

Institute for Visualization and Interactive Systems

University of Stuttgart  
Universitätsstraße 38  
D-70569 Stuttgart

Masterarbeit

# **Robotic Surgery Training in AR: Multimodal Record and Replay**

Fabian Raphael Krauthausen

**Course of Study:** Softwaretechnik  
**Examiner:** Prof. Dr. Michael Sedlmair  
**Supervisor:** Katherine J. Kuchenbecker, Ph.D.

**Commenced:** October 20, 2020

**Completed:** May 20, 2021



## Acknowledgments

I would first like to thank my thesis examiner *Prof. Dr. Michael Sedlmair* of the Institute for Visualization and Interactive Systems at the University of Stuttgart. Thank you for always taking the time to talk about the progress of the thesis whenever it was needed and for steering the thesis in a more meaningful direction.

I would also like to thank my supervisor *Katherine J. Kuchenbecker* of the Haptic Intelligence Department at the Max Planck Institute for Intelligent Systems. Thank you for enabling me to pursue such a meaningful project.

I would also like to thank my tutor *Maria-Paola Forte* of the Haptic Intelligence Department at the Max Planck Institute for Intelligent Systems. Thank you for all your help throughout the whole thesis process and for always helping out whenever it was needed.

I would also like to thank *Bernard Javot* of the Haptic Intelligence Department at the Max Planck Institute for Intelligent Systems. Thank you for all your help in manufacturing the training tasks and integrating the sensors.

I would also like to thank *Joey Burns* of the Haptic Intelligence Department at the Max Planck Institute for Intelligent Systems. Thank you for your help with all IT-related questions.

I would also like to thank the chief surgeons who participated in the study of this thesis. Thank you for your participation and effort to make the study possible.

Finally, I would like to thank my parents *Siglinde Krauthausen* and *Herbert Krauthausen*, and my brother *Heiko Krauthausen*. Thank you for always supporting me during my studies and on whichever path I decided to take in life.

This accomplishment would not have been possible without all of you. Thank you.





## Abstract

Robotic surgery enhances surgeon capabilities and lowers patient risk compared to traditional minimally invasive surgery. However, learning robotic surgery can be challenging for novice surgeons. On the one hand, existing physical approaches are time-consuming for expert surgeons, as they require experts to supervise the training process. On the other hand, existing virtual approaches do not provide a physical experience for novice surgeons, resulting in lower skill transfer to real surgeries. To overcome these challenges and to combine the advantages of both approaches, a mixed approach with augmented reality could be used. This approach could relieve expert surgeons from supervising training processes and provide novice surgeons with physical experiences to improve skill transfer to real surgeries. In this work, we develop a multimodal record and replay platform in augmented reality for robotic surgery training using the Intuitive da Vinci Surgical System. With this platform, we aim to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. Moreover, we aim to explore different concepts for motion guidance and surgical skill evaluation in augmented reality.

The developed platform allows expert surgeons to record surgical procedures with multiple modalities and novice surgeons to replay and train with the attained multimodal recordings, including visual, haptic, and auditory feedback. Visual feedback is provided by recording left and right camera streams of a stereo endoscope and directly overlaying the recorded videos on the surgeon's stereoscopic view. Haptic feedback is achieved by recording left and right robotic instrument vibrations with accelerometers attached to the robotic instruments and replaying the recorded vibrations on voice coil actuators mounted to the surgeon handles. Finally, auditory feedback is enabled by recording verbal explanations of a procedure and playing back the recordings on the surgeon speakers. The platform also incorporates motion guidance concepts in the form of ghost tools. For the development of the platform, we used ROS and OpenCV with C++ and Python.

With this platform, we conducted an exploratory study with three chief surgeons to evaluate the concepts of the developed platform and to further explore the influence of different modalities and motion guidance concepts. For this purpose, we manufactured a box trainer with two standardized tasks. Further, we placed a force sensor below the task board and used the previously attached accelerometers on the robotic instruments to explore possibilities for quantitative evaluation of surgical skill performances. In the study, participants were first asked to record performances for the two tasks without and with verbal explanations. Next, they were asked to replay and train with their multimodal recordings. After each phase, they were requested to evaluate their experience and to provide suggestions for improvement.

Overall, we found that multimodal record and replay in augmented reality is a promising approach for robotic surgery training. In terms of modalities, we found that chief surgeons generally prefer a combination of visual and auditory feedback. Regarding the visual feedback, the surgeons prefer a direct overlay while replaying as the focus should be on the recorded performance. In contrast, a corner view is preferred while training as surgeons should focus on their own performance. Regarding the auditory feedback, the surgeons found verbal explanations beneficial in the early stages of training. In terms of motion guidance, our results showed that visual cues are preferred only during challenging sub-tasks of surgical procedures. In contrast, no preference was expressed between ghost tools and trajectories. In the future, we aim to use the recordings of the chief surgeons to conduct a user study with novice surgeons to assess the effects of the developed training platform on their learning curve.



# Contents

<b>1</b>	<b>Introduction</b>	<b>15</b>
1.1	Robot-Assisted Minimally Invasive Surgery . . . . .	16
1.1.1	Surgical Approaches . . . . .	16
1.1.2	da Vinci Surgical System . . . . .	17
1.2	Training in Robotic Surgery . . . . .	19
1.2.1	Visual, Auditory, Haptic and Multimodal Feedback . . . . .	21
1.2.2	Motion Guidance . . . . .	21
1.2.3	Augmented and Virtual Reality . . . . .	22
1.2.4	Physical, Augmented and Virtual Simulators . . . . .	23
1.2.5	Training Curricula . . . . .	26
1.2.6	Training Approaches and Processes . . . . .	27
1.3	Aim of the Work . . . . .	30
1.3.1	RQ1 - Multimodal Record and Replay in AR . . . . .	30
1.3.2	RQ2 - Motion Guidance and Skill Evaluation in AR . . . . .	30
1.4	Outline . . . . .	31
<b>2</b>	<b>State of the Art</b>	<b>33</b>
2.1	Multimodal Feedback for Robotic Surgery Training . . . . .	33
2.1.1	Visual Feedback . . . . .	34
2.1.2	Auditory Feedback . . . . .	37
2.1.3	Haptic Feedback . . . . .	37
<b>3</b>	<b>System Development</b>	<b>41</b>
3.1	Existing Systems . . . . .	41
3.1.1	Hardware Setup . . . . .	41
3.1.2	Software Basis . . . . .	41
3.1.3	VerroTouch for Haptic Feedback . . . . .	43
3.2	Frameworks and Libraries . . . . .	45
3.2.1	ROS . . . . .	45
3.2.2	OpenCV . . . . .	46
3.3	Software Development Process . . . . .	47
3.3.1	Goals . . . . .	47
3.3.2	Requirements . . . . .	47
3.3.3	Architecture . . . . .	50
3.3.4	Development . . . . .	57
3.3.5	Integration of VerroTouch . . . . .	66
<b>4</b>	<b>Chief Surgeon Study</b>	<b>69</b>
4.1	Research Questions . . . . .	69
4.1.1	RQ1 - Multimodal Record and Replay in AR . . . . .	69

4.1.2	RQ2 - Motion Guidance and Skill Evaluation in AR . . . . .	70
4.2	Experimental Setup . . . . .	71
4.3	Participants . . . . .	77
4.4	Procedure . . . . .	78
4.5	Iterations . . . . .	80
<b>5</b>	<b>Results</b>	<b>83</b>
5.1	Integrability of AR Platform in Training Processes . . . . .	83
5.1.1	Integration Possibilities in Common Training Processes . . . . .	83
5.1.2	Training Task Satisfaction . . . . .	85
5.1.3	Additional Suggestions . . . . .	86
5.2	Record Functionality . . . . .	87
5.2.1	Modality Preference for Recording . . . . .	87
5.3	Replay Functionality . . . . .	88
5.3.1	Visual Overlay Satisfaction for Replaying . . . . .	88
5.3.2	Modality Preference for Replaying . . . . .	88
5.3.3	Additional Suggestions . . . . .	89
5.4	Train Functionality with Motion Guidance . . . . .	89
5.4.1	Visual Overlay and Motion Guidance Preference for Training . . . . .	90
5.4.2	Modality Preference for Training . . . . .	91
5.4.3	Training Mode Preference . . . . .	92
5.4.4	Additional Suggestions . . . . .	93
5.5	Sensors for Surgical Skill Evaluation . . . . .	94
5.5.1	Sensor Preference . . . . .	94
5.5.2	Metric Suggestions . . . . .	95
5.5.3	Measurements . . . . .	95
<b>6</b>	<b>Discussion</b>	<b>97</b>
6.1	Main Findings with Implications . . . . .	97
6.1.1	For RQ1 - Multimodal Record and Replay in AR . . . . .	97
6.1.2	For RQ2 - Motion Guidance and Skill Evaluation in AR . . . . .	99
6.2	Limitations and Challenges . . . . .	101
6.3	Future Work . . . . .	101
<b>7</b>	<b>Conclusion</b>	<b>107</b>
	<b>Bibliography</b>	<b>109</b>
<b>A</b>	<b>Appendix</b>	<b>115</b>
A.1	Informed Consent Document . . . . .	116
A.2	Demographic Questionnaire . . . . .	122
A.3	Interview Structure . . . . .	124
A.4	Questionnaire for Recording . . . . .	129
A.5	Questionnaire for Replaying . . . . .	132
A.6	Questionnaire for Training . . . . .	136
A.7	Questionnaire for Motion Guidance . . . . .	141
A.8	Questionnaire for Surgical Skill Evaluation . . . . .	144

## List of Figures

1.1	Envisioned robotic surgery training process with AR . . . . .	16
1.2	da Vinci Si HD Surgical System . . . . .	18
1.3	da Vinci Si HD dual console . . . . .	18
1.4	da Vinci instrument arm with EndoWrist instruments and surgeon handles . . . . .	20
1.5	da Vinci stereo endoscope with stereoscopic vision . . . . .	20
1.6	Representation of Reality-Virtuality Continuum . . . . .	23
1.7	Physical simulator dry lab tasks . . . . .	25
1.8	Common VR simulators . . . . .	26
1.9	VR simulator dry lab tasks . . . . .	27
1.10	Physical simulator FRS tasks with box trainer . . . . .	27
1.11	Simplified robotic surgery training process . . . . .	29
2.1	Effectiveness of feedback strategies for motor learning . . . . .	34
2.2	Visual representation of forces on robotic instrument tips . . . . .	35
2.3	Graphical force feedback on robotic instruments . . . . .	36
2.4	Ghost tools of expert surgeon during mentoring a robotic surgery training . . . . .	37
2.5	VR simulator motion guidance for knot tying task . . . . .	38
2.6	VR simulator motion guidance for suturing task . . . . .	38
2.7	Multimodal motion guidance system . . . . .	40
3.1	Concept of visual AR . . . . .	42
3.2	VerroTouch system components . . . . .	44
3.3	VerroTouch signal flow of one channel for transforming surgical tool vibrations to haptic and auditory representations . . . . .	44
3.4	VerroTouch signal flow of one channel for recording surgical tool vibrations . . . . .	45
3.5	ROS Computation Graph . . . . .	45
3.6	Use-case diagram of AR system . . . . .	49
3.7	Package diagram of AR system . . . . .	51
3.8	Node-topic diagram of AR system . . . . .	52
3.9	Sequence diagram for record expert . . . . .	54
3.10	Sequence diagram for replay expert . . . . .	55
3.11	Sequence diagram for train expert . . . . .	56
3.12	Record functionality of stereo endoscope videos, instrument vibrations, and verbal explanations for knot tying task . . . . .	59
3.13	Record functionality of stereo endoscope videos, instrument vibrations, and verbal explanations for suturing task . . . . .	60
3.14	Replay functionality of overlaid stereo endoscope videos, instrument vibrations, and verbal explanations for knot tying task . . . . .	61

3.15	Replay functionality of overlaid stereo endoscope videos, instrument vibrations, and verbal explanations for suturing task . . . . .	62
3.16	Train functionality of overlaid stereo endoscope videos with initial green filter for knot tying and suturing task . . . . .	63
3.17	Train functionality of overlaid stereo endoscope videos without green filter for suturing task . . . . .	64
3.18	Detect and highlight functionality of left stereo endoscope video for instruments and grippers . . . . .	65
3.19	VerroTouch record and replay signal flow of one channel . . . . .	66
3.20	VerroTouch acceleration sensors on robotic instruments . . . . .	67
3.21	VerroTouch main receiving unit . . . . .	67
3.22	VerroTouch record and replay unit with voltage divider, voltage amplifier and switch	67
3.23	Voice coil actuators on surgeon handles . . . . .	68
4.1	Experimental setup with the clinical da Vinci Si HD Surgical System . . . . .	72
4.2	Box trainer with additional door and LEDs . . . . .	74
4.3	Training task knot tying . . . . .	76
4.4	Training task suturing . . . . .	76
4.5	Mini40-E Transducer force sensor below task board . . . . .	77
4.6	VerroTouch acceleration sensors on robotic instruments . . . . .	77
4.7	Procedure for chief surgeon study . . . . .	78
5.1	Sample recorded force and acceleration data . . . . .	96
6.1	Confirmed robotic surgery training process with AR . . . . .	98
6.2	Suggested training process for robotic surgery . . . . .	100
6.3	Envisioned left and right corner view for training functionality . . . . .	104
6.4	Motion guidance concepts from VR simulator for knot tying task . . . . .	104
6.5	Motion guidance concepts from VR simulator for suturing task . . . . .	105
6.6	Draft of GUI . . . . .	105
6.7	Envisioned recording process with GUI . . . . .	105

## List of Tables

1.1	Benefits and drawbacks of MIS compared to open surgery . . . . .	17
1.2	Benefits and drawbacks of robotic surgery compared to MIS . . . . .	17
1.3	Benefits and drawbacks of physical simulators . . . . .	24
1.4	Benefits and drawbacks of AR simulators . . . . .	24
1.5	Benefits and drawbacks of VR simulators . . . . .	24
1.6	VR simulator FRS tasks with standardized procedures . . . . .	28
4.1	Demographic information of participants . . . . .	78
4.2	Iterations of experimental setup . . . . .	80
4.3	Iterations of AR system . . . . .	81
4.4	Iterations of questionnaires . . . . .	81
4.5	Iterations of procedures . . . . .	82





# Acronyms

**AR** Augmented Reality. 22

**CV** Computer Vision. 46

**dV** da Vinci. 26

**dVSS** da Vinci Skills Simulator. 26

**FRS** Fundamentals of Robotic Surgery. 26

**GEARS** Global Evaluative Assessment of Robotic Skills. 23, 24

**GEARS** Robotic Objective Structured Assessment of Technical Skills. 24

**MIS** Minimally Invasive Surgery. 16

**RAMIS** Robot-Assisted Minimally Invasive Surgery. 16

**ROS** Robot Operating System. 45

**RoSS** Robotic Surgery Simulator. 26

**RQs** Research Questions. 30

**VR** Virtual Reality. 22



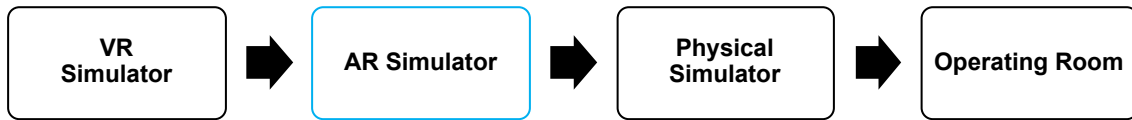
# 1 Introduction

Robotic surgery enhances surgeon capabilities and lowers patient risk compared to traditional minimally invasive surgery [SV09]. However, learning robotic surgery can be challenging for novice surgeons [SBKN17]. On the one hand, existing physical approaches are time-consuming for expert surgeons, as they require experts to supervise the training process [LMK+12]. On the other hand, existing virtual approaches are not providing a physical experience for novice surgeons, resulting in lower skill transfer to real surgeries [CMD+20]. To overcome these challenges and to combine the advantages of both approaches, a mixed approach with augmented reality could be used. This approach could relieve expert surgeons from supervising training processes and provide novice surgeons with physical experiences to improve skill transfer to real surgeries.

Hence, an augmented reality approach could be used to fill the gap between existing physical and virtual approaches (see Figure 1.1) due to the following reasons:

- **Combining Advantages of VR and Physical Simulators:** AR simulators could combine the advantages of VR and physical simulators. Through this, expert surgeons could be relieved from supervising training processes, and novice surgeons could be provided with physical experiences to improve skill transfer to real surgeries.
- **Independent of Underlying Surgical Robot:** AR simulators could be used independently of the underlying surgical robot, as these simulators only access the vision pipeline of the stereo endoscope and the surgeon's stereoscopic view of the instruments and task materials. Through this, AR simulators could be used for a variety of surgical robots.
- **Time Efficient for Expert Surgeons:** AR simulators could enable expert surgeons to record relevant procedures one time and train various novice surgeons with the recording without the need to supervise the training process.
- **Usage of Existing Surgical Recordings for Novice Surgeons:** AR simulators could be used to train novice surgeons with existing stereoscopic recordings of expert surgeons performing real surgeries.
- **Improved Skill Transfer for Novice Surgeons:** AR simulators could be used for novice surgeons to increase skill transfer from trained tasks to real surgeries and hence reduce the required training time with physical simulators and expert surgeons.
- **Learning from Best Expert Surgeons for Novice Surgeons:** AR simulators could enable all novice surgeons to learn procedures from the best expert surgeons.
- **Training Availability for Novice Surgeons:** AR simulators could enable novice surgeons to train procedures at any time, as these simulators do not require supervision by expert surgeons for direct one-to-one teaching and are hence always available for novice surgeons.

- **Learning Possibility for all Novice Surgeons:** AR simulators could enable all novice surgeons to learn robotic surgery, as these simulators do not require supervision by expert surgeons for direct one-to-one teaching and are hence equally usable for every novice surgeon with interest in learning robotic surgery.



**Figure 1.1:** Envisioned robotic surgery training process with AR.

In this chapter, an introduction to robotic surgery and its training as well as to key concepts such as augmented reality, multimodal feedback, and motion guidance will be provided.

## 1.1 Robot-Assisted Minimally Invasive Surgery

Robot-Assisted Minimally Invasive Surgery (RAMIS), also known as *robot-assisted surgery* or *robotic surgery*, is a surgical approach in which surgeons use robotic systems to perform surgeries through small incisions in the patient's body, which is generally known as Minimally Invasive Surgery (MIS) or *laparoscopic surgery*. Thereby robotic surgery enhances surgeon capabilities and lowers patient risk compared to traditional MIS [SV09]. To perform robotic surgery, the surgeon sits at a console and moves manipulators to remotely control robotic arms (see Figure 1.2) [GS00].

Robotic surgery is mostly applied within the following surgical specialties [DMS+21]:

- **Urology:** Surgeries focusing on the male and female urinary tract and the male reproductive system.
- **Gynecology:** Surgeries focusing on the female reproductive system.
- **Visceral Surgery:** Surgeries focusing on the abdominal tract. This surgical specialty is also known as *general abdominal surgery*.

This concept will be presented in more detail in the following.

### 1.1.1 Surgical Approaches

For performing surgeries, there exist different approaches. In the following, these different surgical approaches will be introduced and compared to each other regarding the benefits and drawbacks of each:

- **Open Surgery:** Open surgery is a classical surgical approach that requires large incisions in patients' bodies. Here, the surgeon opens large areas in the patient's body to operate with open view and access. Through this, surgeons can directly see the surgical field for visual feedback and can also directly manipulate tissue for haptic feedback [WGJD08].

- **MIS:** MIS is a surgical approach that tries to minimize incision size in patients' bodies to improve wound healing, reduce associated pain, and decrease infection risks. Here, the surgeon performs the surgery through small incisions using laparoscopic instruments [GS00]. Through this, surgeons can only see the surgical field through displays resulting in limited visual feedback and can also only manipulate tissue with laparoscopic instruments resulting in distorted haptic feedback [WGJD08]. The benefits and drawbacks of MIS compared to Open Surgery can be seen in Table 1.1.
- **Robotic Surgery:** Robotic surgery is a surgical approach that uses robotic systems through small incisions in patients' bodies. Here, the surgeon sits at a console and moves manipulators to control robotic arms remotely [GS00]. Through this, surgeons can only see the surgical field through stereoscopic displays resulting in distorted 3D visual feedback, and can also only manipulate tissue with robotic instruments controlled by handles resulting in missing haptic feedback [WGJD08]. The benefits and drawbacks of robotic surgery compared to MIS can be seen in Table 1.2.

MIS compared to Open Surgery	
Benefits	Drawbacks
Smaller incision size Reduced pain levels Reduced wound healing and recovery time Reduced infection risk	Long and thin laparoscopic instruments Hand tremor Constrained operation space Constrained view

**Table 1.1:** Benefits and drawbacks of MIS compared to open surgery [WGJD08].

Robotic Surgery compared to MIS	
Benefits	Drawbacks
Improved vision, digital zoom and camera stability Improved control and dexterity with motion scaling Improved ergonomics for long surgeries Filtered hand tremor Reduced patient risk Possibility for telesurgery and telementoring	High costs Large size of system No haptic feedback Long training process for novice surgeons

**Table 1.2:** Benefits and drawbacks of robotic surgery compared to MIS [GS00; SV09; WGJD08].

### 1.1.2 da Vinci Surgical System

The *da Vinci Surgical System* (by Intuitive Surgical Inc., Sunnyvale, USA) is a robotic platform designed to perform complex surgeries using a minimally invasive approach. It is a seven-degree-of-freedom teleoperated system allowing surgeons to manipulate robotic instruments while sitting at a console away from the patient (see Figure 1.2). It commonly consists of a *surgeon console*, a *patient cart*, and a *vision cart* [GS00].

The *da Vinci Si HD Surgical System* is one recent version of the *da Vinci Surgical Systems*. It also consists of the previously described *surgeon console*, *patient cart* and *vision cart*. In addition, a second surgeon console can be connected to the surgical system to operate in dual console mode. This second surgeon console is also commonly called *dual console*. It allows two surgeons to perform surgical procedures with the same robotic platform by sharing control over the robotic instruments (see Figure 1.3) [Sur12].

These components and the characteristics of the *da Vinci Si HD Surgical System* will be described in more detail in the following.



**Figure 1.2:** *da Vinci Si HD Surgical System*. Adapted from Intuitive Surgical Inc. [Sur12].



**Figure 1.3:** *da Vinci Si HD dual console*. Reprinted from Intuitive Surgical Inc. [Sur12].

**Surgeon Console** The surgeon console is the component that enables the surgeon to control the robotic arms and the stereo endoscope. It further enables the surgeon to see the left and right video streams of the stereo endoscope. Hence, it is the user interface for the surgeon which allows moving handles to remotely control the robotic arms for surgery [Sur12].

It consists of two *surgeon handles*, a *stereo viewer*, a *touchpad*, a *left-side and right-side pods*, and a *footswitch panel*. The *surgeon handles* allow the surgeon to control the robotic arms and the stereo endoscope (see Figure 1.4). The *stereo viewer* enables the surgeon to see the left and right video channels of the stereo endoscope (see Figure 1.5). The *touchpad*, *left-side and right-side pods*, and *footswitch panel* provide further functionalities such as controlling surgical activities, ergonomic controls, power and emergency stop buttons [Sur12].

**Patient Cart** The patient cart is the component containing the robotic arms with instruments to perform surgery with and the stereo endoscope to see the surgical scene. Hence, it is controlled by the surgeon to perform surgery with [Sur12].

It consists of *setup joints*, three *instrument arms*, one *camera arm*, and one *motor drive*. The three *instrument arms* are the interface for *EndoWrist* instruments to manipulate patient tissue with and are controlled by the surgeon handles (see Figure 1.4). Common *EndoWrist* instruments are *large needle drivers* for holding tissue or sutures, and *monopolar curved scissors* for cutting tissue or sutures. The one *camera arm* is the interface for the *3D endoscope* to provide left and right video channels for the surgeon and are also controlled by the surgeon handles (see Figure 1.5) [Sur12].

**Vision Cart** The vision cart is the component for processing visual information of the stereo endoscope. Hence, it is the video processing point of the system [Sur12].

It consists of a *core*, an *instrument control box*, an *illuminator*, an *endoscope with HD stereo camera head and HD camera control unit*, and a *touchscreen*. The *endoscope* used by the da Vinci Si HD is a 3D endoscope, also called *stereo endoscope*, with either straight (0 degree) or angled (30 degree) tip. When combined with the *HD stereo camera head* and the *HD camera control unit*, it provides left and right video channels for the surgeon (see Figure 1.5) and can also be controlled by the surgeon handles. It offers a 60-degree-field-of-view, a 6-10x magnification, and interlaced videos to double the perceived frame rate. The *touchscreen* allows to control system settings and to view the surgical image [Sur12].

## 1.2 Training in Robotic Surgery

The concept of training is commonly based on *motor learning*, which describes a lasting change of motor performance achieved by training [SRRW13]. Hence, training in robotic surgery describes a lasting change of motor performance in regard to performing surgical procedures with robotic systems achieved by training. However, learning robotic surgery can be challenging for novice surgeons [SBKN17]. This concept will be presented in more detail in the following.



**Figure 1.4:** da Vinci instrument arm with EndoWrist instruments and surgeon handles. Left: da Vinci instrument arm with attached EndoWrist instrument. Top right: EndoWrist instruments. Bottom right: Surgeon handles. Reprinted from Intuitive Surgical Inc. [Sur12].



**Figure 1.5:** da Vinci stereo endoscope with stereoscopic vision. Left: Stereo endoscope. Right: Stereo viewer.



### 1.2.1 Visual, Auditory, Haptic and Multimodal Feedback

Different modalities can be used to convey information to the user, whereas a modality is a stimulus for one human sensory organ [SRRW13]. The usage of these modalities for feedback, either on their own or in combination, will be elaborated in the following.

**Unimodal Feedback** Unimodal feedback uses one modality to provide feedback to the user. The most common modalities include the following [SRRW13]:

- **Visual:** The surgeon can see feedback.
- **Auditory:** The surgeon can hear feedback.
- **Haptic:** The surgeon can feel feedback.

**Multimodal Feedback** Multimodal feedback uses more than one modality to provide feedback to the user. The most common combinations of modalities include the following [SRRW13]:

- **Audio-visual:** Combination of auditory and visual feedback.
- **Visuo-haptic:** Combination of visual and haptic feedback.
- **Audio-haptic:** Combination of auditory and haptic feedback.
- **Audio-visuo-haptic:** Combination of auditory, visual, and haptic feedback.

**Haptic Feedback** Haptic feedback consists of both *tactile* and *kinesthetic information*, which are common when exploring or manipulating objects. In the human body, these information are obtained through *tactile* and *kinesthetic perception* respectively, which can be described as following [SRRW13; WGJD08]:

- **Tactile Perception:** Perception of sensations on the skin, such as vibration, pressure, texture, and temperature. Thereby tactile perception relies on receptors in the skin, in particular on mechanoreceptors. This perception is also called *discriminative touch* or *cutaneous sense*.
- **Kinesthetic Perception:** Perception of sensations related to the pose of the body, such as position, movement and forces. Thereby kinesthetic perception relies on receptors in muscles and tendons.

### 1.2.2 Motion Guidance

Motion guidance provides information for the user to correctly perform a movement. Traditionally, motion guidance was offered by trainers or physiotherapists through verbal or physical advice and correction while observing movements. However, this approach of an external advisor is not always applicable, as sometimes exercises need to be practiced alone at home, or as the presence of the advisor is not possible due to time or distance constraints. In these unsupervised cases, motion guidance can provide lacking feedback to ensure that movements are executed correctly. Thereby movements can be characterized by *plane*, *range*, *extent*, and *rate* of movements, as well as by

*positions* or *angles* to maintain [TYB+15]. Nowadays, motion guidance is increasingly used for learning purposes when real-world training is safety-critical or expensive, such as preparing medical students for surgery [YAM+20].

In motion guidance, different modalities can be used to convey information to the user. Similar to providing feedback to the user as described in Section 1.2.1, these consist of *visual motion guidance*, *auditory motion guidance*, *haptic motion guidance*, and *multimodal motion guidance* [YAM+20]. *Visual motion guidance* is mostly the prevalent modality to guide the movements of users. It can correct the body parts of users by emphasizing limbs for awareness correction, by superimposing target trajectories, or by showing scores. *Haptic motion guidance* can be used to correct body part positions of users. For this purpose, primarily vibration actuators are used for tactile actuation. *Auditory motion guidance* can be used to communicate 3D positions or directions to users [SFO+12].

For *visual motion guidance*, different *perspectives* can be used to visualize the scene for the user. These consist of the *first-person perspective*, the *third-person perspective*, and the *mirror-person perspective* [YAM+20]. Further, the following approaches can be used to guide the movements of users:

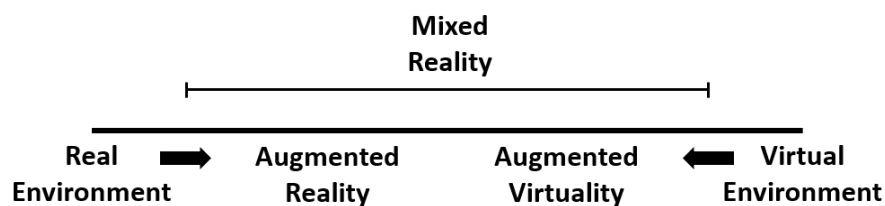
- **Video Tutorials:** Live or prerecorded video tutorials can be used to convey tasks and motions effectively. However, it can be difficult for users to follow the instructions provided in video tutorials precisely. This is because motion paths and velocities need to be translated from 2D screens into 3D motions [YAM+20].
- **Trajectories:** Trajectories are visualized movement paths for the user to follow. This type of motion guidance provides the user with information about the desired path of positions over time as future paths with directions to follow. It further allows the user to compare the own performance with the predefined path. Thereby, trajectories are often used for guiding limbs or instruments along a predefined path with corrective adjustments whenever users deviate from the desired path. Here, trajectories can be visualized on 2D screens or 3D in-air [SK04; YAM+20].
- **Ghost Tools:** Ghost tools are visualized structures to show the user a goal position and orientation. This type of motion guidance provides the user with information about desired position and orientation at the current time as current posture information to follow. Thereby, ghost tools are often used for goal postures of limbs, gestures, or instruments. Here, ghost tools can be visualized on 2D screens or 3D in-air [JSC+17].

### 1.2.3 Augmented and Virtual Reality

Augmented Reality (AR) and Virtual Reality (VR) are commonly defined within a Reality-Virtuality continuum as illustrated in Figure 1.6. This continuum can be described as follows [Azu97; MTUK95]:

- **Real Environment:** Environment consisting only of real objects, such as viewing real-world scenes in person.
- **Mixed Reality:** Environment consisting of both real and virtual objects.

- **Augmented Reality:** Real environment with real objects and integrated virtual objects. This subclass belongs to the larger class of Mixed Reality.
- **Augmented Virtuality:** Virtual environment with virtual objects and integrated real objects. This subclass belongs to the larger class of Mixed Reality.
- **Virtual Environment:** Environment consisting only of virtual objects, such as viewing computer graphic simulations via a monitor or immersive display. This subclass is also commonly called *Virtual Reality*.



**Figure 1.6:** Representation of Reality-Virtuality Continuum. Adapted from Milgram et al. [MTUK95].

*AR systems* can also be defined as systems that integrate virtual objects into real environments with real objects. Through this, it appears to the user as if virtual objects coexist with real objects in the real environment. Thereby the following properties can be defined for AR systems [ABB+01]:

- **Augmentation:** AR systems combine real and virtual objects in real environments.
- **Interaction:** AR systems run interactively and in real-time.
- **Registration:** AR systems register real and virtual object with each other.

### 1.2.4 Physical, Augmented and Virtual Simulators

For training in robotic surgery, hands-on practical skill development is most important for novice surgeons. Here, surgical simulation offers an alternative to training robotic surgery directly in the operating room. The benefits of simulators for robotic surgery training include the possibility to train procedures repeatedly, without expert surgeons, and also without potential harm to clinical patients [SBKN17].

There exist different simulator types to help novice surgeons learn robotic surgery. These simulator types consist of *physical simulators*, *AR simulators*, and *VR simulators* [LGL16; SBKN17]. An important concept for simulators is called *skill transfer*. It describes the rate at which skills acquired within simulator environments can be transferred to real environments, such as to real surgeries on patients [CMD+20]. Another important concept is called *surgical skill evaluation*. For VR simulators, common simulators provide an objective skill assessment with real-time feedback for novice surgeons and summarized assessments for expert surgeons [LGL16]. For physical simulators, structured human grading is commonly used to evaluate surgeon skills. Here, expert surgeons watch the performances of novice surgeons and rate their performance using validated skill assessment tools. Common skill assessment tools include the *Global Evaluative Assessment of Robotic Skills*

(GEARS) [GGS+12] and the *Robotic Objective Structured Assessment of Technical Skills (GEARS)* [SGG+14]. However, these skill assessment tools are based on qualitative expert evaluations, leading to subjective ratings for novice surgeons and consumed time for expert surgeons [LGL16].

The benefits and drawbacks of each simulator type are summarized in Table 1.3, Table 1.4, and Table 1.5. In the following, each simulator type will be described in more detail.

<b>Physical Simulator</b>	
Benefits	Drawbacks
High skill transfer to real surgeries Real tasks with real interactions Realistic haptic feedback Cost-effective setup	Requires constant supervision by expert surgeon Less flexibility due to physical tasks Less interactivity due to missing virtual content Subjective skill assessment Damaging of surgical robots

**Table 1.3:** Benefits and drawbacks of physical simulators [CMD+20; LGL16; SBKN17].

<b>AR Simulator</b>	
Benefits	Drawbacks
Medium skill transfer to real surgeries Requires only initial introduction of expert surgeon Real tasks with real interactions Possibility for interactivity due to virtual content Possibility for objective skill assessment Possibility for less costs	Effort to combine physical and virtual world Less flexibility due to physical tasks Missing skill assessment protocol Damaging of surgical robots

**Table 1.4:** Benefits and drawbacks of AR simulators [LGL16].

<b>VR Simulator</b>	
Benefits	Drawbacks
Requires no time of expert surgeons Flexibility due to virtual tasks Interactivity due to virtual content Objective skill assessment No damaging of surgical robots Less costs	Lower skill transfer to real surgeries Virtual tasks with virtual interactions Unrealistic haptic feedback

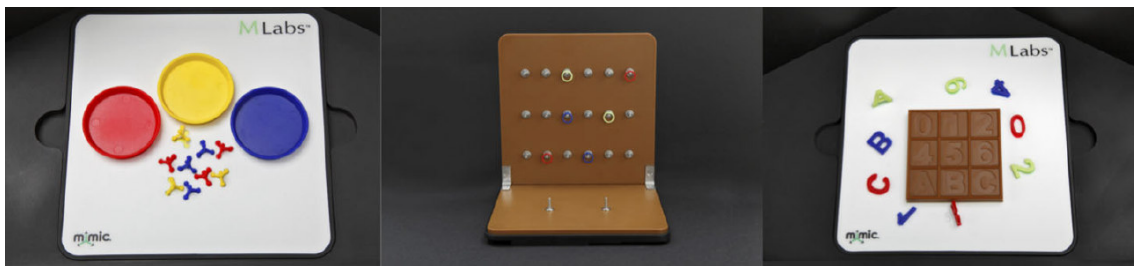
**Table 1.5:** Benefits and drawbacks of VR simulators [CMD+20; LGL16; SBKN17].

