Computers in Biology and Medicine 145 (2022) 105429

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A comparison of input devices for precise interaction tasks in VR-based surgical planning and training

Mareen Allgaier [a,*,1], Vuthea Chheang [a,1], Patrick Saalfeld [a], Vikram Apilla [a], Tobias Huber [b], Florentine Huettl [b], Belal Neyazi [c], I. Erol Sandalcioglu [c], Christian Hansen [a], Bernhard Preim [a], Sylvia Saalfeld [a]

[a] *Department of Simulation and Graphics, Faculty of Computer Science, University of Magdeburg, Universitätsplatz 2, 39106, Magdeburg, Germany*
[b] *Department of General, Visceral and Transplant Surgery, University Medicine of the Johannes Gutenberg-University Mainz, Mainz, Germany*
[c] *Department of Neurosurgery, University Hospital Magdeburg, Magdeburg, Germany*

ARTICLE INFO

Keywords:
Virtual reality
Input devices
Surgical planning
Surgical training

ABSTRACT

To exploit the potential of virtual reality (VR) in medicine, the input devices must be selected carefully due to their different benefits. In this work, input devices for common interaction tasks in medical VR planning and training are compared. Depending on the specific purpose, different requirements exist. Therefore, an appropriate trade-off between meeting task-specific requirements and having a widely applicable device has to be found. We focus on two medical use cases, liver surgery planning and craniotomy training, to cover a broad medical domain. Based on these, relevant input devices are compared with respect to their suitability for performing precise VR interaction tasks. The devices are standard VR controllers, a pen-like VR Ink, data gloves and a real craniotome, the medical instrument used for craniotomy. The input devices were quantitatively compared with respect to their performance based on different measurements. The controllers and VR Ink performed significantly better than the remaining two devices regarding precision. Qualitative data concerning task load, cybersickness, and usability and appropriateness of the devices were assessed. Although no device stands out for both applications, most participants preferred using the VR Ink, followed by the controller and finally the data gloves and craniotome. These results can guide the selection of an appropriate device for future medical VR applications.

1. Introduction

Virtual reality (VR) has been applied for medical applications, comprising various areas such as education and training [1,2], diagnosis [3], surgical planning [4] as well as treatment [5] and rehabilitation [6]. Hereby, VR applications benefit from the possibility of direct interaction with 3D models without harming patients or consuming resources in comparison to e.g. cadaver training. Regarding surgery planning or training, VR facilitates the exploration of different strategies where single steps can be undone or corrected.

Different aspects from the field of human-computer interactions have to be considered to design and implement a VR application. Among them are interaction tasks which are primitive inputs performed by the user [7]. A task is performed with an input device and interaction technique, describing the way of using the device. To have a suitable and effective system, these three aspects must be well-chosen. In our work, we define common interaction tasks for medical VR applications and use them to compare input devices. Different interaction techniques are not compared, but we stick to common design choices. For the choice of input device, the specific tasks related to an application as well as its purpose have to be considered. Based on a specific task, some devices might be more supportive and efficient due to the corresponding hand position. Besides the task, the purpose of the medical application also plays a role. On the one hand, a general device benefits from its flexibility and can be used for several tasks. On the other hand, a highly specialized device such as an endoscope [8] or laparoscopic device [9] that is similar to a real medical instrument, is crucial for medical training applications. Due to this trade-off, we compare different input

* Corresponding author.
E-mail address: mareen.allgaier@ovgu.de (M. Allgaier).
[1] These authors contributed equally.

https://doi.org/10.1016/j.compbiomed.2022.105429
Received 26 January 2022; Received in revised form 6 March 2022; Accepted 19 March 2022
Available online 23 March 2022

devices in the context of medical applications to investigate their suitability for common tasks.

In contrast to existing studies [10–12], we do not just compare devices held in *precision grip* and devices held in *power grip* but, also hand gestures. We focus on pointing tasks [10,12], sketching [11] and other tasks that are common for medical applications. Based on the different tasks we want to examine which device or rather grip is most suitable. Besides measuring the precision, we also use usability questionnaires to assess how well the measured data corresponds to the participant's perception. In the study, we focus on planning and training and selected two relevant and representative use cases: liver surgery planning and craniotomy training.

The use case liver surgery planning, is representative for therapy planning where the user can directly interact with a 3D model as well as 2D image data. Liver surgery is particularly challenging due to the complex vascular structures and the strong blood supply of the liver. Therefore, it usually requires very careful software-based planning based on medical 3D image data. Huettl et al. [13] showed that VR enables a good anatomic orientation and has a better usability as well as user experience than 3D PDFs or 3D printed models. While being specific, the task of creating an incision line and resection plane can easily be applied to other tumor resection surgeries. The second application is for craniotomy training. Craniotomy is the process of temporarily removing a section of the skull to access and treat underlying structures [14]. Afterwards, the bone flap is replaced. This procedure was selected, as it is part of the majority of brain surgeries, such as brain tumor removal and treatment of intracranial vessel diseases like aneurysms.

The two selected use cases are also representative w.r.t. the comprised interactions and tasks. Basic interactions and tasks like selecting, grabbing, scaling and rotating are often part of exploring anatomical structures. Furthermore, the included task of precisely drawing is similar to following a trajectory which is important for medical tasks like navigating in air-filled structures [15]. Another similar task is cutting which is part of every surgery such as in craniofacial surgery [16]. Thus, virtual resection where different cutting strategies can be explored and compared is essential. Hereby, consequences, e.g. the effect on structures at risk like blood vessels, can be analyzed. Consequently, the two applications were chosen based on their relevance and the involved interactions.

In this paper, we analyze relevant medical interactions and investigate which devices are appropriate in the medical context as well as their task-specific benefits and drawbacks. The results of this study can guide further developments of medical VR applications.

2. Related work

There are several virtual surgical planning and training applications in medical areas such as orthopedic surgery [17–19], neurosurgery [20,21], general surgery [22–24] as well as oral and maxillofacial surgery [2,25]. Although previous studies reveal the benefits of immersive VR applications using head-mounted displays (HMDs), not all of the previously mentioned approaches are immersive applications [1,13,26]. Common tasks in these applications are reaming [17,18] and drilling [2], careful positioning of a plate [19], navigating and applying a clip [20], placement of screws [21], and laparoscopic tasks such as touching, grasping, cutting and fine dissection [22–24].

In the next part we describe VR input devices and input techniques, as well as their applications with focus on VR in medicine.

An intuitive way of interacting are hand gestures. Sousa et al. [27] uses gestures to interact with medical image data, whereas Johnsen et al. [28] combines gesture and speech to interact with virtual patients. Instead of using additional cameras, hand interaction can also be implemented with data gloves as Chheang et al. [29] proposes for surgical planning.

Another common type of input devices are controllers usually associated with HMDs. Adams et al. [30] uses controllers to explore and manipulate 2D and 3D image data. Controllers can also be used for specific therapy planning, like clipping of intracranial aneurysms [4]. Furthermore, there are stylus devices for precise interactions such as drawing [11] or fine-motor tasks e.g. in micro-surgical procedures [31,32] or dental surgery training [2]. Stylus devices can either be grounded haptic controllers, such as the Geomagic device, or midair devices, such as VR Ink, Massless Pen, Wacom VR Pen, zSpace Ink, and some non-commercial devices [33–35]. Finally, there are highly specialized devices for training applications. An essential example is the Simball used for laparoscopic interventions [9,36]. Another example is a modified controller to simulate an endoscope [8].

The variety of input devices and applications shows that the choice of an appropriate device has to be considered carefully. To assess benefits and drawbacks, several studies have been conducted. Some of them compare the precision of devices held in power grip and precision grip [10–12] for pointing or sketching tasks. Further studies compare gestures and conventional interaction [37] or gestures and VR controllers [38]. In our comparison, we use devices held in power grip and precision grip as well as hand gesture interactions and additional tasks that are relevant for medical applications such as rotation. Furthermore, we investigate the perceived usability besides precision. In addition, we cover different medical applications and thus interaction tasks, and compare devices that differ in their grip as well as generalizability.

In the following, related work for our use cases liver surgery planning and craniotomy training are presented.

2.1. VR systems for liver surgery planning

Preoperative planning for liver surgery is based on 2D image data acquired from computed tomography or magnetic resonance imaging. Planning with 2D image data is a challenging task because it requires high experience and skills. Most surgical planning systems provide 3D visualizations generated from these image data [39–41]. This allows the physicians to understand complex internal structures, assess the risk areas and improve their confidence. In recent years, the use of VR has advanced in a way that it can be used to provide visualizations and interactions for planning complex patient cases better and faster than the desktop-based approaches [26,42]. In addition to the use of single-user VR for liver surgery planning [43,44], collaborative VR has been utilized to support the team collaboration between users either remotely or physically co-located [45–47].

2.2. VR systems for craniotomy training

For multiple brain surgeries, the initial steps include the positioning of the patient's head and the craniotomy. These should enable an easy access point that facilitates the further steps. The correct position of the patient's head is influenced by several factors: planned surgical trajectory, position of the surgeon, gravity retraction or drainage as well as measures for avoiding potential position-related complications such as air embolism [14].

