Applying Surgeon-centric Design to Advanced VR-based Medical Training

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieurin (Dr.-Ing.)

angenommen durch die Fakultät für Informatik

der Otto-von-Guericke-Universität Magdeburg

von

M.Sc. Mareen Allgaier

geb. am 05.04.1996 in Öhringen

Gutachter:

Prof. Dr. Christian Hansen

Prof. Dr. Torsten Wolfgang Kuhlen

Prof. Dr. Thomas Wolbers

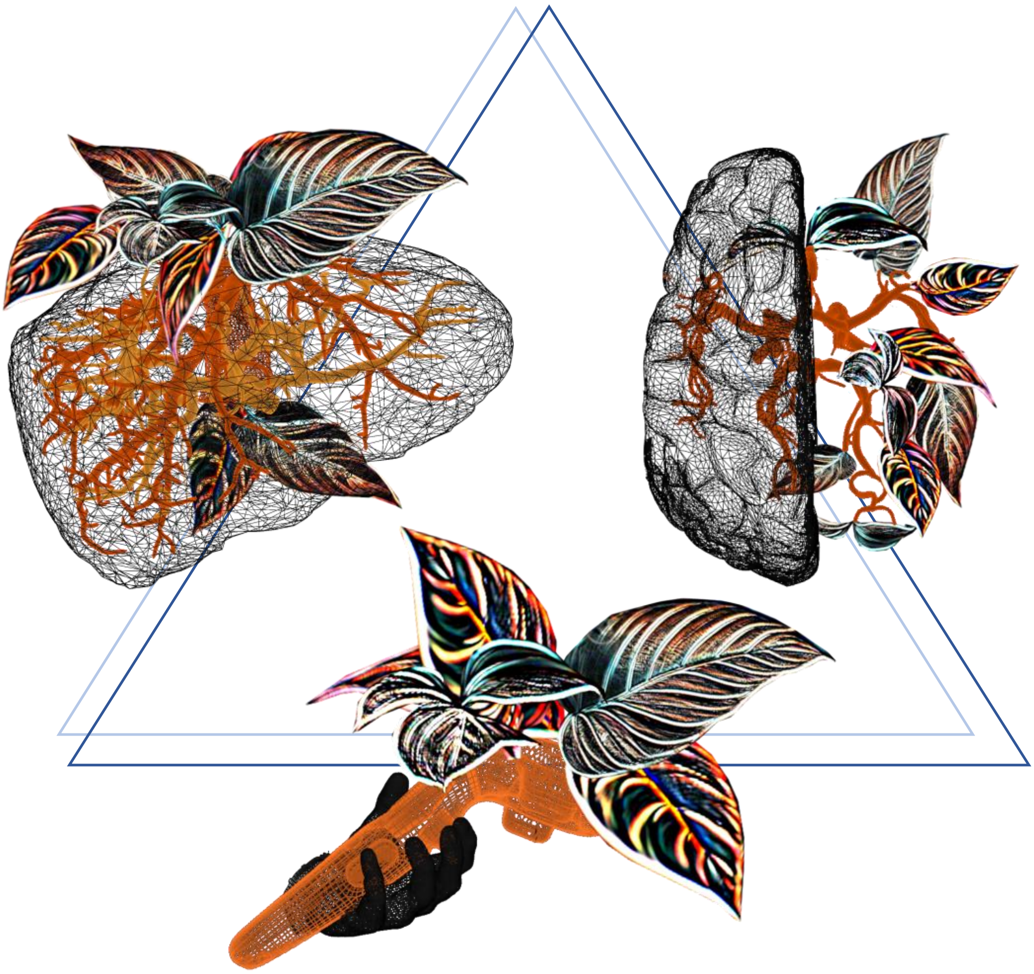
Magdeburg, 05.09.2024

Mareen Allgaier: Applying Surgeon-centric Design to Advanced VR-based Medical Training, Dissertation, © 05.09.2024

Dissertation

Mareen Allgaier

Applying Surgeon-centric Design to Advanced VR-based Medical Training



Abstract

During medical education, physicians need several years to gain various types of competencies. This includes motoric and visuospatial skills as well as declarative, procedural and situative knowledge. To gain practical routine and assurance, frequent training is essential. Because of this, simulations are created that provide specific scenarios that can be repeated and do not harm patients. Simulations can be physical, virtual or hybrid, whereof each of them has its benefits and limitations. Virtual simulations include virtual reality (VR) as an immersive variant to simulate situations that could not be trained in reality. VR-based surgical training has the potential to consider even rare anatomical variants. Consequently, surgeons then have a better competence level when they start to operate on patients. Thus, medical use cases were identified in close collaboration with medical experts where VR simulations are relevant to gain specific skills. These cases are in the two medical disciplines liver surgery and neurosurgery.

In the first part of this thesis, a training application for visuospatial skills in the context of intraoperative ultrasound is presented. This includes the simulation of ultrasound images, haptic feedback and the creation of four different scenarios to train visuospatial skills. The single training scenarios as well as the content and face validity of the training system were evaluated by medical experts. The evaluations revealed that three scenarios are appropriate for training and the main limitation is the haptic input device. Furthermore, one of these scenarios was gamified to provide a better training experience. Therefore, several game elements were discussed and two studies compared *levels of difficulty* and an interactive *kit* with the non-gamified version. The training scenario with the kit benefits from a more interactive user experience. Levels provide good feedback for progress and performance.

In the second part, a VR-based training, which addresses strategical knowledge, is presented. The trainee should gain a better understanding of the relevance of the correct access for microsurgical intracranial operations. The resulting application was evaluated by medical experts, and its usefulness became apparent.

During the development of all application prototypes, the necessity of proper input devices emerged. Therefore, a general comparison of various input devices for medical VR-based applications was conducted. The presented benefits and limitations of the devices should support the choice of a proper input device.

Close collaborations with medical experts were the basis for all presented prototypes. This surgeon-centric design and development ensure the clinical relevance of the prototypes. By addressing specific skills, the prototypes serve as additional training where other training modalities are limited or not applicable.

Zusammenfassung

Während der medizinischen Ausbildung benötigen Ärzte mehrere Jahre um verschiedene Arten von Kompetenzen zu erlangen. Dazu gehören motorische und visuell-räumliche Fähigkeiten sowie deklaratives, prozedurales und situatives Wissen. Um eine praktische Routine und Sicherheit zu erlangen, ist häufiges Training notwendig. Deshalb gibt es Simulationen, welche spezifische Szenarien zum Trainieren bereitstellen. Diese können wiederholt werden und schaden keinem Patienten. Simulationen können physikalisch, virtuell oder hybrid sein, wobei jede Variante ihre Vor- und Nachteile hat. Virtuelle Simulationen schließen auch VR Simulationen als immersive Variante mit ein. Diese simulieren Situationen, welche in der Realität nicht trainiert werden können. VR-basiertes chirurgisches Training hat das Potential, dass auch seltene anatomische Varianten berücksichtigt werden können. Demzufolge haben Chirurgen dann ein höheres Kompetenzlevel wenn sie beginnen am Patienten zu operieren. In enger Zusammenarbeit mit medizinischen Experten wurden Anwendungsfälle identifiziert, bei denen VR Simulationen notwendig sind um spezifische Fähigkeiten zu erlangen. Diese Anwendungsfälle sind in den zwei Disziplinen Leberchirurgie und Neurochirurgie.

Im ersten Teil der Dissertation wird eine Trainingsanwendung für visuell-räumliche Fähigkeiten im Kontext von intraoperativem Ultraschall vorgestellt. Dieser beinhaltet die Simulation von Ultraschallbildern, haptisches Feedback und die Erstellung von vier verschiedenen Trainingsszenarien. Die Szenarien sowie die Inhaltsvalidität und Augenscheinvalidität des Trainigssystems wurden von medizinischen Experten ausgewertet. Die Auswertung zeigt, dass drei der Szenarien für ein Training angemessen sind und die Hauptlimitation das haptische Eingabegerät ist. Darüber hinaus wurde eins der Szenarien gamifiziert, um ein besseres Trainingserlebnis zu erzeugen. Hierfür wurden mehrere Spielelemente diskutiert und zwei Studien vergleichen *Schwierigkeitsstufen* und einen interaktiven *Baukasten* mit der nicht-gamifizierten Version. Das Trainingsszenario mit dem Baukasten profitiert von einem interaktiveren Erlebnis. Schwierigkeitsstufen stellen ein gutes Feedback bezüglich des Fortschrittes und der Leistung dar.

Im zweiten Teil wird ein VR-basiertes Training, welches strategisches Wissen adressiert, vorgestellt. Der Lernende soll ein besseres Verständnis für die Relevanz des richtigen Zugangs bei mikrochirurgischen intrakraniellen Operationen erlangen. Die daraus entstandene Anwendung wurde von medizinischen Experten beurteilt, wobei sich deren Nützlichkeit abzeichnete.

Während der Entwicklung aller Anwendungsprototypen kam die Notwendigkeit passender Eingabegeräte zur Sprache. Dementsprechend wurde ein allgemeiner Vergleich verschiedener Eingabegeräte für medizinische VR-basierte Anwendungen durchgeführt. Die präsentierten Vor- und Nachteile der Geräte sollen als Unterstützung bei der Wahl eines Eingabegerätes dienen.

Enge Kollaborationen mit den jeweiligen medizinischen Experten waren die Basis für alle vorgestellten Prototypen. Chirurgen-orientiertes Design und Entwicklung stellten die kli- nische Relevanz der Prototypen sicher. Durch das Adressieren von spezifischen Fähigkeiten dienen diese als zusätzliches Training wo andere Trainingsmodalitäten limitiert oder nicht umsetzbar sind.

Contents

1

Introduction

Synopsis This chapter motivates the need for VR training simulations and gives context to the thesis. The presentation of the structure includes the research questions and objectives of the individual projects.

1.1 Motivation

Since the purchase of a VR head-mounted display (HMD) has become affordable in the 2010s, they have been used in a wide variety of areas. In the healthcare domain, their application can be divided into four areas [106]: 1) therapy, 2) rehabilitation, 3) training, and 4) prevention. Thus, the target group of these VR applications can either include patients (1 and 2), healthcare specialists (3), or the broad audience (4). Halbig et al. [106] investigated the acceptance of VR among healthcare professionals and asked for use cases where VR might be beneficial. Among others, training and education are mentioned as areas where VR has potential and at the same time little ethical concerns. This includes the acquisition of different competencies with respect to specific interventions, anatomy or interpersonal situations. VR offers a safe training environment, especially for 3D learning content or skills that can only be trained hands-on and allows the users to explore on their own. However, VR also has drawbacks. The review of Mao et al. [191] identified realistic tactile feedback as the most often mentioned limitation. This lack may hinder the transfer of the learned skills to the patients. Because of its benefits and limitations, it is important to select proper use cases and learning objectives where VR can provide additional support.

Most surgical procedures require hands-on training and repetitions. In practice, training directly on the patient during surgery is often not possible due to rare procedures or ethical concerns. In the last years, more and more ways to simulate procedures have been developed, which also led to the new paradigm *see one, simulate many, do one, teach one* instead of *see one, do one, teach one* [325, 341] (refer to Figure 1.1). Accordingly, it is important to closely collaborate with domain experts to define relevant training cases where VR is impactful. But what defines a good training simulation? Harris et al. [112] discuss important factors as well as proper testing and validation of training simulations. Even if usability and user experience might not directly affect training effectiveness, a lack of usability and bad user experience has negative effects on the training simulation. Factors that directly influence the simulation are various types of validity as well as fidelity. Harris et al. [112] state that simulations aim at replicating some aspects of reality or the task to be trained while others are not reproduced (for example, danger). Because of this, it is important to clearly define the *learning objective* and to identify relevant aspects that should be simulated in a faithful way and those which can be neglected. This can only be done in close collaboration with domain experts.

This thesis aims at improving medical education and training by providing additional VR-based training applications using an expert-oriented design. With these applications, the performance of physicians should be increased and patient safety should be improved. To do so, relevant procedures and skills that cannot be trained at the patient due to ethical or practical reasons and where no proper simulations exist were identified: visuospatial skills for intraoperative ultrasound (IOUS) and strategic knowledge for neurosurgical access. For these use cases, VR-based training applications were implemented and evaluated accordingly. Furthermore, parts of the thesis investigate how gamification can be included. This aims at increasing motivation and engagement to improve the training experience and ideally the learning outcome.

[1]Cliparts: https://openclipart.org/. Creative Commons Zero 1.0 License. Last access: 08.11.2023

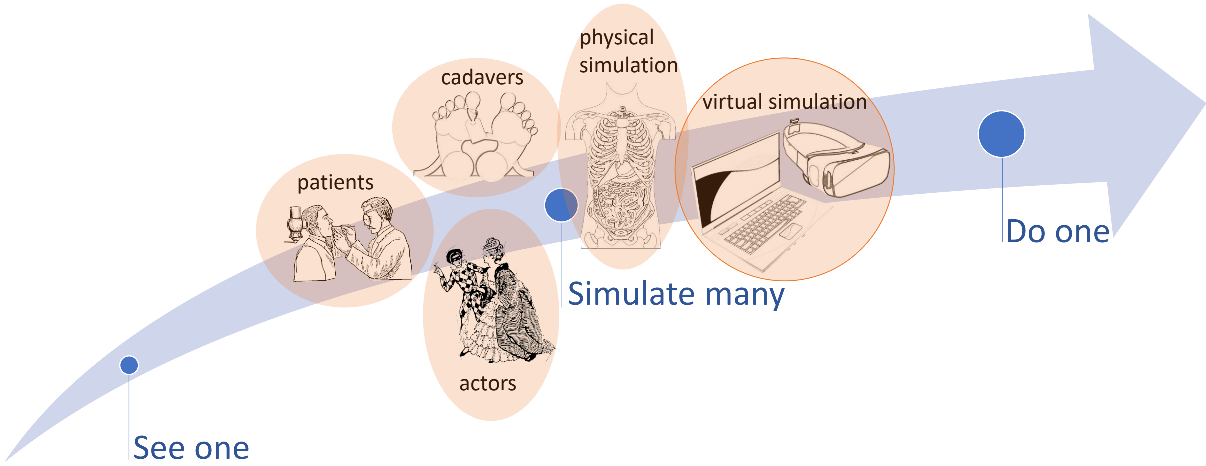


Figure 1.1: Placement of this work in the general medical education. Own figure[1].

1.2 Structure and Contribution

To achieve the aforementioned goals of this thesis it is required to collaborate closely with medical experts. Using this interdisciplinary collaboration, relevant skills to be learned have to be analyzed and considerations how VR can be deployed efficiently to support medical education have to be discussed. As a basis, the *medical background* of this thesis first provides insights into medical training and various modalities that can be used (Section 2.1.1). Because all selected medical use cases that are relevant for subsequent projects are either in the field of liver surgery or address brain diseases, there will be a medical background about the liver (Section 2.1.2) as well as the brain (Section 2.1.3). The medical background is followed by a *technical background* that covers the definition and distinction of VR (Section 2.2.1), VR technologies (Section 2.2.2), and input devices (Section 2.2.3). In *related work* (Chapter 3), a brief overview of VR applications in medicine is given and the concept of gamification is introduced.

The main part of the thesis is divided into three parts: The two selected skills with their exemplary medical area *VR-based Visuospatial Skill Training in Liver Surgery* and *VR-based Strategic Knowledge Training in Neurosurgery*, and an *Excerpt on Input Devices*. Both medical parts start with a related work section (Section 4 and Section 7). The other sections present the following contributions and research questions:

LiVRSono (Section 5) Ultrasound (US) is a common imaging modality, however, it requires visuospatial skills to understand the cross-sections and to properly use it. These skills can only be trained hands-on, which is not possible in the case of IOUS which is, for example, used during liver surgery. Because of this, VR offers an additional and safe training modality. In close collaboration with liver surgeons, different training scenarios were identified and implemented. This requires an US simulation that meets both the medical requirements and the performance requirements of a VR application. To investigate face and content validity, an expert study and a pilot study for learning outcome were conducted.

RQ 1: What are suitable training scenarios of a VR-based application to train visuospatial skills for IOUS?

Gamification (Section 6) For a proper training application, the learning content is fundamental. However, emotions have an influence on attention, memory, motivation and behavior [319]. Thus, positive emotions, such as fun, and increased motivation can positively influence the learning experience and outcome. This behavioral change can be achieved by gamification. Because of this, the previously presented training application LiVRSono was used as basis to investigate the appropriateness of game elements for this use case. Therefore, common game elements and their suitability are discussed, and the effects of selected game elements are investigated. This is done by a broad-audience study in combination with a medical expert study.

RQ 2: Which game elements are appropriate for a VR-based training for visuospatial skills for IOUS?

Microsurgical Interventions (Section 8) In the case of microsurgical interventions, such as treating aneurysms, the access is a crucial part. The access influences all subsequent steps and impairs treatment choices. Because of this, a VR application to train this understanding was developed and evaluated with experts.

RQ 3: To what extent can VR provide an additional training possibility for microsurgical procedures in case of aneurysms?

Input Devices (Section 9) When developing a VR application, the question of which input device is appropriate arises. Using two different but relevant medical VR applications, input devices that vary in the grip style and degree of specialization were compared. With several common user tasks, their suitability was assessed with a large user study.

RQ 4: Which benefits of input devices with varying grip styles should be considered when developing medical VR applications?

The dissertation is completed with an overall conclusion in Section 10, including a summary, limitations and future work of this thesis.

When developing VR-based training applications, several aspects can be investigated. In this work, the focus is not on high-fidelity interactions that include soft tissue deformation or other physical processes. Aspects, such as presence or personal avatars, are also not investigated in this thesis. As described above, the focus is the selection of relevant cases and providing appropriate applications, respectively. Thereby, some projects have the research focus on gamification and its influence.

2

Background

Synopsis The background consists of the medical and technical background. The medical background presents medical education and training modalities and summarizes the required information about the two medical use cases. The technical background explains and delineates the term VR and gives an overview of tracking methods and required hardware.

2.1 Medical Background

The first part describes the current state of medical education and training, including the acquired skills, knowledge and different modalities. This is followed by the medical backgrounds for the liver and the brain.

2.1.1 Training Modalities - How to Train Physicians

During medical education, trainees have to gain different types of knowledge and skills which can be acquired with a variety of training modalities. These competencies are summarized in Figure 2.1. They can be divided into *non-technical* and *technical skills*. Non-technical skills include *social skills*, such as teamwork and communication, *cognitive skills*, such as situation awareness and decision-making, and *personal resource skills*, which comprise stress management and coping with fatigue [84]. The specific technical skills highly depend on the discipline. Besides *theoretical knowledge*, *fine-motor skills* and *visuospatial skills* are included, because they are often mentioned related to surgical competencies [119, 264]. Motor skills are always required when the physician uses medical instruments to interact with the anatomy. During minimally invasive interventions, they cannot directly see the internal structures when interacting with them but have to rely on cameras, which might provide unfamiliar views. With visuospatial skills and anatomical understanding, they are able to create a mental spatial representation of the interior structures [119]. These skills are also required for the understanding of medical image data, which are two-dimensional representations of three-dimensional anatomical structures. Besides practical skills, there is also knowledge that has to be acquired. For this, the four types presented by Jong et al. [141] are used:

- *Declarative knowledge* is static knowledge about facts, theories and concepts. In surgery, this is, for example, the knowledge about anatomical landmarks.
- *Situational knowledge* is knowledge about problem situations. This knowledge enables the solver to filter relevant features out of the problem statement. An example are experiences that are employed when a medical case is similar to previous cases and relevant information can be adapted and reused.
- *Procedural knowledge* contains actions or manipulations helping the problem solver move from one problem state to another. Knowing how to perform different steps and courses of action during intervention is an example of procedural knowledge.
- *Strategic knowledge* is used to organize the problem-solving process. It structures and directs stages the problem solver has to go through to reach the solution. Strategic knowledge is, among others, essential for diagnosis, but also for surgery planning to decide on the best surgical approach.

Common methods for the acquisition of practical skills are the traditional *see one, do one* approach, the use of cadavers and animals, and simulated patients by actors [61, 270]. For the first method, real patients are used as training objects which limits the anatomical variation to the available cases in each hospital. Training on real patients is always a safety risk and ethically questionable. However, this approach has the benefit of

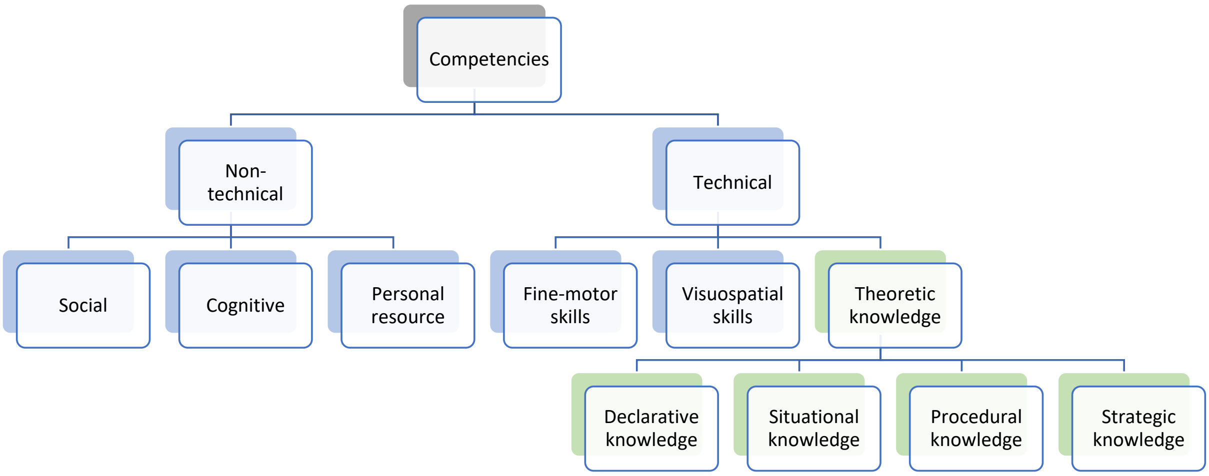


Figure 2.1: Skills and knowledge that are acquired during medical education. Own figure.

completely realistic cases and scenarios, and includes social interaction which should not be neglected, for example, in case of medical examinations. Ethical concerns also arise when using cadavers in regard to treating them in a respectful manner, given the different cultural and religious backgrounds. The haptic is realistic and no patient is harmed, but cadavers are expensive and difficult to maintain [270].

With technical progress, normal lectures are extended by online lectures and videos. Thereby, various learning programs arose. Tutorial systems are used to impart new learning content, whereas training systems do not provide new knowledge but consolidate existing knowledge or train skills by repetition. According to Mönch's [209] overview, simulations are the most complex form of learning programs. By emulating real procedures, knowledge can be tested in practice. To train practical skills in a safe environment, physical as well as virtual simulations were developed. On the one hand, simulations are particularly beneficial for procedures involving children, sensitive tasks or those where the body must be opened or injured, and procedures requiring visual and auditory perception. On the other hand, simulations have limitations when it comes to haptic and sensual perception and biomedical procedures such as tissue behavior or blood flow. With respect to haptic feedback, physical simulations are superior to virtual simulations. They also do not have the drawbacks of uncomfortable and cumbersome VR HMDs as well as cybersickness. However, virtual simulations benefit from visual aspects, such as visualizations, to include additional information or revealing underlying structures. Virtual simulations enable the user to undo and redo single steps and to repeat procedures without harming patients or consuming resources. It is also possible to easily include a wide variety of anatomical variations or surgical anatomy in case of surgical training.

All prototypes that are proposed in this thesis are training simulations. Consequently, they aim at simulating a procedure or provide a training task close to reality, and thus, theoretic preknowledge is required. Once a training simulation is developed, it has to be tested and validated. For this, several aspects can be considered [46, 112]:

- *Face validity* describes whether the simulation feels and looks realistic, which can be assessed by self-reports from experts regarding plausibility.
- *Content validity* measures whether the content is appropriate to reach the learning objective in the real world.

- *Construct validity* is given if the simulation is able to distinguish different performances. This means it should distinguish between novices and experts of the real-world task and improvement should be visible.
- *Fidelity* comprises subtypes, such as physical, physiological or emotional fidelity. It states whether the simulation represents and simulates the physical elements, cognitive and perceptual features, and emotional responses in a realistic way.

Most evaluations of the training simulations presented in the main chapters address face and content validity.

2.1.2 The Liver - Largest Solid Abdominal Organ

Visceral surgery encompasses surgical treatments of all organs of the digestive tract and the abdomen, and therefore, it also includes liver surgery. Liver surgery is a particularly challenging surgical field, as the liver contains many vessels and the biliary structures, and thus, bleedings are very likely to occur. For this reason, liver surgery is a relatively young field [82].

The liver (greek: Hepar) is the largest solid abdominal organ with a weight of around $1.5\,kg$ [208] and a median liver span at the midclavicular line of $14.5\,cm$ for men and $13.4\,cm$ for women [165]. However, the size and shape can vary a lot. The liver comprises a complex vasculature including one arterial system and two venous systems, the portal vein and the hepatic veins [244]. The arterial system supplies oxygenated blood to the liver, whereas the portal veins bring nutrient rich blood from the spleen, pancreas and guts [244]. The hepatic veins drain into the inferior vena cava. Although there are various concepts for subdividing the liver into independent segments, they all agree that there are sections within the liver that have their own vascular and biliary system [80]. Depending on the concept, the size, shape, and number of subdivisions vary. Despite being controversial, the so-called *Couinaud* classification is a widely used and common classification [33, 80]. It defines eight liver segments based on the third order branch of the portal vein [260]. The resulting segments are displayed in Figure 2.2. Because the blood supplies of these segments are independent of each other, they can be used as margins for resection. The *International Hepato-Pancreato-Biliary Association* (IHPBA) used this classification to propose a standardized terminology for the hepatic anatomy and liver resections [300]. However, the segments and their exact borders differ greatly between patients [277].

2.1.2.1 Hepatic Lesions

Diseases that occur in the liver range from vascular diseases (for example shunts and thromboses) to medicamentous-toxic harms, alcoholic diseases, hepatitis, hepatic cirrhosis, to tumors as well as metastases. In this work, only tumors and metastases are relevant. These lesions can either be benign (non-cancerous) or malignant (cancerous). Benign lesions, which are often detected incidentally, can be separated into the groups: regenerative lesions and true neoplastic lesions [29]. Lesions belonging to the first group usually do not increase in volume. Because of this, risk is not increasing and treatment

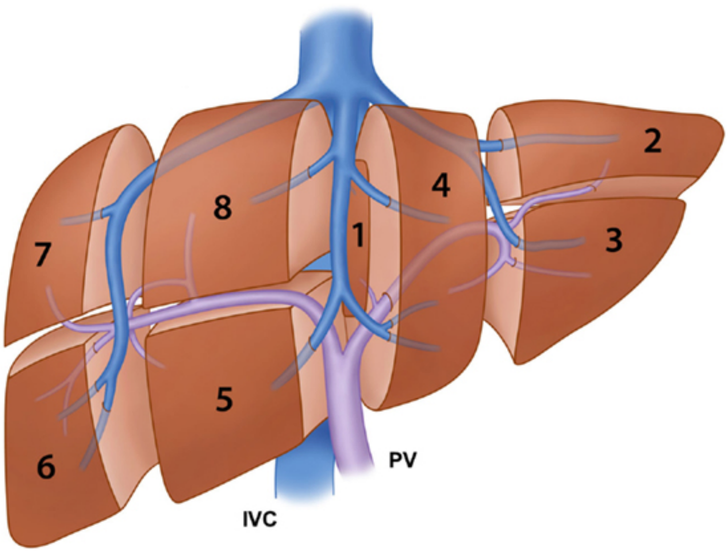


Figure 2.2: Schematic liver segments. PV = Portal vein, IVC = Inferior vena cava. Image from Orcutt et al. [226] and available under a CC BY 4.0 license. No changes were made.

is not required. Lesions of the second group tend to increase their volume, and thus, have a higher risk of complications such as bleeding or malignant transformation [29]. Examples of benign lesions are cysts and focal nodular hyperplasia. Malignant lesions occur either due to a primary tumor, called hepatocellular carcinoma (originating from liver cells) or cholangiocarcinoma (originating from cells from the hepatic bile ducts), or metastases, such as colorectal liver metastases. Primary liver cancer is relatively rare (incidence (cases per 100,000 population per year) in 2018: 3.5 for women and 10.3 for men), however, due to its poor prognosis, it is one of the most frequent causes of cancer death in Germany (relative five-year survival rate: 14 % for women and 18 % for men) [166].

2.1.2.2 Treatment of Neoplasms

As the focus of this work is on neoplasms, such as hepatocellular carcinoma, cholangiocarcinoma and liver metastases, this section concentrates on treatment possibilities for these lesions.

Surgical resection Especially in the case of major resection, the capability of the liver to regenerate up to 80 % of its volume is exploited. Criteria for surgical treatment of liver metastases usually include that there is no disseminated tumor disease, it is a controlled primary tumor and that it is possible to leave tumor-free margins [149]. Regarding hepatocellular carcinoma, resection depends on the coexistence of other liver diseases. Besides these criteria, the patient's physiological condition as well as the disease stage and biology have to be considered. Because of this and the individual anatomy with respect to blood vessels and lesion location, each resection has to be planned carefully using various information including, for example, image data. Part of surgical resection can be IOUS to locate the vessels and lesions, and to find previously undiscovered lesions [163]. With the help of, for example, vascular landmarks, a segment-by-segment examination is performed. If a lesion is identified, the segmental location, proximity to vessels, and maximum size should be recorded [328]. One can either use non-contrast IOUS or contrast-enhanced IOUS. Depending on the literature, non-contrast IOUS leads to 33-42 % additionally identified metastases and contrast-enhanced IOUS to a further 7-20 % [328]. Other literature states that in up to 40 % of patients, additional lesions can be found, which leads to a different surgical strategy for resection in 25-50 % of the cases [62]. This makes IOUS an essential intraoperative imaging modality not just for open surgery but also for laparoscopic surgery. However, Walker et al. [328] summarize limitations that might explain why routine use of IOUS is uncommon. First, it takes time to properly scan the liver and assess lesions. Second, specific US equipment is required and has to be set up. Third, which is stated as the main limitation, is its operator dependence. Around 50 IOUS examinations are considered to be required to gain operator competency and sufficient training is missing [104].

As an alternative to open surgery, laparoscopic surgery is a minimally invasive method. This method can be divided into three approaches [40, 228]:

- Pure laparoscopic liver resection: The whole procedure is performed through ports. Usually around three to six ports are used, with a minimum of two required ports: one for the flexi-tip or 30 °camera, and a working port for instruments, such as an US probe or stapler [312].
- Hand-assisted laparoscopic liver resection: In this case, an elective hand port is used, facilitating the procedure. This method is considered a bridge between the open and pure laparoscopic method.
- Hybrid hepatectomy: After pure or hand-assisted laparoscopic liver mobilization, a mini-laparotomy (surgical opening of the abdomen) is performed to conduct the parenchymal transection [63].

The benefits of laparoscopic surgery are shorter operation time, shorter regeneration and hospitalization duration, reduced postoperative pain, and less blood loss. However, this technique requires a longer learning curve due to the complex hand-eye coordination [184]. At the German Liver Tumor Center Leipzig, which was one of the first teams adapting to laparoscopic liver surgery, 57 % of tumor resections were performed openly and 43 % were performed laparoscopically in the years 2018 and 2019 [302].

Liver transplantation According to the oncological guideline program of the German Cancer Society, liver transplantation is the therapy method with the best long-term results and survival rate [20]. Here, the whole liver is replaced by a donor liver. The main limitation in this case is the availability of donor organs and that only a few patients are eligible for this treatment.

Other treatment methods Alternatively to surgical resection, local ablation or transarterial approaches can be used as interventional methods [20].

Local ablation can be thermally conducted via radiofrequencies or microwaves. The most common way to insert the instrument is via percutaneous access. Using real-time image guiding and visualizations of the surrounding structures, the instrument is inserted. Then, the tumor is destroyed by heat induced either by radiofrequencies or microwaves. Uhlig et al. [320] found that, regarding follow-up treatment, the ablation performed better than surgical resection; however, regarding overall survival, surgical resection was superior in the full cohort.

If neither surgical treatment nor ablation is possible, transarterial approaches, such as transarterial chemoembolization or transarterial radioembolization, are used. These methods are used as palliative therapy or for downstaging, which means the tumor size is reduced such that one of the other therapies is possible [20]. Besides the mentioned surgical and interventional approaches there are systemic treatments, such as chemotherapy and immunotherapy [20].

The choice of the therapy method depends on various factors. Factors include, for example, the tumor type, how advanced the tumor is at diagnosis as well as the condition of the liver and of the patient. Besides characteristics related to the disease and patient, the specialization or competence of the clinical center also plays an important role.

2.1.3 The Brain - Human Computation Center

The second medical field that is considered is neurosurgery. This discipline deals with lesions in the brain, spinal cord as well as spine. In this work, only lesions in the brain are relevant. Between the brain and skull there are three cerebral membranes: the pia mater, arachnoidea, and dura mater (from inside to outside) [123]. The skull is made of six neural bones and eight facial bones [92]. These bones are connected by sutures, which are fibrous joints. The brain surface is characterized by prominent fissures, such as the Sylvian fissure dividing the frontal and parietal lobes from the temporal lobe, and more individual sulci [92]. The blood supply of the brain is provided by the internal carotid and vertebral arteries [195]. Between these arteries, there is a circular formation of arteries called circle of Willis (CoW).

2.1.3.1 Lesions - Intracranial Aneurysms

Similar to the liver, there is a large variety of lesions and diseases that can occur in the brain. These can range from dementia, infections, inflammatory diseases, such as meningitis, and neoplastic diseases, such as tumors, to cerebrovascular diseases. The latter comprises stroke, stenosis, intracranial aneurysm (IA), and vascular malformations.

IAs are pathological dilatations at weakened sections of the arterial wall within the cerebral vasculature [269]. In Central European countries, they have a prevalence, which is the proportion of a population having this characteristic, of 3.2 % [324]. The detection and assessment of IAs is crucial because they have the potential to rupture, resulting in a fatal subarachnoid hemorrhage. Only nine cases per 100.000 persons per year will get such a bleeding due to an IA, however, in around 35 % of cases it has fatal consequences [257]. An IA that is likely to rupture can be treated microsurgically or using endovascular techniques. All methods aim at preventing blood flow between the IA and parent vessel [269]. The treatment methods also have in common that they require several years of experience and specialization.

2.1.3.2 Treatment of Intracranial Aneurysms

Microsurgical Treatment The microsurgical process starts with positioning the head, which is then fixated in a clamp, such as the Mayfield head clamp. There are different surgical approaches; in the following, the pterional approach is described as an example. First, the skin is opened to expose the bone. Using a craniotome, the desired area of the bone is cut. Then, the dura mater has to be opened before the brain can be dissected to reveal the aneurysm.

Endovascular Treatment If an IA is treated endovascularly, a catheter is traditionally inserted using the transfemoral access. Alternatively, the transradial artery in the wrist can be used [332]. This catheter is then navigated into the arteries of the brain and the aneurysm. Several devices can be inserted via the catheter. A common technique is coiling, where a wire is inserted, leading to occlusion. Concerning wide-neck aneurysms, coiling reaches its limits in comparison to clipping [27]. Because of this, currently new techniques are developed. One of the earliest techniques uses a balloon that supports the coil [201] (Figure 2.3a). Alternatively, a stent can be inserted into the parent artery (Figure 2.3b). In contrast to these two devices, newer methods do not aim at an instant but a delayed occlusion [201]. These include flow diverters as an intraluminal device (Figure 2.3c). Blood flow is still possible through the device leading to stasis and thrombosis of the aneurysm. Novel devices are, for example, the Woven EndoBridge device that is placed into the aneurysm sac leading to thrombosis (Figure 2.3d).

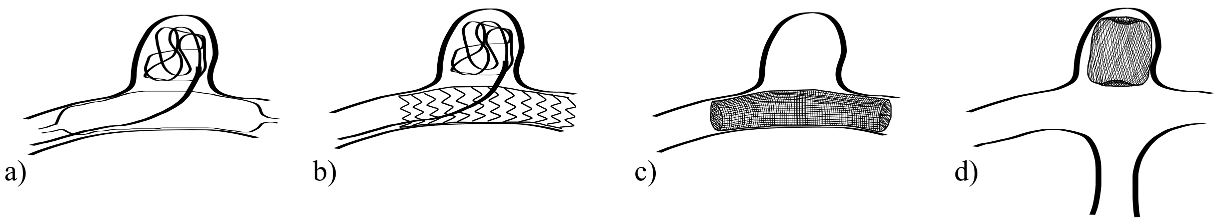


Figure 2.3: Endovascular techniques to treat IAs. a) Balloon-assisted coiling; b) stent-assisted coiling; c) flow diverter; d) WEB device. Own figure.

2.2 Technical Background

In this section, VR is defined and delineated from other technologies such as augmented reality (AR). Afterwards, types of VR and their technical setup are described. In the last section, different types of input devices are presented.

2.2.1 Mixed Reality - Dimensions of Reality

Among virtual simulations, one can differentiate between conventional desktop-based simulations and so-called *XR* simulations. XR is often defined as *eXtended Reality* [121], defining a continuum from a real environment to a completely virtual environment with AR and augmented virtuality in between. Rauschnabel et al. [250] criticize that the terms

AR, *VR*, *mixed reality (MR)* and *extended reality (XR)* are used inconsistently. By performing a literature research, they grouped different definitions of these terms into four views (see Figure 2.4):

- MR-dominant view: MR includes everything between the real world and a fully virtual experience.
- VR-dominant view: Here, AR is a subtype of VR.
- MR-centered view: Similar to the MR-dominant view, the real world and VR are on opposite sites. However, MR is in between AR and augmented virtuality and is not an umbrella term for those two.
- Extended reality view: In this case, XR (extended reality) is the umbrella term including AR and VR and MR is a subform of AR.

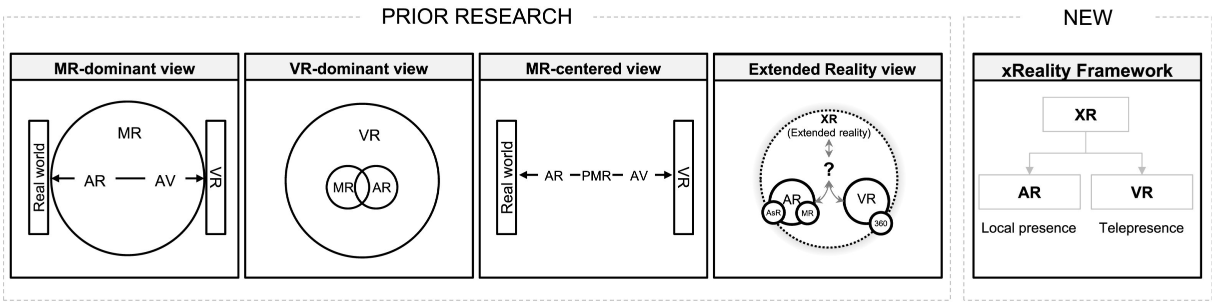


Figure 2.4: Summary of different definitions of new reality formats into four views. AR = augmented reality; VR = virtual reality; MR = mixed reality; PMR = pure mixed reality; AV = augmented virtuality; AsR = assisted reality; 360 = 360 degree content. Image from by Rauschnabel et al. [250] and available under a CC BY 4.0 license. No changes were made.

The MR-dominant view seems to be the most used definition. Rauschnabel et al. [250] criticize that this definition groups VR and AR under the term MR. Looking at the different design goals and user experiences of these two approaches, grouping VR and AR might be problematic. Using XR as an umbrella term for AR and VR is also misleading, because in VR the reality is not extended but replaced, and thus, the definition 'extended reality' excludes VR. Instead of defining an AR/VR-continuum, Rauschnabel et al. [250] propose an *xReality framework* where the 'x' in xReality (XR) is just a placeholder and does not stand for 'extended' (see Figure 2.5). The framework divides XR into AR and VR based on whether the physical environment is part of the experience. Furthermore, AR represents a 'local presence' continuum between assisted reality and MR. It describes the extent to which a user perceives the content as actually present. In this framework, VR encompasses the 'telepresence' continuum ranging from atomistic VR to holistic VR. Telepresence is defined as the extent to which the user feels present in the virtual world. Using this framework, the prototypes proposed in this work are atomistic VR applications focusing on training specific tasks and procedures rather than on completely realistic environments where the user feels completely present.

In the context of VR, the two terms *presence* and *immersion* are fundamental [290]. Slater et al. [289] define presence as the sense of 'being there' in the virtual environment and further divide into *place illusion* and *plausibility*. The first refers to the impression of being in the virtual environment despite the knowledge that one is still in the real world. The second implies that the user has the impression that events are actually happening

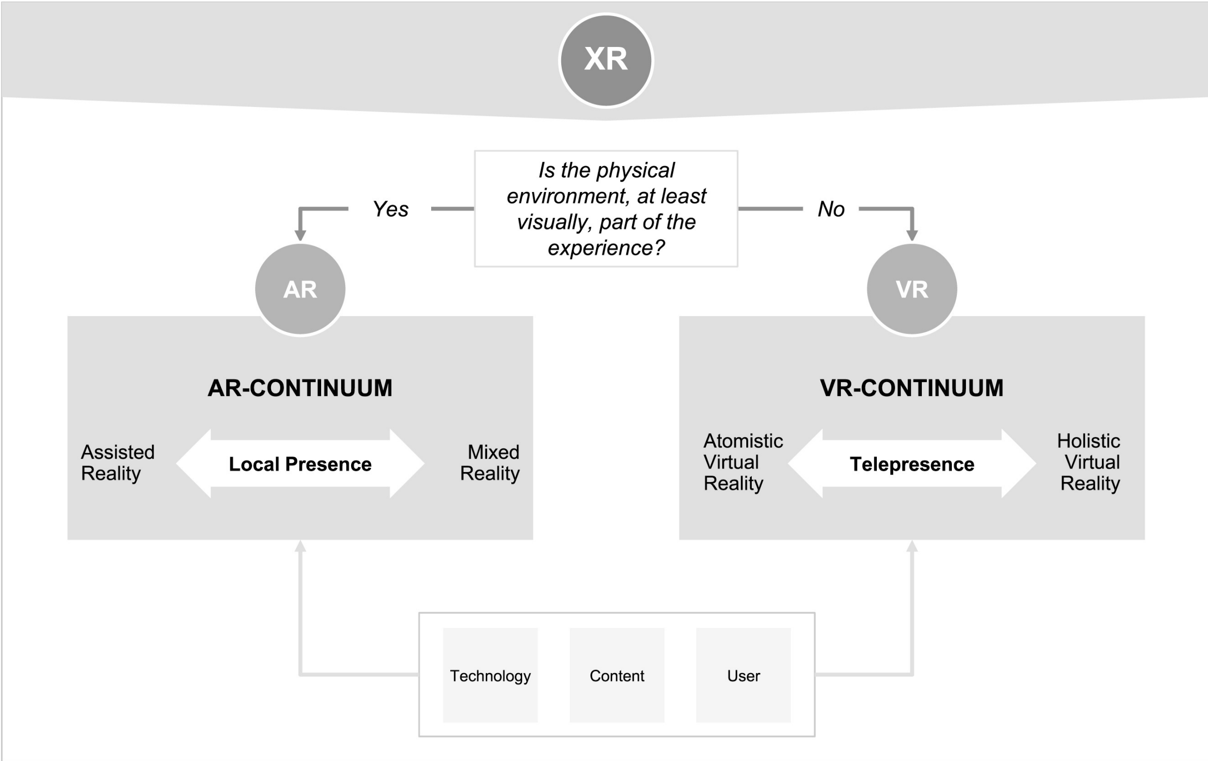


Figure 2.5: New definition of XR. Image from Rauschnabel et al. [250] and is available under a CC BY 4.0 license. No changes were made.

(for example, virtual humans responding to the user) and that all aspects of the virtual environment match reasonably in the specific context.

In contrast to presence which is subjective, immersion is an aspect of VR that is objective. For example, immersion comprises the field of view, system conditions such as interpupillary distance, resolution, and tracking [290].

Besides these two terms, there is also *body ownership*. Depending on the context, it might be relevant to have a virtual representation of the user. Visuomotor and visuotactile synchrony by providing synchronized movements and tactile feedback lead to embodiment: the impression that the virtual body is the own [290].

The model proposed by Skarbez et al. [287] extends the previously described model such that presence is a function of place illusion, plausibility illusion and social presence. Latoschik et al. [174] discussed these two models and came up with an alternative model based on congruence and plausibility. If the sensory input, such as cognition and perception, is congruent, it will result in plausibility. In contrast to the previous models, they order the typical XR-related experiences, such as presence, on the same hierarchical level and all of them are influenced by plausibility.

2.2.2 Virtual Reality - How to See the Non-existing

Despite the various definitions or because of different definitions, VR literature also contains a lot of papers referring to non-VR technology, such as stereoscopic displays, as VR technology [305]. Whenever the terms VR or immersive VR are used in this thesis, it can either be VR using an HMD or a Cave Automatic Virtual Environment (CAVE), but all proposed prototypes employ HMDs.

A CAVE is a room where the user is surrounded by projected images [70]. To provide the correct perspective and stereo vision, the user is captured using tracked stereoscopic glasses. In contrast to this, VR HMDs enable a spatial perception by using one display per eye, showing slightly different images [75]. Using different methods of tracking, the position as well as rotation of the HMD and corresponding input devices in the virtual world are given. In Table 2.1, both modalities are compared and the characteristics of the *HTC Vive Pro Eye*, which was used for most prototypes and evaluations, are listed. There is no guarantee for integrity, but the range of the different specifications is still visible.

Table 2.1: Technical specifications of a CAVE and corresponding stereoscopic glasses, HMDs and the HTC Vive Pro Eye.

Characteristic	CAVE	HMDs[1,2]	*HTC Vive Pro Eye*[3,4]
Weight	< 100 g [65]	500 - 1000 g [65]	800 g
Field of view	Full	horizontal: 89 ° - 160 °(*Varjo VR-2*, *Pimax 5K Plus*); vertical: 85 ° - 117 °(*Varjo Aero*, *Pimax 8K*)	horizontal: 107 °; vertical: 107 °
Price	10.000 € - 100.000 € [65]	400 € - 4.000 € [65]	˜1.400 €
Resolution	highly depend on the projectors	870 × 500 - 3840 × 2160 (per eye) (*Carl Zeiss Cinemizer OLED*, *Pimax Vision 8K+*)	1440 × 1600 (per eye)

To track the position of the HMD, two tracking methods can be used. *Outside-in* tracking requires additional sensors that detect the HMD's position from outside, while with *inside-out* tracking the HMD can track its position due to sensors that are applied directly on the HMD [75]. Consequently, *outside* and *inside* refers to the active sensors that track the position of the HMD and controllers. The HTC Vive Pro Eye, for example, uses the *inside-out* tracking by receiving infrared rays from external lighthouse stations. Because of the external lighthouse stations that serve as markers, this system belongs to the so-called marker-based *inside-out* method. Another way to use *inside-out* tracking is by equipping the HMD with cameras that do not require external markers, such as with the Oculus Rift S, which is then called marker-less *inside-out* tracking [210].

Seeing and being in a virtual environment can also cause adverse reactions, such as headache, nausea or vertigo [75]. Because of a variety of influencing factors, the term *cybersickness* was established. Reasons causing cybersickness involve:

[1] Richard Musil, HMD Geometry Database: https://risa2000.github.io/hmdgdb/ Last access: 09.11.2023

[2] Rory Brown, VR Compare: https://vr-compare.com/vr Last access: 09.11.2023

[3] HTC Corporation, Taiwan: https://www.vive.com/us/product/vive-pro-eye/specs/ Last access: 09.11.2023

[4] Rory Brown, VR Compare: https://vr-compare.com/headset/htcviveproeye Last access: 09.11.2023

- the used displays leading to eye strain and headache and
- contradictory sensory information (sensory conflict theory) when the user is virtually moved without a physical movement or the other way round (motion sickness).

However, symptoms similar to those caused by motion sickness can also occur when the user is not moving (virtually and physically). Although the reasons are not completely investigated, the following factors that influence cybersickness were identified: the individual person (age, sex, VR experience,...), the VR system (contrast, flicker, frame rate, tracking problems,...) and the user's degree of movement control [75].

2.2.3 Input Devices - How to Interact with the Virtual World

For interaction with virtual objects, one can either directly or indirectly interact with them. The former requires hand or finger tracking, which can be done by cameras such as the *Leap Motion Controller*[5]. Another possibility is to use gloves where each finger is tracked individually. Gloves using an exoskeleton, such as the *TESLAGLOVE*[6], include force feedback that restricts the fingers from closing when colliding with an object. Using tracking, not just hand gestures can be used as input but also the whole body can be tracked. Other natural interactions, such as speech or foot gestures, will not be covered in this thesis.

The most common interaction device for VR is the VR controller which is usually enclosed with every HMD. Although they can vary in shape, they have the following main buttons in common: a trigger button (index finger), a grip button (palm), a joystick or trackpad (thumb), a primary button and a secondary button. They also have in common that they can provide haptic feedback by vibrating and that they are held in power grip.

When talking about grips, they can be divided into two types based on functionality: *power grip* and *precision grip*. Using the power grip, an object can be held firmly and forces can be applied, whereas the precision grip supports accurate actions [342]. Both types can be further divided, which can be seen in Figure 2.6. An example of an input device held in precision grip is the *VR Ink*[7].

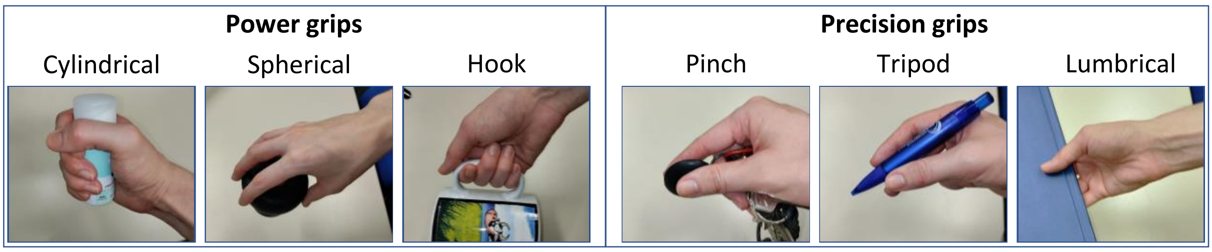


Figure 2.6: Grip types. Own figure based on Yang et al. [342].

[5]Ultraleap, United States: https://www.ultraleap.com Last access: 09.11.2023
[6]Teslasuit, VR Electronics Ltd, United Kingdom: https://teslasuit.io Last access: 09.11.2023
[7]Logitech, Switzerland: https://www.logitech.com/de-de/promo/vr-ink.html Last access: 25.03.2024

Finally, one can use various objects in combination with a tracking system, such as the lighthouse system and Vive tracker, to interact with the virtual world by providing the position and rotation of the tracked object. As already mentioned, some devices like the VR controller provide haptic feedback by vibrating. However, there are also grounded haptic devices that provide force feedback when colliding with a surface. Using different physical properties, various materials can be simulated. Most of these haptic devices have six degrees of freedom (DOFs), three positional and three rotational; only a few devices have six DOFs force feedback [317]. Another important characteristic of these devices is their workspace. The *Geomagic Touch*[8] has a relatively small workspace with $160\,mm \times 120\,mm \times 70\,mm$, but there are also devices with a workspace of up to $1330\,mm \times 1020\,mm \times 575\,mm$ such as the *Virtuose 6D* [317]. When choosing a haptic device, one should consider the resolution. For example, devices used for microsurgical procedures require movement tracking with sub-millimeter precision. For many other applications this might not be the case.

Furthermore, there are devices specialized for a single interaction task, such as navigation, where, for example, treadmills can be used [18].

Technical Setup and Software Used in the Projects All prototypes presented in this thesis were developed with the game engine Unity[9] Version 2021.3.5f1. For the development and evaluation of the first prototype of the craniotomy training, an *Oculus Quest* was used. For all other prototypes and the user studies, the employed HMD was the *HTC Vive Pro Eye*.

[8] 3D Systems, United States: https://de.3dsystems.com/haptics-devices/touch Last access: 25.03.2024

[9] Unity Technologies, San Francisco, USA: https://unity.com/de Last access: 12.04.2024

3

Related Work on Medical VR Applications and Gamification

Synopsis This section serves to provide an overview of related work in the two areas *VR in medicine* and *gamification*. Both areas do not include a complete literature research but should give an impression of the wide variety of applications.

3.1 VR - Immersion in Medicine

This section serves to give an overview of how VR is used in medicine. Medical VR applications can either have medical experts, such as nurses or surgeons, or patients as users. Areas where VR is used by patients are diagnosis, patient education, rehabilitation and therapy, and pain management. Applications for experts are for training and planning [137]. Keywords used for publications indicate that *rehabilitation* and *simulation* are the two main focus areas [345].

3.1.1 Patients as Target Group

Starting with diagnosis, VR can be used to diagnose or assess diseases. The natural display of 3D objects and navigation in VR can be exploited to assess spatial memory [28] or entorhinal cortex functionality [125]. Thus, by providing a more natural test environment, the diagnosis of various neurological disorders, such as Alzheimer's disease should be facilitated. Additional to providing a realistic test environment, VR also provides a safe environment, which is important in case of, for example, road-crossing for assessing unilateral spatial neglect [326].