### Physical Simulators

Physical simulators use the actual robotic surgical system to train on physical objects in the physical world. Hence, novice surgeons train with physical objects in the physical world. These simulators are widely used in current training processes for robotic surgery and represent an essential step before training in operating rooms [SBKN17].

Common physical simulation models include the following [SBKN17; ZLH+19]:

- **Dry Lab Simulation:** In dry lab simulation, novice surgeons train with the actual robotic surgical system on inanimate models (see Figure 1.7). This model represents the simplest training model for novice surgeons. Due to the usage of inanimate models, it is a cost-effective and reproducible method for training robotic surgery. However, inanimate models are often very simplified compared to human anatomies and hence only provide very different experiences to performing procedures on patients in operating rooms.
- **Wet Lab Simulation:** In wet lab simulation, novice surgeons train with the actual robotic surgical system on animal parts, human body parts, and live animals. This model represents a more realistic training model for novice surgeons. Due to the usage of animal and human body parts, the model provides very similar experiences to performing procedures on patients in operating rooms. However, animal and human body parts usage often result in high costs, ethical concerns, and regulatory issues. As a consequence, wet lab simulation is only available to a limited number of novice surgeons.



**Figure 1.7:** Physical simulator dry lab tasks. Reprinted from Mimic Simulation [Simd].

### AR Simulators

AR simulators use the actual robotic surgical system to train with physical trainer boxes in the physical world, augmented with virtual content. Hence, novice surgeons train with physical objects and overlaid virtual content in the physical world. These simulators are still in the research stages for robotic surgery training, whereas some are already applied for MIS training. As such, there are currently no standard AR training simulators used for robotic surgery training in practice [LGL16].

### VR Simulators

VR simulators generally only use the surgeon console from robotic surgical systems in combination with their own created components or create their own surgeon consoles to train on virtual objects in the virtual world. These simulators are also widely used in current training processes for robotic surgery and represent an essential step before training with actual robotic surgical systems on physical simulators or in operating rooms [SBKN17].

Common VR training simulators include the *da Vinci Skills Simulator (dVSS)* (by Intuitive Surgical Inc.), the *da Vinci (dV)-Trainer* (by Mimic Technologies), the *RobotiX Mentor* (by 3D Systems, formerly Symbionix), and the *Robotic Surgery Simulator (RoSS)* (by Simulated Surgical Systems) (see Figure 1.8). These VR training simulator use virtual models for dry lab simulation (see Figure 1.9), wet lab simulation, and clinical simulation [RMT+14; SBKN17].

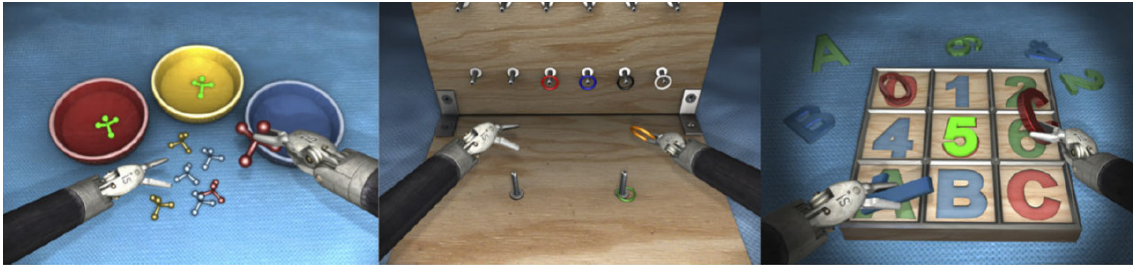
Current literature confirms the validity of the *RobotiX Mentor* (by 3D Systems, formerly Symbionix) [APHP20; EWA+21; WAR+16], the *dVSS* (by Intuitive Surgical Inc.), and the *dV-Trainer* (by Mimic Technologies) [HGVB18; LMK+12; TGS+16] as valuable tools for training and assessment of trainee skills in robotic surgery. However, current literature could not confirm similar validity for the *RoSS* (by Simulated Surgical Systems) [TGS+16].



**Figure 1.8:** Common VR simulators. Top left: *dVSS*. Top right: *dV-Trainer*. Bottom left: *RobotiX Mentor*. Bottom right: *RoSS*. Adapted from Intuitive Surgical Inc., Mimic Simulation, 3D Systems formerly Symbionix, and Simulated Surgical Systems [Simb; Simc; Sur; Sys].

### 1.2.5 Training Curricula

For training robotic surgery, there exist different standardized training curricula of fundamental skills for novice surgeons. The most know curriculum for training robotic surgery is called Fundamentals of Robotic Surgery (FRS).

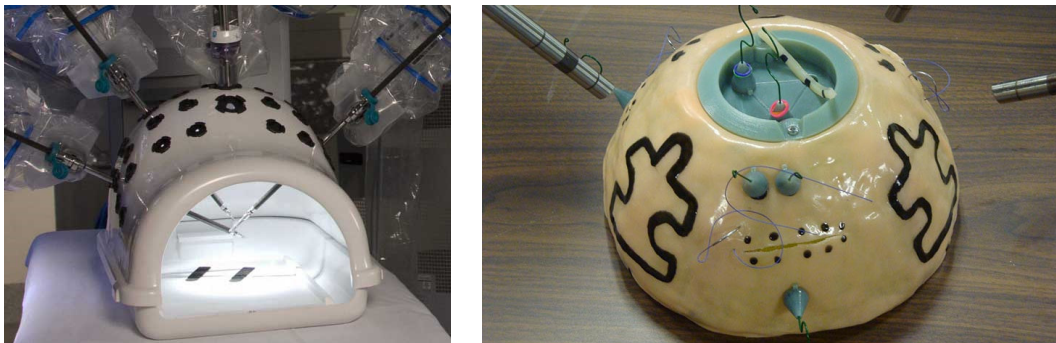


**Figure 1.9:** VR simulator dry lab tasks. Reprinted from Mimic Simulation [Simd].

**FRS** *FRS* is a standardized training and assessment curriculum of fundamental skills to safely and efficiently perform robotic surgery. It is used to train novice surgeons via didactic lectures, psycho-motor skills labs, and team training activities. It is further used to document basic knowledge, fundamental skills, and competency of surgeons in robotic surgery [SPS14].

This curriculum was developed in two years in a series of consensus conferences by various international robotic surgery experts, behavioral psychologists, medical educators, and statisticians. It was further supported and reviewed by representatives of the leading surgical societies involved in robotic surgery [SPS14].

For the training and assessment of the *FRS* curriculum, a physical box trainer and dome model with the *FRS* tasks was designed (see Figure 1.10) [Rob]. These *FRS* tasks are often used for physical dry lab simulation [SBKN17], and are also widely adopted by VR simulators such as by the *RobotiX Mentor* (by 3D Systems, formerly Symbionix) (see Table 1.6) [Sima].

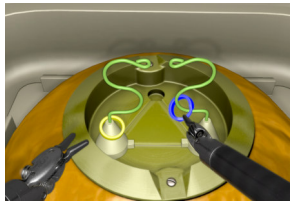
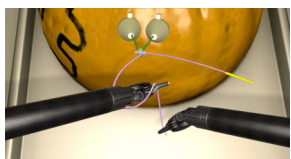
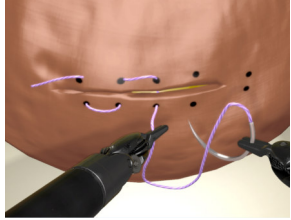
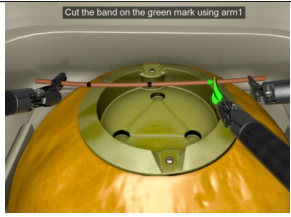
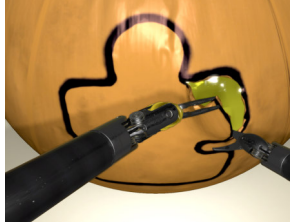
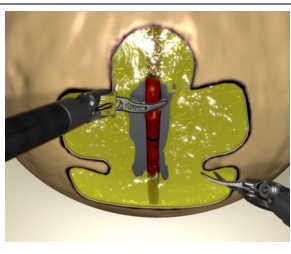


**Figure 1.10:** Physical simulator *FRS* tasks with box trainer. Left: Box trainer. Right: Dome model with *FRS* tasks. Reprinted from Fundamentals of Robotic Surgery [Rob].

### 1.2.6 Training Approaches and Processes

Currently, there exists no validated and uniformly accepted training curriculum across training programs for robotic surgery. Hence, there are no standardized training processes for novice surgeons before applying robotic surgery clinically to patients [SSL+20]. As a consequence, hospitals and their departments are required to create their own training processes. These training



	<p><b>Task 1: Ring Tower Transfer</b></p> <ul style="list-style-type: none"> <li>• Remove ring from top right middle tower and place on lower left side tower</li> <li>• Remove ring from top left middle tower and place on lower right side tower</li> </ul>
	<p><b>Task 2: Knot Tying</b></p> <ul style="list-style-type: none"> <li>• Tie surgeon's knot to approximate two eyelets until touching each other</li> <li>• Back up knot with two more throws</li> </ul>
	<p><b>Task 3: Railroad Track</b></p> <ul style="list-style-type: none"> <li>• Perform horizontal mattress suturing through target points to approximate tissue</li> <li>• Anchor needle by passing through final two target points twice</li> </ul>
	<p><b>Task 4: 4th Arm Cutting</b></p> <ul style="list-style-type: none"> <li>• Switch control between instruments to use mono polar scissors</li> <li>• Cut vein transversely at hash marks</li> </ul>
	<p><b>Task 5: Puzzle Piece Dissection</b></p> <ul style="list-style-type: none"> <li>• Cut puzzle piece pattern between lines</li> <li>• Avoid incising underlying tissue or cutting outside of lines</li> </ul>
	<p><b>Task 6: Vessel Energy Dissection</b></p> <ul style="list-style-type: none"> <li>• Dissect through fat layer to expose vessel</li> <li>• Coagulate vessel at two points</li> <li>• Cut vessel between two coagulated points</li> </ul>

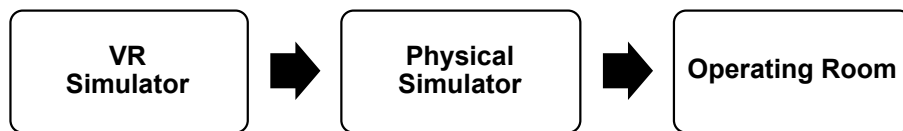
**Table 1.6:** VR simulator FRS tasks with standardized procedures. Adapted from 3D Systems formerly Symbionix [SDK+18; Sima].

processes commonly follow a progression of observing, assisting, performing under supervision, and finally, independent practice [SBKN17]. Hence, training processes typically share a common structure which can be described as follows [RMT+14; SBKN17; ZLH+19]:



1. **Online Learning:** In online learning, novice surgeons typically use videos to discuss procedures and methods to avoid complications and to further receive hands-on team-based training on patient positioning, port positioning, and robotic docking.
2. **VR Simulator Training:** In VR simulators, novice surgeons typically train basic and advanced tasks in virtual environments. Thereby basic tasks commonly consist of simple routines such as cutting, suturing and grasping, whereas advanced tasks commonly consist of more sophisticated routines on vascular and bowel models.
3. **Physical Simulator Dry Lab Training:** In dry lab physical simulation, novice surgeons train basic and advanced tasks with the physical robotic surgical system on inanimate models. Here, novice surgeons learn initial console skills such as camera and clutch control. Novice surgeons further learn basic motor skills such as bi-manual handle movement to control the robotic instruments for object transfer, cutting, and suturing.
4. **Physical Simulator Wet Lab Training:** In wet lab physical simulation, novice surgeons train advanced tasks with the physical robotic surgical system on animal parts, human body parts, and live animals. Here, novice surgeons learn how to handle tissue and further understand tissue reactions to robotic instrument touch.
5. **Operating Room Training:** In operating room training, novice surgeons train previously attained skills on patients under the supervision of expert surgeons as mentors. First, novice surgeons only watch performances of expert surgeons as *observers*. Then, novice surgeons assist expert surgeons as *bedside assistants*. Finally, novice surgeons train together with expert surgeons as *console surgeons*. Often, the *dual console* (by Intuitive Surgical Inc.) is used for this purpose. Here, novice surgeons perform the simplest part of procedures and progressively take on more challenging procedures. The role of the mentor also progresses, from the proctor who supervises and allows the novice surgeon to practice procedures from time to time to the preceptor who only steps in when required.

A simplified robotic surgery training process is illustrated in (see Figure 1.11) and can be described as follows. First, novice surgeons train simple to advanced procedures in virtual environments with VR simulators without supervision by a mentor. Then, novice surgeons train progressively more challenging procedures in physical environments with physical simulators on inanimate anatomical models, animal parts, and human parts. Finally, novice surgeons train clinical procedures in physical environments on clinical patients using the *dual console* (by Intuitive Surgical Inc.) under the supervision of a mentor.



**Figure 1.11:** Simplified robotic surgery training process.

## 1.3 Aim of the Work

In this work, we develop a multimodal record and replay platform in augmented reality for robotic surgery training using the Intuitive da Vinci Surgical System. With this platform, we aim to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. Moreover, we aim to explore different concepts for motion guidance and surgical skill evaluation in augmented reality. For this purpose, we address the following Research Questions (RQs):

- **RQ1:** Is multimodal record and replay in augmented reality a beneficial approach for training in robotic surgery?
- **RQ2:** How could motion guidance and surgical skill evaluation in augmented reality benefit training in robotic surgery?

### 1.3.1 RQ1 - Multimodal Record and Replay in AR

The first aim of this work is to use the developed platform to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. For this purpose, we introduce the following *areas of interest* to answer the stated *RQ1* for multimodal record and replay in augmented reality:

- **Integrability of AR Platform in Training Processes:** How can the developed AR platform be *integrated* in existing training processes for robotic surgery?
- **Record Functionality:** How do chief surgeons prefer to *record* performances with multiple modalities in AR for robotic surgery training?
- **Replay Functionality:** How do chief surgeons prefer to *replay* performances with multiple modalities in AR for robotic surgery training?

### 1.3.2 RQ2 - Motion Guidance and Skill Evaluation in AR

The second aim of this work is to use the developed platform to explore different concepts for motion guidance and surgical skill evaluation in augmented reality for training in robotic surgery. For this purpose, we introduce the following *areas of interest* to answer the stated *RQ2* for multimodal training with motion guidance and skill evaluation in augmented reality:

- **Train Functionality with Motion Guidance:** How do chief surgeons prefer to *train* performances with multiple modalities and motion guidance in AR for robotic surgery training?
- **Sensors for Surgical Skill Evaluation:** Which sensors can be used for surgical skill evaluation in robotic surgery training?

## 1.4 Outline

This work is organized as follows: In Chapter 1, we introduce robotic surgery and its training as well as key concepts such as augmented reality, multimodal feedback, and motion guidance. In Chapter 2, we present relevant literature for robotic surgery training. In Chapter 3, we describe the development of a robotic surgery training approach in augmented reality with multimodal record and replay. In Chapter 4, we present the conducted study of this work which evaluates the concepts of the developed approach. In Chapter 5, we report the results of this study. In Chapter 6, we discuss the results and also point out possible directions to guide future research. In Chapter 7, we provide a conclusion of this work in robotic surgery training with augmented reality and multimodal record and replay.



## 2 State of the Art

In this chapter, relevant literature for robotic surgery training will be presented.

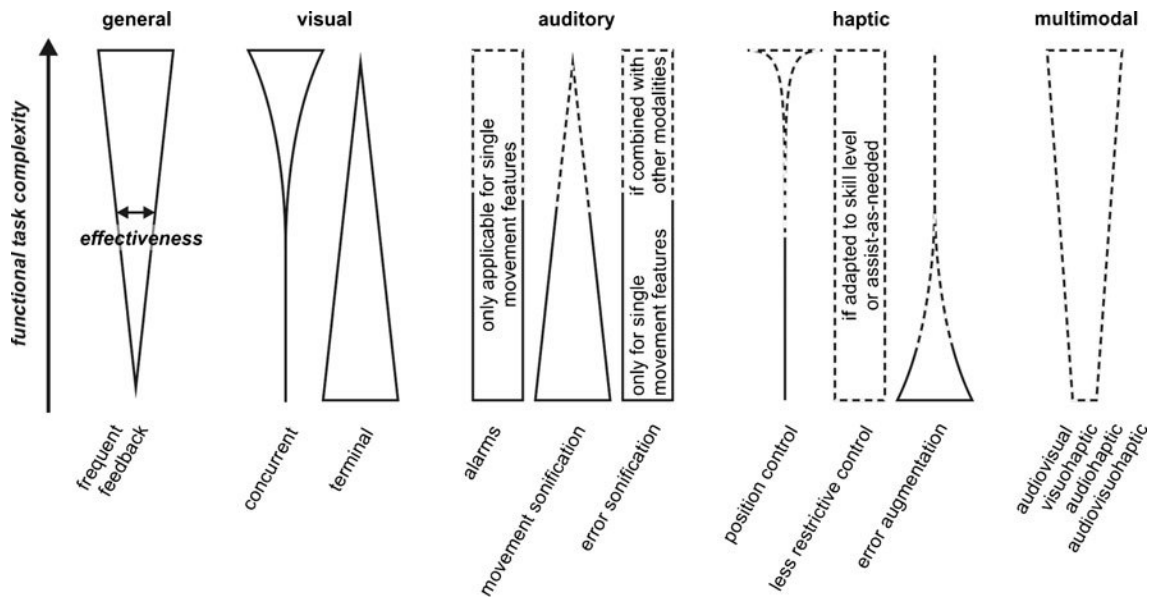
### 2.1 Multimodal Feedback for Robotic Surgery Training

To address the potential of unimodal and multimodal feedback for motor learning, Sigrist et al. [SRRW13] conducted an extensive review. The authors elaborated design criteria for effective visual, auditory, haptic, and multimodal feedback (see Figure 2.1). These design criteria include the following:

- **Functional Task Complexity:** The *functional task complexity* can be defined by the number of sessions required to master a task, the degrees of freedom, and the realism of a task. Following this definition, a task can be classified as simple if it can be mastered in a single session, has only one degree of freedom, and appears artificial.
- **Feedback Strategy:** The *feedback strategy* can be defined by the used *feedback modality* and the *feedback purpose*. Here, the used feedback modalities include general, visual, auditory, haptic, and multimodal feedback. The feedback purpose differs per feedback modality. For general feedback, it includes frequent feedback. For visual feedback, it includes concurrent and terminal feedback. For auditory feedback, it includes alarms, movement sonifications, and error sonifications. For haptic feedback, it includes position control, less restrictive control, and error augmentation.

The authors experimentally found that frequent and visual feedback are effective feedback strategies for complex tasks in motor learning. Further, the authors stated that visual feedback is already widely explored for simple and complex tasks and that auditory feedback is well explored for simple tasks in motor learning.

However, the authors only hypothesized about the effectiveness of auditory, haptic, and multimodal feedback strategies for complex tasks in motor learning. For *haptic feedback*, the authors stated that many concepts are suggested but that studies and systematic evaluations are missing. Here, the authors suggest to evaluate haptic feedback to augment movement errors in complex tasks. For *multimodal feedback*, the authors stated that it is a promising approach and hypothesized that the effectiveness of multimodal feedback increases with raising functional task complexity. Here, the authors hypothesized that *audio-visual feedback* can enhance perception, and that *visuo-haptic feedback* can be effective for spatio-temporal learning. As surgical procedures imply a high functional task complexity, multimodal feedback might represent an effective feedback strategy for robotic surgical training.



**Figure 2.1:** Effectiveness of feedback strategies for motor learning. Solid lines: Experimentally confirmed. Dashed lines: Hypothesized. Broader shape: More effective. Reprinted from Sigrist et al. [SRRW13].

### 2.1.1 Visual Feedback

Visual feedback and visual motion guidance have been traditionally investigated in rehabilitation. In this application field, it has been shown that visual cues help in maintaining movement paths [SK04]. For robotic surgery training, research groups have shown that force-based visual feedback reduces applied forces on sutures [BOK+04] and tissues [RAB+08]. Further, research groups have investigated the effects of simple motion guidance concepts in robotic surgery teaching [JSC+17] and learning [CMP+21; MVL+20].

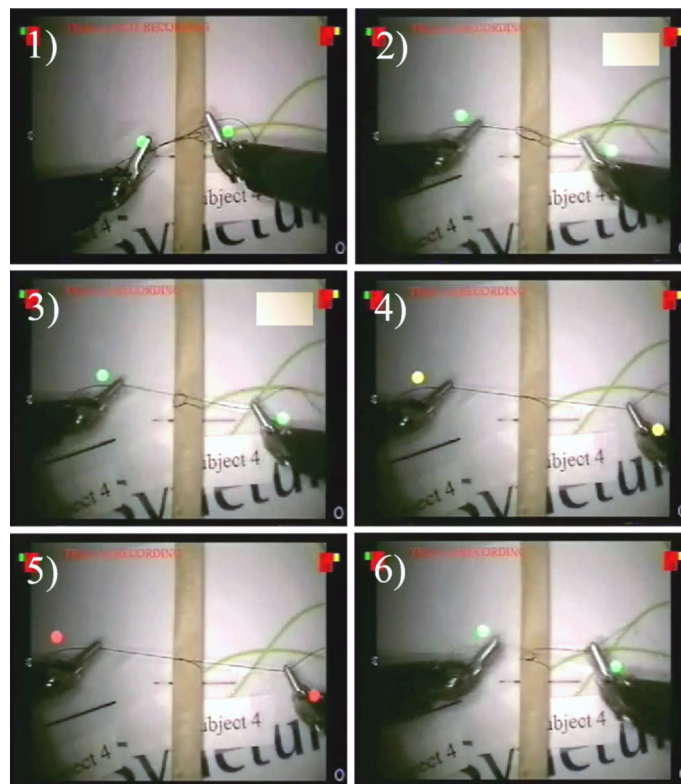
#### Force-based Visual Feedback

Bethea et al. [BOK+04] conducted a study to evaluate the effect of haptic feedback on forces exerted on sutures during knot tying tasks in robot-assisted surgery. For this purpose, the authors used a da Vinci Surgical System, manufactured a tension measuring device for knot tying tasks, and provided force feedback via sensory substitution as a visual color bar scale. The authors found significantly less forces applied to sutures when providing haptic feedback via sensory substitution compared to not providing haptic feedback. The use of haptic feedback also resulted in more consistently applied forces to sutures and in significantly reduced breakage of fine suture materials.

Akinbiyi et al. [ARS+06] proposed an augmented reality system to present force information through sensory substitution. The proposed system consisted of force-sensing robotic instruments, a kinematic tool tracker, and a graphic display that overlays a visual representation of the exerted forces on the instrument tips. The color of the overlaid circles provided information about the applied forces (see Figure 2.2). In the study, the authors integrated the system with the da Vinci

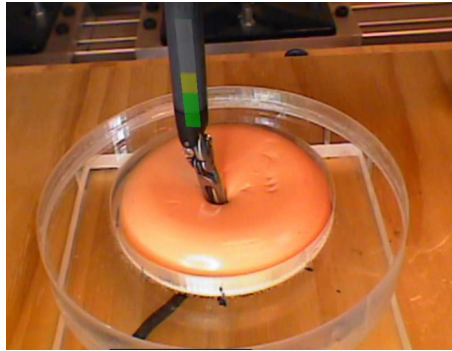
Surgical System and evaluated its concepts on a phantom knot-tying task. The authors found that the visual force-feedback decreased the number of broken sutures as well as the number of loose knots and hence leads to a more consistent application of forces.

Similar effects were found by the study of Reiley et al. [RAB+08], which also investigated the effect of visual force feedback on surgeon performance. Here, the authors also used the graphical overlay to display applied forces on the da Vinci stereo viewer. The authors found reduced suture breakage, decreased applied forces, and more consistently applied forces for novice robotic surgeons. However, for experienced robotic surgeons, the authors did not find significant differences in performance. As a consequence, the authors hypothesized that visual force feedback primarily benefits novice surgeons.



**Figure 2.2:** Visual representation of forces on robotic instrument tips. Reprinted from Akinbiyi et al. [ARS+06].

Later, Gwilliam et al. [GMV+09] investigated the effect of direct haptic feedback and graphical force feedback on surgeon performance in robotic surgery. Their approach consisted of a position-exchange controller with dynamics compensation for direct haptic feedback and a force estimator displayed via a tooltip tracking bar graph for graphical force feedback in augmented reality (see Figure 2.3). In a study, the authors compared direct haptic feedback only, graphical force feedback only, combined direct haptic and graphical force feedback, and no feedback. The authors found that direct haptic feedback minimizes applied forces to the tissue and that combined haptic and graphical force feedback reduces subject task error.



**Figure 2.3:** Graphical force feedback on robotic instruments. Reprinted from Gwilliam et al. [GMV+09].

### Visual Motion Guidance

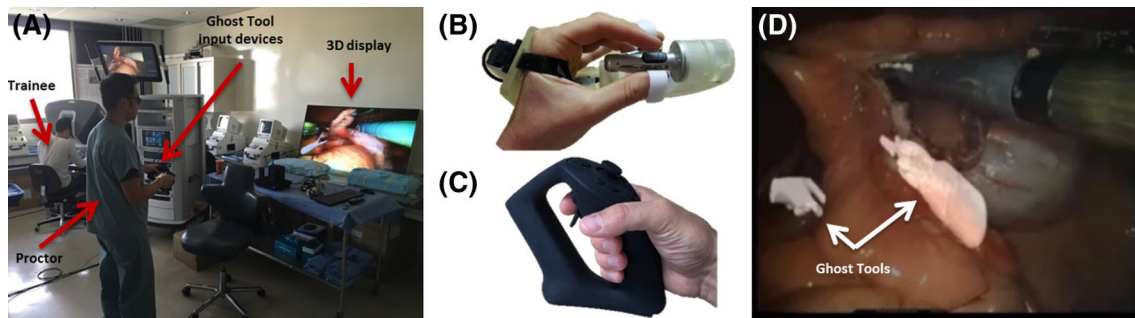
Motion guidance approaches in mixed reality have been widely explored for application areas such as physical exercises, martial arts, physiotherapy, and rehabilitation, as well as for repair and maintenance tasks. Here, mixed reality approaches have demonstrated their benefits of changing viewpoints and displaying instructions in the 3D space [YAM+20].

For general visual motion guidance, Yu et al. [YAM+20] conducted three user studies to investigate the influence of perspective, visual encoding, and motion characteristics on the design of mixed reality motion guidance systems to guide human body motions. For this purpose, the authors designed two virtual reality motion guidance systems. The authors found that the first-person perspective for motion guidance leads to improved performance in terms of motion accuracy and time for all visible motions. This finding can be explained through the usage of the same coordinate systems for user body and guidance paths, which avoids additional cognitive load for mapping motions between different coordinate systems.

For visual motion guidance in robotic surgery training, Jarc et al. [JSC+17] conducted a user study to investigate if semi-transparent ghost tools of an expert surgeon overlaid on the novice surgeon's stereo viewer are beneficial for mentoring (see Figure 2.4). For this purpose, the authors used wireless input devices for the expert surgeon to move 3D ghost tools to augment the novice surgeon's view. Using objective measures of expert surgeon performance based on hand movements, button presses, and questionnaires, the authors found that expert surgeons exploited the augmented capabilities of provided 3D ghost tools during training scenarios. The authors further found that expert surgeons and novice surgeons evaluated the ghost tools as effective for mentoring, whereas novice surgeons found the ghost tools more helpful than expert surgeons. Hence, the authors stated that advanced mentoring technologies such as 3D ghost tools should be further explored to improve training in robotic surgery.

Later, Caccianiga et al. [CMP+21] conducted a study to combine visual and haptic feedback with visual motion guidance. For this purpose, the authors developed a virtual reality task resembling needle-driving using the surgeon console of the da Vinci Research Kit. The authors found significant improvements in task completion capabilities, accuracy, and error reduction with visual guidance.





**Figure 2.4:** Ghost tools of expert surgeon during mentoring a robotic surgery training. a) Experimental setup. b) Custom wireless input devices. c) Commercial input devices. d) Ghost tools in the surgeon’s view. Reprinted from Jarc et al. [JSC+17].

Common VR simulator such as the *RobotiX Mentor* (by 3D Systems, formerly Symbionix) and the *dVSS* (by Intuitive Surgical Inc.) mostly integrate simple forms of visual motion guidance for training novice surgeons (see Figure 2.5 and Figure 2.6).

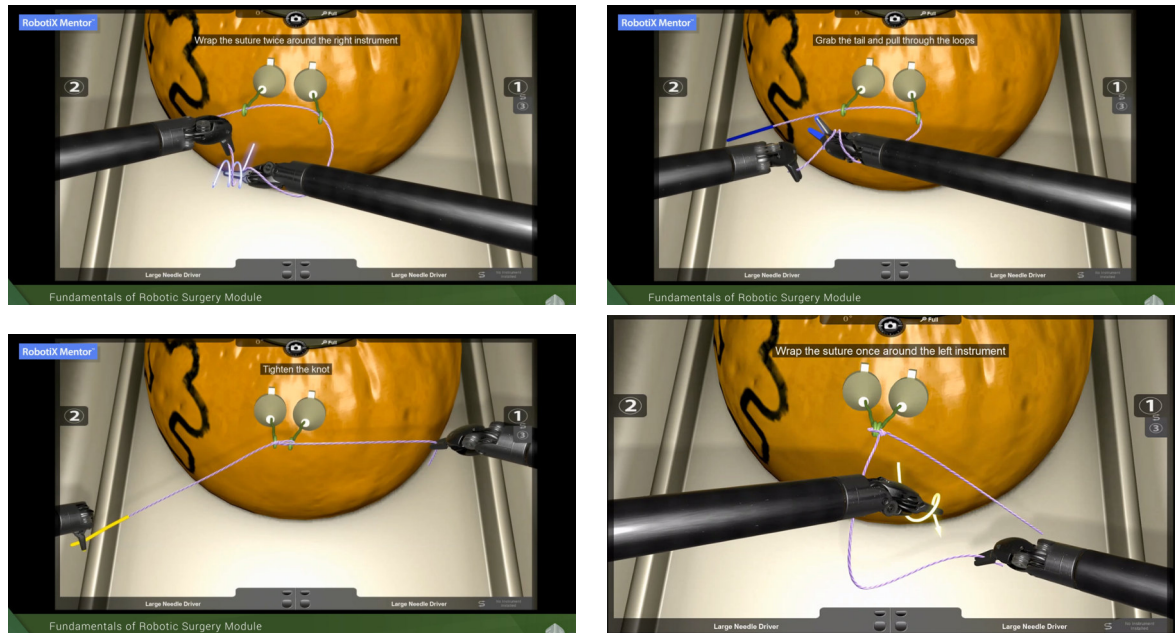
To advance motion guidance concepts in VR simulators for effective teaching and coaching, Malpani et al. [MVL+20] conducted a study to investigate the effect of real-time teaching cues and coaching modes on surgical skill acquisition. Further, the authors investigated additional metrics to assess surgical technique skills besides metrics such as time and motion efficiency. For this purpose, the authors developed an automated VR teaching framework for robotic surgery with six teaching cues for a needle passing task using a non-commercial research version of the *dVSS* (by Intuitive Surgical Inc.). The teaching cues were designed as graphical overlays to demonstrate the ideal surgical technique, such as trajectories to follow. In the study, the authors found that automated teaching cues significantly improve surgical technique metrics compared to self-learning in VR simulator environments for needle passing tasks.

### 2.1.2 Auditory Feedback

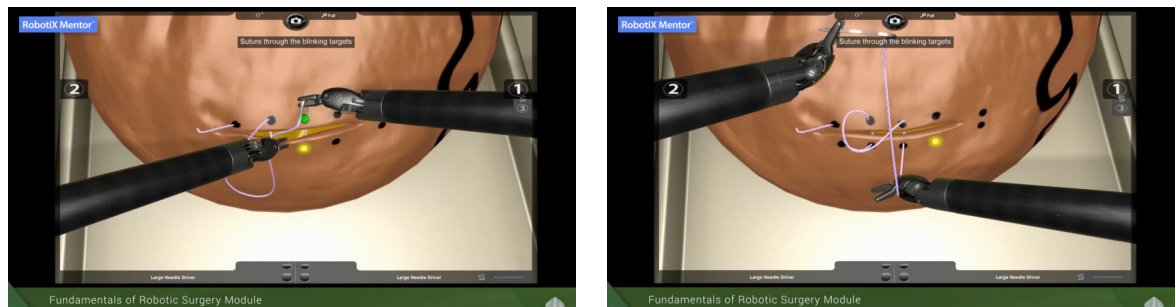
Auditory feedback has been mainly investigated in sports and rehabilitation. Here, it has been shown that auditory feedback is beneficial for cyclic movements as the human auditory system is sensitive for rhythms [SJMT19]. For robotic surgery training, auditory feedback has been used to convey robotic instrument vibrations to surgeons and non-surgeons in order to improve awareness of tool contacts [KK15; MGS+11]. Further, auditory feedback has been used to impose temporal constraints to task executions without overloading the visual domain [CMD+20].

