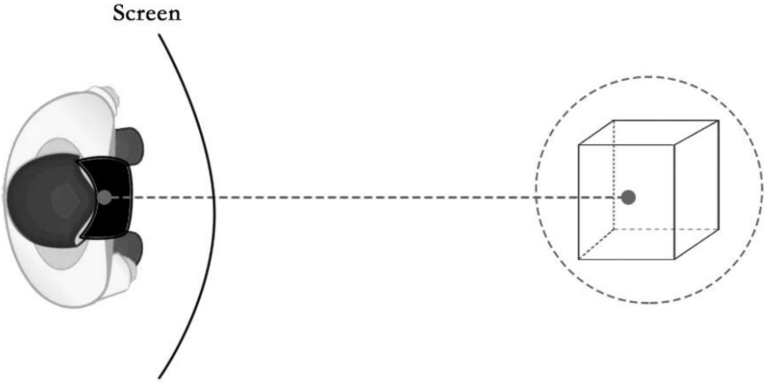


Figure 5. Definition of eye movement gaze collision interaction mode in 3D environment.

The line of sight moves within a certain range around the object. The farther the object is from the observation point, the greater the error range. When the number of fixation frames f is greater than the set threshold, it is considered that it is used to select the object.

3.3. Voice Interaction

The Voice interaction system is a natural input form that can complete accurate information input. Its processing flow can be divided into front-end signal processing and back-end recognition. To obtain better speech information, the front-end processing and back-end recognition convert the sound signal into text content according to the acoustic and language model. By calling the speech recognition module of iFLYTEK, this paper converts the speech signal collected from the HTC Vive Pro helmet into the text to complete the interaction with the system. To distinguish the unconscious operation of the user, the wake-up word of voice interaction is designed. The information obtained is valid only after the user speaks the wake-up word. The specific flow is shown in the figure 6.

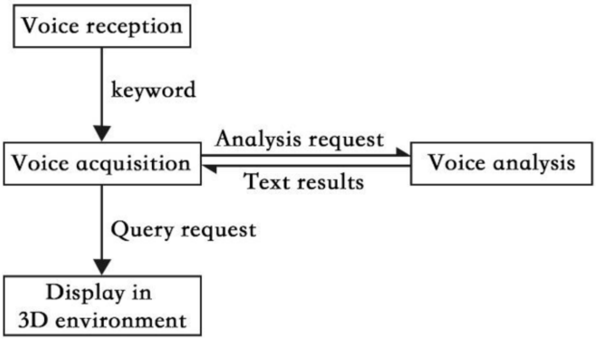


Figure 6. Voice interaction process.

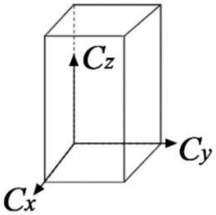
4. 3D Modeling of Virtual Environment

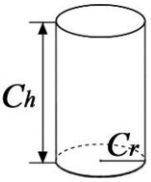
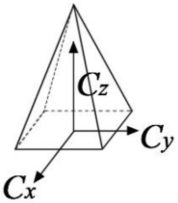
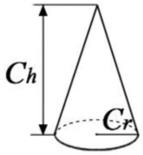
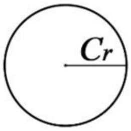
The virtual environment includes two elements: scene and object. The scene is where users interact with interactive objects and system scenes through media, including background, lighting and projection, layout, and material. Establishing an appropriate scene in the three-dimensional environment can enhance the user's sense of immersion, improve the system experience, and assist the user in completing the task more smoothly. In addition, the environment also contains object elements, including entity objects, graphical user interface (GUI) objects, media objects, and auxiliary objects. The object elements provide feedback on the behavior and let the user understand their behavior or the information provided by the system. The most important element in the 3D modeling environment is the solid object, the visual carrier of the model.

4.1. Geometric Feature Classification

The geometric features contained in parts can be divided into columns, cones, and spheres. Columns can be divided into prisms and cylinders, and cones can be divided into pyramids and cones. The table 2 shows different kinds of geometric features and their key parameters.

Table 2. Classification and description of basic geometric features.

Classification	Name	Graph	Feature description
column	prism		The side and bottom surfaces are both planes, with multiple sides and two bottom surfaces, and the bottom surfaces are parallel.

	cylinder		The side surface is curved, and the bottom surface is plane. It has one side surface and two bottom surfaces, which are parallel to each other
cone	Pyramid		The sides and bottom surfaces are planes, with multiple sides and a bottom surface.
	cone		The side surface is curved and the bottom is plane with a side and a bottom.
ball	ball		There is only one surface, and this surface is a curved.