VR's benefit of intuitively displaying complex 3D structures is also used for patient education. VR videos as well as interactive 3D objects or scenes are used to reduce anxiety and to increase patient understanding and satisfaction. Two reviews [281, 321] show that using 3D models, sometimes with additional image data, in combination with explanations of the corresponding physician is the most common method.

In contrast to this, rehabilitation makes use of the immersive environment and presence to elicit realistic behavior as well as emotional responses [313]. A broad field within rehabilitation is limb rehabilitation, which is relevant for patients with various disorders such as multiple sclerosis [334] and stroke [42]. For cognitive rehabilitation, VR applications often focus on mild cognitive impairment, traumatic brain injury, Alzheimer's disease and stroke, but there are also applications concerning, for example, depression and attention deficit hyperactivity disorder [118]. The benefit of creating a realistic scenario is also used for assessing and treating various specific phobias, such as fear of spiders, agoraphobia, which is anxiety in situations (for example public transport) that are perceived as unsafe, and social phobia [87]. One common way of treatment is exposure. VR exposure treatment offers a safe and cheap possibility to experience specific situations, such as preparing a meal or crossing a street, with difficulty adapted to the patient and under the guidance of the therapist [78]. Lütt et al. [188] emphasized the benefit of offering a less costly and complex setup in the case of cue exposure therapy to assess alcohol craving in persons with alcohol use disorder [188].

The last field where VR is used by patients is during therapy, however, not for therapeutic reasons but for pain management. In this case, being immersed and present in a virtual world is used as distraction from reality, or more precisely from pain, which can occur during wound dressing changes, interventions, labor, or which can be chronic [60].

3.1.2 Medical Experts as Target Group

For medical experts, VR is mainly used for planning and education or training. Sometimes the differentiation between planning and training is not that clear. For example if a planning application includes the execution of the procedure, one could also say that one can train for this specific case. However, usually the intention whether it is for planning or training affects the choice of interactions, visualizations and the workflow of the application. Furthermore, the included medical data varies. For training, selected cases that are representative are used, whereas for planning, the patient-specific case has to be processed. For both training and planning, there is a wide range of medical disciplines where VR is used [41, 170].

3.1.2.1 Planning

Additionally to the 2D image data, VR-based planning applications provide interactive 3D models of the anatomy. The benefit of having an additional 3D model is that this visualization of anatomy corresponds to the view of anatomy during surgery, whereas the 2D images require cognitive resources to translate these into a 3D mental model [337, 344]. In contrast to viewing a 3D model using a monoscopic display, VR offers a stereoscopic view with natural interactions. This helps in treatment decision-making as well as planning the chosen treatment such as access planning.

Besides showing a 3D representation and the image data, planning applications might include various interactions or have specific characteristics to facilitate planning. Exemplary interactions and functionalities are summarized in the following:

- Collaboration to allow remote or co-located planning [56, 285]
- Drawing cutting lines or distances [56, 157, 253, 285]
- Adding (screws or clips) or removing (resection volume) [56, 157, 161, 253, 299]
- Visualizations for risk estimation [16, 56, 161]
- Annotations and highlighting [16, 263]

Although there are no clear results showing that planning with VR leads to better clinical outcome, surgeons rated it favorably [170].

3.1.2.2 Training

Besides motivational aspects due to immersion, VR-based training benefits from simulating the real situation, thus, creating a more authentic training scenario. Especially for cases where no or limited training possibilities are available because of safety aspects or accessibility, VR-based training should be considered. It benefits from trial and error learning, exploration, direct feedback, and additional visualizations as well as information that can provide more insights or serve as motivational aspects.

In the following, exemplary applications are presented to get an impression of the different skills and knowledge types that can be trained and learned as well as the various medical areas where VR-based training is used.

The difference between non-technical skills and technical skills is sometimes not that obvious. For example situation awareness belongs to cognitive, and thus, non-technical skills, however, there is also situational knowledge. Situation awareness as a non-technical skill means understanding and perceiving a situation, whereas situational knowledge is used to react to the situation using technical knowledge and experience. Sometimes it is not possible to separate them, such as in communication [284] and teamwork [54] training, which also requires situation-specific technical knowledge. Because of this and as the focus of this thesis is on medical-related and not on personal-related skills, training applications for non-technical skills are not comprised in the following.

Starting with the technical skills, *motor skills* are included whenever the user has to use a medical instrument, for example, for cutting [155] or more advanced interactions such as with laparoscopic devices [55]. *Visuospatial skills* are involved whenever the user has to work with medical image data. This includes US-based navigation or examination [286, 304], or interpreting other imaging modalities such as X-ray [178].

Declarative knowledge, for instance anatomical facts, is very fundamental in medical education and often learned with 2D sketches of complex 3D anatomical structures. To facilitate the learning of 3D correlations and spatial arrangements, VR is frequently used. Thereby, static anatomical parts, such as arrangements or variations [274, 278, 238], or dynamic processes, such as the development of anatomy [273], can be visualized. *Situative* and *strategic knowledge* are often addressed in applications where the user has to make several decisions, such as in the case of first responder training [39, 185, 234], and is also enlarged with every real and every virtual patient one has to treat. An example where *procedural knowledge* is trained is Kockwelp et al.'s [159] VR application for brain death determination where several steps have to be passed through to make a final decision. Other examples are first-aid training or clinical observations [31].

Despite the amount of immersive VR training applications, there are only few studies evaluating the training outcome. These can either include subjective feedback regarding increased competency, confidence as well as perceived increased knowledge [19, 243, 268], or objective measurements with respect to performance or knowledge [69, 154, 182, 243, 128]. According to Mao et al's. [191] literature review, common measures to compare VR with a control group were time to completion, procedure-specific checklist or knowledge tests, confidence scales, performance scales, and efficiency. The results of the review indicate that VR can improve surgical training. However, one has to consider that positive results from pre-post studies that use the simulator as test environment, arose due to practice effect. This aspect will also be part of later discussions. Huber et al.'s [128] comparison of an immersive VR training with a non-immersive laparoscopy training revealed that although participants using the VR training needed more time and had more errors, they were exhilarated by the high level of immersion and appreciated the increasing attractiveness of the simulation. These aspects might lead to a more frequent use which is essential for training.

Furthermore, training and learning systems require a balanced level of difficulty, not just for evaluation purposes, but also to prevent frustration and to enable improvement. In the case of evaluations, this is referred to as *floor* and *ceiling* effect.

Merely offering a training system does not guarantee a good learning outcome. In addition to delivering appropriate learning content and employing effective teaching methods, it is essential to take into account various other factors. Gomez et al. [98] highlighted that perceived motivation and enjoyment significantly correlate with perceived learning. Other influencing factors are repetition and practice. Kang et al. [147] studied repetitive learning and concluded that repetitions spaced out over time positively influence memory, problem solving skills, and the transfer of learning. While extrinsic motivational factors, such as mandatory courses and grades, can ensure repetitions, intrinsic factors like enjoyment and self-motivation tend to be more impactful. One approach to increase these intrinsic factors is by gamification, which will be presented in the following section.

3.2 Gamification - Is it Just for Fun

The basic idea of gamification is to motivate people, steer their behavior and increase fun. Thus, gamification can be used in various digital but also non-digital areas. Especially in learning applications, these aspects are desired, because the attitude and behavior positively influence the learning outcome. Consequently, gamification might have positive effects on the learning outcome [265]. In the last years, several definitions of *gamification* arose. One popular definition is:

'Gamification' is the use of game design elements in non-game contexts.

(Deterding et al., 2011, p.10)

For game elements, there are also many lists and classifications. Common game elements are included in Reeves and Read's [251] *Ten Ingredients of Great Games* and summarized in Figure 3.1.

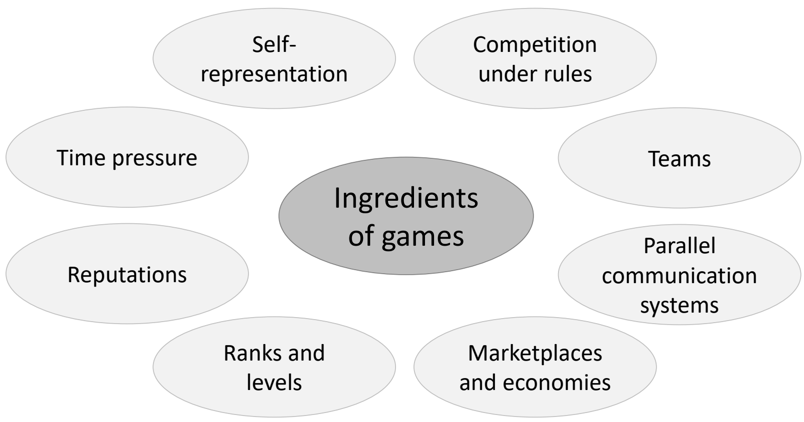


Figure 3.1: Game ingredients. Own figure based on Reeves and Read's [251].

Deterding et al. [73] define game elements as 'elements that are characteristic to games' and thus found in most games. But they can also be classified due to various aspects. For example, Chou [58] presents the Octalysis Framework, which consists of eight core drives, such as epic meaning, to which specific game elements can contribute.

In several survey papers [127, 203], two observations are demonstrated. First, some studies reveal positive impacts of gamification [172, 230], while others indicate no significant effects or even negative consequences [91, 314]. Second, particularly in educational contexts, common practices involve the frequent use of points, leaderboards, and badges. Due to this trend, it is often called 'pointification' rather than gamification. According to Westera [336], these reward mechanisms predominantly lead to extrinsic motivation instead of intrinsic motivation.

Another term that often appears in the context of gamification is *serious games*. Although both concepts integrate game elements, there is a clear distinction according to Bedwelll et al.'s [26] game attribute taxonomy. Serious games encompass all game attributes, while gamification is the application of individual game attributes or a limited selection of several attributes [171]. In the context of learning, both serious games and gamification share the common goal of enhancing learning outcome. Serious games directly influence learning by providing instructional content and incorporating game characteristics, whereas gamification impacts learning by influencing the learner's behavior or attitude [171]. Because serious games cannot be separated into 'the application itself' and 'the gamification aspects', studies evaluate the serious game as a whole and compare the game to other training and learning modalities. In contrast, a gamified system necessitates the evaluation of the individual game elements and their combinations [160].

Instead of presenting one gamification approach, it is also possible to incorporate tailored gamification. Gabele [89] points out the different frameworks and approaches that all use some kind of user/player categorization and include suitable game elements based on the different playing preferences. This aims at a more personalized experience and higher performance.

To limit the vast amount of applications, first, the focus is on gamified VR-based applications and those, where game elements are described and which include a study comparing the game element or gamified system to a control group, are presented in more detail. Because of the limited amount of approaches using VR in medical education and training, the subsequent section also includes non-immersive virtual systems that are gamified.

In education, most gamification research is conducted in social sciences and engineering/computing [127], however, gamified VR-based systems can range from mathematics over assembly training to breastfeeding.

3.2.1 Gamified VR-based Educational Systems

Affine transformations, an important mathematical concept, can be taught using a gamified application proposed by Oberdörfer et al. [223], where transformations have to be used to escape a room. Levels, points, and achievements are included. Their study compared the VR version to a desktop version with the result that VR is in favor of the desktop version.

The assembly training of a drum set presented by Palmas et al. [230] includes a progress bar, points, sound feedback, and a particle effect. The number and severity of errors were positively affected by gamification for novice VR users. However, the time spent in the training was in favor of the control group. Similarly, there are VR puzzles for anatomy education [99, 238]. However, they did not include further game elements or compare it to conventional education methods.

The serious game proposed by Hartfill et al. [114] uses the game dynamics from *Beat Saber* to learn vocabulary. Compared to the conventional flashcard method, the VR game has significantly lower scores in terms of recognition and recall rate. However, due to perceived fun and higher motivation, it may be useful in the long run.

Tang et al. [306] compared a non-gamified breastfeeding VR-based training with a gamified version comprising scores, feedback effects, timer, and badges. Using a within-subject study, they assessed player experience and whether gamification reduces the opportunity for reflection on breastfeeding. The gamified version scored significantly higher with respect to immersion, ease of control, clarity of goals, challenge, progress feedback, and enjoyment. The quantitative data could not reveal a reduction of reflection on breastfeeding due to gamification. However, qualitative feedback indicated that participants focused more on their achievements and scores than on the feeding process and their behavior. This is also supported by a significantly higher score of the gamified version concerning the experience of reflection on performance. Gamification leads to significantly higher temporal demand and effort and some participants stated that this version is less serious.

A medical education system to learn dental morphology is presented by Quispe et al. [247]. Using narrative and a landscape including non-player characters, medical students should learn the dental morphology, answer a questionnaire, and afterwards assemble a set of teeth.

Tashiro et al. [307] propose a VR-based training system for novice neurosurgeons for microscopic suturing. Speed, accuracy and carefulness are used to include feedback and scores. Besides feedback regarding the score and time, a colored circle provides feedback for the tip position, and a bar indicates the motion of the entire gauze. Performing a within-subject design study with ten participants, each participant had to go through the experiment twice, once with feedback including the two visualizations, scores and time, and once without. Measures included the *user experience questionnaire* and *system usability scale*. Both conditions, with additional feedback and without, provide effective training without a significant difference between them. The subjective results show that the users perceived the score as motivating.

Süncksen et al. [303] gamified a VR-based and desktop-based training for X-ray imaging and handling C-arm computed tomography (CT). In this case, gamification includes levels of difficulty and points incorporated in a non-medical and a medical setting. For evaluating the training, they used the *user experience questionnaire* and additional questions regarding the user interface (UI), enjoyment and usefulness for medical education. The only question referring to the gamification was whether they enjoyed the game part. This was answered with yes by eight of nine non-medical participants.

3.2.2 Gamified Medical Systems

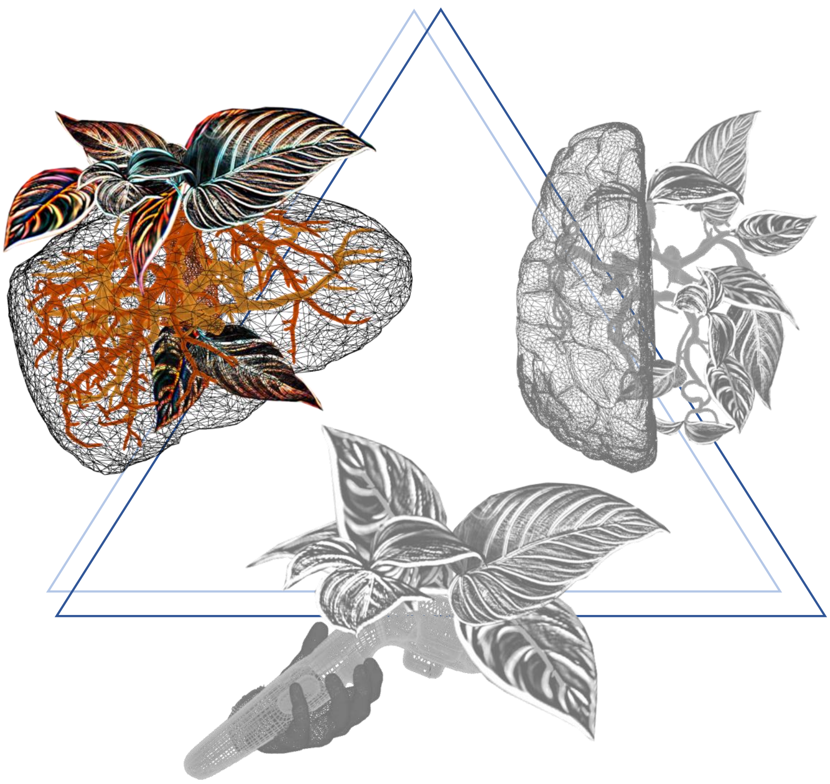
Since the amount of gamified systems for medical education using VR is very limited, this section includes desktop-based systems, too. In medical education, gamified approaches predominantly use question-based game-like techniques, including quizzes and question banks [101, 183, 267, 327]. These are primarily incorporated in undergraduate lectures. Here, pointification is again a widespread trend. Evaluations typically assess the entire system rather than individual game elements and do not compare the system with a non-gamified control group.

In addition to the described gamified systems, serious games, for example, for phlebotomy training [88], chest tube insertion [117], anatomy education [238], or laparoscopy [133] are investigated and show an increase in performance. Some studies directly compared the serious game with conventional lectures and revealed an increased learning outcome [190]. However, these are complete games and no individual game elements are described and investigated.

In the field of gamification for patients, the border between gamification and serious games is blurred and most approaches entitle their work as serious games. These systems can address many different aspects, such as physiological training [38, 50, 94, 206, 116, 237, 272, 343], psychological and neurological disorders [21, 329, 347], or developmental disorders, such as autism spectrum disorder [216, 311], as well as cognitive rehabilitation and therapy [49, 90, 214].

Part I

VR-based Visuospatial Skill Training in Liver Surgery



4

Related Work on VR Applications in Liver Surgery

Synopsis Because the first part of the dissertation focuses on training in liver surgery, this chapter provides relevant literature. Therefore, it is subdivided into planning and training systems. Although liver surgery planning is not the focus of this thesis, an overview is given because there are much more planning than training systems. As the chosen use case is in the field of US, training applications for US in various medical areas are presented, too.

4.1 Liver Intervention Planning

In liver surgery, most virtual systems focus on planning rather than training. Because of this, the first section gives an overview of virtual planning systems. Although there are also desktop-based systems, the literature search was restricted to VR systems.

Most of these systems focus on liver resection, thus planning the resection plane that separates the remaining parts from the parts to be removed. Reitinger et al. [254] started to combine virtual resection with VR, however, it is not immersive. They use a large active-stereo back-projection wall together with shutterglasses. For an appropriate planning, they included manual branch labeling, liver segment approximation, safety margin calculation, and spatial analysis tools. Based on this, an anatomical resection is possible by selecting individual segments. When choosing an atypical resection, meaning that the resection is not along the segment boundaries but along the lesion, three shapes for resection can be selected: a plane, a sphere and a deformable plane. To assess the resected liver, distance, volume and angular measurements are provided.

Kenngott et al. [152] evaluated a VR planning system with 158 participants using the Oculus Rift HMD. This system provides a 3D model of the anatomical structures, imaging data as well as patient data in a minimal, futuristic environment. Their results show that most participants agreed that it is possible to assess complex cases faster (85%) and more comprehensively (94%). Although they entitle it ‘planning system’, they asked about the training potential. Above 80% of medical students and residents see potential for medical student training. However, this prototype does not include performing a virtual resection and its evaluation.

Another immersive VR planning system is proposed by Chheang et al. [56]. In their collaborative system, the user is situated in a planning room, where they can explore the CT imaging data and the 3D representation of the liver. Similar to Konrad-Verse et al.’s [162] desktop-based application, a virtual resection is initialized by placing a resection plane which can then be deformed individually. The deformation can either be done on the single CT layers or directly by manipulating–bulging by pushing and pulling–the plane at the 3D model. An additional color scheme highlights the safety margin to the tumor. Medical experts emphasized the potential of the system to visualize and assess anatomical relations and to identify safety-critical areas of surgically complex cases.

In a newer prototype, Chheang et al. [53] compared interacting with the plane using a Bezier surface interaction with the previously mentioned direct deformation. In contrast to this, Reinschluessel et al. [253] included a drawing function to indicate cuts, a volume tool to mark the volume that should be removed using four predefined shapes, and a clip function to clip blood vessels. All medical experts stated that still having the 3D model in mind while performing the surgery increases confidence and reduces stress. Using the 3D model for planning facilitates the understanding of distances and spatial relations. Furthermore, anatomical variants can be seen very quickly.

Positive results and comments of experts during evaluations of the presented planning systems emphasize the benefits of visualizing complex 3D anatomical structures in VR.

4.2 VR-based Training and Education in Liver Surgery

Besides the planning systems, there are also immersive VR training systems. Here, most approaches focus on laparoscopic interventions. By presenting a collaborative environment, Chheang et al. [54] concentrated on the social skills teamwork and communication. Using two training scenarios during laparoscopic surgery, anesthesiologists and surgeons can practice communicating when one of them recognizes a problem that the other person has to react to.

In laparoscopic surgery, there are many training applications for various interventions and studies comparing the VR-based training using a laparoscopic input device [55, 128, 86, 239]. These training applications aim at improving motor skills, because laparoscopy requires spatial orientation and hand-eye coordination due to operating with the elongated instruments [322]. Another example for training motor skills is Kanzira et al.'s [148] non-immersive needle intervention application. During needle interventions, such as liver biopsy, a needle has to be inserted precisely. However, this is challenging due to different tissues and movements. Chheang et al. [57] combine their previous works in a large virtual hospital where users can collaboratively train the whole procedure ranging from planning the surgery to the actual intervention. Thus, besides motor skills, mental skills and communication can be trained.

Instead of training, there are also approaches for liver anatomy education. Displaying 3D anatomy in a VR environment, especially a collaborative system, can be used for various types of knowledge. Depending on the instructor or provided information, declarative knowledge, such as anatomical structures, or strategic knowledge, such as therapy planning, can be addressed. One example is Schott et al.'s [274] multi-user application for liver anatomy. They focused on decision-making by including case-specific information regarding the treatment.

This section shows that most training applications related to the liver are within laparoscopic or needle interventions, thus, mostly focusing on motor skills.

4.3 Virtual Ultrasound Training

As the application proposed in Chapter 5 and used in Chapter 6 focuses on IOUS, this section presents virtual–desktop-based and VR-based–US training applications in various medical areas.

Other areas, where virtual US training is advantageous are, for example, obstetrics and gynecology [232]. Studies presented by Madsen et al. [189] and Al-Memar et al. [202] have confirmed that training with the Scantrainer[1] led to improved performance in transvaginal US examinations. This desktop-based simulation uses two monitors and an US probe. The

[1] Intelligent Ultrasound (trading name of MedaPhor Ltd.), UK: https://www.intelligentultrasound.com/ Last access: 15.04.2024

US image is shown on one monitor and a virtual patient including the probe position on the abdomen is displayed on the other monitor.

Another medical use case that requires careful training is US-based needle intervention. The focus of Mastmeyer et al.'s work [194] is on a realistic simulation of US imaging, including breath and deformation while inserting the needle. Ni et al. [218] chose the use case of US-guided biopsy in their desktop-based training. With the two haptic devices *Phantom Omni* and *Phantom Premium* they simulate the US probe and a needle. Similarly, Barnouin et al. [23] use the haptic device *Geomagic Touch* for their needle insertion simulation. They aim at simulating realistic US imaging by using textures and raytracing as well as displacement functions for tissue deformations.

Orr et al. [227] analyzed the influence of transabdominal US simulators integrated in the curriculum. To assess performance, they consider various general skills, including equipment usage, hygiene and ergonomics, along with specific US techniques. In the context of liver US, the evaluated skills cover, amongst others, scanning the entire volume, capturing images in different planes to maximize the visualization, and identifying the portal vein with color Doppler US. The study results show that, compared to the control group, the simulation leads to a significantly higher confidence with regard to diagnostic images of the liver and the identification of pathologies.

In contrast to the presented desktop-based approaches, there are also some immersive VR simulations. By using a phantom with a VR environment, Bublak et al. [37] presented a training that combines cardiopulmonary resuscitation and US imaging. Thereby, physically separated users have their own phantoms and are joined in a virtual environment with the same virtual patient. Their US simulation incorporates the hardware from *Schallware*[2]. With this training simulation communication skills are addressed.

Johnson et al. [140] presented a semi-immersive VR application with the focus on task analysis and appropriate performance metrics for US-guided liver biopsy. With the performance metrics, including targeting, probe usage time and mean needle length in beam, they were able to show significant differences between different levels of expertise of the participants.

Another use case is US guidance for vascular access, which was chosen by Shenoy et al. [280]. To interact with the virtual US probe and needle, the user employs normal VR controllers. Furthermore, they included gamification by adding levels and a score based on metrics, such as puncturing the target structure. In the study, ten medical students had to rate their confidence before and after training with the system. The self-reported confidence increased by the training. Other VR-based approaches include, for example, a competency test for contrast-enhanced US [136] or specific user tasks, such as identifying the position of a previously shown lesion [304]. However, the latter example does not include an evaluation of the system, thus, the validity of the system cannot be assessed.

In the context of US-guided needle intervention for anesthesia, Simon et al. [286] addressed the lack of training simulations focusing on hand-eye coordination. To interact with the virtual US probe and needle, they used two *Geomagic Touch* devices. During the evaluation, 18 anesthesiologists with varying experience in US-guided needle insertion had to perform the tasks, including exploring the interior, locating the nerve, inserting

[2]Schallware GmbH, Germany: https://www.schallware.de/ Last access: 12.04.2024

the needle, injecting the anesthetic, and following its spread. Questions for face and content validity were inspired by previous works and are not validated. The moderate answers for face validity revealed that the main drawback is the realism of the haptic feedback of the needle, whereas the realism of the environment and the realism of the vein deformation achieved the highest scores. The content validation shows that the simulation is not sufficient to train this procedure, however, it is a promising training tool for hand-eye coordination. Although the simulation has a medical context, the complexity of the anatomy is not quite clear. Furthermore, their training simulates the medical workflow, however, it lacks specific tasks for users, resulting in a lack of feedback and guidance.

Chuan et al. [59] also chose the use case of US-guided anesthesia. For the US imaging, they recorded data while scanning a phantom including various positions and rotations. These data also included needle insertion. The recorded data was then incorporated into a VR-based training simulating an operating room (OR). Using VR controllers for the US probe as well as needle, the user has to choose a proper insertion point. In a study with 38 participants they were able to differentiate between experts and novice participants.

In summary, the majority of existing research regarding US training is in US-guided interventions and gynecology. Thereby, the main focus is either on the US simulation itself or on the assessment of performance and learning outcome when integrating the simulation into the curriculum. The latter assesses general handling skills based on tasks that are deduced from clinical routine, like scanning the whole anatomy or identifying important structures. However, only a limited number of US simulations directly target visuospatial skills.

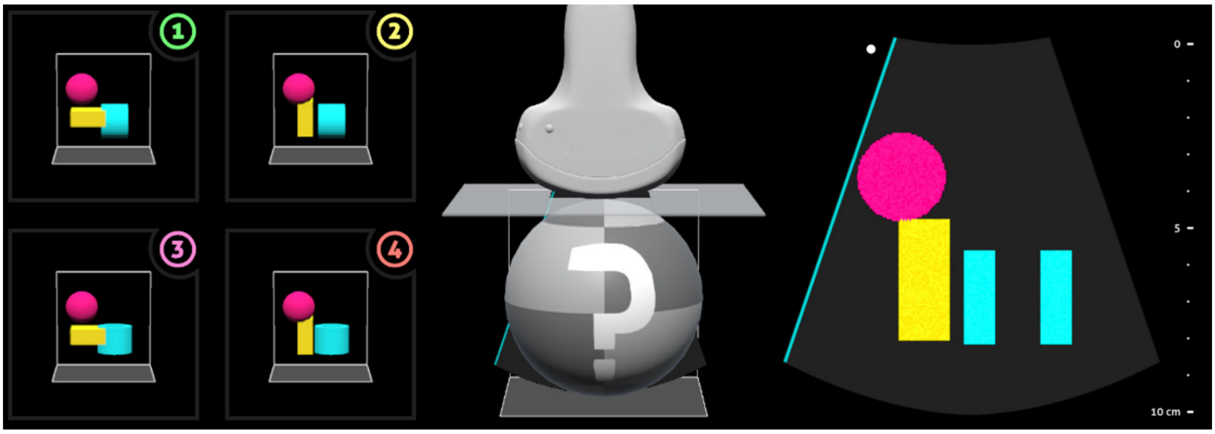
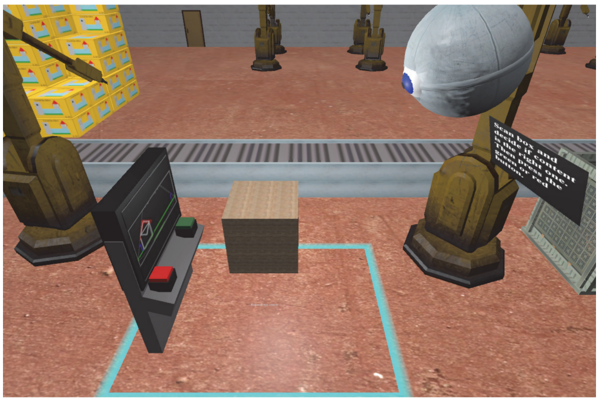


Figure 4.1: Desktop-based US training with geometric scenes and mouse input. Image from Mayer et al. [196] © 2021 IEEE.

Figure 4.2: US training simulations.

To provide a simulation that includes training- and learning-related aspects, Law et al. [175] proposed an US simulator with a didactic system. Their desktop application provides haptic feedback by using the *Phantom Omni*. In addition to the US simulation, 3D anatomical models, and an annotation system for didactics are integrated. The didactic system adds labels for all anatomical structures, which are visible in the current US image.

Mayer et al. [196] presented a desktop-based US game that trains visual-spatial relations. In this case, there is no medical context, instead they use a geometric scene containing few objects, such as spheres, cuboids and tubes (see Figure 4.2). The training comprises four different minigames with tasks concerning the understanding of the US image and probe handling. By testing the visuospatial ability before and after training, they investigated the learning outcome. Although no significant differences between a control group and a game group could be found, which might be explained due to a ceiling effect, the game group improved more on average than the control group.



(a) VR-based US training in a toy factory. Image from Byl et al. [44] © 2018 IEEE.

(b) Desktop-based US training in an underwater world. Image from Olgers et al. [225] and available under a CC BY 4.0 license. No changes were made.

Figure 4.3: US training simulations.

A similar approach was introduced by Byl et al. [44], however, their game is VR-based. Instead of a medical setting, they use a toy factory as an environment (see Figure 4.3a). Using cross-sectional images, the user has to identify mispacked boxes in a specific time span. The evaluation focused on general usability and player experience.

Olgers et al. [225] proposed a desktop-based game where the player is situated in an underwater world. Using this context, the player has to collect hidden coins (see Figure 4.3b). With this task, they want to train probe manipulation to receive the desired US image. 18 experts revealed that the probe manipulation with the 3D-printed probe is realistic and the game is fun to play. Performance differences between experts and novices could not be found. The simulations that directly address visual-spatial skills are summarized in Table 4.1.

Table 4.1: Comparison of approaches focusing on training the understanding of US images and the corresponding 3D structures.

Approach	Modality	US simulation	Context	Input device
Law [175]	Desktop	Raytracing	Medical	Haptic device
Mayer [196]	Desktop	Contour+noise+shadows	Geometries	Mouse, keyboard
Byl [44]	iVR	Contour+noise	Toys	VR controller
Olgers [225]	Desktop	-	Underwater	3D printed probe

In addition to the mentioned approaches, there is also the commercial VR-based training simulation *VitaSim*[3]. It can be used for US for catheter placement [17], trauma assess-

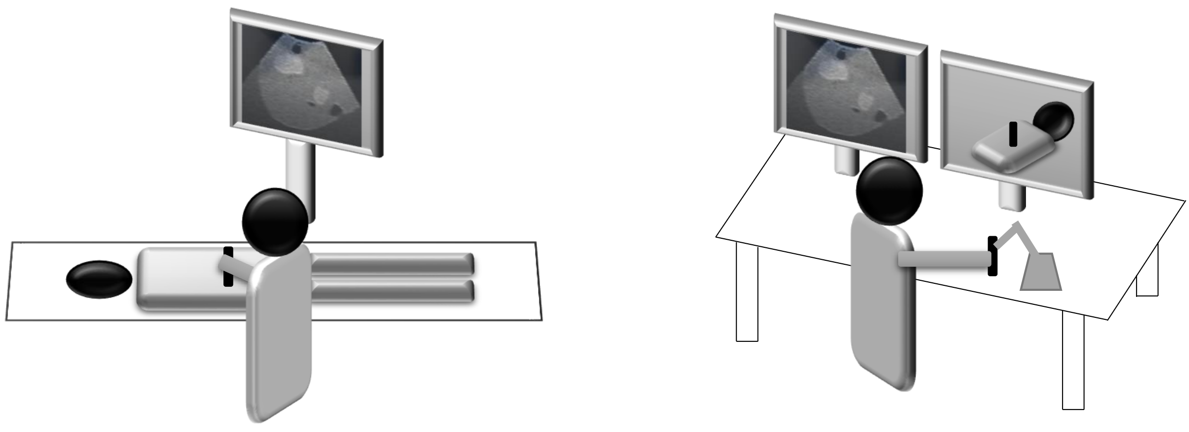
[3]VitaSim, Denmark: https://www.vitasim.dk/ Last access: 12.04.2024

ment [144], or the lung [173]. For the US imaging recordings are used. With a VR controller, the virtual US probe can be placed at predefined locations where it can be tilted.

Besides US, the necessity to train visuospatial skills is also mentioned in the context of other medical image modalities. For X-ray imaging, Lenz et al. [178] developed a serious VR game. Various minigames are proposed where the user, for example, has to recognize or count objects in a sealed box. An overall score is calculated using the amount of required X-ray images, required time, and the amount of false answers. Furthermore, the three modes story, challenge and single game as well as five difficulty levels were included. They vary the difficulty, the amount of answers, the amount of objects in the boxes, and the X-ray representation (color or grayscale).

The visuospatial challenge in case of inspecting angiographies is addressed by Allgaier et al. [8]. They present a gamified VR-based training where an angiography image is given and the user has to identify the corresponding location on a 3D model of the CoW. Included game elements are, for example, a score, time, health points, and achievements. In contrast to this, Ropinski et al. [259] addressed this difficulty by adding visualizations to the angiography image to support the perception. This is done by, for example, color schemes that encode depth information or highlight edges. Another approach is a virtual mirror that displays 3D data in a 2D angiography [331]. These last two techniques directly modify the image to support the perception, however, they are not integrated into clinical routine. Consequently, training for physicians to handle this challenge is necessary.

This literature research shows that there are US applications, however, they often do not include tasks addressing specific skills but simulate the general workflow without additional information or feedback. Those applications that focus on training these skills tend to lack medical context. Furthermore, many approaches are desktop-based, which does not reflect the real situation regarding hand-eye coordination (illustrated in Figure 4.4). In desktop-based systems, there is a deviation between where the user moves the US probe and the virtual patient and the moving probe on the patient.



(a) Setting in reality and VR applications.

(b) Setting of desktop-based applications.

Figure 4.4: Setting regarding hand-eye coordination in (a) (virtual) reality and (b) desktop-based systems. Own figure.

5

LiVRSono - Virtual Reality Training with Haptics for Intraoperative Ultrasound

Synopsis This chapter starts with highlighting the motivation and contribution of the proposed US training application. After presenting a brief physical background regarding US and related work, requirements are set up. The development includes an US simulation as well as training scenarios. Finally, the evaluation and results are presented and discussed.

This chapter is partially reused with permission of IEEE and has been published under a CC BY 4.0 license in:

Mareen Allgaier, Florentine Huettl, Laura Isabel Hanke, Hauke Lang, Tobias Huber, Bernhard Preim, Sylvia Saalfeld, and Christian Hansen. "LiVRSono - Virtual Reality Training with Haptics for Intraoperative Ultrasound", *IEEE ISMAR*, Sydney, Australia, pp. 980-989, 2023, doi: 10.1109/ISMAR59233.2023.00114.

Florentine Huettl, Mareen Allgaier, Sylvia Saalfeld, Laura Isabel Hanke, Bernhard Preim, Hauke Lang, Christian Hansen and Tobias Huber. "Liver IOUS 2.0: Training and Gaming in virtueller Realität", *Z Gastroenterol*, 61(08): e602, 2023, doi: 10.1055/s-0043-1772139.

5.1 Contribution

IOUS provides real-time information without ionizing radiation. It is used to locate and characterize lesions, evaluate the vasculature, assess the planned surgical margin, and guide operative procedures [186]. The US probe is directly placed on the organ surface so that no overlying structures influence the image. Surgeries are carefully planned preoperatively; however, the intraoperative situation differs due to significant deformations of the organ [120]. IOUS is also superior to preoperative data in terms of detecting small lesions.

The most common limitation of US is its operator dependence [104, 328]. The main challenge during the procedure is to build a mental model of the organ based on the US image and the position of the probe on the organ [104, 219], which is illustrated in Figure 5.1. Thus, the surgeon needs two skills. First, they need a good hand-eye coordination. Second, they have to understand where on the US screen an anatomic structure is represented and link it to its respective location in the organ using a spatial mental model [104]. This hand-eye coordination and visuospatial skill can only be trained hands-on and the lack of proper training and education is often mentioned as a limitation of IOUS [104, 328, 333]. IOUS is used for example during surgery in the liver, kidney, pancreas, and during brain surgery [186, 207].

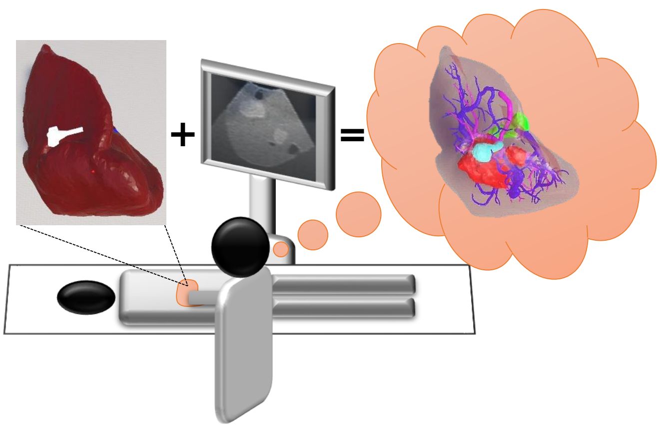


Figure 5.1: Illustration of the challenge to create a spatial mental model based on the US probe position and US image. Image from Allgaier et al. [9] and available under a CC BY 4.0 license. No changes were made.

US training systems in general can either be physical, for example, using 3D printed models [229], or virtual [227]. A physical phantom has the benefit of haptics and the US image can either be obtained by using an US-capable model and a real US probe, or a simulated US image using patient data, such as CT data. Because printing a 3D liver is expensive, it would be unfeasible to print many patient-specific livers. If only one non-patient-specific liver is used, there is either no variation or the US image does not fit to the physical model, leading to confusion when moving the probe on the liver surface. These restrictions do not exist for virtual systems, and haptics can be included by a haptic device. As described and illustrated before, an immersive VR environment can simulate the situation and the hand-eye coordination in a more realistic way. Furthermore, immersive VR and its realistic setting can improve the learning experience [129, 181, 191, 243] and can evoke realistic physical responses, as well as behavioral changes due to different environmental conditions [115].

To advance hand-eye coordination and manual skills, haptics play a crucial role. As the physician moves the US probe directly on the organ and is looking at the US image, they depend on the haptic feedback. This is particularly important because the surgeon has to apply light pressure to avoid air between the probe and the organ. Thereby, the hand position should be similar to the real hand position [13]. Consequently, the haptic device *Geomagic Touch* was employed.

In the following, *LiVRSono* (**Li**ver **V**irtual **R**eality **Sono**graphy), an immersive VR training system for IOUS for the liver, is proposed. IOUS is the gold standard for navigation during liver resection [142]. The creation of a patient-specific mental model of the liver anatomy is especially difficult due to the complex vasculature and interpersonal variations [301].
The main goal of training should be to increase performance and to enable transfer to the real situation. Because of this, the following research question emerged:

RQ 1: What are suitable training scenarios of a VR-based application to train visuospatial skills that are required for IOUS?

Before the training outcome can be evaluated, it is important to examine whether the training application itself emulates the real setting or at least the most important aspects that are relevant for the learning goal. Accordingly, face and content validity are investigated [46].

To provide a plausible and appropriate IOUS simulation for this use case, the following contributions can be summarized:

- The learning goal *anatomical and 3D understanding and orientation* and workflow were analyzed in detail with liver surgeons. The focus is on visuospatial skills and hand-eye coordination but not directly on probe manipulation such as learning an appropriate sequence of movements.
- Four training scenarios were defined and implemented based on the intraoperative workflow and mental task.
- *LiVRSono* was evaluated with physicians with varying expertise regarding IOUS. The main part of the evaluation is a study focusing on the realism and the meaningfulness of the training scenarios. An additional pilot study serves to get first impressions of the learning outcome.

5.2 Design

Before presenting related work on simulations of US images and summarizing US training applications, a brief overview of the physical background of US is given.

5.2.1 Physics of Ultrasound

Without going into mathematical and physical detail, this section provides the required background for US. US imaging is based on US waves propagating through tissues.

Thereby, the higher the frequency, the better the image resolution but the smaller the penetration depth [246]. Besides the properties of the US waves, tissues have different acoustic impedances. This describes the resistance of the tissue to the propagation of waves [246]. While propagating through tissue, several phenomena occur:

- Absorption: Wave energy is converted into heat energy [246].
- Reflection: Part of the US signal is reflected at interfaces and captured, resulting in an image. This occurs at interfaces of two tissues with different impedance values [1].
- Scattering/Dispersion: Because of inhomogeneities of tissues, some waves are dispersed [246].

These phenomena lead to an attenuation of the US waves, however, absorption is the main cause [246]. Other characteristics of US images are, for example, speckle artifacts or speckle noise leading to the granular appearance of US images [72]. This noise is caused by an interference of scattered echoes and has a multiplicative behavior that strongly correlated with non-Gaussian statistics [76]. It is a common problem in US images. Other artifacts that can occur in US images include, for example, shadows behind solid structures and mirror artifacts at tissues with a high acoustic impedance [1].

5.2.2 Related Work

First, methods to simulate US images are presented, and second, important aspects of already presented US training systems are highlighted.

5.2.2.1 Ultrasound Simulation

To simulate US, there are *interpolative approaches* and *generative approaches*. Interpolative approaches use prerecorded 3D US volumes that are resliced. In these approaches, differences between the actual US probe position and the position from where the image was acquired lead to incorrect direction-dependent artifacts, such as shadows [71, 97]. In contrast, direction-independent features, such as tissue texture, are very realistic. Another disadvantage is the restriction by data acquisition.

Generative approaches are based on other image modalities such as CT or magnetic resonance imaging (MRI) data or on mesh models. The methods vary a lot regarding accuracy and costs. Accurate and most realistic methods solve wave equations such as the Green's functions or Westervelt equation [150]. These methods require several hours to simulate the image. Less accurate but faster methods first create a slice and then simulate the US using, for example, texture synthesis with radial blur [194, 252, 348], convolution [93] or ray-tracing [176, 318]. There are also approaches combining, for example, convolution and ray-tracing [43, 266] aiming to exploit the advantages of both methods. Starkov et al. [298] combined the interpolative approach with ray-tracing for a transvaginal US. Thereby, the generated US image of the target structure is fused with a background US volume acquired in vivo. Another approach uses ray-tracing with deep learning. Vitale et al. [323] use generative adversarial neural networks (GANs) that synthesize a new image based on an input image to simulate abdominal US. As input, they use CT data and a voxel-wise segmentation of the organs. Based on these, a ray-tracing approach is used to get a synthetic US image. A more realistic US image is then retrieved using a CycleGAN. A GAN is also used by Chen et al. [51] to simulate IOUS of the liver using preoperative MRI data. For such a framework, they had 4130 pairs of MRI and IOUS images.

Methods using convolution or texture synthesis are the most performant ones. Ray-tracing approaches sometimes are also real-time capable, however, a frame rate of about $15 - 30fps$ [43, 298] might be sufficient for desktop applications but not for VR applications. For training a GAN, a large training dataset is required, which might not be available.

The choice of an appropriate approach strongly depends on application-specific requirements. For example, studies for probe design, the training of the diagnosis of non-obvious lesions, or the understanding and perception of minor differences in US images require realistic simulations. The focus of the presented project is on learning an overall orientation, anatomical understanding and interpretation of the US image. Therefore, the decision was made to use a fast texture-based approach suitable for real-time interaction in an immersive virtual environment. However, by continuously including expert feedback in the development process, it was ensured that the US simulation is plausible enough to enable a proper training.

5.2.2.2 Virtual Ultrasound Training

Only few approaches for US training include important aspects for training such as feedback, annotations, or motivational factors [175, 196, 225] (refer to Chapter 2). However, those are either not included in a medical context [44, 225] or are not VR-based for a proper hand-eye coordination [175, 196, 225].

Mayer et al. [196] include the following four training tasks to address visual-spatial skills:

- Identify the correct scene with 3D models using US.
- Identify the correct US image without seeing the scene.
- Identify the correct US probe position and rotation that creates the given US image.
- Identify how the probe has to be moved to create a given US image.

With these tasks, they want to improve the understanding of the correlation of the US probe and the resulting image, the transfer of 2D and 3D as well as the understanding of the probe movements and the resulting changes in the US image. As *LiVRSono* has a similar learning goal, the training scenarios partially derived from their tasks and were adapted to the specific medical use case.

Similar to these approaches, this project aims at targeting the challenge of creating the spatial mental correlation between the US image and the 3D model. In contrast to Byl et al.'s [44] immersive VR training and Mayer et al.'s [196] and Olgers et al.'s [225] desktop-based learning games, the proposed system and training scenarios emulate the real situation including probe handling and, thus, hand-eye coordination as well as visuospatial skills in an immersive surgical setting. Hereby, the focus is on the clinically relevant application IOUS in liver surgery.

5.2.3 Requirements

The requirements for *LiVRSono* were established based on intensive discussions with the clinical development team with different levels of experience in IOUS. Thereby, different implementations of haptic feedback and important aspects of the US image were considered. For a better understanding, a real IOUS procedure was observed. The following requirements can be classified into two categories: plausibility and training.

Plausibility This refers to the US simulation, the handling of the probe, and the setting. Harris et al. [112] emphasized that 'A simulation aims to capture key features of the real task and environment, rather than exactly emulate or imitate it' (p.3). Thus, it is important to identify these key features and to simulate them in a plausible way. In the proposed training, the US image should exhibit the main physical phenomena and should have a similar appearance, but aspects such as the exact gray values are less important for spatial understanding. Furthermore, the handling of the US probe should be similar to the real one. This includes haptics to perceive the curvature of the surface as well as a real hand position. Finally, the setting should be similar to a real surgery, meaning that the user should stand next to the patient's abdomen with the US monitor on the other side of the patient. Thus, the following requirements arose:

- R1 The US simulation should be plausible, including a proper image section, the most relevant physical phenomena, and the main functionalities such as depth regulation. Using the above presented US background and other US simulations, attenuation, noise, reflections and blur to simulate low resolution should be included.
- R2 The user should be able to move the US probe in a realistic manner and there should be haptic feedback when touching the liver surface.
- R3 The whole setting in the VR environment should be similar to a real setting.

These requirements are difficult to measure. The simulated US could be compared with real images, however, this would require real images and simulated images of the same patient and with the exact same probe position as well as orientation. Because of this, the requirements are considered to be met if experts evaluate them positively.

Training This requires appropriate training scenarios and the assessment of the user's performance. Consequently, there is an additional requirement:

- R4 The training should include training scenarios that vary regarding their anatomical focus or mental transfer, but they should all require the building of a mental model and spatial orientation.

5.3 Development

This section is divided into the two main parts of the training application. First, it is explained how the US images are simulated using 3D anatomical models. Second, the training environment and the training scenarios addressing the underlying learning objective are described.

5.3.1 Ultrasound Simulation

In consultation with medical experts, the decision was made to use rigid models to reduce complexity. Otherwise, realistic deformation and interaction with the deformable liver would be necessary. Furthermore, the US simulation also has to automatically adapt to the current shape of the model leading to a much higher computational effort. This would be more relevant for a realistic handling of the liver and surgical procedure than for the learning goal of the presented prototype.

The developed real-time US simulation uses 3D surface models that are also displayed in the virtual environment. The 3D models are segmentations provided by the medical expert team. The following describes how US images are generated based on these models and the whole procedure is summarized in Figure 5.2.

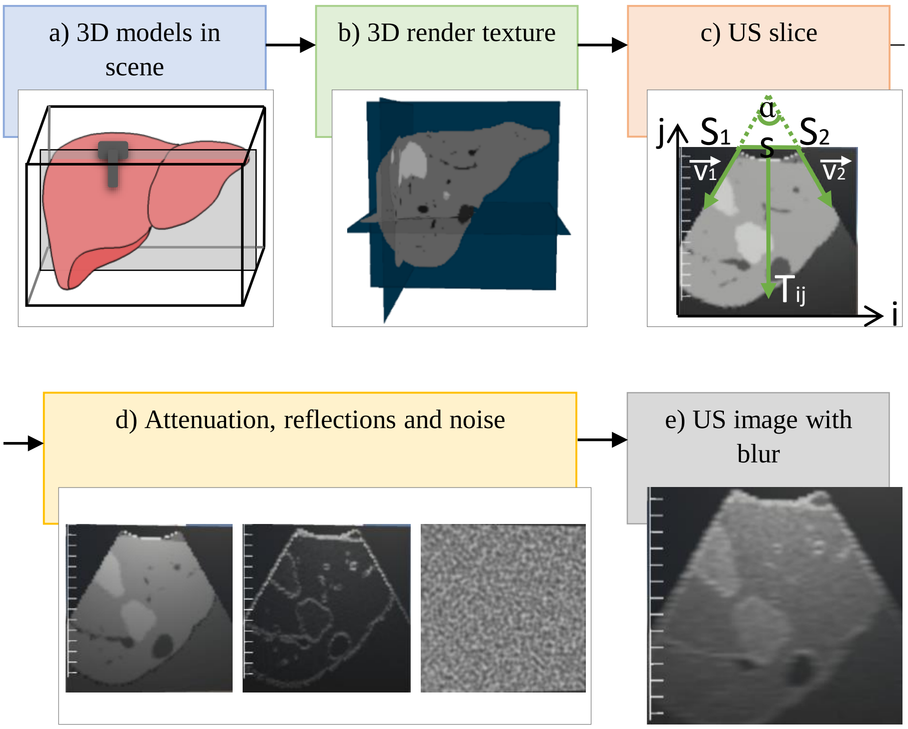


Figure 5.2: Simulation of real-time US based on 3D models. Image from Allgaier et al. [9] and available under a CC BY 4.0 license. The layout and colors were changed slightly.

5.3.1.1 Generation of 3D Render Texture

To simulate US during runtime, a 3D render texture is required (refer to Unity's render texture[1]). This texture is created using the anatomical 3D models. The texture, which has a resolution of $128 \times 128 \times 128$, is aligned with the bounding box of the liver and each voxel is assigned a corresponding tissue type. This requires duplicating the models and inverting the normals to detect collisions from the inside of the models. By iterating through the render texture, ray casting is used to identify in which tissue the current voxel lies. The first object it hits from the inside is the object or tissue it belongs to. To identify which tissue was hit, layers with the tissue names have to be assigned to the anatomical models manually (refer to Unity's layer system[2]). The gray values of the tissues are approximated using real US images as a template leading to the allocation of RGB values shown in Table 5.1, and resulting in an US image shown in Figure 5.2b. This preprocessing step has to be done once for each patient case that is included.

Table 5.1: Settings of the US image.

Characteristics	Value
Tissues	RGB values
Liver	(0.47, 0.47, 0.47)
Inferior vena cava	(0.1, 0.1, 0.1)
Portal vein	(0.1, 0.1, 0.1)
Vein	(0.1, 0.1, 0.1)
Gallbladder	(0.1, 0.1, 0.1)
Lesion	(0.77, 0.77, 0.77)
Parameters	
Probe size (s)	$5cm$
Probe angle (α)	60°
Penetration depth (p)	$[1, 20]cm$

5.3.1.2 Slice shader

The render texture is used to simulate the US image via shaders. First, a slice representing the current US image has to be created. This is done based on the current US probe position and orientation. The position is mapped to the bounding box and is scaled to the range of $[0, 1]$. Then, the field of view is adapted to match a common US image. A texel T_{ij} lies within the field of view if it meets the following conditions:

$$(-v1_i * (T_j - S1_j) + v1_j * (T_i - S1_i)) < 0 \tag{5.1}$$

$$(-v2_i * (T_j - S2_j) + v2_j * (T_i - S2_i)) > 0 \tag{5.2}$$

[1]Unity render texture: https://docs.unity3d.com/Manual/class-RenderTexture.html Last access: 07.03.2024

[2]Unity layer system: https://docs.unity3d.com/ScriptReference/LayerMask.html Last access: 07.03.2024

If both conditions are fulfilled, the texel T_{ij} lies between the two vectors $v1$ and $v2$ defining the trapezoid together with the two upper corners $S1$ and $S2$. The trapezoid depends on the probe size s and the angle α (refer to Figure 5.2c and Table 5.1).

5.3.1.3 Attenuation

If the conditions 5.1 and 5.2 are met, attenuation is included based on the exponential function:

$$I_\alpha^k = I_i^k e^{-\beta df} \tag{5.3}$$

where I_i^k is the intensity of the incoming beam and I_α^k of the output beam. β is the absorption coefficient of the material, d the distance traveled in the material and f the frequency of the wave. Because no material-specific parameters are used, the attenuation is simulated using a general absorption coefficient, the length to the current texel ST_{ij} and the penetration depth p, which is inversely proportional to the frequency, to simulate the attenuation. The resulting slice is shown in Figure 5.2d.