### 2.1.3 Haptic Feedback

Haptic feedback and haptic motion guidance have been recently investigated in combination with visual feedback to broaden the usefulness of simulators for robotic surgery training. Here, studies showed that haptic feedback improves training effects, fidelity, and realism of simulators, especially for novice surgeons [RDP20]. For providing haptic feedback in robotic surgery training,



**Figure 2.5:** VR simulator motion guidance for knot tying task. Reprinted from 3D Systems formerly Symbionix [Sima].



**Figure 2.6:** VR simulator motion guidance for suturing task. Reprinted from 3D Systems formerly Symbionix [Sima].

research groups mainly used haptic feedback devices such as voice coil actuators [MGS+11], tactile actuators [SFO+12], squeezing devices [BFCK17], pneumatic balloon-based systems [WGN+16], or combinations of those [APT+19].

For training in MIS, Pinzon et al. [PBZ16] conducted a review to investigate the role of haptic feedback for skill acquisition and training, as well as to discuss haptic simulation in relation to surgical performance. The authors found that force feedback is the most beneficial method for tissue identification in MIS. Further, the authors found that novice surgeons in the early stages of surgery training benefit most from haptic feedback. In the end, the authors emphasized the need to understand how haptic feedback can be used in surgical education to increase training efficiency, performance, and safety.

For training in robotic surgery, Rangarajan et al. [RDP20] conducted a systematic literature review to investigate the role and efficiency of haptic feedback in VR simulators for robotic surgical education. Overall, the authors found that haptic feedback improves VR simulators in terms of training effects, fidelity, and realism. In particular, the authors found that VR simulators with haptic feedback were significantly more effective for skill training compared to VR simulators without haptic feedback. Here, the authors found improved learning curves, reduced time to proficiency, and faster task completion time. These results were particularly found for novice surgeons in the early stages of training. However, the authors also stated that haptic simulators could be too expensive at the present time and that further research and cost-benefit analyses are needed to determine whether haptic simulators are a surgical necessity.

As already described for visual motion guidance, Caccianiga et al. [CMP+21] conducted a study to combine visual and haptic feedback with visual motion guidance. Here, the authors found a significant improvement in error reduction with combined visual and haptic feedback.

### **Vibration-based Haptic Feedback**

McMahan et al. [MGS+11] investigated preferences and effects of vibration-based auditory and haptic feedback on surgical task performance. For this purpose, the authors used a previously developed sensing and actuating device from Kuchenbecker et al. [KGM+10] called *VerroTouch* for the da Vinci Surgical System. With this prototype, the authors provided combinations of auditory or haptic feedback based on robotic instrument vibrations by mounting customized acceleration sensors on the robotic instruments and custom voice coil actuators on the surgeon handles. The authors conducted a user study with 11 surgeons to compare visual-only feedback, visual with auditory feedback, visual with haptic feedback, and visual with auditory and haptic feedback. With preference ratings and survey responses, the authors found that surgeons preferred the inclusion of robotic instrument vibrations in general, which could be explained through an improved concentration and situational awareness by strengthening the connection between surgeons and robotic instruments. However, the authors did not find a preference for which sensory modality best represented robotic instrument vibrations. Also, the authors did not find a significant difference in task performance.

Similar effects were found by the study of Koehn et al. [KK15], which investigated the preference of surgeons and non-surgeons for combinations of vibration-based auditory and haptic feedback. Here, the authors also used *VerroTouch* for the da Vinci Surgical System. The authors found that almost all subjects preferred receiving feedback based on tool vibrations in general. However, the authors did not find a preference between haptic and audio feedback together and haptic feedback only. In addition, the authors found that vibration-based feedback improves the awareness of tool contacts and provides valuable tactile information for surgeons and non-surgeons.

### **Force-based Haptic Feedback**

Wagner et al. [WHS02] conducted a study to evaluate the effect of contact force feedback on tissue damage for robot-assisted blunt dissection tasks. For this purpose, the authors used a stylus-based robot, built a force sensor into the instrument shaft, and provided force feedback as if the stylus was attached directly to the proximal end of the instrument shaft. The authors found significantly

less average force magnitudes applied to tissues when providing force feedback compared to not providing force feedback. The authors also found significantly reduced peak force magnitudes applied to tissues and significantly less tissue damage when providing force feedback.

Later, Wottawa et al. [WGN+16] conducted a study to evaluate the effect of grasping force feedback on tissue damage for robot-assisted surgery. For this purpose, the authors used a da Vinci Surgical System, mounted grasping force sensors on the grippers of the da Vinci robotic instruments, and balloon actuators on the da Vinci surgeon handles. The authors found that haptic feedback significantly decreased grasping forces for novice and expert surgeons. The authors also found that grasping forces significantly correlated to tissue damage, resulting in significantly reduced tissue damage.

To investigate the effect of wrist-squeezing haptic feedback on exerted forces in robotic surgery training, Brown et al. [BFCK17] conducted a study with novice surgeons using inanimate training tasks. For this purpose, the authors developed a wrist-squeezing haptic feedback system and used it with the da Vinci Surgical System. The authors found that haptic feedback of exerted forces significantly reduced exerted forces on task materials. Further, the authors found that exerted forces were still significantly reduced even after removing the haptic feedback system. Hence, the authors suggested that wrist-squeezing force feedback has the potential to minimize exerted forces for novice surgeons in robotic surgery training.

### Haptic Motion Guidance

Schonauer et al. [SFO+12] investigated the design space for motion guidance. For this purpose, the authors developed a motion guidance system for mixed reality and telepresence applications (see Figure 2.7). Using full-body tracking data acquired by motion capture, the system can provide real-time visual, vibro-tactile, and pneumatic feedback in 3D to guide movements towards desired motions, including position, direction, and velocities.



**Figure 2.7:** Multimodal motion guidance system. a) Haptic feedback to guide user motions for following remote teachers. b) Motion capture. c) Vibration actuators. d) Concept of haptic feedback for motion guidance regarding direction and speed. Reprinted from Schonauer et al. [SFO+12].

## 3 System Development

In this chapter, the development of a robotic surgery training approach in augmented reality with multimodal record and replay will be presented.

### 3.1 Existing Systems

This work is based on already existing systems, which will be briefly presented in the following.

#### 3.1.1 Hardware Setup

The hardware setup of this work is based on a previously conducted Master's thesis [For18] and abstract [FK19]. Thereby, it consists of the following components:

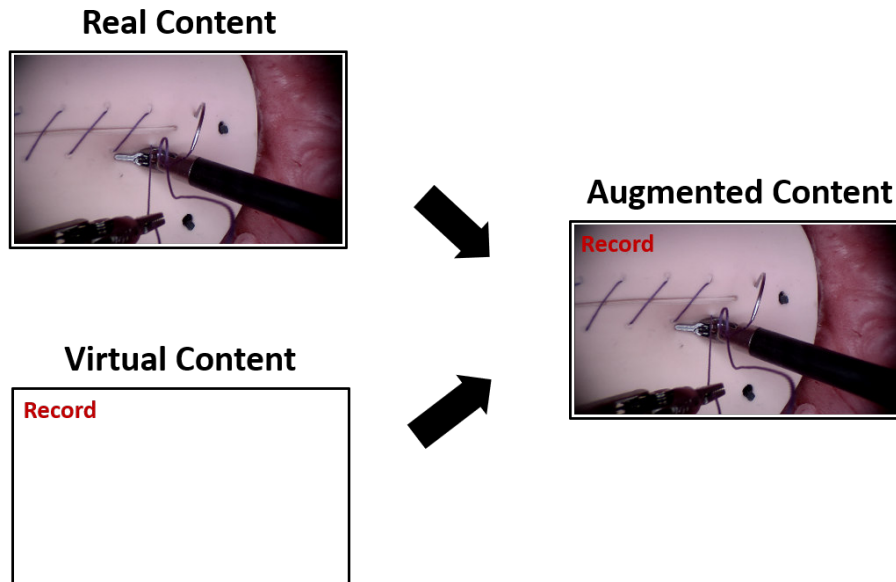
- **Clinical da Vinci Si HD Surgical System:** The da Vinci Si HD surgical system introduced in Section 1.1.2 is the surgical robot used for training of robotic surgery. It is a clinical version of the surgical system and consists of a surgeon console, a vision cart, and a patient cart.
- **Workstation Computer:** The workstation computer is the processing point of the system. It is an *Alienware Aurora R7 Desktop* [Del].
- **DeckLink Quad 2:** The DeckLink Quad 2 video capture and playback card (by Blackmagic Design) [Bla] is used to achieve visual augmented reality. It is built into the workstation computer and is connected to the da Vinci vision cart via input and output cables.
- **Laptop Computer:** The laptop computer is the recording point of the system.
- **Box Trainer:** The box trainer is the created environment for training procedures in robotic surgery. It was created to ensure realistic training environments with consistent lighting conditions. It consists of a globe with an LED ring, a foam ring, and an embedded task plate inside with a piece of simulated tissue. The box trainer itself is attached to a tilting table.

#### 3.1.2 Software Basis

The software of this work is also based on the previously conducted Master's thesis [For18] and abstract [FK19]. It mainly consists of a keying functionality to enable visual augmented reality, which will be described in more detail in the following.

### Visual Augmented Reality

Visual augmented reality was accomplished by first accessing the camera stream for real world content, followed by overlaying it with a transparent frame with virtual content. Through the composition of real content and virtual content, augmented content is created (see Figure 3.1).



**Figure 3.1:** Concept of visual AR. Adapted from Forte et al. [FK19].

The technique used to superimpose virtual content on real content is called *keying*. This technique was implemented by integrating two SDI Splitters and a DeckLink Quad 2 video capture and playback card into the standard video pipeline between the left and right da Vinci camera control units and the left and right da Vinci cores. Through this technique, it is possible to first receive images from the da Vinci in ROS, to then read or key the images via the DeckLink Quad 2, and to finally send the augmented images from ROS to the da Vinci. Thereby keying is achieved by the following steps:

1. **da Vinci Camera Control Units:** The left and right da Vinci camera control units send images from the da Vinci stereo endoscope to the left and right SDI splitters.
2. **SDI Splitters:** The left and right SDI splitters split the received images to enable reading and keying for the DeckLink Quad 2 video capture and playback card.
3. **DeckLink Quad 2:** The DeckLink Quad 2 video capture and playback card performs keying and sends the resulting augmented images to the left and right da Vinci Cores.
4. **da Vinci Cores:** The left and right da Vinci cores display the received augmented images on the left and right display of the da Vinci surgeon viewer.

## Voice Control

The voice controls were developed to improve the usability of AR functionalities. Thereby the following decisions have been made:

- **Lexical Language Model:** Lexical language models only recognize words from their known dictionary and always take the closest matching words. Through this, voice commands with only a small set of words are very robust against conversations within the operating room, assuring that functions are not activated by accident (false positive).
- **Three-Word Commands:** Three-word commands are long enough to reduce the probability of false-positive function activation and yet short enough to be usable for surgeons to remember and use during a surgical procedure.
- **Activation Word:** Activation words are also useful to reduce the probability of false-positive function activation. The activation word should be selected from a set of uncommon words that are not likely to be used in conversations within operating rooms. The selected word is then used as the first word of the three-word command sequences.

These concepts were implemented by using the following technologies:

- **PyAudio:** *PyAudio* [Cam] is a library providing Python bindings for PortAudio, a cross-platform audio I/O library. This library allows the use of Python to record, process, and play audio signals with a Linux platform.
- **Sphinx Decoder:** *Sphinx Decoder* [CMUa] is a continuous speech recognition engine, which works speaker-independently and can be used for large vocabularies. For the decoder, a lexical and language model is required.
- **Sphinx Knowledge Base Tool:** *Sphinx Knowledge Base Tool* [CMUb] is a web-based tool to compile text-based components into the required lexical and language models for the decoder.

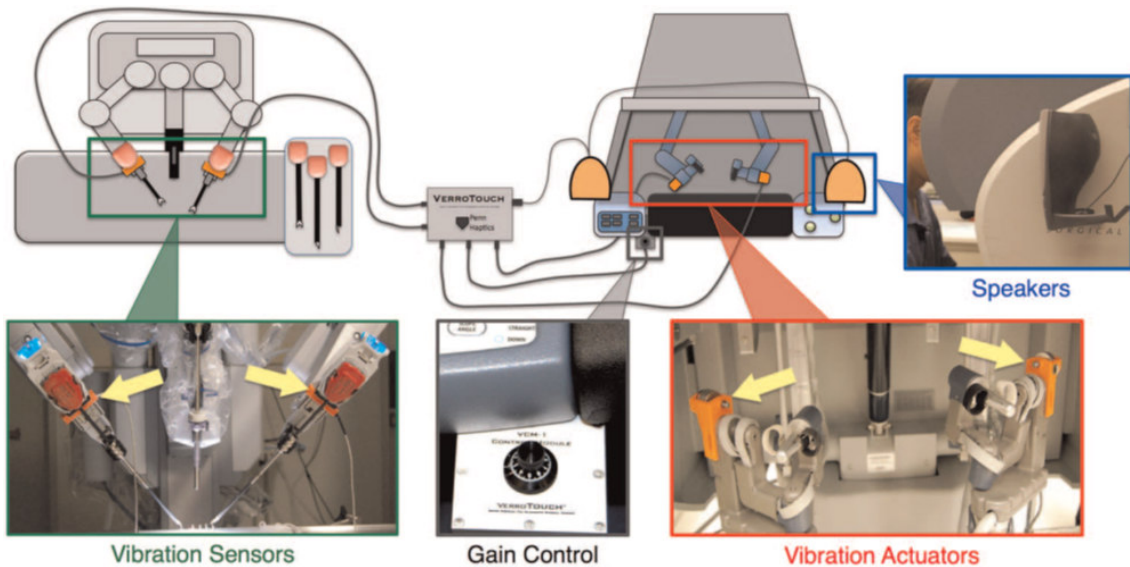
### 3.1.3 VerroTouch for Haptic Feedback

The haptic feedback of this work is based on a previously developed prototype called *VerroTouch* [KGM+10]. This prototype allows surgeons to feel the vibrations of the robotic instruments at the handles of the surgeon console. It can further allow surgeons to hear the vibrations of the robotic instruments at the speakers of the surgeon console [MGS+11]. In addition, it can also be extended to record the vibrations of the robotic instruments [MGC+13]. To achieve these functionalities, *VerroTouch* consists of acceleration sensors, a main receiving unit, voice coil actuators, and optional speakers or DVD recorders (see Figure 3.2). The signal flow of these system components can be seen in Figure 3.3, which illustrates the measurement of robotic instrument vibrations and their respective transformation to haptic and auditory representations. The signal flow for recording the vibrations of the robotic instruments can be seen in Figure 3.4. These system components, as well as their signal flows, can be described as follows:

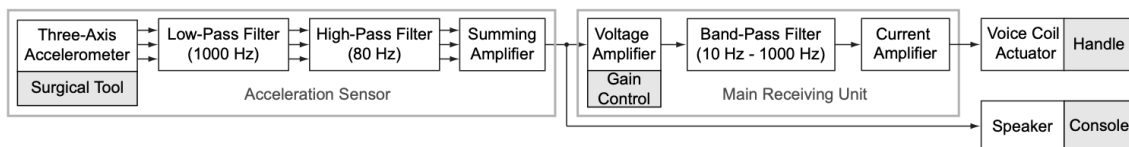
- **Acceleration Sensors:** Three-axis accelerometers are attached to the left and right robotic instruments to sense their vibrations. These vibrations are sensed as three separate signals, filtered by a low-pass filter and a high-pass filter, and added up to one signal by a summing amplifier.



- **Main Receiving Unit:** The main receiving unit is connected to the acceleration sensors at the robotic instruments, the voice coil actuators at the surgeon handles, and the speakers at the surgeon console. It first amplifies the voltage of the summed acceleration sensor signal, which amplification can be controlled by a gain control. Then, it filters the amplified signal by a band-pass filter and finally converts the voltage of the signal into a current.
- **Voice Coil Actuators:** Voice coil actuators are attached to the left and right handles of the surgeon console. This allows surgeons to feel the vibrations of the left and right robotic instruments.
- **Speakers:** Speakers can be connected directly after the acceleration sensor, bypassing the main receiving unit. This allows surgeons to hear the vibrations of the left and right robotic instruments.
- **DVD Recorder:** A DVD recorder can be connected directly after the acceleration sensor, bypassing the main receiving unit as well. This allows surgeons to record vibrations of the left and right robotic instruments.

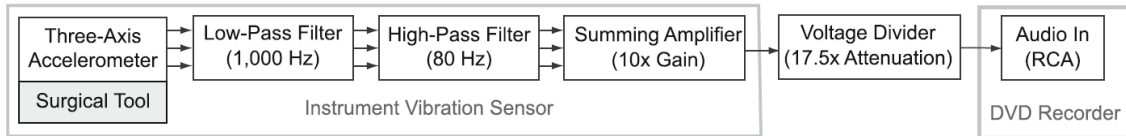


**Figure 3.2:** VerroTouch system components. Reprinted from McMahan et al. [MGS+11].



**Figure 3.3:** VerroTouch signal flow of one channel for transforming surgical tool vibrations to haptic and auditory representations. Note that robotic instruments are called surgical tools in the diagram. Reprinted from McMahan et al. [MGS+11].





**Figure 3.4:** VerroTouch signal flow of one channel for recording surgical tool vibrations. Note that robotic instruments are called surgical tools in the diagram. Reprinted from McMahan et al. [MGC+13].

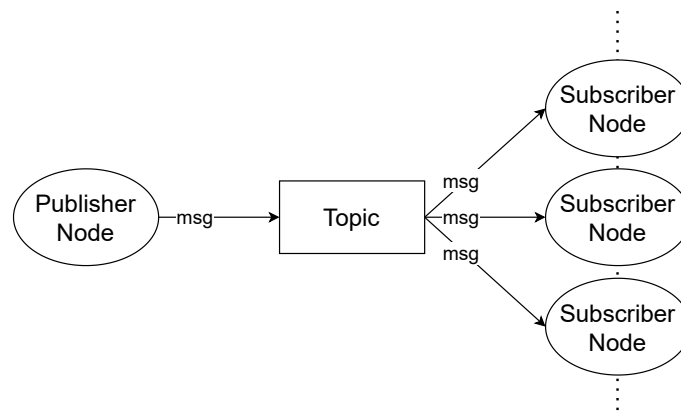
## 3.2 Frameworks and Libraries

For developing an AR-based training simulator for robotic surgery, existing frameworks and libraries can be used. The most known frameworks and libraries for this purpose will be briefly introduced in the following.

### 3.2.1 ROS

Robot Operating System (ROS) is a common framework used for developing robot applications. Within this framework, packages can be shared and reused. The framework and its packages are mainly based on the *programming languages* C++ and Python [ROSa; ROSd].

**Architectural Patterns** ROS is based on the architectural pattern *publish-and-subscribe* between *nodes* via *topics* and *messages* (see Figure 3.5). Through this pattern, a node only knows the topics it publishes and the topics it subscribes to. Hence, a node does not know any other node or if other nodes subscribe to its published topics. This characteristic results in *loose coupling* between nodes, which increases reuse-ability and maintain-ability for nodes and packages [ROSc].



**Figure 3.5:** ROS Computation Graph. A message is denoted as msg. Adapted from ROS [ROSc].

### 3.2.2 OpenCV

OpenCV is a common library used for developing Computer Vision (CV) applications. The library and its functionalities are mainly based on the programming languages C++ and Python [Opea].

**Computer Vision Functionalities** OpenCV provides various functionalities for computer vision applications. Thereby computer vision is enabling computer systems to perceive the world through cameras. By doing so, images are perceived as two-dimensional matrices with one-to-many channels, and values for each pixel ranging between zero-to-255 [Opeb].

These images can be represented in different color schemes. The most common color schemes include the following [Opeb]:

- **RGB:** Three image channels representing red, green and blue.
- **RGBA:** Four image channels representing red, green, blue and alpha channel for transparency.
- **HSV:** Three image channels represent hue, saturation and value.
- **Grayscale:** One image channel representing grayscale.

For *image processing*, the provided functionalities include the following [Opeb]:

- **Blurring:** Blurring is the smoothing of images by averaging neighboring pixel values.
- **Masking:** Masks are contours in binary images used to keep certain areas of the image at their original value while setting other areas of the image to zero.
- **Erosion:** Erosion is shrinking the contour of a binary image by a kernel.
- **Dilation:** Dilation is expanding the contour of a binary image by a kernel.
- **Opening:** Opening is the sequential execution of erosion followed by dilation to erase external structures of the foreground object, such as noise in the binary image.
- **Closing:** Closing is the sequential execution of dilation followed by erosion to erase internal structures of the foreground object, such as little holes in the binary image.

For *object detection*, these include the following [Opeb]:

- **Thresholding:** Thresholding is the transformation of an image into a binary image. For this purpose, a fixed value is used as a threshold to decide for each pixel if its value is set to zero or one. If the pixel value is above the threshold, it is commonly set to one, whereas if the pixel value is below the threshold, it is commonly set to zero.
- **Color Segmentation:** Color segmentation is the transformation of an image into a binary image using a color value-based threshold.
- **Principal Angle Analysis:** Principal angle analysis is a procedure to extract important features from images. Common features to extract are principal angles and mass centers of binary images representing foreground objects.

### 3.3 Software Development Process

The software development process of this work follows a goal-oriented iterative waterfall model. First, we define the goals of this work. Second, we elaborate on the requirements of the system, including use-cases, functional and non-functional requirements. Third, we design an architecture to fulfill the requirements, including package, node-topic, and sequence diagrams. Fourth, we describe the development of the system, including its AR functionalities and haptic feedback integration. These steps will be discussed in more detail in the following.

#### 3.3.1 Goals

The first goal of this work is to develop a multimodal record and replay platform in augmented reality for robotic surgery training using the da Vinci Surgical System. The second goal is to use this platform to answer the RQs stated in Section 1.3. To accomplish these goals, the following concepts which were previously introduced in Section 1.1 and Section 1.2 are used:

- **Robotic Surgery Training**
- **AR Simulator with Record and Replay**
- **Multimodal Feedback**
- **Motion Guidance**

#### 3.3.2 Requirements

The requirements of the system evolve in subsequent steps. First, a stakeholder analysis is conducted. Next, possible use-case scenarios between the stakeholders and the system are elaborated. Finally, functional as well as non-functional requirements are derived.

##### Stakeholder

Stakeholders are all persons who have an interest in the system due to different reasons. Here, we focus on the users of the system, which include the following:

- **Novice Surgeon:** Surgeons with limited or no prior experience in performing robotic surgery.
- **Expert Surgeon:** Surgeons with at least two years of experience in performing and teaching robotic surgery.

#### Use-Case Scenarios

Use-case scenarios are possible interactions between the user and the system. Here, the system should enable expert surgeons to record surgical procedures with multiple modalities and novice surgeons to replay and train with the multimodal recordings, including visual, haptic, and auditory feedback.

As such, one use-case scenario of the system can be described as follows:

1. **Expert Surgeon:** Record a surgical procedure with multiple modalities.
2. **Novice Surgeon:** Replay the multimodal recording of the expert surgeon.
3. **Novice Surgeon:** Train with the multimodal recording of the expert surgeon.

**Use-Case Diagram** Use-case diagrams represent all possible interactions between the user and the system. The use-case diagram of this system is illustrated in Figure 3.6.

#### Functional Requirements

Functional requirements are all functionalities which the system should provide for the user. Here, these functional requirements consist of the following:

- **Use Voice Commands:** The system should enable novice and expert surgeons to use voice commands.
- **Record Own:** The system should record the performance of novice surgeons when training expert performances.
- **Record Expert:** The system should enable expert surgeons to record their performances.
- **Replay Own:** The system should enable novice surgeons to replay their own performances.
- **Replay Expert:** The system should enable novice surgeons to replay expert performances.
- **Replay Both:** The system should enable novice surgeons to replay their own and expert performances at the same time.
- **Train Expert Uninterrupted:** The system should enable novice surgeons to train expert performances in an uninterrupted training mode.
- **Train Expert Interrupted:** The system should enable novice surgeons to train expert performances in an interrupted training mode.
- **Train Expert with Ghost Tools:** The system should enable novice surgeons to train expert performances with ghost tools as motion guidance.
- **Train Expert with Trajectories:** The system should enable novice surgeons to train expert performances with trajectories as motion guidance.
- **Get Performance Evaluation:** The system should enable novice surgeons to receive an evaluation of their performance after training expert performances.

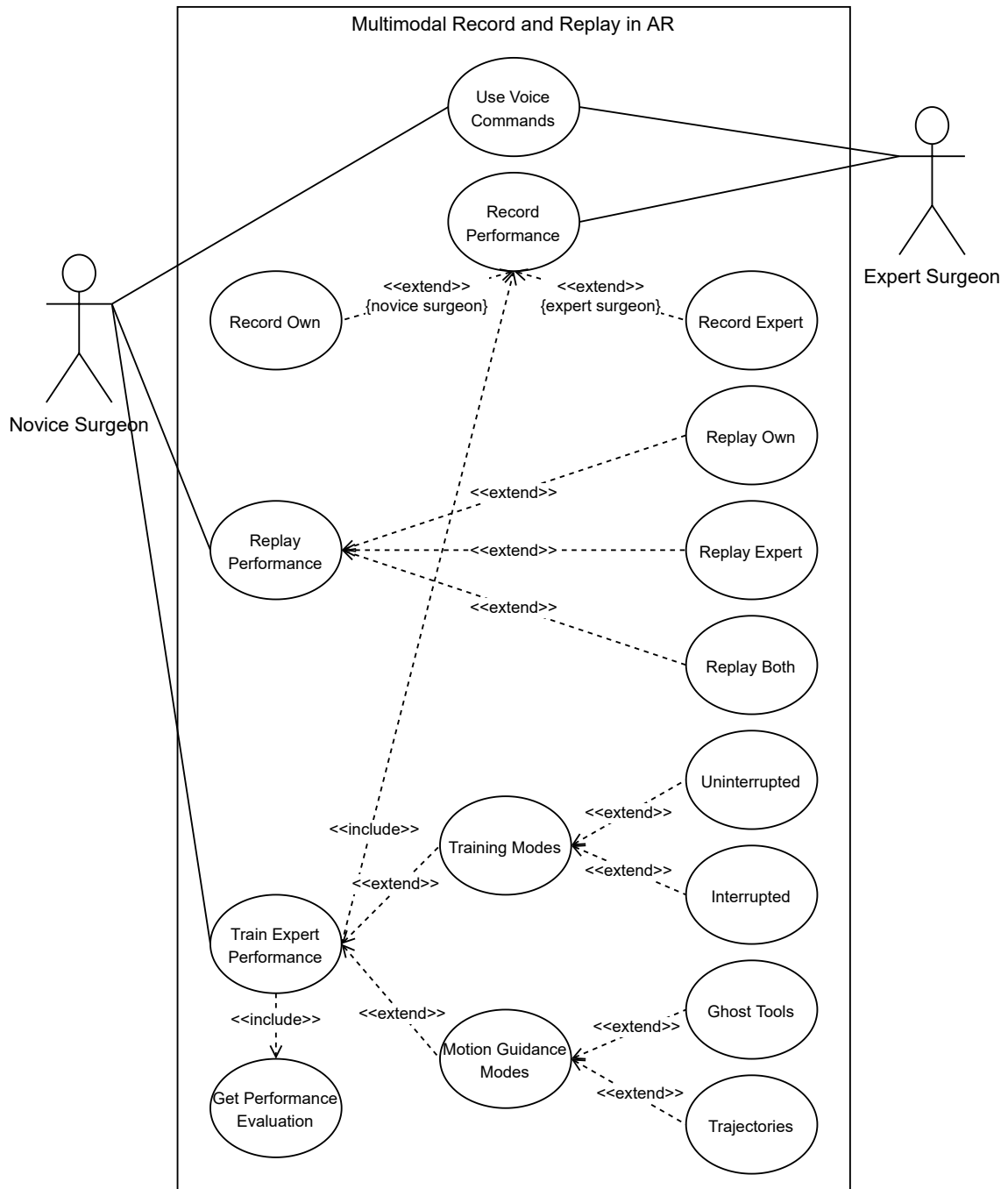


Figure 3.6: Use-case diagram of AR system.

**Non-Functional Requirements**

Non-functional requirements are all properties which the system should fulfill while providing the previously described functionalities. Here, the non-functional requirements consist of the following:

- **Research Prototype for RQs:** The system is intended to answer research questions and to allow surgeons to train under laboratory conditions. It is not intended for real surgeries on human patients.
- **Low Computation Time:** The computation time of the system per image processing cycle has to remain below 33 ms to allow fluent replaying of videos at 30 FPS.
- **User Friendly:** The system should allow surgeons a usage without the need of looking away from the surgical scene. This implies the usage of voice commands.

### 3.3.3 Architecture

The architecture of the system is designed to fulfill the functional and non-functional requirements elaborated in Section 3.3.2. For this purpose, a package diagram, a node-topic diagram, and sequence diagrams are created, including rationales for documenting architectural decisions.

#### Package Diagram

Package diagrams represent all packages within a software project with their communication between each other. The package diagram of this system is illustrated in Figure 3.7 and the functionality of each package can be described as follows:

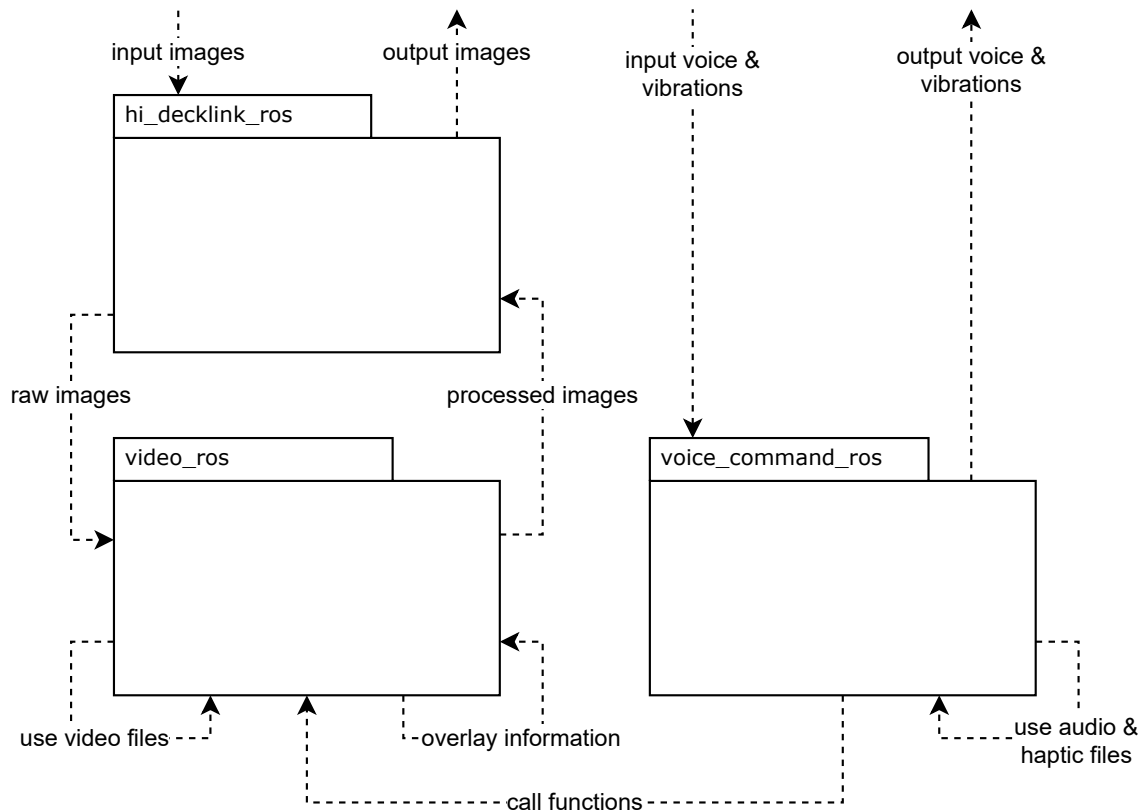
- **hi\_decklink\_ros:** The package contains all drivers required for receiving images from the da Vinci in ROS, and respectively for sending images from ROS to the da Vinci, via the Blackmagic DeckLink Quad 2 (see Section 3.1.2).
- **video\_ros:** The package contains all functionality related to recording, processing, and replaying of video files.
- **voice\_command\_ros:** The package contains all functionality related to voice commands as well as recording, processing, and replaying of audio and haptic files.

#### Node-Topic Diagram

Node-topic diagrams are an adaption of the previously introduced *ROS computation graph* in Section 3.2.1. These diagrams represent all nodes within their respective packages and their communication via topics between each other. The node-topic diagram of this system is illustrated in Figure 3.8. In the following, each package of the node-topic diagram will be described in more detail, including its nodes and published topics.

In the package *hi\_decklink\_ros*, the functionality of each node with its published topics can be described as follows:

- **endoscope/left/camera:** The node for input of left images from the da Vinci stereo endoscope.
  - **endoscope/left/image\_raw:** The topic for sending left images.
- **endoscope/right/camera:** The node for input of right images from the da Vinci stereo endoscope.



**Figure 3.7:** Package diagram of AR system.

- **endoscope/right/image\_raw**: The topic for sending right images.
- **camera\_left**: The node for output of left images to the da Vinci stereo viewer.
- **camera\_right**: The node for output of right images to the da Vinci stereo viewer.

In the package *video\_ros*, the functionality of each node with its published topics can be described as follows:

- **recorder**: The node for recording images as video files.
- **playback**: The node for replaying images from video files.
  - **overlay\_output\_left**: The topic for sending left images to overlay on actual left images from the da Vinci stereo endoscope which are then displayed in the left da Vinci stereo viewer.
  - **overlay\_output\_right**: The topic for sending right images to overlay on actual right images from the da Vinci stereo endoscope which are then displayed in the right da Vinci stereo viewer.
  - **function/output\_write**: The topic for sending opacity values to specify the opacity of the images which are overlaid on actual images of the da Vinci stereo endoscope and the displayed in the da Vinci stereo viewer.

- **function/record**: The topic for sending record commands to record novice performances while training.
- **tool\_detector**: The node for detecting and highlighting tools and grippers in video files.

In the package *voice\_command\_ros*, the functionality of each node can be described as follows:

- **voice\_control**: The node for enabling voice commands from the da Vinci microphone, and for recording and replaying audio files.
  - **function/rehearsal**: The topic for sending voice commands to activate AR functionalities.
- **haptic\_control**: The node for recording and replaying haptic files. This node is located within the voice command package as VerroTouch transforms measured robotic instrument vibrations to haptic and auditory representations (see Section 3.1.3). Here, the auditory representations are used for recording and replaying haptic files.