Based on a specific pathology, such as brain tumors the surgeons have to decide on an appropriate craniotomy. This procedure can be trained with a VR-based system.

For intracranial aneurysm (IA) clipping, which are critical pathologies in brain arteries, and brain tumor removal, multiple virtual applications exist, but often with focus on the clipping or tumor removal itself and not on the previous steps [4,20,31,32,48–50]. In case of brain tumors, several approaches have shown that the planning of the location and size as well as the understanding of the relation between tumor and bone were improved by semi-immersive virtual applications using stereoscopic displays [49,50]. Regarding IA clipping training, some applications also include virtual drilling and the craniotomy [20,31,32,48]. In contrast to physical simulations, these applications are virtual [31,32,48], mainly using stereoscopic displays, or hybrid [20], using a VR headset and printed model. Except for the hybrid application, they are not immersive, whereas in our application the user is surrounded by a virtual operating room. Several immersive VR approaches have shown that an immersive environment leads to a high sense of being present and high motivation that might result in a more frequent use [1,26].

3. Materials and methods

In the following, first the two applications are described, followed by the motivation of the choice of input devices, and finally the tasks comprised in the two applications are explained.

3.1. Medical use cases

Both applications are based on previous work and were developed in close collaboration between medical engineers and doctors [45,51]. The individual steps of the applications are illustrated in Fig. 1.

3.1.1. Liver surgery planning

Such a system usually starts with the segmentation of the liver and tumor. In our application there are already different patient cases and their corresponding segmentations integrated. Accordingly, the user first chooses a patient dataset and the corresponding liver model, and 2D medical image data are provided for exploration. As common for these planning systems, the user may draw lines on the liver model to initialize a virtual resection. Based on this, we determine the origin and directions of the resection surface by a principle component analysis according to Konrad et al. [52]. After that, the virtual resection is initialized, and a risk map visualization to the liver tumor is projected [53]. The modification of the virtual resection can be realized by two methods. It can be directly deformed on the 3D surface as well as on the 2D line representation on the 2D image slices. After modification, the user can estimate the resection and remaining volume. The resulting reconstructed 3D model representations for each part are highlighted with different colors and their volumes are displayed.

3.1.2. Craniotomy training

Although our craniotomy training application focuses on the use case IAs, the VR-based system can be easily adapted to other neurosurgical interventions. In the workflow, the user first has to explore healthy brain arteries and place a target structure (see Fig. 1). Based on this, they have to decide on a proper approach. Then, the user enters the virtual operating room, which is based on Huber et al. [36]. Here, the user has to position the patient's head. In a real surgery, the patient head is then fixated which is not necessary in the virtual application. Before cutting the bone, skin incision and the dissection of muscles takes place. In consultation with our clinical partners, we skipped this step and directly proceed with removing a section of the bone which is more relevant for training. This procedure is simulated by sketching the contour of the craniotomy hole that should be cut.

The skull model is a simplified version of the segmentation of a patient's computed tomography angiography data. The simplification mainly aimed at reducing artifacts, non-relevant anatomical structures, and the amount of vertices of the mesh. For training purposes, there are also some craniotomy templates that can serve to train a specific location. All templates were drawn by a neurosurgeon with the help of a desktop application. In a microscopic scene, the brain can be opened to evaluate the choice of the craniotomy location.

3.2. Input devices

The tasks as well as the purpose of the applications affect the choice of the input device. Our comparison comprises four input devices in total (see Fig. 2), three per application:

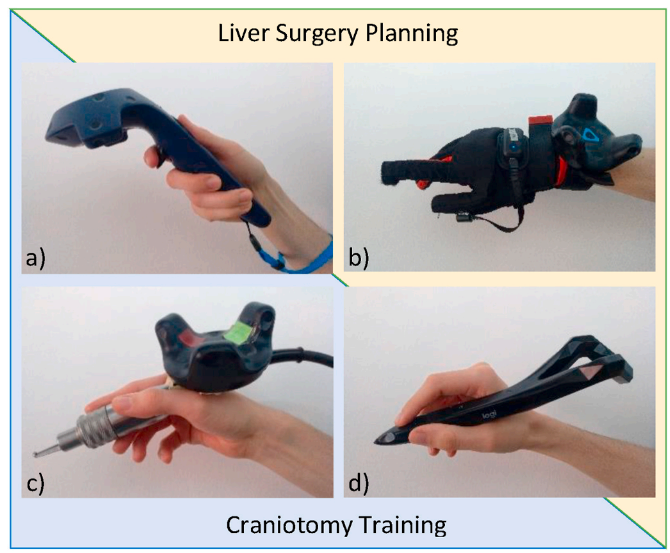

Fig. 2. Input devices: a) VR controllers, b) Manus data gloves, c) VR Ink, d) craniotome. a), b) and d) are included in the liver planning, whereas a), c) and d) are used for craniotomy training.

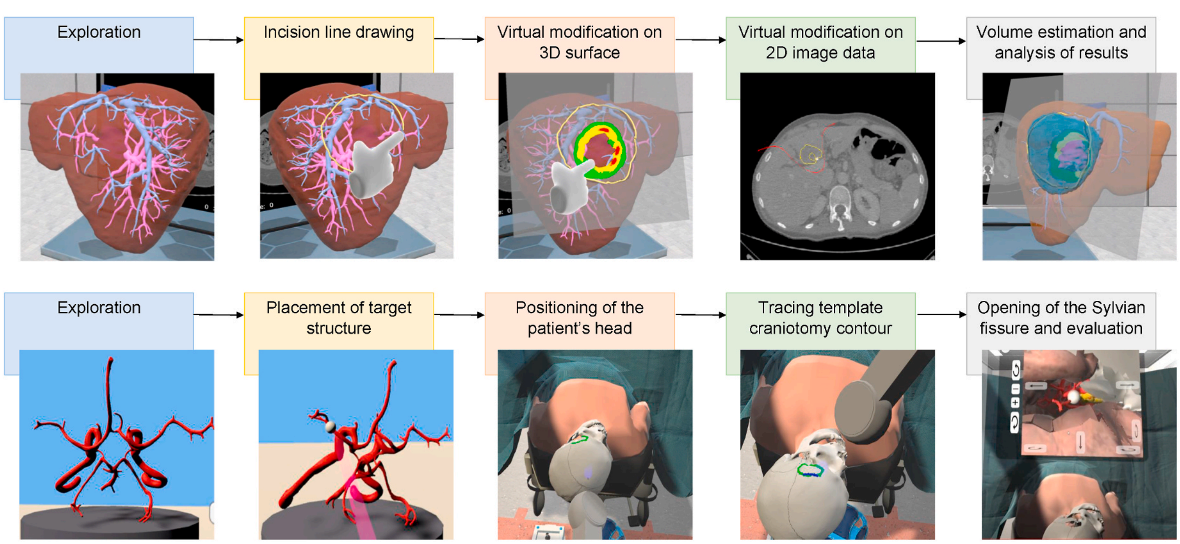

Fig. 1. Individual steps of liver surgery planning (top) and neurosurgery training (bottom) and associated interactions.

- VR controllers (both applications),
- VR Ink (both applications),
- Data gloves (liver surgery planning), and
- Craniotome (craniotomy training)

The VR controllers are used in both applications as they serve as the baseline device (Fig. 2a). Head-mounted displays always involve controllers. Thus, they are widely available, cheap and familiar to people using VR frequently. For the implementation and study, the HTC Vive Pro Eye (HTC Corporation, Taiwan) and the corresponding controllers were used. Although there are small differences between VR controllers, they have the power grip in common and have a similar setup.

The second device that is used in both applications is the VR Ink (Logitech, Switzerland) (Fig. 2d). This device was chosen because one of the main tasks in both applications is to draw, and the natural hand position for drawing cannot be provided by a controller but by a pen-like device such as the VR Ink. Especially in the craniotomy training application, a correct and realistic hand position during performing the craniotomy is crucial to build up correct muscle memory. For applications aiming to train surgeons, it is important to have a realistic device. Otherwise, according to our medical cooperation partners, frequent training with an input device held in a non-realistic hand position can lead to incorrect muscle memory. Due to muscle memory, specific motor tasks do not require conscious effort after enough repetitions. Especially for novice surgeons, an incorrect muscle memory would be fatal. However, with an appropriate device, VR systems can be used to repeat specific skills and to build up muscle memory accordingly.

Depending on the tasks, an input device held with a power grip or an input device held with a precision grip can be more useful. The power grip is known from VR controllers, but several studies have shown that the precision grip is much more precise and has a lower error rate regarding selection [10–12]. Especially for medical applications, precision is important when interacting with small or vulnerable structures. There are also other stylus devices which include haptic feedback. We decided against grounded haptic devices, as they are locally bound and an immersive VR application benefits from being able to move around. In the training application, one could use the haptic device for performing a craniotomy locally on the skull, but it is not suitable for other interactions. In this case, an additional controller would be necessary. We are aware of the importance and benefits of grounded haptic devices especially for surgical training. Nevertheless, our comparison focuses on the hand position and thus including haptic feedback would be an additional variable to investigate separately.