5.3.1.4 Reflections

Reflections occur at boundaries of tissues of different acoustic impedances. The larger the difference, the more reflection occurs. The reflection also depends on the angle of incidence. A wave hitting the interface perpendicularly results in the highest reflection. If the angle is smaller, the wave is deflected away from the probe. Because tissue information is neglected, the reflection shader calculates the absolute difference between the current texel and the texel above. The texel above T_{above} is not $T_{i,j+1}$ but the texel lying on the vector between the current texel T_{ij} and the US probe S_{ij}: $T_{above} = T_{ij} + (S_{ij} - T_{ij}) * TexelSize.j$. The output (see Figure 5.2d) is then added to the image.

5.3.1.5 Noise shader

In the next step, noise is added to the US image. US images exhibit the above-presented speckle noise. Because of the inhomogeneity of tissues, waves are scattered, leading to interferences that create the speckle pattern. Consequently, different soft tissues and diseases cause different speckle patterns [95]. As tissue information is not considered, a general noise was added. Inspired by Mastmeier et al. [194], Perlin noise [235] was used[3]. Using the same origin and axes as for the slice shader, a single noise slice is created and the corresponding texels of the main texture and noise texture are multiplied (see Figure 5.2d).

[3] K. Takahashi, Perlin Noise for Unity (2015): https://github.com/keijiro/PerlinNoise

5.3.1.6 Blur shader

The lower the frequency, which is $5MHz$ in the underlying medical case [167], the larger the penetration depth and the lower the resolution of the image. This is also the case with increasing distance to the probe. Furthermore, there is radial blur in US images. Because radial distortion, which occurs the more convex a probe is, was neglected, a 1D horizontal blur is used to simulate the mentioned effects. Therefore, the average of the surrounding pixels along the x-axis is calculated. The strength of the blur effect scales with the texel distance to the probe (as radial blur is more visible the further away the pixel is from the center), and the current penetration depth (to simulate the resulting smoother edges due to lower image resolution) was added.

5.3.1.7 Further Considerations

The last part of creating a realistic US is to zoom the image so that only the part within the current depth is visible. The scale next to the US also has to be adapted to show the corresponding size of one centimeter.

Shadows are also a typical component of US images. However, as they appear behind bones or air, and thus are not crucial in intraoperative liver US, they were neglected after consultation with liver surgeons. The final US image can be seen in Figure 5.2e. With the described slice shader, attenuation including depth regulation, and the other phenomena, *R1* is addressed.

5.3.1.8 Ultrasound Probe

Figure 5.3: *Geomagic Touch* Device with a 3D printed attachment, mimicking the US probe. Image from Allgaier et al. [9] and available under a CC BY 4.0 license. No changes were made.

To address *R2*, the user can move an US probe using the *Geomagic Touch*, on the liver of a patient approximated by a torso [153]. The *Geomagic Touch* has six DOFs of movement, three DOFs for force feedback, and a workspace of approximately $160W \times 120H \times 70D\,mm$; thus, a restricted workspace compared to a real US probe. As the input device is pen-like and the US probe is not held like a pen, an attachment to enable a more realistic hand position was designed and 3D printed (see Figure 5.3). This approximation was modeled based on images and product information of IOUS probes. A corresponding US probe representation was included into the virtual environment. Virtual hands were not included because only the position of the US probe is relevant and studies revealed that visualizing hands is not necessary when performing motor tasks in VR [255].

For the haptic feedback the *Haptics Direct for Unity V1* from *3D Systems* is used and a haptic material with a stiffness of 0.073 is assigned to the liver surface. This value was determined with the clinical development team. The device has a nominal position resolution of $> 450\,dpi$ and a refresh rate of $1\,kHz$.

5.3.2 Training Environment

In *LiVRSono*, the user is situated in a virtual OR adapted from Huber et al. [129]. To fulfill *R3*, an US monitor was modeled and placed according to a real scenario. Direct interactions with the liver, such as slightly lifting it, and feeling different health states would require very realistic deformations and haptics and is not addressed by the proposed training system.

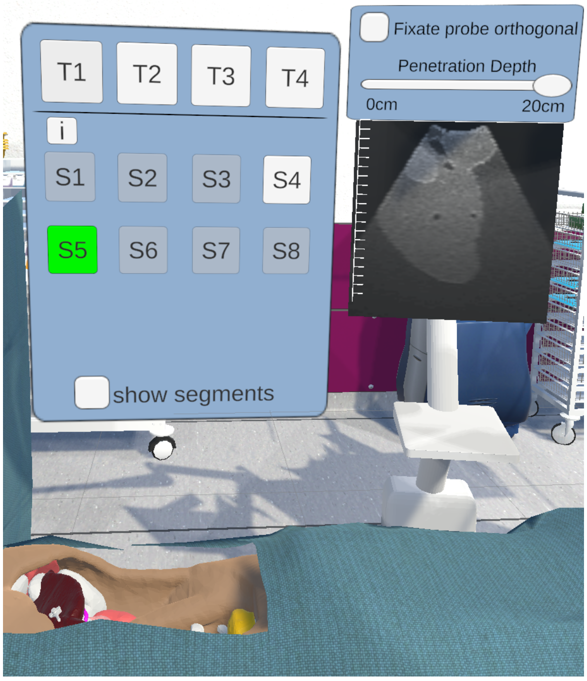
The learning tasks address the following aspects:

- Building a spatial mental 3D model
- Understanding and interpreting (with respect to anatomical understanding, not diagnosis) the US
- Orientating within the liver using US
- Hand-eye coordination.

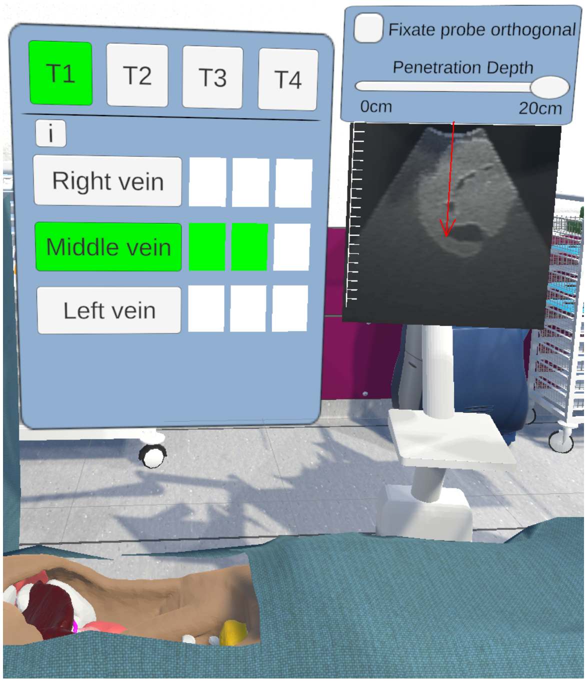
Based on the learning goal and the tasks proposed by Mayer et al. [196], four training scenarios were identified. The scenarios and their suitability for liver surgeons were discussed with medical experts of varying levels of experience. *Scenario 1* and *scenario 2* are based on the IOUS workflow which consists of the three steps (1) identification of hepatic veins, (2) identification of portal veins and their branches, and (3) the systematic scanning of the whole liver parenchyma [3, 104], whereas *scenario 3* and *scenario 4* focus on the spatial mental model. The order of the scenarios was chosen based on their level of difficulty after consultation with the experts. Consequently, the first scenario refers to the third step of the workflow and the second scenario refers to the first two steps.

Scenario 1 This scenario focuses on the systematic scanning of the whole liver parenchyma. In order to train procedural skills and hand-eye coordination, the user has to scan and identify all segments (see Figure 5.4a). This is done by placing the probe on the liver surface and clicking the correct button for the specific segment where the probe lies on. The segments were manually defined in an additional application where the user can mark the eight segments.

Scenario 2 Finding and following important anatomical structures trains the anatomical understanding and deduction of 3D anatomical structures from 2D US images, as well as the orientation within the liver. Although for *scenario 1*, one also has to find the vessels for identifying the segments, *scenario 2* was rated as more difficult because of tracing the vessels. Therefore, it was placed as the second scenario and not as the first scenario which would be in accordance with the IOUS workflow. To select a vessel, the user has to start at the hepatic vein star and click the corresponding button of the



(a) Scenario 1 - identification of liver segments. The user has to select the segment on which the probe currently is.



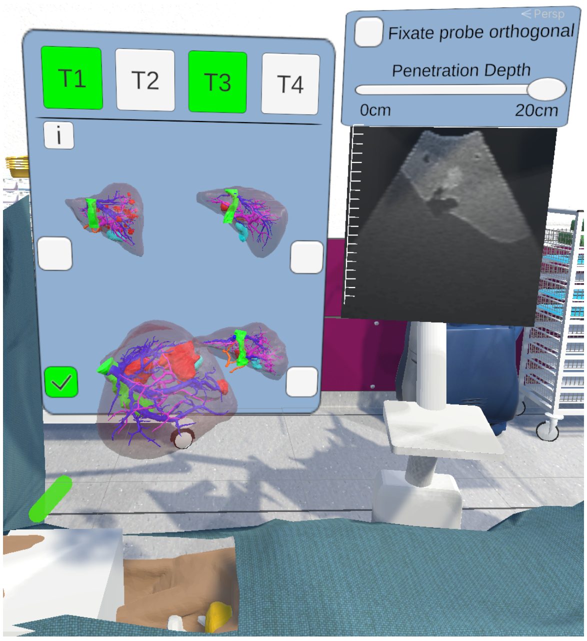
(b) Scenario 2 - scanning veins. The user has to trace veins by placing the consecutive segments in the middle of the US image.

Figure 5.4: Training scenarios 1 and 2. Image from Allgaier et al. [9] and available under a CC BY 4.0 license. The image was split into two.

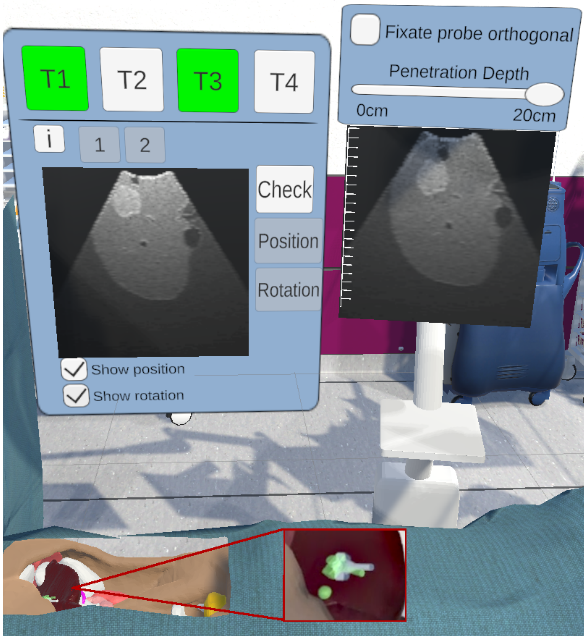
vein they want to trace (see Figure 5.4b). By placing the subsequent vein parts in the middle of the US image (highlighted by the red arrow), the progress bar of the vein fills up.

Scenario 3 The deduction of 3D models from 2D US images is also trained with the third scenario. Based on Mayer et al.'s [196] learning tasks, the user has to scan a liver and has to determine which semi-transparent 3D liver model corresponds to the US images. In order to prevent the user from determining the correct liver based on the liver surface and outer appearance, a white cube occludes the liver in the abdomen (see Figure 5.5a). To recognize details of the inner structures of the semi-transparent models, the user can grab and rotate them with the controller.

Scenario 4 With the last scenario, users can train to interpret an US image, as well as the relation between probe position and orientation and the US image. This scenario is a combination of learning games two, three and four of Mayer et al. [196]. In the proposed scenario, one US image is given; the user has to interpret the image and create a mental model to place the US probe in the immersive VR environment in the same position and orientation to reproduce the image (see Figure 5.5b). The position and orientation are checked separately. If necessary, the user can get hints. The position can be shown by enabling a sphere on the liver, and the orientation can be shown by a ghost view of the probe, showing the correct orientation attached to the virtual probe.



(a) Scenario 3 - transfer US to 3D model. Based on the interior structures, the user has to select the corresponding 3D model.



(b) Scenario 4 - probe handling. The user has to find the correct probe position and rotation to recreate the given image. A green sphere and ghost view can be enabled for support.

Figure 5.5: Training scenarios 3 and 4. Image from Allgaier et al. [9] and available under a CC BY 4.0 license. The image was split into two and in b) the inlay image showing the probe was added.

5.4 Impact

During the development, a resident assessed the US simulation several times to adjust it, for example, by adapting the amount of noise and blur. Despite this, the focus of this evaluation is to investigate whether *LiVRSono* would be useful as additional training. Before the learning outcome and training effect can be evaluated, the application has to be assessed regarding its general suitability. The focus was on two aspects that were also relevant regarding the requirements: plausibility and training. For these aspects, an expert study was conducted with eleven medical experts. Additionally, a short pilot study was conducted to get an impression of the training effect after using *LiVRSono* multiple times.

5.4.1 Apparatus

For the implementation and the two studies, the HTC Vive Pro Eye was used. The studies were conducted with a laptop with the following properties: NVIDIA GeForce RTX 2080 Super with Max-Q graphics card, Intel Core i7-10850H 2.70 GHz CPU, and 32 GB RAM. Once the 3D render texture is precalculated, scanning the liver with the *Geomagic Touch* and using the described IOUS simulation is possible with about 90 fps.

5.4.2 Setup and Procedure

In accordance with the two studies, the setup and procedures are presented separately in the following two sections.

5.4.2.1 Expert Study

The medical experts have varying experience regarding IOUS. Consequently, feedback from six experts of IOUS that can rate the plausibility, and from five members of the target group of the training system was collected. The experiences and demographics are summarized in Table 5.2. The limited number of participants is caused by the specific use case that requires more knowledge and practical experience with surgical liver resection than a medical student has. Complex liver surgery is very demanding. Thus, only a subset of experienced surgeons performs this type of surgery. Because of the limited number of participants and to get in-depth feedback, a qualitative analysis was included.

Table 5.2: Characteristics of experts ($n = 11$). Table from Allgaier et al. [9] and available under a CC BY 4.0 license. The table was split into two columns.

Characteristics	Value	Mean
Age	[28-59]	42
25-34	3	(27%)
35-44	4	(36%)
45-54	2	(18%)
55-64	2	(18%)
Gender		
Male	6	(55%)
Female	5	(45%)
Medical Experience		
Resident	3	(27%)
Specialist	3	(27%)
Attending	4	(36%)
Chief physician	1	(10%)

Characteristics	Value	Mean
IOUS Experience		
None	2	(18%)
5-20 times	3	(27%)
>200 times	2	(18%)
>1000 times	4	(36%)
Experience with VR		
None	2	(18%)
Less than 15 times	6	(55%)
More than 15 times	3	(27%)

After a short introduction presenting the learning objective, each expert first became familiar with the VR environment, haptic device, and interactions. Keeping the learning goal in mind, they had to rate the plausibility using a 5-point Likert scale (1 = not at all, 5 = completely). The exact questions were:

- Is the setting realistic/plausible enough for the learning objective? Setting refers to the US monitor position, user position and patient position.
- Is the handling of the US probe realistic/plausible enough?
- How realistic is the haptic feedback?
- Is the US image realistic/plausible enough?

Afterwards, they were asked to explain their assessment by indicating which aspects were realistic enough and which were not. They were also encouraged to think aloud and to mention problems, and positive aspects. In the second part, they had to go through all scenarios to state whether they are helpful for the learning goal using a 5-point Likert scale. Thereby, they were again encouraged to mention improvements.

5.4.2.2 Pilot Study - Learning Outcome

Six persons with no medical background participated in the additional pilot study regarding learning outcome. This limited number was because the idea was to reserve the target group, which is difficult to recruit, for a large study assessing the learning outcome and comparing it to the current learning method. Before such a study is possible, appropriate cases, difficulties, and feedback from this evaluation have to be included. Because of this, only the third task was used, which was also rated the best and which does not require anatomy knowledge.

The six participants had a pre-test, three training sessions with four tasks each plus a repetition of the previous sessions on following days, and a post-test. Due to technical problems during the first test, a similar test had to be conducted before starting the second training session, which will be referred to as 'pre-test'. This results in the following procedure:

- Session 1: (Cancelled Pre-test) and Training with difficulty level 1 (varying amount of lesions)
- Session 2: Pre-test, repeating difficulty level 1, training with difficulty level 2 (varying positions of lesions)
- Session 3: Repeating difficulty level 1 and 2, training with difficulty level 3 (varying positions and sizes)
- Session 4: Post-test

During the tests, errors and time per task were recorded and the participants had to answer questions based on the competence item of the standardized Intrinsic Motivation Inventory (IMI) [261] plus two additional questions with a 5-point Likert scale. The additional questions directly refer to the learning objective. Another two questions were asked in the post-test where they had to rate the perceived learning effect. The additional questions were:

- I think I can safely navigate within the liver.
- I think that I can quickly orient myself within the liver.
- Post-test: I think my US skills, for example, understanding the images and orienting in the liver have improved as a result of the practice sessions.
- Post-test: If I had to scan a liver now, I think I can use the skills and strategies I have practiced.

Before each test, the participants were told that both error and time are measured, but also that it was more important to be right than to be fast.

5.4.3 Results

Similar to the previous section, the results of the expert study are followed by the results of the pilot study.

5.4.3.1 Expert Study

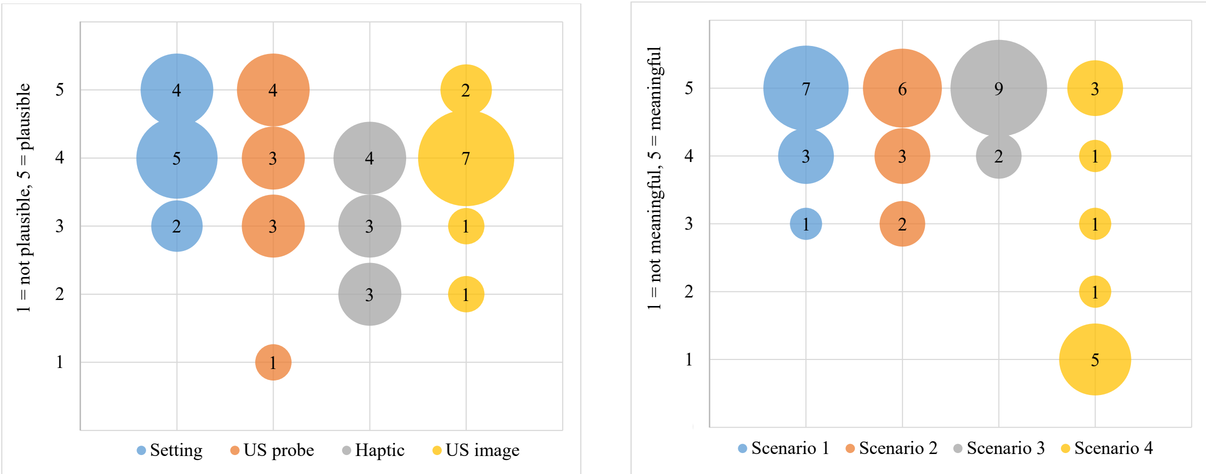
Plausibility In general, the setting as well as the US were rated as realistic, which can be seen in Figure 5.6a. Regarding the general setting, some participants mentioned that it would be better and easier if the US monitor was more to the left or directly in the viewing direction. However, the current setup is similar to the intraoperative setting. It was also mentioned that the monitor position could be a way to include various levels of difficulty. Regarding the virtual patient, the abdomen should be opened further to reveal the whole liver. In Figure 5.5a, it can be seen that the liver was partially covered.

The 3D printed attachment simulating an US probe was sometimes mentioned positively, but experts also emphasized that they use a different one and that the real probe was more cumbersome to handle. One problem regarding the US probe was that the device juddered when there was an indentation on the liver surface. However, the main problem was the restricted workspace. Because of this, experts in particular were not able to scan the liver in the same way they would during surgery. Another aspect concerning the input device was the height. To enable a proper height of the device, it should have been placed on a height-adjustable surface.

The haptic feedback was the least realistic aspect; however, many participants who rated it as not realistic mentioned that it is still supportive and better than without haptic feedback. Only one participant would prefer to have no haptic feedback.

Feedback regarding the US image was very positive. Some experts mentioned that it should be less noisy and edges should be a bit sharper. Two aspects should be improved to increase realism. First, the US image should only be visible if there is contact with the liver surface. Second, it is possible to rotate the probe around its longitudinal axis. This is not possible to such an extent in real life, and the behavior of the simulation was not correct. This was recognized by some of the experts.

Training Scenarios The detailed rating of the scenarios is summarized in Figure 5.6b. Nearly all participants appreciated the first scenario. To further improve the learning experience, some participants suggested the possibility of seeing all liver segments. This could also be included in an additional training room where the liver could be inspected using transparent colors and grabbing interactions. For more clinical relevance, one participant suggested modifying the scenario so that the user has to indicate in which segments metastases are located.



(a) Rating of the plausibility of different aspects.

(b) Rating of meaningfulness and efficacy of scenarios.

Figure 5.6: Results of the expert study. Image from Allgaier et al. [9] and available under a CC BY 4.0 license. Two images were merged into one.

In general, the second scenario was rated as effective. However, there are two important aspects that have to be changed to enable proper training. It was very difficult to place the segments of the vein in the middle of the US image. Instead, it was suggested to point and click on the corresponding vessels using the controller. As mentioned before, the workspace of the input device is limited. Due to this, experts in particular were not able to trace the veins using the method they use during surgery.

The third scenario was the most preferred one. Comments show that the white cube occluding the liver model is not relevant, especially when more similar cases have been integrated. Various difficulties could be included if cases with different amounts of metastases or different courses of vessels are used. Furthermore, the liver models in the menu should be rotated because during surgery the liver is seen from a ventral and not dorsal perspective.

The last scenario was mostly rated as inappropriate. The reason given for this was that there is no benefit in simulating a given US image. Sometimes it is also possible to create a very similar image or an image showing the same relevant structures without having the probe at the same position. Alternatively, a scenario to count metastases or to show a given structure would be more effective and clinically relevant. In this context, some participants mentioned including CT data, because surgeons usually have a mental model of the liver based on the preoperative data. Accordingly, a scenario could also be to find metastases that are not visible in the CT data.

5.4.3.2 Pilot Study - Learning Outcome

In the pilot study, three participants had no errors in the pre-test as well as in the post-test. Each of the tests had twelve tasks. Three participants reduced their errors: from three to one, from four to one, and from two to zero. Additionally to reduced or constant errors, five participants were also faster in the post-test by 13.4s on average per liver scan. One participant needed 30 s more on average per liver scan. However, this participant reduced their errors from four to one, which is more important than being fast. The detailed results are summarized in Table 5.3. Using the eight questions q_n, the competence difference between the pre- and post-test is calculated by:

$$Comp_{diff} = \frac{\sum_{n=1}^{8} q_{n,post} - q_{n,pre}}{8} \quad (5.4)$$

All participants had an increased competence (see Table 5.3). They also answered the question regarding perceived improvements in their US skills with a 5. For the question regarding applying their skills, three participants gave a 4 and three gave a 5.

Three participants used the comment field to emphasize the following aspects:

- Repetitions were very helpful in producing a learning effect.
- They recognized their perceived learning effect and had an increased confidence in performing the US.
- They developed strategies on their own.
- Especially small differences in lesion positions such as the height within the liver were much easier to identify after the training sessions.

Table 5.3: Errors (E), the average time (t) per liver scan in seconds of the pre- and post-test, and the competence difference ($Comp_{diff}$). Table from Allgaier et al. [9] and available under a CC BY 4.0 license. No changes were made.

	E_{pre}	E_{post}	t_{pre}	t_{post}	t_{post} - t_{pre}	$Comp_{diff}$
P1	3	1	26	24	-2	0.875
P2	4	1	41	72	31	1.375
P3	0	0	29	24	-5	0.125
P4	0	0	80	40	-40	0.625
P5	2	0	58	41	-17	1.375
P6	0	0	32	29	-3	0.75

5.4.4 Discussion

The discussion of the presented results is separated into the individual investigation aspects: US simulation and training scenarios. Furthermore, the evaluation and generalization of the project are discussed and future work is presented.

5.4.4.1 Ultrasound Simulation

As described, the simulated US image is not complex and does not include, among other things, deformations. Nevertheless, the image was rated as realistic enough for the learning goal, showing that a fast simulation can be used for training. The US simulation can be further improved by requiring contact with the liver, less noise in the image, and restricting the probe rotation along the longitudinal axis. Although most participants appreciated the US image and its quality, some mentioned that the image, especially the edges, is blurred too much. Because of this varying subjective feedback, directly comparing the US image with a real image might be helpful in giving a more objective result. Before doing so, one could also try out other methods that include more information, such as tissue characteristics, and incorporate more artifacts.

Similarly to Simon et al. [286], the presented evaluation revealed that the haptic feedback is the main limitation of face validation. Although the haptic feedback does not realistically simulate touch sensations from a real liver, nearly all participants stated that the haptic feedback is supportive. The mentioned jiggling and limited workspace are distracting and also limit the user's performance. For an appropriate training, it is crucial to reduce these drawbacks. This can be done by either setting the initial device position such that the whole liver can be reached, or alternatively by scaling the movement. However, this is only possible to a certain degree while still preserving realism and an appropriate level of difficulty. If this is not possible, another input device or haptic feedback simulation might be more useful, and alternative products such as the *Emerge Wave-1*[4] should be considered in the future. Furthermore, a direct comparison of having haptic feedback versus no haptic feedback should be considered. The drawbacks of using a 3D-printed liver were already discussed in the introduction. However, one could investigate the impact of the mentioned mismatch of the printed and virtual models and compare the two input modalities. Using no haptic feedback might work with the current state. However, when considering the requirement that the probe has to have contact with the liver surface to create an US image, it will probably be very difficult and exhausting to scan the liver.

Most participants rated their position in relation to the patient and to the US monitor positively. The height of the input device was sometimes not appropriate. This can be solved by using a height-adjustable surface, as the operating table is also adjusted to the surgeon's height. Some participants also mentioned that they would prefer to have the monitor in their field of view, which would differ from the real setting. Including different monitor positions might be a way to include varying difficulties. Having the monitor not in the field of view is much more difficult due to hand-eye coordination. This is a great benefit of VR: the three components–patient, user and monitor–can be arranged in the correct way. Using a normal desktop, such as in Law et al. [175], cannot provide this. Only Byl et al. [44] provide an immersive VR environment; however, here the user did not have to handle an US probe.

5.4.4.2 Training Scenarios

Three of four scenarios were rated as helpful with minor improvements. As there is no clinical benefit to the last scenario, and the handling of the US probe can also be learned with other scenarios, it is recommended to remove or replace this scenario. In a simpler setting, such as in Mayer et al. [196], this scenario might work, but in a liver with many vessels and metastases, different probe positions and rotations might also lead to similar images. Some participants would prefer a scenario where the user has to find either all metastases or a specific one. For this scenario, the user has to scan the whole liver. Thus, they need a good orientation and understanding to differentiate whether a metastasis is a new one or one they have already scanned but seen from another perspective. This would also

[4]Emerge, USA: https://shop.emerge.io/ Last access: 15.04.2024

have a high clinical relevance because there are metastases that are not visible in the CT data and therefore have to be found with US. Alternatively, this could be included in the first scenario. Instead of simply stating which segment they are scanning, the user could also have to count the metastases in the current segment.

Another aspect of a meaningful training are the included medical cases. Although *LiVR-Sono* includes four cases, only one case was used for the evaluation due to time restrictions. However, this case is a difficult one for the first scenario because the gallbladder was already removed. In everyday clinical practice, the gallbladder is used for orientation within the liver, since it attaches to segment five of the liver. The four liver cases are very different in regard to their diseases (such as the number of metastases leading to a relatively easy third scenario). It was not part of the evaluation to include appropriate training cases regarding anatomy and difficulty, but this might be considered for further studies assessing, for example, the learning outcome.

5.4.4.3 Evaluation

Regarding the expert evaluation, questions directly referring to the application instead of standard questionnaires, such as questionnaires concerning usability or task load, were employed. The reason for this was the very limited time of surgeons. It was not possible to include more questionnaires and that is why specific questions were preferred to more general ones. Furthermore, it can be assumed that if the usability as well as technical aspects would not be sufficient, the participants would mention this as the think-aloud method was used and their task was to assess the quality of the training application. Questions referring to the mental and physical load would only be beneficial in comparison to the real IOUS, otherwise, a high load could indicate that the application is too difficult or it could reflect the real situation. However, in a large user study comparing the training to a control group, these standardized questionnaires are recommended. It is also necessary to emphasize that this study only assesses face and content validity and that the questionnaires used are not validated. Further studies must be conducted to evaluate other types of validity, such as construct, or content validity. Aside from validity, the fidelity of the simulation could also be assessed [112].

In the pilot study difficult tasks, such as where the exact position of a lesion (in relation to the vessels) has to be determined, were not included because of the missing medical background of the participants. Due to the reduced difficulty, the ceiling effect could be observed as three participants had no error in the first test (and also the second test). However, the reduced time, errors, and the questionnaire in particular showed positive results regarding the increased confidence and competence. Comments especially emphasize the learning effect regarding spatial orientation and exact lesion positions. The study reveals that something new can be learned after training with the system. Since the tasks require visuospatial skills, it can be hypothesized that these are trained. However, further studies are necessary to confirm whether this target skill can be learned and to investigate whether it can be transferred to the real procedure using statistical analyses. In a more recent work presented by Junge et al. [144] the learning effect of a VR-based US training for trauma assessment was compared to a historical control group using a screen-based simulation with a haptic trainer. Acquired competencies were assessed using the simulator from Schallware. The VR-based training *VitaSim* comprised the placement of the US probe, image optimization, and interpretation of normal and pathologic findings. This was implemented by presenting basic knowledge, training cases and finally combining given diagnosis with the correct virtual patient. The VR-based training was non-inferior to the control group with respect to the acquired competencies. How a similar evaluation concerning learning outcome can be designed for the presented case is described in future work.

5.4.4.4 Generalization

As described in the introduction, IOUS is also used for other applications than the liver. Although the proposed application is specialized for the liver, some aspects might be generalizable to other applications. In other organs, such as the pancreas or kidney, IOUS is also used to identify tumors, metastases, or other anatomical structures like the duct. Consequently, the third training scenario with its previously discussed variations is applicable. The first two scenarios are specialized for the liver and might be removed or adapted for other organs.

5.4.4.5 Future Work

A summary of the main aspects that should be improved to enable a proper training environment is shown in Table 5.4.

Table 5.4: Summary of essential improvements. Table from Allgaier et al. [9] and available under a CC BY 4.0 license. The table was slightly expanded.

	Improvements
US image	Sharpen edges, reduce noise
Haptic	Smooth the surface, workspace
Probe	Add more and different US attachments
Setting	Adjust the height of the input device in relation to the user
Scenario 1	Show segments for learning
Scenario 2	Remove the arrow and select the vessel via controller
Scenario 3	Change liver orientation
Scenario 4	Remove or replace

Furthermore, a training system requires different levels of difficulty. These can be included by incorporating:

- different sizes, locations, and amount of metastases,
- different levels of the information shown to, or asked of, the users. For example, whether the users have to differentiate the livers in the third scenario based on the number of metastases, their location or simply based on the vessels. In the second scenario, the precision of tracing the veins could be varied.
- further scenarios, such as measuring the size of a metastasis or showing its relation to vessels, and

- including cases with common anatomical variations and cases with rare anatomy.

After including varying difficulties, future studies should assess the learning outcome with the target group when using *LiVRSono* over a longer period. By comparing a group regularly using *LiVRSono* to a control group with no additional training, one could measure whether the aspects described in Section 5.3 could be the result of learning from the system. However, there is a difficulty of a lacking ground truth or test environment. If the user's skills are tested within *LiVRSono*, similar to the presented pilot study, it only states that one can improve within the system by training with this system. There are the following alternatives:

- Using a phantom for US training: This would be a realistic setting, however, these phantoms are very expensive and there might be a limited amount of liver cases. Creating own liver phantoms is also very time and resource-consuming, and requires knowledge in this area.
- Using a molded liver in combination with an US simulation: This enables a large amount of liver cases, however, the same US simulation would be used for training and testing, which also limits the transfer of the results to the real situation.
- Using the real situation: Regarding the validity of the results, this option would be the best. In this case, there are ethical concerns as well as an enormous effort. In addition to the amount of study participants several surgeries to test the participants would be necessary, which is an organizational effort and requires additional time during surgery. Using real patients as pre- and post-tests has the drawback that the test cases vary a lot with respect to difficulty due to interpersonal variations. There is also no ground truth. One possibility would be that an experienced surgeon assesses the performance of the participant which might also include bias. However, this alternative is the only way to investigate whether the learned skills can be transferred to real IOUS.

All these possible studies would require that regular training sessions with LiVRSono are included in the clinical routine.

As emphasized in the introduction, *LiVRSono* focuses on IOUS for the liver but addresses the general need to train visuospatial skills for US and the lack of training possibilities for IOUS. In the future, the proposed training system can be adapted to other applications, such as IOUS for kidneys or pancreas. The adaptation would include adjusting the grey values of tissues (if necessary) and replacing workflow-related training scenarios. This can also include other medical disciplines such as neurosurgery [74, 110], however, first the necessity of such a training and the requirements have to be verified.

5.5 Conclusion

With *LiVRSono*, the need for training systems to train the mental skills that are necessary for IOUS is addressed. The proposed immersive VR system for liver surgeons benefits from a real-time US simulation, a modified haptic input device, as well as a virtual OR, which improve the learning experience by providing a setting similar to the real situation. Furthermore, training scenarios were identified based on the application-specific workflow and the transfer between the US image and 3D anatomy. Using the proposed system, drawbacks of the chosen input device, as well as important improvements of the training scenarios to enhance the training without harming real patients were identified.

6

Gamification Concepts for a VR-based Visuospatial Training for Intraoperative Liver Ultrasound

Synopsis After describing the motivation and contribution, this chapter proceeds with relevant literature. Using one training scenario, various game elements are discussed to identify proper ones for an US training simulation. The development includes the implementation of the selected game elements. Finally, this chapter presents two studies to evaluate and compare the elements regarding motivation and player experience.

This chapter is partially reused with permission of © 2024 ACM and has been published in:

Mareen Allgaier, Florentine Huettl, Laura Isabel Hanke, Tobias Huber, Bernhard Preim, Sylvia Saalfeld, and Christian Hansen. "Gamification Concepts for a VR-based Visuospatial Training for Intraoperative Liver Ultrasound", *CHI EA '24*, Association for Computing Machinery, NY, USA, Article 175, 1–8, 2024, https://doi.org/10.1145/3613905.3650736 .

6.1 Contribution

In the previous chapter, an appropriate VR-based training system including various training scenarios has been presented. However, simply providing a training system does not automatically imply a good learning outcome. In Chapter 3, other factors, such as perceived motivation, enjoyment, repetition and practice, that influence the learning outcome are mentioned [98, 147]. To foster these aspects, the term 'gamification' was already introduced in this chapter. In education, gamification can lead to an increased motivation, engagement as well as enjoyment [109]. However, there are also studies revealing the negative effects of gamification [91, 314].

As presented in Section 3.2, there are many gamified systems in education. In medical education, most systems address *declarative knowledge* and are designed for use in lectures. Other medical systems are designed for patients, such as rehabilitation or diagnosis. Consequently, there is a lack of gamified practical training systems that are designed for physicians.

This literature research also showed that there is a general lack of studies investigating the effects of individual game elements and the effects of combinations of them. The few studies investigating the effects in detail apply them to general use cases, such as an online quiz [198], and/or focus on frequently used game elements like badges and leaderboards [34]. Even for the game elements that are explored more often, it is questionable whether the findings can be transferred to other gamified systems with a different context and activities [108, 160, 256]. Different contexts, technologies, methods as well as the implementation of gamification might influence their effects, which might also explain the contradictory results of some studies. Literature introducing gamified medical training using VR tends to evaluate the system itself employing common questionnaires, such as the *user experience questionnaire* and *system usability scale*, however, this does not prove any positive effects of gamification. Instead, it is recommended to first develop a training application focusing on the training itself, considering aspects, such as relevant training scenarios, appropriate interactions, proper feedback and balanced difficulties. A good training application should be meaningful and user-friendly even without additional game elements.

In the second step, the application can be gamified to increase motivation and improve the result in the best case. These two development steps should be separated and evaluated individually because gamification makes no sense if the training itself is not meaningful and a bad implementation of gamification can also impair the training. Because of this, a part of the previously described training *LiVRSono* (refer to Chapter 5) has been selected to investigate the following research question:

RQ 2: Which game elements are appropriate for a VR-based training for visuospatial skills for IOUS?

Contributions include the analysis and discussion of proper game elements for VR-based US training. The final selection of game elements includes common elements, such as levels, but also gamified interactions to solve the task, which can also be used in similar training systems. A broad-audience study compares the game elements as well as the combination of them with a control group. An additional within-subject design study with medical students directly compares the game elements. The studies include multiple standardized questionnaires as well as own questions specifically asking for reasons of motivation and the choice for a specific game element. The two studies aim at investigating the appropriateness and preference of the game elements as well as how they influence the user's motivation, experience, and performance.

6.2 Design

To recap, the first section summarizes relevant literature concerning gamification and visuospatial skill training. After presenting the training scenario that was gamified, the target group and desired behavior are analyzed to discuss various game elements and their suitability in the presented scenario.

6.2.1 Related Work

Relevant literature for this project includes related work regarding gamification and related work for visuospatial skill training. For the first, refer to Section 3.2. Related work concerning visuospatial skill training for US is presented in Chapter 4. Three of the mentioned approaches and the commercial *VitaSim* are gamified or serious games and thus are described below with a focus on gamification.

Mayer et al. [196] proposed a game consisting of four minigames where the user, for instance, has to find the correct geometric 3D scene using US. Different levels are included based on varying complexity of the 3D scene, the probe movements that are required, and the visualization mode (different colors, grayscale and a simulated US). To guide the user through the game, the game was divided into different campaigns, consisting of different challenges. To process from one campaign to the next, the user has to collect a specific amount of medals by correctly answering rounds. By providing an overview scene, the user can track their progress. They did not compare the game to a non-gamified version or any other training, they compared the learning outcome of a group playing this game twice to a control group not playing the game, using the Santa Barbara Solids Test (SBST). The group playing the game showed a greater improvement from the pre- to post-test, but not significantly.

Byl et al. [44] presented a similar serious game that also lacks medical context, but is VR-based. By correctly performing US in a given time, the user gets gaming points, which are then used for a high score list. The evaluation investigated the attractiveness and efficiency of the game using qualitative feedback from nine users.

The third approach presented by Olgers et al. [225] is desktop-based and takes place in an underwater world where players have to collect hidden coins. The user has to use an US probe to find all coins. Levels and scores used for local competition are included game elements. With 42 participants, they were able to show that their game possesses content and face validity.

The already presented commercial VR-based training *VitaSim* was gamified in a recent project [173]. Using the module for lung US, a gamified version was compared with a non-gamified one. The gamified version comprises quizzes and training missions with which experience points can be gained to unlock subsequent rooms. The choice of these game elements was not further motivated. They found no significant differences regarding test scores between the gamified and non-gamified groups and the game elements were not assessed individually.

Referring to Section 3.2 and the above presented approaches, it is recognizable that there are many gamified systems in education. However, in medical education, most systems address *declarative knowledge* and are designed for the use in lectures like quizzes. Other medical systems are designed for patients, such as for rehabilitation or diagnosis. Consequently, there is a lack of gamified practical training systems that are designed for physicians. Furthermore, only a subset of the gamified approaches includes game design decisions and an evaluation that compares the individual game elements and uses a control group. There are no known studies investigating individual game elements for visuospatial training for US.

6.2.2 Training Scenario

To reduce the complexity of the study, the training scenario of *LiVRSono* which was rated as most appropriate and meaningful and is also similar to other US training applications was selected [44, 196]. Here, the user has to scan a virtual liver using the haptic device *Geomagic Touch* and select the correct 3D liver model from a UI. To avoid a selection based on the outer appearance, training cases that only differ regarding the interior structures were created. For the study, only the amount and positions of lesions varied. Further tasks might include various lesion sizes and courses of vessels. With this scenario, building a mental model and understanding the internal structures and their spatial relations is trained. Probe manipulation is not the focus of this training.

6.2.3 Conceptualization of Game Elements

Although there is a lot of literature, such as surveys and proposed frameworks for gamification, the challenge of finding a suitable approach still exists [5]. Different frameworks and design principles emphasize defining objectives and the mission, knowing the player, and defining the target behavior and motivation [168, 213, 335].

Objectives and Mission The overall objective of the training system is to train the creation of a spatial mental model when using IOUS. More precisely, this means the user should learn the anatomical and spatial relations, interpreting the US image (Which anatomical structure is visible? Where in the organ is the current US image?), a systematic and appropriate handling of the US probe, as well as the correlation and transfer of the 2D US image and 3D model.

Target Group One method in user-centered design is the so-called persona [66]. Cooper et al. [67] describe personas as 'composite archetypes' of, for example, users based on the behaviors and motivations of real people. In the area of human-computer interaction, interaction designers can emphasize with them, leading to user-oriented decisions based on their defined motivation, goals and feelings [67]. Based on a survey with seven liver surgeons, the following user persona arose. The user is a liver surgeon with no or little IOUS experience. They are familiar with the liver anatomy and have observed IOUS several times during liver surgery. Due to limited time, they could not use a training system for a long time at a stretch but prefer shorter training sessions. The user is not a regular gamer (two out of seven liver surgeons play video games several times a week). The gaming type using the HEXAD questionnaire with seven liver surgeons reveals that the user is predominantly an achiever, socializer and philanthropist.

Behavior and Motivation During the training task, the user should scan the liver systematically to recognize conspicuous anatomical aspects, such as the number and location of lesions, vessel variations, or the proximity of lesions and vessels. Based on this, they have to select the correct 3D model. The intrinsic motivation of the users should be to improve their IOUS skills to become a better surgeon with the consequences of, for example, feeling better prepared for their first actual IOUS on a real patient and of providing better patient care. Chou [58] also emphasizes: 'It's not just what game elements you put in – it's how, when, and most importantly, why these game elements appear' (p.19). Because of this, one has to think about how you want the user to feel when using the gamified system. In the presented case, the user should feel ambitious and curious to learn or train a (new) skill and should also feel competent.

After an introduction of the training system *LiVRSono*, liver surgeons were asked about their thoughts regarding different game elements that could be included in the training. Thereby, Huang's [127] list of game elements served as basis. From this list, the following game elements were excluded beforehand:

- Quests: The idea is to divide a task into sub-tasks. As the tasks in the training system are already very small and trial and error learning as well as exploration should be supported, quests were excluded. For more procedural training scenarios, such as scenario one, one could include a small quest according to the normal workflow, such as finding the hepatic vein star and the next vessel segment.
- Badges and awards: This element is used to honor the completion of a task and to show status. However, it is important that the badges and awards provide a benefit for the user, otherwise they do not motivate [91].
- Avatars: Individualization of the virtual representation creates a stronger feeling of relatedness. As the VR application is in first person view, the own character cannot be seen by oneself (and by others if it is no multiuser application). Consequently, the benefit of avatars would probably disappear.
- Adaptivity and non-linear navigation: This element requires some kind of story or context for actions with different consequences. In the training system are only independent tasks and actions which make non-linear navigation difficult.

Leaderboards and competition were assessed together. In Table 6.1, the remaining game elements and the opinions of liver surgeons are summarized.

Table 6.1: Game elements and comments of liver surgeons.

Game element	Description	Opinions
Points and experience	Quantification of actions and performance and good for tracking progress, performance assessment for self-improvement and competition.	'Motivating', 'Good', 'Good for improving performance', 'not necessary', 'punishing errors because otherwise, people don't think, they just click'
Competition and Leaderboard	Comparing performance	'Might have positive aspects', 'Competition is always great', 'Unrealistic, but probably fun', 'Competitive, but global score is sufficient (no multiuser)', 'Nice but not necessary', 'Might be fun for training'
Collaboration	Interpersonal communication	'During surgery one has to do it alone as well', 'Teams would be good'
Levels and advancement	Implicate expertise and skill	'Helpful (including different livers and tumors (locations, sizes,...))', 'Including different tasks, such as measuring sizes could be included', 'More common and more seldom cases could be used'
Time	Used to encourage fast decisions and adds a new challenge	'Makes no sense (there is no time pressure during surgery)', 'Not important (in real surgery), but due to learning improvement one should become faster', 'Motivating', 'Good for orientation (how fast is an expert), but there should be no time pressure', 'Good'
Narrative	Includes some kind of story and context for the actions	'Not necessary, might be distracting', 'Not relevant clinically or for the learning task', 'Might be nice to create a whole game', 'Might create a more serious setting', 'Not relevant, maybe a narrator who presents what the user currently sees', 'Even if there is no clinical benefit, it would be nice to have a context and might be better for learning and remembering because of the additional association', 'Would be nice to know what kind of tumor'
Responsive feedback	To reveal the current state and changes in the current state	'The included one was good', 'Auditory feedback would be nice',

Considering the given conditions such as target group, desired behavior and medical feedback, the choice of game elements from the remaining ones is discussed in the following paragraphs.

Points and experience Although most surgeons would appreciate points, they were neglected in the comparison, because they are very investigated in the gamification research. Westera [336] advises that a scoring system in educational serious games should not create a *performance attitude*. Instead, a *learning attitude* should be provided by avoiding stress factors and enabling reflecting on decisions and achievements. In this use case, the term 'points' can, for example, be replaced with 'correct diagnoses' to bear reference to the medical relevance and thus intrinsic motivation.

Leaderboard and competition When including points, leaderboards are often the next logical step including a social aspect of points [168]. When implementing a leaderboard, it is recommended to limit it to a few users performing a bit better and a bit worse than the current user to not demotivate them [168]. Otherwise, users with poor performance are in danger of being demotivated, but these are especially the users that should be motivated by the system to train more frequently. One could also distinguish between a global leaderboard and a local one among friends. Leaderboards were not included, but it was assessed whether a leaderboard is desired.

Collaboration Regarding competition, the feedback was ambiguous. As mentioned by an expert, during surgery they have to scan the liver on their own as well. Moreover, part of the visuospatial challenge is to navigate the probe properly; it is not feasible to have a second person who can only see the US image. In practice, collaboration would probably not work due to the limited time described in the target user persona.

Levels and advancement Another element are levels of difficulty. Starting with easier tasks and gradually increasing the difficulty is essential to not demotivate the user. Letting them know at which level they currently are and thus rating their performance is another way to assess their performance and motivate them. Increasing the difficulty gradually also makes sure that the users can further improve their skills. The experts preferred different levels and proposed various ways to vary the difficulty.

Time Another performance-oriented game element is time or time pressure. Here, the medical experts' opinions were different. As during a real liver surgery it is more important to scan properly than being fast, time pressure was not included. The surgeons also stated that time for scanning is not essential because it only accounts for a fraction of the time that the whole operation takes. Furthermore, time pressure prevents explorative learning, taking time to understand the handling and procedure, and reconsidering decisions for in-depth understanding. Especially users with no or little experience in IOUS need time for orientation, understanding, and handling. Time pressure only leads to stress or even panic and might negatively influence the learning outcome [336]. However, some experts mentioned that time can be used to assess improvement. For unexperienced users, it might also be a good orientation to see how fast experts are.

Responsive feedback Feedback is not considered a game element, but a necessary aspect to reflect on one's performance and to improve. Feedback in terms of red and green for false and correct answers was appreciated in a previous version of the prototype. As auditory feedback was desired by some experts, two different sounds for false and correct answers were incorporated. However, this game element is part of the base version and not part of the comparison.

Creativity and exploration This game element is not listed in Huang et al.'s [127] meta analysis, but was inspired by Chou's [58] *Octalysis Framework* which includes the two core drives *empowerment of creativity* and *ownership*. The first core drive about creativity and exploration leads to long-lasting motivation. For example, drawing triggers this core drive because the user can be creative and immediately gets feedback. Ownership and possession is the core drive, for example, for collecting stamps and puzzle pieces or virtual goods. In the underlying scenario, it is difficult to include creativity, since there is only one correct answer and there is a limited amount of data. However, the aim is to address creativity by providing a *kit* with various anatomical parts of the liver which the user has to combine to build the current model actively. This might create a feeling of ownership and possession similar to puzzles. This game element refers to the previously mentioned curiosity the user should feel.

Narrative An underrated game element is narrative, story or meaning. Chou [58] describes it as the user 'believes they are doing something greater than themselves' (p.25). This core drive is especially essential during the *discovery and onboarding phase* of a gamified system. Because of this and to emphasize the context and importance of the skill, a start scene conveys the higher goal of the training system. However, this was also included in the base version.

6.3 Development

In this section, the implemented game elements are summarized, their design is described in detail and the different groups of the comparative study are demonstrated.

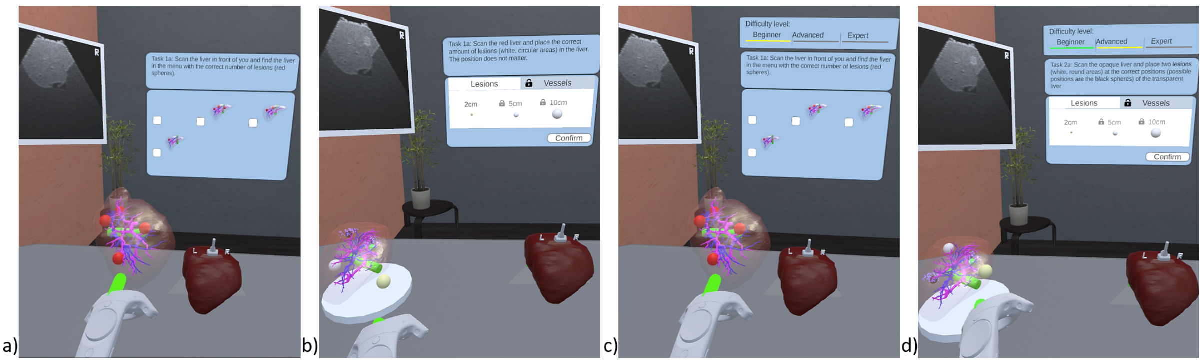


Figure 6.1: Gamified liver US: a) common UI interaction, b) assembling a liver using a kit, c) UI interaction with difficulty levels, and d) kit interaction with levels. Image from Allgaier et al. [10] © 2024 ACM. No changes were made.

Basis For simplicity, the base version consists of a clean office-like room with a patient bed and a virtual liver, which has to be scanned (see Figure 6.1). Because other anatomical structures than the liver and its interior structures are not considered for the US simulation, they were also discarded. The US monitor is on the left side for right-handed users and on the right side for left-handed users. Although this does not reflect reality, it facilitates the situation for non-experts. To reveal the relevance of this task, the base version includes a short narrative context directly at the beginning after introducing how to interact. The dialog is as follows: *'Our liver surgeon, Dr. Peters, is asking for your support. To remove liver lesions correctly (surgically), it is important to determine the number of lesions and their positions with the help of US.*
Diagonally in front of you, you see an US monitor. Use the haptic input device to move the US probe on the liver surface (in front of you on the patient bench). By moving the probe systematically, you can get an overview of the inside of the liver.
Dark structures with white borders are vessels. Bright, large circles represent lesions. In the following, you will be given various tasks and you should try to solve them correctly'.
Depending on the study group, different tasks to grab a liver or lesion are presented to become familiar with these interactions. The dialog ends with: *'Are you ready to help Dr. Peters with the diagnosis? Then start with the tasks'.*

In the base version, the user has to scan the liver and select the correct 3D liver from the UI (refer to Figure 6.1 a and c). The models can be grabbed to get a closer look. After selecting, the checkbox either turns red or green depending on whether it is right or wrong. Accordingly, audio feedback is played. The user can only proceed with the next task when solving the task correctly. This results in a base version that serves as a control group that might already count as gamified. However, it is only feasible to compare additional game elements with an appropriate base version, which includes, for example, feedback. If the base version is not suitable for training in general, it is very likely that any gamified version performs better.

Levels The first additional game element is levels of difficulty (refer to Figure 6.1 c and d). The aim is to investigate whether the awareness of which level the user is currently on makes a difference in motivation. Consequently, all groups have the same training cases with two easier and two more difficult tasks. However, only the groups with levels know that they are currently at the beginner level. This is achieved by several hints. Firstly, instead of ending the dialog with '*Are you ready to help Dr. Peters with the diagnoses? Then start with the tasks*' it says '*Are you ready to help Dr. Peters with the diagnosis? Then start with the beginner level*'. Secondly, above the menu, the user can see the three levels *beginner*, *advanced* and *expert* and that they have not completed them yet. Third, after finishing the tasks, a pop-up congratulates on finishing the beginner level and asks whether the user is ready to proceed with the advanced level. The same would be for the transition between the advanced and expert level.

Kit The second game element is the so-called liver kit (refer to Figure 6.1 b and d). To provide a more interactive task and to give the impression of assembling the correct liver, the user does not have to just select the liver from the UI. Instead, there is a second, semi-transparent liver next to the liver they have to scan. This semi-transparent liver has four possible positions where the user can place a lesion. This lesion has to be grabbed from the kit menu and can be placed at one of four predefined positions. To facilitate the placement, a snapping function is included and the position where it will snap to is highlighted in green. After placing lesions, the user has to confirm their answer, which leads to immediate feedback by turning red or green. To tease at higher levels, the kit menu indicates that there might be more choices later on by including different sizes of lesions and vessels. These are currently locked.

