Combining Gesture and Voice Control for Mid-air Manipulation of CAD Models in VR Environments

Markus Friedrich, Stefan Langer and Fabian Frey
Institute of Informatics, LMU Munich, Oettingenstraße 67, 80538 Munich, Germany

Keywords: CAD Modeling, Virtual Reality, Speech Recognition, Constructive Solid Geometry.

Abstract: Modeling 3D objects in domains like Computer Aided Design (CAD) is time-consuming and comes with a steep learning curve needed to master the design process as well as tool complexities. In order to simplify the modeling process, we designed and implemented a prototypical system that leverages the strengths of Virtual Reality (VR) hand gesture recognition in combination with the expressiveness of a voice-based interface for the task of 3D modeling. Furthermore, we use the Constructive Solid Geometry (CSG) tree representation for 3D models within the VR environment to let the user manipulate objects from the ground up, giving an intuitive understanding of how the underlying basic shapes connect. The system uses standard mid-air 3D object manipulation techniques and adds a set of voice commands to help mitigate the deficiencies of current hand gesture recognition techniques. A user study was conducted to evaluate the proposed prototype. The combination of our hybrid input paradigm shows to be a promising step towards easier to use CAD modeling.

1 INTRODUCTION

Current Computer Aided Design (CAD) modeling tools use mainly mouse control-based object manipulation techniques in predefined views of the modelled objects. However, manipulating 3D objects with a 2D input device (mouse) on a 2D output device (monitor) always necessitates a complex transfer, making it difficult for beginners to grasp the needed interaction concepts.

With the advent of affordable Virtual Reality (VR) devices and robust hand gesture recognition systems, more possibilities are at the fingertips of interaction designers and researchers. The goal is to leverage the potential of these immersive input methodologies to improve intuitiveness, flattening the learning curve, and increasing efficiency in the 3D modeling task.

However, hand gesture recognition systems face many challenges in productive environments. Compared to expert devices like 3D mouses, these systems lack robustness as well as precision. Moreover, constant arm movements can cause fatigue syndroms like the Gorilla Arm Syndrom (LaValle, 2017), making it hard to use systems over an extended period of time. This is where we hypothesize that a hybrid approach, combining gesture recognition and voice control, is beneficial. While using hands for intuitive model part manipulation, voice commands replace complex to recognize gestures with simple to say and easy to memorize word commands.

In this work, we propose a novel interaction concept, implemented as a prototype, which combines hand gesture recognition and voice control for CAD model manipulation in a VR environment. The prototype uses gesture-based mid-air 3D object manipulation techniques where feasible and combines it with a set of voice commands to complement the interaction concept. CAD models are represented as a combination of geometric primitives and Boolean set-operations (so called Constructive Solid Geometry (CSG) trees) enabling transformation operations that are more intuitive for beginners. The whole system is evaluated in a user study, showing its potential. The paper makes the following contributions:

- A new interaction concept for intuitive, CSG tree-based, CAD modeling in VR leveraging the strengths of both, gesture- and voice-based interactions.
- A prototypical implementation of the interaction concept with off-the-shelf hard- and software.
- A detailed user study proving the advantages of the proposed approach.

The paper is structured as follows: Essential terms are explained in Section 2. Related work is discussed in Section 3 which focuses on mid-air manipulation techniques. This is followed by an explanation of the concept (Section 4) which is evaluated in Section 5.

Friedrich, M., Langer, S. and Frey, F.
Combining Gesture and Voice Control for Mid-air Manipulation of CAD Models in VR Environments.
DOI: 10.5220/0010170501190127
In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 2: HUCAPP*, pages 119-127
ISBN: 978-989-758-488-6

Since the proposed prototype opens up a multitude of different new implementation and research directions, the paper concludes with a summary and a short description of possible future work (Section 6).

2 BACKGROUND

2.1 Construction Trees

Construction or CSG trees (Requicha, 1980) is a representation and modeling technique for geometric objects mostly used in CAD use cases. Complex 3D models are created by combining primitive geometric shapes (spheres, cylinders, convex polytopes) with Boolean operators (union, intersection, difference, complement) in a tree-like structure. The inner nodes describe the Boolean operators while the leave nodes represent primitives. The CSG tree representation has two big advantages over other 3D model representations: Firstly, it is memory-saving, and secondly, it is intuitive to use.