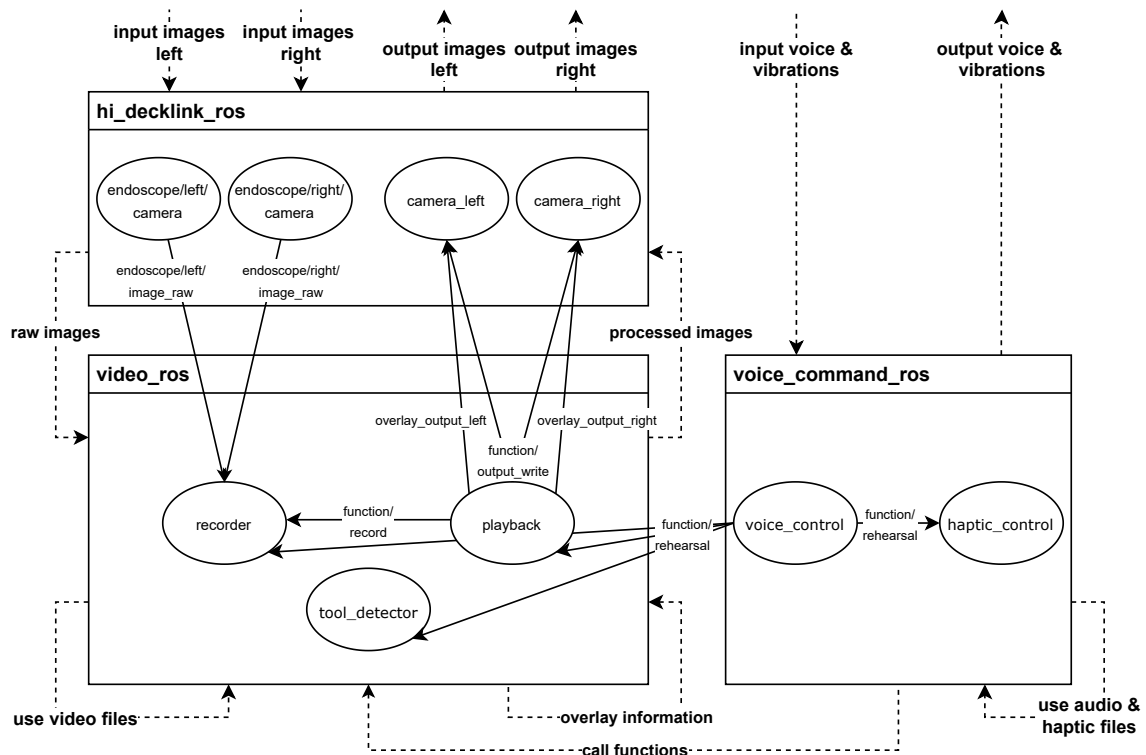


Figure 3.8: Node-topic diagram of AR system.

### Sequence Diagrams

Sequence diagrams visualize the dynamic control flow between nodes for specific use-case scenarios. The following sequence diagrams will be elaborated in more detail:

- **Record Expert**: The sequence diagram for the use-case *record expert*.



- **Replay Expert:** The sequence diagram for the use-case *replay expert*.
- **Train Expert Uninterrupted:** The sequence diagram for the use-case *train expert uninterrupted*.

For comprehensibility reasons, the vision pipeline is simplified in the sequence diagrams.

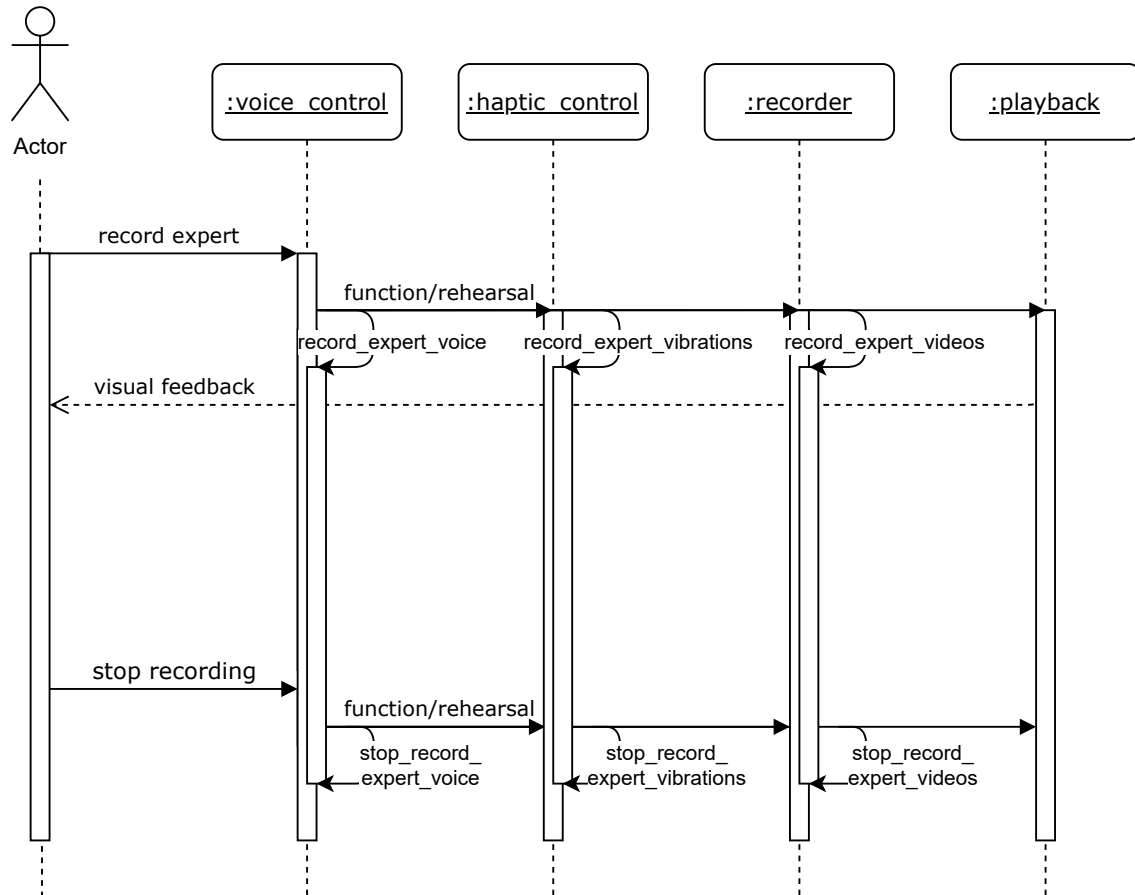
The simplification for the node *recorder* contains that it directly receives left and right input images of the da Vinci stereo endoscope. In the extended version, the node *recorder* receives left and right input images via the topics *endoscope/left/image\_raw* and *endoscope/right/image\_raw*, which are published by the nodes *endoscope/left/camera* and *endoscope/right/camera*.

The simplification for the node *playback* contains that it directly sends left and right images with overlaid information to the da Vinci stereo viewer. In the extended version, the node *playback* first creates transparent left and right images, then overlays information on these transparent images, and finally publishes the left and right transparent images with overlaid information via the topics *overlay\_output\_left* and *overlay\_output\_right*. Further, the node publishes an opacity value via the topic *function/output\_write* to specify the opacity with which the left and right images with overlaid information should be overlaid on the actual images of the da Vinci stereo endoscope, which are then displayed on the da Vinci stereo viewer. These steps are summarized as *visual feedback* in the sequence diagrams.

**Record Expert** The sequence diagram of the use-case *record expert* is illustrated in Figure 3.9 and can be described as follows:

1. **Actor:** The actor uses the voice command *record expert*.
2. **voice\_control:** The voice control node recognizes the voice command *record expert* and publishes the voice command via the topic *function/rehearsal*. The node further starts the recording of verbal explanations.
3. **haptic\_control:** The haptic control node receives the voice command *record expert* and starts the recording of left and right robotic instrument vibrations.
4. **recorder:** The recorder node receives the voice command *record expert* and starts the recording of left and right stereo endoscope images.
5. **playback:** The playback node receives the voice command *record expert* and provides *visual feedback* to the *Actor* by overlaying the text *Record Expert* on the left and right images displayed in the da Vinci stereo viewer.
6. **Actor:** The actor uses the voice command *stop recording*.
7. **voice\_control:** The voice control node recognizes the voice command *stop recording* and publishes the voice command via the topic *function/rehearsal*. The node further stops the recording of verbal explanations.
8. **haptic\_control:** The haptic control node receives the voice command *stop recording* and stops the recording of left and right robotic instrument vibrations.
9. **recorder:** The recorder node receives the voice command *stop recording* and stops the recording of left and right stereo endoscope images.

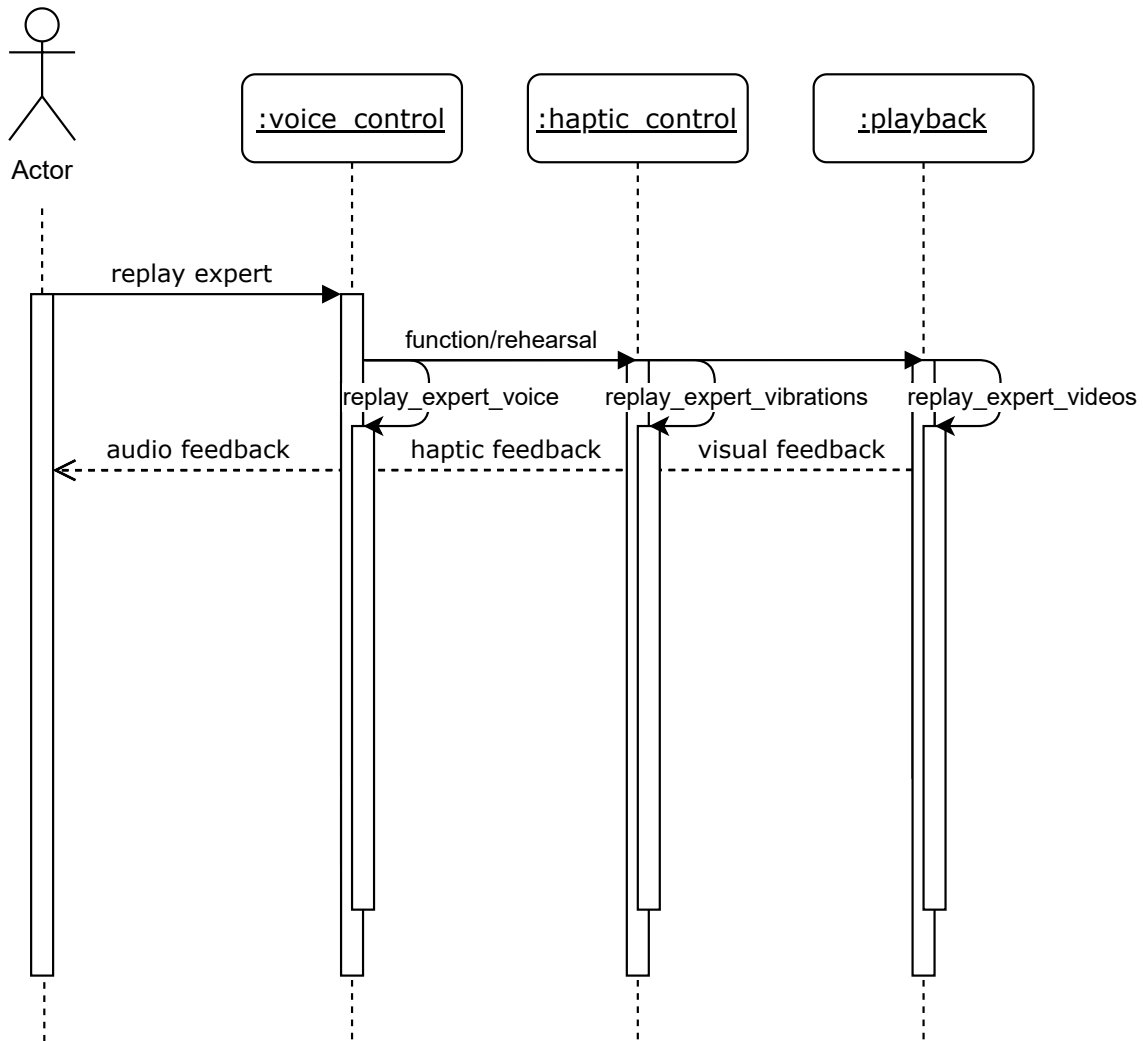
- 10. playback:** The playback node receives the voice command *stop recording* and stops providing *visual feedback* to the Actor.



**Figure 3.9:** Sequence diagram for record expert.

**Replay Expert** The sequence diagram of the use-case *replay expert* is illustrated in Figure 3.10 and can be described as follows:

- 1. Actor:** The actor uses the voice command *replay expert*.
- 2. voice\_control:** The voice control node recognizes the voice command *replay expert* and publishes the voice command via the topic *function/rehearsal*. The node further starts the replaying of verbal explanations as *audio feedback* for the Actor until the audio file is finished.
- 3. haptic\_control:** The haptic control node receives the voice command *replay expert* and starts the replaying of left and right robotic instrument vibrations as *haptic feedback* for the Actor until the haptic files are finished.
- 4. playback:** The playback node receives the voice command *replay expert* and starts the replaying of left and right images as *visual feedback* for the Actor by overlaying the replayed left and right images on the left and right images displayed in the da Vinci stereo viewer with high opacity until the video files are finished.



**Figure 3.10:** Sequence diagram for replay expert.

**Train Expert** The sequence diagram of the use-case *train expert* is illustrated in Figure 3.11 and can be described as follows:

1. **Actor:** The actor uses the voice command *train expert*.
2. **voice\_control:** The voice control node recognizes the voice command *train expert* and publishes the voice command via the topic *function/rehearsal*.
3. **playback:** The playback node receives the voice command *train expert* and starts the replaying of the first left and right image as *visual feedback* for the Actor by overlaying the replayed first left and right image on the left and right images displayed in the da Vinci stereo viewer with low opacity.
4. **Actor:** The actor uses the voice command *begin training*.
5. **voice\_control:** The voice control node recognizes the voice command *begin training* and publishes the voice command via the topic *function/rehearsal*.

6. **playback:** The playback node receives the voice command *begin training* and publishes this information via the topic *function/record*. Then, the node starts the replaying of the left and right images as *visual feedback* for the *Actor* by overlaying the replayed left and right images on the left and right images displayed in the da Vinci stereo viewer with low opacity.
7. **recorder:** The recorder node receives the information that the *Actor* chose to begin the training and starts the recording of left and right stereo endoscope images.
8. **playback:** When the video files are finished, the playback node publishes this information via the topic *function/record* and stops providing *visual feedback* to the *Actor*.
9. **recorder:** The recorder node receives the information that the replayed video files are finished and stops the recording of left and right stereo endoscope images.

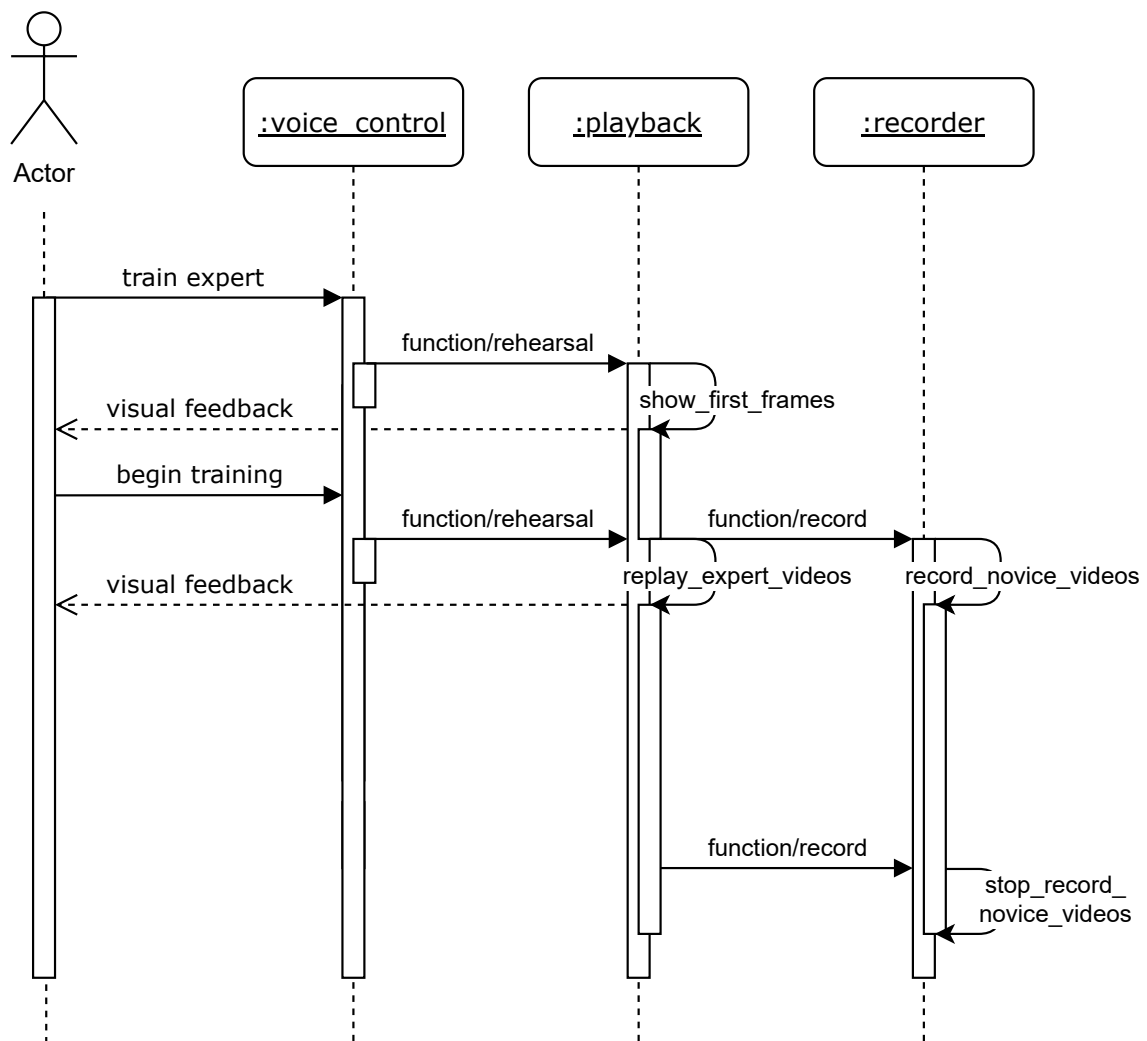


Figure 3.11: Sequence diagram for train expert.

## Rationale

The purpose of rationales is to document architectural decisions. The following criteria were considered during the creation of the architecture:

- **Synchronization:** To enable synchronization between left and right image streams, the video recorder and playback nodes were created as combined nodes for left and right image streams. This architectural decision allows controlled access to the left and right image streams from within the same node, without uncontrollable multi-threading through multiple ROS nodes, as it would have been the case for separate nodes for left and right image streams.
- **Loose Coupling:** The nodes should only communicate via messages without any required knowledge of the other nodes.
- **High Cohesion:** The communication between nodes via messages should be reduced to a minimum, which results in high cohesion within each node.
- **Simplicity:** The nodes and messages should be kept as simple as possible to improve *understand-ability*, which in turn results in high *extends-ability* of the system for future research.

### 3.3.4 Development

The system is developed on the basis of the elaborated requirements in Section 3.3.2 and the created architecture in Section 3.3.3. First, we implement voice commands to enable the user to activate AR functionalities with their voice. Then, we develop the AR functionalities of the system and describe their underlying basic functionalities. Finally, we describe algorithms to accomplish these basic functionalities in more detail.

#### Voice Commands

The voice commands of the system are implemented to enable users to activate AR functionalities with their voice. Through this, we create a user friendly system for surgeons as described in the non-functional requirements in Section 3.3.2.

To activate the voice commands, we first chose a *trigger word*. On the one hand, trigger words reduce the chance for false-positive activation of voice commands. On the other hand, trigger words also personalize the developed system for the user. With this in mind, we chose the trigger word *SEER*, representing *Surgical Expert Emulator in Augmented Reality* and also resembling the *Seer*.

The hereby created *voice commands* consist of the following:

- **SEER Record Expert**
- **SEER Stop Recording**
- **SEER Replay Mine**
- **SEER Replay Expert**

- **SEeAR Replay Both**
- **SEeAR Stop Replaying**
- **SEeAR Train Expert**
- **SEeAR Begin Training**
- **SEeAR Train Expert Piece-wise**
- **SEeAR Next Piece**
- **SEeAR Stop Training**

#### **AR Functionalities**

The AR functionalities of the system are developed in order to achieve the functional requirements described in Section 3.3.2. In particular, the developed system should enable expert surgeons to record surgical procedures with multiple modalities and novice surgeons to replay and train with the multimodal recordings, including visual, haptic, and auditory feedback. In addition, the system should also incorporate motion guidance concepts through the display of ghost tools.

Hence, the following *modalities* should be included:

- **Visual Feedback:** Visual feedback should be provided by recording left and right streams of a stereo endoscope and directly overlaying the recorded videos on the surgeon's stereoscopic view of the instruments and task materials.
- **Haptic Feedback:** Haptic feedback should be achieved by recording left and right robotic instrument vibrations with accelerometers attached to the robotic instruments and replaying the recorded vibrations on actuators attached to the surgeon handles.
- **Auditory Feedback:** Auditory feedback should be enabled by recording verbal explanations of a procedure and playing back the recordings on the surgeon speakers.

To provide the described behavior and modalities, the following underlying *basic functionalities* of the system are required:

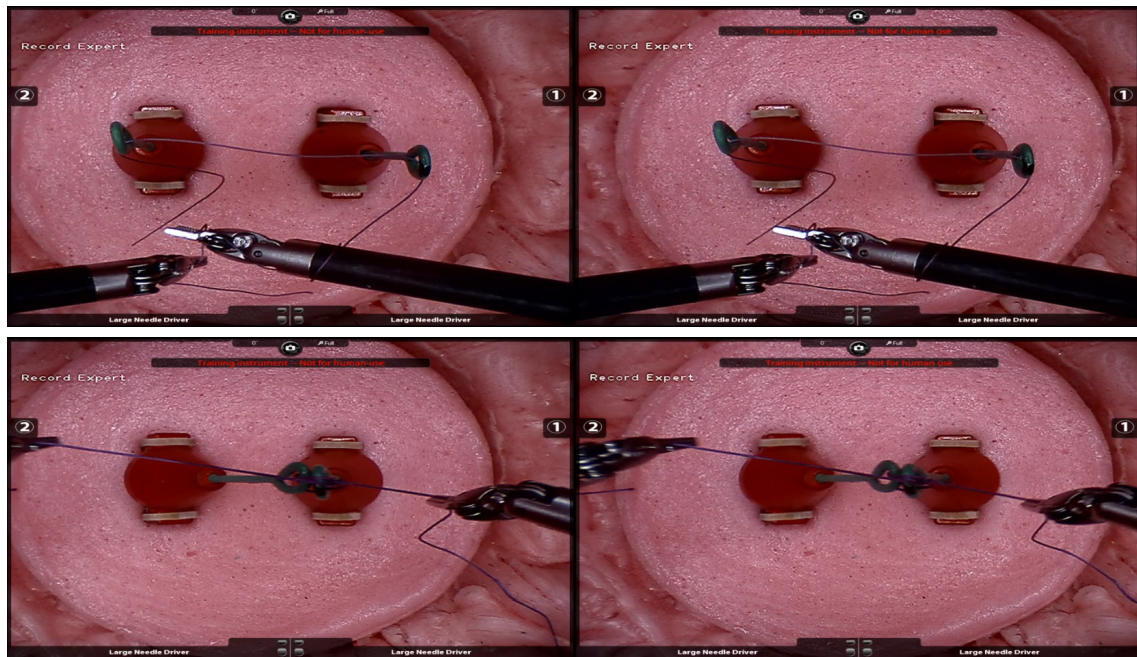
- **Record Videos**
- **Record Instrument Vibrations**
- **Record Verbal Explanations**
- **Replay Videos**
- **Replay Instrument Vibrations**
- **Replay Verbal Explanations**
- **Train with Uninterrupted Mode**
- **Train with Piece-wise Mode**

By combining these underlying basic functionalities, the following *AR functionalities* are achieved:

- **Record Expert**
- **Replay Expert**
- **Train Expert with Uninterrupted Mode**
- **Train Expert with Piece-wise Mode**

**Record Expert** For record expert, the focus lies on the expert performance. As such, only the own tools are *visible*. Here, the videos, instrument vibrations, and verbal explanations of the expert performance are recorded. The recording of expert performances can be seen in Figure 3.12 for knot tying and in Figure 3.13 for suturing. It includes the following basic functionalities:

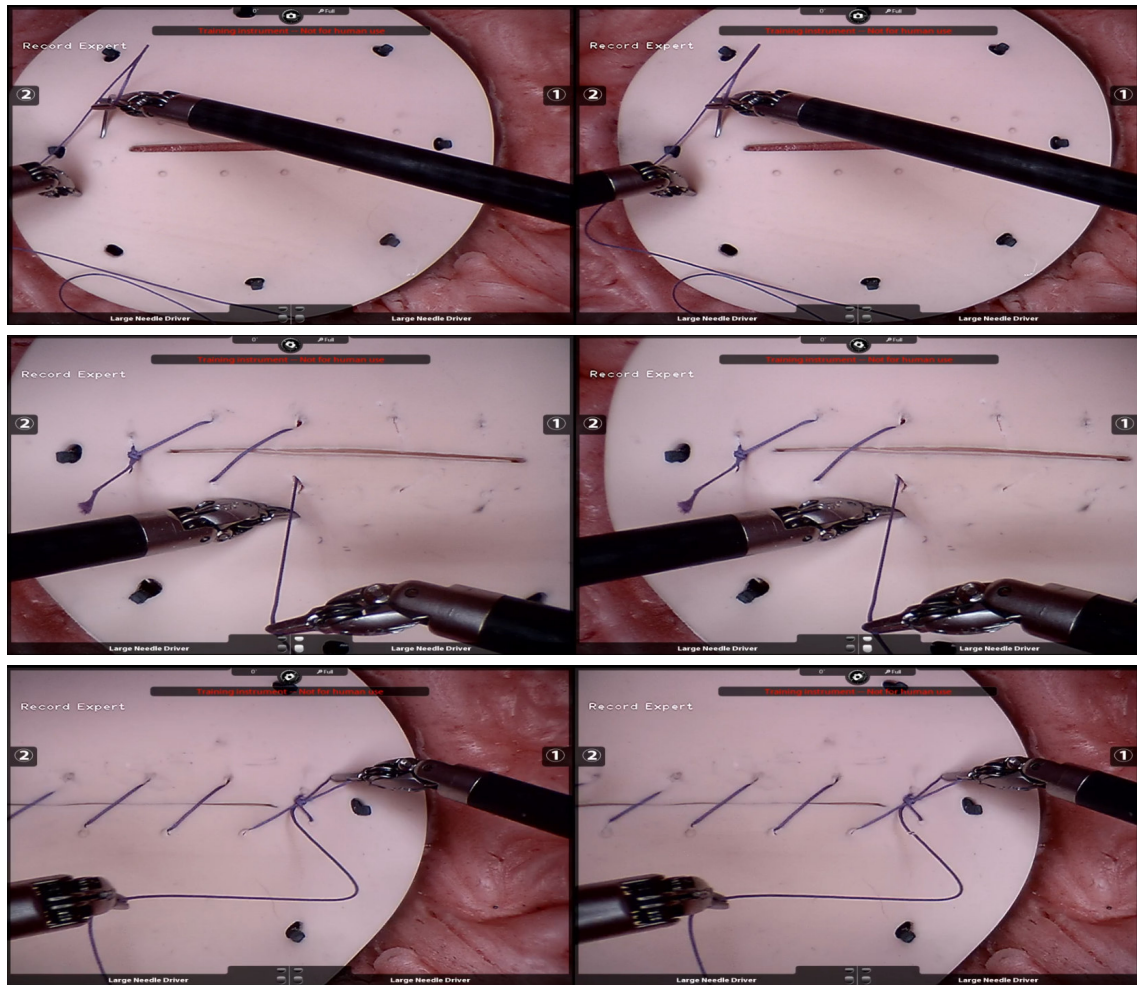
- **Record Expert Videos**
- **Record Expert Instrument Vibrations**
- **Record Expert Verbal Explanations**



**Figure 3.12:** Record functionality of left and right stereo endoscope videos, left and right instrument vibrations, and verbal explanations for knot tying task.

**Replay Expert** For replay expert, the focus lies on the replayed expert recordings. As such, the recorded video is *highly visible*, whereas the own tools are *less visible*. Here, the recorded instrument vibrations and verbal explanations of the expert performance are replayed as well. The replaying of expert performances can be seen in Figure 3.14 for knot tying and Figure 3.15 for suturing. It includes the following basic functionalities:

- **Replay Expert Videos**



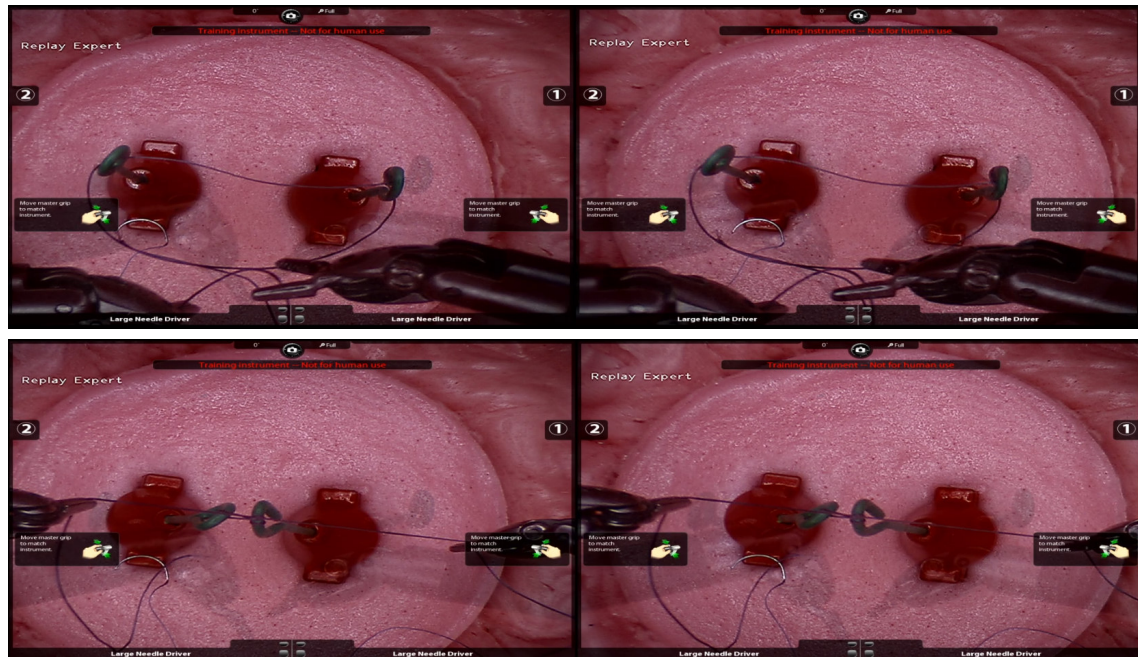
**Figure 3.13:** Record functionality of left and right stereo endoscope videos, left and right instrument vibrations, and verbal explanations for suturing task.

- **Replay Expert Instrument Vibrations**
- **Replay Expert Verbal Explanations**

**Train Expert with Uninterrupted or Piece-wise Mode** For train expert, the focus lies on the own performance. As such, the own tools are *highly visible*, whereas the expert tools are *less visible*. The training with expert performances in uninterrupted or piece-wise mode can be seen in Figure 3.16 with initial green filter for knot tying and suturing and in Figure 3.17 without green filter for suturing. It includes the following basic functionalities:

- **Record Trainee Videos**
- **Replay Expert Videos**
- **Train with Uninterrupted Mode**
- **Train with Piece-wise Mode**





**Figure 3.14:** Replay functionality of overlaid left and right stereo endoscope videos, left and right instrument vibrations, and verbal explanations for knot tying task.

### Additional AR Functionalities

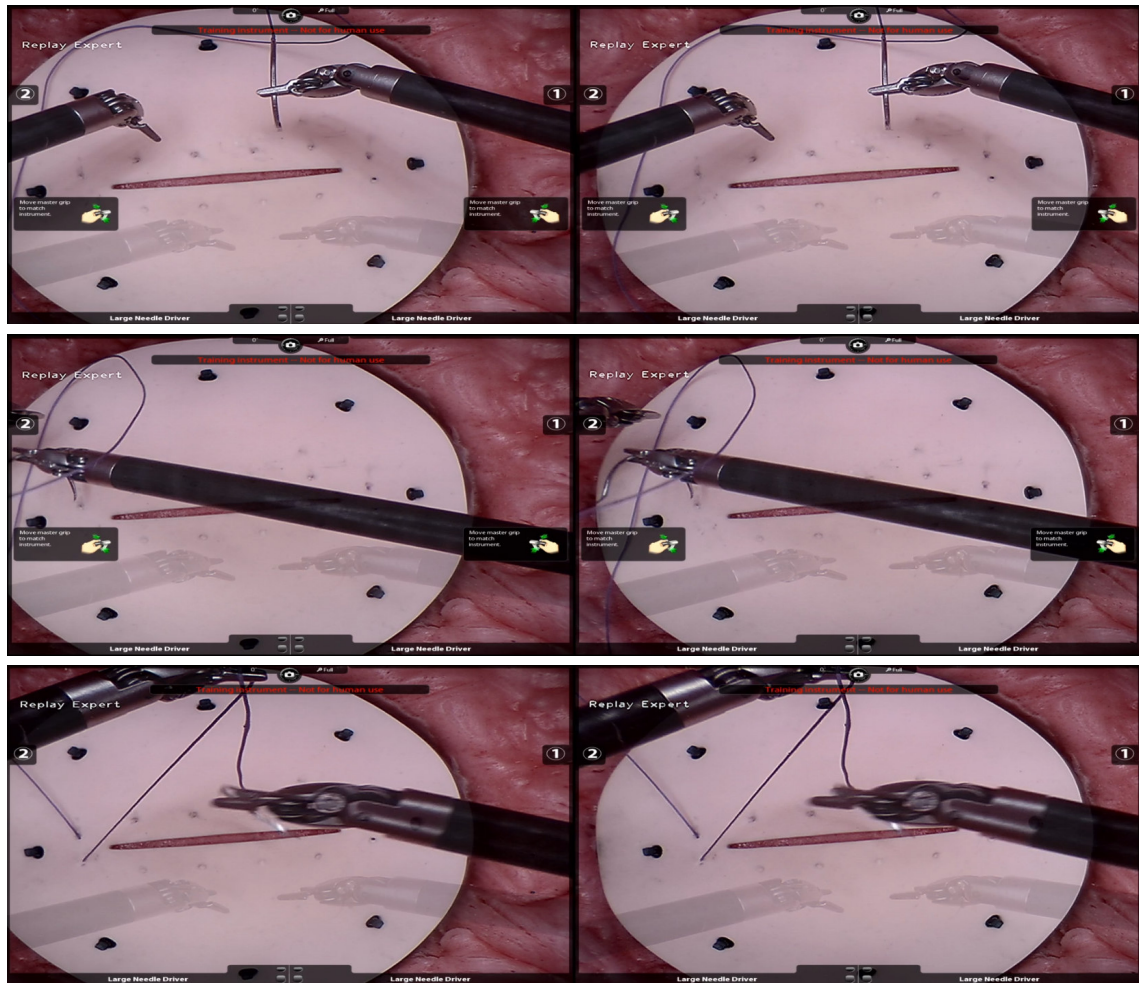
In addition to the developed AR functionalities of the system, we developed further AR functionalities for future use-cases. In particular, the system should also enable novice surgeons to replay their own performance and to compare their own performance with expert performances. In addition, the system should also include detection and highlighting of tools and grippers.