4.2. Multimodal Interactive Fusion Modeling

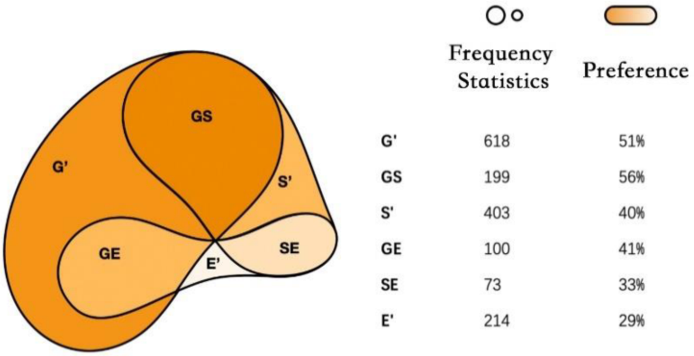
Multi-modal interaction can organically combine different interaction modes by using one or more interaction modes separately, simultaneously, and coupled. To achieve the purpose of improving user experience and overlapping information bandwidth. To obtain the interaction form for the different tasks, we use the heuristic method to analyze the adaptation degree between the interaction of different modes and the modeling task to establish the interaction mode of multi-modal coupling. The tasks are shown in the table 3 below.

Table 3. Classification and description of 3D interactive tasks.

Classification	Level	Name
Create feature	Solid object	Prism/ Cylinder/ Pyramid/ Cone/ Ball
Edit feature	Choice	Single choice\ Multiple choice\ Select all\ Deselect

	Global Edit	Group\ Contact group\ Delete
Operation	Move	Free\Along X, Y, Z axis
	Rotate	Along X, Y, Z axis
Boolean operation	Two solid objects	Sum\ Difference\ Union\ Intersection
other	Menu operation	Open menu\ Close menu

There are six interaction modes: gesture / eye movement / voice / gesture & eye movement / gesture & voice / Voice & eye movement. The following experiments are designed. We invited ten users to test the usability of multimodal interaction modeling. The test aims to collect user preferences for interactive forms and score them with a multidimensional scale, as shown in the table. We obtained 2280 applicability scores of different models and 910 interactive description data, of which 380 interactive data were attached with four-dimensional scoring data. The selection frequency of different interaction modes is shown in the figure 7 below.



	Frequency Statistics	Preference
G'	618	51%
GS	199	56%
S'	403	40%
GE	100	41%
SE	73	33%
E'	214	29%

Figure 7. Area based interactive method selection Venn graph.

To further obtain the interaction modes adapted to different tasks, tasks are divided into the following types so that users can try different interaction modes for the same task and evaluate them. The evaluation methods are shown in the table 4, and the results are shown in the table 5.

Table 4. Evaluation indicators and evaluation criteria for the applicability of different interaction modalities.

Evaluation object	**Evaluation dimension**	**Metrics**
modality	Applicability	The practicability of the modality to the task. (Inappropriate)1~5 (Appropriate)
interactive	Matching	The degree to which the modality interacts matches the task. (Mismatch)1~7(Match)
	Fatigue	The degree of fatigue felt when finishing this task by the modality. (Non-Fatigue)1-7(Fatigue)
	Difficulty	How difficult it feels to operate this interaction (Not Difficult)1-7(Difficulty)
	Fluency	The fluency of this interaction (Not fluency)1-7(Fluency)

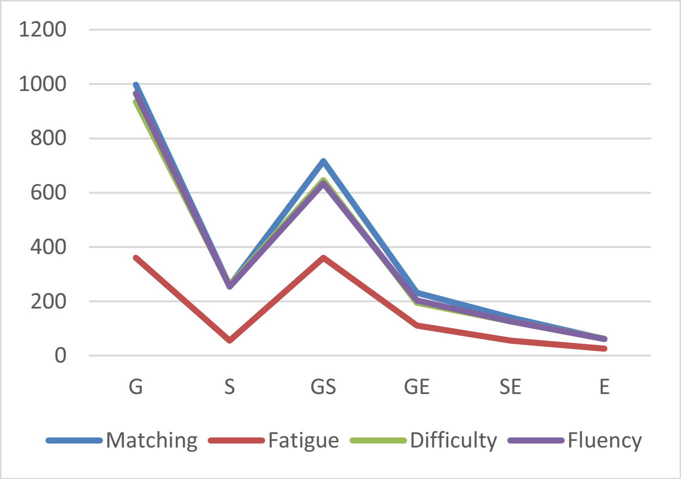


Figure 8. Summary of user preferences for different interaction modes.

Table 5. Matching degree of different interaction modes to tasks.

Create feature	Solid object	G	GS	S	GE	SE	E
Edit feature	Choice	▦	▦	□	■	■	■
	Global Edit	▤	■	▤	▦	▤	□
Operation	Move	■	▤	□	░	□	□
	Rotate	▦	■	▦	▤	▦	□
Boolean operation	Two solid objects	▦	▦	▤	▤	░	□
other	Menu operation	▦	▦	□	■	■	■

■ very matched	▦ relatively matched	▤ loosely matched	░ less matched	□ mismatched

The experimental results show (figure 8) that gesture interaction is the most widely used and has a strong combination with other forms of interaction. In contrast, eye movement interaction needs to rely on other forms of interaction, and voice interaction is the most auxiliary to gesture interaction. This paper defines the modeling interaction in a virtual environment as the form with gesture interaction as the core and voice interaction, and eye movement interaction as the auxiliary.