2.2 Mid-air Interactions for 3D Object Manipulation

The manipulation of 3D objects in a virtual scene is a fundamental interaction in immersive virtual environments. Manipulation is the task of changing characteristics of a selected object using spatial transformations (Bowman and Hodges, 1999). Translation, rotation, and scaling are referred to as basic manipulation tasks (Bowman et al., 2004). Each task can be performed in any axis direction. In general, a single transformation in one specific axis is defined as a degree of freedom (DOF). Thus, a system that provides a translation-only interface on all three axes supports three DOFs whereas a system that offers all three manipulation tasks on all axes has 9 DOFs (Mendes et al., 2019).

In so-called mid-air interactions, inputs are passed on through the user's body, including posture or hand gestures (Mendes et al., 2019). This kind of interaction provides a particularly immersive and natural way to grab, move, or rotate objects in VR which allows for more natural interfaces that can increase both, usability and user performance (Caputo, 2019). In this paper, the term mid-air manipulation refers specifically to the application of basic transformations to virtual objects in a 3D environment using hand gestures. Apart from the allowed degrees of freedom, existing techniques can be classified by the existence of a separation between translation, rotation, and scaling (Mendes et al., 2019) which also affects manipulation precision. Furthermore, two additional interaction categories exist: Bimanual and unimanual. Bimanual interfaces imply that users need both hands to perform the intended manipulation, whereas applications using unimanual interfaces can be controlled with a single hand (Mendes et al., 2019).

3 RELATED WORK

Object Manipulation. The so-called Handle Box approach (Houde, 1992) is essentially a bounding box around a selected object. The Handle Box consists of a lifting handle, which moves the object up and down, as well as four turning handles to rotate the object around its central axis. To move the object horizontally, an activation handle is missing and instead, the user can easily click and slide.

In (Conner et al., 1992), so-called Virtual Handles that allow full 9 DOF control are proposed. The handles have a small sphere at their ends, which are used to constrain geometric transformations to a single plane or axis (Mendes et al., 2019). The user selects the manipulation mode with the mouse button. During rotation, the initial user gesture is recognized to determine the rotation axis. Both techniques are initially designed for classic mouse/screen interactions but are seamlessly transferable to hand gesture/VR environments.

Mid-air Interactions. Caputo et al. state that users often like mid-air interactions more than other methods, finding the accuracy in manipulation sufficiently good for many proposed tasks (Caputo, 2019). First systems used gloves or other manual input devices providing the functionality to manipulate or to grab objects. In (Robinett and Holloway, 1992), a system for translating, rotating, and scaling objects in VR is proposed. The formalization of the object manipulation task in three sub tasks, namely (object) selection, (object) transformation and (object) release was proposed in (Bowman and Hodges, 1999). One of the most influential techniques in this field is the Hand-centered Object Manipulation Extending Ray-casting (HOMER) method (Bowman and Hodges, 1997). After selecting an object with a light ray, the hand moves to the object. The hand and the object are linked and the user can manipulate the object with a virtual hand until the object has been dropped. After that, the connection is removed and the hand returns to its natural position.

Metaphors for Manipulation. A bimanual 7 DOF manipulation technique is the Handlebar introduced by Song et al. (Song et al., 2012). It is adopted by multiple real-world applications. The Handlebar uses the relative position and movement of both hands to identify translation, rotation, scaling, and to map the transformation into a virtual object. The main strength of this method is the intuitiveness due to the physical familiarity with real-world actions like rotating and stretching a bar. However, holding an arm position may exhaust the user (Gorilla Arm Syndrome (LaValle, 2017)).

In (Wang et al., 2011), manipulations (in particular translation and rotation) are separated. Translation is applied by moving the closed hand and rotation by using two hands around the three main axes, mimicking the physical action of rotating a sheet of paper. The most direct unimanual method is the Simple Virtual Hand metaphor (Caputo, 2019) where objects are selected by grasping them. After the selection, movement and rotation of the user's hand are mapped directly to the objects. This method is characterized by its high intuitiveness. However, accurate, reliable and robust hand tracking is required. Kim et al. extended this idea in (Kim and Park, 2014) by placing a sphere around the selected object and additionally enable the user to control the reference frame during manipulation which shows great improvements, in particular for rotation.