Next, with the Manus data gloves (Manus VR, Netherlands) (Fig. 2b), we included a device for intuitive hand interaction, especially for tasks such as pointing, selecting and scaling. To avoid the mentioned incorrect muscle memory problem, we decided to use the gloves only for the liver surgery planning application and not for craniotomy training, as they differ the most from the real device.

The idea for the fourth option also arose because of muscle memory and was chosen due to the craniotomy training. Since the best device for a training application is the actual instrument, we applied a Vive tracker on a craniotome (Fig. 2c), which is mostly used by surgeons to perform a craniotomy. However, in contrast to the other two devices, it is highly specialized to this task. Accordingly, they can be compared to find a compromise between providing a realistic hand position and being broadly applicable.

3.3. Task description

In the following, the tasks that both applications have in common and the application-specific tasks as well as their corresponding interactions are described. Furthermore, we identify the device with the highest interaction fidelity, which describes how exact real world actions are reproduced in an interactive system [54]. Here, the biomechanical symmetry, input veracity and control symmetry are considered [55]. In the following we focus on the biomechanical symmetry, as the others depend on the quality of the input devices or the implemented transfer to system effects and not on the general type of device.

3.3.1. Selecting and pointing

Most tasks that require the user to point and/or select occur when interacting with the user interface (UI). This includes clicking on buttons but also moving a slider. Besides the UI, in the craniotomy training application the users have to point at the arteries to place a target structure for craniotomy. All mentioned tasks are carried out via the trigger or trackpad of the controller, the primary button of the VR Ink or pointing and touching the button with the index finger when using the gloves.

The craniotome cannot be used for pointing and selecting. For these interactions, a VR controller held in the non-dominant hand has to be used. For pointing and selecting (e.g. pressing a button or pointing on image data) one usually uses the index finger in real-world. This is only given with the data gloves, which is therefore the device with the highest fidelity. Nevertheless, the fidelity of the other two devices is not that low, as pointing or selecting with the help of a pointing stick, held in power grip, or pen, held in precision grip, are also common.

3.3.2. Drawing

In contrast to pointing and selecting, the users do not only have to point at one location, but to precisely move their hands. This task is included in both applications when drawing the incision line on the liver and craniotomy contour, respectively. Precisely drawing the incision line midair is equivalent to planning using 3D printed models or highlighting the incision line on 2D image data. In the craniotomy training, precisely drawing the craniotomy contour simulates the procedure of cutting the bone.

The drawing tasks vary in their difficulty, as the craniotomy contour is much smaller than the incision line. Additionally, the drawing in the craniotomy training application was at a fixed location, whereas in the liver surgery planning application, the drawing was in mid-air and the user was able to change the position, rotation and scale of the liver and thus the incision line. For the controllers, VR Ink and gloves, the same interactions as for the selection are used. The craniotome in contrast only has to be moved close to the skull. For this task, only the pen-like devices offer the natural hand position, whereby the craniotome also provides the realistic weight for the craniotomy procedure. Thus, the VR Ink and craniotome have the highest interaction fidelity.

The last three tasks are specific for either the craniotomy training or the liver surgery planning.

3.3.3. Scaling

The liver can be scaled by grabbing with both devices and moving the devices towards or away from each other. Using the controller or VR Ink, the grip buttons have to be pressed while moving, whereas for the gloves the users have to make a fist. Since two devices are necessary, the VR Ink is used with a controller in the second hand. Only with the gloves, the user really grabs the model by making a fist. The pulling movement is the same with all devices. Accordingly, the gloves have the highest interaction fidelity.

3.3.4. Rotation via hand

The patient's head is rotated by grabbing it via grip buttons of the controller or VR Ink and rotating the hand holding the device. With both devices, the wrist rotation leads to a high fidelity, but due to the power grip the controller has a higher fidelity concerning the hand position.

3.3.5. Rotation via trackpad

To rotate the arteries during target selection, the trackpad of the controller or VR Ink can be used. For this task, both devices have a low fidelity, as the interaction differs a lot from rotating e.g. a 3D printed artery model. However, this interaction method is familiar due to steering via joystick.

4. Evaluation and results

The focus of our evaluation was on the input devices and their suitability for the different tasks. In the following, the hypotheses, the setup and procedure are described.

4.1. Hypotheses

The hypotheses for our study arose from the specified tasks and the described interaction fidelities of the input devices, resulting in the following:

H1. For selecting as well as grabbing and scaling, data gloves are most suitable due to the high biomechanical symmetry leading to a natural interaction.

H2. Concerning drawing, the VR Ink and craniotome perform best regarding precision due to the precision grip and are most appreciated by the users because of high interaction fidelity.

H3. For both rotations the controller is more intuitive due to its power grip leading to a more natural wrist rotation with higher biomechanical symmetry. Regarding the rotation via trackpad, the controller is more suitable due to its larger trackpad.

4.2. Setup

The evaluation was conducted with 22 participants, whose relevant data are listed in Table 1. Besides the common data such as gender, age and handedness, we also asked for their professional background and their experiences with VR. For VR experience, we have chosen three categories: "never used VR", "used VR several times", and "using it regularly". Although the tasks and interactions are based on medical applications and their specific requirements, the study was designed in a way that medical expertise is not necessary. Thus, non-experts can evaluate the suitability and appropriateness of the devices for the specific tasks, too. For example, the participants should trace an already existing incision line on the liver.

Concerning the resection plane in the liver surgery planning application, the task was not to create a medically correct resection plane, but to create a virtual resection removing the tumor completely while sparing liver tissue. Four medical experts participated: two neurosurgeons (1–5 and 11–15 years of experience), one medical student with previous knowledge in neurosurgery and one nurse.

The procedure described in the following is shown in Fig. 3. After stating their personal data, the participants were introduced to the first application. To avoid a bias caused by learning effects, the order of the applications and devices was randomized. For each drawing task, the participants had a maximum of five attempts. During the study we focused on two standardized questionnaires for an objective assessment. After completing the workflow with one device, the participants were asked to answer questions based on the *NASA Task Load Index (TLX)* [56]. These questions serve as an indicator for the mental and physical load, their success in fulfilling the given tasks and how hard they had to work to accomplish their level of performance. All questions except for temporal demand and frustration were used. The available time and thus time pressure, was the same for all devices. Consequently, temporal demand would not provide insights for the device comparison. Moreover, frustration caused by a device is mainly based on low usability which was assessed via a separated usability questionnaire.

Questions due to the *Simulator Sickness Questionnaire (SSQ)* [57] were asked after completing each application to get an impression regarding cybersickness. The first part of the post-questionnaire was an adapted version of a standardized usability scale [58]. According to our research focus, we adapted the questions in order to focus on the device instead of the system. Three questions that could not be adapted properly were discarded, resulting in seven questions concerning whether the participants would use the device frequently, the ease and complexity of using the devices, and the required learning effort. Subsequently, the users had to compare the devices according to their suitability to perform the different tasks. Finally, they were asked to rank the devices for each application.

Table 1
Characteristics of participants ($n = 22$).

Characteristics	Value	Mean
Age [years, mean (range)]	[15–38]	27.64
Gender		
Male	11	(50%)
Female	8	(36.36%)
Non-binary	3	(13.64%)
Background		
Medical experts	4	(18.18%)
Computer scientists and engineers	17	(77.27%)
Pupils	1	(4.55%)
Experience with VR		
Never used before	5	(22.73%)
Used several times	14	(63.64%)
Regular use	3	(13.64%)
Handedness		
Left	1	(4.55%)
Right	21	(95.45%)

4.3. Measurements

Besides the questionnaires, quantitative data was measured and calculated.

User performance w.r.t. the tasks involved in liver surgery planning was recorded and measured. First, the *task completion time* was recorded when the user pressed the button on the UI to start the task until the user pressed the button to complete the virtual resection modification. *Drawing resets* described how many times the user pressed the reset button to redraw the incision lines (recall Fig. 1) because of not being satisfied with the result. In addition, *drawing attempts* were counted when the user attempts to draw the incision lines on the liver surface. *Deforming attempts* described the number of attempts for virtual resection modification on both the 3D surface and 2D line representation. An attempt is defined as pressing the corresponding button and beginning the drawing or deformation. When releasing the button or being too far away for drawing or deforming, the attempt is stopped. Last but not least, *deforming error* measured how accurately the participants modified the virtual resection compared to the reference model with regard to the remaining volume of the liver.

For both applications, the *precision* and *Hausdorff distance* of the drawing results were calculated. In the craniotomy training application, the precision P of tracing a given craniotomy contour, drawn by an expert, was calculated. Both contours are defined by a set of colored vertices of the skull's triangle mesh. Based on this, the precision is calculated by:

$$P = N_{correct}/N_{drawn} \tag{1}$$

Hereby, $N_{correct}$ is the number of correctly drawn vertices, which are the drawn vertices that are also included in the given contour. N_{drawn} is the number of all drawn vertices. Consequently, the precision indicates how many vertices of the user's contour also belong to the given contour.