From these game elements, the four groups summarized in Table 6.2 and Figure 6.1 arose.

Table 6.2: The four groups of the study design: Control group (CG), level group (LG), kit group (KG), and kit and level group (KLG).

	Feedback	Narrative	Levels	Kit
CG	x	x		
LG	x	x	x	
KG	x	x		x
KLG	x	x	x	x

6.4 Impact

This section is divided into two studies: a broad audience study and a medical target group study. In both studies, two tasks with two sub-tasks, which means two different liver cases but the same task description, were used. In the first task, they had to count the number of lesions in the liver. In the case of the kit, the positions of the placed lesions were not considered. In the second, more challenging task, there was a liver with two lesions and the participants had to identify their positions. The order of the two livers in each task was randomized.

6.4.1 Broad Audience Study

To investigate the effects of the individual game elements, a broad audience study was conducted with the four presented groups (refer to Table 6.2). A between-subject design was chosen to investigate whether gamification leads to an increased motivation and performance without having a direct comparison to the non-gamified version. When using VR, it is important to also use VR as control group and just remove the gamification elements but still provide a complete application with, for example, appropriate feedback. Otherwise, possible differences between the gamified versions and the control group might be due to VR but not due to gamification.

As the actual target group includes residents or higher because of the required medical knowledge and experience regarding IOUS for liver, it would be very difficult to reach a proper participant number with the target group. Providing all necessary medical information in the instruction and simplifying the tasks as well as the environment ensures that participants without medical background are appropriate. Using participants not belonging to the target group is a limitation, however, it is very difficult to recruit that many participants belonging to the target group. Furthermore, it is assumed that motivational differences should also be visible with participants without medical background. With these adaptions, the study was conducted at the open day of the university to get a diverse group of participants. Results from participants with more than two contradictory answers were removed. All other contradictory answers were removed only for the two ambiguous items. Another two participants were excluded from the VR results as well as the questionnaire results because of a trial and error strategy without understanding the functionality of US which was recognizable due to their behavior. This resulted in 78 valid participants. The participants' demographics are summarized in Table 6.3. This also includes the distribution of player types. Therefore, the occurrences of the six types as primary types were counted, for example, a participant being primarily an achiever as well as a player counts for both. Achiever and philanthropist being the most frequent player type and disruptor the least frequent is in accordance with Tondello et al. [315]. The distribution of the other three types differs.

Table 6.3: Characteristics of participants ($n = 78$). Table from Allgaier et al. [10] © 2024 ACM. The table was split into the two studies and two columns.

Characteristics	Value	Mean
Age [years]	[16-65]	
below 21	9	(11.5%)
21-25	28	(35.9%)
26-30	19	(24.4%)
31-35	12	(15.4%)
36-40	4	(5.1%)
41-45	3	(3.8%)
46-50	0	(0%)
51-55	0	(0%)
56-60	2	(2.6%)
61-65	1	(1.3%)
Gender		
Male	45	(57.7%)
Female	33	(42.3%)

Characteristics	Value	Mean
Experience with VR		
None	18	(23.1%)
Less than 15 times	48	(61.5%)
More than 15 times	12	(15.4%)
HEXAD player types		
Achiever	42	(53.8%)
Philanthropist	41	(52.6%)
Player	25	(32.1%)
Socializer	24	(30.8%)
Free Spirit	16	(20.5%)
Disruptor	1	(1.3%)

6.4.1.1 Study Procedure

The group assignment was equally distributed and assigned randomly. Before the actual procedure started, the participants were informed about how to wear a VR HMD and about the medical relevance of the training. This also includes the information that US always shows a cross section of the scanned objects and that there is no time pressure when solving the tasks. In the VR application, the participants were led through a short instruction, including the previously described narrative context. All relevant interactions, such as pressing UI buttons, grabbing and placing objects as well as the US monitor and US probe, are described.

Afterwards, they had to master the two tasks described above. After finishing them, participants in the level group were informed that they start with the beginner level, and they were notified when they had mastered this and could proceed to the advanced level. After finishing the tasks, all participants were informed that they could not proceed due to time restrictions. The total amount of liver cases was limited to four because otherwise, the whole study procedure would last more than 20 to 30 minutes. A longer study procedure was not feasible because most participants took part during an open day at the university. The open day was chosen to recruit a diverse group of participants instead of using, for example, only STEM students.

This practical part is followed by some follow-up questions within the VR environment regarding their motivation to proceed and engagement. By including them in VR, there is no break in-between and the participants answer them with their current feelings. Although standardized questionnaires were used later on, questions which are not validated were asked as well. The reason is that the questions in standardized questionnaires ask for motivation and effort in general, whereas the other questions aim at directly asking whether they want to continue and, if so, why. Therefore, the participants had to rate the following statements in the VR environment:

- Q1: If there was still time, I would like to continue.
- Q2: I want to continue to test my skills on more tasks.
- Q3: I want to continue to discover and try more possibilities (different sizes of lesions, vessels,...).
- Q4: I do not want to repeat or continue playing it.
- Q5: It excites me to try out more difficulty levels.
- Q6: It was important for me to solve the tasks correctly.
- Q7: I would like to know how many diagnoses (tasks) I got right/wrong.
- Q8: Tick the answers that apply (multiple choice).
  - I want to put my score on a leaderboard and compare myself to others.
  - I want to compare myself with others, but do not want to enter my result.
  - I want to save the result for myself so that I can improve.
  - I don't care about my score.

Except for Q8, the statements were answered using a 7-point Likert scale. After the general statement Q1, reasons to continue were assessed. Reasons could be to test one's skills (Q2), to explore medical variations (Q3), or to try more difficult tasks (Q5). Q5 was only asked if the person belongs to one of the two groups using levels. The levels as well as the kit might indicate that there are more and other tasks to test out, which might influence Q2. Q3 directly addresses the desire to explore medical variations which are more prominent in the kit than in the UI. Q6 assesses the ambition of the participant and Q7 was included to investigate the demand for a score. With the last statement, the demand of a leaderboard was evaluated.

After finishing the VR part, the participants were asked to fill out the following standardized questionnaires:

- IMI [261]. The subscales interest/enjoyment, perceived competence and effort/importance were used as these are the most appropriate in the present case.
- Mini Player Experience Inventory (miniPXI) [105]. This questionnaire tries to investigate how player experience a game using different functional and psychosocial constructs.
- Hexad-12 [164], a short version of the gamification user types Hexad scale. With this questionnaire it can be investigated whether single game elements are only appreciated by a specific player type and it can be ensured that different types were part of this study. Hexad-12 differentiates between the player types achiever, philanthropist, player, socializer, free spirit, and disruptor.
- SBST [64]. This spatial ability test was chosen to investigate correlations between this test and the training tasks to validate that mastering the US tasks is related to having good spatial abilities.

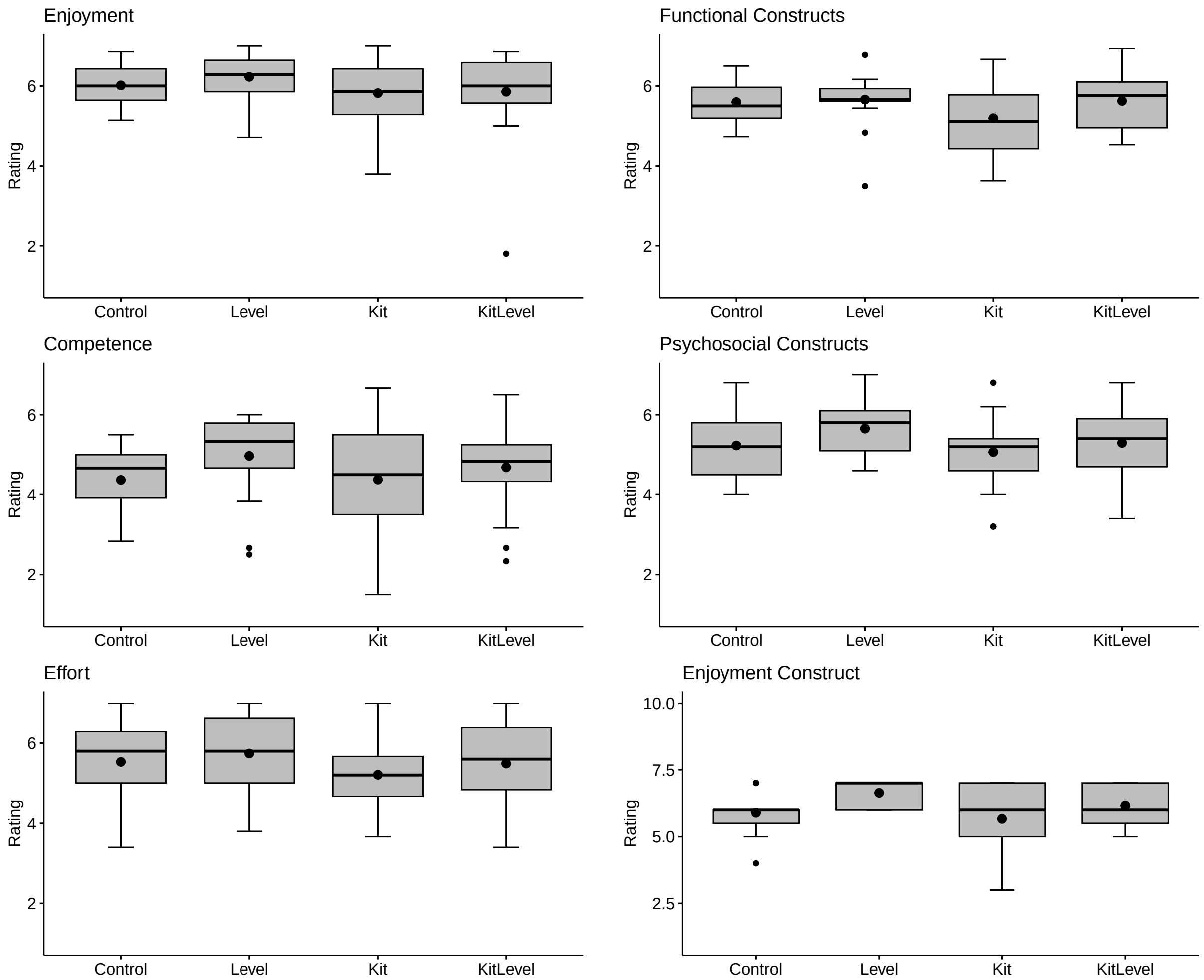
An alternative to the SBST would be the *Schnitte* test [245]. Mayer et al. [196] proposed this test because it measures higher levels of spatial abilities than the SBST and thus the ceiling effect would be more unlikely. However, this test takes 45 minutes, which would be too long. Even the SBST takes approximately 10 minutes. Consequently, this test was voluntary during the open day, resulting in 71 participants conducting the SBST.

6.4.1.2 Results

For all investigations a one-way ANOVA was performed if normality and sphericity–the homogeneity of variances in a between-subjects variance analysis–were given. Normality was tested using the Shapiro-Wilk's method, whereas sphericity was tested with the Levene test. If normality and sphericity could not be confirmed, the non-parametric Kruskal-Wallis test was used after consultation with the university's statistical advisory service. For pairwise comparison in case of significant differences, the common Dunn's test is used.

IMI For the three subscales enjoyment, competence, and effort no significant differences could be found. Descriptive analysis shows that LG has the highest mean for enjoyment with 6.23 ± 0.61 followed by CG 6.02 ± 0.53, KLG 5.81 ± 1.15, and KG 5.82 ± 0.8 (refer to Figure 6.2a). In the case of competence, LG also achieved best results with 4.97 ± 1.03 followed by KLG 4.68 ± 1.06, CG 4.37 ± 0.85, and KG 4.38 ± 1.44. Participants belonging to LG put more effort into the tasks with a mean score of 5.74 ± 1.05 followed by CG 5.53 ± 1.11, KLG 5.49 ± 1.11, and KG 5.21 ± 0.84. However, as no significance could be found, these descriptive results might only indicate a trend.

MiniPXI There were no significant differences when analyzing the overall miniPXI as well as the functional and psychosocial constructs separately. Looking at single constructs of the miniPXI, the Kruskal-Wallis test revealed significant differences ($H(3, N = 78) = 10.52, p = 0.0146$) regarding the enjoyment construct. The pairwise analysis with Dunn's test reveals differences between LG and CG (p=0.006) as well as LG and KG (p=0.004).



(a) Results of the three IMI subscales enjoyment (top), competence (middle), and effort (bottom).

(b) Results of the functional and psychosocial construct of the miniPXI.

Figure 6.2: Results of the (a) IMI and (b) miniPXI. Own figure based on Allgaier et al. [10].

For this construct, the LG has a mean of 6.63 ± 0.5, followed by KLG 6.16 ± 0.83, CG 5.67 ± 1.28 and KG 5.67 ± 1.28. For the non-significant results (Figure 6.2b) regarding the functional constructs, the mean of the LG 5.66 ± 0.64 is slightly higher than the mean of the KLG 5.63 ± 0.69 and CG 5.6 ± 0.54. KG has a mean of 5.19 ± 0.92. For the psychosocial constructs, LG 5.65 ± 0.65 also performs best with smaller differences between the KLG 5.29 ± 0.94, CG 5.23 ± 0.76, and KG 5.07 ± 0.89.

SBST Using the Pearson coefficient, a positive correlation $r = 0.33$ between errors in VR and errors in the SBST with a p-value of $p = 0.005$ was found. Of the 71 participants, one was excluded as an outlier because of less than ten correct answers in the SBST.

VR For the questions asked in the VR application, no statistically significant differences were found. The descriptive results revealed that participants of the LG rated their desire to continue highest, directly followed by KLG (see Table 6.4). However, participants of the CG and KG also stated that they would like to continue. The LG also achieved the highest scores for the two follow-up questions Q2 and Q3. Here, the KG reached the second place. When asking whether they want to try out more levels, LG is slightly in front of KLG, but in both cases more levels are desired. These results are just a trend and have to be further evaluated because no significances exist.

Table 6.4: Mean and standard deviation for Q1-Q3. The color indicates the order of mean values, however, there are no significant differences. Own table based on Allgaier et al. [10].

Question	CG (n=19)	LG (n=19)	KG (n=21)	KLG (n=19)
Q1: Continue	6.12 ± 0.86	6.42 ± 1.43	6.05 ± 1.2	6.32 ± 1.16
Q2: Continue to improve	5.79 ± 0.92	6.47 ± 0.84	6 ± 1.38	5.84 ± 1.17
Q3: Continue for variations	6 ± 1.25	6.32 ± 1	6.24 ± 1.14	6.21 ± 1.08
Q5: More levels		6.42 ± 1.12		6.16 ± 0.83

With mean values ranging from 6.26 ± 1.05 for CG to 6.62 ± 0.59 for KG, the participants stated that it was important for them to solve the tasks correctly. Getting feedback regarding the correctly solved diagnoses (Q7) was also desired throughout all groups (CG: 5.74 ± 1.19 - KG: 6.33 ± 1.35).

With 41%, nearly half of the participants want to put their score on a leaderboard and compare themselves with others. Only 18% want to compare with others without entering their results. About half of the participants (53%) would like to save their results for themselves to improve and 18% do not care about their score. Furthermore, these results were analyzed considering the different player types (see Figure 6.3). To do so, the amount of participants who have a specific primary player type–no matter whether they have one or several primary types–and who would like to, for example, enter their results in a leaderboard was investigated. Because of this, the sum of the related numbers is larger than 100% and the sum of the primary types is larger than the total participant number. For philanthropists, socializers, achievers, and free spirits the largest proportion wants to save the result for themselves, whereas all players and disruptors want to enter their score. This corresponds to Marczewski's [192] suggestion that leaderboards are a game element for the HEXAD-type player. However, saving the results for personal improvement was desired by a majority of each HEXAD-type, except for disruptors where only one participant belonged to: philanthropists 54%, socializers 67%, achievers 62%, players 48%, free spirits 56% (refer to Figure 6.3).

For analyzing the error count, one outlier who had nine and ten errors in the last two tasks was removed. There were no significant differences regarding error count. The descriptive results in Table 6.5 show that the KLG had in total least errors, followed by KG. Most errors occurred in the CG.

Table 6.5: Amount of errors in the VR application. Color indicates the order of mean values, however, no significances exist. Own table based on Allgaier et al. [10].

Task	CG (n=19)	LG (n=19)	KG (n=20)	KLG (n=19)
1	0.63 ± 0.96	0.74 ± 1.05	0.20 ± 0.41	0.16 ± 0.37
2	0.26 ± 0.45	0.21 ± 0.54	0.35 ± 0.81	0.53 ± 0.70
3	1.16 ± 1.68	0.42 ± 1.22	1.05 ± 1.73	0.53 ± 1.17
4	1.11 ± 1.76	1.11 ± 1.70	0.80 ± 1.54	0.16 ± 0.50
Total	3.16 ± 2.52	2.47 ± 2.63	2.40 ± 3.41	1.37 ± 1.89

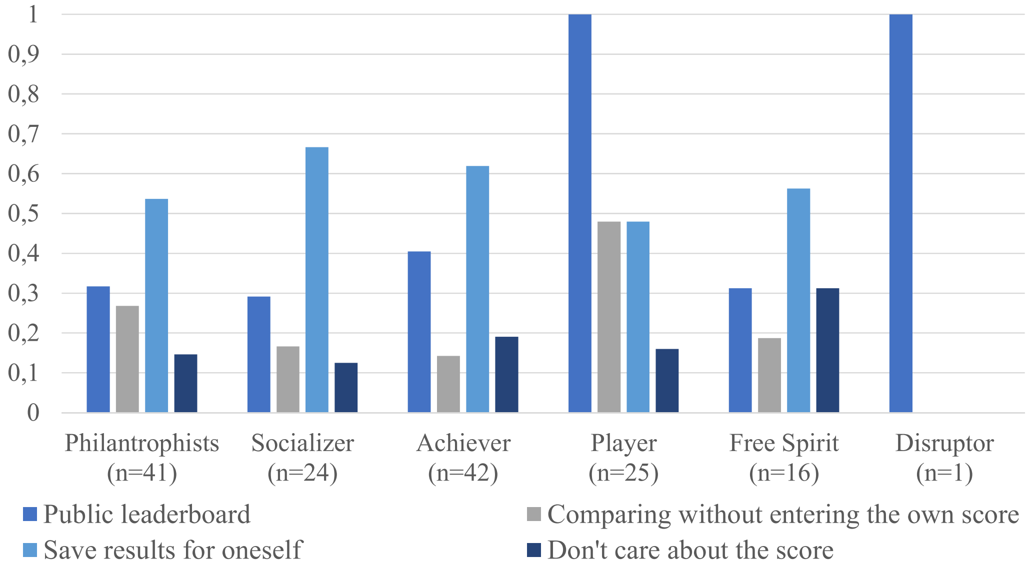


Figure 6.3: Proportion of participants, having the corresponding player type as primary type (multiple primary types per person are possible), who chose the answers of Q8. Own figure based on Allgaier et al. [10].

6.4.2 Medical Target Group Study

The broad audience study did not reveal significant differences between the groups. Because of the positive results for all groups, one assumption is that the participants are motivated and engaged by the VR and haptic technology, and thus, no significant differences are visible. This observation would be in line with Thamrongrat et al.'s [310] study where they compared gamified AR applications with a non-gamified AR application. In such cases, positive effects of gamification might be visible after a certain amount of time when the technology itself is not that novel anymore. Because of this, a second study was conducted using the within-subject design to directly compare the game elements. In this study, medical students participated (refer to Table 6.6). Similar to the other study participants, the most prominent player types are achiever and philanthropist.

Table 6.6: Characteristics of the medical students ($n = 10$). Table from Allgaier et al. [10] © 2024 ACM. The table was split into the two studies and two columns.

Characteristics	Value	Mean
Age [years]	[21-27]	
21-25	8	(80%)
26-30	2	(20%)
Gender		
Male	1	(10%)
Female	9	(90%)
Experience with VR		
None	6	(60%)
Less than 15 times	4	(40%)
More than 15 times	0	(0%)

Characteristics	Value	Mean
HEXAD player types		
Achiever	4	(40%)
Philanthropist	6	(60%)
Player	2	(20%)
Socializer	0	(0%)
Free Spirit	2	(20%)
Disruptor	0	(0%)

After an introduction including the medical relevance and explanation of the procedure, the first part compares the UI with the kit. To do so, they had to solve one sub-task of the first task with either the UI interaction (CG) or kit interaction (KG) and subsequently one sub-task with the other version. To avoid influencing them by continuing to show one version, they had to answer the following questions on paper and not in the VR environment. Although the IMI subscale enjoyment was employed, it was done in a non-validated way by directly comparing both versions. With this method, the aim was to avoid ceiling effects due to the technology, thus having more distinguishable results. Additionally, own questions were used to detect the version that is preferred for training and to assess different benefits, such as fun, motivation, or concentration. The questions are the following:

- Which version would you use for training?
- Which version is more fun?
- Which version is more motivating?
- With which version are you more focused?
- Which version is more suitable for learning this US?

Participants have to choose between the following answers: 'Clearly the user interface', 'Rather the user interface', 'Uncertain', 'Rather the kit', and 'Clearly the kit'.

In the second part, the game element level was investigated. According to the two-factor analysis, both versions–the control and kit– were presented with levels. After completing both levels with each of the interactions, they had to answer the following questions in a randomized order using a 7-point Likert scale:

L1 The level display stimulates my ambition.

L2 The level display helps to assess my performance.

L3 For training, I would like to have the level display.

L4 The additional text at the beginning and end of the level is motivating.

L5 I did not care about the level display.

L6 The level display is distracting.

L7 The level display and additional text put too much pressure on me.

For the orders of versions a balanced Latin square was used. The study procedure completed with a discussion to get insights into their preferences of game elements and reasons for their choice.

6.4.2.1 Results

There is no significant difference between the UI and the kit regarding enjoyment. The UI achieved a better enjoyment score two times, whereas the kit was rated higher six times. Two participants rated both the same. In general, the UI achieved an average enjoyment score of 6.07 ± 0.55 and the kit of 6.31 ± 0.47 of maximum 7 points, resulting in a positive result for both interactions. The only statement that was rated ≤ 4 in 6/10 answers for the kit as well as the UI is the statement 'While I was doing this activity, I was thinking about how much I enjoyed it'.

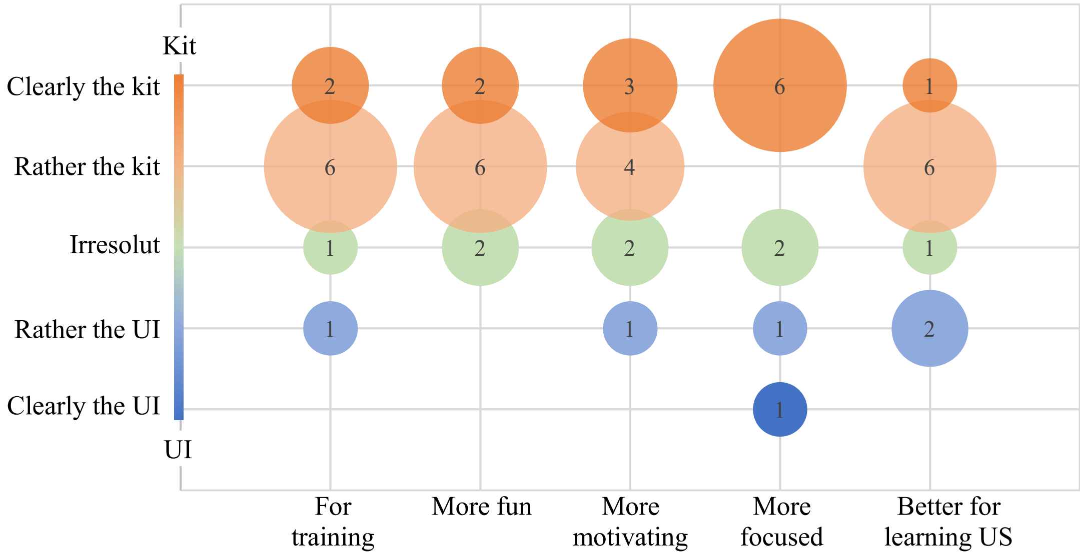


Figure 6.4: Amount of participants rating for the two interaction variants. Image from Allgaier et al. [10] © 2024 ACM. The image was split into two.

In the direct comparison, nearly all participants preferred the kit over the UI (refer to Figure 6.4). Only one participant stated they would use the UI for training. However, when finally asking which one they preferred and why, this participant commented that they first preferred the UI but that later on the kit was better because of the greater opportunity for interactions. The additional comments revealed that the participants liked the kit because it is more interactive. Some also mentioned that placing the lesions was more effective than just looking at the models in the UI, and that it stimulated three-dimensional thinking. Other comments were that they preferred the kit because it is more fun due to the direct consequence of interactions, and that considering where to place a lesion is better than predefined answers. According to the participants, one has to be more careful and concentrated with the kit. The follow-up questions, which interaction they preferred and why, were answered always in favor of the kit. Only one participant mentioned that the UI is more comfortable but not better than the kit. Regarding difficulty, there was a variety of opinions. Some participants mentioned that one has to concentrate and orient more using the kit, whereas others stated that the kit is more comparable to the scanned liver because of its orientation.

Regarding the levels, there are no significant differences between the UI and the kit. Because of this, Figure 6.5 only shows the results for levels when using the kit, which was the preferred interaction. In the discussion, eight participants emphasized that they would appreciate the levels in such a training application because they provide feedback regarding the progress and performance, make the application less monotonous, and are motivating. One participant did not perceive an increase in difficulty and one participant did not like the text when finishing one level.

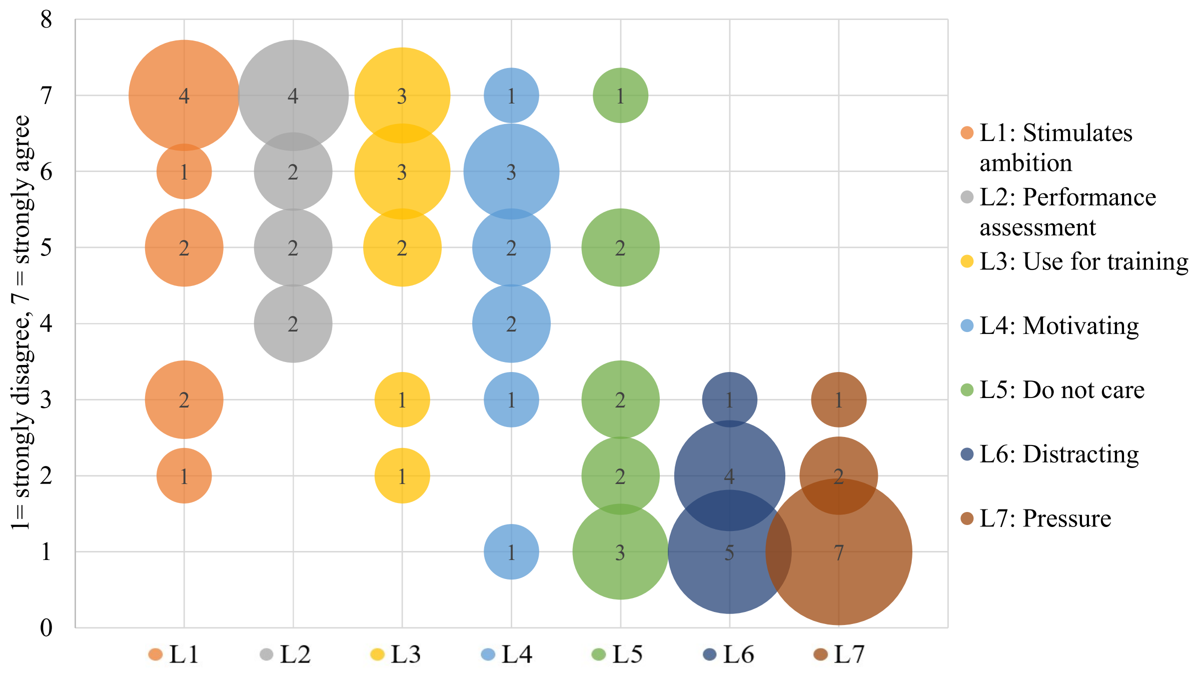


Figure 6.5: Amount of participants rating for the level questions using the kit. Image from Allgaier et al. [10] © 2024 ACM. The image was split into two and a legend was added.

6.4.3 Discussion

The presented comparison of the non-gamified VR application with gamified versions found significant differences regarding the *PXI* enjoyment construct. This is in accordance with the findings by Tang et al. [306]. The previously presented results show that the LG achieved significantly better scores. Although this is the only significant difference that could be found in the broad audience study, the LG also achieved slightly better results than the other groups regarding the functional and psychosocial constructs. This was also observed for three subscales of the IMI. The KG achieved the lowest scores. The descriptive results also indicate positive effects of the game elements regarding the desire to continue and the performance. However, they have to be confirmed in future studies.

The positive correlation between the SBST and the errors in the VR application indicates that the VR application reflects the visuospatial skills related to the understanding of cross-sectional imaging such as US.

In the medical target group study, no significant differences regarding enjoyment could be found. However, when directly asking which way of interacting they prefer, the kit was the clear favorite. Discussions with the participants revealed that the kit is preferred because of its interactive character and direct consequence, which coincides with the described intention of implementing the kit. Some also noticed that it stimulates three-dimensional thinking and concentration.

This study also shows that levels are appreciated and not distracting. Additional comments revealed that levels are good but the difficulties have to be well-chosen. More experienced participants regarding US did not recognize an increase in difficulty, and thus, levels might be confusing. However, they emphasized that with more difficult tasks, levels would be good for learning. Accordingly, levels are appreciated for training because they provide feedback regarding learning progress. Additionally, for some participants, levels increase motivation, ambition and fun.

To highlight some of the limitations of the presented investigation, Westera's [336] summary of several limitations of studies investigating gamification and serious games is used. Similar to the findings written in the contribution, they mention the lack of control groups. Accordingly, the presented comparison includes a control group that has a fully developed version including basics, such as feedback. The next limitation refers to the measurements and also applies to the used study design. The most meaningful to evaluate would be to measure the learning outcome, which was also discussed above. By asking what participants think, several biases might appear:

- If the researchers are part of the development team, confirmation bias might be induced.
- Voluntary participation leads to participants who might have positive expectations which might be reflected in the answers. This is called Hawthorne effect.
- Using own students might also add dependencies. This was not the case in the presented study.
- Introducing an innovative tool might lead to the novelty effect. This was already discussed after presenting the broad audience study. In this case, the assumption is not that the novelty effect leads to positive results of gamification, but that this effect leads to no significant results, because even the control group uses novel technology (haptics and VR).

For the medical target group study, another limitation is the participant group. First, ten participants are a very limited amount and second, nine out of ten participants were female, which might be an influencing factor concerning game preferences. Nevertheless, the comments revealed that most participants stated similar reasons why they prefer the kit and levels. This consistency might indicate that there is a general benefit of these game elements in this scenario and that the results do not just reflect personal preferences.

Although attempts were made to ensure that the difficulty of the tasks is similar for all groups, there might be differences between the UI selection and kit. The more advanced interaction could also lead to a higher mental demand–especially for novice VR users–as well as physical demand, which might negatively influence the miniPXI and IMI. However, comments during the second study revealed that the kit was perceived as easier, especially for the second task where the positions mattered. The recorded errors did not show that one of the groups is easier than the others.

In general, the reactions and comments of the medical students revealed that they liked the application and that they would appreciate such a training possibility in their curriculum. This highlights the need for additional US training but also the openness of prospective physicians to such a training modality. As these students were not specialized in liver surgery, the comments reveal that a VR-based US training in general is appreciated. The selection of game elements was based on the specific use case, however, the game elements are not specialized to this and can also be used in a similar VR-based training using different organs or medical scenarios.

6.4.3.1 Future Work

Motivation is affected by various aspects and perceived task difficulty might be one of these. Therefore, it could be interesting to examine how motivating the game elements are if people are struggling with the current level, or, on the other hand, how they may motivate the repetition of a mastered level for more practice.

To assess whether the higher motivation when using levels is a long-lasting effect, a longitudinal study has to be conducted. Mazarakis [197] also emphasized that joy and usefulness of gamification might decrease over time. After a longer training period it could also be evaluated whether the game elements lead to an increased learning outcome. This was assessed by Larsen et al. [173], however, they found no differences between the gamified and non-gamified version. A longitudinal study might also eliminate the likely positive influence of the new technology. As the question regarding leaderboard achieved positive results, the next prototype could include this as an optional game element. For such a study, the target group should be used, and it is recommended to include it in their daily routine in order to investigate the application and effects of game elements under real circumstances. While doing so, it would also be better to make the training voluntary and investigate the duration and repetitions instead of, for example, asking whether they want to proceed. Such a study would also be better in terms of the above-mentioned biases.

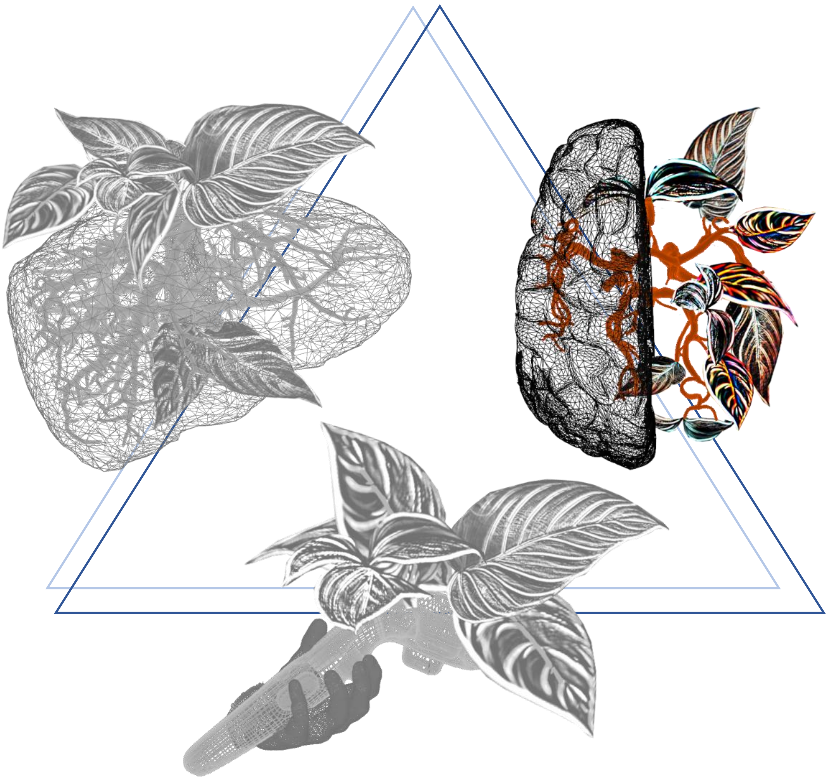
When performing a longitudinal study, possible misuse of gamification could be assessed. Gamification might also promote unintended and undesired behavior, addiction or undesired competition that influence the training experience negatively [103]. Thus, the gamified system has to be investigated to ensure that the training goal is still in the focus and users are not distracted from it by gamification.

6.5 Conclusion

Although gamification is widespread, there is a lack of thoughtful design decisions and evaluations in the field of medical VR applications. Using a VR-based training task for visuospatial skills in IOUS for liver, the appropriateness of several game elements was analyzed and discussed. Two studies compared *levels* and a *kit* for interacting, as well as their combination with a control group. While the broad audience study found significantly positive results for the levels regarding enjoyment, the qualitative feedback of the medical target group study revealed that the kit is the more desired way of interacting due to its more interactive character. Levels are appreciated for training, because they serve as performance assessment and progress feedback. These insights can be used for further studies and can be transferred to similar training scenarios.

Part II

VR-based Strategic Knowledge Training in Neurosurgery



7

Related Work on Computer-Assisted Neurosurgery

Synopsis Due to a blurred border between planning and training systems, the research focus and challenges concerning the literature research are highlighted first. The presented literature is then divided according to the intraoperative steps of the chosen intervention: craniotomy and clipping.

7.1 Research Focus

This chapter presents relevant literature regarding computer-assisted neurosurgery. The focus is on training systems, however, due to the mentioned blurry border, planning systems are also included. There are only few training systems in this specific area and in some publications both terms–training and planning–are used. Systems for intraoperative support which would also imply challenges like registration or physics of deformation because of brain shift [292] are not included in the following.

The term *Virtual Reality* is used very inconsistently (refer to Section 2.2.1). Approaches are often named VR systems, however, they are not immersive but use stereoscopic displays. This is even more misleading when the system and its technical details are not described properly. Because of this and the lack of immersive VR-based systems in this area, the following research includes non-immersive virtual simulations, too.

The following sections are divided according to the steps of a skull surgery: First, the patient has to be prepared, including a proper head position, and an appropriate access has to be chosen. The opening of the skull, which is called *craniotomy*, and the opening and dissection of underlying structures are subsequent steps. The next part focuses on systems for the surgical treatment of IAs via clipping.

7.2 Access and Craniotomy Training

There are few approaches focusing on the first steps of a skull surgery. Raabe et al. [248] point out the importance of careful positioning which is the basis for the subsequent surgery. A proper head position can reduce complications, facilitate the access according to the planned trajectory and provide a comfortable position for the surgeon.

A non-immersive simulation was already introduced in 2010 [271]. However, this simulation is restricted to drilling the burr holes. Therefore, a drill handle is attached to a haptic device to interact with the virtual patient shown on the monoscopic display. Their study revealed a positive transfer from training with the virtual model to drilling a physical model.

Neyazi et al. [217] proposed an immersive VR simulation, where the user is situated in a virtual OR. In this environment, they provide different training situations: A healthy CoW and five artificially modified CoW with different aneurysms at different locations. Based on the chosen situation, the user has to rotate the patient's head along two axes via sliders in the UI. The size and location of a circular craniotomy hole can be changed to properly reveal the underlying vascular structures. During the evaluation with the *Think-Aloud method* and one experienced neurosurgeon the immersive experience was highly appreciated. The expert emphasized the different craniotomy locations and sizes based on the underlying aneurysm and that it is advantageous to explore these situations in a virtual environment. However, the craniotomy shape should not be limited to a round shape but arbitrary shapes should be possible. Furthermore, patient-specific data and other anatomical structures, especially important landmarks, are relevant.

Starting as research projects from Poston and Serra [240] in the 1990s, the now so-called *Dextroscope*[1] is a widespread commercial VR pioneering system for neurosurgical planning [158]. This system is semi-immersive, meaning that the users' hands are immersed in the patient data. Variations of the Dextroscope are the *Dextrobeam* using a large stereoscopic screen for a group of people, the *DextroVision* which can be used on the users' desktop, and the *DEX-Ray* which is an intraoperative extension by providing an AR navigation system [158]. Stadie et al. [295] used the Dextroscope system for several years to plan neurosurgical interventions related to tumors, traumatic injuries, blood vessels and the spine. With the system, they could simulate different intraoperative head positions and thereby explore different surgical corridors. Consequently, in most cases they simulated the position of the craniotomy. In their later work [296], they compared the VR planning system with intraoperative image-guided navigation for craniotomy localization. The size and position of the craniotomy can be identified and measured accurately using the VR system. The usage of the Dextroscope for craniotomy and aneurysm clipping was also described by Wong et al. [338]. The patient's head can be positioned according to selected angles and a hole can be drilled. A clip can be positioned to define a proper clip angle, but the actual clipping is not implemented. Thus, there is neither a haptic feedback nor does the aneurysm deform.

Besides the Dextroscope, there is the commercial, semi-immersive system *Immersive Touch*[2]. In this system, the user can draw a craniotomy outline and place a clip [4]. Aneurysm exposure with opening the Dura and Sylvian Fissure was not integrated. This is similar to the semi-immersive system, also using the *Geomagic Touch* device, which was developed at *RISC Software*[3] [96].

Instead of a stereoscopic display, Munawar et al. [215] use a VR HMD as a microscope in a virtual skull base surgery simulation. Their system includes various burr types and incorporates haptic as well as auditory feedback when drilling.

Shono et al. [283] also included further steps after cutting the bone. In the dissection phase, the user can dissect the arachnoid membrane and trabecula, and the brain can be retracted.

Apart from virtual simulations, there are also hybrid ones combining physical and virtual simulations. Vite et al.'s [309] simulation combines a urethane skull model with VR. First, the physical skull model has to be fixated with a Mayfield head clamp. Because it is a physical skull model, the craniotomy is performed only once. The cutting process was supervised by an expert. The alignment of the virtual space with the physical space is achieved through the registration of landmarks. Therefore, the user has to mark the craniotomy contour with the haptic device *Geomagic Touch*, which is then projected on the virtual skull to create the corresponding virtual craniotomy hole.

Other virtual systems use AR to directly visualize important structures on the patient's skull to assist during the actual procedure [339]. Accordingly, the surgeon can see which paths include less important structures. As the evaluation focused on tracking accuracy and different color maps, one physical skull that did not correspond to the virtual data was used and head movements were not considered.

[1] Volume Interactions, Ltd.: no commercial operations anymore
[2] Immersive Touch, Chicago, USA: https://www.immersivetouch.com/ Last access: 08.04.2024
[3] RISC Software GmbH, Austria: https://www.risc-software.at/ Last access: 08.04.2024

7.3 Clipping Training for Intracranial Aneurysms

In contrast to systems for craniotomy training, there are several systems for aneurysm clipping. Alaraj et al. [4] assessed the clipping procedure of the commercial *Immersive Touch* system with 17 experts. For the study, a semi-immersive stereoscopic monitor-mirror system and a haptic stylus, the Geomagic Touch, were used. The majority of participants think that this system is helpful concerning anatomy, education, preparing for surgery and finding an appropriate surgical approach. Due to an unfamiliar depth perception, nine participants had difficulties grasping and interacting using the clip. The haptic feedback was rated as realistic by only 12 % of the participants.

In contrast to the *Immersive Touch*, the system developed by *RISC Software* uses real medical instruments, such as a forceps, applied to the haptic device [83]. Additionally, they included a blood flow simulation to calculate the residual aneurysm filling and degree of major branch stenosis, which is then used to quantitatively assess the clipping result. Other criteria to assess the clipping are the time needed, the number of clips and the frequency of repositioning clips. Gmeiner et al. [96] evaluated this system with 18 experts and obtained similar results as the previously presented study. The results indicated that for anatomical understanding and education the *Immersive Touch* is highly appreciated. However, only one third of the experts found that the haptic interaction was truly satisfactory and adequate.

The above-mentioned Dextroscope was also used in a study proposed by Kockro et al [157]. Three neurosurgical departments employed the Dextroscope with either a monitor-mirror system or a stereoscopic display. It was used for intervention planning by placing already closed clips at the aneurysm. The system lacked clip application by closing and opening the clip as well as vessel deformation.

In the clipping phase of Shono et al.'s [283] simulation, applying a clip leads to vessel deformation. For the deformation, they assigned rigs to the brain, artery, and vein models. The deformation of the parent arteries and aneurysms was realized using the *PhysX engine NVIDIA*. Similar to the previous approaches, they use a stereoscopic display. Interacting with the virtual clip is realized by a 3D-printed forceps and the motion capture device *Leap Motion®*[4]. Additionally, they included common sounds of an OR to increase the sensation. One limitation is that every step of the simulation was carried out by one neurosurgeon who is also the first author of this work. The same person also recorded the preoperative simulation and did the postoperative analysis.

Another work focusing on biomechanical modeling is presented by Vite et al. [308]. They used the SOFA (Simulation Open Framework Architecture [81]) to simulate a deformable brain and aneurysms. The evaluation includes only one expert neurosurgeon who confirmed the realism of the deformations.

Steineke et al. [299] compared preoperative traditional planning based on image data and 3D image reconstructions with preoperative planning in *The SuRgical Planning*[5]

[4]Ultraleap, San Francisco, USA: https://www.ultraleap.com/ Last access: 08.04.2024
[5]Surgical Theater, Inc., Ohio, USA: https://surgicaltheater.com/ Last access: 09.04.2024

platform. The latter includes virtual drilling for assessing the craniotomy location and size as well as the virtual selection and placement of clips. The procedure time after planning with VR was significantly lower, namely 246 min compared to 328 min on average.

In the hybrid system proposed by Theodoro-Vite et al. [309], a VR headset mimics the microscope and therefore was positioned at a fixed place. Using the Geomagic Touch device, the user can interact with virtual instruments.

The desktop-based system introduced by Allgaier et al. [11] aims at providing an easily available training that does not require special technical devices such as VR HMDs, stereoscopic displays or haptic devices. Accordingly, no haptic feedback is provided. To compensate for the lack of depth cues, visualizations are included. In this simulation, the user first has to define the craniotomy hole and then moves a selected clip to the aneurysm. When closing the clip, the vessels deform which is highlighted by additional visualizations. Distance visualizations include a color map on the vessel surface, a cylinder showing the shortest distance between the clip and vessels and multiple semi-transparent rays displaying all distances shorter than a threshold. The visualizations for the vessel deformation show the displacement magnitude and direction of the vertices of the surface mesh of the vessels. Here, a color map, rays, and two types of glyphs (arrows and drops) are compared via an online survey with eleven participants with medical background. The color map was the favorite visualization concerning criteria such as the information it conveys, occlusion and being intuitive. The distance visualizations were compared by one neurosurgeon who rated the semi-transparent rays best.

Instead of focusing on planning the surgical procedure, it is also possible to concentrate on the clip selection. Schwandt et al. [276] used the previously mentioned Dextroscope to determine a proper clip from a clip collection. Closing the clip or tissue deformation are not included, however, their results indicate that the presurgical clip selection based on the system could be transferred to the OR.

Another method to use VR for training is by providing 360° videos of real surgeries [36]. Without any interaction possibilities, the users could watch the recorded surgeries. Study results show that this approach is perceived as a good complement to neurosurgical training. In addition to the presented (semi-)immersive training systems, there are also educational systems focusing on anatomy, the understanding of complex 3D structures, and case analysis [22, 279].

The conclusions and requirements drawn from this research are presented in Section 8.2.

8

VR-based Training of Microsurgical Intervention for Intracranial Aneurysms

Synopsis This chapter presents a training system for microsurgical clipping of IAs. Thereby, the focus is on proper access and understanding its importance, however, it also includes an extension for clipping. After presenting the contribution, the system design is explained and the implementation is described. The chapter is completed with a conclusion.

This chapter is partially reused with permission of Springer Nature and has been published under a CC BY 4.0 license in:

Mareen Allgaier, Amir Amini, Belal Neyazi, I. Erol Sandalcioglu, Bernhard Preim and Sylvia Saalfeld. "VR-based Training of Craniotomy for Intracranial Aneurysm Surgery", *Int J Comput Assist Radiol Surg* 17, pp. 449–456, 2022.

As well as under permission of De Gruyter and has been published under a CC BY 4.0 license in:

Mareen Allgaier, Belal Neyazi, I. Erol Sandalcioglu, Bernhard Preim and Sylvia Saalfeld. "Immersive VR training system for clipping intracranial aneurysms", *Curr Dir Biomed Eng* 8(1), pp. 9-12, 2022.

8.1 Contribution

IAs can be treated minimally-invasively or via micro-surgical clipping (refer to Section 2.1.3). Since the last decades, more and more IAs have been treated minimally-invasively, as this method has several advantages like a reduced operation time [177]. Despite this shift of treatment, there are cases that have to be treated surgically due to complex circumstances, for example, aneurysms located at the middle cerebral artery (MCA) often have to be clipped [169]. The decreasing number of clipping procedures has the negative effect that surgeons and trainees gain less practical experience.

Due to the combination of little practical experience and complex cases, a strong need for improved training possibilities arises. During the last years, several training methods including physical simulations [193], virtual applications [4, 96, 283, 308] and hybrid methods [309], were presented for IA clipping. In this project, a VR training system that has the advantage that surgeons and trainees can easily explore different approaches without destroying models and with the possibility to undo steps was implemented. Consequently, this method is less resource-intensive than cadavers or physical simulations. Concerning virtual simulations, VR is a common method to provide realistic and immersive training in surgery and interventional radiology. Furthermore, virtual applications provide better exploration possibilities, such as scaling, rotation or semi-transparent rendering. They also benefit from the possibility of saving intermediate steps and to visualize additional information. Another virtual training possibility would be to use AR. In AR, the main benefit is the combination of real models with virtual models, such as displaying a virtual aneurysm in a real skull. Because this would also require more resources as the real skull model would be destroyed, a VR-based system was preferred. Thus, a VR simulation using an HMD was developed to exploit the benefits of immersion and to answer to following research question:

RQ 3: To what extent can VR provide an additional training possibility for microsurgical procedures in case of aneurysms?

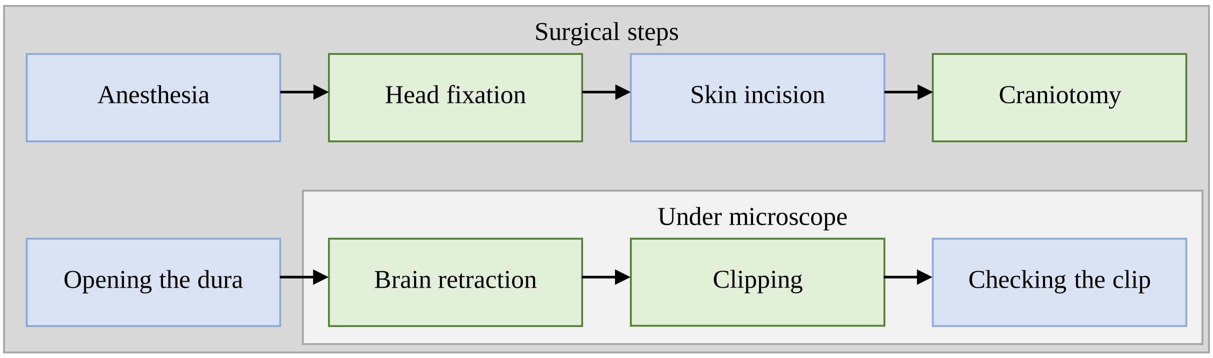


Figure 8.1: Main steps of surgical clipping. Steps highlighted in green are included in the proposed VR training simulation. Own figure based on Allgaier et al. [7] and Allgaier et al. [12].

To recap the surgical procedure, Figure 8.1 summarizes the main steps, whereby green indicates the steps included in the proposed simulation. This is just an approximation of the procedure, whereas more detailed steps are not mentioned due to the complexity of the surgery. As presented in Chapter 7, many existing aneurysm simulations focus on the actual clipping procedure. Only few include previous steps like opening the fissure, performing a craniotomy and the head alignment. However, these steps are the most important ones, as they enable easy and proper access and thus facilitate further steps [248]. The correct position can reduce bleeding, but also provides the most relaxing position for the surgeon. Thus, there are many factors influencing the positioning of the head: planned surgical trajectory, position of the surgeon, gravity retraction or drainage as well as measures for avoiding potential position-related complications like air embolism [248]. The focus of the proposed training system is on the most important factor: the planned surgical trajectory [248]. Based on a given IA, surgeons have to decide on an approach and an appropriate craniotomy. This procedure and the correlations, and thus strategic knowledge regarding access, can be trained with such a simulation. The skin incision and opening of the dura are also necessary steps, however, they do not influence access planning. According to the medical cooperation partners, they try to make the incision behind the hairline due to cosmetical reasons, which is usually possible without refraining from the desired craniotomy. The dura can be opened at any craniotomy hole. Thus, these steps were excluded as they do not directly contribute to the learning objective of getting an understanding of how important the correct craniotomy is to have proper access to a lesion. In a follow-up project, the clipping process with additional problem-specific visualizations to complete the surgical procedure was included. With this, the user can get an impression of whether the craniotomy is sufficient.

8.2 Design

Because the relevant literature has already been presented, the first section summarizes the main findings. Afterwards, the requirements of the simulation are set up.

8.2.1 Related Work

The relevant literature is presented in detail in Chapter 7. In summary, it can be derived that:

- Most systems are semi-immersive using a stereoscopic display or use VR HMDs in a static position simulating the microscope.
- Most systems focus on the clipping procedure and clip placement rather than the access and craniotomy.
- With haptic devices and a proper clip placement, fine-motor skills are addressed.

8.2.2 Requirements

To develop an effective training and planning simulation, requirements were set up during multiple meetings with a senior neurosurgeon and two novice neurosurgeons. An initial meeting served as a basis to gain knowledge about the general procedure and important steps. During the development, several meetings served to optimize the virtual surrounding, the interactions and the models. Based on these discussions, the following requirements were extracted:

*R*1 Head positioning: A head inclination along two axes via intuitive hand rotation should be available. Longitudinal inclination should be possible up to 90 degrees and transversal inclination up to 30 degrees.

*R*2 Craniotomy: The user should be able to draw an individual shape.

*R*3 IAs suitable for clipping should be included.

*R*4 For training purposes and improved exploration of vessels and IA, the user should be able to adjust the transparency of the skull and brain.

*R*5 Simplification of the anatomical models is feasible, but the following important landmarks should be represented: (a) optical nerve: It serves as an important reference to all other surrounding structures and should be identified early in the exposure [204]. (b) pterion: As this region is a relevant cranial landmark, the sutures of the skull are displayed [248].