In order to demonstrate the advantages of DOF separation, Mendes et al. applied the Virtual Handles concept (Conner et al., 1992) to VR environments (Mendes et al., 2016). Since Virtual Handles allow users to choose a single axis, transformations in unwanted axes are improbable. However, transformations in more than one axis take a bit more time. We use this concept for basic transformations in our prototype (see Figure 4a).

Multimodal Object Manipulation. Chu et al. (Chu et al., 1997) presented a multimodal interface for CAD Systems in VR as early as 1997, concluding that voice commands and hand gestures are superior to eye tracking. Zhong et al. (Zhong and Ma, 2004) propose an approach for constraint-based manipulations of CAD models in a VR environment. The authors implemented a prototype for a three-level model comprising of a high-level model for object definition, a mid-level, hybrid, CSG and Boundary Representation model for object creation, and a polygon-level based model for visualization and interaction. Zheng et al. (Zheng et al., 2000) also propose a VR-based CAD system for geometric modeling. In order to enhance the human-machine interaction, they add an electronic data glove as an additional input device for constructing, destroying and freeform-creating objects. The authors claim these interaction techniques to be more suitable and intuitive to humans than traditional manipulation devices. Lee et al. (Lee et al., 2013) conducted a usability study of multimodal inputs in Augmented Reality environments. The authors compare speech-only and gesture-only approaches to a hybrid solution and find that the multimodal approach is a more satisfying interaction concept for users. Xue et al. (Xue et al., 2009) improve the voice command based input model for CAD software by making the interaction more natural through natural language processing. Kou et al. (Kou et al., 2010), as well, focus on natural language processing, utilizing a template matching approach, enabling the system to process unknown expressions as well.

4 CONCEPT

In order to simplify the modeling process as much as possible, we utilized mid-air manipulation techniques without any controllers or gloves - hand tracking only. The main feature of our technique is the use of only the grasping gesture. This simple gesture allows the usage of low-cost hand-tracking sensors, more specifically the Leap Motion Controller sold by Ultraleap. More complex instructions are operated via voice control. To the best of our knowledge, this is the first approach that combines mid-air manipulation techniques with voice commands in order to create an intuitive and fast way of manipulating CSG-based 3D objects in VR.

4.1 System Overview

The proposed system consists of hardware and software components and is detailed in Figure 1. The hand tracking controller (*Leap Motion Controller*) is mounted on the VR headset which is an off-the-shelf *HTC Vive* that uses laser-based head tracking and comes with two 1080x1200 90Hz displays. The speech audio signal is recorded using a wearable microphone (*Headset*). A central computing device (*Workstation*) takes the input sensor data, processes it, applies the interaction logic and renders the next frame which is then displayed in the VR device's screens.

The software architecture consists of several subcomponents. The *Gesture Recognition* module translates input signals from the *Leap Motion sensor* into precise hand poses and furthermore recognizes hand gestures by assembling temporal chains of poses. The *Speech Recognition* module takes the recorded speech audio signal and translates it into a textual representation. The CSG model, which is later used for editing,

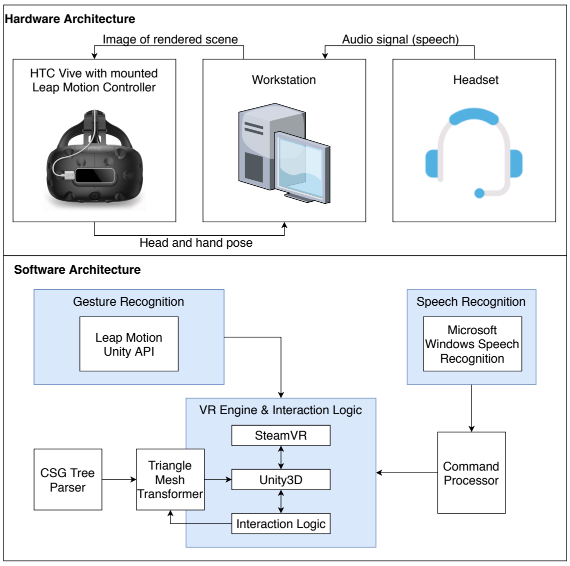


Figure 1: Architecture overview of the developed system.