To provide the described *additional AR functionalities*, the following underlying *additional basic functionalities* of the system are required:

- **Detect Tools**
- **Detect Grippers**
- **Highlight Tools**
- **Highlight Grippers**

Through combining the previously developed *basic functionalities* and the *additional basic functionalities*, the following *additional AR functionalities* are achieved:

- **Replay Mine**
- **Replay Both**
- **Detect and Highlight Tools**
- **Detect and Highlight Gripper**



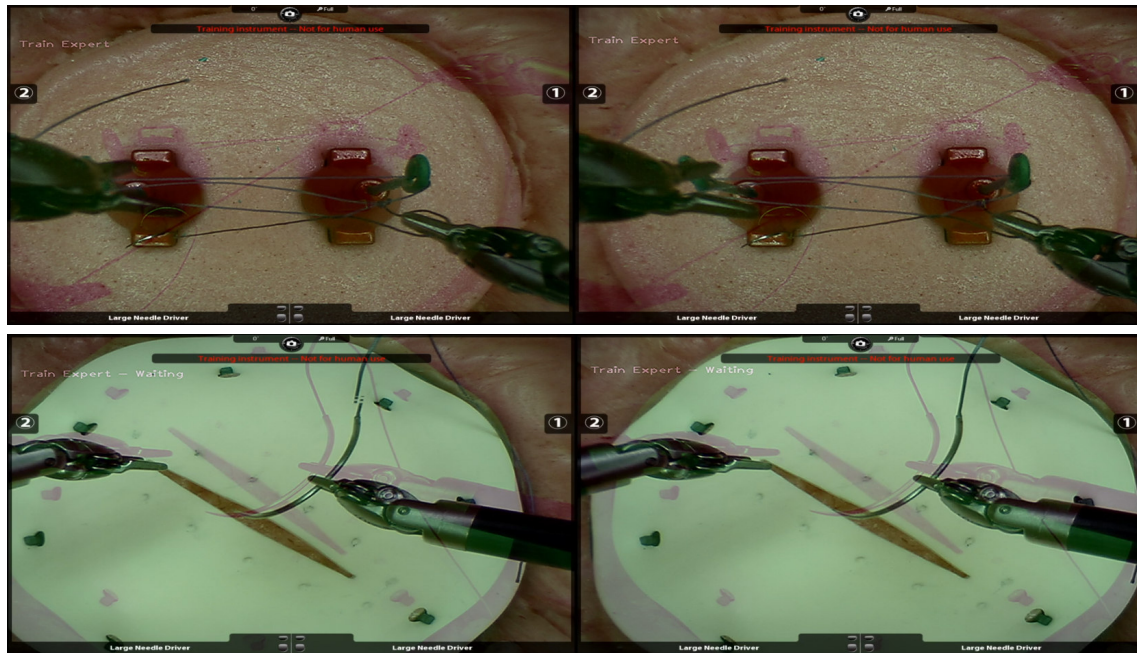
**Figure 3.15:** Replay functionality of overlaid left and right stereo endoscope videos, left and right instrument vibrations, and verbal explanations for suturing task.

**Replay Mine** For replay mine, the focus lies on the replayed trainee recordings. As such, the recorded video is *highly visible*, whereas the own tools are *less visible*. The replaying of trainee performances can be seen in Figure 3.14 for knot tying and Figure 3.15 for suturing. It includes the following basic functionalities:

- **Replay Trainee Videos**

**Replay Both** For replay both, the focus lies equally on the replayed trainee and expert recordings. As such, the recorded trainee and expert videos are *equally visible*. The replaying of trainee and expert performances at the same time includes the following basic functionalities:

- **Replay Expert Videos**
- **Replay Trainee Videos**



**Figure 3.16:** Train functionality of overlaid left and right stereo endoscope videos with initial green filter for knot tying and suturing task. Top: Knot tying task. Bottom: Suturing task.

**Detect and Highlight Tools or Grippers** The detection and highlighting of tools or grippers can be seen in Figure 3.18 and includes the following basic functionalities:

- **Detect Tools**
- **Detect Grippers**
- **Highlight Tools**
- **Highlight Grippers**

### Algorithms

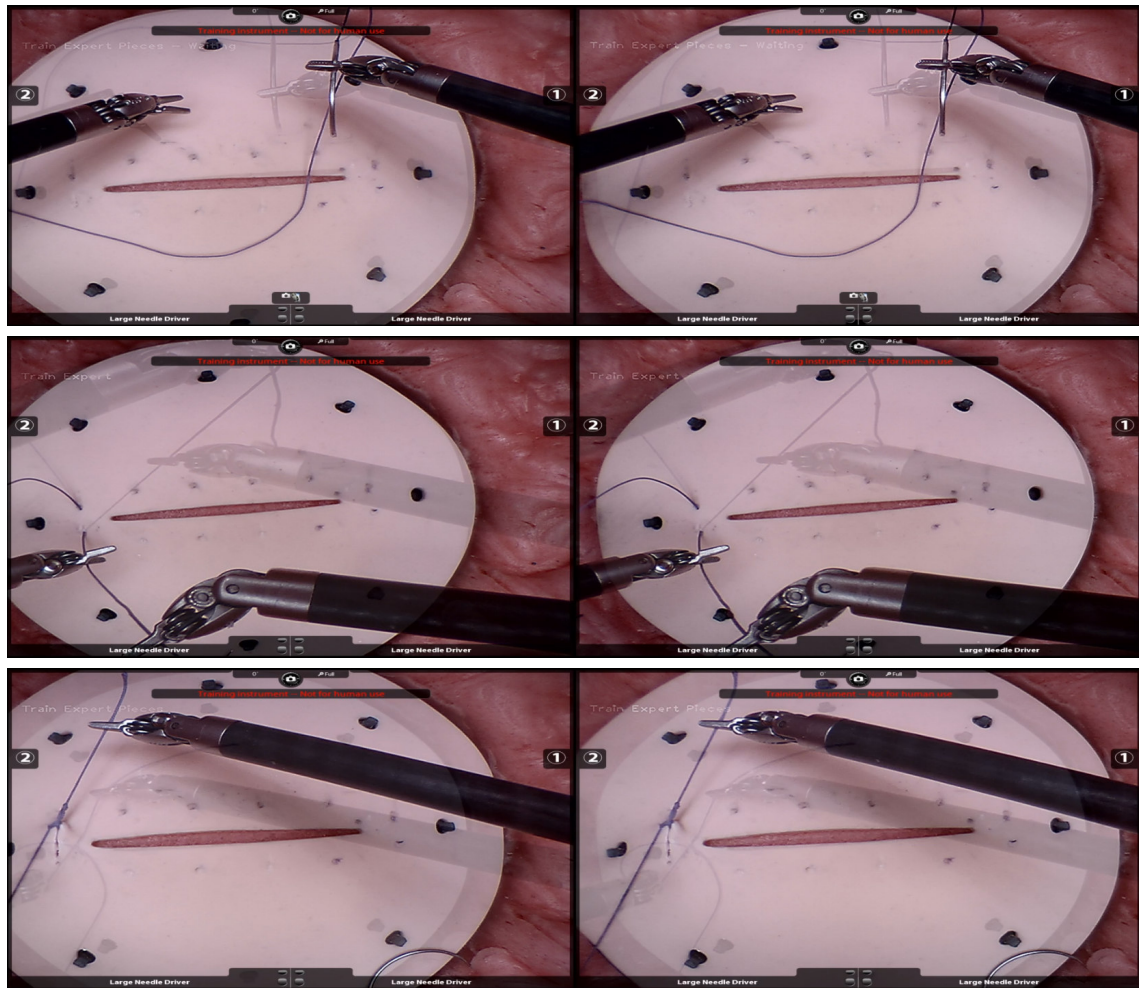
To accomplish the AR functionalities of the system including their underlying basic functionalities as described in Section 3.3.4 and the non-functional requirements described in Section 3.3.2, non-trivial algorithms were required. These algorithms are based on terms introduced in Section 1.2 and Section 3.2 and will be described in more detail in the following.

On the one hand, the following algorithms were implemented to be executed *online* (i.e., during the run-time of the system):

- **Stereo Camera Synchronization**
- **Video Processing Optimization**

On the other hand, the following algorithms were implemented to be executed *offline* (i.e., after the run-time of the system):

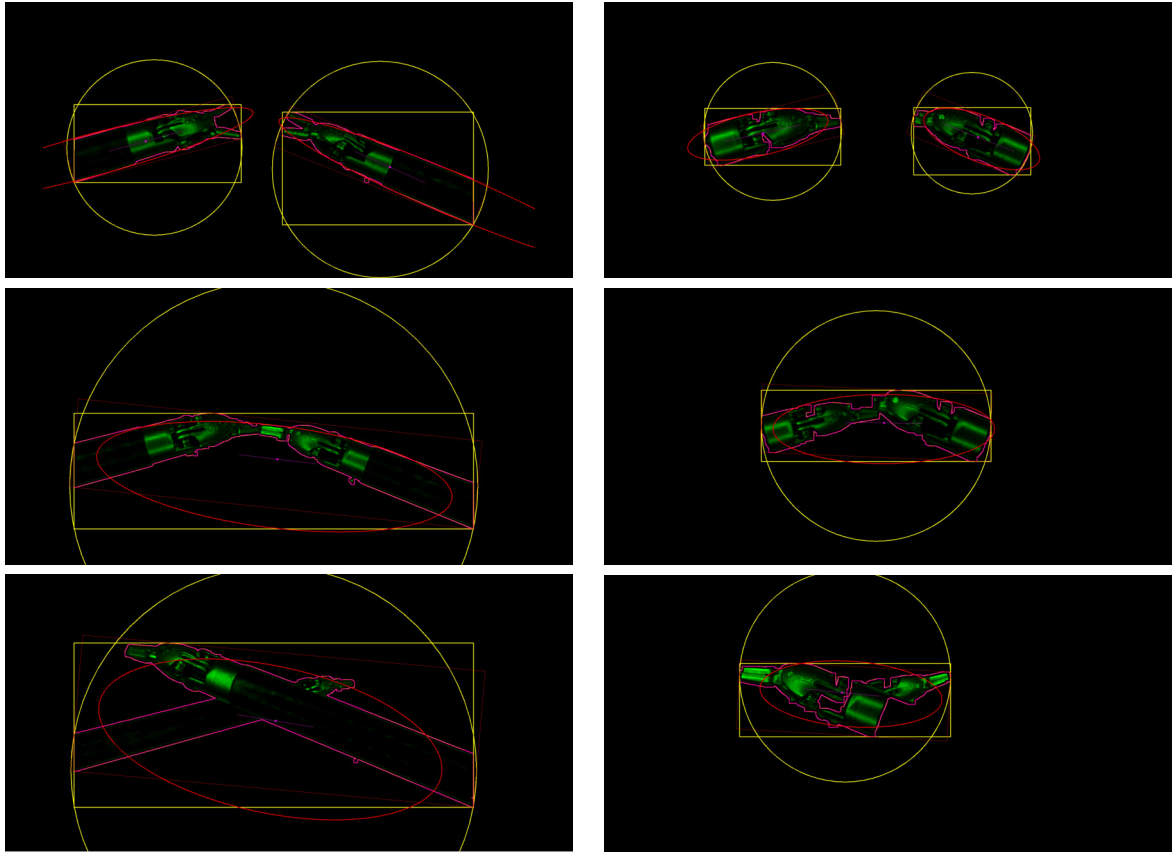




**Figure 3.17:** Train functionality of overlaid left and right stereo endoscope videos without green filter for suturing task.

- **Tool and Gripper Detection and Highlighting**

**Stereo Camera Synchronization** The *problem* in stereo camera synchronization was to ensure that left and right image streams were both equally progressing frame by frame for recording video files. This problem was not trivial as left and right images were received independently from each other. That is because left and right images were published by two publisher threads of the nodes *endoscope/left/camera* and *endoscope/right/camera* via two topics and were received via two callback threads in the node *recorder*. Due to this, there existed two publisher threads, two callback threads, and one main thread. Hence, there was no shared time, no shared memory, and no control over time or order of image arrival. The *solution* to this problem was inspired by *snapshot algorithms* and consisted of the following steps. First, two threads were created within the node *recorder* for left and right image streams. Second, a global timer with 33 ms steps was created



**Figure 3.18:** Detect and highlight functionality of left stereo endoscope video for instruments and grippers. Left: Instruments. Right: Grippers. Top: Separate condition. Middle: Touching condition. Bottom: Intersecting condition.

within the main thread of the node *recorder*. Third, after each time step, the latest received left and right image was written to the video file. Through this, it was ensured that left and right image streams were recorded with the closest frames available at each time step.

**Video Processing Optimization** The *problem* in video processing was to ensure that the *processing time* of each image-cycle remains below 33 ms, to stay within 30 FPS for fluent video recording and replaying as stated in Section 3.3.2. The *solution* to this problem consisted of the following steps. First, we identified that *processing time* is the summation of *computation time* for video operations and *file-access time* for read-write operations. Then, we identified that the bottle neck in *processing time* is the *file-access time*. Finally, we implemented *multi-threading* for left and right video files to reduce *file-access time*. Through this, the *processing time* of each image cycle was almost reduced by half the time, resulting in mean image cycles of 22 ms.

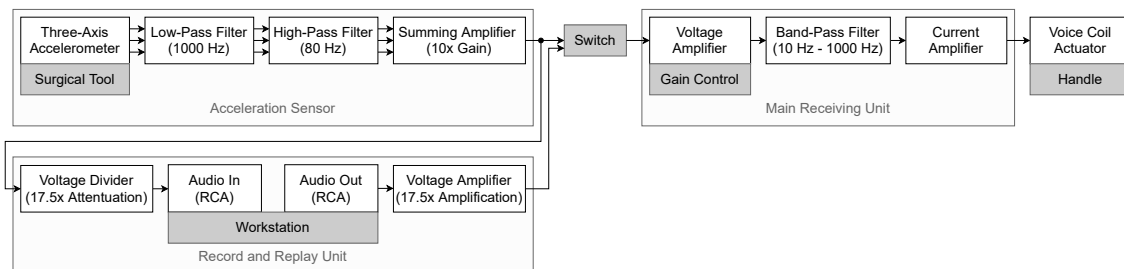
**Tool and Gripper Detection and Highlighting** The *problem* in tool and gripper detection and highlighting for motion guidance was to create a mask that only covers the tool or the gripper. Another problem was to ensure that the algorithm also works for cases in which the tools or grippers

touch or overlap each other. The *solution* to this problem consisted of the following steps. First, color segmentation was used to create a mask by separating the tools from the background, including different filters for darker and reflecting parts. Next, morphological transformations such as erosion, dilation, closing, and opening were used on the mask. Through this, external structures and internal structures of the tools were erased, such as noise outside the tools and holes within the tools, and contours covering the whole tool were achieved. Then, bounding boxes, contours, principal angles, and center of masses of the masks were calculated. Finally, the mask was applied to the original image to separate the tools from the background, and the calculated components were visualized. Then, another color segmentation with morphological transformations and subsequent steps was applied to separate the grippers from the tools following the previously described approach. Through this, a robust tool and gripper detection and highlighting was achieved for separate, touching, and overlapping conditions.

### 3.3.5 Integration of VerroTouch

The haptic feedback of the platform was developed on the basis of the introduced prototype *VerroTouch* described in Section 3.1.3. The following steps were performed to integrate *VerroTouch*. First, a signal flow for recording and replaying robotic instrument vibrations was designed (see Figure 3.19). Second, the system components of *VerroTouch* were added to the developed platform. These system components include two acceleration sensors attached to the robotic instruments (see Figure 3.20) and two actuators mounted to the surgeon handles (see Figure 3.23), whereas acceleration sensors and actuators are both connected to the main receiving unit (see Figure 3.21). Third, additional system components were created (see Figure 3.22) in order to achieve the designed signal flow:

- **Voltage Divider:** A voltage divider was required after the acceleration sensor to match the voltage input of the workstation computer in order to record instrument vibrations as sound input.
- **Voltage Amplifier:** A voltage amplifier was required after the workstation computer to match the voltage input of the main receiving unit in order to replay instrument vibrations as sound output.
- **Switch:** A switch was required to change between recording and replaying signal flows.



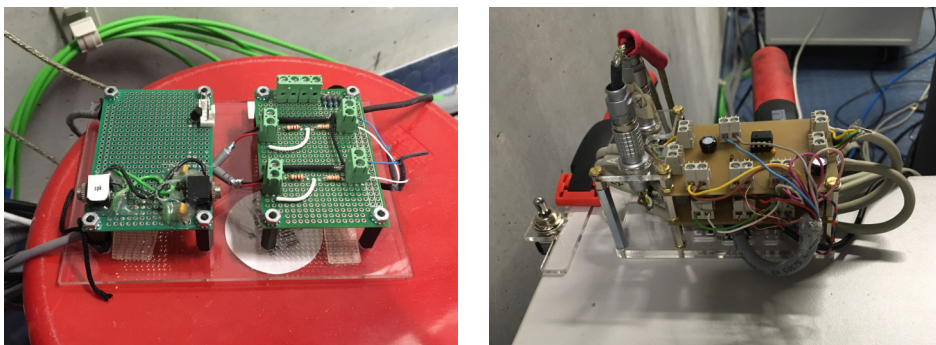
**Figure 3.19:** VerroTouch record and replay signal flow of one channel. Note that robotic instruments are called surgical tools in the diagram for consistency with the authors of VerroTouch.



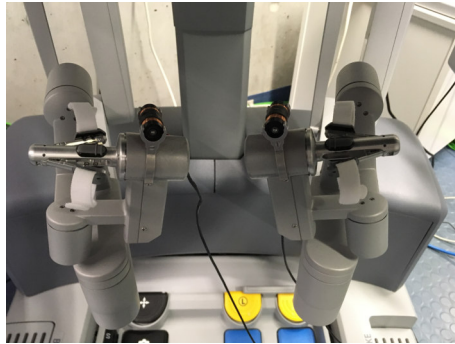
**Figure 3.20:** VerroTouch acceleration sensors on robotic instruments.



**Figure 3.21:** VerroTouch main receiving unit.



**Figure 3.22:** VerroTouch record and replay unit with voltage divider, voltage amplifier and switch. Left: Voltage divider. Right: Voltage amplifier and switch.



**Figure 3.23:** Voice coil actuators on surgeon handles.



## 4 Chief Surgeon Study

In this chapter, the conducted study of this work will be presented in more detail.

### 4.1 Research Questions

To answer the RQs of this work, we conducted an exploratory human-subject study with three chief surgeons. Within this chief surgeon study, we addressed the stated RQs from Section 1.3:

- **RQ1:** Is multimodal record and replay in augmented reality a beneficial approach for training in robotic surgery?
- **RQ2:** How could motion guidance and surgical skill evaluation in augmented reality benefit training in robotic surgery?

#### 4.1.1 RQ1 - Multimodal Record and Replay in AR

As stated in Section 1.3.1, the first aim of this work was to use the developed platform to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. For this purpose, the following *areas of interest* were introduced:

- **Integrability of AR Platform in Training Processes:** How can the developed AR platform be *integrated* in existing training processes for robotic surgery?
- **Record Functionality:** How do chief surgeons prefer to *record* performances with multiple modalities in AR for robotic surgery training?
- **Replay Functionality:** How do chief surgeons prefer to *replay* performances with multiple modalities in AR for robotic surgery training?

In the following, we introduce further *categories of interests* to answer the stated *RQ1* for multimodal record and replay in augmented reality.

#### Integrability of AR Platform in Training Processes

To understand the integrability of the developed AR platform in existing training processes, we investigated different aspects such as common training approaches and processes as well as integration possibilities for the AR platform. We further investigated the satisfaction with the manufactured training tasks and the chosen trigger word.

For the *integrability of the AR platform in training processes*, the following *categories of interest* can be defined:

- **Integration Possibilities in Common Training Processes:** What are common training approaches and processes? Are VR, AR, and physical simulators used? How could the developed AR platform be integrated into these existing training processes?
- **Training Task Satisfaction:** How satisfying are the manufactured tasks?

### Record Functionality

To understand the recording of performances, we investigated different aspects such as modalities. Thereby, the focus for the recording functionality is on the live performance. This allows expert surgeons to concentrate on the performance to record.

For the *record functionality*, the following *categories of interest* can be defined:

- **Modality Preference for Recording:** What modalities are preferred for *recording* surgical procedures?

### Replay Functionality

To understand the replaying of performances, we investigated different aspects such as modalities. Thereby, the focus for the replaying functionality is on the previously recorded performance. This allows novice surgeons to concentrate on the replayed performance.

For the *replay functionality*, the following *categories of interest* can be defined:

- **Visual Overlay Satisfaction for Replaying:** How satisfying is the direct visual overlay for *replaying* surgical procedures?
- **Modalities Preference for Replaying:** What modalities are preferred for *replaying* surgical procedures?

### 4.1.2 RQ2 - Motion Guidance and Skill Evaluation in AR

As stated in Section 1.3.2, the second aim of this work was to use the developed platform to explore different concepts for motion guidance and surgical skill evaluation in augmented reality for training in robotic surgery. For this purpose, the following *areas of interest* were introduced:

- **Train Functionality with Motion Guidance:** How do chief surgeons prefer to *train* performances with multiple modalities and motion guidance in AR for robotic surgery training?
- **Sensors for Surgical Skill Evaluation:** Which sensors can be used for surgical skill evaluation in robotic surgery training?

In the following, we introduce further *categories of interests* to answer the stated *RQ2* for multimodal training with motion guidance and skill evaluation in augmented reality.

### Train Functionality with Motion Guidance

To understand the training of performances with motion guidance, we investigated different aspects such as training modes and modalities as well as motion guidance concepts such as ghost tools and trajectories. Thereby, the focus for training performances with motion guidance is on the live performance. This allows novice surgeons to concentrate on the performance to train.

For the *train functionality with motion guidance*, the following *categories of interest* can be defined:

- **Visual Overlay and Motion Guidance Preference for Training:** How satisfying is the direct visual overlay for *training* surgical procedures? What motion guidance concepts are preferred for *training* surgical procedures?
- **Modality Preference for Training:** What modalities are preferred for *training* surgical procedures?
- **Training Mode Preference:** What training mode is preferred for *training* surgical procedures?

### Sensors for Surgical Skill Evaluation

To understand sensors that could be used for surgical skill evaluation, we investigated exerted forces on the task board and vibrations of the robotic instruments.

For the *sensors for surgical skill evaluation*, the following *categories of interest* can be defined:

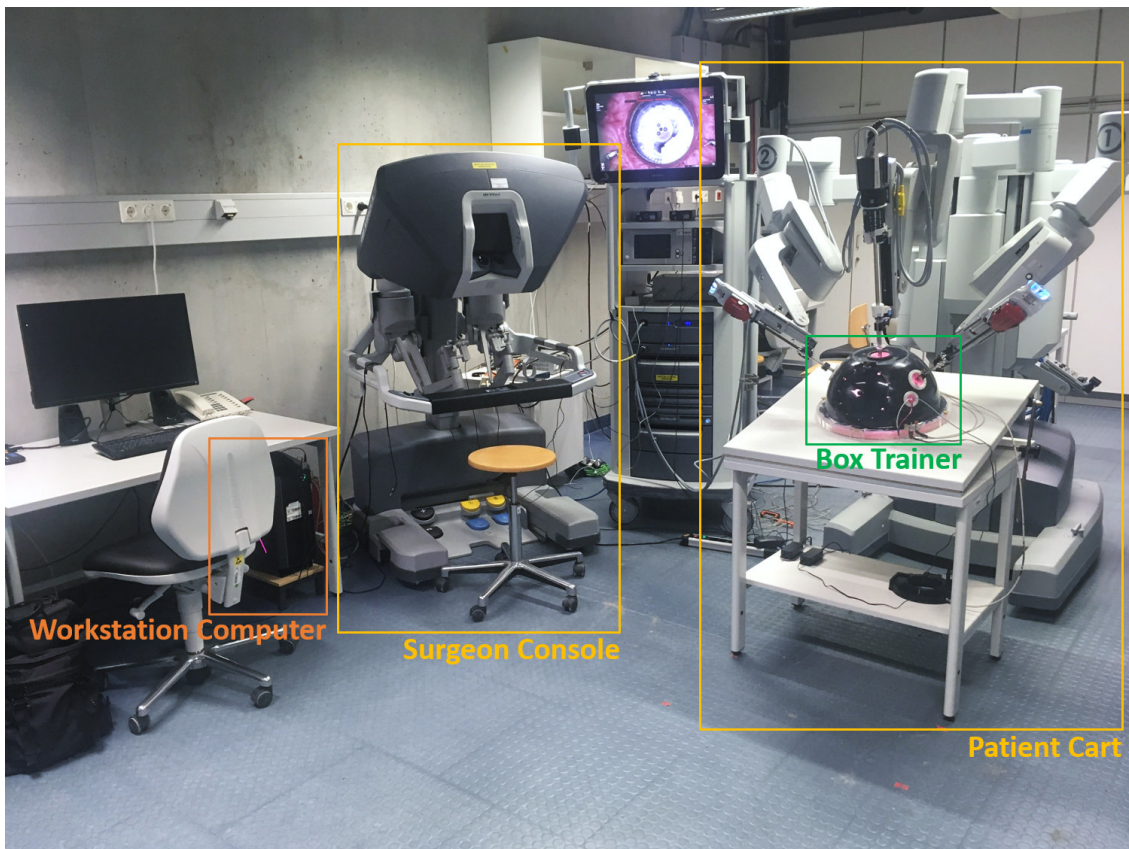
- **Sensor Preference:** What sensors are preferred for surgical skill evaluation? Which further sensors could be used?
- **Metric Suggestions:** Which metrics could be used for surgical skill evaluation?

## 4.2 Experimental Setup

To conduct the human-subject study of this work, we used the developed platform of Section 3.3.4 and the existing hardware setup of Section 3.1.1. First, we extended the existing box trainer and manufactured two standardized training tasks. Then, we placed a force sensor below the task board and used the previously attached accelerometers on the robotic instruments to explore possibilities for quantitative evaluation of surgical skill performances. Hence, the experimental setup of this work consists of the following components (see Figure 4.1):

- **Clinical da Vinci Si HD Surgical System:** The da Vinci Si HD surgical system introduced in Section 1.1.2 is the surgical robot used for training of robotic surgery. It is a clinical version of the surgical system, is connected to the workstation computer, and consists of a surgeon console, a vision cart, and a patient cart.
- **Workstation Computer:** The workstation computer is the processing point of the system for multimodal record and replay in augmented reality. It recognizes voice commands, processes images, and outputs virtual content.

- **Laptop Computer:** The laptop computer is the recording point of the system for the force sensor, the accelerometers, and the augmented image streams of the da Vinci. It records the experiment by capturing images acquired by the stereo endoscope with augmented views. It also records the voice of the experimenter and the surgeon.
- **Box Trainer:** The box trainer is the created environment for training procedures in robotic surgery. It is attached to a tilting table.
- **Training Tasks:** The training tasks are the procedures to be trained in robotic surgery. They are placed within the box trainer.
- **Force Sensor:** The force sensor measures the exerted forces of the robotic instruments on the task board. It is placed below the training tasks.
- **Accelerometer:** The accelerometer measure the vibrations of the robotic instruments. They are attached to the robotic instruments.



**Figure 4.1:** Experimental setup with the clinical da Vinci Si HD Surgical System. Note that the laptop computer is located further left on the table.

### Workstation Computer

The workstation computer is the processing point of the system for multimodal record and replay in augmented reality. As already described in Section 3.1.1, it is an *Alienware Aurora R7 Desktop* [Del] with the following components:

- **CPU:** Intel(R) Core(TM) i7-8700K CPU @ 4.60GHz (overclocked across all cores), 6 Cores, 12MB Cache
- **GPU:** NVIDIA GeForce GTX 1080 Ti with 11GB GDDR5X memory
- **Operating System:** Ubuntu 16.04 LTS
- **Video Capture and Playback Card:** Blackmagic Design DeckLink Quad 2 [Bla]

### Laptop Computer

The laptop computer is the recording point of the system for the force sensor, the accelerometers, and the augmented image streams of the da Vinci. It is an *Dell Inc. Latitude 7390 2-in-1* with the following components:

- **CPU:** Intel(R) Core(TM) i7-8650U CPU @ 1.90GHz, 2112 Mhz, 4 Cores, 8 Logical Processors
- **Operating System:** Microsoft Windows 10 Enterprise LTSC

In addition, the following software is executed for data collection with sensors and recording image and audio streams:

- **MATLAB:** The software *MATLAB* is commercial software with an integrated numerical computing environment and programming language [Mata], which is commonly used for sensor processing and data collection [Matb]. Here, it is used to record the exerted force on the task board from the force sensor.
- **OBS Studio:** The software *Open Broadcast Software (OBS) Studio* is a free open-source software for video recording and live streaming [Con]. Here, it is used to record the augmented left and right image streams from the da Vinci surgeon's stereoscopic view of the instruments and task materials as well as the left and right robotic instrument vibrations from the accelerometers.

### Box Trainer

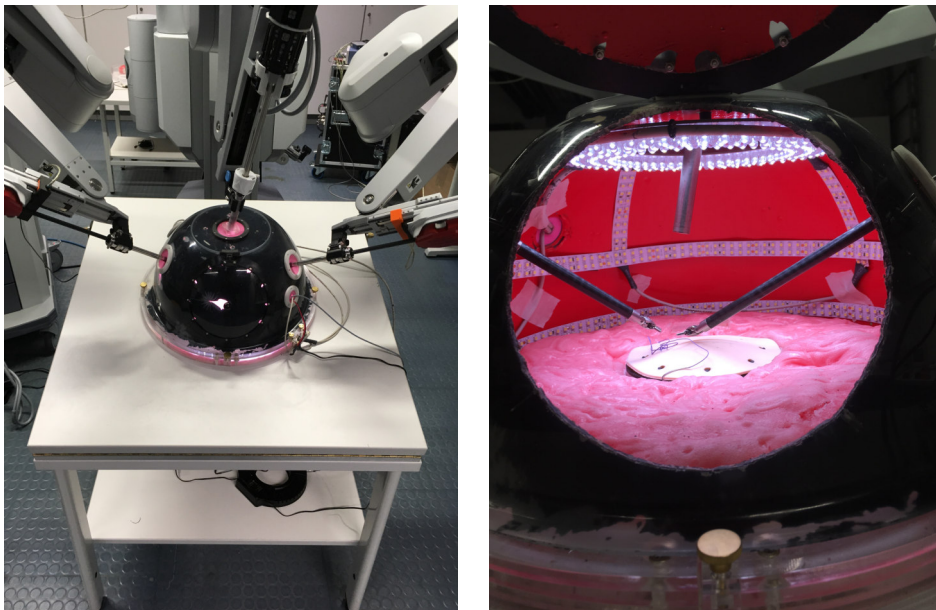
The box trainer is the created environment for training procedures in robotic surgery. As already described in Section 3.1.1, it was created to ensure realistic training environments with consistent lighting conditions and initially consisted of the following parts:

- **Globe:** The globe was built using a rigid bowl upside down. The bowl was painted black and five circular windows were drilled into it. The windows were filled with membranes containing small holes in their centers to allow internal instrument access for the stereo endoscope, the left and right robotic instruments, and the left and right assistant instruments.

- **LED Ring:** The LED ring was attached to the top of the globe from within to ensure consistent lighting conditions.
- **Foam Ring:** The foam ring was created to ensure a realistic training environment that resembles the environment within the human body. It was prepared with *Smooth-On Soma Foama 25*, which is a soft two-component platinum-cured silicone casting foam, and *UVO Colorant*, which is a red coloring substance.
- **Task Plate:** The task plate was created to allow switching tasks to be trained in robotic surgery.
- **Simulated Tissue:** The simulated tissue was created to ensure a realistic training task that resembles real tissue within the human body. It was also prepared with the described soft two-component silicone casting foam and the red coloring substance.

We extended the existing box trainer and made the following modifications (see Figure 4.2):

- **Door:** We cut a door in the globe using a heated knife to melt the plastic. Through this, it is possible to switch tasks within the box trainer without removing the globe from the tilted table.
- **LEDs:** We attached additional LED rings to ensure more consistent lighting conditions without shadows of the robotic instruments.