Precision is also calculated for the incision line in the liver surgery planning application. In this case, there is no underlying discrete mesh and thus no discrete positions are available. That is why one point of the template line and one point of the drawn line are considered the same if the distance is smaller than 5 mm. This threshold arises from the segment length when drawing. Drawing one point results in a tubular segment with a length of 5 mm and a radius of 2 mm. Thus, a threshold

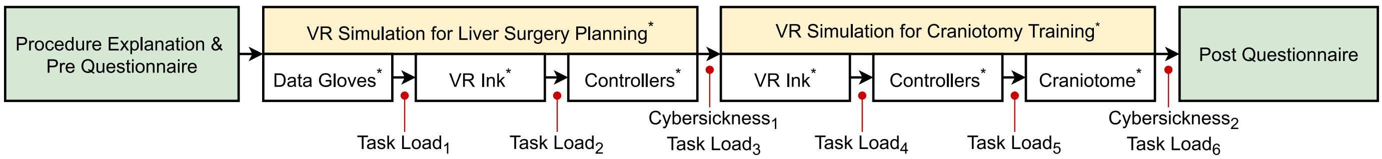

Fig. 3. Study procedure. The order of the devices and applications (marked with *) was randomized. Red: Mid-questionnaires.

of 5 mm, used on a liver that has a size of approximately 500 mm, is appropriate.

In addition to precision, the Hausdorff distance H was calculated to include a measure of the proximity of the drawn and given craniotomy contours and incision lines, respectively. H is defined as the largest distance of all minimum distances between the two curves A and B [59]:

$$\begin{aligned} H(A,B) &= max(h(A,B), h(B,A)) \\ h(A,B) &= max_{a\in A} min_{b\in B} \|a-b\| \end{aligned} \quad (2)$$

4.4. Results

This section describes the statistical analysis as well as the results of the different questionnaires.

4.4.1. Statistical analysis

To get a statistical insight into the obtained data and to see whether there are significant differences between the input devices with regard to the user performance, one-way repeated measures ANOVA were conducted [60]. The corresponding statistical results are summarized in Table 2. Fig. 4 illustrates the identified effects. Additionally, the post-hoc analyses using pairwise t-test with *Bonferroni* correction to compare between the input devices were performed after the results of significant effects [61].

4.4.1.1. Liver surgery planning. For the liver surgery planning application, we found statistically significant effects on drawing precision ($p = 0.0034$) and deforming attempts ($p = 0.00378$). We further evaluated these variables with the post-hoc analysis. The results for drawing precision revealed significant differences between the controller and the data gloves ($t = 3.37$, df $= 21$, $p < 0.009$), and between the VR Ink and the data gloves ($t = -3.48$, df $= 21$, $p < 0.007$). Regarding the deforming attempts, the results of the post-hoc analysis showed significant differences between the controller and the data gloves ($t = 3.57$, df $= 21$, $p < 0.005$), and between the VR Ink and the data gloves ($t = -3.57$, df $= 21$, $p < 0.005$).

For the other variables, there were no statistically significant effects. Hence, these null hypotheses could not be rejected. According to the descriptive results, the controllers required lower task completion time compared to the data gloves and VR Ink. It also required fewer attempts and resets of drawing the incision line. The results also indicate that deforming with VR Ink provided less error compared to the other devices. With respect to the Hausdorff distance, the controller ($\varnothing = 30.2$ mm) has a high average, but as shown in Fig. 5, the median of the controller and VR Ink ($\varnothing = 15.9$ mm) are close. The average Hausdorff distance of data gloves has a large distance, higher median, and a wider interquartile range ($\varnothing = 28.17$ mm). We found that the high average of the controller might be caused by three outliers that are above 100 mm; the data gloves have one outlier and with the VR Ink there is no distance higher than 100 mm.

Table 2
Summary of the ANOVA's results of the input devices (* denotes statistical significance ($p < 0.05$)).

Variable	df	F	p	η^2
Liver surgery planning				
Task completion time	2	1.343	0.268	0.04
Drawing resets	2	2.198	0.119	0.07
Drawing attempts	2	2.908	0.062	0.08
Precision	2	6.229	0.0034*	0.17
Hausdorff distance	2	1.102	0.339	0.03
Deforming attempts	2	6.101	0.00378*	0.16
Deforming error	2	0.023	0.977	0.000726
Craniotomy training				
Precision	2	7.136	0.00161*	0.18
Hausdorff distance	2	1.366	0.263	0.04

4.4.1.2. Craniotomy training. We found statistically significant effects regarding the precision between the input devices ($p < 0.00161$) (see also Fig. 5). The post-hoc analysis revealed statistically significant differences between the controller and the craniotome ($t = 2.73$, df $= 21$, $p < 0.038$), and between the VR Ink and the craniotome ($t = -3.72$, df $= 21$, $p < 0.004$). The craniotomy contours drawn with the VR Ink are the most precise ($\varnothing = 63.00\%$) (see also Fig. 5). Slightly less precise are the contours drawn with the controller ($\varnothing = 60.34\%$). The difference to the last device, the craniotome ($\varnothing = 53.60\%$), is much larger than between the VR Ink and the controller. Regarding the Hausdorff distance, there is no statistically significant difference. Nevertheless, the VR Ink ($\varnothing = 5.33$ mm) performs best, followed by the controller ($\varnothing = 6.29$ mm) and craniotome ($\varnothing = 6.63$ mm). It is also noticeable that the VR Ink has the smallest interquartile range and no Hausdorff distance higher than 10 mm. In contrast, the controller has five distances higher than 10 mm and the data gloves have four.

4.4.2. Questionnaire results

The participants were asked to answer the mid- and post-questionnaires (see Fig. 3). The results of the questionnaires were analyzed descriptively in the following.

4.4.2.1. Task load. Regarding mental demand, the VR Ink ($\varnothing = 7.2$, $\sigma = 3.87$) and controller ($\varnothing = 7.52$, $\sigma = 4.29$) performed better than the craniotome ($\varnothing = 8.09$, $\sigma = 4.61$) and data gloves ($\varnothing = 8.5$, $\sigma = 4.35$). The results of the physical demand were in the same order with VR Ink ($\varnothing = 6.78$, $\sigma = 3.5$), followed by controllers ($\varnothing = 7.93$, $\sigma = 5.08$) and finally data gloves ($\varnothing = 9.82$, $\sigma = 5.18$) and craniotome ($\varnothing = 11.95$, $\sigma = 4.7$). According to this order, the success of accomplishing the tasks and effort were rated with the same order.

4.4.2.2. Cybersickness. From the questionnaire the following symptoms were adopted: fatigue ($\varnothing = 1.45$, $\sigma = 0.73$), drowsiness ($\varnothing = 1.11$, $\sigma = 0.44$), headache ($\varnothing = 1.3$, $\sigma = 0.93$), eyestrain ($\varnothing = 1.61$, $\sigma = 1.04$), sweating ($\varnothing = 1.39$, $\sigma = 0.62$), nausea ($\varnothing = 1.2$, $\sigma = 0.59$), and blurred vision ($\varnothing = 1.27$, $\sigma = 0.54$). Each of them was rated with a 5-point Likert scale, where one equals no occurrence. The two symptoms that occurred most frequently, which means they were rated greater than 1, are eyestrain and fatigue. Six out of 22 participants rated them with a three or higher.

4.4.2.3. Post-questionnaire. The questionnaire, which was filled after completing all tasks with all devices, uses a 5-point Likert scale with one for disagreement and five for agreement.

Usability Regarding the statement *When using the system frequently, I would like using this device*, the VR Ink (liver: $\varnothing = 4.14$, $\sigma = 0.97$; craniotomy: $\varnothing = 4.41$, $\sigma = 0.98$) would be the device of choice in both applications, followed by the controllers (liver: $\varnothing = 3.73$, $\sigma = 1.17$;

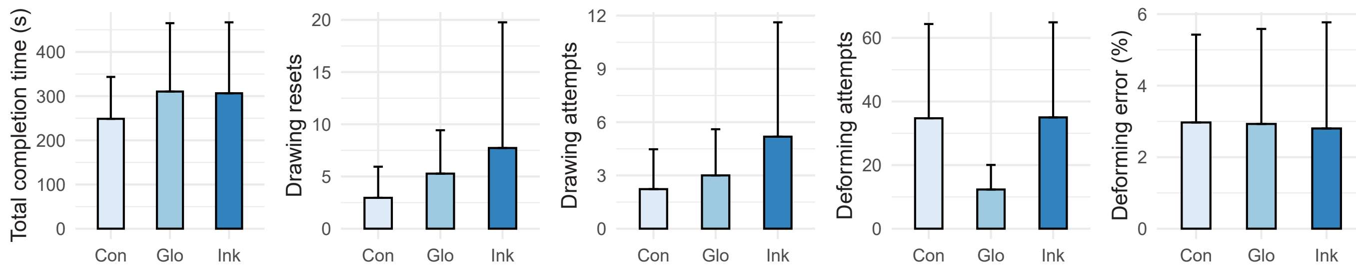

Fig. 4. Statistical results of input devices for liver surgery planning (Con: controllers; Glo: data gloves; Ink: VR Ink).