*R*6 The virtual environment should be realistic enough to support immersion and not distract the user.

*R*7 Derived from *R*6, it is also important that the user's posture and position in relation to the patient is similar to those during a real surgery. This mainly affects the transition between the craniotomy and microscopic view.

8.3 Development

This section first describes the technical setup. Afterwards, the necessary models, such as anatomical structures and medical instruments, are introduced. With this as basis, the workflow of the training simulation is presented.

8.3.1 Simulation Setup

The focus of the simulation is on head positioning and craniotomy because these aspects strongly drive the following surgery. In contrast to the previously-mentioned simulations that include these steps as well, the presented prototype does not employ haptic devices and a stationary stereoscopic display. Instead, it uses a VR HMD and a virtual OR to create a more immersive training experience. Consequently, the simulation is not bound to a fixed workstation but can be used wherever free space of about 2×2 meters, an appropriate PC or laptop and a VR HMD are available. This is especially advantageous for usage in hospitals.

During the development, an Oculus Quest and the corresponding controllers were used, but due to the XR Interaction Toolkit, transferability to other VR glasses is possible. In a previous work [217] an adapted virtual OR similar to Huber et al. [129] was used, which

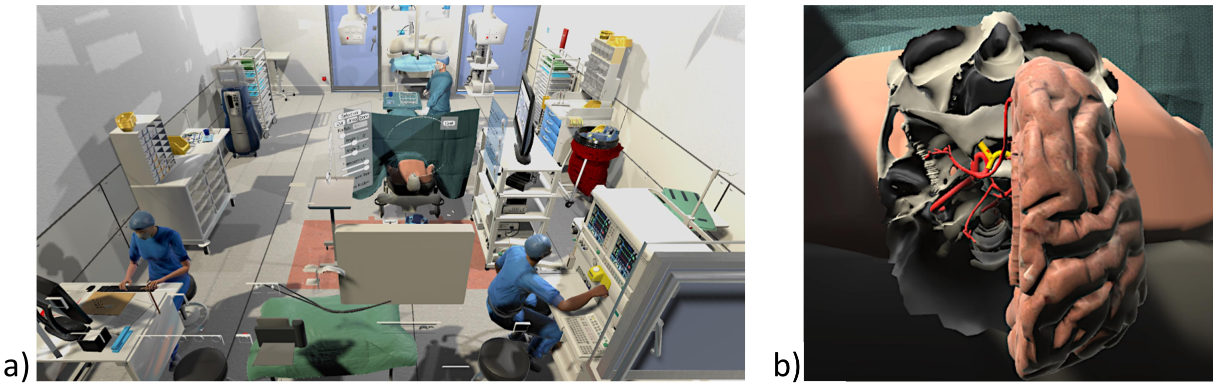


Figure 8.2: a) virtual OR b) lower part of the skull, right brain half, CoW (red) and optic nerve (yellow). Image from Allgaier et al. [7] and available under a CC BY 4.0 license. No changes were made.

was expanded in this work. As Figure 8.2 a) shows, the patient was replaced by an upper body without a head, and a segmented and processed spine and skull, including brain and vessels.

As the craniotomy training and clipping extension were evaluated separately, the input device changed for the extension based on the feedback received for the craniotomy training. The *VR Ink* was chosen because this device is more similar to the medical instrument, which should be held like a pen and not held in power-grip [248].

8.3.2 Models

One relevant craniotomy with respect to MCA aneurysms is the pterional craniotomy which is used for the transsylvian approach [100]. The pterion is defined as the anatomical region near the temple where the four bones frontal, parietal, temporal and sphenoid meet. For a pterional craniotomy, the patient is positioned supine, which means that the body and legs lie straight. Afterwards, the head is positioned and fixated [233]. Additionally to the pterional craniotomy, other approaches, such as the lateral supraorbital approach and the minipterional approach, can be used. The described procedure of the specific case served as a basis for the presented simulation. Nevertheless, further IA locations and approaches can be trained. For the simulation, the required models are a skull, a CoW, aneurysms and clips.

A complete CoW, extracted from a healthy person's MRI data, serves as a basis for all IA models. The surface model was segmented with a customized workflow [262]. To meet $R3$, the MCA aneurysms from a previous project were used [11]. Furthermore, the user can also set a small sphere at a desired location at the CoW to approximate a target structure. This enables an easy way to train different situations of various locations of target structures. As mentioned, the predefined MCA bifurcation IAs are based on the same CoW, thus located at the same position not just on the CoW but also in relation to other anatomical structures. Besides the location at the CoW, the characteristics of the specific CoW also matter. Since the length of the M1 segment is an important criterion for the decision about the surgical strategy in MCA aneurysms [79], different CoW configurations for the individual target placement were provided.

The skull is a segmentation of a patient's CT angiography data that was further processed to reduce artifacts and to simplify the geometry. For simplicity, most of the inner parts of the skull were removed. Additionally, the sutures of the skull were added. As the optic nerve is an important landmark that should be identified after opening the Sylvian Fissure [233], a short fragment of it protruding from the optic canal of the sphenoid bone was modeled. With the optic nerve and the previously described skull model, R5 is met. Furthermore, the sphenoid bone was included to complete this region. Regarding the brain, a free model[1] was adapted to fit into the skull. The models can be seen in Figure 8.2 b).

The extension also comprises the clips used in a previous work [11]. All clips are based on L-aneurysm clips from the clip company Peter Lazic[2]. From these, the cooperation partners from the neurosurgical department selected clips that are similar to frequently used ones at the University Hospital Magdeburg. Consequently, during the previous project 18 clip models were modeled using the product catalogue of the company [6].

8.3.3 Workflow

After selecting an M1 segment configuration and placing the target structure at the desired position at the CoW, the user is situated in the virtual OR.

Head positioning The first step of the workflow is the head positioning (Figure 8.3a). There are two possibilities to rotate the patient's head. Either via two sliders in the menu or via hand rotation, which is the more intuitive and realistic way (recall $R1$). Therefore, the user has to point at the head and hold the grab button while rotating the hand in the desired direction.

Craniotomy By drawing directly on the skull, the users define the contour, size and location of the area they want to remove (recall $R2$). To draw a closed contour, they can adjust the line thickness, or radius of the 'pen'. After defining a seed point in the closed contour, the area is filled by a region growing algorithm (Figure 8.3b, center). Exploring the structures and finding an appropriate location and size of the hole is supported by the adjustable transparency of the skull and brain (recall $R4$).

Microscopy After confirming the hole, the simulation switches to the microscopic view (Figure 8.3c). When using a microscope during surgery, the surgeon's viewing direction is straightforward, horizontal to the floor, while the hands are operating at the hole. Thus, they are not looking in the direction of their hands. To simulate this, and thereby realizing $R7$, a virtual screen is placed in front of the user and above the patient, displaying the microscopic view. Similar to a real surgery, the microscope camera can be adjusted and moved via an interface next to the display.

[1] Free 3D, uploaded by bejek_2812: https://free3d.com/3d-model/brain-18357.html Last access: 27.03.2024

[2] Peter Lazic GmbH Microsurgical Innovations, Germany: https://www.lazic.de/en/ Last access: 09.04.2024

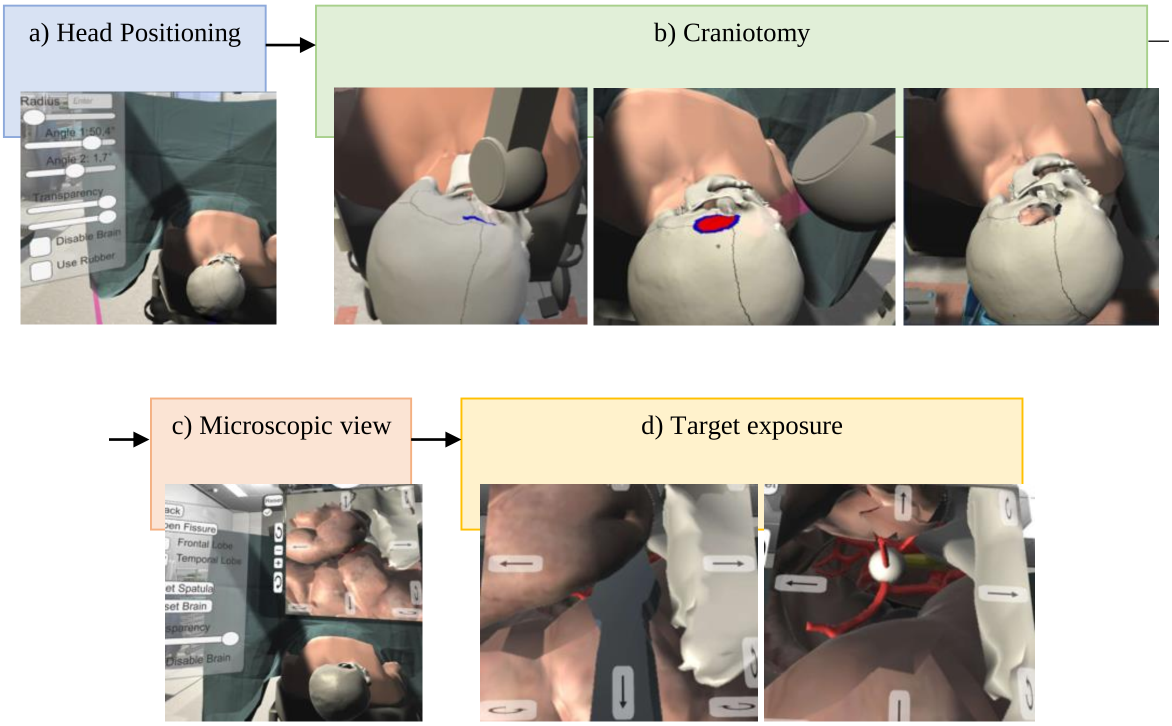


Figure 8.3: Overview of the craniotomy training. (a) Head positioning, (b) drawing the craniotomy hole, (c) microscopic view, (d) opening the Sylvian Fissure to expose the target structure (white sphere) and the optic nerve (yellow structure). Image from Allgaier et al. [7] and available under a CC BY 4.0 license. The layout was changed slightly and a figure in step b) was added.

Assessment To assess the position and size of the hole as well as the trajectory to the IA, the user can open the Sylvian Fissure (Figure 8.3d). As the procedure of opening the fissure was not the focus, it is simplified by the following procedure: Taking a spatula and placing it at the fissure. Via the interface, the frontal and temporal lobes can be deformed separately. To simulate deformations either the finite element method (FEM) or the mass-spring model (MSM) can be used. Despite various adaptions, FEM still has the limitation of high computational costs [179]. Because the VR setup requires real-time interactions, the less accurate but faster surface MSM is used. The implementation of an MSM used for vessel deformation when applying a clip is described in detail in previous works [6, 11]. In general, an MSM assumes that there are masses (in this case at the positions of the surface mesh's vertices) that are connected with springs. Each spring between mass i and j has a resting length l_{ij}^0 and a spring constant k_s describing its stiffness. By applying a force to a mass, the spring is stretched or compressed. Using Hooke's law, the following formula calculates the force vector f_{ij}^t acting on the attached masses i and j (with their positions x) at time step t:

$$f_{ij}^t = k_s * \frac{x_j - x_i}{|x_j - x_i|} * (|x_j - x_i| - l_{ij}^0) \tag{8.1}$$

In addition to the described springs, a spring from each mass to its initial position is used in the deployed model for volume preservation [6]. After calculating all forces acting on the masses, their new positions have to be determined using an integration method. The used MSM deploys the Verlet integration which is an explicit method, thus being fast at the cost of having stability problems and being prone to failure. This decision is also based on Mor's [211] comparison of integration methods. Similar to Halic et al. [107] the damping k_d was included in the integration step, resulting in the following equation to calculate the new mass position $x(t + \Delta t)$:

$$x(t + \Delta t) = x(t) + (x(t) - x(t - \Delta t)) * (1 - k_d) + f(t)/m * \Delta t * \Delta t \quad (8.2)$$

The described MSM was adopted for the use case of brain deformation by adjusting the parameters. Because the parameters are not directly related to material parameters of the tissues, they were set up experimentally resulting in the following: spring constant $k_s = 0.9$, spring constant backwards to the initial position $k_{sBack} = 0.3$, and damping constant $k_d = 0.8$. After the brain is retracted employing the described MSM, the user can assess their chosen approach via self-assessment.

Clipping In the extension, the user has to choose a clip from a menu where the clip properties such as opening angle and length can be seen. The chosen clip can then be applied to the aneurysm. Due to a lack of depth cues because of missing surrounding structures, the navigation is difficult. To facilitate the navigation, rays indicating distances between the clip and vessels are included [11] (see Fig. 8.4a). Once the clip is placed, closing is visualized by highlighting the area the clip would hit (see Fig. 8.4c). Thus, users can asses whether the clip is placed in the desired way and which parts of the vessels and aneurysm are affected. Finally, the clip can be closed leading to a deformation of the vessels. For this, an NVIDIA Flex softbody surface representation with empirically determined physical parameters (see Table 8.1) is applied to the vessels.

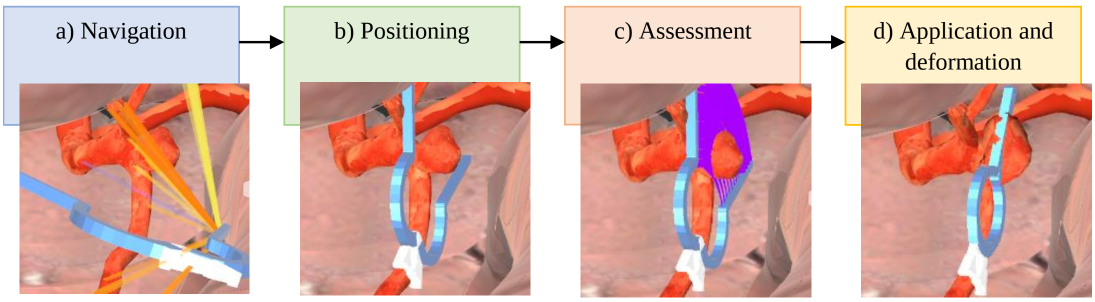


Figure 8.4: Overview of the clipping extension. (a) Clip navigation with supportive distance rays, (b) clip placement, (c) clip assessment, (d) clip closure and resulting deformation. Image from Allgaier et al. [12] and available under a CC BY 4.0 license. The layout was changed slightly and a figure in step a) was removed.

Table 8.1: Parameters used for NVIDIA Flex soft container for vessel deformation. Table from Allgaier et al. [12] and available under a CC BY 4.0 license. No changes were made.

Simulation Parameters	
Substep Count	9
Iteration Count	15
Gravity	(0,0,0)
Radius	0.002
Solid Rest	0.0005
Fluid Rest	0.0002

Common Parameters	
Static Friction	3
Dynamic Friction	2
Particle Friction	1
Max Speed	340282.3
Max Acceleration	340282.3
Damping	50
Collision Distance	0.0002

8.4 Impact

The two prototypes –the craniotomy training and clipping extension– were evaluated separately. This section is divided accordingly and presents the respective setup, results and discussion. Both evaluations were conducted with a small number of participants, however, they are neurosurgeons or medical students at the neurosurgical department, and thus, experts and potential users. Because this group of people is small (due to the specialization) and hard to reach (due to the clinical routine), a formative evaluation based on a method developed by Jakob Nielsen in 1989 was chosen. With *discount usability engineering*, he presented a concept that uses only up to five participants, focuses on qualitative studies and uses the think-aloud method [220, 221]. Although this method has limitations and is not as good as traditional, more expensive methods, Jakob Nielsen showed that it provides valuable insights by still identifying most problems.

8.4.1 Evaluation of the Craniotomy Training

This section presents the study setup, results and discussion of the prototype including the craniotomy.

8.4.1.1 Setup

The aim of this simulation was not to replace existing training modalities, but rather to provide an additional training possibility. Thus, the investigation should determine whether the craniotomy training can serve as additional training and whether the expected advantages of using VR HMDs were recognized and appreciated by the participants. To evaluate the mentioned aspects qualitatively and in detail, the think-aloud method was used. Thereby, thoughts and comments of the participant are recorded and in-depth questions can be asked. In addition to this, a small survey collected comparative opinions and an overall impression. Finally, the precision of drawing a contour with a controller was assessed.

The evaluation was conducted with four participants with different levels of experience, see Table 8.2 for relevant data about the participants. It is important to note that one participant used the simulation tool, and VR HMDs in general, for the first time, whereas the others had tried it before. The study was performed separately with each participant. After a short introduction to the tool, the participants were asked to use it on their own. Thereby, they should enter the IA selection, select a CoW and place the target sphere. Subsequently, they were asked to explore the interface and interaction possibilities and to perform a craniotomy. After they were satisfied with the hole, they were requested to set up an appropriate microscopic view to evaluate the craniotomy hole. As during the whole procedure the think-aloud method was used, they were asked to comment on their actions, highlight difficulties and state what they appreciate or what can be improved. After completing the VR procedure, they had to fill out a questionnaire concerning the previous task, which comprised three parts: Immersion, training and load.

Immersion This part focuses on whether the virtual OR was appreciated and why. Therefore, the following description was given: '*As a comparison you can imagine the same tool without the virtual OR, just using a head 'floating' in front of you*'. Given this task, they had a multiple-choice selection with the following items:

- It looks nice/appealing
- It is more fun to use the tool
- It provides a better training situation
- It is more realistic
- I am more concentrated
- I take the tool/training possibility more seriously
- I feel present in the virtual environment

As the immersion and presence in the virtual environment depend on the plausibility, the participants should also rate the plausibility of the environment, anatomical models and interactions, such as head rotation, drawing the craniotomy contour, adjusting the microscope and opening the fissure.

Training This part serves to get insights on whether the tool could and would be used for training purposes and in what way it would be helpful. The following questions had to be answered with a 5-point Likert scale:

- Would you use the tool in order to learn/train?
- Would you use the tool in order to teach?
- Do you think this simulation can improve novice surgeons'/students' anatomical understanding?
- Do you think this simulation can help with becoming aware of the importance of the right craniotomy hole/approach?
- Do you think you would feel better prepared for a real surgery/gain self-confidence when using the tool more often?

Table 8.2: Overview of relevant participant's data. Table from Allgaier et al. [7] and available under a CC BY 4.0 license. No changes were made.

Age	Gender	Level of Experience	Handedness
31-34	Male	Neurosurgeon: 1-5 years	Right
35-40	Male	Neurosurgeon: 11-15 years	Right
26-30	Male	Medical student	Left
21-25	Female	Medical student	Right

Load The last part of the questionnaire was the *NASA Task Load Index (TLX)* [113], which is commonly used to assess the load, effort and frustration when solving a task using a software.

In addition to the questions, the precision of drawing the craniotomy was assessed. To provide an appropriate training system, it should be precise enough to create realistic craniotomy contours, but also to not frustrate the user. To examine whether a user is able to draw an intended contour, the participants had the task of tracing four predefined contours. They are at two different locations with two different line thicknesses each. The templates were drawn by a person with an appropriate medical background and with mouse as input device.

8.4.1.2 Results

The rating of plausibility can be seen in Figure 8.5. According to this, the most plausible part is the head rotation whereas the microscope was rated as least plausible. For more detailed and qualitative insights, the following summarizes the main results obtained by the think-aloud method. They are sorted with respect to the workflow and requirements.

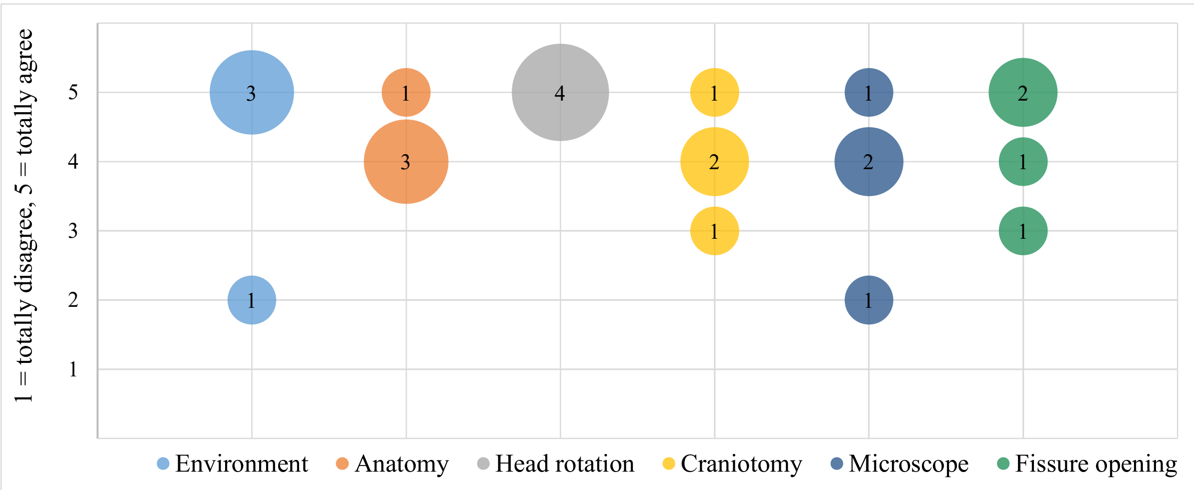


Figure 8.5: Participants' rating of the plausibility of the different aspects. The size of the circles and the number in circles denote the number of ratings. Own figure based on Allgaier et al. [7].

Regarding the *head positioning*, only the participant using the simulation for the first time had difficulties. However, after becoming familiar with the controller and its buttons, all participants rated it as appropriate. The *craniotomy* was well accepted, but there are two things that should be improved: When using a controller, the hand and arm position is different from the real position. A pen-like device would be more natural. Furthermore, only the outer part of the skull can be removed, but it should be possible to mill the sphenoid bone, too.

All participants liked the additional display for the *microscopic view* and the movement and rotation possibilities. Nevertheless, some suggestions for improvement were mentioned. The interaction with the microscope would be better via joystick or hand rotation instead of UI buttons. It would also be more realistic if the user has to switch between the two views via a button, because usually it is not possible to see the microscopic view and the patient just by changing the viewing direction.

Concerning the *target exposure*, the participants appreciated the opening of the brain depending on the placement of the spatula and that both brain lobes can be retracted separately. However, two spatulas, different sizes, and a Leyla retractor that can be moved via hand movement would be more natural.

Other important suggestions and feedback were:

- Including sounds to create an even more immersive atmosphere.
- Placing a sphere to simulate various IA locations was strongly appreciated.
- Feedback, for example, when rotating the head or retracting the brain too much, would be welcome. This could also be in terms of gamification to increase motivation, fun and ambition.
- Some given craniotomy contours could be integrated for learning purposes.

All suggested improvements are summarized and categorized in Table 8.3.

Table 8.3: Summary of essential improvements.

	Improvements
Interactions	Milling the sphenoid bone; pen-like input device; microscope adjustment via joystick or hand rotation; including a second spatula and retractor for brain retraction
Environment	Surrounding and microscopic view should not be visible at the same time; common OR sounds for more realism
Learning	Feedback (e.g., when retracting too much); gamification for more fun, motivation, and ambition; provide contours from experts to learn and compare

The questionnaire assessed whether and why the participants like the virtual environment in contrast to a simulation where only a three-dimensional head is available. The average value using a 5-point Likert scale (0 equals disagreement, 5 equals agreement) of whether they like the OR is 4.5. The reasons why they prefer an immersive OR are displayed in Figure 8.6. All participants agreed that it is a better training situation, more fun, and looks nice and appealing, whereas only two participants agreed that they felt present and took it more seriously because of the immersive environment. In Figure 8.7, the ratings of statements concerning the usage and possible benefits of the simulation are shown. The greatest approval was given to the statements that it helps to become aware of the importance of the craniotomy hole, followed by the statements that they would use the tool for teaching and that it can improve the anatomical understanding.

In the NASA TLX questionnaire, a 20-point Likert scale (zero equals very low, 20 equals very high) was used. The results can be summarized as follows. The participants were successful in accomplishing their task ($\varnothing = 16.5$, $\sigma = 2.06$) and did not have to work that

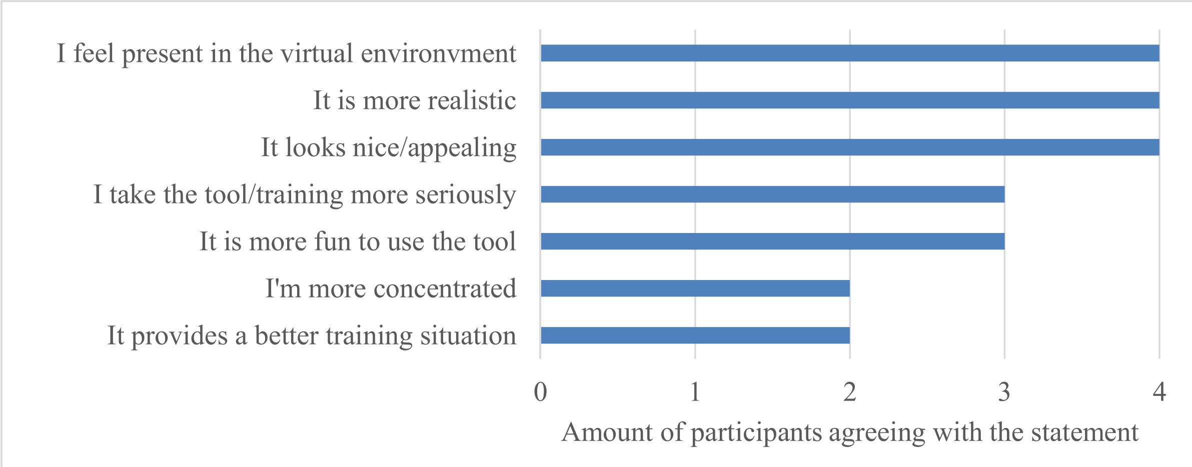


Figure 8.6: Rating of the four participants why they like the virtual OR in contrast to a simulation with just a head. Image from Allgaier et al. [7] and available under a CC BY 4.0 license. The layout has been changed slightly.

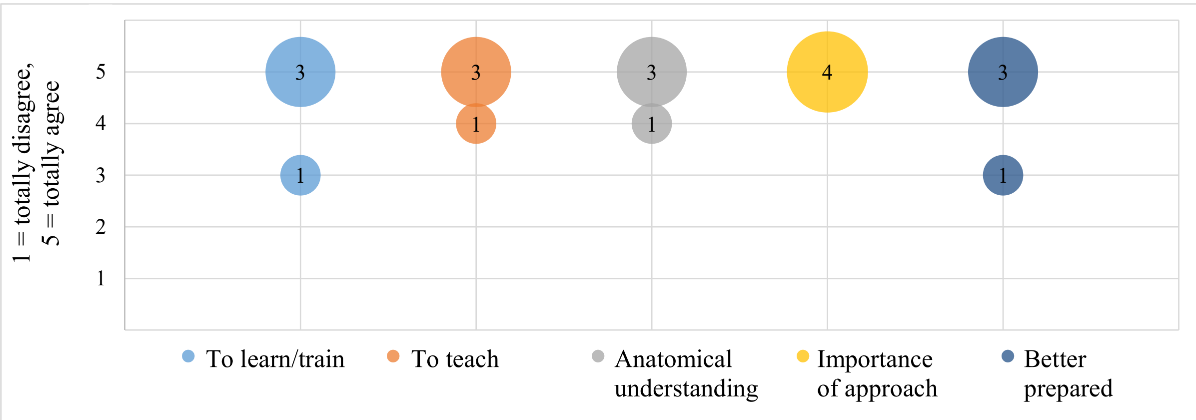


Figure 8.7: Rating of statements regarding the usage and benefits of the simulation. Own figure based on Allgaier et al. [7].

hard to accomplish their level of performance ($\varnothing = 8.25$, $\sigma = 4.323$). The mental load ($\varnothing = 11.5$, $\sigma = 4.33$) was rated higher than the physical demand ($\varnothing = 9$, $\sigma = 5.24$), but both values are in the middle range.

Regarding the craniotomy, the participant's contour was compared with the template based on the surface dice similarity coefficient (surface DSC) and the Hausdorff distance. The surface DSC indicates the overlap of two contours [222]:

$$\text{surfaceDSC(A, B)} = \frac{2 * TP}{2 * TP + FP + FN} \tag{8.3}$$

using true positive (TP), false positive (FP) and false negative (FN) vertices and the two contours A and B. Additionally, the Hausdorff distance H, indicating the proximity of two contours, was calculated. It is the largest distance of all minimum distances between two curves [132]:

$$\begin{aligned} H(A, B) &= max(h(A, B), h(B, A)) \\ h(A, B) &= max_{a \in A} min_{b \in B} ||a - b|| \end{aligned} \tag{8.4}$$

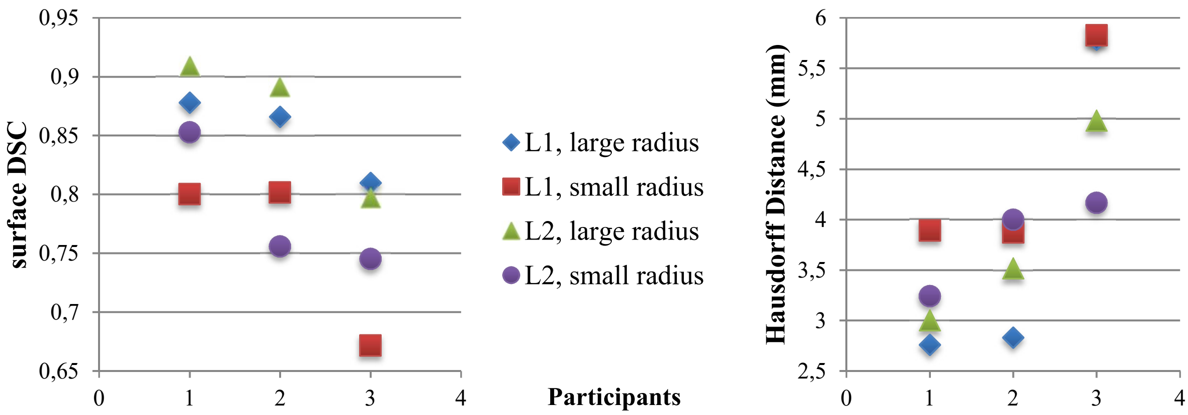


Figure 8.8: Results of the evaluation with surface DSC (left) and Hausdorff distance (in mm; right). L1 and L2 are the two different locations. Image from Allgaier et al. [7] and available under a CC BY 4.0 license. No changes were made.

The equation $h(A, B)$ calculates the maximum distance of all minimum distances between vertices of two contours. The average Hausdorff distance is $3.89\,mm$ and the average surface DSC is 0.81. The corresponding results can be seen in Figure 8.8. Due to technical problems, the contours of only three participants were available. It can be seen that for all results except for the Hausdorff distance of one participant, the results of the contours with large radius achieved better results than the ones with small radius. Furthermore, two participants achieved relatively similar results whereas the third participant had more difficulties with this task, resulting in a Hausdorff distance of almost $6\,mm$ for one of the craniotomy locations.

8.4.1.3 Discussion

During evaluation it could be observed that the participant using the simulation for the first time and having no previous experience with VR HMDs, had most difficulties. Nevertheless, even in this short time span progress regarding the interactions and drawing a smooth contour could be seen. Thus, using the simulation requires a certain degree of training.

The surface DSC and Hausdorff distance of drawing the contour revealed better results for the larger radius than the smaller radius. Regarding drawing in general, tracing a fine line is much more difficult than tracing a wider line. Unsteady hand control has a much larger influence when drawing a thin line with a pen with a small radius. The differences in the results concerning the radius indicate that the simulation can emulate this natural difference in difficulty. However, to interpret the specific numbers, a comparison with results from performing a given craniotomy with a real craniotome and physical skull would be necessary. Because the participants mentioned that drawing the contour has a suitable level of difficulty, the results can be rated as precise enough for the current learning objective but should be investigated in the future.

With the proposed prototype, it could be demonstrated that users appreciate the virtual OR and the resulting realism and immersion leading to an increased fun factor and high motivation. Consequently, this kind of simulation offers different advantages than physical simulations, cadaver training or non-immersive virtual training systems and is an appropriate and additional training possibility. These advantages can be further improved by implementing the feedback concerning interactions and the input device, resulting in a more realistic training experience. However, the presented simulation only considers the position of the lesion for access, but more factors influence the access, such as other intracranial structures along the surgical corridor.

Regarding the two learning objectives *learning the importance of the surgical approach* and *improving anatomical understanding*, the presented VR-based training benefits from being non-destructive. Thus, the users can freely explore and try different approaches with the possibility to undo and redo single steps.

Compared to the IA clipping training approaches presented in Chapter 7, the proposed approach benefits from the immersive environment, whereas the others benefit from the pen-like device and haptic feedback. Having an input device held in a realistic grip might be more important during the clipping procedure than during head positioning and craniotomy. During clipping, fine-motor skills are required whereas the proposed training addresses strategic knowledge concerning the initial steps of the surgery. Regardless of the input device, the importance of the right craniotomy placement is demonstrated.

Future Work One major aspect that is lacking in the presented protoype is feedback or assessment. To provide a proper training simulation, the simulation itself has to be plausible and include the necessary parts. If this is given, it is essential to provide the user with feedback to assess their performance and foster improvement. Chan et al. [48] present a review where they summarize performance metrics that are used to assess skills in various neurosurgical VR simulations, including:

- distance to target: The distance between the final position of a device and the intended target,
- time: Either for a whole procedure or single phases,
- tool movement: Instrument movements, such as velocity or acceleration,
- pressure and force: with a haptic system, the force applied to tissue can be measured, and
- virtual surgical outcome: Whether the virtual surgery was successful including, for example, blood loss or extent of resection.

In a follow-up discussion with a neurosurgeon, who was not part of this project's development and evaluation, suitable performance metrics that might be included in such a craniotomy training were enquired. Thereby, the subsequent three criteria arose:

- Size of the craniotomy: The user should receive feedback regarding the size of the craniotomy hole in relation to the size of the lesion. This applies to superficial lesions.
- Highest point: The craniotomy should be the highest point of the skull to prevent leakage.
- Central position: The craniotomy should be centered over the lesion.

These aspects and exceptions have to be discussed with more neurosurgeons to avoid bias. Once the aspects are confirmed by several experts, they have to be included in the application. Therefore, different ways to visualize the criteria and conveying the feedback can be investigated. One first, straightforward visualization to assess the craniotomy size and whether it is central can be seen in Figure 8.9.

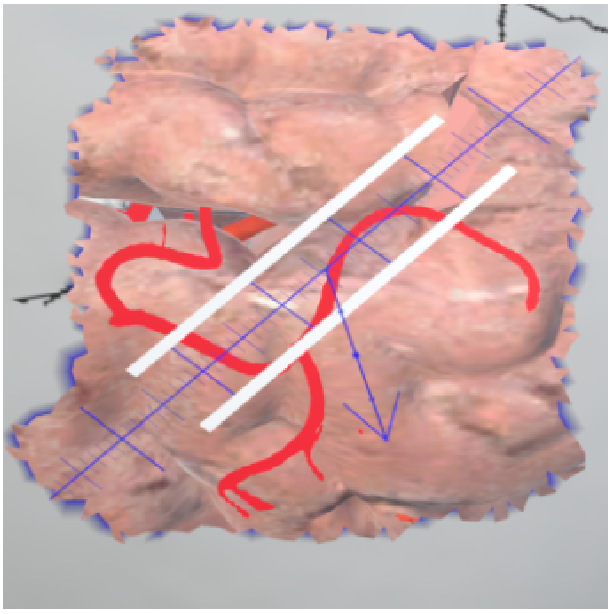


Figure 8.9: White lines show the size of the lesion projected to the craniotomy surface. The blue ruler supports size recognition and the blue arrow indicates the offset between the craniotomy center and center of the lesion. Own figure.

Such metrics could also be used to create a gamified training by including, for example, scores or achievements. After including an assessment, its validity has to be proven. According to the neurosurgeon, the given steps could also be extended by including the adjustment of the microscope (whether it has to be bent strongly) or applying the Mayfield clamp to fixate the head (the connections always should be parallel to the ground and proper application of the pins). Further steps, for example, skin incision can also be included.

Despite the positive trend indicated by the results, several improvements were mentioned. Because the microscopic view was rated as least realistic, this should be improved first. This includes the proposed hand interaction instead of UI interaction. Furthermore, the view should be more separated from the OR view. This could be done by hiding the surrounding when using the microscope, however, concerns arise with this idea: A static view even when (slightly) rotating the head might lead to cybersickness. Because of this, it can be considered to blur the surrounding instead of completely hiding it or to hide the microscopic view whenever the user moves their head beyond a certain threshold.

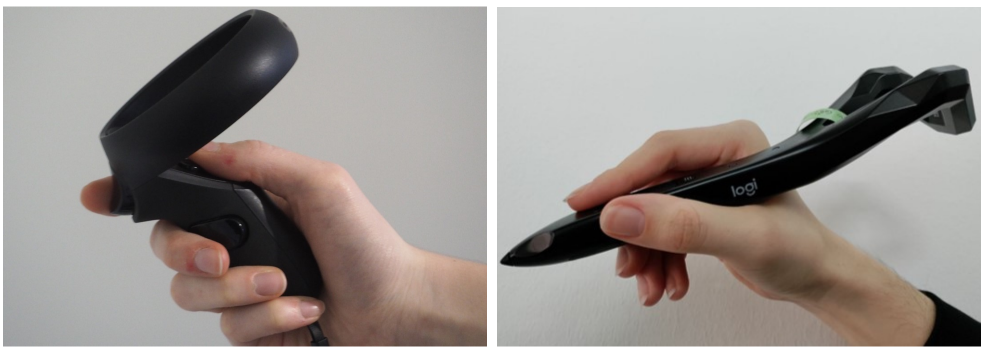
After including the mentioned improvements, a more meaningful study concerning the simulation's validity is necessary. After refining the simulation based on the given feedback, a larger study assessing the learning outcome is required. Furthermore, the transfer of the learned to a real surgery should be investigated. In further steps, patient-specific data could be included, resulting in more and various anatomical models, including different access strategies. Despite of the chosen use case of IAs, the simulation could also be used for other surgeries that include a craniotomy, like brain tumor removal. With respect to IAs, further important aspects that were mentioned by the neurosurgeons could be included. Before the craniotomy step, the skin incision and correct opening of the skin can also be part of the training. Regarding the vessel and aneurysm exposure, more details like the dura could be included. Proceeding with the surgery, the clipping process can be implemented, comprising aspects such as the pulse for pulse synchronous clipping. After applying one or multiple clips, there should be the possibility to evaluate the clipping to verify that the aneurysm is sealed off and the normal blood flow in the parent artery is preserved.

8.4.2 Evaluation of the Clipping Extension

Before presenting the evaluation of the extension, improvements that were realized based on the evalution of the craniotomy are introduced. The actual evaluation includes the study setup, results and discussion.

8.4.2.1 Improvements

Based on the feedback from the previous evaluation, the input device from normal VR controller to the VR Ink was changed. This pen-like device is more similar to the real medical instrument than a device held in power-grip, such as the VR controller (shown in Figure 8.10). This should prevent a wrong muscle memory and more precise interactions are possible with this precision grip. Furthermore, the previously described last step of the workflow and the presented visualizations were added (refer to Section 8.3.3).



(a) Oculus Quest controller. (b) VR Ink.

Figure 8.10: The VR controller used for the craniotomy (a) and the VR Ink used for the clipping extension (b). Own figure.

8.4.2.2 Setup

The virtual clipping was qualitatively assessed by two neurosurgeons: One male senior neurosurgeon (S) with 13 years of neurosurgical experience who often used VR, and one female novice neurosurgeon (N) with one year of neurosurgical experience who has used VR a few times. The evaluation started with a short description of the tasks and interactions. Before using the application, they were asked to complete the Simulator Sickness Questionnaire (SSQ-pre) [151]. Subsequently, they had to go through the application. Thereby, they were verbally assisted and reminded which buttons they have to use. After they completed the workflow, they were asked to fill out a questionnaire comprising the following parts:

1. Simulator Sickness Questionnaire (SSQ-post) [151]
2. NASA Task Load Index (NASA-TLX) to assess mental and physical demand [113]
3. Questions referring to clipping, deformation and clip assessment:
   - Is the deformation sufficiently realistic to provide an effective training? If not, why and what is missing?
   - Is the deformation important? Or is the clip placement with the additional visualization sufficient?
   - Is an effective training with this system possible? If not, why and what is missing? Which aspects can be trained, which cannot?
   - How should the clipping (the position of the clip) be evaluated within the system? 1) visual, explorative assessment 2) numeric feedback of how much of the aneurysm is closed 3) filling the vessels with blood particles (no real blood simulation)
4. Questions referring to further improvements
   - Including a narrative/emotional context by providing personal patient information would...
     - influence the medical approach (yes, no, maybe)
     - influence (subconsciously) the performance and approach (yes, no, maybe)
     - increase the realism of the simulation (yes, no, maybe)
     - involve the user more emotionally and thus motivate them more (yes, no, maybe)
   - Multiuser in terms of working together in the virtual environment would...
     - motivate and increase the training effect (yes, no, maybe)
     - lead to a more frequent use (yes, no, maybe)
   - Multiuser in terms of competing in the virtual environment would...
     - motivate and increase the training effect (yes, no, maybe)
     - lead to a more frequent use (yes, no, maybe)
   - With which criteria can the whole procedure be evaluated and compared?
   - Are there further requirements that are not fulfilled by the simulation or would improve the simulation?
   - The following aspects or steps complicate the clipping of aneurysms (in a real surgery)

5. Igroup Presence Questionnaire [275] and the following three additional questions to assess presence and immersion:
   - Due to immersion I could concentrate better.
   - Due to immersion I take the task more seriously.
   - Due to immersion I am more motivated to solve the task correctly.

8.4.2.3 Results

In the *SSQ* one participant (S) mentioned an increase from 'none' to 'slight' regarding the symptoms 'general discomfort' and 'eye strain' after using the VR system.

The results of the *NASA-TLX*, where usually a 20-point Likert scale is used, show that for one participant the task was more mentally demanding (N: 15/20, S: 15/20) than physically demanding (N: 11/20, S: 18/20). Both of them had to work hard to accomplish their level of performance (N: 14/20, S: 19/20). The question of how frustrated they were (N: 2/20, S: 14/20) and following-up questions indicate that the novice surgeon did not have more difficulties placing the clip compared to real microsurgical procedures. However, the senior neurosurgeon emphasized that the virtual clipping is much more difficult due to jittering, leading to higher frustration. This could be because of a missing physical skull which they usually use to stabilize their hands or due to the tracking.

Clipping Regarding the clipping procedure, both participants liked the deformation and rated it as realistic enough for the training application. They also emphasized that it is a crucial part of a training system. However, it would be good to have the possibility to modulate the speed of the clip application. Usually, surgeons do not just apply the clip but close it carefully, observe the deformation and open it again to reposition it if necessary. Showing the area affected by the clip helped the surgeons to discern the clip location in the 3D space. Furthermore, it should be possible to apply multiple clips.

Clip assessment For this, different possibilities were discussed. The first possibility would be to include visual exploration of the clipped aneurysm. Both rated this approach as very useful, as this would be similar to an angiography with which one can check whether the aneurysm is sealed off completely. Furthermore, they would appreciate an additional numeric output indicating how much of the aneurysm ostium is closed (for example as percentage). The third possibility was rated a bit less helpful than the others. Here, it was proposed to fill the vessels with blood. One participant mentioned that with this one they can see whether the parent artery is still open. However, it is difficult to simulate realistic blood flow during runtime of a VR real-time application due to time-consuming mesh and convergence requirements [126]. It would be possible to add particles that will move through the vessels but without any calculated or measured flow field, they will probably not behave realistically. Further aspects that can be used to evaluate and compare the clipping results are time and how strongly the brain is retracted.

Improvements Here, general improvements were discussed. First, collaborative VR, where two or more users can use the simulation at the same time, was proposed. They can either work together and discuss different approaches, or work in a competitive mode where the users compete against each other. Both participants think that both modes would lead to an increased motivation and learning effect. However, the feedback and reactions of the senior neurosurgeon show that he would prefer the competitive mode.

Moreover, the brain retraction should be limited because such large deformations are not possible in reality. The novice neurosurgeon stated the consistence and volume of the brain as a challenge of the access during clipping surgery. A further aspect that could be included is puls synchronous clipping. In this context, the novice neurosurgeon also mentioned that including sound of an electrocardiogram would make the experience more realistic.

Immersion The last part of the questionnaire was about immersion. Both participants agreed that they felt present in the virtual space and they were captivated by the virtual world. However, they were aware of the real world and paid attention to it. This could be due to verbal instructions from outside. The virtual world was rated as real and the virtual experience was rated as moderate consistent with the real world. Nevertheless, the virtual world was perceived as an imaginary world.

The three statements additional to the Igroup Presence Questionnaire were used to get an impression of whether the participants think that immersion can improve the training experience. The statement 'Due to immersion I can concentrate better on the exercise' was rated with 4 (S) and 5 (N), with 1=fully disagree and 5=completely agree. Whether they take the exercise more seriously was rated with 3 (S) and 5 (N). The last statement 'Due to immersion I'm more motivated to solve the exercise well' was rated with 4 (S) and 5 (N).

8.4.2.4 Discussion

The proposed training system differs from previous approaches on the basis of providing an immersive environment using a virtual OR for aneurysm clipping training. Consequently, the prototype provides a motivating and engaging experience, which is essential for effective training. With the presented prototype, an immersive environment is combined with IA clipping using a soft body deformation. Thus, users can try out different strategies in a more realistic scenario. Although the input device is no medical instrument, it was important to provide a similar device regarding the hand position. However, having a completely virtual system without grounded haptic devices has the drawback of having no supportive surface which is available in a real clipping procedure. Using a virtual skull also has the drawback that the user is not restricted when moving the clip because their real hands do not collide with the skull. The evaluation focused on qualitative feedback, which also provides important feedback for further development. However, having more participants would lead to statistical results. The main challenge in this case is to get enough participants, as the target group is limited to neurosurgeons.

Future Work To address the jittering problem, a 3D printed skull with a large craniotomy hole can be used. This physical model has to be adjusted according to the virtual head placement or the virtual head has to be registered with the physical skull after placement. A more simplified and robust alternative (if various virtual skull models are incorporated) would be a sphere or ellipsoid with a hole. Although it does not represent the virtual head, it might be possible that the user will not recognize the difference if it only serves to support the hand. It might be more misleading if the virtual hole size does not correspond to the physical hole size. However, to provide a correct hole, the user directly has to cut the physical representation. The senior neurosurgeon mentioned a jittering. If this is still present with a supportive surface, it might be due to tracking inaccuracy.

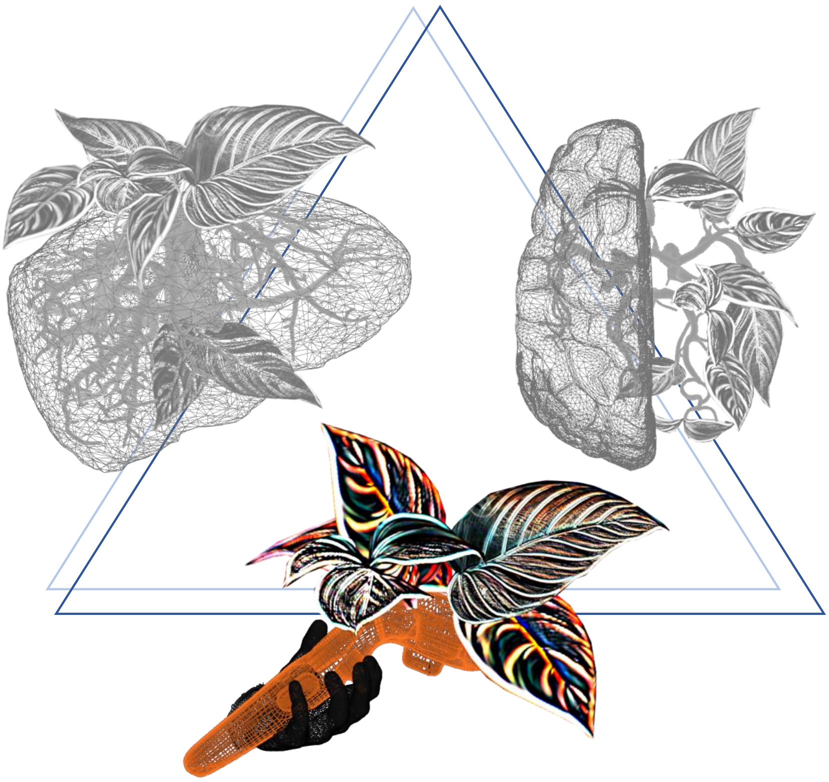
Similar to the future work regarding the craniotomy, including feedback and a larger user study are essential for future prototypes.

8.5 Conclusion

The proposed VR system for craniotomy benefits from using an HMD, leading to an immersive training possibility where the user feels present in the virtual OR. The results indicate that this simulation can be used for training and teaching to improve anatomical understanding, to become aware of the importance of a correct craniotomy hole, and to feel better prepared for real surgery. Furthermore, the clipping extension benefits from a pen-like device and real-time deformation for the vessels. Both qualitative evaluations provided detailed insights into further improvements, for example, concerning proper feedback and assessment or improving the learning process by gamification or reference contours. Thus, an appreciated prototype was presented and various improvements for further development were identified.

Part III

Excerpt on Input Devices



9

A Comparison of Input Devices for Precise Interaction Tasks in VR-based Surgical Planning and Training

Synopsis After presenting the contribution, related work is briefly introduced and the selection of use cases is motivated. Afterwards, both applications, the input devices, and interaction tasks are presented. This is followed by the study, the results and a discussion.

This chapter is partially reused with permission of ©2022 Elsevier and has been published in:
Mareen Allgaier*, Vuthea Chheang*, Patrick Saalfeld, Vikram Apilla, Tobias Huber, Florentine Huettl, Belal Neyazi, Erol I Sandalcioglu, Christian Hansen, Bernhard Preim, Sylvia Saalfeld. "A comparison of input devices for precise interaction tasks in VR-based surgical planning and training", *Comput Biol Med*, 145, pp. 105429, 2022, doi: 10.1016/j.compbiomed.2022.105429.

A portion of this chapter has been published in the following paper ©2022 IEEE. It is reused with permission from Mareen Allgaier, Vuthea Chheang, et al.
Mareen Allgaier*, Vuthea Chheang*, et al. "A Comparison of Input Devices for Precise Interaction Tasks in VR-based Surgical Planning and Training", *Proc. of IEEE VRW*, pp. 674-675, 2022, doi: 10.1109/VRW55335.2022.00188.
* These authors contributed equally: *Mareen Allgaier* Implementation of craniotomy simulation, study design, study conduct, paper writing. *Vuthea Chheang* Implementation of resection simulation, study design, statistical analysis, paper writing.

9.1 Contribution

During the development of all previously described projects, the choice of an appropriate input device was permanently present. As presented in Section 3.1, VR is used for various purposes, such as training or planning, in healthcare. Independent of the use case and purpose, it is common to interact with 3D models. Thus, different aspects from the field of human-computer interaction have to be considered to design and implement a VR application. Among them are *interaction tasks* which are primitive inputs performed by the user [135]. 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 this work, common interaction tasks for medical VR applications are defined and used to compare input devices. Different interaction techniques are not compared, but common design choices are employed. 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 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 [139] or laparoscopic device [68] that is similar to a real medical instrument, is important for medical training applications. Due to this trade-off, different input devices in the context of medical applications are compared to investigate their suitability for common tasks.

In contrast to existing studies [24, 45, 236], the presented study does not just compare devices held in *precision grip* and devices held in *power grip*, but also hand gestures (refer to Section 2.2.3 for grips and interactions). Besides pointing tasks [24, 236] and sketching [45], the study includes other tasks that are common for medical applications. Based on the different tasks the investigation aims at determining which device or rather which grip is most suitable. Besides measuring the precision, usability questionnaires are used to assess how well the measured data corresponds to the participant's perception. In the study, the focus is on planning and training, for which two relevant and representative use cases were selected: a liver surgery planning and a craniotomy training. Using these use cases, the following research question is investigated:

RQ 4: Which benefits of input devices with varying grip styles should be considered when developing medical VR applications?

This project aims at analyzing relevant medical interactions and investigating which devices are appropriate in the medical context as well as determining their task-specific benefits and drawbacks. In a user study, benefits and drawbacks and the usability of input devices based on user performance and subjective feedback were identified. These results can guide further developments of medical VR applications.

9.2 Design

This section starts with related work. Because an overview of common input devices and types of grips was already presented (refer to Section 2.2.3), the literature presents examples for which types of medical applications these devices can be used for and other comparative studies. For literature concerning the two medical use cases, consult Chapter 4 and Chapter 7. In the second part, the chosen medical use cases are introduced.

9.2.1 Related Work

There are several virtual surgical planning and training applications in medical areas such as orthopedic surgery [102, 124, 231], neurosurgery [309, 340], general surgery [55, 86, 180] as well as oral and maxillofacial surgery [145, 212]. Although previous studies reveal the benefits of immersive VR applications using HMDs, not all of the previously mentioned approaches are immersive applications [128, 131, 243]. Common tasks in these applications are reaming [124, 231] and drilling [145], careful positioning of a plate [102], navigating and applying a clip [309], placement of screws [340], and laparoscopic tasks such as grasping, cutting and fine dissection [55, 86, 180]. Because of the various use cases, some exemplary devices and for which medical VR applications they might be used are presented. Afterwards, other comparisons of input devices and their findings are described.