**Figure 4.2:** Box trainer with additional door and LEDs.

### Training Tasks

The training tasks are the procedures to be trained in robotic surgery. For this purpose, we manufactured two standardized tasks based on the FRS tasks introduced in Section 1.2.5. Here, the following tasks have been chosen and manufactured:

- Knot Tying:** We created a knot tying task using an elastic band-based approach (see Figure 4.3). First, we 3D-printed the transparent plate with two elevations. The transparent plate was screwed to the underlying task plate to ensure the transition of pushing and pulling forces to the force sensor beneath. Second, we 3D-printed the left and right cylinders and used red spray paint to color them. The cylinders were attached on the two elevations of the transparent plate using two elastic bands for each. Third, we prepared the left and right wires with the eyelets and used green spray paint to color them. The wires were inserted in the cylinders and glued with instant glue at the insertion hole to restrict rotation. Fourth, we 3D-printed a round phantom with two elevations for creating the simulated tissue with two placeholders for the two elevations of the transparent plate. Finally, we created a simulated foam tissue using the 3D-printed round phantom and the described soft two-component silicone casting foam with 20 grams of each component and the red coloring substance with half a drop. As robotic *EndoWrist* instruments, we used two *large needle drivers*. As suturing material, we used *POLYSORB CV-23 1/2 17 mm* needles with 3-0 75 cm suture.
- Suturing:** We created a suturing task using a hook-based approach (see Figure 4.4). First, we 3D-printed the transparent plate to screw on the underlying task plate. Second, we 3D-printed the black hollow cylinder with eight hooks to screw on the transparent plate. The hooks were required to attach artificial skin on top. Third, we 3D-printed the black elevation piece to place within the black hollow cylinder. The elevation piece was required to elevate the foam in the center to create tension on the hooked artificial skin so that the cut opens up. Fourth, we laser-cut a linear cut in the middle of the artificial skin sample and further laser-cut eight round holes around the outer edges so that the eight hooks of the black elevation piece could fit through. We further laser-engraved twelve equally distributed dots alongside the linear cut as insertion points for the suturing needle. The material used for the artificial skin was silicone-based blank tattoo practice skin sheets with two mm thickness from *ATOMUS*. Finally, we created a simulated foam tissue to put beneath the artificial skin using the described soft two-component silicone casting foam with 30 grams of each component and the red coloring substance with one drop. As robotic *EndoWrist* instruments, we used two *large needle drivers*. As suturing material, we used *VICRYL Plus VCP340H CT-1 1/2c 36 mm* needles with 0 70 cm suture.

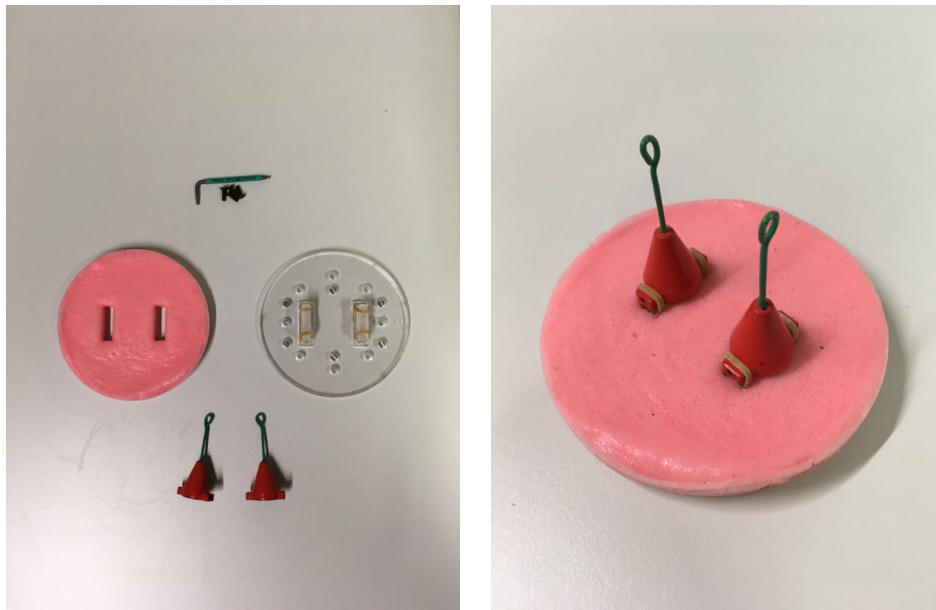
### Force Sensor

The force sensor measures the exerted forces of the robotic instruments on the task board. This functionality is commonly accomplished by transducers converting applied forces into measurable electrical outputs based on hydraulic, pneumatic, piezoelectric, or capacitive principles. Here, we used a *Mini40-E Transducer* (by ATI Industrial Automation) [Aut] and placed it below the task board (see Figure 4.5) to explore possibilities for quantitative evaluation of surgical skill performances.

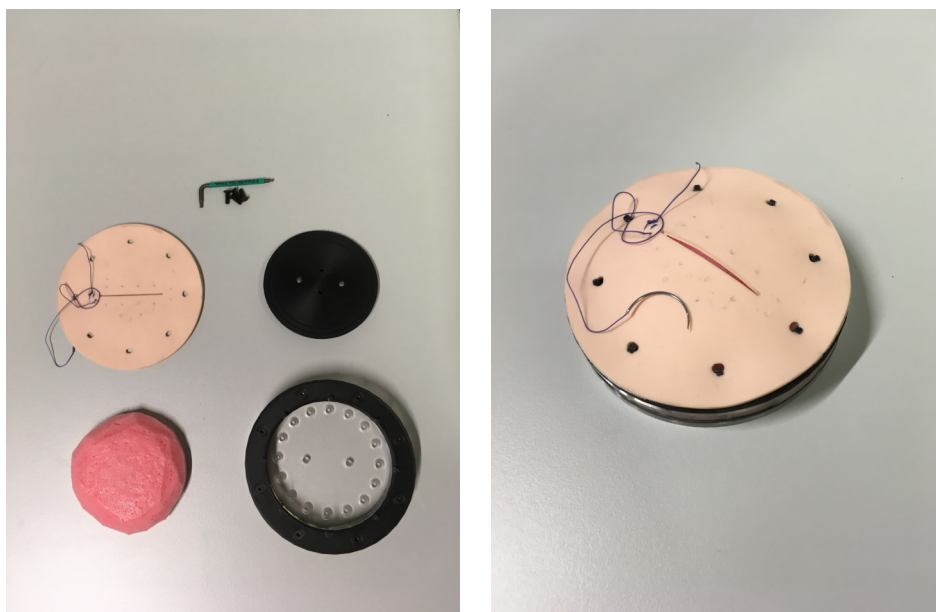
### Accelerometer

The two accelerometers measure the accelerations of the left and right robotic instruments and hence also their vibrations. This functionality is commonly accomplished by using known masses and converting their velocity change rates into measurable electrical outputs. Here, we used the





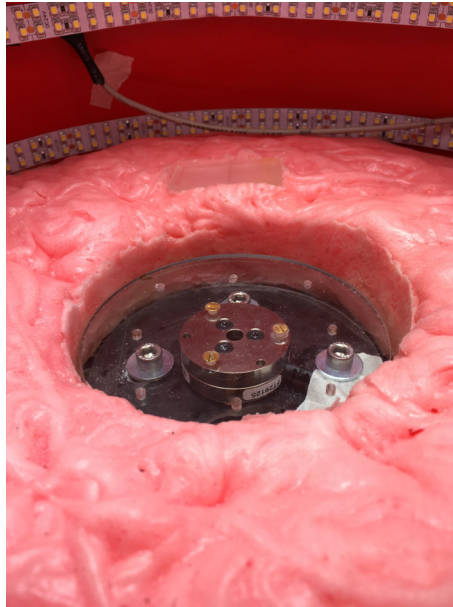
**Figure 4.3:** Training task knot tying.



**Figure 4.4:** Training task suturing.

previously integrated *VerroTouch* (see Section 3.3.5) with its accelerometers attached to the robotic instruments (see Figure 4.6) to explore possibilities for quantitative evaluation of surgical skill performances.





**Figure 4.5:** Mini40-E Transducer force sensor below task board.



**Figure 4.6:** VerroTouch acceleration sensors on robotic instruments.

### 4.3 Participants

Three chief surgeons participated in the human-subject study of this work. No participant was excluded as all participants fulfilled the inclusion criterion of at least two years of clinical experience in robotic surgery. The demographic information of the participants is summarized in Table 4.1. For *open surgery*, the mean clinical experience was 30.67 years, with a mean of 250 cases performed in the last year and a mean teaching experience of 18.67 years. For *MIS*, the mean clinical experience was 13.33 years, with a mean of 66.67 cases performed in the last year and a mean teaching experience of 10 years. For *robotic surgery*, the mean clinical experience was six years, with a mean of 100 cases performed in the last year and a mean teaching experience of 5.33 years. All participants had at least two years of clinical and teaching experience in robotic surgery. One participant was specialized in visceral surgery, one in urological surgery, and one in cardio-thoracic surgery. One participant stated to have no experience in MIS as it is not used in cardiology.

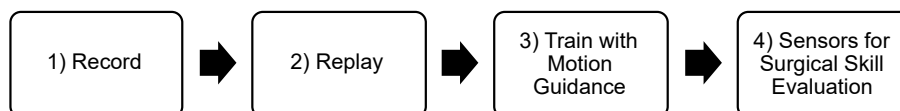
	S1	S2	S3
<b>Gender</b>	Male	Male	Male
<b>Handedness</b>	Right	Right	Right
<b>Age</b>	40 - 49 years	≥ 60 years	50 - 59 years
<b>Surgical Specialty</b>	Visceral	Urological	Cardio-thoracic
<b>Clinical Experience (Open Surgery)</b>	~ 20 years	~ 40 years	32 years
<b>Teaching Experience (Open Surgery)</b>	~ 10 years	~ 20 years	26 years
<b>Cases Last Year (Open Surgery)</b>	~ 300 cases	~ 200 cases	~ 250 cases
<b>Clinical Experience (MIS)</b>	~ 10 years	~ 30 years	0 years
<b>Teaching Experience (MIS)</b>	~ 10 years	~ 20 years	0 years
<b>Cases Last Year (MIS)</b>	~ 200 cases	0 cases	0 cases
<b>Clinical Experience (Robotic Surgery)</b>	2 years	14 years	2 years
<b>Teaching Experience (Robotic Surgery)</b>	2 years	12 years	2 years
<b>Cases Last Year (Robotic Surgery)</b>	~ 100 cases	~ 100 cases	~ 100 cases

**Table 4.1:** Demographic information of participants.

#### 4.4 Procedure

The procedures of the human-subject study of this work were approved by the Ethics Council of the Max Planck Society under the Haptic Intelligence framework protocol number F015A. All participants provided their informed consent to participate in the study prior to data collection. A payment of eight euros per hour was offered to all participants for their participation.

The participants completed the study in a lab environment with the described experimental setup of Section 4.2. The study was self-timed and took between 90 to 120 minutes. First, participants filled out an informed consent document (see Appendix A.1) and a demographic questionnaire (see Appendix A.2). Then, participants were seated in front of the da Vinci surgeon console and were given the possibility to adjust the height of the da Vinci stereo viewer and the da Vinci surgeon handles to their preferences. Next, participants received in-experiment instructions and practiced the voice commands by reading every command once. Finally, participants used the developed record and replay platform in augmented reality of Section 3.3.4 to complete sequential blocks of one to two trials in each phase (i.e., record, replay, train with motion guidance) per manufactured training task (i.e., knot tying and suturing). The procedure of the chief surgeon study is illustrated in Figure 4.7, and the complete structure with all questions can be seen in Appendix A.3.



**Figure 4.7:** Procedure for chief surgeon study.

After each phase, participants were requested to fill out questionnaires (see Appendix A.4, Appendix A.5, Appendix A.6, Appendix A.7, and Appendix A.8) and to answer interview questions (see Appendix A.3) in order to evaluate their experience with the concepts of the developed platform, and to provide suggestions for improvement. In addition, we also measured exerted forces on the task board and vibrations of the robotic instruments during the recording and training of performances.

### **Record Procedure**

In the study, participants were first asked to record performances for the two tasks without and with verbal explanations. In particular, participants recorded one to two performances for each of the two manufactured training tasks without and with verbal explanations per modality condition (i.e., only live visual feedback, live visual and live haptic feedback, live visual feedback with verbal explanations). On a trial, participants performed one repetition of the prepared task. The user experience and suggestions for improvements were acquired with a questionnaire (see Appendix A.4) and a structured interview (see Appendix A.3). The design was a 2 (verbal explanations: without vs. with)  $\times$  2 (haptic feedback: without vs. with) within-subjects design.

### **Replay Procedure**

Next, participants were asked to replay with their multimodal recordings. In particular, participants replayed one to two performances for each of the two manufactured training tasks per modality condition (i.e., only recorded visual feedback, recorded visual and recorded haptic feedback, recorded visual and recorded auditory feedback, recorded visual and recorded haptic and recorded auditory feedback). The user experience and suggestions for improvements were acquired with a questionnaire (see Appendix A.5) and a structured interview (see Appendix A.3). The design was a 2 (auditory feedback: without vs. with)  $\times$  2 (haptic feedback: without vs. with) within-subjects design.

### **Train with Motion Guidance Procedure**

Then, participants were asked to train with their multimodal recordings. In particular, participants trained one to two performances for each of the two manufactured training tasks per training mode and modality condition (i.e., uninterrupted training mode, piece-wise training mode, recorded visual feedback only, recorded visual feedback and live haptic feedback, recorded visual feedback and recorded auditory feedback, recorded visual feedback and live haptic feedback and recorded auditory feedback). Afterward, we showed different motion guidance concepts to the participants. For training with motion guidance, the user experience and suggestions for improvements were acquired with questionnaires (see Appendix A.6 and Appendix A.7) and a structured interview (see Appendix A.3). Afterward, we explained different motion guidance concepts to the participants. The design was a 2 (training mode: uninterrupted vs. piece-wise)  $\times$  2 (haptic feedback: without vs. with)  $\times$  2 (motion guidance: ghost tools vs. trajectories)  $\times$  2 (sensors: force sensor vs. accelerometers) within-subjects design.

### Sensors for Surgical Skill Evaluation Procedure

Finally, we explained the currently used sensor concepts for surgical skill evaluation to the participants. For sensors for surgical skill evaluation, the user opinions and suggestions were acquired with a questionnaire (see Appendix A.8) and a structured interview (see Appendix A.3).

## 4.5 Iterations

Over the course of the chief surgeon study, two iterations of the experimental setup, the developed AR system, the questionnaires, and the procedures have evolved. For the experimental setup, the differences between both iterations are summarized in Table 4.2. For the developed AR system, the differences between both iterations are summarized in Table 4.3. For the questionnaires, the differences between both iterations are summarized in Table 4.4. For the procedures, the differences between both iterations are summarized in Table 4.5.

In regard to questionnaires and procedures, the *second iteration* of documents is attached in the appendix. In regard to participants, Subject 1 completed the study with the *first iteration*, whereas Subject 2 and Subject 3 completed the study with the *second iteration*.

Experimental Setup	
First Iteration	Second Iteration
<p><b>Knot Tying Task</b></p> <ul style="list-style-type: none"> <li>• Open bottom of metallic wire eyelet</li> <li>• Magnetic fixation of cylinder to base</li> <li>• Dry suture</li> </ul> <p><b>Suturing Task</b></p> <ul style="list-style-type: none"> <li>• Thinner skin</li> <li>• Dry suture</li> </ul> <p><b>Task Board</b></p> <ul style="list-style-type: none"> <li>• Not screwed to plate of force sensor, so tasks lift up when being pulled</li> </ul> <p><b>Laptop Computer</b></p> <ul style="list-style-type: none"> <li>• Video and force recording (no vibration recording)</li> </ul>	<p><b>Knot Tying Task</b></p> <ul style="list-style-type: none"> <li>• Closed bottom of metallic wire eyelet</li> <li>• Elastic band fixation of cylinder to base</li> <li>• Wet suture</li> </ul> <p><b>Suturing Task</b></p> <ul style="list-style-type: none"> <li>• Thicker skin</li> <li>• Wet suture</li> </ul> <p><b>Task Board</b></p> <ul style="list-style-type: none"> <li>• Screwed to plate of force sensor, so tasks do not lift up when being pulled</li> </ul> <p><b>Laptop Computer</b></p> <ul style="list-style-type: none"> <li>• Video, force and vibration recording</li> </ul>

**Table 4.2:** Iterations of experimental setup.

<b>AR System</b>	
First Iteration	Second Iteration
<p><b>Record Functionality</b></p> <ul style="list-style-type: none"> <li>• Recording performances without verbal explanations</li> <li>• Recording of only video (no haptic or audio)</li> </ul> <p><b>Replay Functionality</b></p> <ul style="list-style-type: none"> <li>• Replaying performances with only visual feedback</li> </ul> <p><b>Train Functionality</b></p> <ul style="list-style-type: none"> <li>• Training performances with green filter and equally visible own tools and expert tools as visual feedback</li> <li>• Only uninterrupted training mode</li> </ul>	<p><b>Record Functionality</b></p> <ul style="list-style-type: none"> <li>• Recording performances without and with verbal explanations</li> <li>• Recording of video, haptic and audio</li> </ul> <p><b>Replay Functionality</b></p> <ul style="list-style-type: none"> <li>• Replaying performances with visual, haptic and auditory feedback</li> </ul> <p><b>Train Functionality</b></p> <ul style="list-style-type: none"> <li>• Training performances with no green filter and more visible own tools than expert tools as visual feedback</li> <li>• Uninterrupted and piece-wise training mode</li> </ul>

**Table 4.3:** Iterations of AR system.

<b>Questionnaires</b>	
First Iteration	Second Iteration
<p><b>Replay Functionality</b></p> <ul style="list-style-type: none"> <li>• Opinion about adding haptic feedback</li> <li>• Opinion about adding auditory feedback</li> </ul> <p><b>Train Functionality</b></p> <ul style="list-style-type: none"> <li>• Opinion about uninterrupted training mode only</li> </ul>	<p><b>Replay Functionality</b></p> <ul style="list-style-type: none"> <li>• Opinion about haptic and auditory feedback</li> </ul> <p><b>Train Functionality</b></p> <ul style="list-style-type: none"> <li>• Opinion about uninterrupted and piece-wise training mode</li> </ul>

**Table 4.4:** Iterations of questionnaires.

<b>Procedures</b>	
First Iteration	Second Iteration
<p><b>Record Procedure</b></p> <ul style="list-style-type: none"> <li>• Record performance with only visual feedback</li> <li>• Record performance with visual and live haptic feedback</li> </ul> <p><b>Replay Procedure</b></p> <ul style="list-style-type: none"> <li>• Replay performance with only visual feedback</li> <li>• Replay performance with visual and haptic feedback</li> </ul> <p><b>Train Procedure</b></p> <ul style="list-style-type: none"> <li>• Train performance with only visual feedback</li> <li>• Train performance with visual and live haptic feedback</li> <li>• Train performance with only uninterrupted training mode</li> </ul>	<p><b>Record Procedure</b></p> <ul style="list-style-type: none"> <li>• In addition, record performance with visual feedback and verbal explanations</li> </ul> <p><b>Replay Procedure</b></p> <ul style="list-style-type: none"> <li>• In addition, replay performance with visual and auditory feedback</li> <li>• In addition, replay performance with visual, haptic and auditory feedback</li> </ul> <p><b>Train Procedure</b></p> <ul style="list-style-type: none"> <li>• In addition, train performance with piecewise training mode</li> </ul>

**Table 4.5:** Iterations of procedures.

## 5 Results

In this chapter, the results of the chief surgeon study will be presented.

The qualitative data analysis to obtain the results included questionnaire and interview responses of participants regarding their experience with the concepts of the developed platform and their suggestions for improvements. It further included observations of the participants using the developed platform. Finally, it also included measured exerted forces on the task board and vibrations of the robotic instruments during the recording and training of performances.

As already discussed in Section 4.5, two iterations of the experimental setup, the developed AR system, the questionnaires, and the procedures evolved throughout the chief surgeon study. In regard to participants, Subject 1 completed the study with the *first iteration*, whereas Subject 2 and Subject 3 completed the study with the *second iteration*. The results of the conducted study will be presented in more detail in the following.

### 5.1 Integrability of AR Platform in Training Processes

For the integrability of the developed AR platform in existing training processes, the following results can be reported.

#### 5.1.1 Integration Possibilities in Common Training Processes

For integration possibilities in common training processes, we investigated *own training processes*, *own training tasks*, *own robotic surgical systems*, and *recommended integration of AR approach*.

**Own Training Processes** For the own training process, S1 stated the following process:

1. **VR Simulator:** The VR simulator *dVSS* (by Intuitive Surgical Inc.) is used until the defined metric benchmarks are achieved by the novice surgeon, such as a certain percentage of accuracy.
2. **Physical Simulator:** The *dual console* (by Intuitive Surgical Inc.) is used together with expert surgeons on simulated tasks.
3. **Operating Room:** The *dual console* (by Intuitive Surgical Inc.) is used together with expert surgeons in the operating room.

S2 stated the following process:

1. **VR Simulator:** The VR simulator *dVSS* (by Intuitive Surgical Inc.) is used until the defined metric benchmarks are achieved by the novice surgeon, such as a certain percentage of accuracy.
2. **Physical Simulator:** The *dual console* (by Intuitive Surgical Inc.) is used together with expert surgeons for training *simple procedures on anatomical models and animal organs*.
3. **Operating Room:** The *dual console* (by Intuitive Surgical Inc.) is used together with expert surgeons for training *clinical procedures on patients*.

S3 stated the following process:

1. **VR Simulator:** The VR simulator *dVSS* (by Intuitive Surgical Inc.) is used until the defined metric benchmarks are achieved by the novice surgeon, such as 30 hours of training time.
2. **Physical Simulator:** The *dual console* (by Intuitive Surgical Inc.) is used together with expert surgeons on *anatomical models*, whereas the expert surgeon performs procedures and gives control to the novice surgeon from time to time.
3. **Operating Room:** The *dual console* (by Intuitive Surgical Inc.) is used together with expert surgeons on *clinical patients*, whereas the novice surgeon performs procedures and the expert surgeon observes and takes control in case of an emergency as a backup.

For the VR simulator *dVSS* (by Intuitive Surgical Inc.), S3 added that it is well developed and that it enables training without expert surgeons, which is beneficial as the time of expert surgeons is quite limited. However, the subject also stated that VR simulators are never realistic enough, as it is just different within the human body.

For the *dual console* (by Intuitive Surgical Inc.), S3 added that it is also well developed. Here, the subject pointed out that the dual console enables expert surgeons to give control to novice surgeons from time to time. Through this, novice surgeons can practice important parts of procedures and gather experience on real patients, while expert surgeons are still there as a backup if something goes wrong. The subject further pointed out that the dual console allows expert surgeons to take over control of the third robotic arm from time to time so that expert surgeons can assist novice surgeons in practicing or performing challenging parts of procedures.

**Own Training Tasks** For own training tasks, S1 stated that suturing is rarely used in visceral surgery. S2 stated to train novice surgeons on anatomical models and animal organs for urological surgery. S3 stated that much suturing is used in Cardio-thoracic surgery. The subject further stated not to know the *FRS curriculum* for robotic surgery training.

**Own Robotic Surgical Systems** For the own robotic surgical system, S3 stated to use the *da Vinci Xi*. Here, the subject added that the 3D vision improved compared to the *da Vinci Si HD*, and also that the instruments improved as suture does not get caught in the joints anymore.



**Recommended Integration of AR Approach** For integrating the developed AR approach in common training processes, S3 suggested the following training process for novice surgeons:

1. **VR Simulator:** Training with the VR simulator *dVSS* (by Intuitive Surgical Inc.) without expert surgeons, until defined metric benchmarks are achieved.
2. **AR Simulator:** Training with the developed AR simulator without expert surgeons until defined metric benchmarks are achieved.
3. **Physical Simulator:** Training with the *dual console* (by Intuitive Surgical Inc.) together with expert surgeons on *anatomical models and animal organs*, whereas expert surgeons perform procedures and give control to novice surgeons from time to time.
4. **Operating Room:** Training with the *dual console* (by Intuitive Surgical Inc.) together with expert surgeons on *clinical patients*, whereas novice surgeons perform procedures and expert surgeons observe and take control in case of emergencies as a backup.

Here, S3 recommended to integrate the developed AR simulator after virtual training with the VR simulator *dVSS* (by Intuitive Surgical Inc.) and before first clinical procedures on patients with the *dual console* (by Intuitive Surgical Inc.). Further, S3 added to like the AR approach and to believe that the approach would fit very well in between the VR simulator and the operating room with the dual console.

### 5.1.2 Training Task Satisfaction

For training task satisfaction, we investigated *knot tying task satisfaction* and *suturing task satisfaction*.

**Knot Tying Task Satisfaction** For the *first iteration* of the knot tying task, S1 recommended to close the eyelet of the metallic wire at the bottom. Further, S1 recommended to use a longer standard suture. Also, S1 noticed that the robotic instruments were magnetically charged, which attracted the metallic needle to the robotic instruments and hence disturbed at performing tasks. For the *second iteration* of the knot tying task, we replaced the magnetic approach with an elastic band approach. After this modification, S2 and S3 stated to be satisfied with the manufactured task.

**Suturing Task Satisfaction** For the *first iteration* of the suturing task, S1 noticed that the artificial skin easily breaks when sutures are tightened to close the gap in the tissue. Due to this, S1 recommended to use either thicker skin which is more robust or thinner suture which breaks before the skin. For the *second iteration* of the suturing task, S2 and S3 stated to be satisfied with the manufactured task. Further, S2 suggested to use thinner sutures which break easier when being pulled too strong in order to avoid tissue damage. In addition, S3 stated that the chosen skin material is not realistic enough, as the resistance is too high and the stability too low compared to real human skin. Through this, the skin material breaks at too low forces when the suture is being pulled. Here, S3 recommended to look for already produced medical suturing pads. In the end, the subject stated that the task is well suited for practicing realistic situations and that it is more beneficial than virtual training programs.

**Further Suggestions** For the *first iteration* of both tasks, S1 noticed that dry suture is not realistic within the human body and recommended to use wet sutures instead. Through this, the suture is able to stick to surfaces as within the human body, which surgeons tend to use to fixate parts of the suture from time to time. For the *second iterations* of both tasks, S2 suggested to use different strengths of sutures to train the perception of forces of novice surgeons. In particular, S2 recommended to move from thicker towards thinner sutures, such as the sequence of 0, 2-0, 3-0, 4-0, and 5-0. Through this, increasing difficulty levels of the same task could be achieved. Furthermore, S2 suggested to extend the tasks by placing a scale behind or beside the tasks. By doing so, the subject also recommended to instruct novice surgeons to tie the suture as close as a specific centimeter number on the scale.

**Observations** For the *first iteration* of both tasks, we noticed during the study of S1 that both tasks lift up from the task board when the suture is being pulled. For the *second iteration* of both tasks, we screwed both tasks to the task board. Through this, the tasks did not lift up when the suture was pulled.

### 5.1.3 Additional Suggestions

Beyond the questionnaire and interview responses, the subjects also provided additional suggestions for improvements. These suggestions will be presented in the following.

**Weight of Actuators on Surgeon Handles** During the study, S1 and S3 noticed that the surgeon handles are heavier than usual and sink down when not being hold. Here, the subjects stated that this behavior is not typical and that it distracts them from performing surgical procedures.

**Stereo Endoscope with Angled Tip** During the study, S1 stated to prefer the stereo endoscope with the angled (30 degrees) tip instead of the straight (0 degrees) tip.

**Additional Training Tasks** During the study, S3 recommended to add more realistic training tasks from real operating rooms, such as scenes from real surgical procedures as tasks. Here, the subject suggested to train these realistic tasks on anatomical models with the developed AR simulator. Also, S3 recommended to enable departments to record their own procedures to create their own department-specific tasks. Here, the subject argued that every department performs procedures differently and has its own methods. As an example, the subject stated that their cardio-thoracic department uses an alternative suturing method. In this method, a cut is closed by suturing from both ends towards the middle of the cut.

**Trigger Word Satisfaction** For the chosen trigger word *SEeAR*, S1 and S3 stated to have no specific opinion about it. S2 stated always to have to think about *to sear* instead of *the seer*.

**Usage of Voice Commands** During the study, S3 stated to not like voice commands in general. First, the subject argued that voice commands never work correctly. Then, the subject explained that voice commands always need to be remembered, which requires mental effort and hence distracts from performing surgical procedures. Here, the subject state that voice commands keep the hands free but the mind busy. As a solution, the subject suggested to use shorter voice commands. As an alternative, the subject recommended to only use voice commands to activate or deactivate AR functionalities in general and to then use physical switches or buttons at the surgeon handles or pedals to trigger specific AR functionalities.

**Observations** For the *usage of the stereo endoscope* while performing a procedure, we noticed during the study of S1 that the subject started to move the endoscope intuitively a few minutes after instructing not to move it. We observed the same behavior with S2 and S3.

## 5.2 Record Functionality

For the record functionality, the following results can be reported.

### 5.2.1 Modality Preference for Recording

For recording modality preferences, we investigated *opinion about live haptic feedback* and *ranking between modalities*.

**Opinion about Live Haptic Feedback** For *live haptic feedback*, S1 stated to have no specific opinion about performing tasks with or without live haptic feedback. S2 stated not to notice a difference in performing tasks with or without live haptic feedback. Here, S2 hypothesized that visual compensation is probably sufficient for thick sutures such as 0, 2-0, and 3-0. Also, S2 noticed strong vibrations on the surgeon handles when moving the robotic instruments faster. Here, S2 stated to want to move the robotic instruments fast without being warned about it by strong vibrations on the surgeon handles. S2 further stated only to want to be warned about collisions of the robotic instruments or about strong exerted forces on tissue or suture. S3 stated that the difference in performing tasks with or without live haptic feedback was minimal. The subject explained that the provided live haptic feedback only enables to feel collisions of instruments, which is not essential to feel as it can also be seen through vision. Here, the subject further explained that the provided live haptic feedback does not enable feeling the applied pulling force at the suture, which would be important to feel as it can not be seen through vision.

**Ranking between Modalities** S1, S2, and S3 stated that to prefer only vision over combined vision and live haptic.

### 5.3 Replay Functionality

For the replay functionality, the following results can be reported.

**General Opinion about Replaying Expert Performances** For *replaying expert performances*, S1 stated it to be beneficial for robotic surgery training. S2 also stated that replaying expert performances is beneficial for robotic surgery training. Here, S2 added that this functionality might be especially useful for conveying the feeling under adrenaline circumstances. S3 also stated that replaying expert performances is beneficial for training in robotic surgery.

#### 5.3.1 Visual Overlay Satisfaction for Replaying

For visual overlay satisfaction, we investigated *opinion about direct visual overlay*.

**Opinion about Direct Visual Overlay** For the *direct visual overlay*, S1 stated to be satisfied with it when replaying expert performances. S2 stated to be satisfied with the direct visual overlay as well. Here, the subject expressed that the functionality would be more helpful when it would be possible to differentiate the intensity of pressure. S3 stated that the direct visual overlay is helpful as well. Here, the subject explained that the visual overlay allows correct manipulations also in repeated ways, which is essential for training.

#### 5.3.2 Modality Preference for Replaying

For replaying modality preferences, we investigated *opinion about adding haptic and auditory feedback*, *opinion about replayed haptic and auditory feedback*, and *ranking between modalities*.

**Opinion about Adding Haptic Feedback** For the *first iteration* of replaying, S1 stated to have no specific opinion about the possibility to add haptic feedback.

**Opinion about Replayed Haptic Feedback** For the *second iterations* of replaying, S2 stated that the replayed haptic feedback could be smoother and that it feels as if the instruments would move over a rough surface. Here, S3 stated that the replayed haptic feedback is unnecessary at this evolution level. Further, the subject stated to have not been aware of the replayed haptic feedback while recording the performance.

**Opinion about Adding Auditory Feedback** For the *first iteration* of replaying, S1 stated that adding auditory feedback such as verbal explanations could be beneficial for teaching robotic surgery to novice surgeons.

**Opinion about Replayed Auditory Feedback** For the *second iteration* of replaying, S2 stated that the replayed auditory feedback is a good idea and works well. Here, S3 stated that the replayed auditory feedback is helpful for the first and second repetition, while it might be disturbing afterward for repeated examination.

**Ranking between Modalities** For the *first iteration* of replaying, S1 stated the following preference ranking:

1. **Recorded Visual Feedback only**
2. **Recorded Visual and Haptic Feedback**

For the *second iteration* of replaying, S2 and S3 stated the following preference ranking:

1. **Recorded Visual and Auditory Feedback**
2. **Recorded Visual Feedback only**
3. **Recorded Visual, Haptic, and Auditory Feedback**
4. **Recorded Visual and Haptic Feedback**

Here, S3 added that visual feedback is more important than verbal explanations as auditory feedback.

### 5.3.3 Additional Suggestions

Beyond the questionnaire and interview responses, the subjects also provided additional suggestions for improvements. These suggestions will be presented in the following.

**Adding Replay Own Training Performances** For adding *replay own training performances*, S1 stated that it could be beneficial for novice surgeons training in robotic surgery, as repetition is essential. S2 stated that adding the functionality to replay own training performances is definitely helpful and compared it to training in sports. S3 also stated to prefer to add replay own training performances.

**Adding Replay Both Trainee and Expert Performance** S3 suggested to add *replay both trainee and expert performance together* as targeted training. The subject explained that a competitive view would help compare which steps were already performed well and which steps the expert is still doing better. Here, the subject suggested to use a vertical split screen showing the own training performance on the left side and the expert performance on the right side.

## 5.4 Train Functionality with Motion Guidance

For the train functionality with motion guidance, the following results can be reported.

**General Opinion about Training with Expert Performances** For *training with expert performances*, S1 stated it to be beneficial for robotic surgery training. S2 and S3 also stated that training with expert performances is beneficial for robotic surgery training. Here, S3 added that this functionality could be helpful for step-by-step training of complex procedures and that the possibility to provide instructions directly inside the console could be beneficial.

### 5.4.1 Visual Overlay and Motion Guidance Preference for Training

For visual overlay and motion guidance preference, we investigated *opinion about direct visual overlay as ghost tools*, *opinion about trajectories*, and *ranking between motion guidance concepts*.

**Opinion about Direct Visual Overlay as Ghost Tools** For the *first iteration* of direct visual overlay as ghost tools, S1 stated to be unsatisfied with this functionality when training with expert performances. Here, S1 expressed the opinion that a direct overlay of a replayed performance while training the same performance in parallel is confusing, as it is difficult to distinguish between own and replayed robotic instruments when both are moving simultaneously.