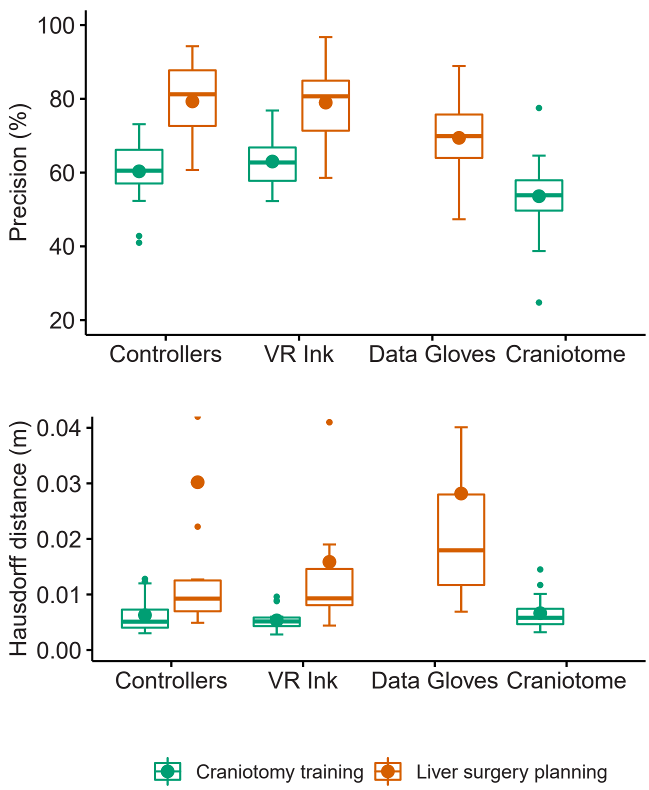

Fig. 5. Precision and Hausdorff distance of input devices for liver surgery planning and craniotomy training.

craniotomy: $\varnothing = 3.18$, $\sigma = 1, 15$). The data gloves ($\varnothing = 2.68$, $\sigma = 1.43$) and craniotome ($\varnothing = 2.27$, $\sigma = 1.17$), would not be chosen frequently.

Concerning the statement *I found interacting with this device unnecessarily complex*, rating with one equals 'not unnecessary complex'. The data gloves were rated as the most complex device ($\varnothing = 2.36$, $\sigma = 1.26$), followed by the craniotome ($\varnothing = 2.18$, $\sigma = 1.43$). The controllers and VR Ink were slightly more complex in the craniotomy training application than in the liver surgery planning application, and the controllers (liver: $\varnothing = 2.00$, $\sigma = 1.18$; craniotomy: $\varnothing = 2.09$, $\sigma = 1.03$) are more complex than the VR Ink (liver: $\varnothing = 1.64$, $\sigma = 0.81$; craniotomy: $\varnothing = 1.73$, $\sigma = 0.73$). These results are also reflected in the questions whether the devices are easy to use and easy to learn. Only for the question regarding ease of use, the craniotome was rated worse than the data gloves. This pattern is only different concerning the statement *I needed to learn a lot of things before I could get going with this device*. Here, the craniotome was rated as the one with the least learning effort.

Applicability The next part of the questionnaire focused on the applications of the devices. Most participants think that controllers ($\varnothing = 4.18$, $\sigma = 1, 11$), VR Ink ($\varnothing = 4.41$, $\sigma = 1, 07$) and data gloves ($\varnothing = 4.09$, $\sigma = 1, 04$) can also be used for other medical applications, whereas the craniotome ($\varnothing = 2.68$, $\sigma = 1, 39$) is not applicable for other scenarios.

Tasks Subsequently, the previously described tasks were used to find the most appropriate device for each specific task. All results are displayed in Fig. 6. The suitability for UI interaction, including *selecting and pointing* was rated best for the VR Ink ($\varnothing = 4.41$, $\sigma = 0.58$), followed by the controllers ($\varnothing = 4.36$, $\sigma = 1.02$) and data gloves ($\varnothing = 3.45$, $\sigma = 1.2$). The post-hoc analysis shows significant differences between the controllers and data gloves ($p < 0.0103$) and between the VR Ink and data gloves ($p < 0.0065$). There was no significant difference between the controllers and VR Ink. The craniotome was not rated, as these interactions are not possible with this device. Regarding *drawing* the craniotomy contour and incision line, the VR Ink ($\varnothing = 4.77$, $\sigma = 0.42$) is the clear favorite of the participants. In the middle range are the controllers ($\varnothing = 3.73$, $\sigma = 1.14$) and data gloves ($\varnothing = 3.27$, $\sigma = 1.32$), and the craniotome ($\varnothing = 2.95$, $\sigma = 1.11$) comes in last. We found statistically significant differences between the VR Ink and controllers ($p = 0.011$), between the VR Ink and data gloves ($p < 8.3e^{-5}$), and between the VR Ink and craniotome ($p < 1.6e^{-6}$). For *scaling* the liver, the controllers ($\varnothing = 4.41$, $\sigma = 0.58$) were rated as the most suitable device, followed immediately by the data gloves ($\varnothing = 4.36$, $\sigma = 0.83$). The VR Ink ($\varnothing = 3.86$, $\sigma = 1.01$) in combination with one controller was the least suitable device. We found no statistically significant differences between the devices regarding the scaling. The suitability of *rotating* the patient's head via grip button and hand rotation with the controller ($\varnothing = 4.09$, $\sigma = 0.79$) and with the VR Ink ($\varnothing = 4.00$, $\sigma = 0.9$) was assessed as almost similar. For the rotation of the vessels via trackpad, the controller ($\varnothing = 4.50$, $\sigma = 0.58$) was rated as more appropriate than the VR Ink ($\varnothing = 3.95$, $\sigma = 1.15$). No statistically significant differences were found between the devices regarding this interaction task.

General assessment Finally, the participants were asked to rank the devices for each application, see Fig. 7. For liver surgery planning, 64% of the participants preferred the VR Ink. 23% preferred the gloves most, and 14% the controllers. For craniotomy training, the VR Ink was also most preferred by 73% of the participants. Both controllers and craniotome with one controller were preferred by 14% of all participants.

4.4.2.4. General feedback. Some participants gave general feedback during the study or left a comment in the questionnaire. Below, these are summarized and categorized according to the specific device.

Regarding the *VR Ink*, several participants mentioned problems reaching the buttons. Most had to change their grip to reach the trackpad and had difficulties balancing the VR Ink while using it. Using the grip button of the VR Ink as well as the controllers was cumbersome and not intuitive for several participants.

One participant compared the weight of the VR Ink (63 *g*) with the weight of the craniotome (407 *g*) and stated that a pen-like device with a weight in between these two would be perfect (controller: 203 *g*). In craniotomy training where the contour is much smaller than in the liver surgery planning, the VR Ink is too light, resulting in hand jittering that influences the line tracing too much.

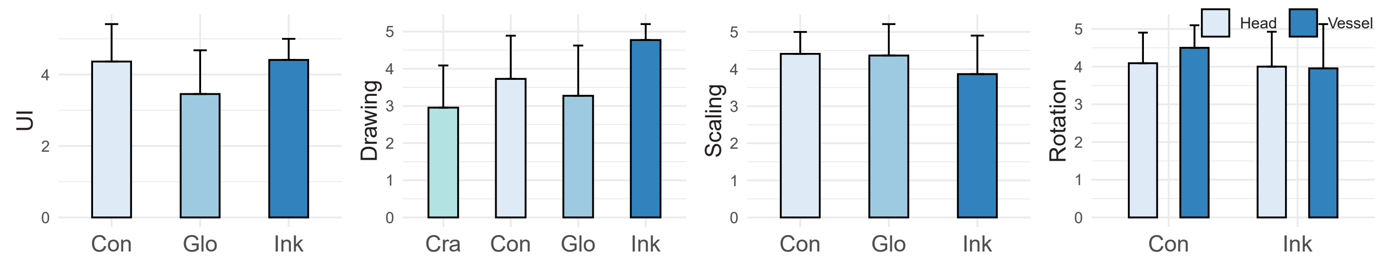

Fig. 6. Average rating of the suitability regarding the different tasks (Con: controllers; Glo: data gloves; Cra: craniotome; Ink: VR Ink).