is parsed by the *CSG Tree Parser* based on a JSON-based input format and transformed into a triangle mesh by the *Triangle Mesh Transformer*. This step is necessary since the used rendering engine (*Unity3D*) is restricted to this particular geometry representation. The main component (*VR Engine & Interaction Logic*) reacts on recognized gestures and voice commands, executes corresponding actions (*Interaction Logic*), and handles VR scene rendering (*Unity3D*, *SteamVR*). For example, if the user applies a basic scaling transformation to a model part consisting of a sphere, the geometric parameters of the sphere (in this case its radius) are changed based on the hand pose. Then, the model is re-transformed into the triangle mesh representation, rendered, and the resulting image is sent to the VR device.

4.2 Interaction Concept

The interaction concept is based on hand gestures (visible as two virtual gloves that visualize the user's hand and finger poses) and simple voice commands. It furthermore consists of two tools, the model tool and the tree tool, as well as the information board serving as a user guidance system.

4.2.1 Model Tool

The model tool is used for direct model manipulation and is active right after system startup. It can be divided in two different input modes, the selection mode and the manipulation mode. Modes can be switched using the voice command 'select' (to selection mode) and one of the manipulation commands 'scale', 'translate' and 'rotate' (to manipulation mode). Figure 2 details the interaction concept of the model tool.

The main reason for separating interactions that way is the ambiguity of hand gestures. In some cases, the recognition system was unable to separate between a model part selection and a translation or rotation transformation since both use the same grasp gesture. The grasp gesture was chosen since initial tests which involved basic recognition tasks for all supported gestures revealed that it is the most robustly recognized gesture available.

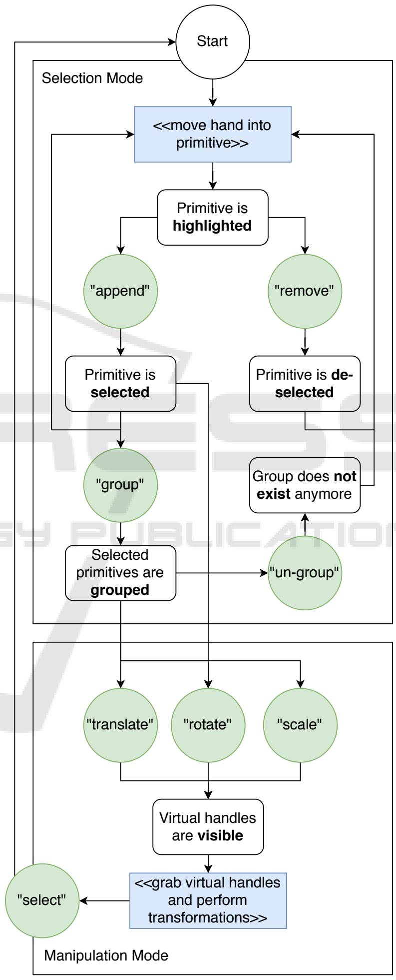
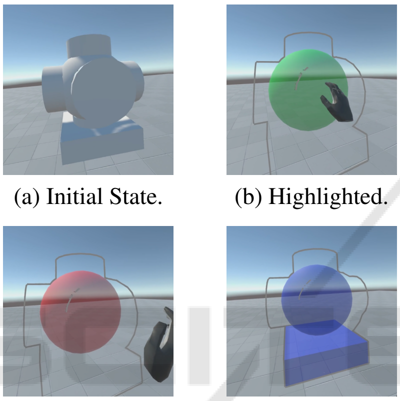


Figure 2: Interaction concept of the model tool.