Hand gestures An intuitive way of interacting are hand gestures. Sousa et al. [291] use gestures to interact bimanually with medical image data. One hand controls render properties, such as brightness and the cutting plane, while the other hand can manipulate the volume, such as scaling and rotating. Render properties are changed by moving the hand forwards and backwards as well as moving the hand to the left and right. For scaling, the pinch gesture is used and for rotation around the vertical axis, the hand has to be rotated, whereas pitch rotation is done by moving the hand through the forward-backward axis. For recognition, gestures have to be performed on a multitouch frame on the desk surface. Another way to detect gestures is by data gloves with, for example, a Steam VR tracker as Chheang et al. [52] proposed for surgical planning. With 2D image data and a 3D representation of the liver, the user can define a resection line. As mentioned in the technical background (see Section 2.2.3), there are also gloves that include force feedback to apply resistance. Boutin et al. [32] employed such a device for external ventricular drain placement. During this procedure, the neurosurgeon has to choose an optimal burr hole to insert a catheter.

VR controllers Another common type of input devices are controllers because they are usually associated with HMDs. Adams et al. [2] use controllers to explore and manipulate 2D and 3D image data. Controllers can also be used for specific therapy planning, like IA clipping [299]. For the individual buttons of VR controllers refer to Section 2.2.2.

Stylus devices Furthermore, there are stylus devices for precise interactions, such as drawing [45] or fine-motor tasks, for example, in micro-surgical procedures [4, 96] or dental surgery training [145]. In the latter case, the stylus device–a Geomagic Touch–is used as the dental drill. Stylus devices can be grounded haptic controllers, such as the mentioned Geomagic Touch, or midair devices, such as the VR Ink, Massless Pen[1], Wacom VR Pen[2], zSpace Ink[3], and some non-commercial devices [134, 146, 258].

Task-specific devices Finally, there are highly specialized devices for training applications. One example is the Simball[4] used for laparoscopic interventions [68, 129]. John et al. [139] simulate an endoscope to train handling and navigating it in the context of gastrointestinal procedures. As basis, they used an HTC Vive controller in a self-constructed holder and attached a Microsoft Surface Dial[5] to replicate the control knobs. For the endoscope shaft, they applied an HTC Vive tracker to a rod which can be pushed forward and pulled back.

Studies The variety of input devices and applications shows that the choice of an appropriate device has to be considered carefully. To assess the benefits and drawbacks, several studies have been conducted. Some studies compared the precision of devices held in power grip and precision grip [24, 45, 236] for pointing or sketching tasks. In Batmaz et al.'s [24] study, participants had to point at spheres with the VR Ink in precision grip as well as power grip (like a stick). Twelve participants had to perform the task with both grip styles and three different distances to the target and different target sizes. Significant differences between the grips were found for the error rate, with fewer errors for the precision grip. Subjective results showed that the majority preferred the precision grip over the power grip.

Cannavò et al. [45] compared an HTC Vive controller with a VR Ink in a within-subject design study with eleven participants. Therefore, they used a VR setting where the participants had to draw a curve mid-air and a hybrid setting where the participants were in a VR environment but had to draw on a physical surface. With these conditions, the participants had to draw circles, horizontal and vertical lines. Regarding deviation, the VR Ink performed significantly better than the controller in general and in the VR setting but not in the hybrid setting. They also investigated the drawing plane orientation and found significantly better results for the VR Ink in terms of deviation when the plane is sideways. Although no significant differences regarding usability were visible, participants rated the VR Ink significantly better with respect to ease of use, comfortableness and naturalness.

Another study by Pham et al. [236] compared a mouse, VR controller and pen for distant pointing. The objective results of twelve participants revealed that with the controller the movement time was significantly larger than with the other two devices. The controller also led to a significantly higher error rate than the mouse and pen. Furthermore, the

[1] Massless Corp, USA: https://www.massless.io/ Last access: 10.04.2024
[2] Wacom, Japan: https://www.wacom.com/de-de Last access: 10.04.2024
[3] zSpace, USA: https://zspace.com/ Last access: 10.04.2024
[4] Surgical Science, Sweden: https://surgicalscience.com/simulators/simball-box/ Last access: 10.04.2024
[5] Microsoft Corp, USA: https://www.microsoft.com/de-de/ Last access: 10.04.2024

controller had a significantly lower throughput than the other two devices and the pen had a faster cursor manipulation than the controller. Subjective results showed that the pen was significantly better regarding perceived cursor speed and ease of interaction than the controller. Concerning comfortableness, the pen and mouse were rated significantly better than the controller.

There are also studies comparing gestures and conventional interactions, such as a joystick or mouse [122]. Hettig et al. [122] used gestures and conventional devices for scrolling (decrement and increment) through medical image data and rotation around three axes. For gestures, they used two systems, one where the user has to wave in or out for scrolling and one where the user has to place the hand in two different areas of the field of view. The first variant achieved significantly better results than the second one with respect to task completion time. However, for the rotation task, both gesture variants performed significantly worse than the other conventional methods.

In the presented comparison, 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, are considered. The perceived usability and precision are assessed. In addition, the study covers different medical applications and thus interaction tasks and compares devices that differ in their grip as well as generalizability.

9.2.2 Medical Use Cases

For the comparison of input devices, two use cases that are different regarding their medical use case and purpose were chosen:

- Liver surgery planning, and
- craniotomy training.

Although these are specific medical use cases, they are still representative due to the comprised interactions. Basic interactions and tasks like selecting, grabbing, scaling and rotating are often part of exploring anatomical structures. Furthermore, the included task of precisely drawing (a resection line and craniotomy contour) is similar to following a trajectory which is important for medical tasks like navigating in air-filled structures [282]. Another similar task is cutting, which is part of most surgeries, for example, craniofacial surgery [346]. Because cutting or resecting is a general task, virtually planning a resection is not just relevant in liver surgery. Exploring and analyzing different cutting strategies considering risk structures is an important part of various medical areas. Consequently, the two applications were chosen based on their relevance and the involved interactions. The following two sections motivate the relevance of VR planning and training for the two use cases.

9.2.2.1 VR Systems for Liver Surgery Planning

Preoperative planning for liver surgery is based on 2D image data acquired from CT or MRI. 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. 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 [205, 242, 330]. 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 [187, 243].

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. While being specific, the task of creating an incision line and resection plane can easily be applied to other tumor resection surgeries.

9.2.2.2 VR Systems for Craniotomy Training

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

For VR-based systems for IAs the previous chapter already concludes that most systems focus on the clipping itself and not on the access and craniotomy and that they often use stereoscopic displays (refer to Chapter 7). In the case of brain tumors, some 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 [224, 294].

Consequently, the use case craniotomy planning is relevant because it is part of many surgeries. Similar to the resection planning application, it comprises interactions, such as cutting and drawing, that are often included in virtual medical simulations.

9.3 Development

This chapter first describes the workflow of the two applications. This is followed by the motivation of the choice of input devices, and finally the tasks comprised in the two applications are explained.

9.3.1 Medical Use Cases

Both applications are based on previous work and were developed in close collaboration with liver surgeons and neurosurgeons, respectively [7, 56]. The individual steps of the applications are illustrated in Figure 9.1.

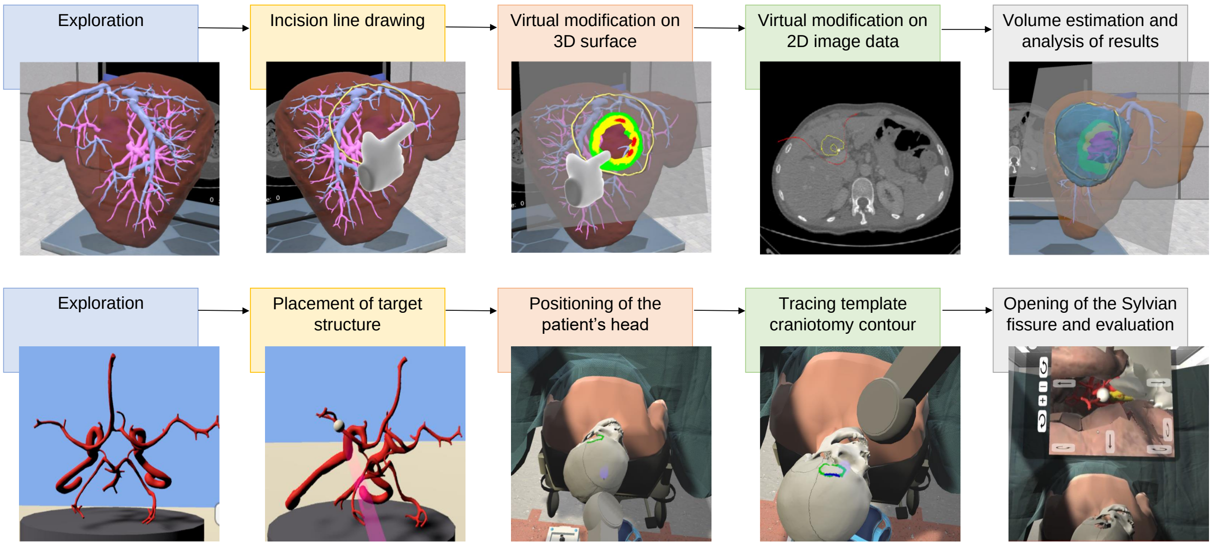


Figure 9.1: Individual steps of liver surgery planning (top) and neurosurgery training (bottom) and associated interactions. Image from Allgaier et al. [13] © 2022 Elsevier. No changes were made.

Liver Surgery Planning This system already provides patient cases which include the medical image data and segmentations of the liver and its interior structures. Accordingly, the user first chooses a patient dataset. As common for such planning systems, the user can draw lines on the liver model to initialize a virtual resection. For a proper evaluation, the line was given and the user has to trace it. Based on this, the origin and directions of the resection surface are determined by a principle component analysis according to Konrad-Verse et al. [162]. After that, the virtual resection is initialized, and a risk map visualization to the liver tumor is projected [111]. The modification of the virtual resection can be realized by two methods. It can be directly deformed on the 3D surface–by pushing and pulling–as well as on the 2D line representation on the 2D image slices. After modification, the resection and remaining volume can be estimated. The resulting reconstructed 3D model representations for each part–the resected and remaining–are highlighted with different colors and their volumes are displayed.

Craniotomy Training The employed application was already presented in Chapter 8, nevertheless, the workflow will be briefly summarized. First, the user has to explore healthy brain arteries and place a target structure which is approximated by a sphere. Based on this, they have to decide on a proper approach. Therefore, the user has to position the patient's head and proceeds with removing a section of the bone. This procedure is simulated by sketching the contour of the craniotomy hole that should be cut. Similar to the other application, a craniotomy contour, which the user has to trace, was given.

9.3.2 Input Devices

The interaction tasks as well as the purpose of the applications, whether it is for planning or training, affect the choice of the input device. In the presented comparison, four input devices are investigated, three per application: VR controller and VR Ink are included in both applications and additionally the data gloves are used for liver surgery planning and the craniotome is used for craniotomy training (see Figure 9.2).

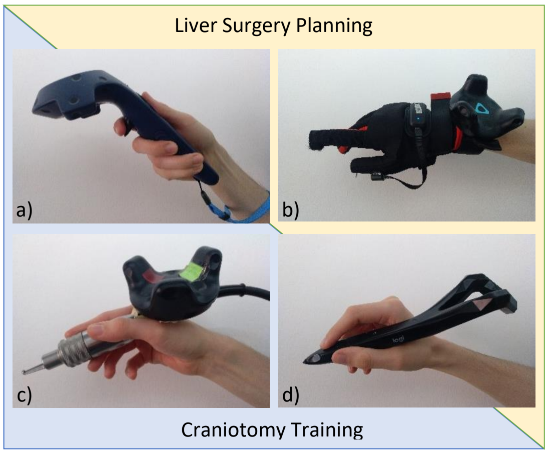


Figure 9.2: Input devices: a) VR controllers, b) Manus data gloves, c) craniotome, d) VR Ink. a), b) and d) are included in the liver planning, whereas a), c) and d) are used for craniotomy training. Image from Allgaier et al. [13] © 2022 Elsevier. No changes were made.

The *VR controllers* are used in both applications as they usually come with HMDs, and thus, serve as the baseline device (Figure 9.2a). They are widely available, cheap and familiar to people using VR frequently. For the implementation and study, the HTC Vive Pro Eye and the corresponding controllers were used.

The second device that is used in both applications is the *VR Ink* (Figure 9.2d). The reason for this device is that the main task of both applications is drawing and the natural hand position for drawing cannot be provided by a controller but by a pen-like device. Especially in the craniotomy training application, a correct and realistic hand position is crucial to build up a correct muscle memory. Frequent training with a device held in a non-realistic way can lead to incorrect muscle memory, which would be particularly fatal for novice surgeons. Furthermore, the precision grip is known to facilitate more precise interactions [24, 45, 236]. In particular for medical applications, precision is important when interacting with small or vulnerable structures. There are also other stylus devices which include haptic feedback. These grounded haptic devices were not included, 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. However, the importance and benefits of grounded haptic devices, especially for surgical training, should not be neglected. Nevertheless, the comparison focuses on the grip style, which would be the same as for the VR Ink, and thus including haptic feedback would be an additional variable to be investigated separately.

With the *Manus data gloves*[6] (Figure 9.2b), hand interactions were provided which are very intuitive for tasks, such as selecting, pointing and scaling. To avoid the mentioned incorrect muscle memory, the gloves are only used for the planning application and not for craniotomy training, as they differ the most from the real device.

[6]Manus, Netherlands: https://www.manus-meta.com/ Last access: 10.04.2024

The fourth device arose to support the right muscle memory and was chosen for the craniotomy training. Because the best device for a training application is the actual instrument, a Vive tracker was applied on a *craniotome* (Figure 9.2c), which is used by surgeons to perform a craniotomy. However, in contrast to the other two devices, it is highly specialized for this task. Accordingly, they can be compared to find a compromise between providing a realistic hand position and being broadly applicable.

9.3.3 Task Description

In the following, the tasks–some are included in both applications, some are application-specific–and their corresponding interactions are described. An overview is given in Figure 9.3. Furthermore, the device with the highest interaction fidelity, which describes how exact real-world actions are reproduced in an interactive system [199], is identified. Here, the biomechanical symmetry, input veracity and control symmetry are considered [200]. The following focuses on biomechanical symmetry, as the others depend on the quality of the input devices or the translation of user actions to system effects and not on the general type of device. Biomechanical symmetry describes to what extent the interaction reproduces real-world body movements required for the specific task. Thereby, it considers the involved body segments, the required motion and forces.

Selecting and pointing Most tasks that require pointing and/or selecting occur when interacting with the UI. This includes interacting with buttons and sliders, which is involved in both applications. Additionally, in the craniotomy training application, the users have to point at the arteries to place a target structure. The liver planning application requires pointing and selecting for resection plane deformation. 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 in the real world (for example, pressing a button or pointing at image data) one usually uses the index finger. 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 a pen held in precision grip, are also common.

Drawing In contrast to pointing and selecting, the users do not only have to point at one location, but precisely move their hands. This task is included in both applications when drawing the incision line on the liver and craniotomy contour, respectively. 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, 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 additionally 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 only one of the two applications.

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. As two devices are necessary, the VR Ink is used together with a controller. Only with the gloves, the user really makes a grab movement. The pulling movement is the same with all devices. Accordingly, the gloves have the highest interaction fidelity.

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.

Rotation via trackpad To rotate the brain vessels 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, for example, a 3D printed artery model. However, this interaction method is familiar due to steering via joystick.

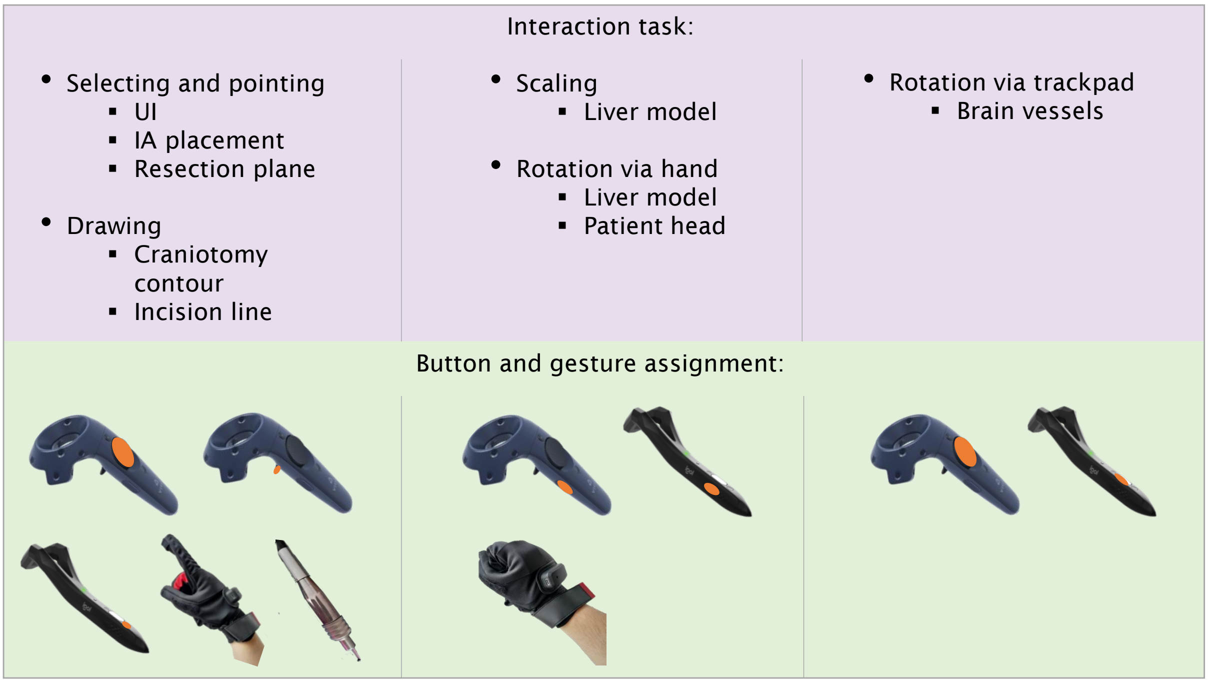


Figure 9.3: Summary of interaction tasks and the corresponding button and gesture assignment. Orange highlights the button that has to be pressed on the controller and VR Ink. Own figure based on Allgaier et al. [13].

9.4 Impact

The focus of the evaluation was on the input devices and their suitability for the different tasks. Before conducting the user study, hypotheses, which arose from the specified tasks and the described interaction fidelities of the input devices, were set up:

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 VR Ink is more suitable due to its larger trackpad.

9.4.1 Setup and Procedure

After describing the participants' characteristics, the procedure and included questionnaires are presented. In the second part, the quantitative measurements are described.

9.4.1.1 Study Procedure

The evaluation was conducted with 22 participants (see Table 9.1). Besides the common data such as gender, age and handedness, their professional background and experiences with VR were recorded. For VR experience, the following three categories were used: 'never used VR', 'used VR several times', and 'using it regularly'.

Table 9.1: Characteristics of participants ($n = 22$). Participants with STEM background mainly are computer scientists and engineers. Image from Allgaier et al. [13] © 2022 Elsevier. The table was split into two columns.

Characteristics	Value	Mean
Age [years]	[15-38]	27.64
Gender		
Male	11	(50%)
Female	8	(36.4%)
Non-binary	3	(13.6%)
Background		
Medical experts	4	(18.2%)
STEM	17	(77.3%)
Pupils	1	(4.6%)

Characteristics	Value	Mean
Experience with VR		
Never used before	5	(22.7%)
Used several times	14	(63.6%)
Regular use	3	(13.6%)
Handedness		
Left	1	(4.6%)
Right	21	(95.5%)

Although the tasks and interactions are based on medical applications and their specific requirements, the study tasks were 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, for drawing the incision line or craniotomy contour, they had to trace a given line and contour, respectively. 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. Accordingly, the majority of participants had a background in computer science and engineering. 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, which is described in the following, is summarized in Figure 9.4. 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, two standardized questionnaires were employed. After completing the workflow with one device, the participants were asked to answer questions based on the *NASA Task Load Index (TLX)* [113]. 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 separate usability questionnaire. Thus, the questionnaire contained questions for mental demand, physical demand, performance and effort.

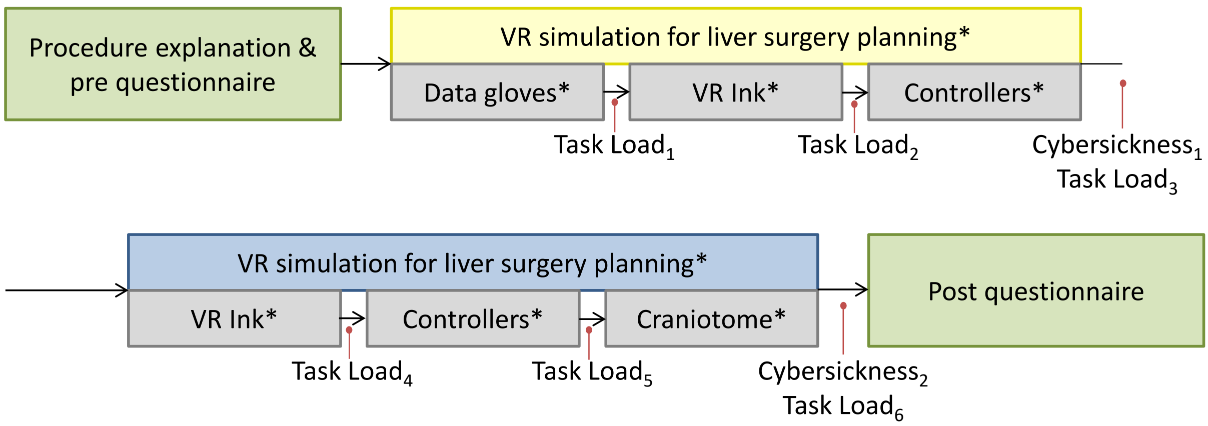


Figure 9.4: Study procedure. The order of the devices and applications (marked with *) was randomized. Red: Mid-questionnaires. Image from Allgaier et al. [13] © 2022 Elsevier. The figure was split into two lines.

Questions of the *Simulator Sickness Questionnaire (SSQ)* [151] 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 [35]. According to the research focus, the questions had to be adapted in order to focus on 'the device' instead of 'the system', resulting in the following statements:

- When using the system frequently, I would like using this device.
- I found interacting with this device unnecessarily complex.
- I thought the device was easy to use.
- I would imagine that most people would learn to use this device very quickly.
- I found the device very cumbersome to use.
- I felt very confident using the device.
- I needed to learn a lot of things before I could get going with this device.

Each statement has to be rated for each input device per application, thus, the controller and VR Ink have to be rated separately for each application. The following three statements that could not be adapted properly were discarded:

- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.

Subsequently, the users had to compare the devices according to their suitability to perform the single tasks. Finally, they were asked to rank the devices for each application. All ratings employ a 5-point Likert scale.

9.4.1.2 Measurements

Besides the questionnaires, quantitative data was measured and calculated. In the liver surgery planning application, the performance of the drawing task was recorded. *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 because of not being satisfied with the result. In addition, *drawing attempts* were counted when the user attempted to draw the incision line 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. Additionally, *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* P and *Hausdorff distance* H of the drawing results were calculated. In the craniotomy training application, the given contours and the contours drawn by the user are defined by a set of colored vertices of the skull's triangle mesh. Based on this, the precision is calculated by:

$$P = N_{correct}/N_{drawn} \tag{9.1}$$

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 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 [132]:

$$\begin{aligned} H(A,B) &= max(h(A,B), h(B,A)) \\ h(A,B) &= max_{a \in A} min_{b \in B} ||a-b|| \end{aligned} \tag{9.2}$$

9.4.2 Results

The presented results are separated into the measured performance data and the results of the questionnaires.

9.4.2.1 Measurements

To get a statistical insight into the obtained data and to see whether there are significant differences between the input devices with regard to user performance, one-way repeated measures ANOVA were conducted [47]. The corresponding statistical results are summarized in Table 9.2. Additionally, the post-hoc analyses using pairwise t-test with *Bonferroni* correction to compare the input devices were performed after the results of significant effects [30].

Table 9.2: Summary of the ANOVA's results of the input devices (* denotes statistical significance ($p < 0.05$)). Image from Allgaier et al. [13] © 2022 Elsevier. The table was merged horizontally.

	Liver surgery planning				Craniotomy training			
Variable	df	F	p	η^2	df	F	p	η^2
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	2	7.136	0.0016*	0.18
Hausdorff distance	2	1.102	0.339	0.03	2	1.366	0.263	0.04
Deforming attempts	2	6.101	0.0038*	0.16				
Deforming error	2	0.023	0.977	0.0007				

Liver surgery planning The results of drawing and deformation measurements for liver surgery planning are illustrated in Figure 9.5. Precision results are shown in Figure 9.6. Statistically significant effects on drawing precision ($p = 0.0034$) and deforming attempts ($p = 0.0038$) were found. These variables were further evaluated 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 the 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 Figure 9.6, 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 ($\varnothing = 28.17\,mm$). It could be assumed 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 $100\,mm$.

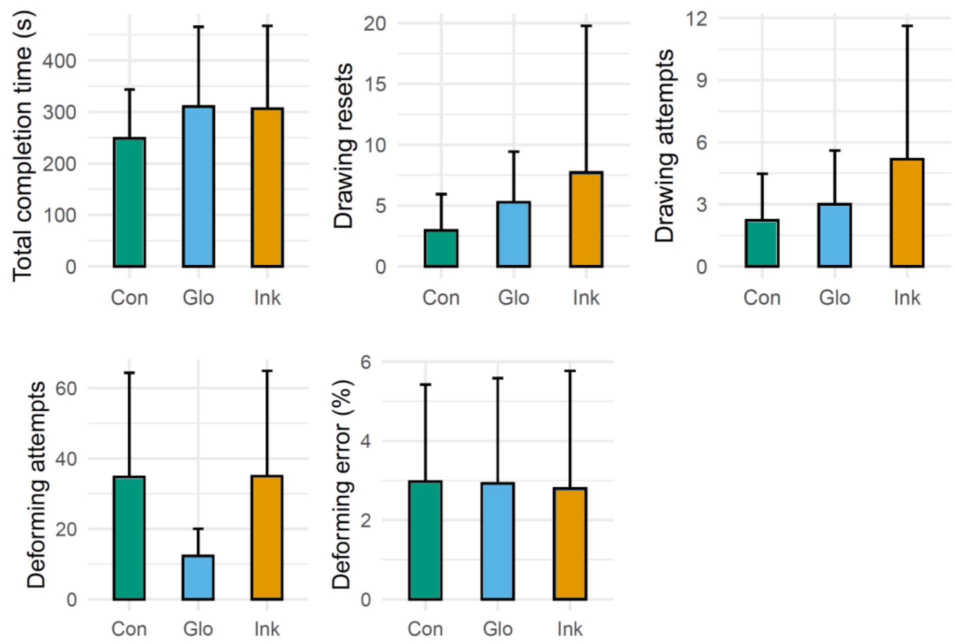


Figure 9.5: Statistical results of input devices for liver surgery planning (Con: controllers; Glo: data gloves; Ink: VR Ink). Image from Allgaier et al. [13] © 2022 Elsevier. The image was split into two lines and the color scheme was changed.

Craniotomy training Statistically significant effects regarding the precision between the input devices ($p < 0.00161$) were found (see also Figure 9.6). 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 ($\varnothing = 63.00\%$). 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.

9.4.2.2 Questionnaire results

The participants were asked to answer the mid- and post-questionnaires (see Figure 9.4). The results of the questionnaires were analyzed descriptively in the following.

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 ($\varnothing = 11.95$,

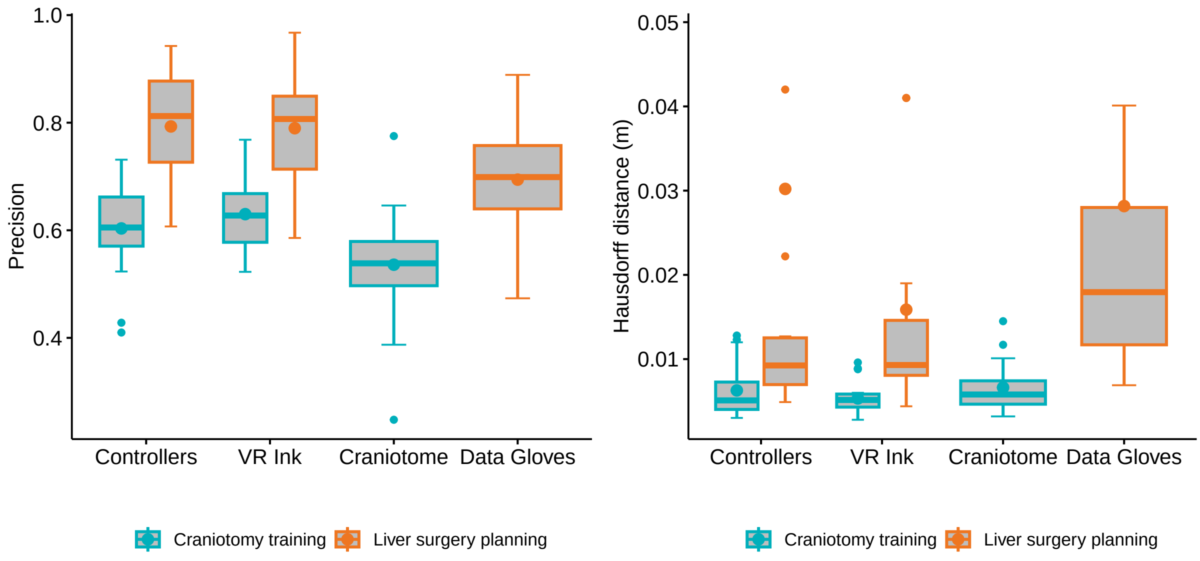


Figure 9.6: Precision and Hausdorff distance of input devices for liver surgery planning and craniotomy training. Image from Allgaier et al. [13] © 2022 Elsevier. The color scheme was changed slightly.

$\sigma = 4.7$). The success of accomplishing the tasks and effort were rated in the same order.

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 ($\varnothing = 1.27$, $\sigma = 0.54$). They were rated with a 5-point Likert scale, where one equals no occurrence. The symptoms that occurred most frequently, meaning they were rated greater than one, are eyestrain, sweating, and fatigue.

Post-questionnaire This questionnaire uses a 5-point Likert scale, one equals disagreement and five equals agreement. The results are presented in the following.

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: $\varnothing = 3.73$, $\sigma = 1.17$; 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 of 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 Figure 9.7.

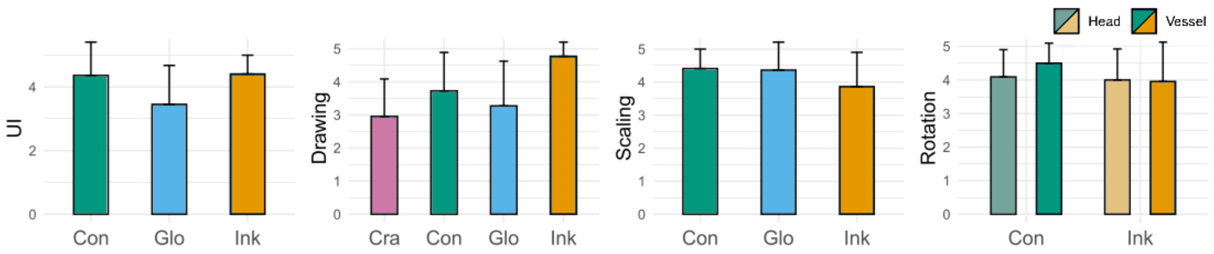


Figure 9.7: Average rating of the suitability regarding the different tasks (Con: controllers; Glo: data gloves; Cra: craniotome; Ink: VR Ink). Image from Allgaier et al. [13] © 2022 Elsevier. The color scheme was changed.

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. The analysis revealed 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. There were no statistically significant differences between the devices regarding 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 Figure 9.8. 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.

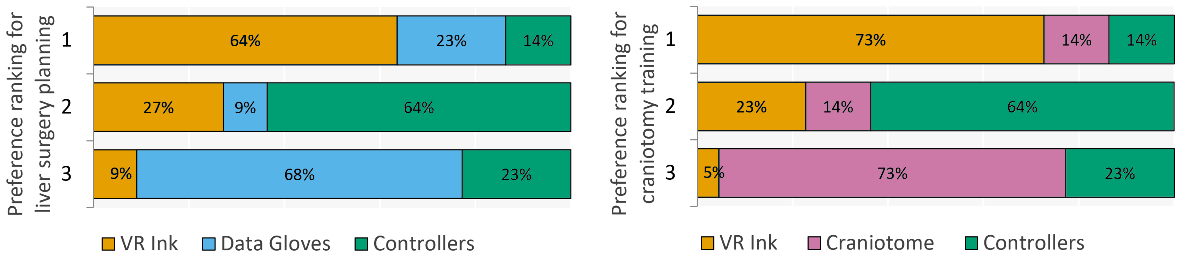


Figure 9.8: Results of the ranking of the most preferred devices for liver surgery planning (left) and craniotomy training (right). Image from Allgaier et al. [13] © 2022 Elsevier. The color scheme was changed.

General feedback Some participants gave general feedback during the study or left a comment on the questionnaire. They were summarized and categorized according to the specific devices.

Regarding the *VR Ink*, several participants mentioned problems reaching the buttons. To reach the trackpad, most had to change their grip leading to difficulties in balancing the VR Ink during usage. But also 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 a good compromise between being heavy enough for stabilization and being light enough to avoid effort (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. 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.

Concerning the *data gloves*, the main issue was drawing. For several participants it was not intuitive to draw with the index finger. Additionally, 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 thumb is moved away from the index finger, the drawing is stopped. Many participants had problems with this mechanism because it is not intuitive and sometimes it is even more cumbersome due to a not correctly tracked thumb leading to strange thumb positions.

With regard to 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, which complicates 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.

Hypotheses The hypotheses stated in Section 9.3.3 are mainly refuted: Concerning $H1$, 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 their low interaction fidelity, familiar due to devices in daily life such as a presenter. $H2$ 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 $H3$, the VR Ink and controller are approximately equally suitable. Only for rotation via trackpad, the controller performs slightly better.

9.4.3 Discussion

In comparison to previous studies [24, 45, 236] showing that the precision grip leads to statistically higher precision and less errors, the presented study does not show a significant difference between the VR Ink and the controller with respect to precision. Comparing the Hausdorff distance, it is visible that 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.

A more recent study was published by Rantamaa et al. [249]. The medical tasks included object manipulation (rotation and translation) and precise object marking. They compared mouse interaction, hand interaction and a combination of VR controller (for manipulation) with an ink (for marking). The hand interaction was significantly better than the mouse in terms of marking accuracy. However, similar to the results of the presented study, the controller/stylus interaction is significantly preferred over the other two interactions in daily use and is rated as the most natural, easy and accurate method for object manipulation.

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 two of three 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, much more tracking issues were observed 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 the 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 [25].

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 the high number of drawing attempts might be the different hand positions. When sketching, people often do not draw one line, but instead make several small strokes. The number of drawing resets is also the highest with the VR Ink and the 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 be that these devices are not as familiar to users as the controller. Consequently, for these devices, the participants need more time to become familiar. Regarding the 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 assessed to exclude that, for example, 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, it can exclude that cybersickness has an impact on the evaluation of 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.

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 this bias is compensated but it would still be helpful for the participants to first 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 in the sense of achieving a medically correct result. The results of the deformation only compare the deformed plane with the example plane via volume, but do not consider anatomical aspects, such as 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 concerning complexity, they comprise representative tasks that are common in VR-based medical applications. The involved task *drawing* is (as explained in Section 9.2.2) similar to other medical tasks. However, based on the given scenario, the task complexity may differ. 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 minimally invasive surgery, such as laparoscopic surgery, the surgeon has a limited field of view and operating space [77]. Furthermore, depth perception is complicated leading to impaired hand-eye coordination [316]. The task complexity can also vary depending on aspects, such as interaction space or size of execution. For example, the craniotomy contour is much smaller and finer than the incision line on the large liver model. Increasing the complexity of the applications might 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.

9.5 Conclusion

A comparative study of input devices regarding their suitability to accomplish precise interaction tasks in two medical applications–aliver surgery planning and craniotomy training–was presented. 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 regard to the 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.

10

Final Conclusion

Synopsis This chapter summarizes the main contributions of this thesis by answering the research questions. Afterwards, limitations that apply to all projects and the work in general are discussed. The chapter concludes with ideas and remarks for future work.

10.1 Summary

This section is structured by the two competencies that were addressed with the presented training applications. For each project, the research question is repeated and answered.

Visuospatial skills For training these skills the use case IOUS for liver surgery was chosen. With two projects, the following two research questions were answered:

RQ 1: What are suitable training scenarios of a VR-based application to train visuospatial skills for IOUS?
RQ 2: Which game elements are appropriate for a VR-based training for visuospatial skills for IOUS?

In close collaboration with liver surgeons, a VR training environment including an US simulation, virtual OR and haptic device for scanning a liver was developed. Based on the intraoperative workflow and the learning objective of *building a mental model using US*, four training scenarios were designed and implemented. An evaluation with medical experts revealed that three of four scenarios are considered to be meaningful and qualitative comments gave insights on how to improve them. The first scenario provides an overview of the liver segments which requires the understanding of the vascular courses. Comments suggest asking for the segment in which the lesion lies or transferring the inspection of the liver and its segments into an additional learning room. In the second scenario, the vascular courses have to be understood and precise probe manipulation is required to trace them. Concerning this scenario, the interaction of selecting the correct vessel should be changed to pointing at the vessel on the US image. In the third scenario, the correct liver has to be selected. The experts suggested ways to include various difficulties. Only the fourth scenario, where the user has to reproduce a given US image, should be replaced or omitted because it is not clinically relevant. To answer RQ 1, one can summarize that the first three scenarios are suitable for the training objective.

In a second project, it was investigated how the training experience, using one of the above scenarios, can be enhanced by gamification. Therefore, several common game elements were analyzed and discussed. The chosen elements–a level display and a kit interaction–were compared in a broad audience study and a medical target group study. Although only few significant differences could be found in the broad audience study, the medical target group study revealed a trend. In the broad audience study, the enjoyment construct of the miniPXI revealed significantly better results in the level group than in the control group and the kit group. Moreover, scores are appreciated for personal improvement and nearly half of the participants would use a leaderboard. Participants of the second study preferred the kit interaction due to its interactive character leading to more fun and being (subjectively) more effective by stimulating three-dimensional thinking. They also mentioned that placing lesions and not having a predefined selection leads to higher concentration. Regarding levels, most participants appreciated them because of the progress and performance feedback. Furthermore, the levels make the application less monotonous and were rated as motivating. However, the levels have to be balanced to be effective.

These insights can be used for further studies, such as a longitudinal study, or as basis for similar training scenarios. Based on the presented analysis of game elements and the study results, the answer to RQ 2 is that the proposed kit and levels are appreciated and thus are appropriate game elements for the chosen use case.

Strategic knowledge This knowledge type was addressed in the training application for neurosurgery. The main objective was that the user understands the importance of a correct craniotomy hole. For this, the investigation was whether VR can provide additional training using the following research question:

RQ 3: To what extent can VR provide an additional training possibility for microsurgical procedures in case of aneurysms?

As basis, a training environment for access training, which was later extended by the clipping procedure, was developed. According to subjective feedback from medical experts, such an immersive VR training application benefits from the following aspects: free exploration, more fun, and visualization possibilities. The latter should at least include feedback related to performance and could be implemented using gamification. In addition to these benefits, they stated that this application can be used to learn the importance of a correct craniotomy approach. The evaluation also revealed minor improvements that should be implemented in further prototypes, such as more interactions or the microscopic view. Besides these minor improvements, general limits of such VR applications were mentioned. These include haptic feedback, realistic deformations, and realistic anatomy. Although these aspects could be implemented in a better and more realistic way, they will probably always be a limitation of VR in comparison to physical simulations or cadaver training. Consequently, regarding RQ 3, VR provides a good training situation where the importance of the access can be learned and explored. Though, physical properties such as haptic feedback, realistic anatomies and deformations might be limited.

Input devices Besides the specific training applications, the thesis includes an excerpt on input devices. Because the question of which input device should be used for a specific application was present in all projects, the investigation aimed at answering the following research question:

RQ 4: Which benefits of input devices with varying grip styles should be considered when developing medical VR applications?

Using two medical VR applications–a liver surgery planning and craniotomy training application–, four input devices with different grip styles were compared: a VR controller (power grip), VR Ink (precision grip), craniotome (precision grip, only for the craniotomy training), and data gloves (hand interaction, only for the liver surgery planning). For the comparison, common interaction tasks, such as selecting, rotating, scaling, and drawing were considered. In both applications, the VR Ink and controller performed significantly better than the third device with respect to precision. These two were also rated significantly better for UI interaction. Regarding the precise task drawing, the VR Ink was significantly better than the other devices. This was also the device that was preferred most. These results can serve to guide the selection of an input device, however, the context and specific application as well as its purpose have to be considered, too. To answer RQ 4 one can summarize that the VR Ink and controller are beneficial for precise interaction tasks, the VR Ink is most suitable for drawing, and it is favored by the participants for both applications.

Overall summary In summary, this thesis presented training applications and research results for two medically relevant skills or knowledge types that are not as often addressed as, for example, declarative knowledge for anatomy education or motor skills. With the applications, it could be shown that VR provides an important additional training modality for these skills. Although the exemplary use cases are very specific, the insights might be easily adapted to similar cases, for example, IOUS for other organs or craniotomy training for other brain lesions. To provide proper applications and evaluations, experts were asked for feedback during the development and (for most projects) recruited for evaluation.

10.2 Limitations

Additionally to the limitations described for each project, in this section, some general limitations are emphasized.

Medical experts The first limitation is the number of cooperation partners. No matter which project and which medical discipline, close cooperation have taken place with medical experts, however, the medical experts were from one group, such as the neurosurgeons of the University Hospital Magdeburg and the liver surgeons of the University Hospital Mainz. Consequently, already the medical input during the idea generation and conceptualization is based on only one medical group. Furthermore, all decisions that the medical experts have to make during the development are also only based on one group. Literature reports on differences regarding surgical methods between hospitals as well as surgeons [297, 288]. However, it is unclear how large these differences are and to what extent personal preferences play a role during conceptualization and development of the proposed prototypes. Ideally, the group of medical experts includes experts from different hospitals. To limit the overreliance on one clinical partner, scientific literature was carefully analyzed, for example, to verify which training goals are considered essential for IOUS.

Nevertheless, the main motivation, for example, of the visuospatial skills in the case of IOUS is also mentioned in the literature.

At the moment, all experts are part of the projects, meaning that they are co-authors of the publications. Depending on the project, these experts or colleagues from the same working group evaluated the applications. Because of this, it would be even more important to include experts from other hospitals and experts who are not included in publications, resulting in probably more unbiased feedback during evaluations. In practice, this is very difficult because of two reasons. First, the proposed projects require a certain level of experience. Thus, medical students are not part of the target group and some of the trained procedures are not done in all hospitals. This limits the possible experts a lot. Second, physicians are very busy and most of them do their research in their leisure time, which makes it very difficult to find physicians who spend their time as medical experts during the project development without having any benefit, such as being co-authors. University hospitals at least provide a good infrastructure and support for research, whereas in non-university hospitals this support and appreciation for research is not given, which excludes these hospitals as cooperation partners.

Evaluations Another limitation regarding the evaluations besides the small number of participants are the used questionnaires. In some previous chapters, it is already described why no validated questionnaires were used. Looking at other presented VR simulations, it can be seen that non-validated questionnaires are often used for assessing face and content validity. Because the questions are very specific, it is not feasible to use the same questions in another context to have comparable results.

Although the presented training applications were evaluated to a certain extent–mainly the content and face validity–, all approaches lack an evaluation regarding the learning outcome and learning transfer. VR systems for medical education should be evaluated with respect to the knowledge gain. This is often done by a comparison with traditional or non-immersive methods [241]. However, in the case of surgical training, this is more difficult. It would require a proper inclusion of the training application in the clinical routine. The literature including a comparison of VR simulations with traditional simulations is quite limited [85]. Sometimes there is no state-of-the-art training, and assessing the skills at the patient is ethically questionable. Even if there are other training methods, there is also the question whether it is meaningful to compare these to investigate which one is better. Sometimes they address different learning objectives and should be used as complementary and not as replacements. Most of the presented evaluations were based on subjective results. Especially in a longitudinal study performance, error and time could be measured to provide more objective results.

Using *Kirkpatrick's model* [156] with the four levels of training evaluation shown in Figure 10.1, all presented evaluations, except for the pilot study of *LiVRSono*, addressed level 1. It would be best to evaluate on level 4, however, this is the most difficult one to conduct.

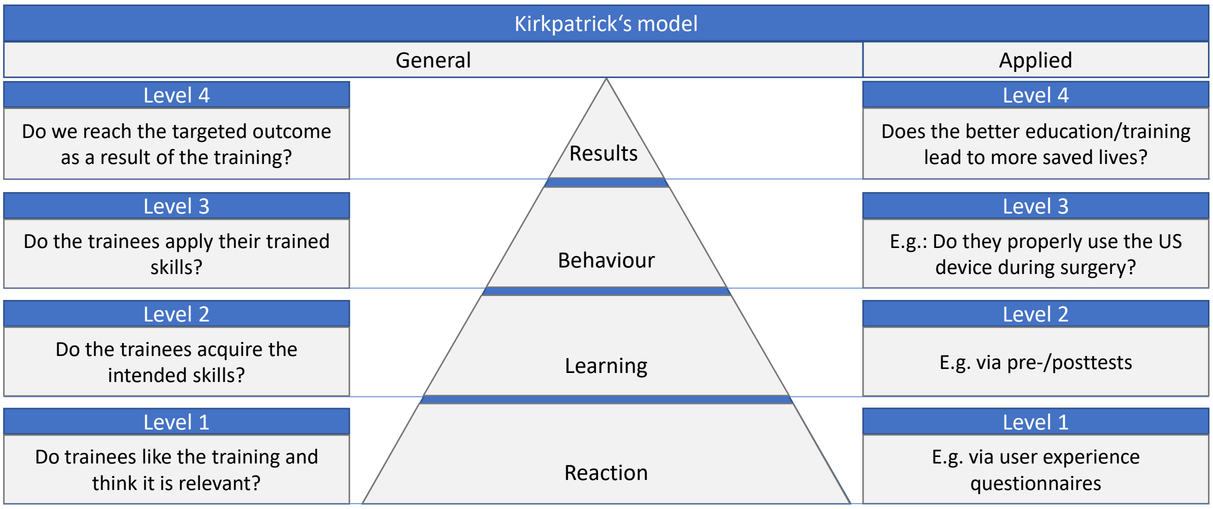


Figure 10.1: Kirkpatrick's model showing four levels to evaluate a training system. Own figure based on Kirkpatrick's model [156].

10.3 Future Work

Besides application-specific improvements and the mentioned evaluations for future work that were described in the respective chapters, this section will present more general prospects of VR-based surgical training. One aspect that was neglected in most presented prototypes, is soft tissue deformation. The difficulty regarding deformation is to create realistic and physically correct deformations. Therefore, many tissue-specific information and various information, for example, regarding the health condition of the current case have to be considered. Even if this could be solved, the necessary calculations are more time-consuming and computationally intensive, which is problematic with the performance of VR-based simulations. Other physical effects that were neglected include, for example, blood flow simulations. Here, the same challenges are present, which can be addressed in the future.

The next steps regarding the IOUS training include the improvements and adjustments mentioned for the training scenarios as well as for the US simulation. After this, the insights from the gamification project can be incorporated. This comprises the kit interaction for training scenario three. In close cooperation with the medical experts, a proper level design and scoring system have to be developed. Besides improving the existing parts, such as the microscopic view, the microsurgical training requires feedback and performance assessment. Again in consultation with medical experts, the described three criteria have to be implemented. To provide complete training applications, suitable training cases–for both applications–have to be prepared and included. Suitable cases means diverse, representative and a proper amount. Furthermore, Mönch [209] considers a surgical training system as successful, if the following three factors are given:

- Effective imparting of knowledge: Because medical students and physicians have a stressful daily routine, it is important to learn and train efficiently in a short time. Because the presented training systems address aspects for which no or little training possibilities are available, the content itself is relevant. Although the learning outcome was not evaluated, the systems could be considered as effective because of their relevance. Efficiency was also not assessed, however, a VR training system does not require a large setup.

- Motivation: In the presented gamification project, motivation was addressed. As assumed in this project, the new technology might lead to a high motivation at the beginning. For long-term motivation, gamification might be helpful, however, this was not assessed in a longitudinal study.
- Users have to trust the system regarding correctness and relevance: The applications were developed in close cooperation with medical experts ensuring correctness and relevance. If these experts use the system for their students, they will probably trust it. To increase trust, larger evaluations showing the effectiveness or incorporating the experts visibly might be possible. The latter can be done by, for example, including a virtual physician as an expert.

After all necessary and possible improvements are incorporated and the applications are well evaluated, including a longitudinal study revealing positive learning effects, the training applications have to be integrated into the curriculum. For example, the University of Münster included a VR-based training for brain death diagnosis into their curriculum to compare it with phantom training [143]. The integration requires a proper setup: several rooms equipped with HMDs and a monitoring room from where the medical students are observed and guided if needed. However, Jiang et al. [138] state that only few studies reported on the integration of VR-based training into the curriculum. Accordingly, guidance and information on how to adopt these training systems in the medical curriculum have to be established.

The presented prototypes and other related work have shown that VR has potential for medical education and training. As mentioned in the introduction, the prototypes are not meant to replace existing methods. Instead, the respective learning and training objectives have to be clearly identified to successfully combine the different modalities. Through additional VR-based training systems trainees can explore and learn independently, which also relieves the medical staff. Besides the logical and effective order of training contents and modalities, a proper technological infrastructure is required. From the above-mentioned example from the University of Münster, one can derive that several rooms equipped with HMDs and (technical) support are required. Because HMDs are further developed and become more affordable as well as comfortable, the technological effort will probably be reduced in the future.

List of Publications

[7] M. Allgaier, A. Amini, B. Neyazi, I. E. Sandalcioglu, B. Preim, and S. Saalfeld. “VR-based training of craniotomy for intracranial aneurysm surgery”. In: *Int J Comput Assist Radiol Surg* 17 (2022), pp. 449–456. DOI: 10.1007/s11548-021-02538-3.

[8] M. Allgaier, T. J. Härtel, S. Zubel, M. Thormann, D. Behme, B. Preim, and S. Saalfeld. “Gamified Virtual Reality Training for Visuospatial Ability in Neuroradiology”. In: *Curr Dir Biomed Eng* 10.2 (2024), pp. 25–28. DOI: doi:10.1515/cdbme-2024-1058.

[9] M. Allgaier, F. Huettl, L. I. Hanke, T. Huber, B. Preim, S. Saalfeld, and C. Hansen. “LiVRSono-Virtual Reality Training with Haptics for Intraoperative Ultrasound”. In: *Proc. of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2023, pp. 980–989. DOI: 10.1109/ISMAR59233.2023.00114.

[10] M. Allgaier, F. Huettl, L. I. Hanke, T. Huber, B. Preim, S. Saalfeld, and C. Hansen. “Gamification Concepts for a VR-based Visuospatial Training for Intraoperative Liver Ultrasound”. In: *Proc. of CHI EA ’24: Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–8. DOI: 10.1145/3613905.3650736.

[12] M. Allgaier, B. Neyazi, I. E. Sandalcioglu, B. Preim, and S. Saalfeld. “Immersive VR training system for clipping intracranial aneurysms”. In: *Curr Dir Biomed Eng* 8.1 (2022), pp. 9–12.

[13] M. Allgaier*, V. Chheang*, P. Saalfeld, V. Apilla, T. Huber, F. Huettl, B. Neyazi, I. E. Sandalcioglu, C. Hansen, B. Preim, and S. Saalfeld. “A Comparison of Input Devices for Precise Interaction Tasks in VR-Based Surgical Planning and Training”. In: *Comput Biol Med* 145 (2022). *These authors contributed equally. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2022.105429.

[14] M. Allgaier*, V. Chheang*, P. Saalfeld, V. Apilla, T. Huber, F. Huettl, B. Neyazi, I. E. Sandalcioglu, C. Hansen, B. Preim, and S. Saalfeld. “A Comparison of Input Devices for Precise Interaction Tasks in VR-based Surgical Planning and Training”. In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. *These authors contributed equally. 2022, pp. 674–675. DOI: 10.1109/VRW55335.2022.00188.

[15] M. Allgaier*, L. Spitz*, D. Behme, A. Mpotsaris, P. Berg, B. Preim, and S. Saalfeld. “Design of a virtual data shelf to effectively explore a large database of 3D medical surface models in VR”. In: *Int J Comput Assist Radiol Surg* 18 (2023). *These authors contributed equally. DOI: 10.1007/s11548-023-02851-z.