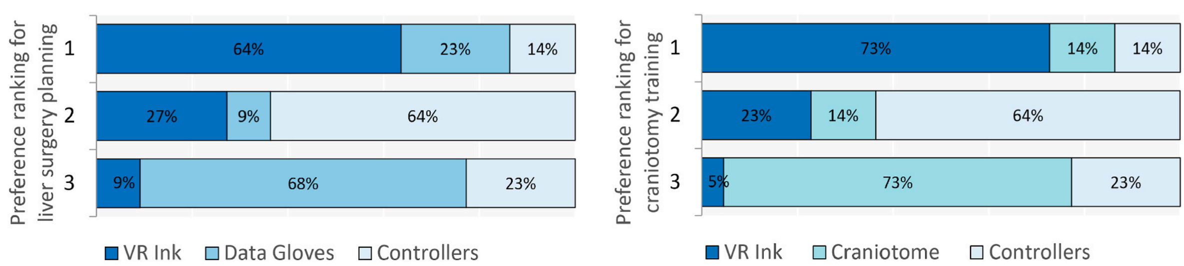

Fig. 7. Results of the ranking of the most preferred devices for liver surgery planning (left) and craniotomy training (right).

Another participant mentioned that they like the virtual VR Ink model used in the liver surgery planning application, where the currently pressed button is highlighted. With the *data gloves* the main issue was drawing. For several participants it was not intuitive to draw with the index finger. Additionally, the drawing has to be started and stopped via the thumb. If the thumb is directed forward like the index finger, drawing is enabled. However, if the distance of the thumb moves far from the index finger, the drawing is stopped. Many participants had problems with this mechanism since it is not intuitive and sometimes it is even more cumbersome due to a not correctly tracked thumb leading to strange thumb positions.

Concerning the *craniotome*, participants stated two difficulties. First, it is much more heavy than the other devices and thus unfamiliar, especially for non-experts. Second, due to tracking issues, the virtual model sometimes jittered, thus complicating tracing the line. Nevertheless, some non-experts stated that they think using the real device might be a better training for novice neurosurgeons than the other devices. For training, the medical experts also emphasized that the *controller* would lead to an incorrect muscle memory in contrast to the other two devices.

4.4.2.5. Hypotheses. The hypotheses stated in subsection 3.3 are mainly refuted: Concerning *H*1, the data gloves are as suitable as the controller for scaling, but for selecting and pointing the controller and VR Ink performed better. One reason for this could be that these two devices are, despite of their low interaction fidelity, familiar due to devices in daily life such as a presenter. *H*2 is refuted as there is no statistically significant difference between the VR Ink and controller concerning precision. The craniotome does not achieve good precision results. However, the VR Ink is significantly favored in the questionnaire. Regarding the rotations addressed in *H*3, the VR Ink and controller are approximately equally suitable. Only for rotation via trackpad, the controller performs slightly better.

4.5. Discussion

In comparison to previous studies [10–12] showing that the precision grip leads to statistically higher precision and less errors, our study does not show a significant difference between the VR Ink and the controller with respect to precision. But regarding the Hausdorff distance, with the VR Ink there are less outliers than with the controller. Additionally, the questionnaire shows that the participants preferred the VR Ink for drawing - the task for which most precision is required.

On the contrary, the craniotome, which is also held in precision grip, achieved the worst precision results. But these results can mainly be explained by two reasons. First, the craniotome is significantly heavier, which can be seen in the high physical demand, than the other devices. Only the neurosurgeons are used to handling such a heavy device precisely. Accordingly, for 2/3 of the neurosurgical experts, the craniotome was the best device regarding precision. The differences between the devices are very small for all three of these experts. Of course the weight leads to high realism which is a great benefit in a training application. However, this device can only be used for the specific task of cutting the bone. For all other tasks that might be included in an application, it is not appropriate. This leads to either inappropriate devices, or the user having to change the device for each task. The results regarding applicability also reflect this.

The second reason for the results are tracking issues. Although the applied tracker is based on the same principle as the other devices, we faced much more tracking issues with the craniotome, leading to a jittering virtual object. Even a neurosurgeon who performed best with the craniotome mentioned that it is not precise enough due to the jittering. In general, one has to be careful to not occlude the sight between tracker and lighthouse base station. Consequently, an appropriate fixation of a tracker to a medical instrument has to be thoroughly considered. Further studies would require determining the reason for jitter and possible compensations [62].

The additional measurements for the liver surgery planning showed that with the VR Ink participants had more drawing attempts than with the other devices. Comparing it with the precision, it can be concluded that the number of attempts does not imply low precision. One possible reason for a high number of drawing attempts might be the different hand position. When sketching, people often did not draw one line, but instead made several small strokes. The number of drawing resets is also the highest with the VR Ink and smallest with the controller, but it is difficult to state a reason for this or to assess it. Reasons for resetting could be non-satisfying results, but also the feeling to improve and perform better. These observations can be a starting point for further investigations to figure out the reasons behind this. The number of deforming attempts of the data gloves is significantly lower than with the other two devices. One reason could be the mentioned issues with stopping the deformation with the thumb.

The completion time was high for the VR Ink and for the gloves. One reason could that these devices are not as familiar to users as the controller. Consequently, for these devices, the participants need more time to become familiar with. Regarding deformation of the resection plane, the few drawing attempts of the data gloves are conspicuous. However, due to the combination of attempts and errors, one cannot conclude that fewer attempts correlate with less or more errors.

The cybersickness was gathered to exclude that e.g. one application performs significantly worse and thus influences the results regarding the devices. The number of participants suffering from symptoms is low and the symptoms are also not rated that high. Consequently, we can exclude correlations between cybersickness and the performance of the input devices.

Concluding all results, the VR Ink and controllers do not show statistically significant differences, except for the drawing task, where the VR Ink performs significantly better. Nevertheless, for most participants the VR Ink is the device of choice. This overall ranking coincides with the results from the usability questionnaire, where the VR Ink always reached slightly better results than the controller. In conclusion, the participants felt confident using this device.

4.5.1. Limitations

Although the study was conducted with 22 participants, only four of them were medical experts. For further insights it would be necessary to conduct the study with more physicians, especially physicians of general surgery who would use the liver surgery planning application. The fact that the neurosurgeons performed best with the craniotome in contrast to most of the non-expert participants emphasizes the importance of this. Three persons are not enough for a statistically meaningful result, but this shows that further investigations with the target group of the applications could give more insights. Nevertheless, as the tasks were designed in a way that non-experts can accomplish them, the study still provides insights into the benefits and drawbacks of the devices.

Furthermore, additional time for exploring and understanding the tasks, especially for the liver surgery planning application where the time is measured, should have been included. For most people it was difficult to understand the task immediately; consequently, a clear learning effect was visible. Some participants also mentioned their learning effect. Due to randomization of the device order we compensate this bias, but it would still be helpful for the participants to have one test run instead of measuring the time directly.

For the liver resection plane, the comments during the study revealed that it was difficult for non-experts to figure out how to deform it appropriately. Our results only compare the deformed plane with the example plane via volume, but do not consider e.g. the distance to important vessels. This would be necessary for an evaluation or a study with medical experts, but is not applicable for non-experts.

Although neither use case is extremely demanding w.r.t. complexity, they comprise representative tasks that are common in VR-based medical applications. The involved task *drawing* is (as stated in Section 1) similar to other medical tasks. However, based on the given scenario, the task complexity may differ. E.g. for the same task, such as cutting, the difficulty is different based on the chosen type of intervention. During open surgery a surgeon has a direct view of the operative field and medical instrument, whereas in minimal invasive surgery such as laparoscopic surgery the surgeon has a limited field of view and operating space [63]. Furthermore, depth perception is not possible leading to an impaired hand-eye coordination [64]. The task complexity can also vary depending on aspects such as interaction space or size of execution. E.g. the craniotomy contour is much smaller and finer than the incision line. Increasing the complexity of the applications would not necessarily increase the complexity of the tasks. Consequently, further research could focus on comparing different levels of task complexity and investigate whether this would lead to different results. However, one has to be careful as more complex tasks and thus more specialized tasks may no longer be representative.

Further studies comparing input devices in the medical context could also include haptic devices that are able to simulate the resistance of objects and tissue to provide a wider variety of possible devices.

5. Conclusion

We have presented a comparative study of input devices regarding their suitability to accomplish precise interaction tasks in two medical applications: liver surgery planning and craniotomy training. The user study with medical experts and non-experts shows that it is essential to consider the devices based on the tasks and focus of each specific use case. The user performance, questionnaire results as well as the qualitative participant's feedback revealed the following trends:

- The VR Ink and controller are superior in regards to drawing precision.
- The descriptive results show that the VR Ink performs slightly better regarding usability.

The results reveal the benefits and drawbacks of the different devices for precise interaction tasks in VR-based surgical planning and training and can provide good assistance when selecting an appropriate device for specific medical tasks.

CRediT authorship contribution statement

Mareen Allgaier: Conceptualization of this study, Methodology, Software, Investigation, Writing - Original Draft. **Vuthea Chheang:** Conceptualization of this study, Methodology, Software, Investigation, Writing - Original Draft, Formal analysis. Patrick Saalfeld: Conceptualization of this study, Writing - Review & Editing. **Vikram Apilla:** Software, Writing - Review & Editing. **Tobias Huber:** Resources, Writing - Review & Editing. **Florentine Huettl:** Resources, Writing - Review & Editing. **Belal Neyazi:** Resources, Writing - Review & Editing. **I. Erol Sandalcioglu:** Resources, Writing - Review & Editing. **Christian Hansen:** Conceptualization of this study, Writing - Review&Editing, Supervision. **Bernhard Preim:** Conceptualization of this study, Writing - Review & Editing, Supervision. **Sylvia Saalfeld:** Conceptualization of this study, Writing - Review & Editing, Supervision.