Selection Mode. The initial application state can be seen in Figure 3 (a): The triangle mesh of the loaded model is displayed in grey. The user can enter the selection mode by saying the word 'select'. The mode change is additionally highlighted through an on-screen message. Using virtual hands, the user can enter the volume of the model which highlights the hovered primitives in green (Figure 3 (b)). This imitates the hovering gesture well known from desktop-based mouse interfaces. Once highlighted, the user can append the primitive to the list of selected primitives by using the voice command 'append'. Selected primitives are rendered in red (Figure 3 (c)). This way, multiple primitives can be selected. In order to remove a primitive from the selection, the user's hand must enter the primitives volume and use the voice command 'remove'. Multiple selected primitives can be grouped together with the voice command 'group' which is useful in situations where the selected primitives should behave like a single primitive during manipulation, e.g., when a rotation is applied. A group is displayed in blue (see Figure 3 (d)) and can be dissolved by saying 'un-group'.

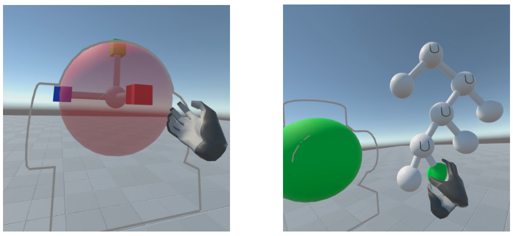


(a) Initial State. (b) Highlighted.

(c) Selected. (d) Grouped.

Figure 3: Illustration of all possible states of a primitive.

Manipulation Mode. For the manipulation of selected or grouped primitives (in the following: a model part), three basic transformations are available: translation, rotation and scaling. The currently used transformation is selected via a voice command ('translate', 'rotate' and 'scale'). The manipulation mode is entered automatically after invoking one of these commands. The selected transformation is highlighted through the displayed Virtual Handles (Conner et al., 1992) (three coordinate axes with a small box at their ends, pointing in x-, y-, and z-direction as shown in Figure 4a).
Rotation. Per-axis rotations are done by grabbing the small boxes at the end of each coordinate axis and performing a wrist rotation. The rotation is directly applied to the corresponding model part axis. Alternatively, the sphere displayed at the center of the coordinate system can be grabbed and rotated. The sphere rotations are directly applied to the model part. This by-passes the restrictions of per-axis rotations and allows for faster manipulation at the cost of precision.
Translation. Per-axis translations work like per-axis rotations. However, instead of rotating the wrist, the grabbed boxes can be moved along the corresponding axis which results in a direct translation of the model part. The sphere in the center of the coordinate axis can be grabbed and moved around without any restrictions. Resulting translations are directly applied to the model part.
Scaling. Per-axis scaling works similar to per-axis translation. The use of the sphere at the center of the coordinate system is not supported since it cannot be combined with a meaningful gesture for scaling. The transformation to use can be switched in manipulation mode by invoking the aforementioned voice commands. If the primitive selection or group should be changed, a switch to the selection mode is necessary.



(a) Virtual Handles. (b) Highlighting.

Figure 4: Virtual handles (a) and the tree tool (b).

4.2.2 Tree Tool

The tree tool displays a representation of the model's CSG tree using small spheres as nodes. It appears above the user's left hand when the user holds it upwards. Each leaf node corresponds to a primitive, the inner nodes represent the Boolean operators. Operation nodes have textures that depict the operation type ($\cup, \cap, -$). The user can change the operation type of a node by grabbing the corresponding sphere and invoking one of the following voice commands ('change to union', 'change to inter', 'change to sub'). The tree tool also allows highlighting multiple primitives at once by grabbing their parent node. Once primitives are highlighted, their corresponding nodes are displayed in green as well (see Figure 4b).

4.2.3 Information Board

The Information Board depicts the current state of the application, the manipulation task, and all voice commands including their explanations. The board is always visible as a billboard and helps the user to memorize voice commands and to be aware of the current interaction mode.

5 EVALUATION

A usability study was conducted in order to evaluate the proposed interaction concept. Its goal was to validate whether the concept is easy to understand, also for novices in the field of VR and CAD, and whether the combination of hand gesture- and voice control is perceived as a promising idea for intuitive CAD modeling.