[130] F. Huettl, M. Allgaier, S. Saalfeld, L. Hanke, B. Preim, H. Lang, C. Hansen, and T. Huber. "Liver IOUS 2.0: Training and Gaming in virtueller Realität". In: *Z Gastroenterol.* Vol. 61. 2023, e602. DOI: `10.1055/s-0043-1772139`.

[293] L. Spitz, M. Allgaier, A. Mpotsaris, D. Behme, B. Preim, and S. Saalfeld. "Segmentation of Circle of Willis from 7T TOF-MRI data and immersive exploration using VR". In: *Curr Dir Biomed Eng* 8.1 (2022), pp. 129–132. DOI: `doi:10.1515/cdbme-2022-0033`.

References

[1] F. M. Abu-Zidan, A. F. Hefny, and P. Corr. “Clinical ultrasound physics”. In: *J Emerg Trauma Shock* 4.4 (2011), pp. 501–503. DOI: 10.4103/0974-2700.86646.

[2] H. Adams, J. Shinn, W. G. Morrel, Noble J., and 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. DOI: 10.1117/12.2506178.

[3] R. B. Adams. “Ultrasound scanning techniques”. In: *Surg Open Science* 10 (2022), pp. 182–207. DOI: 10.1016/j.sopen.2022.09.002.

[4] A. Alaraj, C. Luciano, D. Bailey, A. Elsenousi, B. Roitberg, A. Bernardo, P. Banerjee, and F. Charbel. “Virtual Reality Cerebral Aneurysm Clipping Simulation With Real-Time Haptic Feedback”. In: *Neurosurg* 11.2 (Jan. 2015), pp. 52–58. DOI: 10.1227/NEU.0000000000000583.

[5] M. M. Alhammad and A. M. Moreno. “Gamification in software engineering education: A systematic mapping”. In: *J Sys Softw* 141 (2018), pp. 131–150. ISSN: 0164-1212. DOI: https://doi.org/10.1016/j.jss.2018.03.065.

[6] M. Allgaier. “Distance and Force Visualisations for Aneurysm Clipping Simulations”. Masterthesis. Dept. of Computer Science, Oct. 14, 2020. URL: https://www.vismd.de/wp-content/uploads/2022/08/Masterthesis_Allgaier.pdf.

[7] M. Allgaier, A. Amini, B. Neyazi, I. E. Sandalcioglu, B. Preim, and S. Saalfeld. “VR-based training of craniotomy for intracranial aneurysm surgery”. In: *Int J Comput Assist Radiol Surg* 17 (2022), pp. 449–456. DOI: 10.1007/s11548-021-02538-3.

[8] M. Allgaier, T. J. Härtel, S. Zubel, M. Thormann, D. Behme, B. Preim, and S. Saalfeld. “Gamified Virtual Reality Training for Visuospatial Ability in Neuroradiology”. In: *Curr Dir Biomed Eng* 10.2 (2024), pp. 25–28. DOI: doi:10.1515/cdbme-2024-1058.

[9] M. Allgaier, F. Huettl, L. I. Hanke, T. Huber, B. Preim, S. Saalfeld, and C. Hansen. “LiVRSono-Virtual Reality Training with Haptics for Intraoperative Ultrasound”. In: *Proc. of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 2023, pp. 980–989. DOI: 10.1109/ISMAR59233.2023.00114.

[10] M. Allgaier, F. Huettl, L. I. Hanke, T. Huber, B. Preim, S. Saalfeld, and C. Hansen. “Gamification Concepts for a VR-based Visuospatial Training for Intraoperative Liver Ultrasound”. In: *Proc. of CHI EA ’24: Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–8. DOI: 10.1145/3613905.3650736.

[11] M. Allgaier, B. Neyazi, B. Preim, and S. Saalfeld. "Distance and Force Visualisations for Improved Simulation of Intracranial Aneurysm Clipping". In: *Int J Comput Assist Radiol Surg* 16 (2021). DOI: 10.1007/s11548-021-02413-1.

[12] M. Allgaier, B. Neyazi, I. E. Sandalcioglu, B. Preim, and S. Saalfeld. "Immersive VR training system for clipping intracranial aneurysms". In: *Curr Dir Biomed Eng* 8.1 (2022), pp. 9–12.

[13] M. Allgaier*, V. Chheang*, P. Saalfeld, V. Apilla, T. Huber, F. Huettl, B. Neyazi, I. E. Sandalcioglu, C. Hansen, B. Preim, and S. Saalfeld. "A Comparison of Input Devices for Precise Interaction Tasks in VR-Based Surgical Planning and Training". In: *Comput Biol Med* 145 (2022). *These authors contributed equally. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2022.105429.

[14] M. Allgaier*, V. Chheang*, P. Saalfeld, V. Apilla, T. Huber, F. Huettl, B. Neyazi, I. E. Sandalcioglu, C. Hansen, B. Preim, and S. Saalfeld. "A Comparison of Input Devices for Precise Interaction Tasks in VR-based Surgical Planning and Training". In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* *These authors contributed equally. 2022, pp. 674–675. DOI: 10.1109/VRW55335.2022.00188.

[15] M. Allgaier*, L. Spitz*, D. Behme, A. Mpotsaris, P. Berg, B. Preim, and S. Saalfeld. "Design of a virtual data shelf to effectively explore a large database of 3D medical surface models in VR". In: *Int J Comput Assist Radiol Surg* 18 (2023). *These authors contributed equally. DOI: 10.1007/s11548-023-02851-z.

[16] Z. Aminolroaya, S. Wiebe, W. Willett, C. B. Josephson, G. McLeod, and F. Maurer. "Streamlining Epilepsy Surgery Planning Rounds with Virtual Reality". In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 2023, pp. 161–165. DOI: 10.1109/VRW58643.2023.00040.

[17] N. L. Andersen, R. O. Jensen, S. Posth, C. B. Laursen, R. Jørgensen, and O. Graumann. "Teaching ultrasound-guided peripheral venous catheter placement through immersive virtual reality: An explorative pilot study". In: *Med* 100.27 (2021), e26394. DOI: 10.1097/MD.0000000000026394.

[18] C. Anthes, R. J. García-Hernández, M. Wiedemann, and D. Kranzlmüller. "State of the art of virtual reality technology". In: *Proc. of IEEE Aerospace Conference.* 2016, pp. 1–19. DOI: 10.1109/AERO.2016.7500674.

[19] K. Atli, W. Selman, and A. Ray. "A Comprehensive Multicomponent Neurosurgical Course with use of Virtual Reality: Modernizing the Medical Classroom". In: *J Surg Educ* 78.4 (2021), pp. 1350–1356. DOI: 10.1016/j.jsurg.2020.11.003.

[20] Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe AWMF). *S3-Leitlinie Diagnostik und Therapie des Hepatozellulären Karzinoms und biliärer Karzinome, Langversion 2.0, AWMF Registernummer: 032/053OL.* Access: 23/08/22. 2021. URL: https://www.leitlinienprogramm-onkologie.de/fileadmin/user_upload/Downloads/Leitlinien/HCC/Version_2/LL_HCC_bili%C3%A4re_Karzinome_Langversion_2.0.pdf.

[21] R. Bahar, R. Taki, and S. Yalcýn. "Caregiver: An Application for The First Step in Alzheimer's Disease Early Diagnosis". In: *Proc. of International Conference on Intelligent User Interfaces.* 2022, pp. 34–37. DOI: 10.1145/3490100.3516468.

[22] S. Banerjee, T. Pham, A. Eastaway, W. F. Auffermann, and E. P. Quigley. “The Use of Virtual Reality in Teaching Three-Dimensional Anatomy and Pathology on CT”. In: *J Digit Imaging* 36 (2023), pp. 1279–1284. DOI: 10.1007/s10278-023-00784-2.

[23] C. Barnouin, F. Zara, and F. Jaillet. “A Real-time Ultrasound Rendering with Model-based Tissue Deformation for Needle Insertion”. In: *Proc. of Computer Graphics Theory and Applications (GRAPP)*. 2020, pp. 235–246. DOI: 10.5220/0008947302350246.

[24] A. U. Batmaz, A. K. Mutasim, and 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. DOI: 10.1109/VRW50115.2020.00012.

[25] A. U. Batmaz, M. R. Seraji, J. Kneifel, and W. Stuerzlinger. “No Jitter Please: Effects of Rotational and Positional Jitter on 3D Mid-Air Interaction”. In: *Proc. of the Future Technologies Conference (FTC), Volume 2*. 2021, pp. 792–808. DOI: 10.1007/978-3-030-63089-8_52.

[26] W. L. Bedwell, D. Pavlas, K. Heyne, E. H. Lazzara, and E. Salas. “Toward a Taxonomy Linking Game Attributes to Learning: An Empirical Study”. In: *Simul and Gaming* 43.6 (2012), pp. 729–760. DOI: 10.1177/1046878112439444.

[27] R. Belavadi, S. V. R. Gudigopuram, C. C. Raguthu, H. Gajjela, I. Kela, C. L. Kakarala, M. Hassan, and I. Sange. “Surgical Clipping Versus Endovascular Coiling in the Management of Intracranial Aneurysms”. In: *Cureus* 13.12 (2021). DOI: doi:10.7759/cureus.20478.

[28] J. Belger, M. Blume, M. Akbal, P. Chojecki, J. de Mooij, M. Gaebler, F. Klotzsche, S. Krohn, M. T. Lafci, E. Quinque, J. Tromp, A. Villringer, C. Finke, and A. Thöne-Otto. “The immersive virtual memory task: Assessing object-location memory in neurological patients using immersive virtual reality”. In: *Neuropsychol Rehabil* 0.0 (2023), pp. 1–29. DOI: 10.1080/09602011.2023.2256957.

[29] J. Belghiti, F. Cauchy, V. Paradis, and V. Vilgrain. “Diagnosis and management of solid benign liver lesions”. In: *Nat Rev Gastroenterol Hepatol* 11.12 (2014), pp. 737–749. ISSN: 1759-5053. DOI: 10.1038/nrgastro.2014.151.

[30] Y. Benjamini and Y. Hochberg. “Controlling the false discovery rate: a practical and powerful approach to multiple testing”. In: *J R Stat Soc Ser B Methodol* 57.1 (1995), pp. 289–300. DOI: 10.1111/j.2517-6161.1995.tb02031.x.

[31] H. Berg, E. Prasolova-Førland, and A. Steinsbekk. “Developing a virtual reality (VR) application for practicing the ABCDE approach for systematic clinical observation”. In: *BMC Med Educ* 23.1 (2023), p. 639. DOI: 10.1186/s12909-023-04625-2.

[32] J. Boutin, J. Kamoonpuri, R. Faieghi, J. Chung, S. de Ribaupierre, and R. Eagleson. “Smart haptic gloves for virtual reality surgery simulation: a pilot study on external ventricular drain training”. In: *Front Robot AI* 10 (2024). DOI: 10.3389/frobt.2023.1273631.

[33] W. E. Brant and C. A. Helms. *Fundamentals of Diagnostic Radiologys*. 3rd ed. Lippincott Williams & Wilkins, 2007. ISBN: 978-0-7817-6135-2.

[34] P. Bräuer and A. Mazarakis. “Badges or a leaderboard? How to gamify an augmented reality warehouse setting”. In: *Proc. of GamiFIN Conference*. 2019, pp. 229–240.

[35] J. Brooke. “SUS: A quick and dirty usability scale”. In: *Usability Eval. Ind.* 189 (1995).

[36] D. M. Bruening, P. Truckenmueller, C. Stein, J. Fuellhase, P. Vajkoczy, T. Picht, and G. Acker. “360° 3D virtual reality operative video for the training of residents in neurosurgery”. In: *Neurosurg Focus* 53.2 (2022), E4. DOI: 10.3171/2022.5.FOCUS2261.

[37] T. Bublak, M. Bofferding, C. Olson, H. Jonas, J. Rademeier, and C. Hansen. “A Virtual Environment for Emergency Ultrasound Training during Cardiopulmonary Resuscitation”. In: *Proc. of Mensch und Computer*. ACM, 2022, pp. 600–602. DOI: 10.1145/3543758.3547516.

[38] I. Buchem, S. Vorwerg, O. Stamm, K. Hildebrand, and Y. Bialek. “Gamification in Mixed-Reality Exergames for Older Adult Patients in a Mobile Immersive Diagnostic Center: A Pilot Study in the BewARe Project”. In: *Proc. of International Conference of the Immersive Learning Research Network (iLRN)*. 2021, pp. 1–8. DOI: 10.23919/iLRN52045.2021.9459337.

[39] K. Bucher, T. Blome, S. Rudolph, and S. von Mammen. “VReanimate II: training first aid and reanimation in virtual reality”. In: *Comput Educ J* 6.1 (2019), pp. 53–78. DOI: 10.1007/s40692-018-0121-1.

[40] J. F. Buell, D. Cherqui, D. A. Geller, N. O’Rourke, D. Iannitti, I. Dagher, A. J. Koffron, M. Thomas, B. Gayet, H. S. Han, and et al. “The international position on laparoscopic liver surgery: The Louisville Statement, 2008”. In: *Ann Surg* 250.5 (2009), pp. 825–830. ISSN: 0003-4932. DOI: 10.1097/sla.0b013e3181b3b2d8.

[41] R. Buettner, H. Baumgartl, T. Konle, and P. Haag. “A Review of Virtual Reality and Augmented Reality Literature in Healthcare”. In: *Proc. of IEEE Symposium on Industrial Electronics & Applications (ISIEA)*. 2020, pp. 1–6. DOI: 10.1109/ISIEA49364.2020.9188211.

[42] J. Bui, J. Luauté, and A. Farnè. “Enhancing Upper Limb Rehabilitation of Stroke Patients With Virtual Reality: A Mini Review”. In: *Front Virtual Real* 2 (2021). DOI: 10.3389/frvir.2021.595771.

[43] B. Burger, S. Bettinghausen, M. Radle, and J. Hesser. “Real-Time GPU-Based Ultrasound Simulation Using Deformable Mesh Models”. In: *IEEE Trans Med Imaging* 32.3 (2013), pp. 609–618. DOI: 10.1109/TMI.2012.2234474.

[44] B. Byl, M. Süncksen, and M. Teistler. “A serious virtual reality game to train spatial cognition for medical ultrasound imaging”. In: *Proc. of Serious Games and Applications for Health (SeGAH)*. 2018, pp. 1–4. DOI: 10.1109/SeGAH.2018.8401365.

[45] A. Cannavò, D. Calandra, A. Kehoe, and F. Lamberti. “Evaluating Consumer Interaction Interfaces for 3D Sketching in Virtual Reality”. In: *Proc. of Interactivity and Game Creation*. 2021, pp. 291–306. ISBN: 978-3-030-73426-8. DOI: 10.1007/978-3-030-73426-8_17.

[46] F. J. Carter, M. P. Schijven, R. Aggarwal, T. Grantcharov, N. K. Francis, G. B. Hanna, and J. J. Jakimowicz. "Consensus guidelines for validation of virtual reality surgical simulators". In: *Surg Endosc* 19.12 (2005), pp. 1523–1532. DOI: 10.1007/s00464-005-0384-2.

[47] J. M. Chambers, A. E. Freeny, and R. M. Heiberger. "Analysis of variance; designed experiments". In: *Chapter 5 of Statistical Models in S.* Wadsworth & Brooks/Cole, 1992, pp. 145–190. ISBN: 9780203738535.

[48] J. Chan, D. J. Pangal, T. Cardinal, G. Kugener, Y. Zhu, A. Roshannai, N. Markarian, A. Sinha, A. Anandkumar, A. Hung, G. Zada, and D. A. Donoho. "A systematic review of virtual reality for the assessment of technical skills in neurosurgery". In: *Neurosurg Focus* 51.2 (2021), E15. DOI: 10.3171/2021.5.FOCUS21210.

[49] T. Chan, R. P. Gauthier, A. Suarez, N. F. Sia, and J. R. Wallace. "Merlynne: Motivating Peer-to-Peer Cognitive Behavioral Therapy with a Serious Game". In: *Proc. ACM Hum.-Comput. Interact.* Vol. 5. CHI PLAY. 2021. DOI: 10.1145/3474677.

[50] Y.-J. Chang, Y.-S. Kang, Y.-S. Chang, H.-H. Liu, C.-C. Wang, and C.-C. Kao. "Designing Kinect2Scratch Games to Help Therapists Train Young Adults with Cerebral Palsy in Special Education School Settings". In: *Proc. of the International ACM SIGACCESS Computers Conference on Computers and Accessibility.* 2015, pp. 317–318. DOI: 10.1145/2700648.2811356.

[51] L. Chen, H. Liao, W. Kong, D. Zhang, and F. Chen. "Anatomy preserving GAN for realistic simulation of intraoperative liver ultrasound images". In: *Comput Methods Programs Biomed* 240 (2023), p. 107642. DOI: 10.1016/j.cmpb.2023.107642.

[52] V. Chheang, V. Apilla, P. Saalfeld, C. Boedecker, T. Huber, F. Huettl, H. Lang, B. Preim, and 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, pp. 768–768. DOI: 10.1109/VRW52623.2021.00268.

[53] V. Chheang, R. Bruggernann, B. Preim, and C. Hansen. "Virtual Resection Planning using Bezier Surface Interactions in Collaborative VR Environments". In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 2023, pp. 166–169. DOI: 10.1109/VRW58643.2023.00041.

[54] V. Chheang, V. Fischer, H. Buggenhagen, T. Huber, F. Huettl, W. Kneist, B. Preim, P. Saalfeld, and C. Hansen. "Toward interprofessional team training for surgeons and anesthesiologists using virtual reality". In: *Int J Comput Assist Radiol Surg* 15.12 (2020), pp. 2109–2118. DOI: 10.1007/s11548-020-02276-y.

[55] V. Chheang, P. Saalfeld, T. Huber, F. Huettl, W. Kneist, B. Preim, and C. Hansen. "Collaborative Virtual Reality for Laparoscopic Liver Surgery Training". In: *Proc. of IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR).* 2019, pp. 1–17. DOI: 10.1109/AIVR46125.2019.00011.

[56] V. Chheang, P. Saalfeld, F. Joeres, C. Boedecker, T. Huber, F. Huettl, H. Lang, B. Preim, and C. Hansen. "A collaborative virtual reality environment for liver surgery planning". In: *Comput Graph* 99 (2021), pp. 234–246. ISSN: 0097-8493. DOI: 10.1016/j.cag.2021.07.009.

[57] V. Chheang, D. Schott, P. Saalfeld, L. Vradelis, T. Huber, F. Huettl, H. Lang, B. Preim, and C. Hansen. “Advanced liver surgery training in collaborative VR environments”. In: *Comput & Graph* (2024). ISSN: 0097-8493. DOI: 10.1016/j.cag.2024.01.006.

[58] Y.-K. Chou. *Actionable Gamification: Beyond Points, Badges and Leaderboards.* CreateSpace Independent Publishing Platform, 2015. ISBN: 9780692858905.

[59] A. Chuan, J. Qian, A. Bogdanovych, A. Kumar, M. McKendrick, and G. McLeod. “Design and validation of a virtual reality trainer for ultrasound-guided regional anaesthesia”. In: *Anaesth* 78.6 (2023), pp. 739–746. DOI: 10.1111/anae.16015.

[60] A. Chuan, J. Zhou, R. Hou, C. Stevens, and A. Bogdanovych. “Virtual reality for acute and chronic pain management in adult patients: a narrative review”. In: *Anesth* 76 (July 2020). DOI: 10.1111/anae.15202.

[61] J. A. Cleland, K. Abe, and J.-J. Rethans. “The use of simulated patients in medical education: AMEE Guide No 42”. In: *Med Teach* 31.6 (2009), pp. 477–486. DOI: 10.1080/01421590903002821.

[62] D. Coco and S. Leanza. “Routine Intraoperative Ultrasound for the Detection of Liver Metastases during Resection of Primary Colorectal Cancer - A Systematic Review”. In: *Maedica* 15.2 (2020), pp. 250–252. ISSN: 1841-9038. DOI: 10.26574/maedica.2020.15.2.250.

[63] F. F. Coelho, W. M. Bernardo, J. A. P. Kruger, V. B. Jeismann, G. M. Fonseca, R. L. Macacari, D. M. Cesconetto, L. A. C. D’Albuquerque, I. Cecconello, and P. Herman. “Laparoscopy-assisted versus open and pure laparoscopic approach for liver resection and living donor hepatectomy: a systematic review and meta-analysis”. In: *HPB* 20.8 (2018), pp. 687–694. ISSN: 1365-182X. DOI: 10.1016/j.hpb.2018.02.379.

[64] C. A. Cohen and M. Hegarty. “Inferring cross sections of 3D objects: A new spatial thinking test”. In: *Learning and Individual Differences* 22.6 (2012), pp. 868–874. DOI: 10.1016/j.lindif.2012.05.007.

[65] T. Combe, J.-R. Chardonnet, F. Merienne, and J. Ovtcharova. “CAVE and HMD: distance perception comparative study”. In: *Virtual Real* (2023). DOI: 10.1007/s10055-023-00787-y.

[66] A. Cooper. “The Inmates are Running the Asylum”. In: *Software-Ergonomie ’99: Design von Informationswelten.* Ed. by E. Arend, E. Eberleh, and K. Pitschke. Vieweg+Teubner Verlag, 1999, pp. 17–17. ISBN: 978-3-322-99786-9. DOI: 10.1007/978-3-322-99786-9_1.

[67] A. Cooper, R. Reimann, and D. Cronin. *About Face 3: The Essentials of Interaction Design.* 3rd. Wiley, 2007. ISBN: 978-0470084113.

[68] P. Crochet, R. Aggarwal, S. Dubb, P. Ziprin, N. Rajaretnam, T. Grantcharov, K. Ericsson, and A. Darzi. “Deliberate Practice on a Virtual Reality Laparoscopic Simulator Enhances the Quality of Surgical Technical Skills”. In: *Ann Surg* 253 (2011), pp. 1216–22. DOI: 10.1097/SLA.0b013e3182197016.

[69] W. K. Crockatt, J. E. Confino, N. J. Kopydlowski, C. M. Jobin, and W. N Levine. “Comparing Skill Acquisition and Validity of Immersive Virtual Reality with Cadaver Laboratory Sessions in Training for Reverse Total Shoulder Arthroplasty”. In: *JB JS Open Access* 8.3 (2023), e22.00141. DOI: 10.2106/JBJS.OA.22.00141.

[70] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R: V. Kenyon, and J. C. Hart. "The CAVE: Audio Visual Experience Automatic Virtual Environment". In: *Commun ACM* 35.6 (1992), pp. 64–72. ISSN: 0001-0782. DOI: 10.1145/129888.129892.

[71] D. d'Aulignac, C. Laugier, and M.C. Cavusoglu. "Towards a realistic echographic simulator with force feedback". In: *Proc. of IEEE/RSJ Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients*. Vol. 2. 1999, 727–732 vol.2. DOI: 10.1109/IROS.1999.812766.

[72] V. Damerjian, O. Tankyevych, N. Souag, and E. Petit. "Speckle characterization methods in ultrasound images – A review". In: *IRBM* 35.4 (2014), pp. 202–213. DOI: 10.1016/j.irbm.2014.05.003.

[73] S. Deterding, D. Dixon, R. Khaled, and L. Nacke. "From Game Design Elements to Gamefulness: Defining Gamification". In: *Proc. of International Academic MindTrek Conference: Envisioning Future Media Environments*. 2011, pp. 9–15. DOI: 10.1145/2181037.2181040.

[74] L. Dixon, A. Lim, M. Grech-Sollars, D. Nandi, and S. Camp. "Intraoperative ultrasound in brain tumor surgery: A review and implementation guide". In: *Neurosurg Rev* 45.4 (2022), pp. 2503–2515. DOI: 10.1007/s10143-022-01778-4.

[75] R. Dörner, W. Broll, P. Grimm, and B. Jung. *Virtual und Augmented Reality (VR/AR) - Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. 2nd ed. Springer Vieweg Berlin, 2019. DOI: 10.1007/978-3-662-58861-1.

[76] C. A. Duarte-Salazar, A. E. Castro-Ospina, M. A. Becerra, and E. Delgado-Trejos. "Speckle Noise Reduction in Ultrasound Images for Improving the Metrological Evaluation of Biomedical Applications: An Overview". In: *IEEE Access* 8 (2020), pp. 15983–15999. DOI: 10.1109/ACCESS.2020.2967178.

[77] O. Elhage, B. Challacombe, A. Shortland, and 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". In: *BJU Int* 115 (2014), pp. 274–281. DOI: 10.1111/bju.12680.

[78] P. M. G. Emmelkamp and K. Meyerbröker. "Virtual Reality Therapy in Mental Health". In: *Annu Rev Clin Psychol* 17.1 (2021), pp. 495–519. DOI: 10.1146/annurev-clinpsy-081219-115923.

[79] G. Esposito, S. Dias, J.-K. Burkhardt, J. Fierstra, C. Serra, O. Bozinov, and L. Regli. "Selection strategy for optimal keyhole approaches for MCA aneurysms: lateral supraorbital versus minipterional craniotomy". In: *World Neurosurg* 122 (Nov. 2018), e349–e357. DOI: 10.1016/j.wneu.2018.09.238.

[80] J. H. D. Fasel and A. Schenk. "Concepts for Liver Segment Classification: Neither Old Ones nor New Ones, but a Comprehensive One". In: *J Clin Imaging Sci* 3 (2013), p. 48. DOI: 10.4103/2156-7514.120803.

[81] F. Faure, C. Duriez, H. Delingette, J. Allard, B. Gilles, S. Marchesseau, H. Talbot, H. Courtecuisse, G. Bousquet, I. Peterlik, and S. Cotin. "SOFA: A Multi-Model Framework for Interactive Physical Simulation". In: *Soft Tissue Biomechanical Modeling for Computer Assisted Surgery*. 2012, pp. 283–321. DOI: 10.1007/8415_2012_125.

[82] E. S. Felekouras, D. C. Kaparelos, and E. Papalambros. “The history of liver surgery, hepatectomy and haemostasis”. In: *Hellenic J Surg* 82 (2010), pp. 280–296. DOI: 10.1007/s13126-010-0046-2.

[83] W. Fenz and J. Dirnberger. “Real-time surgery simulation of intracranial aneurysm clipping with patient-specific geometries and haptic feedback”. In: *Proc of SPIE Medical Imaging: Image-Guided Procedures, Robotic Interventions, and Modeling.* Vol. 9415. 2015, 94150H. DOI: 10.1117/12.2082053.

[84] R. H. Flin, P. O’Connor, and M. Crichton. *Safety at the Sharp End: A Guide to Non-technical Skills*. CRC Press, 2017. ISBN: 9781317059950.

[85] C. L. Foronda, L. Gonzales, M. M. Meese, N. Slamon, M. Baluyot, J. Lee, and M. Aebersold. “A Comparison of Virtual Reality to Traditional Simulation in Health Professions Education A Systematic Review”. In: *Simul Healthc* (2023). DOI: 10.1097/SIH.0000000000000745.

[86] J. G. Frederiksen, S. M. D. Sørensen, L. Konge, M. B. S. Svendsen, M. Nobel-Jørgensen, F. Bjerrum, and S. W. W. Andersen. “Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial”. In: *Surg Endosc* 34 (2020), pp. 1244–1252. DOI: 10.1007/s00464-019-06887-8.

[87] J. R. S. Freitas, V. H. S. Velosa, L. T. N. Abreu, R. L. Jardim, J. A. V. Santos, B. Peres, and P. F. Campos. “Virtual Reality Exposure Treatment in Phobias: a Systematic Review”. In: *Psychiatr Q* 92 (2021), pp. 1685–1710. DOI: 10.1007/s11126-021-09935-6.

[88] T. H. Frøland, I. Heldal, E. Ersvær, and G. Sjøholt. “Merging 360°-videos and Game-Based Virtual Environments for Phlebotomy Training: Teachers and Students View”. In: *Proc. of International Conference on e-Health and Bioengineering (EHB).* 2021, pp. 1–6. DOI: 10.1109/EHB52898.2021.9657611.

[89] M. Gabele. “Development and Design of Software-Based Methods to Promote Motivation of Patients in Cognitive Rehabilitation”. PhD thesis. Otto-von-Guericke-University Magdeburg, 2023.

[90] M. Gabele, J. Weicker, S. Wagner, A. Thoms, S. Hußlein, and C. Hansen. “Effects and Ways of Tailored Gamification in Software-Based Training in Cognitive Rehabilitation”. In: *Proc. of the ACM Conference on User Modeling, Adaptation and Personalization.* 2021, pp. 158–168. DOI: 10.1145/3450613.3456828.

[91] C. M. Gallegos, A. J. Tesar, K. Connor, and K. Martz. “The use of a game-based learning platform to engage nursing students: A descriptive, qualitative study”. In: *Nurse Educ Pract* 27 (2017), pp. 101–106.

[92] M. Gallucci, S. Capoccia, and A. Catalucci. *Radiographic Atlas of the Skull and Brain Anatomy.* Springer-Verlag, 2005.

[93] H. Gao, H.F. Choi, P. Claus, S. Boonen, S. Jaecques, G. Harry Van Lenthe, G. Van Der Perre, W. Lauriks, and J. D’hooge. “A fast convolution-based methodology to simulate 2-D/3-D cardiac ultrasound images”. In: *IEEE Trans Ultrason Ferroelectr Freq Control* 56.2 (2009), pp. 404–409. DOI: 10.1109/TUFFC.2009.1051.

[94] N. V. Garcia Hernandez, S. Buccelli, M. Laffranchi, and L. de Michieli. “Mixed Reality-Based Exergames for Upper Limb Robotic Rehabilitation”. In: *Proc. of Companion of the ACM/IEEE International Conference on Human-Robot Interaction.* 2023, pp. 447–451. DOI: 10.1145/3568294.3580124.

[95] J. F. Gerstenmaier and R. N. Gibson. “Ultrasound in chronic liver disease”. In: *Insights Imaging* 5.4 (2014), pp. 441–455. ISSN: 1869-4101. DOI: 10.1007/s13244-014-0336-2.

[96] M. Gmeiner, J. Dirnberger, W. Fenz, M. Gollwitzer, G. Wurm, J. Trenkler, and A. Gruber. “Virtual Cerebral Aneurysm Clipping with Real-Time Haptic Force Feedback in Neurosurgical Education”. In: *World Neurosurg* 112 (2018), e313–e323. DOI: 10.1016/j.wneu.2018.01.042.

[97] O. Goksel and S. E. Salcudean. “B-Mode Ultrasound Image Simulation in Deformable 3-D Medium”. In: *IEEE Trans Med Imaging* 28.11 (2009), pp. 1657–1669. DOI: 10.1109/TMI.2009.2016561.

[98] E. A. Gomez, D. Wu, and K. Passerini. “Computer-supported team-based learning: The impact of motivation, enjoyment and team contributions on learning outcomes”. In: *Computers & Education* 55.1 (2010), pp. 378–390. DOI: doi.org/10.1016/j.compedu.2010.02.003..

[99] J. Gonzalez, C. Pham, A. Umejiaku, J. Benjamin, and T. Dang. “SkeletonVR: Educating Human Anatomy Through An Interactive Puzzle Assembly”. In: *Proc. of the International Conference on Advances in Information Technology.* 2021. DOI: 10.1145/3468784.3471605.

[100] M. S. Greenberg. *Handbook of Neurosurgery.* 9. Thieme, 2019. ISBN: 9781684201372.

[101] S. Gue, J. Ray, and L. Ganti. “Gamification of graduate medical education in an emergency medicine residency program”. In: *Int J Emerg Med* 15.41 (2022). DOI: 10.1186/s12245-022-00445-1.

[102] A. Gupta, J. Cecil, and MigM.uel Pirela-Cruz. “Immersive Virtual Reality based Training and Assessment of an Orthopedic Surgical Process”. In: *Proc. of IEEE International Systems Conference (SysCon).* 2020, pp. 1–7. DOI: 10.1109/SysCon47679.2020.9381832.

[103] R. Hadi Mogavi, B. Guo, Y. Zhang, E.-U. Haq, P. Hui, and X. Ma. “When Gamification Spoils Your Learning: A Qualitative Case Study of Gamification Misuse in a Language-Learning App”. In: *Proc. of the Ninth ACM Conference on Learning @ Scale.* 2022, pp. 175–188. DOI: 10.1145/3491140.3528274.

[104] E. J. Hagopian. “Liver ultrasound: A key procedure in the surgeon’s toolbox”. In: *J Surg Oncol* 122.1 (2020), pp. 61–69. DOI: 10.1002/jso.25908.

[105] A. Haider, C. Harteveld, D. Johnson, M. V. Birk, R. L. Mandryk, M. Seif El-Nasr, L. E. Nacke, K. Gerling, and V. Vanden Abeele. “MiniPXI: Development and Validation of an Eleven-Item Measure of the Player Experience Inventory”. In: *Proc. ACM Hum.-Comput. Interact.* 6.CHI PLAY (2022). DOI: 10.1145/3549507.

[106] A. Halbig, S. K. Babu, S. Gatter, M. E. Latoschik, K. Brukamp, and S. von Mammen. “Opportunities and Challenges of Virtual Reality in Healthcare – A Domain Experts Inquiry”. In: *Front Virtual Real* 3 (2022). DOI: 10.3389/frvir.2022.837616.

[107] T. Halic, S. Kockara, C. Bayrak, R. Rowe, and B. Chen. “Soft Tissue Deformation and Optimized Data Structures for Mass Spring Methods”. In: *Proc of IEEE International Conference on Bioinformatics and BioEngineering*. 2009, pp. 45–52. DOI: 10.1109/BIBE.2009.64.

[108] S. Hallifax, A. Serna, J.-C. Marty, G. Lavoué, and E. Lavoué. “Factors to Consider for Tailored Gamification”. In: *Proc. of CHI PLAY*. 2019, pp. 559–572. DOI: 10.1145/3311350.3347167.

[109] J. Hamari, J. Koivisto, and H. Sarsa. “Does Gamification Work? – A Literature Review of Empirical Studies on Gamification”. In: *Proc. of Hawaii International Conference on System Sciences*. 2014, pp. 3025–3034. DOI: 10.1109/HICSS.2014.377.

[110] C. Hansen, M. Engelhardt, B. Brendel, S. Winter, J. Eydin, K. Schmieder, and H. Ermert. “Prä- und intraoperative Ultraschallbildgebung in der Neurochirurgie”. In: *Zeitschrift für Medizinische Physik* 17.4 (2007), pp. 273–279. DOI: 10.1016/j.zemedi.2007.06.005.

[111] C. Hansen, S. Zidowitz, F. Ritter, C. Lange, K. Oldhafer, and H. K. Hahn. “Risk maps for liver surgery”. In: *Int J Comput Assist Radiol Surg* 8.3 (2013), pp. 419–428. DOI: 10.1007/s11548-012-0790-6.

[112] D. J. Harris, J. M. Bird, P. A. Smart, M. R. Wilson, and S. J. Vine. “A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training”. In: *Front Psychol* 11 (2020). DOI: 10.3389/fpsyg.2020.00605.

[113] S. G. Hart and L. E. Staveland. “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research”. In: *Adv Psychol* (1988), pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

[114] J. Hartfill, J. Gabel, D. Neves-Coelho, D. Vogel, F. Räthel, S. Tiede, O. Ariza, and F. Steinicke. “Word Saber: An Effective and Fun VR Vocabulary Learning Game”. In: *Proc. of Mensch Und Computer*. 2020, pp. 145–154. DOI: 10.1145/3404983.3405517.

[115] S. Hasanzadeh, N. F. Polys, and J. M. de la Garza. “Presence, Mixed Reality, and Risk-Taking Behavior: A Study in Safety Interventions”. In: *IEEE Trans Vis Comput Graph* 26.5 (2020), pp. 2115–2125. DOI: 10.1109/TVCG.2020.2973055.

[116] S. H. Mohd Hashim, M. Ismail, H. Manaf, and F. A. Hanapiah. “Development of Dual Cognitive Task Virtual Reality Game Addressing Stroke Rehabilitation”. In: *Proc. of the International Conference on Virtual and Augmented Reality Simulations*. 2019, pp. 21–25. DOI: 10.1145/3332305.3332312.

[117] P. Haubruck, F. Nickel, J. Ober, T. Walker, C. Bergdolt, M. Friedrich, B. P. Müller-Stich, F. Forchheim, C. Fischer, G. Schmidmaier, and M. C. Tanner. “Evaluation of App-Based Serious Gaming as a Training Method in Teaching Chest Tube Insertion to Medical Students: Randomized Controlled Trial”. In: *J Med Internet Res* 20.5 (2018), e195. ISSN: 1438-8871. DOI: 10.2196/jmir.9956.

[118] D. He, S. Cao, Y. Le, M. Wang, Y. Chen, and B. Qian. “Virtual Reality Technology in Cognitive Rehabilitation Application: Bibliometric Analysis”. In: *JMIR Serious Games* 10 (4 2022), e38315. DOI: 10.2196/38315.

[119] M. Hegarty, M. Keehner, C. Cohen, and D. Montello. "The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals". In: *Proc. of Applied spatial cognition: From research to cognitive technology*. Oct. 2010, pp. 285–315.

[120] O. Heizmann, S. Zidowitz, H. Bourquain, S. Potthast, H.-O. Peitgen, D. Oertli, and C. Kettelhack. "Assessment of Intraoperative Liver Deformation During Hepatic Resection: Prospective Clinical Study". In: *World J Surg* 34.8 (2010), pp. 1887–1893. ISSN: 1432-2323. DOI: 10.1007/s00268-010-0561-x.

[121] A. Herur-Raman, N. Almeida, W. Greenleaf, D. Williams, A. Karshenas, and J. Sherman. "Next-Generation Simulation—Integrating Extended Reality Technology Into Medical Education". In: *Front Virtual Real* 2 (Sept. 2021). DOI: 10.3389/frvir.2021.693399.

[122] J. Hettig, P. Saalfeld, M. Luz, M. Becker, M. Skalej, and C. Hansen. "Comparison of gesture and conventional interaction techniques for interventional neuroradiology". In: *Int J Comput Assist Radiol Surg* 12.9 (2017), pp. 1643–1653. DOI: 10.1007/s11548-017-1523-7.

[123] A. Homach-Klonisch, T. Klonisch, and J. Peeler. *Sobotta Clinical Atlas of Human Anatomy*. Elsevier GmbH, 20019. ISBN: 978-0-7020-5273-6.

[124] 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, and M. Collins. "Virtual Reality Simulation Facilitates Resident Training in Total Hip Arthroplasty: A Randomized Controlled Trial". In: *J Arthroplasty* 34.10 (2019), pp. 2278–2283. DOI: 10.1016/j.arth.2019.04.002.

[125] D. Howett, A. Castegnaro, K. Krzywicka, J. Hagman, D. Marchment, R. Henson, M. Rio, J. A. King, N. Burgess, and D. Chan. "Differentiation of mild cognitive impairment using an entorhinal cortex-based test of virtual reality navigation". In: *Brain* 142.6 (2019), pp. 1751–1766.

[126] X. Hu, X. Liu, H. Wang, L. Xu, W. Peng, W. Zhang, Z. Niu, and L. Zhang. "A novel physics-based model for fast computation of blood flow in coronary arteries". In: *BioMed Eng OnLine* 22.56 (2023), pp. 1–28. DOI: 10.1186/s12938-023-01121-y.

[127] R. Huang, A. D. Ritzhaupt, M. Sommer, J. Zhu, A. Stephen, N. Valle, J. Hampton, and J. Li. "The impact of gamification in educational settings on student learning outcomes: a meta-analysis". In: *Educ Technol Res Dev* 68.4 (2020), pp. 1875–1901. ISSN: 1042-1629. DOI: 10.1007/s11423-020-09807-z.

[128] T. Huber, M. Paschold, C. Hansen, T. Wunderling, H. Lang, and W. Kneist. "New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff". In: *Surg Endosc* 31 (2017), pp. 4472–4477. DOI: 10.1007/s00464-017-5500-6.

[129] T. Huber, T. Wunderling, M. Paschold, H. Lang, W. Kneist, and C. Hansen. "Highly Immersive Virtual Reality Laparoscopy Simulation: Development and Future Aspects". In: *Int J Comput Assist Radiol Surg* 13 (Nov. 2017), pp. 281–290. DOI: 10.1007/s11548-017-1686-2.

[130] F. Huettl, M. Allgaier, S. Saalfeld, L. Hanke, B. Preim, H. Lang, C. Hansen, and T. Huber. "Liver IOUS 2.0: Training and Gaming in virtueller Realität". In: *Z Gastroenterol*. Vol. 61. 2023, e602. DOI: 10.1055/s-0043-1772139.

[131] F. Huettl, P. Saalfeld, C. Hansen, B. Preim, A. Poplawski, W. Kneist, H. Lang, and 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". In: *Ann Transl Med* 9.13 (2021). DOI: 10.21037/atm-21-512.

[132] D. Huttenlocher, G. Klanderman, and W. Rucklidge. "Comparing Images Using the Hausdorff Distance". In: *IEEE Trans Pattern Anal Mach Intell* 15 (1993), pp. 850–863. DOI: 10.1109/34.232073.

[133] W. M. IJgosse, H. van Goor, and J.-M. Luursema. "Saving robots improves laparoscopic performance: transfer of skills from a serious game to a virtual reality simulator". In: *Surg Endosc* 32.7 (2018), pp. 3192–3199. ISSN: 1432-2218. DOI: 10.1007/s00464-018-6036-0.

[134] 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. DOI: 10.1109/VRW50115.2020.00287.

[135] R. J. K. Jacob. "Human-Computer Interaction: Input Devices". In: *ACM Comput Surv* 28.1 (1996), pp. 177–179. DOI: 10.1145/234313.234387.

[136] N. Jacobsen, J. D. Larsen, C. Falster, C. P. Nolsøe, L. Konge, O. Graumann, and C. B. Laursen. "Using Immersive Virtual Reality Simulation to Ensure Competence in Contrast-Enhanced Ultrasound". In: *Ultrasound Med Biol* 48.5 (2022), pp. 912–923. ISSN: 0301-5629. DOI: 10.1016/j.ultrasmedbio.2022.01.015.

[137] M. Javaid and A. Haleem. "Virtual reality applications toward medical field". In: *Clin Epidemiology Glob Health* 8.2 (2020), pp. 600–605. DOI: 10.1016/j.cegh.2019.12.010.

[138] H. Jiang, S. Vimalesvaran, J. K. Wang, K. B. Lim, S. R. Mogali, and L. T. Car. "Virtual Reality in Medical Students' Education: Scoping Review". In: *JMIR Med Educ* 8.1 (2022), e34860. DOI: 10.2196/34860.

[139] N. W. John, T. W. Day, and 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. DOI: 10.2312/vcbm.20201167.

[140] S. Johnson, C. Hunt, H. Woolnough, M. Crawshaw, C. Kilkenny, D. Gould, A. Sinha, A. England, and P.-F. Villard. "Virtual Reality, Ultrasound-guided Liver Biopsy Simulator: Development and Performance Discrimination". In: *Br J Radiol* 85 (Feb. 2011). DOI: 10.1259/bjr/47436030.

[141] T. Jong and M. G. M. Ferguson-Hessler. "Types and qualities of knowledge". In: *Educ Psychol* 31.2 (1996), pp. 105–113.

[142] I. Joo. "The role of intraoperative ultrasonography in the diagnosis and management of focal hepatic lesions". In: *Ultrason* 34.4 (2015), pp. 246–257. ISSN: 2288-5943. DOI: 10.14366/usg.15014.

[143] A. Junga, P. Kockwelp, D. Valkov, B. Marschall, S. Hartwig, W. Stummer, B. Risse, and M. Holling. "Virtual Reality based teaching – a paradigm shift in education?" In: *Proc. of Deutsche Gesellschaft für Neurochirurgie. 73. Jahrestagung der Deutschen Gesellschaft für Neurochirurgie (DGNC), Joint Meeting mit der Griechischen Gesellschaft für Neurochirurgie*. 2022. DOI: 10.3205/22dgnc538.

[144] K. Junge, Larsen. J. D., S. W. Stougaard, R. O. Jensen, C. Falster, S. Posth, A. B. Nielsen, C. B. Laursen, L. Konge, O. Graumann, and et al. "Education in Focused Assessment With Sonography for Trauma Using Immersive Virtual Reality: A Prospective, Interventional Cohort Study and Non-inferiority Analysis With a Historical Control". In: *Ultrasound Med Biol* 50.2 (2024), pp. 277–284. DOI: 10.1016/j.ultrasmedbio.2023.10.013.

[145] M. Kaluschke, M. Su Yin, P. Haddawy, N. Srimaneekarn, P. Saikaew, and 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. DOI: 10.1109/VRW52623.2021.00069.

[146] S. Kamuro, K. Minamizawa, N. Kawakami, and S. Tachi. "Ungrounded kinesthetic pen for haptic interaction with virtual environments". In: *Proc. of IEEE International Symposium on Robot and Human Interactive Communication*. 2009, pp. 436–441. DOI: 10.1109/ROMAN.2009.5326217.

[147] S. H. K. Kang. "Spaced Repetition Promotes Efficient and Effective Learning: Policy Implications for Instruction". In: *Policy Insights from the Behavioral and Brain Sciences* 3.1 (2016), pp. 12–19. DOI: 10.1177/2372732215624708.

[148] A. Kanzira and X. Chen. "Virtual Reality-Based Simulator of Needle Interventions for Liver Biopsy". In: *Proc. of International Conference on Virtual Reality (ICVR)*. 2022, pp. 77–81. DOI: 10.1109/ICVR55215.2022.9848113.

[149] C. C. Karaliotas. *Liver and Biliary Tract Surgery: Embryological Anatomy to 3D-Imaging and Transplant Innovations*. Springer-Verlag/Wien, 2006. ISBN: 978-3-211-49275-8. DOI: 10.1007/978-3-211-49277-2.

[150] A. Karamalis, W. Wein, and N. Navab. "Fast Ultrasound Image Simulation Using the Westervelt Equation". In: *Proc. of Medical Image Computing and Computer-Assisted Intervention – MICCAI*. 2010, pp. 243–250. ISBN: 978-3-642-15705-9. DOI: 10.1007/978-3-642-15705-9_30.

[151] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. "Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness". In: *Int J Aviat Psychol* 3.3 (1993), pp. 203–220.

[152] H. G. Kenngott, M. Pfeiffer, A. A. Preukschas, L. Bettscheider, P. A. Wise, M. Wagner, S. Speidel, M. Huber, F. Nickel, A. Mehrabi, and 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". In: *Surg Endosc* 36.1 (2022), pp. 126–134. ISSN: 1432-2218. DOI: 10.1007/s00464-020-08246-4.

[153] H. G. Kenngott, J. J. Wünscher, M. Wagner, A. Preukschas, A. L. Wekerle, P. Neher, S. Suwelack, S. Speidel, F. Nickel, D. Oladokun, L. Maier-Hein, R. Dillmann, H. P. Meinzer, and B. P. Müller-Stich. "OpenHELP (Heidelberg laparoscopy phantom): development of an open-source surgical evaluation and training tool". In: *Surg Endosc* 29.11 (2015), pp. 3338–3347. ISSN: 1432-2218. DOI: 10.1007/s00464-015-4094-0.

[154] J. S. Kim, N. Jonas, T. Z. Rizvi, Z. Lin, D. Plewa, C. Ricard, Y. L. Cheah, C. J. Simon, and V. Wright. “Validation of a multidisciplinary virtual reality (VR) robotic surgical curriculum”. In: *J Robotic Surg* 17 (2023), pp. 42495–2502. DOI: 10.1007/s11701-023-01679-8.

[155] J. W. Kim, H. Jeong, K. Kim, D. P. DeMeo, and B. T. Carroll. “Image Based Virtual Reality Haptic Simulation for Multimodal Skin Tumor Surgery Training”. In: *Proc. of International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. 2021, pp. 1–4. DOI: 10.1109/BioSMART54244.2021.9677802.

[156] D. L. Kirkpatrick. “Evaluation of Training”. In: *Proc. of Training and Development Handbook: A Guide to Human Resource Development*. Ed. by R. L. Craig. McGraw-Hill, 1976.

[157] R. A. Kockro, T. Killeen, A. Ayyad, M. Glaser, A. T. Stadie, R. Reisch, A. Giese, and E. Schwandt. “Aneurysm Surgery with Preoperative Three-Dimensional Planning in a Virtual Reality Environment: Technique and Outcome Analysis”. In: *World Neurosurg* 96 (2016), pp. 489–499.

[158] R. A. Kockro and L. Serra. “Patient-Specific Virtual Reality Simulation for Minimally Invasive Neurosurgery”. In: *Comprehensive Healthcare Simulation: Neurosurgery*. 2018, pp. 159–184. DOI: 10.1007/978-3-319-75583-0_13.

[159] P. Kockwelp, A. Junga, D. Valkov, B. Marschall, M. Holling, and B. Risse. “Towards VR Simulation-Based Training in Brain Death Determination”. In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2022, pp. 287–292. DOI: 10.1109/VRW55335.2022.00065.

[160] J. Koivisto and J. Hamari. “The rise of motivational information systems: A review of gamification research”. In: *Int J Inf Manag* 45 (2019), pp. 191–210. ISSN: 0268-4012. DOI: https://doi.org/10.1016/j.ijinfomgt.2018.10.013.

[161] R. Kong, R. Wang, and Z. Shen. “Virtual Reality System for Invasive Therapy”. In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2021, pp. 689–690. DOI: 10.1109/VRW52623.2021.00227.

[162] O. Konrad-Verse, A. Littmann, and B. Preim. “Virtual Resection with a Deformable Cutting Plane”. In: *Simulation und Visualisierung*. 2004, pp. 203–214.

[163] M. Korenkov, C.-T. Germer, and H. Lang, eds. *Viszeralchirurgische Operationen und technische Varianten: Ösophagus - Leber - Pankreas - Endokrine Chirurgie - Metabolische Chirurgie*. 1st ed. 2021. Springer eBook Collection. Springer Berlin Heidelberg, 2021. ISBN: 9783662604755. DOI: 10.1007/978-3-662-60475-5.

[164] J. Krath, M. Altmeyer, G. F. Tondello, and L. E. Nacke. “Hexad-12: Developing and Validating a Short Version of the Gamification User Types Hexad Scale”. In: *Proc. of the CHI Conference on Human Factors in Computing Systems*. 2023. DOI: 10.1145/3544548.3580968.

[165] W. Kratzer, V. Fritz, R. A. Mason, M. M. Haenle, and V. Kaechele. “Factors affecting liver size: a sonographic survey of 2080 subjects”. In: *J Ultrasound Med* 22.11 (2003), pp. 1155–61. DOI: 10.7863/jum.2003.22.11.1155.

[166] Gesellschaft der epidemiologischen Krebsregister e.V. (GEKID) und das Zentrum für Krebsregisterdaten (ZfKD) im Robert Koch-Institut. “Krebs in Deutschland für 2017/2018”. In: *Zentrum für Krebsregisterdaten* (2021). DOI: 10.25646/8353.

[167] J. B. Kruskal and R. A. Kane. "Intraoperative US of the liver: techniques and clinical applications". In: *Radiographics* 26.4 (2006), pp. 1067–84. DOI: 10.1148/rg.264055120.

[168] J. Kumar. "Gamification at Work: Designing Engaging Business Software". In: *Proc. of Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience.* 2013, pp. 528–537.

[169] M. Kumar, K. Karagiozov, L. Chen, S. Imizu, M. Yoneda, T. Watabe, Y. Kato, H. Sano, and T. Kanno. "A Classification of Unruptured Middle Cerebral Artery Bifurcation Aneurysms that can Help in Choice of Clipping Technique". In: *Minim Invasive Neurosurg* 50 (July 2007), pp. 132–9. DOI: 10.1055/s-2007-985153.

[170] L. Lan, R. Q. Mao, R. Y. Qiu, J. Kay, and D. de Sa. "Immersive Virtual Reality for Patient-Specific Preoperative Planning: A Systematic Review". In: *Surg Innov* 30.1 (2023), pp. 109–122. DOI: 10.1177/15533506221143235.

[171] R. N. Landers. "Developing a Theory of Gamified Learning: Linking Serious Games and Gamification of Learning". In: *Simul and Gaming* 45.6 (2014). DOI: 10.1177/1046878114563660.

[172] A. Langbein, D. A. Plecher, F. Pankratz, C. Eghtebas, F. Palmas, and G. Klinker. "Gamifying Stereo Camera Registration for Augmented Reality". In: *Proc. of IEEE International Symposium on Mixed and Augmented Reality Adjunct.* 2018, pp. 125–126. DOI: 10.1109/ISMAR-Adjunct.2018.00049.

[173] J. D. Larsen, R. O. Jensen, P. I. Pietersen, N. Jacobsen, C. Falster, A. B. Nielsen, C. B. Laursen, L. Konge, and O. Graumann. "Education in Focused Lung Ultrasound Using Gamified Immersive Virtual Reality: A Randomized Controlled Study". In: *Ultrasound Med Biol* 49.3 (2023), pp. 841–852. ISSN: 0301-5629. DOI: 10.1016/j.ultrasmedbio.2022.11.011.

[174] M. E. Latoschik and C. Wienrich. "Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach". In: *Front Virtual Real* 3 (2022). DOI: 10.3389/frvir.2022.694433.