For the *second iteration* of direct visual overlay as ghost tools, S2 also stated to be unsatisfied with this functionality. Here, S2 expressed that the direct visual overlay is distracting from the own performance and disturbing. As a possible solution, S2 distinguished two cases. If, on the one hand, the *complete scene* of the expert recording is shown, the subject suggested to instead show the recording on a different screen or in a corner within the visual field to imitate the recording. If, on the other hand, *only parts of the scene* of the expert recording are shown, the subject suggested to only show the needle of the recording without the instruments. Here, the subject argued that only the needle is essential in performing procedures and not the instruments. In this case, the subject suggested to either show the actual needle from the recording or to preferably show a visualization of the needle as a yellow 2D or 3D element with the background.

S3 stated that the training with visual recordings of expert performances is beneficial, but the direct visual overlay as ghost tools is not helpful. Here, the subject suggested to use a vertical split screen showing the own live performance on the left side and the recorded expert performance on the right side. As an alternative, the subject suggested to use a corner view showing the own live performance in the center and the recorded expert performance in a corner. For highlighting, the subject recommended to do it similar to the VR simulators, such as only highlighting tools which should be used next.

**Opinion about Trajectories** For *trajectories*, S1 expressed the opinion that this possibility could be easier for training than ghost tools.

S2 also expressed the opinion that trajectories could be better for training than ghost tools. Here, the subject suggested rather to show the needle trajectory instead of the instrument trajectory. The subject argued that only the needle trajectory is essential in performing procedures and not the instrument trajectory and that surgeons know how to move the instrument as this is obvious. In this case, the subject suggested to show the needle trajectory together with the ghost needle.

S3 stated that trajectories could be difficult for training. Here, the subject explained that every surgeon performs a procedure differently and that even the same surgeon performs a procedure differently each time as movements are never the same. The subject further explained that also basic surgical techniques change over time. In addition, the subject also stated that situations change, and that it is only essential to have the ability to understand situations and to react to changing situations, and that it is not crucial to precisely repeat the same movement every time. For *highlighting*, the subject recommended to do it similar to the VR simulators, such as only highlighting trajectories of challenging parts. Here, the subject explained to prefer a possibility to insert trajectories into videos manually.

**Ranking between Motion Guidance Concepts** For *motion guidance concepts*, S1 stated the following preference ranking:

1. **Trajectories of Tools**
2. **Ghost Tools**

Here, S2 stated the following preference ranking:

1. **Trajectories of Needle with Ghost Needle**
2. **Ghost Needle without Instruments**

Here, S3 state the following preference ranking:

1. **Ghost Tools as Split Screen or Corner View**
2. **Trajectories**
3. **Ghost Tools**

**Observations** For the *first iteration of direct visual overlay as ghost tools*, we noticed during the study of S1 that the subject had difficulties in distinguishing the own robotic instruments from the overlaid expert ghost tools. We further noticed that the subject seemed to be overwhelmed by the direct visual overlay. Here, the visual field of the subject seemed to be too overloaded with information, such as highly visible ghost tools with an applied green filter.

For the *second iteration of direct visual overlay as ghost tools*, we noticed during the study of S2 that the subject moved the stereo endoscope slightly to create an offset between the replayed and own performance. Through this, the subject was able to see the own performance with the replayed performance slightly next to the own performance. We observed the same behavior with S3.

#### 5.4.2 Modality Preference for Training

For training modality preference, we investigated *opinion about live haptic feedback, ranking between modalities, and opinion about adding auditory feedback*.

**Opinion about Live Haptic Feedback** For *training with live haptic feedback*, S1 stated that the training functionality with direct visual overlay and uninterrupted training mode is already confusing enough with only visual feedback. Due to this, S1 expressed the preference to train without live haptic feedback. S2 stated that the live haptic feedback during training is not very helpful, as the vibrations at the surgeon handles distract from the visual perception. S3 stated that the live haptic feedback during training is unnecessary. Here, the subject explained that the vibrations of the robotic instruments are not crucial for training. The subject further pointed out that the force on the suture would be important for training robotic surgery.

**Ranking between Modalities** For *training modalities*, S1, S2, and S3 stated the following preference ranking:

1. **Recorded Visual Feedback only**
2. **Recorded Visual and Live Haptic Feedback**

**Opinion about Adding Auditory Feedback** For *adding auditory feedback*, S1 stated that adding auditory feedback such as verbal explanations to the training functionality could be beneficial for teaching robotic surgery to novice surgeons. S2 stated that adding auditory feedback such as verbal explanations to the training functionality might be fine, but that there are probably individual preferences. S3 stated that adding auditory feedback such as verbal explanations to the training functionality is helpful for the first and second repetition, while it might become disturbing afterward for repeated training.

### 5.4.3 Training Mode Preference

For training mode preference, we investigated *opinion about uninterrupted and piece-wise training mode* and *ranking between training modes*.

**Opinion about Uninterrupted Training Mode** For the *first iteration* of training modes, S1 stated to be unsatisfied with the uninterrupted training mode. Here, S1 expressed the opinion that an uninterrupted training mode forces to train performances in parallel to the replayed performance, which the subject found confusing with the direct visual overlay as described before. As a possible solution, S1 recommended to only replay one step of the recording and then to pause and allow the novice surgeon to perform the replayed step before continuing with the next step. For the *second iteration* of training modes, S2 and S3 also stated to be unsatisfied with the uninterrupted training mode.

**Opinion about Piece-wise Training Mode** For the *second iteration* of training modes, S2 stated to be satisfied with the piece-wise training mode in general. Here, S2 recommended to either use longer time segments for pieces or to allow surgeons to create segments interactively through additional voice commands such as start, repeat, and stop. S3 stated that the piece-wise training mode is helpful. To improve this training mode, the subject suggested that segments should be prolonged to take about 20-30 seconds as otherwise, the segments are too short. As an alternative



solution, the subject recommended that segments be defined per specific exercise, topic, or sub-task and not defined by time. Here, the subject recommended sub-tasks such as inserting a needle, pulling a needle or suture and tying a knot. In addition, the subject recommended to repeat one sub-task until it is mastered before moving to the next sub-task and to enable to repeat sub-tasks as often as wanted.

**Ranking between Training Modes** For the *first iteration* of training modes, S1 stated to prefer the idea of a piece-wise training mode over the uninterrupted training mode. For the *second iteration* of training modes, S2 and S3 stated to prefer the piece-wise training mode over the uninterrupted training mode.

**Observations** For the *uninterrupted training mode* of the *first iteration* of training modes, we noticed during the study of S1 that the subject had difficulties in following the replayed performance. We observed the same behavior with S2 and S3.

For the *piece-wise training mode* of the *second iteration* of training modes, we noticed during the study of S2 that the subject sometimes overused the voice command for showing the following pieces in order to have a continuous replaying of the recording, especially during simple parts of the procedure. We observed the same behavior with S3.

### 5.4.4 Additional Suggestions

Beyond the questionnaire and interview responses, the subjects also provided additional suggestions for improvements. These suggestions will be presented in the following.

**Recommended Training Processes** S3 suggested the following process for training surgical procedures:

1. **Replay Expert Performance**
2. **Train Expert Performance**
3. **Replay Both Trainee and Expert Performance**

Here, the subject explained that first, it is important to watch how to perform a procedure, then to train the procedure, and finally to be able to compare how good the own performance was. Through this, it could be possible to identify weaker parts of performances and to manually evaluate the surgical skills of the own performance.

**Ideal Training Recordings** For the *recordings* used for training, S1 suggested to only use recordings of ideal executed performances, similar to introduction or tutorial videos.

**Visualization for Forces** S2 suggested to have an optical bar showing the forces within the visual field. Here, the subject argued that this extension would be especially useful when using fine sutures, as the finer the suture, the less force is necessary to break it. Further, the subject stated that seeing force through visual feedback is better than feeling force through haptic feedback. Here, the subject argued that surgeons are used to compensate for missing haptic feedback through vision.

### 5.5 Sensors for Surgical Skill Evaluation

For the sensors for surgical skill evaluation, the following results can be reported.

#### 5.5.1 Sensor Preference

For sensor preference, we investigated *opinion about force sensor and accelerometers, ranking between sensor concepts, and additional sensors for surgical skill evaluation.*

**Opinion about Force Sensor** For the *force sensor below the task plate*, S1 stated that exerted forces on the task plate could help evaluate surgical skills, as exerted forces correlate with tissue damage. S2 stated a medium priority for exerted forces on the task plate for evaluating surgical skills, especially for suturing tasks with fine sutures. Here, the subject distinguished between two cases. If, on the one hand, *suture is connected to tissue*, exerted forces on the task plate are important. This is because pulling forces detected are from instruments pulling the suture connected to the tissue, which could result in tissue damage. If, on the other hand, *suture is not connected to tissue*, exerted forces on the task plate are not important. This is because the only pushing and pulling forces detected are from instruments pushing and pulling tissue, which mostly do not result in tissue damage. S3 stated that exerted forces on the task plate could help evaluate surgical skills.

**Opinion about Accelerometers** For the *accelerometers on the robotic instruments*, S1 stated to have no specific opinion about robotic instrument vibrations for evaluating surgical skills. S2 and S3 stated that robotic instrument vibrations are not important for evaluating surgical skills.

**Ranking between Sensor Concepts** S1, S2, and S3 preferred the force sensor below the task board over the accelerometers attached to the robotic instruments for evaluating surgical skills.

**Additional Sensors for Surgical Skill Evaluation** For *additional sensors for surgical skill evaluation*, S2 suggested to use *force sensors at the tool tips of the robotic instruments*. Here, S2 expressed that forces at the robot instrument tooltips are more important than exerted forces on the task board. The subject argued that through detecting forces at the tooltips, forces on the suture could be detected regardless of if the suture is connected to the task or not. Hence, forces at the tooltip would be a good metric for evaluating the breaking of sutures or tissue damage. S3 also suggested to use applied force on the suture for evaluating surgical skills.

### 5.5.2 Metric Suggestions

For metric suggestions, we investigated *additional metrics for surgical skill evaluation*.

**Additional Metrics for Surgical Skill Evaluation** For *additional metrics* for surgical skill evaluation, S1, S2, and S3 suggested to use similar metrics as the VR simulator *dVSS* (by Intuitive Surgical Inc.). S2 and S3 further suggested to use *time* as a metric for evaluating surgical skills.

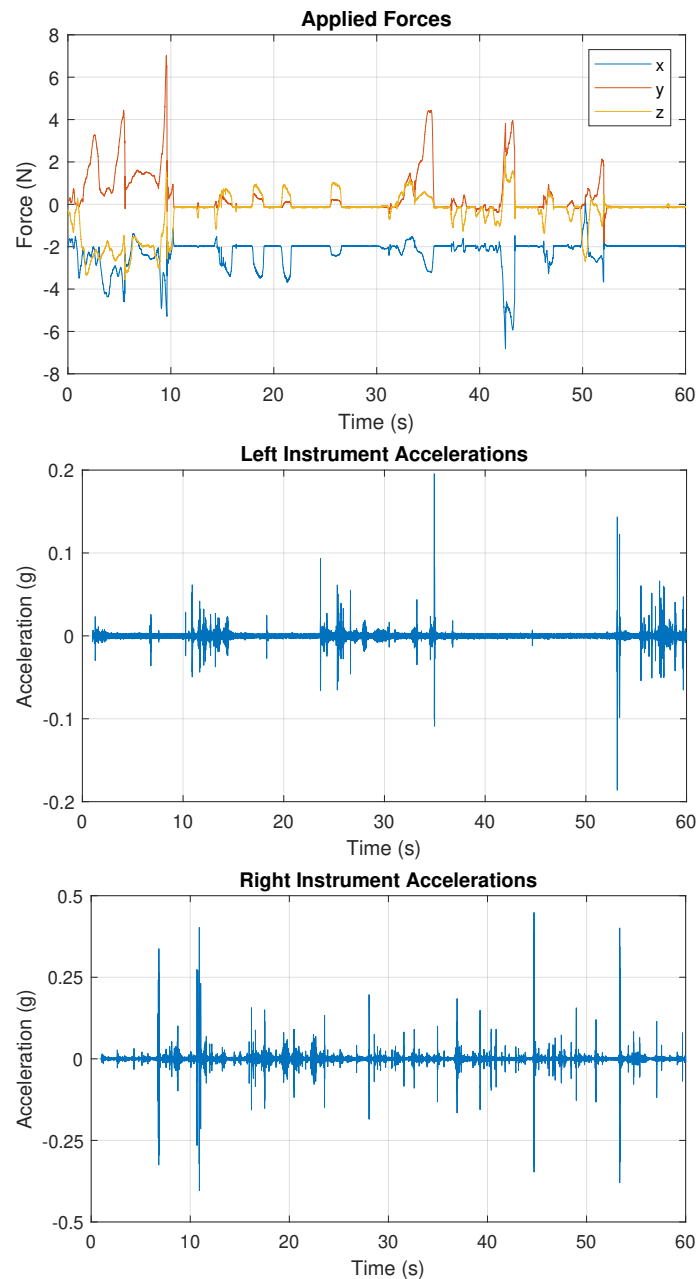
**Normalizing Metrics** S3 recommended to normalize metrics per surgeon. Here, the subject explained that metrics vary from surgeon to surgeon and also from day to day.

### 5.5.3 Measurements

For each trial of the study, we measured applied forces on the task board and robotic instrument vibrations. These measurements will be illustrated in the following.

**Force Sensor for Forces on Task Board** The force sensor measured the applied forces of subjects on the task board for each trial of the study. In the first plot of Figure 5.1, we present a time history of applied forces of S3 while recording a suturing task.

**Accelerometer for Robotic Instrument Vibrations** The accelerometers measured vibrations of left and right robotic instruments for each trial of the study. In the second and third plot of Figure 5.1, we present a time history of left and right robotic instrument vibrations of S3 while recording a suturing task.



**Figure 5.1:** Sample recorded force and acceleration data for Subject 3, for the suturing task under the record expert condition without verbal explanations and without live haptic feedback. Here, the subject pierces the needle through the tissue with the right instrument (2 s - 9 s), pulls the suture through with first the left and then the right instrument (14 s - 26 s), ties a knot by throwing a double loop over the right instrument and then pulling the suture through the double loop with the right instrument (28 s - 35 s), and finally tightens the knot two times (43 s, 52 s).

## 6 Discussion

In this chapter, the results of this work will be discussed. Also, possible directions to guide future research will be pointed out.

As stated in Section 1.3, the first aim of this study was to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. Moreover, the second aim was to explore different concepts for motion guidance and surgical skill evaluation in augmented reality.

Hence, we addressed the following RQs within the chief surgeon study:

- **RQ1:** Is multimodal record and replay in augmented reality a beneficial approach for training in robotic surgery?
- **RQ2:** How could motion guidance and surgical skill evaluation in augmented reality benefit training in robotic surgery?

In regard to the *generalizability* of the found results, it can be noted that all three participants of the human-subject study are chief surgeons with different surgical specialties (i.e., visceral, urological, cardio-thoracic). Hence, the found results can be generalized to several surgical specialties and find broader application.

However, it also has to be noted that chief surgeons are *not the actual user group* for training in robotic surgery. Chief surgeons supervise training processes, but novice surgeons are the actual user group for training in robotic surgery. Hence, the opinion of the chief surgeons is valuable, but it could differ from the opinion of novice surgeons starting to train in robotic surgery. As a consequence, the found results should be discussed with this in mind.

### 6.1 Main Findings with Implications

In this section, the main findings of the conducted chief surgeon study will be presented. Further, the implications of the main findings will be discussed with existing literature to answer the *RQs* of this work.

#### 6.1.1 For RQ1 - Multimodal Record and Replay in AR

As stated in Section 1.3.1, the first aim of this work was to use the developed platform to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. For this purpose, the following *areas of interest* were introduced with subsequent *categories of interests*:

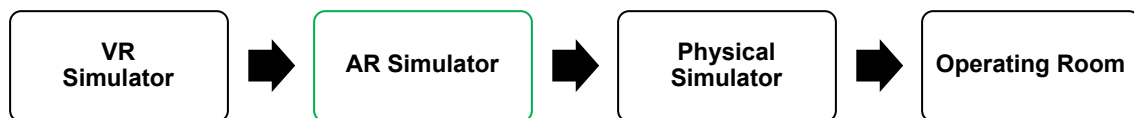
- **Integrability of AR Platform in Training Processes:** How can the developed AR platform be *integrated* in existing training processes for robotic surgery?
- **Record Functionality:** How do chief surgeons prefer to *record* performances with multiple modalities in AR for robotic surgery training?
- **Replay Functionality:** How do chief surgeons prefer to *replay* performances with multiple modalities in AR for robotic surgery training?

In the following, we will discuss the implications of our main findings with existing literature to answer the stated *RQ1* for multimodal record and replay in augmented reality.

### Integrability of AR Platform in Training Processes

Overall, we found that multimodal record and replay in augmented reality is a promising approach for training in robotic surgery. In terms of *integration possibilities in common training processes*, we found that AR simulators could close the gap between virtual and physical simulators in robotic surgery training. In terms of *training task satisfaction*, we found that expert surgeons were satisfied with the two manufactured training tasks. In terms of *stereo endoscope usage*, we found that endoscope movement is essential for performing robotic surgery. Hence, moving the endoscope should be allowed during recording and training, which impacts the choice of the visual overlay. Through this, direct visual overlays during training become impractical due to changing perspectives, and a corner view should be preferred.

For *integration possibilities in common training processes*, we found during the study that AR simulators could close the gap between virtual and physical simulators in robotic surgery training (see Figure 6.1). This result is in accordance with the envisioned robotic surgery training process with AR simulators stated in Chapter 1. As such, we believe that AR simulators should be integrated into common training processes to benefit novice surgeons in training robotic surgery.



**Figure 6.1:** Confirmed robotic surgery training process with AR.

### Record Functionality

Overall, we found that chief surgeons were satisfied with the *record functionality*. In general, we found that expert surgeons preferred to focus on their own performance during recording. In terms of *modality preference for recording*, we found that expert surgeons preferred live visual feedback and did not prefer live haptic feedback.

## Replay Functionality

Overall, we found that chief surgeons were satisfied with the *replay functionality*. In general, we found that expert surgeons preferred to focus on the recorded performance during replaying. In terms of *visual overlay satisfaction for replaying*, we found that a direct overlay is preferred as the focus is on the replayed performance. In terms of *modality preference for replaying*, we found that a combination of visual and auditory feedback is preferred and that haptic feedback based on robotic instrument vibrations is generally not preferred. Here, the visual feedback consists of previously recorded videos. The auditory feedback consists of previously recorded verbal explanations.

For *modality preference for replaying*, we found during the study that chief surgeons generally prefer a combination of visual and auditory feedback for replaying performances, whereas haptic feedback based on robotic instrument vibrations is generally not preferred. This result is not surprising, as experienced surgeons are used to compensate for missing haptic feedback by vision. This effect is also supported by the study of Reiley et al. [RAB+08], which found that visual force feedback resulted in reduced suture breakage, decreased applied forces, and more consistent forces for novice surgeons, whereas no significant difference was found for experienced robotic surgeons. As a consequence, the authors hypothesize that force feedback primarily benefits novice surgeons, which is also supported by a conducted systematic literature review of Rangarajan et al. [RDP20] finding that primarily novice surgeons in the early stages of training benefit most from haptic feedback in simulators resulting in improved training effects, fidelity, and realism. Hence, haptic feedback might provide novice surgeons with a more immersive learning experience, as it enables them to feel what expert surgeons felt while performing a surgical procedure. In addition, haptic feedback might also shorten the learning process for novice surgeons. As such, we believe that haptic feedback should be further investigated in a user study with novice surgeons.

### 6.1.2 For RQ2 - Motion Guidance and Skill Evaluation in AR

As stated in Section 1.3.2, the second aim of this work was to use the developed platform to explore different concepts for motion guidance and surgical skill evaluation in augmented reality for training in robotic surgery. For this purpose, the following *areas of interest* were introduced with subsequent *categories of interests*:

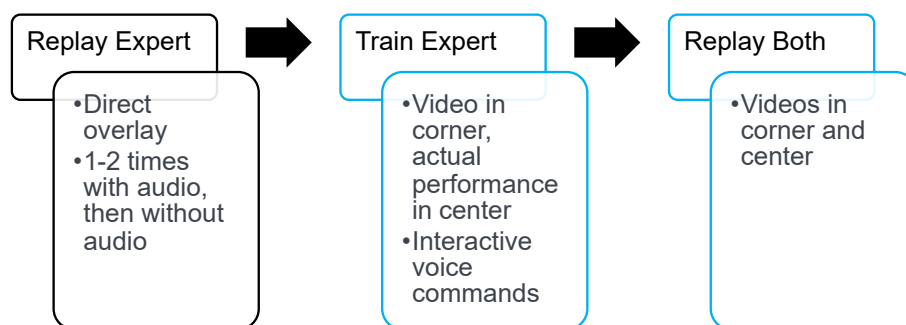
- **Train Functionality with Motion Guidance:** How do chief surgeons prefer to *train* performances with multiple modalities and motion guidance in AR for robotic surgery training?
- **Sensors for Surgical Skill Evaluation:** Which sensors can be used for surgical skill evaluation in robotic surgery training?

In the following, we will discuss the implications of our main findings with existing literature to answer the stated *RQ2* for multimodal training with motion guidance and skill evaluation in augmented reality.

### Train Functionality with Motion Guidance

Overall, we found that chief surgeons were not satisfied with the *train functionality*. In general, we found that expert surgeons preferred to focus on their own performance during training. In terms of *visual overlay and motion guidance preference for training*, we found that a direct overlay is not preferred as the focus is on the own performance. Hence, the visual feedback for training in the center of the visual field should be minimized as far as possible to allow novice surgeons to focus on their own performance. Instead, a corner view is preferred, as it allows novice surgeons to focus on the own performance in the center. We further found that a continuous direct overlay of ghost tools or highlighting of tools or trajectories is generally not preferred. Instead, highlighting of tools or trajectories is only preferred during challenging sub-tasks. In terms of *modality preference for training*, we found that a combination of visual and auditory feedback is preferred and that haptic feedback based on robotic instrument vibrations is generally not preferred. Here, the visual feedback consists of previously recorded videos. The auditory feedback consists of previously recorded verbal explanations. In terms of *training mode preference*, we found that expert surgeons preferred the piece-wise training mode and did not prefer the uninterrupted training mode. We further found that tasks should be divided into logical sub-tasks. Thereby some sub-tasks of tasks are essential and should be emphasized, whereas others are trivial. In line with this finding, we also found that an interactive piece-wise training mode with the possibility to decide when to pause, repeat, and continue sub-tasks is preferred. In terms of *recommended training processes with AR approach*, expert surgeons suggested to replay the expert performance with auditory feedback for the first and second repetition, then to train the task with interactive piece-wise training modes, and finally to replay both own and expert performances at the same time next to each other. Through this, novice surgeons could compare their performances with expert performances (see Figure 6.2).

For *visual overlay and motion guidance preference for training*, we found during the study that for replaying, a direct visual overlay was preferred by chief surgeons, whereas for training, a corner view was preferred. This result can be explained through the different focus objects, as during replaying the focus is on the replayed performance and hence a direct visual is beneficial, whereas during training the focus is on the actual performance and hence a direct visual overlay is distracting. Hence, other visual representations of videos should be explored for training to remain the focus on the actual performance, such as corner views.



**Figure 6.2:** Suggested training process for robotic surgery.



### Sensors for Surgical Skill Evaluation

In terms of *sensor preference*, we found that expert surgeons preferred exerted forces on the task board and did not prefer robotic instrument vibrations. We further found that expert surgeons were interested in exerted forces on the suture. In terms of *metrics suggestions*, we found that time could be an essential metric for evaluating surgical skills. In terms of *measurements*, we found that the measured forces on the task board and vibrations of the robotic instrument match the recorded videos and make sense in regard to performed movements.

## 6.2 Limitations and Challenges

The *current study* is limited in regard to the following points:

- **Small Sample Size:** The current study is only conducted with three subjects. More subjects are needed to generalize the found results to the population of surgeons.
- **Missing User Group:** The current study is only conducted with chief surgeons who supervise training processes in robotic surgery, and the actual user group of novice surgeons training in robotic surgery is missing. Hence, a consecutive study with novice surgeons should be conducted to generalize the results to the actual user group for training in robotic surgery.

The *current developed platform* is limited in regard to the following points:

- **Train Functionality with Direct Overlay:** The current train functionality is limited due to the direct overlay. Through this, the trainee has to move the camera exactly as the expert surgeon. Further, the trainee has to use the same task configuration as the expert surgeon. Finally, the direct overlay seems to be not suitable for training purposes, as it distracts too much from the own performance.

## 6.3 Future Work

There are several possible directions to guide future research when considering the discussed results of the conducted study, including its limitations and the identified limitations of the developed platform.

For improving the *integrability of the AR platform in training processes*, the following steps could be taken:

- **Training Tasks of Real Surgeries:** The training tasks should be extended to include more realistic procedures. This could be achieved by using existing recordings of real surgeries and training them on anatomical models.
- **Training Tasks of Surgical Specialties and Departments:** The training tasks should be extended to include more specific procedures of different surgical specialties and also of different hospitals and departments within surgical specialties. This would enable expert

surgeons to record department-specific procedures and train these specific procedures to novice surgeons of their department or hospital. For this, anatomical models could be used as well.

- **Counterweight for Actuators at Surgeon Handles:** The weight of the actuators with cables at the surgeon handles should be compensated with a counterweight, as then the surgeon handles would remain in place without sinking down when not being held by surgeons.

For enhancing the *replay functionality* for novice surgeons, the following steps could be taken:

- **Replay of Trainee Performance:** The replay functionality should be extended to allow novice surgeons to watch their own training performance.
- **Replay of Expert Performance with Optional Verbal Explanations:** The replay functionality should be extended to work with and without verbal explanations. That is because, for the first few repetitions, verbal explanations might be useful, but then verbal explanations might distract from the actual performance.
- **Replay of Both Trainee and Expert Performance with Corner View:** The replay functionality should be extended to allow novice surgeons to compare their training performance with the expert performance directly. Through this, novice surgeons would be enabled to identify weaker parts of their performance to focus their training on these parts. It would also enable novice surgeons to evaluate their surgical skills manually. This could be achieved by placing one of both performances in the center and the other performance in the upper left corner. In addition, a green filter could be added to the expert performance again to help to distinguish both performances. Further, an option could be provided to switch which performance is currently shown in the center and which in the upper left corner.

For enhancing the *train functionality with motion guidance* for novice surgeons, the following steps could be taken:

- **Piece-wise Training Mode with Sub-tasks:** The piece-wise training mode should be adapted to work with training pieces per sub-task and not with training pieces per time segment.
- **Piece-wise Training Mode with Extended Interactive Voice Commands:** The piece-wise training mode should be extended to include further interactive voice commands such as *pause* and *repeat piece* in addition to the already existing interactive voice commands such as *next piece*. Through the ability to pause the training, surgeons could decide the duration of the shown pieces independently, which would result in pieces-per-sub-task instead of pieces-per-time-duration. This ability to pause would also result in longer pieces for non-important parts and shorter pieces for important parts. Through the ability to repeat pieces, surgeons could watch complicated pieces again, which would enable a more focused training on complicated parts of procedures.
- **Train with Corner View:** The train functionality should be adapted to show the expert performance in the upper left corner instead of showing it as a direct overlay in the center. A green filter could be added to the expert performance again to help to distinguish both performances (see Figure 6.3). This extension would also benefit the generality of the training platform, as, without the direct overlay, there is also no need for exactly matching camera positions or task setups. This way, the trainee can focus on the actual performance, move the camera freely to own preferences, and practice the performance on different task setups

or anatomical models. In addition, this extension also benefits the adaptability of novice surgeons to changing situations, as it allows to train performances on different task setups and not just to repeat movements from a video on one task setup exactly.

- **Train with Optional Verbal Explanations:** The train functionality should be extended to work without and with verbal explanations. That is because, in the first few rounds, verbal explanations might be helpful, but later they might be more disturbing.
- **Train with Visual Force-Feedback:** The train functionality could be extended to include a bar visualization of the applied force of the robotic instruments to the task board, as exerted force implies tissue damage at too high levels.
- **Train with Enhanced Motion Guidance:** Enhanced motion guidance concepts consisting of highlighting the robotic instrument gripper, the needle, or the suture, as well as drawing the trajectory of the robotic instrument tooltip or the needle for important sub-tasks should be implemented similar to existing VR simulators (see Figure 6.4 and Figure 6.5).
- **Vision-based Tracking for Motion Guidance:** Vision-based tracking of the robotic instruments should be implemented to enable efficient highlighting of instrument grippers in real-time.
- **Vision-based Pose Estimation for Motion Guidance:** Vision-based pose estimation of the robotic instruments should be implemented to achieve kinematics, which can be further used for various use-cases such as recording trajectories.

For enhancing the *record functionality* for expert surgeons, the following steps could be taken:

- **Graphical User Interface for Recordings:** A graphical user interface should be developed to allow expert surgeons to modify their own task recordings (see Figure 6.7). This includes splitting the task into sub-tasks and highlighting tools or grippers for important parts (see Figure 6.6).

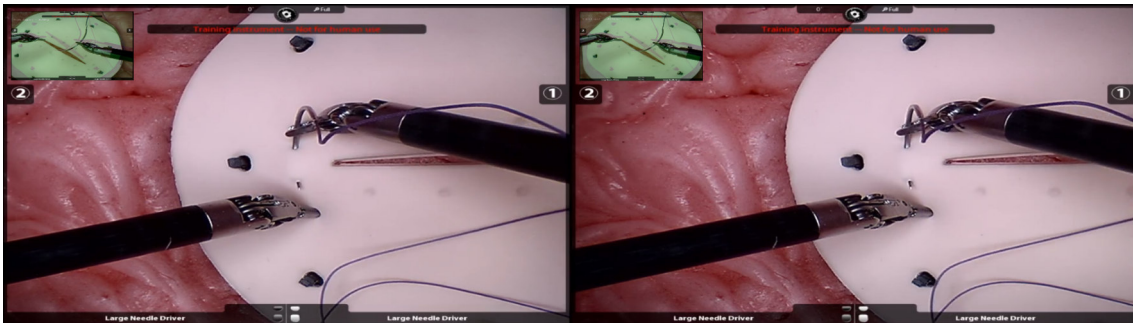
For enhancing the *sensors for surgical skill evaluation* novice and expert surgeons, the following steps could be taken:

- **Sensors for Haptic Feedback:** Additional sensors should be explored as a source of information for haptic feedback, as surgeons also pointed out their interest in forces applied to sutures. Hence, possible sensors could include force sensors at the tooltips of the robotic instruments for cases in which the suture is not connected to the task, in combination with the already integrated force sensor below the task board for cases in which the suture is connected to the task such as suturing.
- **Sensors for Surgical Skill Evaluation:** The already integrated force sensor below the task board and the suggested force sensors at the robotic instrument tooltips should be further explored to evaluate surgeon skill performance automatically.
- **Vision-based Force Estimation:** Vision-based force estimation of the robotic instruments could be implemented to attain exerted forces of the robotic instruments on tissue, which could be further used for visual or haptic feedback of applied forces independently of used tasks or anatomical models.

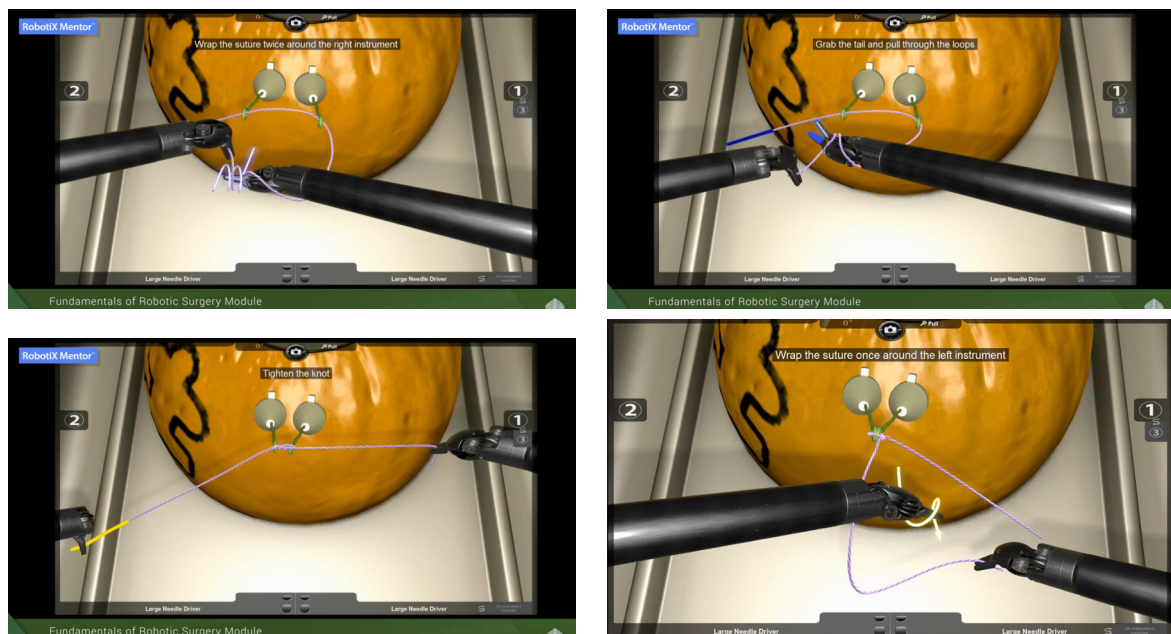
- **Machine Learning:** Different machine learning approaches could be implemented to classify phases of surgical procedures and to evaluate surgical skills based on force and accelerometer measurements and vision.

After considering and implementing the previous suggestions, the following steps should be taken:

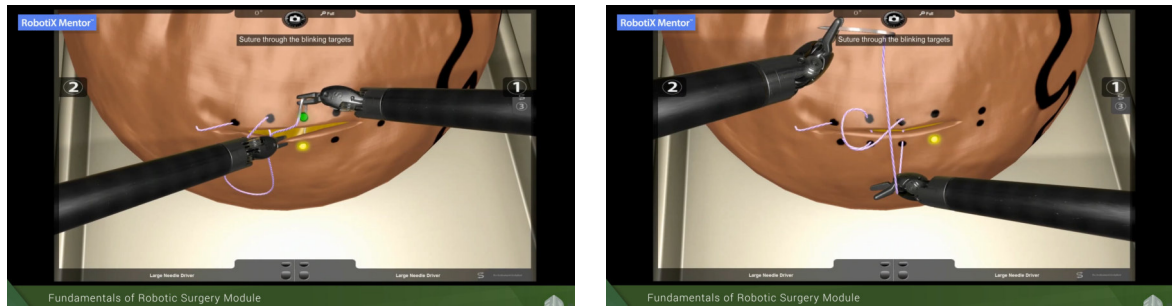
- **User Study:** The recordings of the chief surgeons should be used to conduct a user study with novice surgeons. This study should aim to assess how multimodal record and replay in AR affects the learning curve of novice surgeons. It should also aim to assess how different concepts in motion guidance and surgical skill evaluation affect their learning curves. This evaluation could be achieved by a quantitative analysis of task efficiency, which is the level of profoundness in performing a task, and learning curve, which is the slope in task efficiency over time. These dependent variables could be measured through task completion time, error rate, distance, and manual performance rating.