5.1 Participants

Five student volunteers, three females and two males participated in the study. Two of them never used VR headsets before. Four participants have a background in Computer Science, one participant is a student of Molecular Life Science. Details about the participants are depicted in Table 1.

Table 1: Study participants. Abbreviations: Background in Computer Science (CS), User Experience Design (UXD) and Molecular Life Science (MLS). Experience in Computer Aided Design (CAD) or Virtual Reality (VR).

Gender	Age	Background	VR	CAD
male	30-40	UXD	Yes	No
female	20-30	CS	Yes	Yes
female	20-30	MLS	No	No
female	20-30	CS	No	No
male	20-30	CS	No	No

5.2 User Study Setup

Participants could move around freely within a radius of two square meters. The overall study setting is shown in Figure 5. As an introduction, the participants were shown a simple CAD object (Object 1 in Figure 6) together with the information board to familiarize themselves with the VR environment. Upcoming questions were answered directly by the study director.

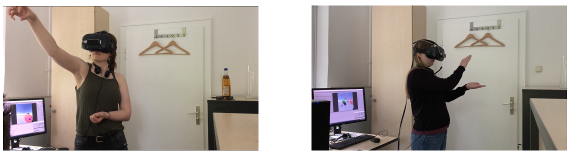


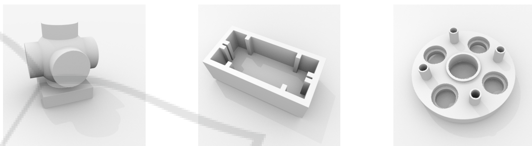
Figure 5: Participants conducting the user study.

After the introduction, the participants were asked to perform selection and modification tasks (see Table 2 for the list of tasks) where they were encouraged to share their thoughts and feelings using a Thinking-Aloud methodology (Boren and Ramey, 2000). The tasks were picked such that all possible interactions with the prototype were covered. After completing these tasks, two more objects (Objects 2 and 3 in Figure 6) were shown to the participants, and they were asked to do the same tasks as with Object 1. The whole process was recorded on video for further research and evaluation purposes.

The participants' actions, comments and experienced difficulties were logged in order to distill the advantages and disadvantages of the proposed interaction concept. In addition, problems that came up during the test could be directly addressed and discussed.

Table 2: Tasks for the participants.

Tasks
Try to select the circle and the cube together.
Deselect the selected primitives.
Select the same primitives using the object tree.
Rotate a primitive by 90 degrees.
Scale a primitive.
Move a cylinder.
Change an operator from union to subtraction.



(a) Object 1. (b) Object 2. (c) Object 3.

Figure 6: User study objects (Friedrich et al., 2019).

5.3 Interviews

After finishing the aforementioned tasks, an interview was conducted in order to gain further insight into the participants' user experience. Each interview started with general questions about pre-existing knowledge of VR and CAD software which should help with the classification of the given answers (see Table 3). Statements were assembled based on the catalogue of questions (Bossavit et al., 2014) (see Table 4) which served as an opener for discussion. Interviews were digitally recorded and then transcribed.

Table 3: Selected questions for the interview.

Questions
How much experience do you have in virtual reality applications?
How much experience do you have in CAD/3D modeling software?
How do you see our control method compared to a mouse based interface?
What is your opinion on pointer devices?

5.4 Results

All results discussed in the following were derived from the interview transcriptions and video record-

Table 4: Statements given to initiate discussion.

Statements
I found the technique easy to understand.
I found the technique easy to learn.
I found it easy to select an object.
I found it easy to group multiple objects.
I found translation easy to do.
I found rotation easy to do.
I found scaling easy to do.
I found the tree tool easy to understand.
The tree tool was useful to understand the structure of the object.
The change of the Boolean operators was easy to understand.

ings. In general, the interaction design was well received by the participants even in this early stage of prototype development. All users stated that they understood the basic interaction idea and all features of the application.

However, it was observed that users had to get used to the separation between selection and manipulation mode which was not entirely intuitive to them. For instance, some participants tried to highlight primitives in the edit mode. In this case, additional guidance was needed. All users stated that the colorization of objects is an important indication to determine the state of the application and which actions can be executed.
The information board was perceived as being very helpful but was actively used by only three participants. Two users struggled with the information board not always being in the field of view which necessitated a turn of the head to see it. Only one user realized that the current state of the application is displayed there as well.