[175] Y. C. Law, T. Knott, S. Pick, B. Weyers, and T. W. Kuhlen. "Simulation-based Ultrasound Training Supported by Annotations, Haptics and Linked Multimodal Views". In: *Proc. of Eurographics Workshop on Visual Computing for Biology and Medicine.* 2015. DOI: 10.2312/vcbm.20151220.

[176] Y. C. Law, S. Ullrich, T. Knott, and T. W. Kuhlen. "Ultrasound Image Simulation with GPU-based Ray Tracing". In: *Proc. of Virtuelle und Erweiterte Realität - Workshop der GI-Fachgruppe VR/AR.* 2011, pp. 183–194.

[177] K. S. Lee, J. Zhang, V. Nguyen, J. Han, J. Johnson, R. Kirollos, and M. Teo. "The evolution of intracranial aneurysm treatment techniques and future directions". In: *Neurosurg Rev* 45.1 (Apr. 2021), pp. 1–25. DOI: 10.1007/s10143-021-01543-z.

[178] F. Lenz, M. Fock, T. Kramer, Y. Petersen, and M. Teistler. "A virtual reality game to support visuospatial understanding of medical X-ray imaging". In: *Proc. of IEEE International Conference on Serious Games and Applications for Health(SeGAH).* 2021, pp. 1–6. DOI: 10.1109/SEGAH52098.2021.9551899.

[179] C. Li, J. Ding, Z. Hong, Y. Pan, and P. X. Liu. “A Surface Mass-Spring Model With New Flexion Springs and Collision Detection Algorithms Based on Volume Structure for Real-Time Soft-Tissue Deformation Interaction”. In: *IEEE Access* 6 (2018), pp. 75572–75597. DOI: 10.1109/ACCESS.2018.2883679.

[180] M. Li, S. Ganni, J. Ponten, A. Albayrak, A.-F. Rutkowski, and J. Jakimowicz. “Analysing usability and presence of a virtual reality operating room (VOR) simulator during laparoscopic surgery training”. In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* 2020, pp. 566–572. DOI: 10.1109/VR46266.2020.00078.

[181] S. Li, J. Cui, A. Hao, S. Zhang, and Q. Zhao. “Design and Evaluation of Personalized Percutaneous Coronary Intervention Surgery Simulation System”. In: *IEEE Trans Vis Comput Graph* 27.11 (2021), pp. 4150–4160. DOI: 10.1109/TVCG.2021.3106478.

[182] R. Lohre, A. J. Bois, G. S. Athwal, D. P. Goel, and Canadian Shoulder and Elbow Society (CSES). “Improved Complex Skill Acquisition by Immersive Virtual Reality Training: A Randomized Controlled Trial”. In: *J Bone Joint Surg Am* 102.6 (2020), e26. DOI: 10.2106/JBJS.19.00982.

[183] J. J. López-Jiménez, J. L. Fernández-Alemán, L. L. González, O. G. Sequeros, B. M. Valle, J. A. García-Berná, A. Idri, and A. Toval. “Taking the pulse of a classroom with a gamified audience response system”. In: *Computer methods and programs in biomedicine* 213.106459 (2022). DOI: doi.org/10.1016/j.cmpb.2021.106459.

[184] E. Lorenz, J. Arend, M. Franz, M. Rahimli, A. Perrakis, V. Negrini, A. A. Gumbs, and R. S. Croner. “Robotic and laparoscopic liver resection—comparative experiences at a high-volume German academic center”. In: *Langenbeck’s Arch Surg* 406.3 (2021), pp. 753–761. ISSN: 1435-2451. DOI: 10.1007/s00423-021-02152-6.

[185] J. Lu, A. Leng, Y. Zhou, W. Zhou, J. Luo, X. Chen, and X. Qi. “An innovative virtual reality training tool for the pre-hospital treatment of cranialmaxillofacial trauma”. In: *Comput Assist Surg* 28.1 (2023), p. 2189047. DOI: 10.1080/24699322.2023.2189047.

[186] M. G. Lubner, L. Mankowski Gettle, D. H. Kim, T. J. Ziemlewicz, N. Dahiya, and P. Pickhardt. “Diagnostic and procedural intraoperative ultrasound: technique, tips and tricks for optimizing results”. In: *J Radiol* 94.1121 (2021), p. 20201406. DOI: 10.1259/bjr.20201406.

[187] A. J. Lungu, W. Swinkels, L. Claesen, P. Tu, J. Egger, and XiaoX.jun Chen. “A Review on the Applications of Virtual Reality, Augmented Reality and Mixed Reality in Surgical Simulation: An Extension to Different Kinds of Surgery”. In: *Expert Rev Med Devices* 18.1 (2020), pp. 47–62. DOI: 10.1080/17434440.2021.1860750.

[188] A. Lütt, N. Tsamitros, T. Wolbers, A. Rosenthal, A. L. Bröcker, R. Schöneck, F. Bermpohl, A. Heinz, A. Beck, and S. Gutwinski. “An explorative single-arm clinical study to assess craving in patients with alcohol use disorder using Virtual Reality exposure (CRAVE)-study protocol.” In: *BMC psychiatry* 23.1 (2023), p. 839. DOI: 10.1186/s12888-023-05346-y.

[189] M. Madsen, L. Konge, L. Nørgaard, A. Tabor, C. Ringsted, A. Klemmensen, B. Ottesen, and M. Tolsgaard. “Assessment of performance and learning curves on a virtual reality ultrasound simulator”. In: *Ultrasound Obstet Gynecol* 44.6 (Apr. 2014), pp. 693–699. DOI: 10.1002/uog.13400.

[190] M. S. Mansoory, M. R. Khazaei, S. M. Azizi, and E. Niromand. “Comparison of the effectiveness of lecture instruction and virtual reality-based serious gaming instruction on the medical students’ learning outcome about approach to coma”. In: *BMC Med Educ* 21.1 (2021), p. 347. DOI: 10.1186/s12909-021-02771-z.

[191] R. Q. Mao, L. Lan, J. Kay, R. Lohre, O. R. Ayeni, D. P. Goel, and D. de SA. “Immersive Virtual Reality for Surgical Training: A Systematic Review”. In: *J Surg Res* 268 (2021), pp. 40–58. DOI: 10.1016/j.jss.2021.06.045.

[192] A. Marczewski. “User Types”. In: *In Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design*. 2015, pp. 65–80. ISBN: 1514745666.

[193] T. Mashiko, K. Otani, R. Kawano, T. Konno, N. Kaneko, Y. Ito, and E. Watanabe. “Development of Three-Dimensional Hollow Elastic Model for Cerebral Aneurysm Clipping Simulation Enabling Rapid and Low Cost Prototyping”. In: *World Neurosurg* 83 (Oct. 2013), pp. 351–61. DOI: 10.1016/j.wneu.2013.10.032.

[194] A. Mastmeyer, M. Wilms, D. Fortmeier, J. Schröder, and H. Handels. “Real-Time Ultrasound Simulation for Training of US-Guided Needle Insertion in Breathing Virtual Patients”. In: *Stud Health Technol Inform* 220 (2016), pp. 219–226.

[195] P. Mastorakos and D. McGavern. “The anatomy and immunology of vasculature in the central nervous system”. In: *Sci Immunol* 4.37 (2019). DOI: 10.1126/sciimmunol.aav0492.

[196] L. Mayer, M. Süncksen, S. Reinhold, S. Bertel, and M. Teistler. “Training visuospatial skills for medical ultrasound imaging with a desktop-based learning game”. In: *Proc. of Serious Games and Applications for Health(SeGAH)*. 2021, pp. 1–6. DOI: 10.1109/SEGAH52098.2021.9551914.

[197] A. Mazarakis. “Gamification Reloaded: Current and Future Trends in Gamification Science”. In: *i-com* 20.3 (2021), pp. 279–294. DOI: 10.1515/icom-2021-0025.

[198] A. Mazarakis and P. Bräuer. “Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements”. In: *Int J Hum-Comput Int* 39.3 (2023), pp. 612–627. DOI: 10.1080/10447318.2022.2041909.

[199] R. McMahan, D. Bowman, D. Zielinski, and R. Brady. “Evaluating Display Fidelity and Interaction Fidelity in a Virtual Reality Game”. In: *IEEE Trans Vis Comput Graph* 18 (2012), pp. 626–33. DOI: 10.1109/TVCG.2012.43.

[200] R. McMahan, C. Lai, and S. Pal. “Interaction Fidelity: The Uncanny Valley of Virtual Reality Interactions”. In: *Virtual, Augmented and Mixed Reality*. Vol. 9740. 2016, pp. 59–70. DOI: 10.1007/978-3-319-39907-2_6.

[201] M. J. MCPheeters, K. Vakharia, S. A. Munich, and A. H. Siddiqui. “Wide-Necked Cerebral Artery Aneurysms : Where Do We Stand ? The evolution and progression of neurointerventional treatments for morphologically challenging aneurysms”. In: *Endovascular Today*. Vol. 18. 2019, pp. 70–79.

[202] M. Al-Memar, S. Saso, S. Bobdiwala, L. Ameye, S. Guha, K. Joash, C. Stalder, S. Sur, K. Moorthy, D. Timmerman, and T. Bourne. "Validation of a virtual reality simulator for the use of transvaginal ultrasonography in gynaecology and early pregnancy". In: *Australas J Ultrasound Med* 20 (May 2017). DOI: 10.1002/ajum.12052.

[203] A. H. S. Metwally, L. E. Nacke, M. Chang, Y. Wang, and A. M. F. Yousef. "Revealing the hotspots of educational gamification: An umbrella review". In: *Int J Educ Res* 109 (2021), p. 101832. ISSN: 08830355. DOI: 10.1016/j.ijer.2021.101832.

[204] F. B. Meyer. *Atlas of Neurosurgery - Basic Approaches to Cranial and Vascular Procedures*. Churchill Livingstone, 1998. ISBN: 0443089515.

[205] Y. Mise, K. Tani, T. Aoki, Y. Sakamoto, K. Hasegawa, Y. Sugawara, and N. Kokudo. "Virtual liver resection: computer-assisted operation planning using a three-dimensional liver representation". In: *J Hepatobiliary Pancreat Sci* 20.2 (2013), pp. 157–164. DOI: 10.1007/s00534-012-0574-y.

[206] H. Miyazaki, W. Kurihara, X. Han, S. Sakaguchi, and K. Kushiyama. "Development of Exergame to Resolve Deconditioning in Children with Orthostatic Dysregulation". In: *Proc. of ACM SIGGRAPH 2022 Posters*. 2022. DOI: 10.1145/3532719.3543212.

[207] A. V. Moiyadi. "Intraoperative Ultrasound Technology in Neuro-Oncology Practice—Current Role and Future Applications". In: *World Neurosurg* 93 (2016), pp. 81–93. ISSN: 1878-8750. DOI: https://doi.org/10.1016/j.wneu.2016.05.083.

[208] D. K. Molina and V. J. DiMaio. "Normal organ weights in men: part II-the brain, lungs, liver, spleen, and kidneys". In: *Am J Forensic Med Pathol* 33.4 (2012), pp. 368–372. DOI: 10.1097/PAF.0b013e31823d29ad.

[209] J. Mönch. "Szenariobasierte Konzeption, Entwicklung und Evaluierung chirurgischer Trainingssysteme". PhD thesis. Otto-von-Guericke Univerity Magdeburg, 2011.

[210] R. Monica and J. Aleotti. "Evaluation of the Oculus Rift S tracking system in room scale virtual reality". In: *Virtual Real* 26 (2022), pp. 1335–1345. DOI: 10.1007/s10055-022-00637-3.

[211] A. B. Mor. "Progressive Cutting with Minimal New Element Creation of Soft Tissue Models for Interactive Surgical Simulation". PhD thesis. Carnegie Mellon University, 2001.

[212] D. Morris, C. Sewell, N. Blevins, F. Barbagli, and K. Salisbury. "A Collaborative Virtual Environment for the Simulation of Temporal Bone Surgery". In: *Proc. of Medical Image Computing and Computer-Assisted Intervention – MICCAI*. 2004, pp. 319–327. ISBN: 978-3-540-30136-3.

[213] B. Morschheuser, L. Hassan, K. Werder, and J. Hamari. "How to design gamification? A method for engineering gamified software". In: *Inf Softw Technol* 95 (2018). DOI: 10.1016/j.infsof.2017.10.015.

[214] F. Mostajeran, A. Kirsten, F. Steinicke, J. Gallinat, and S. Kühn. "Towards Gamified Alcohol Use Disorder Therapy in Virtual Reality: A Preliminary Usability Study". In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 2019, pp. 1471–1476. DOI: 10.1109/VR.2019.8797817.

[215] A. Munawar, Z. Li, N. Nagururu, D. Trakimas, P. Kazanzides, R. H. Taylor, and F. X. Creighton. "Fully immersive virtual reality for skull-base surgery: surgical training and beyond". In: *Int J Comput Assist Radiol Surg* 19.1 (2024), pp. 51–59. DOI: 10.1007/s11548-023-02956-5.

[216] R. S. Najeeb, J. Uthayan, R. P. Lojini, G. Vishaliney, J. Alosius, and A. Gamage. "Gamified Smart Mirror to Leverage Autistic Education - Aliza". In: *Proc. of International Conference on Advancements in Computing (ICAC)*. Vol. 1. 2020, pp. 428–433. DOI: 10.1109/ICAC51239.2020.9357065.

[217] B. Neyazi, P. Saalfeld, P. Berg, M. Skalej, B. Preim, I. E. Sandalcioglu, and S. Saalfeld. "VR Craniotomy for Optimal Intracranial Aneurysm Surgery Planning". In: *Curr Dir Biomed Eng*. 2019, pp. 234–239.

[218] D. Ni, W. Y. Chan, J. Qin, Y.-P. Chui, I. Qu, S. S. M. Ho, and P.-A. Heng. "A Virtual Reality Simulator for Ultrasound-Guided Biopsy Training". In: *IEEE Comput Graph Appl* 31.2 (2011), pp. 36–48. DOI: 10.1109/MCG.2009.151.

[219] D. Nicholls, L. Sweet, and J. Hyett. "Psychomotor Skills in Medical Ultrasound Imaging". In: *J Med Ultrasound* 33.8 (2014), pp. 1349–1352. DOI: 10.7863/ultra.33.8.1349.

[220] J Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers Inc., 1994. ISBN: 9780080520292.

[221] J. Nielsen. "Usability engineering at a discount". In: *Proceedings of the Third International Conference on Human-Computer Interaction on Designing and Using Human-Computer Interfaces and Knowledge Based Systems*. 1989, pp. 394–401. ISBN: 044488078X.

[222] S. Nikolov, S. Blackwell, A. Zverovitch, R. Mendes, M. Livne, J. De Fauw, Y. Patel, C. Meyer, H. Askham, B. Romera-Paredes, and et al. "Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study". In: *J Med Internet Res* 23.7 (2021), e26151.

[223] S. Oberdörfer, D. Heidrich, and M. E. Latoschik. "Usability of Gamified Knowledge Learning in VR and Desktop-3D". In: *Proc. of the CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–13. DOI: 10.1145/3290605.3300405.

[224] M. Oishi, M. Fukuda, N. Yajima, K. Yoshida, M. Takahashi, T. Hiraishi, T. Takao, A. Saito, and Y. Fujii. "Interactive presurgical simulation applying advanced 3D imaging and modeling techniques for skull base and deep tumors". In: *J Neurosurg* 119.1 (2013), pp. 94–105. DOI: 10.3171/2013.3.JNS121109.

[225] T. Olgers, J. Os, H. Bouma, and J. Maaten. "The validation of a serious game for teaching ultrasound skills". In: *Ultrasound J* 14.29 (2022). DOI: 10.1186/s13089-022-00280-8.

[226] S. T. Orcutt, K. Kobayashi, M. Sultenfuss, B. S. Hailey, A. Sparks, B. Satpathy, and D. A. Anaya. "Portal Vein Embolization as an Oncosurgical Strategy Prior to Major Hepatic Resection: Anatomic, Surgical, and Technical Considerations". In: *Front Surg* 3.14 (2016). DOI: 10.3389/fsurg.2016.00014.

[227] K. E. Orr, S. C Hamilton, R. Clarke, M. Y. Adi, C. Gutteridge, P. Suresh, and S. J. Freeman. "The integration of transabdominal ultrasound simulators into an ultrasound curriculum". In: *Ultrasound* 27.1 (2019), pp. 20–30. DOI: 10.1177/1742271X18762251.

[228] Y. Otsuka, M. Tsuchiya, T. Katagiri, Y. Kubota, J. Ishii, T. Maeda, and H. Kaneko. "Indications and technique for laparoscopic liver resection in patients with hepatocellular carcinoma and liver cirrhosis". In: *Hepatoma Res* 2 (2016), pp. 241–247. DOI: 10.20517/2394-5079.2016.11.

[229] A. Pacioni, M. Carbone, C. Freschi, R. Viglialoro, V. Ferrari, and M. Ferrari. "Patient-specific ultrasound liver phantom: materials and fabrication method". In: *Int J Comput Assist Radiol Surg* 10.7 (2014), pp. 1065–1075. DOI: 10.1007/s11548-014-1120-y.

[230] F. Palmas, D. Labode, D. A. Plecher, and G. Klinker. "Comparison of a Gamified and Non-Gamified Virtual Reality Training Assembly Task". In: *Proc. of International Conference on Virtual Worlds and Games for Serious Applications (VS-Games).* 2019, pp. 1–8. DOI: 10.1109/VS-Games.2019.8864583.

[231] D. Panariello, T. Caporaso, S. Grazioso, G. Di Gironimo, A. Lanzotti, S. Knopp, L. Pelliccia, M. Lorenz, and P. Klimant. "Using the KUKA LBR iiwa Robot as Haptic Device for Virtual Reality Training of Hip Replacement Surgery". In: *Proc. of IEEE International Conference on Robotic Computing (IRC).* 2019, pp. 449–450. DOI: 10.1109/IRC.2019.00094.

[232] H. Patel, D. Chandrasekaran, E. Myriokefalitaki, A. Gebeh, K. Jones, Y. B. Jeve, and Midlands Research Collaborative in Obstetrics & Gynecology. "The Role of Ultrasound Simulation in Obstetrics and Gynecology Training: A UK Trainees' Perspective". In: *Simul Healthc* 11.5 (2016), pp. 340–344. DOI: 10.1097/sih.0000000000000176.

[233] S. Peeters and J. July. "Pterional Approach". In: *Neurovascular Surgery: Surgical Approaches for Neurovascular Diseases.* Ed. by J. July and E. J. Wahjoepramono. Springer, 2019, pp. 3–10. ISBN: 978-981-10-8950-3.

[234] O. Peretti, Y. Spyridis, A. Sesis, and G. Efstathopoulos. "Gamified First Responder Training Solution in Virtual Reality". In: *Proc. of International Conference on Distributed Computing in Sensor Systems (DCOSS).* 2021, pp. 295–301. DOI: 10.1109/DCOSS52077.2021.00055.

[235] K. Perlin. "Improving Noise". In: *Proc. of ACM Computer Graphics and Interactive Techniques.* 2002, pp. 681–682. ISBN: 1581135211. DOI: 10.1145/566570.566636.

[236] D.-M. Pham and W. Stuerzlinger. "Is the Pen Mightier than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality". In: *Proc. of VRST.* ACM, 2019, pp. 1–11. ISBN: 9781450370011. DOI: 10.1145/3359996.3364264.

[237] J. F. Pinto, H. R. Carvalho, G. R. R. Chambel, J. Ramiro, and A. Goncalves. "Adaptive gameplay and difficulty adjustment in a gamified upper-limb rehabilitation". In: *Proc. of IEEE International Conference on Serious Games and Applications for Health (SeGAH).* 2018, pp. 1–8. DOI: 10.1109/SeGAH.2018.8401363.

[238] D. Pohlandt, B. Preim, and P. Saalfeld. "Supporting Anatomy Education with a 3D Puzzle in a VR Environment - Results from a Pilot Study". In: *Proc. of Mensch Und Computer 2019*. 2019, pp. 91–102. DOI: 10.1145/3340764.3340792.

[239] M. Portelli, S. F. Bianco, T. Bezzina, and J. E. Abela. "Virtual reality training compared with apprenticeship training in laparoscopic surgery: a meta-analysis". In: *Ann R Coll Surg Engl* 102.9 (2020), pp. 672–684. DOI: 10.1308/rcsann.2020.0178.

[240] T. Poston and L. Serra. "Dextrous virtual work". In: *Commun ACM* 39.5 (1996), pp. 37–45. DOI: 10.1145/229459.229464.

[241] B. Preim, R. Raidou, N. Smit, and K. Lawonn. *Visualization, Visual Analytics and Virtual Reality in Medicine*. 554. Elsevier, 2023. ISBN: 978-0-12-822962-0.

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

[243] Y. Pulijala, M. Ma, M. Pears, D. Peebles, and A. Ayoub. "Effectiveness of immersive virtual reality in surgical training—a randomized control trial". In: *J Oral Maxillofac Surg* 76.5 (2018), pp. 1065–1072. DOI: 10.1016/j.joms.2017.10.002.

[244] E. Quaia. *Imaging of the Liver and Intra-hepatic Biliary Tract: Volume 1: Imaging Techniques and Non-tumoral Pathologies*. Medical Radiology. Springer International Publishing, 2020. ISBN: 9783030389833.

[245] C. Quaiser-Pohl. "The Mental Cutting Test "Schnitte" and the Picture Rotation Test-Two New Measures to Assess Spatial Ability". In: *Int J Test* 3 (Sept. 2003), pp. 219–231. DOI: 10.1207/S15327574IJT0303_2.

[246] C. M. I. Quarato, D. Lacedonia, M. Salvemini, G. Tuccari, G. Mastrodonato, R. Villani, L. A. Fiore, G. Scioscia, A. Mirijello, A. Saponara, and M. Sperandeo. "A Review on Biological Effects of Ultrasounds: Key Messages for Clinicians". In: *Diagnostics* 13.5 (2023), p. 855. DOI: 10.3390/diagnostics13050855.

[247] V. C. C. Quispe, J. C. C. Aguinaga, L. G. D. Huaco, and J. T. Cruz. "DENTORACULUS - A gamified Virtual Reality experience for dental morphology learning". In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 2023, pp. 1007–1008. DOI: 10.1109/VRW58643.2023.00349.

[248] A. Raabe, B. Meyer, K. Schaller, P. Vajkoczy, and P. A. Winkler. *The Craniotomy Atlas*. Thieme, 2019. ISBN: 9783132057913.

[249] H.-R. Rantamaa, J. Kangas, S. K. Kumar, H. Mehtonen, J. Järnstedt, and R. Raisamo. "Comparison of a VR Stylus with a Controller, Hand Tracking, and a Mouse for Object Manipulation and Medical Marking Tasks in Virtual Reality". In: *Appl Sci* 13.4 (2023). DOI: 10.3390/app13042251.

[250] P. A. Rauschnabel, R. Felix, C. Hinsch, H. Shahab, and F. Alt. "What is XR? Towards a Framework for Augmented and Virtual Reality". In: *Comput Hum Behav* 133 (2022), p. 107289. ISSN: 07475632. DOI: 10.1016/j.chb.2022.107289.

[251] B. Reeves and L. Read. *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete*. 2009. ISBN: 9781422155134.

[252] T. Reichl, J. Passenger, O. Acosta, and O. Salvado. "Ultrasound goes GPU: real-time simulation using CUDA". In: *Prog Biomed Opt Imaging - Proc SPIE* 7261 (Feb. 2009). DOI: 10.1117/12.812486.

[253] A. V. Reinschluessel, T. Muender, D. Salzmann, T. Döring, R. Malaka, and D. Weyhe. "Virtual Reality for Surgical Planning - Evaluation Based on Two Liver Tumor Resections". In: *Front Surg* 9 (2022), p. 821060. DOI: 10.3389/fsurg.2022.821060.

[254] B. Reitinger, A. Bornik, R. Beichel, and D. Schmalstieg. "Liver Surgery Planning Using Virtual Reality". In: *IEEE Comput Graph Appl* 26 (2006), pp. 36–47. DOI: 10.1109/MCG.2006.131.

[255] A. Ricca, A. Chellali, and S. Otrnane. "The influence of hand visualization in tool-based motor-skills training, a longitudinal study". In: *Proc. of IEEE Virtual Reality and 3D User Interfaces (VR)*. 2021, pp. 103–112. DOI: 10.1109/VR50410.2021.00031.

[256] C. Richards, C. W. Thompson, and N. Graham. "Beyond Designing for Motivation: The Importance of Context in Gamification". In: *Proc. of CHI PLAY*. 2014, pp. 217–226. DOI: 10.1145/2658537.2658683.

[257] G. J. E. Rinkel and A. Algra. "Long-term outcomes of patients with aneurysmal subarachnoid haemorrhage". In: *The Lancet Neurology* 10.4 (2011), pp. 349–356. ISSN: 1474-4422. DOI: 10.1016/S1474-4422(11)70017-5.

[258] H. Romat, A. Fender, M. Meier, and C. Holz. "Flashpen: A High-Fidelity and High-Precision Multi-Surface Pen for Virtual Reality". In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces*. 2021, pp. 306–315. DOI: 10.1109/VR50410.2021.00053.

[259] T. Ropinski, F. Steinicke, and K. Hinrichs. "Visually Supporting Depth Perception in Angiography Imaging". In: *Smart Graphics*. Ed. by A. Butz, B. Fisher, A. Krüger, and P. Olivier. 2006, pp. 93–104. DOI: 10.1007/11795018_9.

[260] S. Rutkauskas, V. Gedrimas, J. Pundzius, G. Barauskas, and A. Basevicius. "Clinical and anatomical basis for the classification of the structural parts of liver". In: *Medicina (Kaunas, Lithuania)* 42 (Feb. 2006), pp. 98–106.

[261] R. M. Ryan. "Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory". In: *J Pers Soc Psychol* 43.3 (1982), pp. 450–461. DOI: 10.1037/0022-3514.43.3.450.

[262] S. Saalfeld, P. Berg, M. Neugebauer, and B. Preim. "Reconstruction of 3D Surface Meshes for Blood Flow Simulations of Intracranial Aneurysms". In: *Proc of CURAC*. Sept. 2015, pp. 163–168.

[263] A. H. Sadeghi, W. Bakhuis, F. Van Schaagen, F. B S Oei, J. A. Bekkers, A. P. W. M. Maat, E. A. F. Mahtab, A. J. J. C. Bogers, and Y. J. H. J. Taverne. "Immersive 3D virtual reality imaging in planning minimally invasive and complex adult cardiac surgery". In: *Eur Heart J, Digit Health* 1.1 (2020), pp. 62–70.

[264] H. Sadideen and R. Kneebone. "Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning?" In: *Am J Surg* 204.3 (2012), pp. 396–401. DOI: 10.1016/j.amjsurg.2011.12.020.

[265] M. Sailer and L. Homner. “The Gamification of Learning: a Meta-analysis”. In: *Educ Psychol Rev* 32 (2020), pp. 77–112. DOI: 10.1007/s10648-019-09498-w.

[266] M. Salehi, S.-A. Ahmadi, R. Prevost, N. Navab, and W. Wein. “Patient-specific 3D Ultrasound Simulation Based on Convolutional Ray-tracing and Appearance Optimization”. In: *Proc. of Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. 2015, pp. 510–518. ISBN: 978-3-319-24571-3. DOI: 10.1007/978-3-319-24571-3_61.

[267] M. B. Sannathimmappa, V. Nambiar, and R. Aravindakshan. “Learning out of the box: Fostering intellectual curiosity and learning skills among the medical students through gamification”. In: *Journal of education and health promotion* 11.79 (2022). DOI: 10.4103/jehp.jehp_683_21.

[268] M. U. Sattar, S. Palaniappan, A. Lokman, A. Hassan, N. Shah, and Z. Riaz. “Effects of Virtual Reality training on medical students’ learning motivation and competency”. In: *Pak J Med Sci* 35.3 (2019), pp. 852–857. DOI: 10.12669/pjms.35.3.44.

[269] L. E. Savastano, A. Bhambri, D. A. Wilkinson, and A. S. Pandey. “Biology of Cerebral Aneurysm Formation, Growth, and Rupture”. In: *Proc. of Intracranial Aneurysms*. 2018.

[270] R. J. Scalese, V. T. Obeso, and S. B. Issenberg. “Simulation Technology for Skills Training and Competency Assessment in Medical Education”. In: *J Gen Intern Med* 23.1 (2008), pp. 46–49. ISSN: 1525-1497. DOI: 10.1007/s11606-007-0283-4.

[271] M. W. Scerbo, T. R. Turner, E. Newlin-Canzone, D. Meglan, R. Waddington, D. King, and H. Champion. “A Preliminary Evaluation of a Burr Hole Drilling Simulator for Craniotomy”. In: *Proc. of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 54. 2010, pp. 2226–2365. DOI: 10.1177/154193121005402730.

[272] S. Schez-Sobrino, D. Vallejo, D. N. Monekosso, C. Glez-Morcillo, and P. Remagnino. “A Distributed Gamified System Based on Automatic Assessment of Physical Exercises to Promote Remote Physical Rehabilitation”. In: *IEEE Access* 8 (2020), pp. 91424–91434. DOI: 10.1109/ACCESS.2020.2995119.

[273] D. Schott, M. Kunz, T. Wunderling, F. Heinrich, R. Braun-Dullaeus, and C. Hansen. “CardioGenesis4D: Interactive Morphological Transitions of Embryonic Heart Development in a Virtual Learning Environment”. In: *IEEE Transactions on Visualization and Computer Graphics* 29.5 (2023), pp. 2615–2625. DOI: 10.1109/TVCG.2023.3247110.

[274] D. Schott, P. Saalfeld, G. Schmidt, F. Joeres, C. Boedecker, F. Huettl, H. Lang, T. Huber, B. Preim, and C. Hansen. “A VR/AR Environment for Multi-User Liver Anatomy Education”. In: *Proc. of IEEE Virtual Reality and 3D User Interfaces (VR)*. 2021, pp. 296–305. DOI: 10.1109/VR50410.2021.00052.

[275] T. Schubert, F. Friedmann, and H. Regenbrecht. “The Experience of Presence: Factor Analytic Insights”. In: *Presence* 10 (2001), pp. 266–281. DOI: 10.1162/105474601300343603.

[276] E. S. Schwandt, R. A. Kockro, A. Kramer, M. Glaser, and F. Ringel. “Presurgical selection of the ideal aneurysm clip by the use of a three-dimensional planning system”. In: *Neurosurg Rev* 45 (2022), pp. 2887–2894. DOI: 10.1007/s10143-022-01794-4.

[277] D. Selle, B. Preim, A. Schenk, and H.-O. Peitgen. “Analysis of vasculature for liver surgical planning”. In: *IEEE Trans Med Imaging* 21.11 (2002), pp. 1344–1357. DOI: 10.1109/TMI.2002.801166.

[278] J. H. Seo, B. Smith, M. Cook, M. Pine, E. Malone, S. Leal, and J. Suh. “Anatomy builder VR: Applying a constructive learning method in the virtual reality canine skeletal system”. In: *Proc. of IEEE Virtual Reality (VR).* 2017, pp. 399–400. DOI: 10.1109/VR.2017.7892345.

[279] X. Shao, Q. Yuan, D. Qian, Z. Ye, K. I. Zhuang, X. Jiang, Y. Jin, and D. Qiang. “Virtual reality technology for teaching neurosurgery of skull base tumor”. In: *BMC Med Educ* 20.3 (2020). DOI: 10.1186/s12909-019-1911-5.

[280] V. Shenoy, S. I. Khan, E. Lee, and O. Aalami. “Vascular1: Development and Evaluation of a Virtual Reality Ultrasound Guided Vascular Access Training Module”. In: *Proc. of IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 2021, pp. 1519–1523. DOI: 10.1109/BIBM52615.2021.9669886.

[281] T. Shepherd, M. Trinder, and M. Theophilus. “Does virtual reality in the perioperative setting for patient education improve understanding? A scoping review”. In: *Surg Pract Sci* 10 (2022), p. 100101. DOI: 10.1016/j.sipas.2022.100101.

[282] S. Shetty, L. Panait, J. Baranoski, S. Dudrick, R. L. Bell, K. Roberts, and A. J. Duffy. “Construct and Face Validity of A Novel Virtual Reality-Based Camera Navigation Curriculum”. In: *J Surg Res* 177 (2012), pp. 191–195. DOI: 10.1016/j.jss.2011.11.614.

[283] N. Shono, T. Kin, S. Nomura, S. Miyawaki, T. Saito, H. Imai, H. Nakatomi, H. Oyama, and N. Saito. “Microsurgery Simulator of Cerebral Aneurysm Clipping with Interactive Cerebral Deformation Featuring a Virtual Arachnoid”. In: *Oper Neurosurg* 14 (Aug. 2017), pp. 579–589. DOI: 10.1093/ons/opx155.

[284] S. Shorey, E. Ang, J. Yap, E. D. Ng, S. T. Lau, and C. K. Chui. “A Virtual Counseling Application Using Artificial Intelligence for Communication Skills Training in Nursing Education: Development Study”. In: *J Med Internet Res* 21.10 (2019), e14658. DOI: 10.2196/14658.

[285] D. Sibrina, S. Bethapudi, and G. A. Koulieris. “OrthopedVR: clinical assessment and pre-operative planning of paediatric patients with lower limb rotational abnormalities in virtual reality”. In: *Vis Comput* 39 (2023), pp. 3621–3633. DOI: 10.1007/s00371-023-02949-0.

[286] C. Simon, L. Herfort, and A. Chellali. “Design and evaluation of an immersive ultrasound-guided locoregional anesthesia simulator”. In: *Proc. of IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).* 2022, pp. 445–449. DOI: 10.1109/VRW55335.2022.00098.

[287] R. Skarbez, Jr. F. P. Brooks, and M. C. Whitton. “A Survey of Presence and Related Concepts”. In: *ACM Comput Surv* 50.6 (2018), pp. 1–39. ISSN: 0360-0300. DOI: 10.1145/3134301.

[288] B. Skjold-Ødegaard and K. Søreide. “Standardization in surgery: friend or foe?” In: *Br J Surg* 107.9 (2020), pp. 1094–1096. DOI: 10.1002/bjs.11573.

[289] M. Slater. “Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments”. In: *Philos Trans R Soc Lond. B Biol Sci* 364.1535 (2009), pp. 3549–3557. ISSN: 0962-8436. DOI: 10.1098/rstb.2009.0138.

[290] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva. “A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality”. In: *Front Virtual Real* 3 (2022). DOI: 10.3389/frvir.2022.914392.

[291] M. Sousa, D. Mendes, S. Paulo, N. Matela, J. Jorge, and D. S. 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. DOI: 10.1145/3025453.3025566.

[292] G. Soza, R. Grosso, C. Nimsky, P. Hastreiter, and G. Fahlbusch R.nd Greiner. “Determination of the elasticity parameters of brain tissue with combined simulation and registration”. In: *Int J Med Robot Comput Assist Surg* 1.3 (2005), pp. 87–95. DOI: 10.1002/rcs.32.

[293] L. Spitz, M. Allgaier, A. Mpotsaris, D. Behme, B. Preim, and S. Saalfeld. “Segmentation of Circle of Willis from 7T TOF-MRI data and immersive exploration using VR”. In: *Curr Dir Biomed Eng* 8.1 (2022), pp. 129–132. DOI: doi:10.1515/cdbme-2022-0033.

[294] A. T. Stadie and R. A. Kockro. “Mono-Stereo-Autostereo: The Evolution of 3-Dimensional Neurosurgical Planning”. In: *Neurosurg* 72.1 (2013), pp. 63–77. DOI: 10.1227/NEU.0b013e318270d310.

[295] A. T. Stadie, R. A. Kockro, R. Reisch, A. Tropine, S. Boor, P. Stoeter, and A. Perneczky. “Virtual reality system for planning minimally invasive neurosurgery. Technical note”. In: *J Neurosurg* 108.2 (2008), pp. 382–394. DOI: 10.3171/JNS/2008/108/2/0382.

[296] A. T. Stadie, R. A. Kockro, L. Serra, G. Fischer, E. Schwandt, P. Grunert, and R. Reisch. “Neurosurgical craniotomy localization using a virtual reality planning system versus intraoperative image-guided navigation”. In: *Int J Comput Assist Radiol Surg* 6.5 (2011), pp. 565–572. DOI: 10.1007/s11548-010-0529-1.

[297] M. Stark, A. Tinelli, A. Malvasi, and O. Mynbaev. “Clinics in Surgery The Importance of Standardizing Surgical Methods for Comparison between Hospitals and Surgeon”. In: *Clin Surg* 2.1566 (2017), p. 1566.

[298] R. Starkov, L. Zhang, M. Bajka, C. Tanner, and O. Goksel. “Ultrasound simulation with deformable and patient-specific scatterer maps”. In: *Int J Comput Assist Radiol Surg* 14.9 (2019), pp. 1589–1599. DOI: 10.1007/s11548-019-02054-5.

[299] T. C. Steineke and D. Barbery. “Microsurgical clipping of middle cerebral artery aneurysms: Preoperative planning using virtual reality to reduce procedure time”. In: *Neurosurg Focus* 51.2 (2021), E12. DOI: 10.3171/2021.5.FOCUS21238.

[300] S. M. Strasberg, J. Belghiti, P.-A. Clavien, E. Gadzijev, J.O. Garden, W.-Y. Lau, M. Makuuchi, and R.W. Strong. “The Brisbane 2000 Terminology of Liver Anatomy and Resections”. In: *HPB* 2.3 (2000), pp. 333–339. DOI: 10.1016/S1365-182X(17)30755-4.

[301] K. Stucke. *Leberchirurgie: Grundlagen Grenzen Möglichkeiten.* Springer Berlin Heidelberg, 2013. ISBN: 9783662353455.

[302] R. Sucher, H. Guice, S. Recknagel, E. Sucher, K. Semmling, A. Lederer, S. Rademacher, U. Scheuermann, and D. Seehofer. “Analysis of recent open and laparoscopic liver resections in a German high-volume liver tumor center”. In: *Z Gastroenterol* 58.8 (2020), e204. DOI: 10.1055/s-0040-1716283.

[303] M. Süncksen, H. Bendig, M. Teistler, M. Wagner, O. J. Bott, and K. Dresing. "Gamification and virtual reality for teaching mobile x-ray imaging". In: *Proc. of International Conference on Serious Games and Applications for Health (SeGAH).* 2018, pp. 1–7. DOI: 10.1109/SeGAH.2018.8401364.

[304] W.-S. Tai. "A Ultrasonic Examination Simulator Based on Virtual Reality". In: *International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan).* 2023, pp. 513–514. DOI: 10.1109/ICCE-Taiwan58799.2023.10226820.

[305] M. Takac, J. Collett, R. Conduit, and A. De Foe. "Addressing virtual reality misclassification: A hardware-based qualification matrix for virtual reality technology". In: *Clin Psychol Psychother* 28.3 (2021), pp. 538–556. DOI: 10.1002/cpp.2624.

[306] K. Tang, K. Gerling, V. Vanden Abeele, L. Geurts, and M. Aufheimer. "Playful Reflection: Impact of Gamification on a Virtual Reality Simulation of Breastfeeding". In: *Proc. of the CHI Conference on Human Factors in Computing Systems.* 2023. DOI: 10.1145/3544548.3580751.

[307] Y. Tashiro, S. Miyafuji, D.-H. Hwang, S. Kiyofuji, T. Kin, T. Igarashi, and H. Koike. "GAuze-MIcrosuture-FICATION: Gamification in Microsuture Training with Real-Time Feedback". In: *Proc. of the Augmented Humans International Conference 2023.* 2023, pp. 15–26. DOI: 10.1145/3582700.3582704.

[308] S. Teodoro-Vite, C. Domínguez-Velasco, A. F. Hernández-Valencia, J. S. Pérez-Lomelí, and M. Á. Padilla-Castañeda. "Virtual Simulation of Brain Sylvian Fissure Exploration and Aneurysm Clipping with Haptic Feedback for Neurosurgical Training". In: *Augmented Reality, Virtual Reality, and Computer Graphics.* Cham: Springer, 2018, pp. 230–238. ISBN: 978-3-319-95281-9.

[309] S. Teodoro-Vite, J. Pérez-Lomelí, C. Domínguez-Velasco, A. F. Hernández-Valencia, M. Capurso-García, and M. Padilla-Castañeda. "A High-Fidelity Hybrid Virtual Reality Simulator of Aneurysm Clipping Repair With Brain Sylvian Fissure Exploration for Vascular Neurosurgery Training". In: *Simul Healthc* 16.4 (2021), pp. 285–294. DOI: 10.1097/SIH.0000000000000489.

[310] P. Thamrongrat and E. Lai-Chong Law. "Analysis of The Motivational Effect of Gamified Augmented Reality Apps for Learning Geometry". In: *Proc. of the Australian Conference on Human-Computer Interaction.* 2021, pp. 65–77. DOI: 10.1145/3441000.3441034.

[311] D. Theodoropoulos, D. Ioannou, and C. Katsanos. "ToMtool: An Interactive Multimedia Application to Support Training of Emotion Recognition and Theory of Mind Skills to Children with Autism Spectrum Disorder". In: *Proc. of the Pan-Hellenic Conference on Informatics.* 2023, pp. 223–229. DOI: 10.1145/3575879.3575997.

[312] N Thiruchelvam, S. Y. Lee, and A. K. H. Chiow. "Patient and port positioning in laparoscopic liver resections". In: *Hepatoma Res* 7 (2021), pp. 0-22. DOI: 10.20517/2394-5079.2020.144.

[313] G. Tieri, G. Morone, S. Paolucci, and M. Iosa. "Virtual reality in cognitive and motor rehabilitation: facts, fiction and fallacies". In: *Expert Rev Med Devices* 15.2 (2018), pp. 107–117. DOI: 10.1080/17434440.2018.1425613.

[314] A. M. Toda, P. H. D. Valle, and S. Isotani. "The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education". In: *Proc. of Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions.* 2018, pp. 143–156.

[315] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. "The Gamification User Types Hexad Scale". In: *Proc. of CHI PLAY.* 2016, pp. 229–243. ISBN: 9781450344562. DOI: 10.1145/2967934.2968082.

[316] M. Tonutti, D. Elson, G.-Z. Yang, A. Darzi, and M. Sodergren. "The role of technology in minimally invasive surgery: State of the art, recent developments and future directions". In: *Postgrad Med J* 93 (2017), pp. 159–167. DOI: 10.1136/postgradmedj-2016-134311.

[317] A. R. Torabi, A. A. Nazari, E. Conrad-Baldwin, K. Zareinia, and M. Tavakoli. "Kinematic design of linkage-based haptic interfaces for medical applications: a review". In: *Prog Biomed Eng* 3 (2021). DOI: 10.1088/2516-1091/abee66.

[318] M. Tuzer, A. Yazıcı, R. Türkay, M. Boyman, and B. Acar. "Multi-ray medical ultrasound simulation without explicit speckle modelling". In: *Int J Comput Assist Radiol Surg* 13.7 (2018), pp. 1009–1017. DOI: 10.1007/s11548-018-1760-4.

[319] C. M. Tyng. "The Influences of Emotion on Learning and Memory". In: *Front Psychol* 8 (2017), p. 1454. DOI: 10.3389/fpsyg.2017.01454.

[320] J. Uhlig, C. M. Sellers, S. M. Stein, and H. S. Kim. "Radiofrequency ablation versus surgical resection of hepatocellular carcinoma: contemporary treatment trends and outcomes from the United States National Cancer Database". In: *Eur Radiol* 29 (2019), pp. 2679–2689. DOI: 10.1007/s00330-018-5902-4.

[321] S. R. van der Kruk, R. Zielinski, H. MacDougall, D. Hughes-Barton, and K. M. Gunn. "Virtual reality as a patient education tool in healthcare: A scoping review". In: *Patient Educ Couns* 105.7 (2022), pp. 1928–1942. ISSN: 0738-3991. DOI: 10.1016/j.pec.2022.02.005.

[322] L. Villegas, B. E. Schneider, M. P. Callery, and D. B. Jones. "Laparoscopic skills training". In: *Surg Endosc* 17 (2003), pp. 1879–1888. DOI: 10.1007/s00464-003-8172-3.

[323] S. Vitale, J. I. Orlando, E. Iarussi, and I. Larrabide. "Improving realism in patient-specific abdominal ultrasound simulation using CycleGANs". In: *Int J Comput Assist Radiol Surg* 15.2 (2020), pp. 183–192. DOI: 10.1007/s11548-019-02046-5.

[324] M. H. M. Vlak, A. Algra, R. Brandenburg, and G. J. E. Rinkel. "Prevalence of unruptured intracranial aneurysms, with emphasis on sex, age, comorbidity, country, and time period: a systematic review and meta-analysis". In: *The Lancet Neurology* 10.7 (2011), pp. 626–636. ISSN: 1474-4422. DOI: 10.1016/S1474-4422(11)70109-0.

[325] J. Vozenilek, J. S. Huff, M. Reznek, and J. A. Gordon. "See One, Do One, Teach One: Advanced Technology in Medical Education". In: *Acad Emerg Med* 11.11 (2004), pp. 1149–1154. DOI: 10.1197/j.aem.2004.08.003.

[326] S. Wagner, J. Belger, F. Joeres, A. Thöne-Otto, C. Hansen, B. Preim, and P. Saalfeld. "iVRoad: Immersive virtual road crossing as an assessment tool for unilateral spatial neglect". In: *Comput & Graph* 99 (2021), pp. 70–82. DOI: 10.1016/j.cag.2021.06.013.

[327] J. Walker, J. P. Heudebert, M. Patel, J. D. Cleveland, A. O. Westfall, D. M. Dempsey, A. Guzman, A. Zinski, M. Agarwal, D. Long, J. Willig, and R. Lee. "Leveraging Technology and Gamification to Engage Learners in a Microbiology Curriculum in Undergraduate Medical Education". In: *Medical science educator* 32.3 (2022), pp. 649–655. DOI: 10.1007/s40670-022-01552-7.

[328] T. Walker, R. Bamford, and M. Finch-Jones. "Intraoperative ultrasound for the colorectal surgeon: current trends and barriers." In: *ANZ J Surg* 87.9 (2017), pp. 671–676. DOI: doi:10.1111/ans.14124.

[329] C. Wang, G. Chen, Z. Yang, and Q. Song. "Development of a Gamified Intervention for Children with Autism to Enhance Emotional Understanding Abilities". In: *Proc. of the International Conference on Digital Technology in Education.* 2023, pp. 47–51. DOI: 10.1145/3568739.3568749.

[330] D. Wang, D. Ma, M. Lun Wong, Y. Wáng, and J. Xiáng. "Recent advances in surgical planning & navigation for tumor biopsy and resection". In: *Quant Imaging Med Surg* 5.5 (2015), pp. 640–648. DOI: 10.3978/j.issn.2223-4292.2015.10.03.

[331] J. Wang, P. Fallavollita, L. Wang, M. Kreiser, and N. Navab. "Augmented reality during angiography: Integration of a virtual mirror for improved 2D/3D visualization". In: *Proc. of IEEE International Symposium on Mixed and Augmented Reality (ISMAR).* 2012, pp. 257–264. DOI: 10.1109/ISMAR.2012.6402565.

[332] M. Waqas, K. Vakharia, R. H. Dossani, G. B. Rajah, M. K. Tso, A. D. Gong, K. Rho, H. H. Rai, K. V. Snyder, E. I. Levy, A. H. Siddiqui, and J. M. Davies. "Transradial access for flow diversion of intracranial aneurysms: Case series". In: *Interv Neuroradiol* 27.1 (2021), pp. 68–74. DOI: 10.1177/1591019920938961.

[333] S. G. Warner, A. A. Alseidi, J. Hong, T. M. Pawlik, and R. M Minter. "What to expect when you're expecting a hepatopancreatobiliary surgeon: self-reported experiences of HPB surgeons from different training pathways". In: *HPB : the official journal of the International Hepato Pancreato Biliary Association* 17.9 (2015), pp. 785–790. DOI: doi:10.1111/hpb.12430.

[334] A. Webster, M. Poyade, S. Rooney, and L. Paul. "Upper limb rehabilitation interventions using virtual reality for people with multiple sclerosis: A systematic review". In: *Mult Scler Relat Disord* 47 (2021), p. 102610. DOI: 10.1016/j.msard.2020.102610.

[335] K. Werbach and D. Hunter. *For the Win: How Game Thinking can Revolutionize your Business.* Wharton Digital Press, 2012.

[336] W. Westera. "The devil's advocate: identifying persistent problems in serious game design". In: *Int J Serious Games* 9.3 (2022), pp. 115–124. DOI: 10.17083/ijsg.v9i3.547.

[337] L. H. Williams and T. Drew. "What do we know about volumetric medical image interpretation?: a review of the basic science and medical image perception literatures". In: *Cogn Res Princ Implic* 4.1 (2019), p. 21. DOI: 10.1186/s41235-019-0171-6.

[338] G. Wong, C. Zhu, A. Ahuja, and W. Poon. “Craniotomy and clipping of intracranial aneurysm in a stereoscopic virtual reality environment”. In: *Neurosurg* 61.3 (Sept. 2007), pp. 564–568. DOI: 10.1227/01.NEU.0000290904.46061.0D.

[339] Wooning, J. and Benmahdjoub, M. and Walsum, T. van and Marroquim, R. “AR-Assisted Craniotomy Planning for Tumour Resection”. In: *Eurographics Workshop on Visual Computing for Biology and Medicine*. 2021. DOI: 10.2312/vcbm.20211353.

[340] B. Xin, G. Chen, Y. Wang, G. Bai, X. Gao, J. Chu, J. Xiao, and T. Liu. “The Efficacy of Immersive Virtual Reality Surgical Simulator Training for Pedicle Screw Placement: A Randomized Double-Blind Controlled Trial”. In: *World Neurosurg* 124 (2018). DOI: 10.1016/j.wneu.2018.12.090.

[341] B. Yanagawa, R. Ribeiro, F. Naqib, J. Fann, S. Verma, and J. D. Puskas. “See one, simulate many, do one, teach one: cardiac surgical simulation”. In: *Curr Opin Cardiol* 34.5 (2019), pp. 571–577. DOI: 10.1197/j.aem.2004.08.003.

[342] Y. Yang, C. Fermüller, Y. Li, and Y. Aloimonos. “Grasp Type Revisited: A Modern Perspective on A Classical Feature for Vision”. In: *Proc. of IEEE International Conference on Computer Vision and Pattern Recogntion (CVPR)*. 2015, pp. 400–408. DOI: 10.1109/CVPR.2015.7298637.

[343] V. Yazdnian, A. Delavari, H. Moradi, S. Shams, M. Teymouri, and A. Rezaei. “Comprehensive and Gamified Rehabilitation System for Upper-Limb Impairment Treatments”. In: *Proc. of RSI International Conference on Robotics and Mechatronics (ICRoM)*. 2022, pp. 445–452. DOI: 10.1109/ICRoM57054.2022.10025334.

[344] C. T. Yeo, A. MacDonald, T. Ungi, A. Lasso, D. Jalink, B. Zevin, G. Fichtinger, and S. Nanji. “Utility of 3D Reconstruction of 2D Liver Computed Tomography/-Magnetic Resonance Images as a Surgical Planning Tool for Residents in Liver Resection Surgery”. In: *J Surg Educ* 75.3 (2018), pp. 792–797. ISSN: 1931-7204. DOI: 10.1016/j.jsurg.2017.07.031.

[345] A. W. K. Yeung, A. Tosevska, E. Klager, F. Eibensteiner, D. Laxar, J. Stoyanov, M. Glisic, S. Zeiner, S. T. Kulnik, R. Crutzen, O. Kimberger, M. Kletecka-Pulker, A. G. Atanasov, and H. Willschke. “Virtual and Augmented Reality Applications in Medicine: Analysis of the Scientific Literature”. In: *J Med Internet Res* 23.2 (2021), e25499. DOI: 10.2196/25499.

[346] S. Zachow, E. Gladilin, R. Sader, and H.-F. Zeilhofer. “Draw and cut: Intuitive 3D osteotomy planning on polygonal bone models”. In: *Proc. of Computer Assisted Radiology and Surgery*. 2003, pp. 362–369. DOI: 10.1016/S0531-5131(03)00272-3.

[347] H. Zhang, Y. Xu, J. Lin, W. Li, and Z. Shen. “A Serious Mobile Game for Neurodegenerative Diseases Rehabilitation and Risk Estimation”. In: *Proc. of International Conference on Crowd Science and Engineering*. 2022, pp. 103–107. DOI: 10.1145/3503181.3503198.

[348] Y. Zhu, D. Magee, R. Ratnalingam, and D. Kessel. “A virtual ultrasound imaging system for the simulation of ultrasound-guided needle insertion procedures”. In: *Proc. of Medical Image Understanding and Analysis*. 2006.

Acronyms

AR augmented reality

CAVE Cave Automatic Virtual Environment

CoW circle of Willis

CT computed tomography

DOFs degrees of freedom

HMD head-mounted display

IA intracranial aneurysm

IMI Intrinsic Motivation Inventory

IOUS intraoperative ultrasound

miniPXI Mini Player Experience Inventory

MR mixed reality

MRI magnetic resonance imaging

MSM mass-spring model

OR operating room

SBST Santa Barbara Solids Test

UI user interface

US ultrasound

VR virtual reality

XR xReality