Declaration of competing interest

None Declared.

Acknowledgment

This work is partly funded by the Federal Ministry of Education and Research (BMBF) within the Forschungscampus *STIMULATE* under the grant no. 13GW0473A and 16SV8054, and the German Research Foundation (SA 3461/2-1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105429.

References

[1] T. Huber, M. Paschold, C. Hansen, T. Wunderling, H. Lang, W. Kneist, New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff, Surg. Endosc. 31 (2017) 4472–4477.

[2] M. Kaluschke, M. Su Yin, P. Haddawy, N. Srimaneekarn, P. Saikaew, G. Zachmann, A shared haptic virtual environment for dental surgical skill training, in: Proc. Of IEEE Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW), 2021, pp. 347–352, https://doi.org/10.1109/VRW52623.2021.00069.

[3] J. Orlosky, Y. Itoh, M. Ranchet, K. Kiyokawa, J. Morgan, H. Devos, Emulation of physician tasks in eye-tracked virtual reality for remote diagnosis of neurodegenerative disease, IEEE Trans. Visual. Comput. Graph. 23 (2017) 1302–1311.

[4] T.C. Steineke, D. Barbery, Microsurgical clipping of middle cerebral artery aneurysms: preoperative planning using virtual reality to reduce procedure time, Neurosurg. Focus 51 (2021) E12.

[5] A.C. Katz, A.M. Norr, B. Buck, E. Fantelli, A. Edwards-Stewart, P. Koenen-Woods, K. Zetocha, D.J. Smolenski, K. Holloway, B.O. Rothbaum, J. Difede, A. Rizzo, N. Skopp, M. Mishkind, G. Gahm, G.M. Reger, F. Andrasik, Changes in physiological reactivity in response to the trauma memory during prolonged exposure and virtual reality exposure therapy for posttraumatic stress disorder, Psychol. Trauma 12 (2020) 756–764.

[6] E.R. Høeg, J.R. Bruun-Pedersen, S. Serafin, Virtual reality-based high-intensity interval training for pulmonary rehabilitation: a feasibility and acceptability study, in: Proc. Of IEEE Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW), 2021, pp. 242–249, https://doi.org/10.1109/VRW52623.2021.00052.

[7] R.J.K. Jacob, Human-computer interaction: input devices, ACM Comput. Surv. 28 (1996) 177–179.

[8] N.W. John, T.W. Day, T. Wardle, An endoscope interface for immersive virtual reality, in: Proc. Of Eurographics Workshop on Visual Computing for Biology and Medicine, 2020, pp. 25–29, https://doi.org/10.2312/vcbm.20201167.

[9] P. Crochet, R. Aggarwal, S. Dubb, P. Ziprin, N. Rajaretnam, T. Grantcharov, K. Ericsson, A. Darzi, Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills, Ann. Surg. 253 (2011) 1216–1222.

[10] A.U. Batmaz, A.K. Mutasim, W. Stuerzlinger, Precision vs. power grip: a comparison of pen grip styles for selection in virtual reality, in: Proc. Of IEEE Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW), 2020, pp. 23–28, https://doi.org/10.1109/VRW50115.2020.00012.

[11] A. Cannavò, D. Calandra, A. Kehoe, F. Lamberti, Evaluating consumer interaction interfaces for 3D sketching in virtual reality, in: Proc. Interactivity and Game Creation, 2021, pp. 291–306, https://doi.org/10.1007/978-3-030-73426-8_17.

[12] D.-M. Pham, W. Stuerzlinger, in: V.R.S.T. Proc (Ed.), Is the Pen Mightier than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality, ACM, 2019, pp. 1–11, https://doi.org/10.1145/3359996.3364264.

[13] F. Huettl, P. Saalfeld, C. Hansen, B. Preim, A. Poplawski, W. Kneist, H. Lang, T. Huber, Virtual reality and 3d printing improve preoperative visualization of 3d liver reconstructions—results from a preclinical comparison of presentation modalities and user's preference, Ann. Transl. Med. 9 (2021).

[14] A. Raabe, B. Meyer, K. Schaller, P. Vajkoczy, P.A. Winkler, The Craniotomy Atlas, 2019. Thieme.

[15] S. Shetty, L. Panait, J. Baranoski, S. Dudrick, R. Bell, K. Roberts, A. Duffy, Construct and face validity of a novel virtual reality-based camera navigation curriculum, J. Surg. Res. 177 (2012) 191–195.

[16] S. Zachow, E. Gladilin, R. Sader, H. Zeilhofer, Draw and cut: intuitive 3d osteotomy planning on polygonal bone models, in: Proc. Of Computer Assisted Radiology and Surgery, 2003, pp. 362–369, https://doi.org/10.1016/S0531-5131(03)00272-3.

[17] J. Hooper, E. Tsiridis, J. Feng, R. Schwarzkopf, D. Waren, W. Long, L. Poultsides, W. Macaulay, G. Papagiannakis, E. Kenanidis, E. Rodriguez, J. Slover, K. Egol, D. Phillips, S. Friedlander, M. Collins, Virtual reality simulation facilitates resident training in total hip arthroplasty: a randomized controlled trial, J. Arthroplasty 34 (2019) 2278–2283.

[18] D. Panariello, T. Caporaso, S. Grazioso, G. Di Gironimo, A. Lanzotti, S. Knopp, L. Pelliccia, M. Lorenz, P. Klimant, Using the kuka lbr iiwa robot as haptic device for virtual reality training of hip replacement surgery, in: IEEE International Conference on Robotic Computing, IRC), 2019, pp. 449–450, https://doi.org/10.1109/IRC.2019.00094.

[19] A. Gupta, J. Cecil, M. Pirela-Cruz, Immersive virtual reality based training and assessment of an orthopedic surgical process, in: IEEE International Systems Conference, SysCon), 2020, pp. 1–7, https://doi.org/10.1109/SysCon47679.2020.9381832.

[20] S. Teodoro-Vite, J. Pérez-Lomelí, C. Domínguez-Velasco, A.F. Hernández-Valencia, M. Capurso-García, M. Padilla-Castañeda, A high-fidelity hybrid virtual reality simulator of aneurysm clipping repair with brain sylvian fissure exploration for vascular neurosurgery training, Simulat. Healthc. 16 (2021) 285–294.

[21] B. Xin, G. Chen, Y. Wang, G. Bai, X. Gao, J. Chu, J. Xiao, T. Liu, The efficacy of immersive virtual reality surgical simulator training for pedicle screw placement: a randomized double-blind controlled trial, World Neurosurg. 124 (2018).

[22] J.G. Frederiksen, S.M.D. Sørensen, L. Konge, M.B.S. Svendsen, M. Nobel-Jørgensen, F. Bjerrum, S.A.W. Andersen, Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial, Surg. Endosc. 34 (2019) 1244–1252.

[23] V. Chheang, P. Saalfeld, T. Huber, F. Huettl, W. Kneist, B. Preim, C. Hansen, Collaborative virtual reality for laparoscopic liver surgery training, in: IEEE International Conference on Artificial Intelligence and Virtual Reality, AIVR), 2019, pp. 1–17, https://doi.org/10.1109/AIVR46125.2019.00011.

[24] M. Li, S. Ganni, J. Ponten, A. Albayrak, A.-F. Rutkowski, J. Jakimowicz, Analysing usability and presence of a virtual reality operating room (vor) simulator during laparoscopic surgery training, in: IEEE Conference on Virtual Reality and 3D User Interfaces, VR), 2020, pp. 566–572, https://doi.org/10.1109/VR46266.2020.00078.

[25] D. Morris, C. Sewell, N. Blevins, F. Barbagli, K. Salisbury, A collaborative virtual environment for the simulation of temporal bone surgery, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI, 2004, pp. 319–327.

[26] Y. Pulijala, M. Ma, M. Pears, D. Peebles, A. Ayoub, Effectiveness of immersive virtual reality in surgical training—a randomized control trial, J. Oral Maxillofac. Surg. 76 (2018) 1065–1072.

[27] M. Sousa, D. Mendes, S. Paulo, N. Matela, J. Jorge, D.S.o. Lopes, VRRRRoom: virtual reality for radiologists in the reading room, in: Proc. Of ACMCHI Conference on Human Factors in Computing Systems, 2017, pp. 4057–4062, https://doi.org/10.1145/3025453.3025566.

[28] K. Johnsen, R. Dickerson, A. Raij, B. Lok, J. Jackson, M. Shin, J. Hernandez, A. Stevens, D. Lind, Experiences in using immersive virtual characters to educate medical communication skills, in: Proc. Of IEEE Virtual Reality, 2005, pp. 179–186, https://doi.org/10.1109/VR.2005.1492772.