Hand gesture-based model manipulation was very well received by the users. The hover and grab gestures did not need any explanation. Four users had trouble using their hands due to sensor range issues.
The voice control was evaluated positively. However, participants reported that the command recognition did not always work. The overall success rate was 70.7% (see Table 5). Interestingly, both male participants have significantly higher success rates than the three female participants which might be due to different pitches of their voices.

In addition, some command words like 'append' confused the users, because they expected a new primitive to be appended to the object. This indicates that an intuitive voice command design is essential but, at the same time, hard to achieve.

The manipulation technique including the Virtual Handles was also evaluated positively. All users highlighted the intuitiveness of translation and scaling in particular. Two participants missed the functionality of uniform scaling. One participant needed an extra explanation in order to understand the intention of the sphere in the center (axis-free object manipulation).

Table 5: Overview of voice control success rates (success rate $= \frac{\text{recog.}}{\text{recog.+not recog.}}$). Success means, that a command was successfully recognized by the system.

User	recognized	not recog.	Success Rate
P1	60	12	83.3%
P2	52	35	59.8%
P3	49	23	68.1%
P4	55	27	67.1%
P5	36	12	75.0%
	252	109	∅70.7%

The rotation was not fully transparent to the participants. All mentioned that they did not immediately recognize the selected rotation axis and instead of rotating the cubes at the end of the handle tried to move the complete handle.

The tree tool was evaluated positively. All participants used it to understand and manipulate the structure of the object. One participant liked that it was easy-to-use and always accessible. Two participants rated it as an intuitive tool without having an in-depth knowledge of what exactly a CSG tree is. However, one participant could not directly see the connection between the spheres in the tree and the corresponding primitives. Thus, we could observe users hovering random nodes to see the assigned primitive, trying to select specific primitives. Three users had difficulties in understanding the Boolean set-operations.

The participants showed different preferences for highlighting primitives. Two users rather used the tree tool while the others favored the hover gesture-based approach. The selection as well as the de-selection was rated positively mostly for being simple to use. Especially while dealing with more complex objects, it was perceived more difficult to select the desired primitive. This was also due to the size of the elements and due to a smaller distance to the next primitive, which makes hovering more difficult. It was also observed that in unfavorable cases, the selection hand covered the left hand so that the sensor could no longer detect its palm. As a result, the tree disappeared. This made it more difficult for users to select a sphere in the tree.

5.5 Discussion

5.5.1 Limitations

The accuracy of hand tracking is very important. Even though only very simple gestures need to be detected, tracking problems occurred repeatedly for no specific or obvious reason. This led to problems like unwanted transformations, especially when the end of a gesture was not detected correctly. However, the study participants were able to gain comprehensive insights in the application's interaction flow.

Furthermore, the group of participants was rather small and none of them had any experience with existing mid-air interaction techniques. Future versions of the prototype should be tested with more users, also taking different target user groups into account (CAD modelling experts and beginners, ...). In addition, tasks should be performed also using standard CAD modeling software for comparison.

5.5.2 Findings

The reported flat learning curve shows the potential of the proposed approach. After just a few explanations and a few minutes of training, users were able to modify 3D objects, even those participants without VR or CAD modelling experience.

The Virtual Handle approach is a suitable and intuitive three DOF manipulation tool which can be used for translation, rotation, and scaling in virtual environments. For those actions, our prototype needs to be improved, since the implementation of the rotation transformation was not perceived well.

Using additional voice commands for interacting with the system is a good way to simplify the graphical user interface. However, voice interfaces must be designed carefully in order to be intuitive while command recognition must work flawlessly.

The examination and manipulation of objects using the CSG tree structure works well even for novices. The tree tool can be understood and used without having any knowledge about the theory of CSG trees. However, the system quickly reaches its limits if the object consists of many small primitives. Furthermore, some users reported that they had difficulties imagining intersection and subtraction operations in advance. In addition, the more complex the objects become, the more difficult it is to manipulate them using the proposed method.