5. Result

The figure 9 shows the final presentation result of the system, in which different geometric features are established through voice and gesture, respectively. Then the cube is selected through eye movement and moved to the corresponding position, which overlaps with the cylinder. The moving object contacts its adjacent objects, and a complete model is formed through Boolean operations.

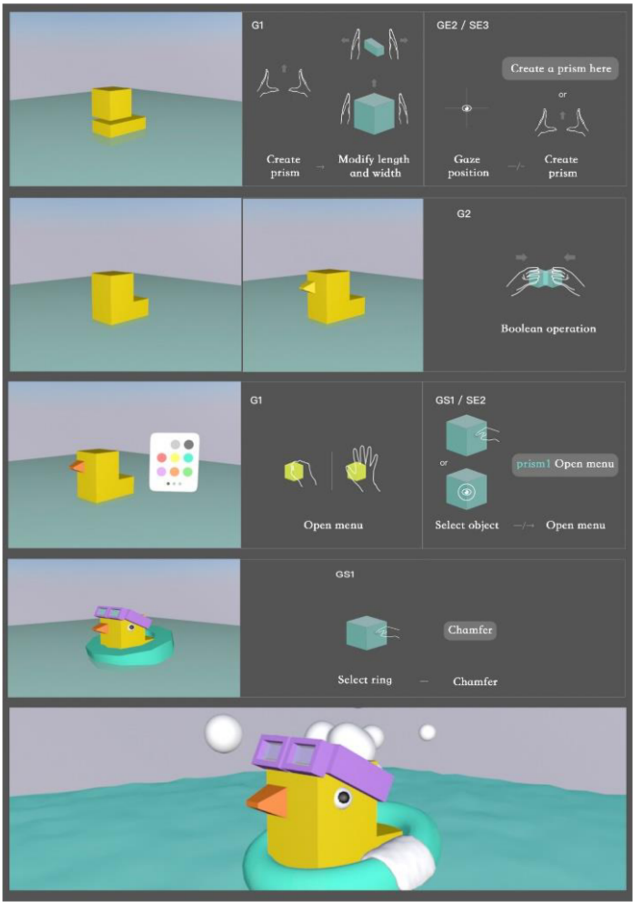


Figure 9. Process demonstration of multimodal modeling program based on unity 3D.

6. Conclusion

Aiming at the modeling process in a virtual reality environment, we use the Unity 3D engine and the combination of gesture, eye movement, voice, and other multimodal interaction methods to obtain the complex 3D model by establishing the basic geometry and Boolean operation on the selected features. The interaction method proposed in this paper can organically combine various interaction forms, quickly and effectively establish a model in the virtual reality environment, and provide a new solution for the interaction mode of virtual reality.

References

[1] Dani TH, Gadh R. Creation of concept shape designs via a virtual reality interface. Computer Aided Design. 1997; 29(8):555-563.

[2] Schkolne S, Pruett M, Schrder P. Surface drawing: creating organic 3D shapes with the hand and tangible tools. Chi Conference on Human Factors in Computing Systems. 2001; 261–268.

[3] CC P Chu, Dani TH, Gadh R. Multi-sensory user interface for a virtual-reality-based computer-aided design system. Computer-Aided Design. 1997; 29(10):1329-1334.

[4] Alkemade R, Verbeek FJ, Lukosch SG. On the efficiency of a VR hand gesture-based interface for 3d object manipulations in conceptual design. International Journal of Human-Computer Interaction. 2017; 33(10-12): 882-901.

[5] Jr LV, Keefe DF. Course: 3D spatial interaction: Applications for art, design and science. ACM. 2011.

[6] Zheng J, Chan K, Gibson I. Desktop virtual reality interface for computer aided conceptual design using geometric techniques. Journal of Engineering Design. 2001; 12(4): 309-329

[7] Robertson B, Radcliffe D. Impact of CAD tools on creative problem solving in engineering design. Computer-aided Design. 2009; 41(3): 136-146.

[8] Pavlovic VI, Sharma R, Huang TS. Visual interpretation of hand gestures for human-computer interaction: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997; 19(7): 677-695.

[9] Cicognani A, Maher ML. Design speech acts. “How to do things with words” In Virtual Communities. Key Centre of Design Computing. 2006,

[10] Hou W, Feng G, Cheng Y. A fuzzy interaction scheme of mid-air gesture elicitation. Journal of Visual Communication and Image Representation. 2019; 64: 102637