[29] V. Chheang, V. Apilla, P. Saalfeld, C. Boedecker, T. Huber, F. Huettl, H. Lang, B. Preim, C. Hansen, Collaborative VR for liver surgery planning using wearable data gloves: an interactive demonstration, in: Proc. Of IEEE Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW, 2021, https://doi.org/10.1109/VRW52623.2021.00268, 768–768.

[30] H. Adams, J. Shinn, W.G. Morrel, J. Noble, B. Bodenheimer, Development and evaluation of an immersive virtual reality system for medical imaging of the ear, in: Proc. Of SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling, vol. 10951, 2019, pp. 265–272, https://doi.org/10.1117/12.2506178.

[31] A. Alaraj, C. Luciano, D. Bailey, A. Elsenousi, B. Roitberg, A. Bernardo, P. Banerjee, F. Charbel, Virtual reality cerebral aneurysm clipping simulation with real-time haptic feedback, Neurosurgery 11 (Suppl 2) (2015) 52–58.

[32] M. Gmeiner, J. Dirnberger, W. Fenz, M. Gollwitzer, G. Wurm, J. Trenkler, A. Gruber, Virtual cerebral aneurysm clipping with real-time haptic force feedback in neurosurgical education, World Neurosurg. 112 (2018) e313–e323.

[33] B. Jackson, OVR stylus: designing pen-based 3D input devices for virtual reality, in: Proc. Of IEEE Virtual Reality and 3D User Interfaces Abstracts and Workshops, VRW), 2020, pp. 13–18, https://doi.org/10.1109/VRW50115.2020.00287.

[34] S. Kamuro, K. Minamizawa, N. Kawakami, S. Tachi, Ungrounded kinesthetic pen for haptic interaction with virtual environments, in: Proc. IEEE International Symposium on Robot and Human Interactive Communication, 2009, pp. 436–441, https://doi.org/10.1109/ROMAN.2009.5326217.

[35] H. Romat, A. Fender, M. Meier, C. Holz, Flashpen: a high-fidelity and high-precision multi-surface pen for virtual reality, in: Proc. Of IEEE VR, 2021, pp. 306–315, https://doi.org/10.1109/VR50410.2021.00053.

[36] T. Huber, T. Wunderling, M. Paschold, H. Lang, W. Kneist, C. Hansen, Highly immersive virtual reality laparoscopy simulation: development and future aspects, Int. J. Comput. Assist. Radiol. Surg. 13 (2018) 281–290.

[37] J. Hettig, P. Saalfeld, M. Luz, M. Becker, M. Skalej, C. Hansen, Comparison of gesture and conventional interaction techniques for interventional neuroradiology, Int. J. Comput. Assist. Radiol. Surg. 12 (2017) 1643–1653.

[38] P. Monteiro, G. Gonçalves, H. Coelho, M. Melo, M. Bessa, Hands-free interaction in immersive virtual reality: a systematic review, IEEE Trans. Visual. Comput. Graph. 27 (2021) 2702–2713.

[39] D. Wang, D. Ma, M.L. Wong, Y.X.J. Wáng, Recent advances in surgical planning & navigation for tumor biopsy and resection, Quant. Imag. Med. Surg. 5 (2015) 640–648.

[40] Y. Mise, K. Tani, T. Aoki, Y. Sakamoto, K. Hasegawa, Y. Sugawara, N. Kokudo, Virtual liver resection: computer-assisted operation planning using a three-dimensional liver representation, J. Hepatobiliary Pancreat. Sci. 20 (2013) 157–164.

[41] B. Preim, D. Selle, W. Spindler, K.J. Oldhafer, H.-O. Peitgen, Interaction techniques and vessel analysis for preoperative planning in liver surgery, in: Proc. MICCAI, 2000, pp. 608–617, https://doi.org/10.1007/978-3-540-40899-4_62.

[42] A.J. Lungu, W. Swinkels, L. Claesen, P. Tu, J. Egger, X. Chen, A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery, Expet Rev. Med. Dev. 18 (2020) 47–62.

[43] H.G. Kenngott, M. Pfeiffer, A.A. Preukschas, L. Bettscheider, P.A. Wise, M. Wagner, S. Speidel, M. Huber, F. Nickel, A. Mehrabi, B.P. Müller-Stich, IMHOTEP: cross-professional evaluation of a three-dimensional virtual reality system for interactive surgical operation planning, tumor board discussion and immersive training for complex liver surgery in a head-mounted display, Surg. Endosc. (2021) 1–9.

[44] B. Reitinger, A. Bornik, R. Beichel, D. Schmalstieg, Liver surgery planning using virtual reality, IEEE Comput. Graph Appl. 26 (2006) 36–47.

[45] V. Chheang, P. Saalfeld, F. Joeres, C. Boedecker, T. Huber, F. Huettl, H. Lang, B. Preim, C. Hansen, A collaborative virtual reality environment for liver surgery planning, Comput. Graph. 99 (2021) 234–246.

[46] O. Bashkanov, P. Saalfeld, H. Gunasekaran, M. Jabaraj, B. Preim, T. Huber, F. Hüttl, W. Kneist, C. Hansen, VR multi-user conference room for surgery planning, in: Proc. Deutschen Gesellschaft für Computer- und Roboterassistierte Chirurgie, vol. 6, 2019, p. 7.

[47] R. Fischer, K.-C. Chang, R. Weller, G. Zachmann, Volumetric medical data visualization for collaborative VR environments, in: Proc. EuroVR, 2020, pp. 178–191.

[48] G. Wong, C. Zhu, A. Ahuja, W. Poon, Craniotomy and clipping of intracranial aneurysm in a stereoscopic virtual reality environment, Neurosurgery 61 (2007) 564–569.

[49] A. Stadie, R. Kockro, Mono-stereo-autostereo: the evolution of 3-dimensional neurosurgical planning, Neurosurgery 72 (Suppl 1) (2013) 63–77.

[50] M. Oishi, M. Fukuda, N. Yajima, K. Yoshida, M. Takahashi, T. Hiraishi, T. Takao, A. Saito, Y. Fujii, Interactive presurgical simulation applying advanced 3D imaging and modeling techniques for skull base and deep tumors, J. Neurosurg. 119 (2013) 94–105.

[51] M. Allgaier, A. Amini, B. Neyazi, I.E. Sandalcioglu, B. Preim, S. Saalfeld, VR-based training of craniotomy for intracranial aneurysm surgery, Int. J. Comput. Assist. Radiol. Surg. 17 (3) (2022) 449–456.

[52] O. Konrad-Verse, A. Littmann, B. Preim, Virtual resection with a deformable cutting plane, in: Proc. SimVis, 2004, pp. 203–214.

[53] C. Hansen, S. Zidowitz, F. Ritter, C. Lange, K. Oldhafer, H.K. Hahn, Risk maps for liver surgery, Int. J. Comput. Assist. Radiol. Surg. 8 (2013) 419–428.

[54] R. McMahan, D. Bowman, D. Zielinski, R. Brady, Evaluating display fidelity and interaction fidelity in a virtual reality game, IEEE Trans. Visual. Comput. Graph. 18 (2012) 626–633.

[55] R. McMahan, C. Lai, S. Pal, Interaction fidelity: the uncanny valley of virtual reality interactions, in: Virtual, Augmented and Mixed Reality, vol. 9740, 2016, pp. 59–70, https://doi.org/10.1007/978-3-319-39907-2_6.

[56] S. Hart, L. Staveland, Development of NASA-TLX (task load index): results of empirical and theoretical research, Adv. Psychol. 52 (1988) 139–183.

[57] R.S. Kennedy, N.E. Lane, K.S. Berbaum, M.G. Lilienthal, Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness, Int. J. Aviat. Psychol. 3 (1993) 203–220.

[58] J. Brooke, SUS: a quick and dirty usability scale, Usability Eval. Ind. 189 (1995).

[59] D. Huttenlocher, G. Klanderman, W. Rucklidge, Comparing images using the hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. 15 (1993) 850–863.

[60] J. Chambers, A. Freeny, R. Heiberger, Analysis of variance; designed experiments, in: Chapter 5 of Statistical Models in S., Wadsworth & Brooks/Cole, 1992, pp. 145–190.

[61] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B Methodol. 57 (1995) 289–300.

[62] A.U. Batmaz, M.R. Seraji, J. Kneifel, W. Stuerzlinger, No jitter please: effects of rotational and positional jitter on 3D mid-air interaction, in: Proc. Of the Future Technologies Conference (FTC), vol. 2, 2021, pp. 792–808, https://doi.org/10.1007/978-3-030-63089-8_52.

[63] O. Elhage, B. Challacombe, A. Shortland, P. Dasgupta, An assessment of the physical impact of complex surgical tasks on surgeon errors and discomfort: a comparison between robot-assisted, laparoscopic and open approaches, BJU Int 115 (2014) 274–281.

[64] M. Tonutti, D. Elson, G.-Z. Yang, A. Darzi, M. Sodergren, The role of technology in minimally invasive surgery: state of the art, recent developments and future directions, Postgrad. Med. 93 (2017) 159–167.