6 CONCLUSION

In this work, a novel, multi-modal mid-air manipulation technique for CSG-based 3D objects was presented. To make the interaction as intuitive as possible, the developed prototype is based on three control elements: Unimanual gesture controls via the demonstrated Virtual Handles, voice control and a virtual widget we named 'the tree tool' for direct CSG tree manipulation. We conducted a qualitative user study which showed that users learn and understand our manipulation method very quickly. For future work, we would like to add missing functionality (e.g., for adding new primitives), evaluate our approach by conducting quantitative and comparative user studies, and extend the voice command concept more in the direction of natural language dialogue systems.

REFERENCES

Boren, T. and Ramey, J. (2000). Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278.

Bossavit, B., Marzo, A., Ardaiz, O., De Cerio, L. D., and Pina, A. (2014). Design choices and their implications for 3d mid-air manipulation techniques. *Presence: Teleoperators and Virtual Environments*, 23(4):377–392.

Bowman, D., Kruijff, E., LaViola Jr, J. J., and Poupyrev, I. P. (2004). *3D User interfaces: theory and practice, CourseSmart eTextbook*. Addison-Wesley.

Bowman, D. A. and Hodges, L. F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 35–ff.

Bowman, D. A. and Hodges, L. F. (1999). Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing*, 10(1):37–53.

Caputo, F. M. (2019). *Gestural interaction in virtual environments: User studies and applications*. PhD thesis, University of Verona.

Chu, C.-C., Dani, T. H., and Gadh, R. (1997). Multimodal interface for a virtual reality based computer aided design system. In *Proceedings of International Conference on Robotics and Automation*, volume 2, pages 1329–1334. IEEE.

Conner, B. D., Snibbe, S. S., Herndon, K. P., Robbins, D. C., Zeleznik, R. C., and Van Dam, A. (1992). Three-dimensional widgets. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 183–188.

Friedrich, M., Fayolle, P.-A., Gabor, T., and Linnhoff-Popien, C. (2019). Optimizing evolutionary csg tree extraction. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '19, page 1183–1191, New York, NY, USA. Association for Computing Machinery.

Houde, S. (1992). Iterative design of an interface for easy 3-d direct manipulation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 135–142.

Kim, T. and Park, J. (2014). 3d object manipulation using virtual handles with a grabbing metaphor. *IEEE computer graphics and applications*, 34(3):30–38.

Kou, X., Xue, S., and Tan, S. (2010). Knowledge-guided inference for voice-enabled cad. *Computer-Aided Design*, 42(6):545–557.

LaValle, S. (2017). *Virtual Reality*, volume 1. Cambridge University Press.

Lee, M., Billinghurst, M., Baek, W., Green, R., and Woo, W. (2013). A usability study of multimodal input in an augmented reality environment. *Virtual Reality*, 17(4):293–305.

Mendes, D., Caputo, F. M., Giachetti, A., Ferreira, A., and Jorge, J. (2019). A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, volume 38, pages 21–45. Wiley Online Library.

Mendes, D., Relvas, F., Ferreira, A., and Jorge, J. (2016). The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 261–268.

Requicha, A. G. (1980). Representations for rigid solids: Theory, methods, and systems. *ACM Comput. Surv.*, 12(4):437–464.

Robinett, W. and Holloway, R. (1992). Implementation of flying, scaling and grabbing in virtual worlds. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 189–192.

Song, P., Goh, W. B., Hutama, W., Fu, C.-W., and Liu, X. (2012). A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1297–1306.

Wang, R., Paris, S., and Popović, J. (2011). 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 549–558.

Xue, S., Kou, X., and Tan, S. (2009). Natural voice-enabled cad: modeling via natural discourse. *Computer-Aided Design and Applications*, 6(1):125–136.

Zheng, J., Chan, K., and Gibson, I. (2000). A vr-based cad system. In *International Working Conference on the Design of Information Infrastructure Systems for Manufacturing*, pages 264–274. Springer.

Zhong, Y. and Ma, W. (2004). An approach for solid modelling in a virtual reality environment. In *Virtual and Augmented Reality Applications in Manufacturing*, pages 15–42. Springer.