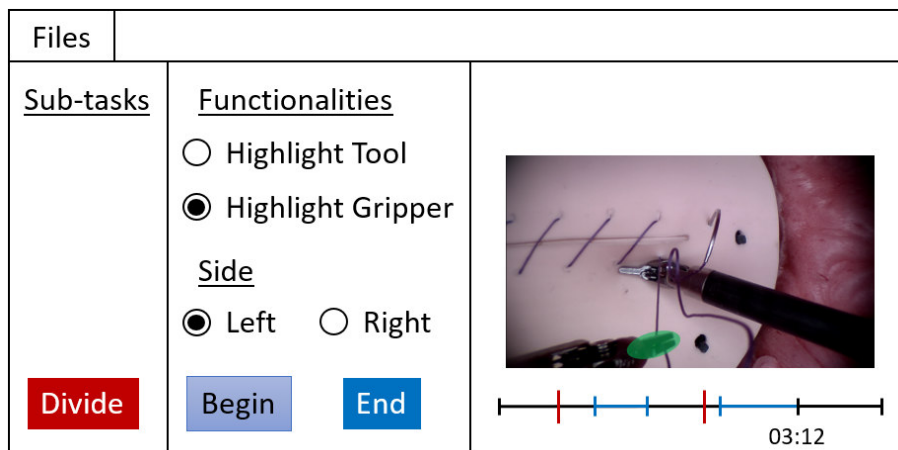
**Figure 6.3:** Envisioned left and right corner view for training functionality.



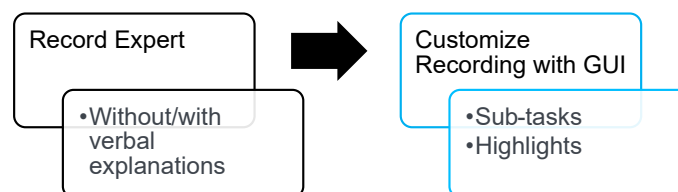
**Figure 6.4:** Motion guidance concepts from VR simulator for knot tying task. Reprinted from 3D Systems formerly Symbionix [Sima].



**Figure 6.5:** Motion guidance concepts from VR simulator for suturing task. Reprinted from 3D Systems formerly Symbionix [Sima].



**Figure 6.6:** Draft of GUI. The divide button creates separate sub-tasks. The begin and end buttons define when to highlight a tool or gripper. Also, it is possible to draw trajectories directly on top of the video.



**Figure 6.7:** Envisioned recording process with GUI.



## 7 Conclusion

First, we presented the concepts of robotic surgery training, augmented reality, multimodal feedback, and motion guidance in Chapter 1. Next, we gave an insight to the state of the art for robotic surgery training in Chapter 2. Then, we described the development of a multimodal record and replay platform in augmented reality for robotic surgery training using the Intuitive da Vinci Surgical System in Chapter 3. The developed platform allows expert surgeons to record surgical procedures with multiple modalities and novice surgeons to replay and train with the multimodal recordings, including visual, haptic, and auditory feedback. The platform also incorporates motion guidance concepts in the form of ghost tools. With this platform, we conducted an exploratory study with three chief surgeons in Chapter 4. The first aim of this study was to investigate if multimodal record and replay in augmented reality is a beneficial approach for training in robotic surgery. Moreover, the second aim was to explore different concepts for motion guidance and surgical skill evaluation in augmented reality. For this purpose, we manufactured a box trainer with two standardized tasks. Further, we placed a force sensor below the task board and used the previously attached accelerometers on the robotic instruments to explore possibilities for quantitative evaluation of surgical skill performances. Afterwards, we reported the results of the study in Chapter 5. Overall, we found that multimodal record and replay in augmented reality is a promising approach for robotic surgery training. In terms of modalities, we found that chief surgeons generally prefer a combination of visual and auditory feedback. Regarding the visual feedback, the surgeons prefer a direct overlay while replaying as the focus should be on the recorded performance. In contrast, a corner view is preferred while training as surgeons should focus on their own performance. Regarding the auditory feedback, the surgeons found verbal explanations beneficial in the early stages of training. In terms of motion guidance, our results showed that visual cues are preferred only during challenging sub-tasks of surgical procedures. In contrast, no preference was expressed between ghost tools and trajectories. Finally, we discussed the implications of the results and pointed out possible directions to guide future research in Chapter 6.





## Bibliography

- [ABB+01] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, B. MacIntyre. “Recent advances in augmented reality”. In: *IEEE computer graphics and applications* 21.6 (2001), pp. 34–47 (cit. on p. 23).
- [APHP20] M. Alshuaibi, C. Perrenot, J. Hubert, M. Perez. “Concurrent, face, content, and construct validity of the RobotiX Mentor simulator for robotic basic skills”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 16.3 (2020), e2100 (cit. on p. 26).
- [APT+19] A. Abiri, J. Pensa, A. Tao, J. Ma, Y.-Y. Juo, S. J. Askari, J. Bisley, J. Rosen, E. P. Dutson, W. S. Grundfest. “Multi-modal haptic feedback for grip force reduction in robotic surgery”. In: *Scientific reports* 9.1 (2019), pp. 1–10 (cit. on p. 38).
- [ARS+06] T. Akinbiyi, C. E. Reiley, S. Saha, D. Burschka, C. J. Hasser, D. D. Yuh, A. M. Okamura. “Dynamic augmented reality for sensory substitution in robot-assisted surgical systems”. In: *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2006, pp. 567–570 (cit. on pp. 34, 35).
- [Aut] A. I. Automation. *F/T Sensor: Mini40*. [https://www.ati-ia.com/products/ft/ft\\_models.aspx?id=Mini40](https://www.ati-ia.com/products/ft/ft_models.aspx?id=Mini40). Accessed: 2021-05-06 (cit. on p. 75).
- [Azu97] R. T. Azuma. “A survey of augmented reality”. In: *Presence: Teleoperators & Virtual Environments* 6.4 (1997), pp. 355–385 (cit. on p. 22).
- [BFCK17] J. D. Brown, J. N. Fernandez, S. P. Cohen, K. J. Kuchenbecker. “A wrist-squeezing force-feedback system for robotic surgery training”. In: *2017 IEEE World Haptics Conference (WHC)*. IEEE. 2017, pp. 107–112 (cit. on pp. 38, 40).
- [Bla] BlackMagicDesign. *DeckLink Quad 2 Technical Specifications*. <https://www.blackmagicdesign.com/products/decklink/techspecs/W-DLK-30>. Accessed: 2021-02-02 (cit. on pp. 41, 73).
- [BOK+04] B. T. Bethea, A. M. Okamura, M. Kitagawa, T. P. Fitton, S. M. Cattaneo, V. L. Gott, W. A. Baumgartner, D. D. Yuh. “Application of haptic feedback to robotic surgery”. In: *Journal of Laparoendoscopic & Advanced Surgical Techniques* 14.3 (2004), pp. 191–195 (cit. on p. 34).
- [Cam] Cambridge-MIT-Institute. *PyAudio*. <http://people.csail.mit.edu/hubert/pyaudio/>. Accessed: 2021-02-01 (cit. on p. 43).
- [CMD+20] G. Caccianiga, A. Mariani, E. De Momi, G. Cantarero, J. D. Brown. “An Evaluation of Inanimate and Virtual Reality Training for Psychomotor Skill Development in Robot-Assisted Minimally Invasive Surgery”. In: *IEEE Transactions on Medical Robotics and Bionics* 2.2 (2020), pp. 118–129 (cit. on pp. 15, 23, 24, 37).

- [CMP+21] G. Caccianiga, A. Mariani, C. G. de Paratesi, A. Menciassi, E. De Momi. “Multi-Sensory Guidance and Feedback for Simulation-Based Training in Robot Assisted Surgery: A Preliminary Comparison of Visual, Haptic, and Visuo-Haptic”. In: *IEEE Robotics and Automation Letters* 6.2 (2021), pp. 3801–3808 (cit. on pp. 34, 36, 39).
- [CMUa] CMUSpeechGroup. *Pocket Sphinx*. <https://github.com/cmuspinx/pocketsphinx>. Accessed: 2021-01-28 (cit. on p. 43).
- [CMUb] CMUSpeechGroup. *Sphinx Knowledge Base Tool*. <http://www.speech.cs.cmu.edu/tools/lmtool-new.html>. Accessed: 2021-01-28 (cit. on p. 43).
- [Con] O. S. Contributors. *Open Broadcaster Software (OBS)*. <https://obsproject.com/>. Accessed: 2021-04-28 (cit. on p. 73).
- [Del] Dell. *Aurora R7 Setup and Specifications*. [https://topics-cdn.dell.com/pdf/alienware-aurora-r7-desktop\\_specifications\\_en-us.pdf](https://topics-cdn.dell.com/pdf/alienware-aurora-r7-desktop_specifications_en-us.pdf). Accessed: 2021-02-02 (cit. on pp. 41, 73).
- [DMS+21] C. D’Ettorre, A. Mariani, A. Stilli, P. Valdastrì, A. Deguet, P. Kazanzides, R. H. Taylor, G. S. Fischer, S. P. DiMaio, A. Menciassi, et al. “Accelerating Surgical Robotics Research: Reviewing 10 Years of Research with the dVRK”. In: *arXiv preprint arXiv:2104.09869* (2021) (cit. on p. 16).
- [EWA+21] J. Ebbing, P. N. Wiklund, O. Akre, S. Carlsson, M. J. Olsson, J. Höijer, M. Heimer, J. W. Collins. “Development and validation of non-guided bladder-neck and neurovascular-bundle dissection modules of the RobotiX-Mentor® full-procedure robotic-assisted radical prostatectomy virtual reality simulation”. In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 17.2 (2021), e2195 (cit. on p. 26).
- [FK19] M.-P. Forte, K. J. Kuchenbecker. *Interactive Augmented Reality for Robot-Assisted Surgery*. Extended abstract presented as an Emerging Technology ePoster at the Annual Meeting of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES). Baltimore, Maryland, USA, Apr. 2019 (cit. on pp. 41, 42).
- [For18] M. P. Forte. “Robust Visual Augmented Reality for Robot-Assisted Surgery”. MA thesis. Politecnico Di Milano, 2018 (cit. on p. 41).
- [GGS+12] A. C. Goh, D. W. Goldfarb, J. C. Sander, B. J. Miles, B. J. Dunkin. “Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills”. In: *The Journal of urology* 187.1 (2012), pp. 247–252 (cit. on p. 24).
- [GMV+09] J. C. Gwilliam, M. Mahvash, B. Vagvolgyi, A. Vacharat, D. D. Yuh, A. M. Okamura. “Effects of haptic and graphical force feedback on teleoperated palpation”. In: *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 677–682 (cit. on pp. 35, 36).
- [GS00] G. S. Guthart, J. K. Salisbury. “The Intuitive/sup TM/telesurgery system: overview and application”. In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*. Vol. 1. IEEE, 2000, pp. 618–621 (cit. on pp. 16, 17).
- [HGVB18] A. M. Hertz, E. I. George, C. M. Vaccaro, T. C. Brand. “Head-to-head comparison of three virtual-reality robotic surgery simulators”. In: *JLS: Journal of the Society of Laparoendoscopic Surgeons* 22.1 (2018) (cit. on p. 26).

- [JSC+17] A. M. Jarc, A. A. Stanley, T. Clifford, I. S. Gill, A. J. Hung. “Proctors exploit three-dimensional ghost tools during clinical-like training scenarios: a preliminary study”. In: *World journal of urology* 35.6 (2017), pp. 957–965 (cit. on pp. 22, 34, 36, 37).
- [KGM+10] K. J. Kuchenbecker, J. Gewirtz, W. McMahan, D. Standish, P. Martin, J. Bohren, P. J. Mendoza, D. I. Lee. “VerroTouch: High-frequency acceleration feedback for telerobotic surgery”. In: *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer. 2010, pp. 189–196 (cit. on pp. 39, 43).
- [KK15] J. K. Koehn, K. J. Kuchenbecker. “Surgeons and non-surgeons prefer haptic feedback of instrument vibrations during robotic surgery”. In: *Surgical endoscopy* 29.10 (2015), pp. 2970–2983 (cit. on pp. 37, 39).
- [LGL16] V. Lahanas, E. Georgiou, C. Loukas. “Surgical simulation training systems: box trainers, virtual reality and augmented reality simulators”. In: *International Journal of Advanced Robotics and Automation* 1.2 (2016), pp. 1–9 (cit. on pp. 23–25).
- [LMK+12] J. Y. Lee, P. Mucksavage, D. C. Kerbl, V. B. Huynh, M. Etafy, E. M. McDougall. “Validation study of a virtual reality robotic simulator—role as an assessment tool?” In: *The Journal of urology* 187.3 (2012), pp. 998–1002 (cit. on pp. 15, 26).
- [Mata] MathWorks. *MATLAB*. <https://www.mathworks.com/products/matlab.html>. Accessed: 2021-05-06 (cit. on p. 73).
- [Matb] MathWorks. *MATLAB and Simulink for Signal Processing*. <https://www.mathworks.com/solutions/signal-processing.html>. Accessed: 2021-05-06 (cit. on p. 73).
- [MGC+13] W. McMahan, E. D. Gomez, L. Chen, K. Bark, J. C. Nappo, E. I. Koch, D. I. Lee, K. R. Dumon, N. N. Williams, K. J. Kuchenbecker. “A practical system for recording instrument interactions during live robotic surgery”. In: *Journal of robotic surgery* 7.4 (2013), pp. 351–358 (cit. on pp. 43, 45).
- [MGS+11] W. McMahan, J. Gewirtz, D. Standish, P. Martin, J. A. Kunkel, M. Lilavois, A. Wedmid, D. I. Lee, K. J. Kuchenbecker. “Tool contact acceleration feedback for telerobotic surgery”. In: *IEEE Transactions on Haptics* 4.3 (2011), pp. 210–220 (cit. on pp. 37–39, 43, 44).
- [MTUK95] P. Milgram, H. Takemura, A. Utsumi, F. Kishino. “Augmented reality: A class of displays on the reality-virtuality continuum”. In: *Telemanipulator and telepresence technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 282–292 (cit. on pp. 22, 23).
- [MVL+20] A. Malpani, S. S. Vedula, H. C. Lin, G. D. Hager, R. H. Taylor. “Effect of real-time virtual reality-based teaching cues on learning needle passing for robot-assisted minimally invasive surgery: a randomized controlled trial”. In: *International journal of computer assisted radiology and surgery* 15 (2020), pp. 1187–1194 (cit. on pp. 34, 37).
- [Opea] OpenCV. *About OpenCV*. <https://opencv.org/about/>. Accessed: 2021-05-06 (cit. on p. 46).
- [Opeb] OpenCV. *OpenCV Image Processing (imgproc module)*. [https://docs.opencv.org/master/d7/da8/tutorial\\_table\\_of\\_content\\_imgproc.html](https://docs.opencv.org/master/d7/da8/tutorial_table_of_content_imgproc.html). Accessed: 2021-05-06 (cit. on p. 46).

- [PBZ16] D. Pinzon, S. Byrns, B. Zheng. “Prevailing trends in haptic feedback simulation for minimally invasive surgery”. In: *Surgical innovation* 23.4 (2016), pp. 415–421 (cit. on p. 38).
- [RAB+08] C. E. Reiley, T. Akinbiyi, D. Burschka, D. C. Chang, A. M. Okamura, D. D. Yuh. “Effects of visual force feedback on robot-assisted surgical task performance”. In: *The Journal of thoracic and cardiovascular surgery* 135.1 (2008), pp. 196–202 (cit. on pp. 34, 35, 99).
- [RDP20] K. Rangarajan, H. Davis, P. H. Pucher. “Systematic review of virtual haptics in surgical simulation: a valid educational tool?” In: *Journal of surgical education* 77.2 (2020), pp. 337–347 (cit. on pp. 37, 39, 99).
- [RMT+14] P. Ramos, J. Montez, A. Tripp, C. K. Ng, I. S. Gill, A. J. Hung. “Face, content, construct and concurrent validity of dry laboratory exercises for robotic training using a global assessment tool”. In: *BJU international* 113.5 (2014), pp. 836–842 (cit. on pp. 26, 28).
- [Rob] F. of Robotic Surgery (FRS). *Final Physical Training Model*. <https://frsurgery.org/training-model/final-physical-training-model/>. Accessed: 2021-04-05 (cit. on p. 27).
- [ROSa] ROS. *About ROS*. <https://www.ros.org/about-ros/>. Accessed: 2021-05-06 (cit. on p. 45).
- [ROSc] ROS. *ROS Concepts*. <http://wiki.ros.org/ROS/Concepts>. Accessed: 2021-05-06 (cit. on p. 45).
- [ROSc] ROS. *ROS Core Components*. <https://www.ros.org/core-components/>. Accessed: 2021-05-06 (cit. on p. 45).
- [ROSc] ROS. *ROS Introduction*. <http://wiki.ros.org/ROS/Introduction>. Accessed: 2021-05-06 (cit. on p. 45).
- [SBKN17] A. N. Sridhar, T. P. Briggs, J. D. Kelly, S. Nathan. “Training in robotic surgery—an overview”. In: *Current urology reports* 18.8 (2017), pp. 1–8 (cit. on pp. 15, 19, 23–28).
- [SDK+18] A. Skinner, D. Diller, R. Kumar, J. Cannon-Bowers, R. Smith, A. Tanaka, D. Julian, R. Perez. “Development and application of a multi-modal task analysis to support intelligent tutoring of complex skills”. In: *International journal of STEM education* 5.1 (2018), pp. 1–17 (cit. on p. 28).
- [SFO+12] C. Schönauer, K. Fukushi, A. Olwal, H. Kaufmann, R. Raskar. “Multimodal motion guidance: techniques for adaptive and dynamic feedback”. In: *Proceedings of the 14th ACM international conference on Multimodal interaction*. 2012, pp. 133–140 (cit. on pp. 22, 38, 40).
- [SGG+14] N. Y. Siddiqui, M. L. Galloway, E. J. Geller, I. C. Green, H.-C. Hur, K. Langston, M. C. Pitter, M. E. Tarr, M. A. Martino. “Validity and reliability of the robotic objective structured assessment of technical skills”. In: *Obstetrics and gynecology* 123.6 (2014), p. 1193 (cit. on p. 24).

- [Sima] 3. S. formerly Symbionix. *Fundamentals of Robotic Surgery (FRS) Module*. <https://symbionix.com/simulators/robotix-mentor/robotix-library-of-modules/fundamentals-of-robotic-surgery-frs/>. Accessed: 2021-04-05 (cit. on pp. 27, 28, 38, 104, 105).
- [Simb] 3. S. formerly Symbionix. *Robotix Mentor Platform*. <https://symbionix.com/simulators/robotix-mentor/robotix-platform/>. Accessed: 2021-05-14 (cit. on p. 26).
- [Simc] M. Simulation. *dV-Trainer*. <https://mimicsimulation.com/dv-trainer/>. Accessed: 2021-05-14 (cit. on p. 26).
- [Simd] M. Simulation. *Study Further Validates Mimic's Dry Labs as Effective Method to Train Robotic Surgeons*. <https://mimicsimulation.com/study-further-validates-mimics-dry-labs-as-effective-method-to-train-robotic-surgeons/>. Accessed: 2021-05-14 (cit. on pp. 25, 27).
- [SJMT19] N. Schaffert, T. B. Janzen, K. Mattes, M. H. Thaut. "A review on the relationship between sound and movement in sports and rehabilitation". In: *Frontiers in Psychology* 10 (2019), p. 244 (cit. on p. 37).
- [SK04] J. A. Saunders, D. C. Knill. "Visual feedback control of hand movements". In: *Journal of Neuroscience* 24.13 (2004), pp. 3223–3234 (cit. on pp. 22, 34).
- [SPS14] R. Smith, V. Patel, R. Satava. "Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 10.3 (2014), pp. 379–384 (cit. on p. 27).
- [SRRW13] R. Sigrist, G. Rauter, R. Riener, P. Wolf. "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review". In: *Psychonomic bulletin & review* 20.1 (2013), pp. 21–53 (cit. on pp. 19, 21, 33, 34).
- [SSL+20] R. M. Satava, D. Stefanidis, J. S. Levy, R. Smith, J. R. Martin, S. Monfared, L. R. Tim-sina, A. W. Darzi, A. Moglia, T. C. Brand, et al. "Proving the effectiveness of the fundamentals of robotic surgery (FRS) skills curriculum: a single-blinded, multi-specialty, multi-institutional randomized control trial". In: *Annals of surgery* 272.2 (2020), pp. 384–392 (cit. on p. 27).
- [Sur] I. Surgical. *da Vinci Skills Simulator*. <https://www.europeanroboticinstitute.org/wp-content/uploads/2013/04/da-Vinci-Skills-Simulator-Brochure.pdf>. Accessed: 2021-05-14 (cit. on p. 26).
- [Sur12] I. Surgical. *da Vinci Si HD Surgical System. User Manual*. 2012 (cit. on pp. 18–20).
- [SV09] H. Schreuder, R. Verheijen. "Robotic surgery". In: *BJOG: An International Journal of Obstetrics & Gynaecology* 116.2 (2009), pp. 198–213 (cit. on pp. 15–17).
- [Sys] S. S. Systems. *ROSS II*. <http://simulatedsurgicals.com/projects/ross/>. Accessed: 2021-05-14 (cit. on p. 26).
- [TGS+16] A. Tanaka, C. Graddy, K. Simpson, M. Perez, M. Truong, R. Smith. "Robotic surgery simulation validity and usability comparative analysis". In: *Surgical endoscopy* 30.9 (2016), pp. 3720–3729 (cit. on p. 26).

- [TYB+15] R. Tang, X.-D. Yang, S. Bateman, J. Jorge, A. Tang. “Physio@ Home: Exploring visual guidance and feedback techniques for physiotherapy exercises”. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2015, pp. 4123–4132 (cit. on p. 22).
- [WAR+16] G. Whittaker, A. Aydin, N. Raison, F. Kum, B. Challacombe, M. S. Khan, P. Dasgupta, K. Ahmed. “Validation of the RobotiX mentor robotic surgery simulator”. In: *Journal of endourology* 30.3 (2016), pp. 338–346 (cit. on p. 26).
- [WGJD08] E. P. Westebring–van der Putten, R. H. Goossens, J. J. Jakimowicz, J. Dankelman. “Haptics in minimally invasive surgery—a review”. In: *Minimally Invasive Therapy & Allied Technologies* 17.1 (2008), pp. 3–16 (cit. on pp. 16, 17, 21).
- [WGN+16] C. R. Wottawa, B. Genovese, B. N. Nowroozi, S. D. Hart, J. W. Bisley, W. S. Grundfest, E. P. Dutton. “Evaluating tactile feedback in robotic surgery for potential clinical application using an animal model”. In: *Surgical endoscopy* 30.8 (2016), pp. 3198–3209 (cit. on pp. 38, 40).
- [WHS02] C. R. Wagner, R. D. Howe, N. Stylopoulos. “The role of force feedback in surgery: analysis of blunt dissection”. In: *Haptic Interfaces for Virtual Environment and Teleoperator Systems, International Symposium on*. Citeseer. 2002, pp. 73–73 (cit. on p. 39).
- [YAM+20] X. Yu, K. Angerbauer, P. Mohr, D. Kalkofen, M. Sedlmair. “Perspective Matters: Design Implications for Motion Guidance in Mixed Reality”. In: *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE. 2020, pp. 577–587 (cit. on pp. 22, 36).
- [ZLH+19] B. Zhao, J. Lam, H. M. Hollandsworth, A. M. Lee, N. E. Lopez, B. Abbadessa, S. Eisenstein, B. C. Cosman, S. L. Ramamoorthy, L. A. Parry. “General surgery training in the era of robotic surgery: a qualitative analysis of perceptions from resident and attending surgeons”. In: *Surgical endoscopy* (2019), pp. 1–10 (cit. on pp. 25, 28).

All links were last followed on May 20, 2021.

# A Appendix

## A.1 Informed Consent Document

Informed Consent Form		Einwilligungserklärung	
<b>Title of Research Project:</b>	<b>Multimodal Record and Replay: AR Training for Robotic Surgery</b>	<b>Titel des Forschungs-Projektes:</b>	<b>Multimodales Record-and-Replay: AR-Training für roboterassistierte Chirurgie</b>
<b>HI Protocol Number:</b>	F015A	<b>HI Protokoll-Nummer:</b>	F015A
<b>Principal Investigator:</b>	<b>Katherine J. Kuchenbecker</b> Director, Haptic Intelligence Department Max Planck Institute for Intelligent Systems Heisenbergstraße 3 70569 Stuttgart, Germany +49 (711) 689-3510 kjk@is.mpg.de	<b>Haupt-Verantwortliche:</b>	<b>Katherine J. Kuchenbecker</b> Direktorin, Abteilung Haptische Intelligenz Max-Planck-Institut für Intelligente Systeme Heisenbergstraße 3 70569 Stuttgart, Deutschland +49 (711) 689-3510 kjk@is.mpg.de
<b>Investigators:</b>	<b>Fabian Krauthausen</b> Master's Thesis Student, Haptic Intelligence Department Max Planck Institute for Intelligent Systems Heisenbergstraße 3 70569 Stuttgart, Germany krauthausen@is.mpg.de  <b>Maria-Paola Forte</b> PhD Student, Haptic Intelligence Department Max Planck Institute for Intelligent Systems Heisenbergstraße 3 70569 Stuttgart, Germany +49 (711) 689-2002 forte@is.mpg.de	<b>Verantwortliche/r:</b>	<b>Fabian Krauthausen</b> Masterarbeits-Student, Abteilung Haptische Intelligenz Max-Planck-Institut für Intelligente Systeme Heisenbergstraße 3 70569 Stuttgart, Deutschland krauthausen@is.mpg.de  <b>Maria-Paola Forte</b> Doktorandin, Abteilung Haptische Intelligenz Max-Planck-Institut für Intelligente Systeme Heisenbergstraße 3 70569 Stuttgart, Deutschland +49 (711) 689-2002 forte@is.mpg.de
<p>This informed consent form is required for you to take part in a human-subject experiment that is part of the research project named above. This research aims to determine whether recording and replaying multimodal videos in augmented reality (AR) can improve training for robot-assisted surgery. Personal data will be collected during the study; however, this information will be stored under a random subject number such as S1 or S2, and it will be kept in a way that makes identification very unlikely (see the description below). By signing this form, you allow us to process your stored data by law according to EU General Data Protection Regulation (DS-GVO), article 6, paragraph 1, lit. a.</p>		<p>Diese Einwilligungserklärung ist erforderlich, damit Sie an einem Experiment mit Menschen teilnehmen können. Das Experiment ist Teil des oben genannten Forschungsprojektes. Ziel dieser Forschung ist es, herauszufinden, ob Record-and-Replay von multimodalen Videos in Augmented Reality (AR) das Training für roboterassistierte Chirurgie verbessern könnte. Im Rahmen der Durchführung der Studie werden personenbezogene Daten erfasst. Diese werden jedoch unter einer zufälligen Teilnehmernummer (z.B. S1 oder S2) vergeben und so aufbewahrt, dass eine Identifizierung sehr unwahrscheinlich ist (Beschreibung siehe unten). Mit Ihrer Unterschrift geben Sie uns die Erlaubnis, Ihre gespeicherten Daten gemäß der EU-Datenschutzverordnung (DS-GVO), Artikel 6 Absatz 1 lit a, zu verarbeiten.</p>	



<p>Your participation is voluntary, which means you can choose whether to participate or not to participate after reading the information provided in this document. This informed consent form describes the study purpose, the possible risks of participation, the financial reward (if any), the tasks that you will undertake if you decide to participate, and the data we will collect. If you are not sure about this experiment, you should ask the principal investigator or an investigator for additional information. You can withdraw from the experiment and revoke your consent at any time without any consequences, even after the end of the experiment. To do so, please contact Fabian Krauthausen, whose contact information is listed on the first page of this document.</p> <p>If any part of this document is unclear to you, please ask the investigator to explain what you do not understand before signing this form. If you do decide to participate in this research, please sign this form at the indicated location on the last page. The investigator will provide you with your own copy that you should keep for future reference.</p>	<p>Ihre Teilnahme ist freiwillig, d.h. Sie können nach dem Lesen der in diesem Dokument enthaltenen Informationen entscheiden, ob Sie teilnehmen möchten oder nicht. Diese Einwilligungserklärung beschreibt den Zweck der Studie, die möglichen Risiken der Teilnahme, die finanzielle Vergütung (gegebenenfalls), die Aufgaben, die Sie übernehmen werden, wenn Sie sich für die Teilnahme entscheiden, und die Daten, die wir sammeln werden. Wenn Sie unsicher sind, sollten Sie die Hauptverantwortliche oder eine/n Verantwortliche/n um zusätzliche Informationen bitten.</p> <p>Sie können jederzeit ohne Konsequenzen von der Teilnahme am Experiment zurücktreten und Ihre Einwilligung widerrufen – auch nach Beendigung des Experiments. Hierzu wenden Sie sich bitte an Fabian Krauthausen. Sie finden seine Kontaktdaten auf der ersten Seite dieses Dokumentes.</p> <p>Wenn Ihnen ein Teil dieses Dokuments unklar ist, bitten Sie die/den Verantwortliche/n bitte, Ihnen vor der Unterzeichnung dieses Formulars zu erklären, was Sie nicht verstehen. Wenn Sie sich für die Teilnahme an dieser Forschung entscheiden, unterschreiben Sie bitte dieses Formular an der angegebenen Stelle auf der letzten Seite. Der/die Verantwortliche/n stellt Ihnen Ihre eigene Kopie zur Verfügung. Bitte bewahren Sie diese Kopie auf, damit Sie eventuell später nachschlagen können.</p>
<p><b>What is the purpose of this study?</b></p> <p>This study is designed to investigate whether multimodal record and replay of task videos and associated haptic feedback in AR could benefit trainees in robot-assisted surgery. In this type of surgery, the surgeon controls the movements of the surgical instruments by moving hand controllers and viewing the surgical scene through a 3D viewer. In particular, we investigate the role of vision, audio, and haptics. The results of this study will help us both understand the value of the developed system's features and define other possible features for training in robot-assisted surgery.</p>	<p><b>Was ist der Zweck dieser Studie?</b></p> <p>Diese Studie soll untersuchen, ob das multimodale Record-and-Replay von Aufgaben-Videos und assoziiertem haptischen Feedback in AR für angehende Chirurgen in der roboterassistierten Chirurgie von Vorteil sein könnte. Bei dieser Art der Chirurgie steuert der Chirurg die Bewegungen der chirurgischen Instrumente durch die Bewegung von Handreglern und betrachtet die Operationsszene durch einen 3D-Viewer. Insbesondere untersuchen wir die Rolle von Vision, Audio und Haptik. Die Ergebnisse dieser Studie werden uns helfen, sowohl den Wert der Funktionen des entwickelten Systems zu verstehen als auch weitere mögliche Funktionen für das Training in roboterassistierter Chirurgie zu definieren.</p>
<p><b>What are the requirements to participate in this study?</b></p> <p>To participate in this study, you must be 18 years or older, speak and understand either English or German well (B2 level), have full use of both hands and arms, and have no current visual and hearing disabilities.</p> <p>You may be a lay person, a surgical trainee, or a surgeon. You are not required to have previous experience with robot-assisted surgery.</p>	<p><b>Was sind die Voraussetzungen für die Teilnahme an dieser Studie?</b></p> <p>Um an dieser Studie teilnehmen zu können, müssen Sie mindestens 18 Jahre alt sein, gut Englisch oder Deutsch sprechen und verstehen können (B2-Niveau), beide Hände und Arme vollständig benutzen können und keine aktuellen Seh- und Hörbehinderungen haben.</p> <p>Sie können entweder Laie, angehender Chirurg oder Chirurg sein. Sie müssen keine Vorkenntnisse in der roboterassistierten Chirurgie haben.</p>

<p><b>How long does it take to complete this study?</b> The study consists of several sub-experiments of different lengths. Thus, its length can vary between 45 minutes and 120 minutes.</p>	<p><b>Wie lange dauert die Teilnahme an dieser Studie?</b> Die Studie besteht aus mehreren Teilversuchen unterschiedlicher Länge. So kann die Dauer zwischen 45 Minuten und 120 Minuten variieren.</p>
<p><b>How many other people will be in this study?</b> There will be up to 40 participants in the study, depending on the outcome of the experimental sessions.</p>	<p><b>Wie viele andere Personen werden an dieser Studie teilnehmen?</b> Je nach Ergebnis der experimentellen Sitzungen werden bis zu 40 Teilnehmer an der Studie teilnehmen.</p>
<p><b>Where will the study take place?</b> You will be asked to come to room 5N07 in the Max Planck Institute for Intelligent Systems located at Heisenbergstraße 3 in Stuttgart, Germany. The time and date of your participation will be determined through communication with an investigator.</p>	<p><b>Wo findet die Studie statt?</b> Sie werden gebeten, in den Raum 5N07 des Max-Planck-Instituts für Intelligente Systeme in der Heisenbergstraße 3 in Stuttgart, Deutschland, zu kommen. Die genaue Uhrzeit und das Datum Ihrer Teilnahme werden in Absprache mit einer/einem Verantwortlichen vereinbart.</p>
<p><b>What is my task?</b> First, you will be asked to fill out a short questionnaire documenting your age, gender, handedness, experience with robotic surgical systems, and surgical experience. This information tells us how the pool of participants compares to the general population. After you finish the demographic questionnaire, you will receive a brief introduction to the experimental task.</p> <p>In each trial of this study, you will be asked to control the da Vinci robot to perform a particular task on inanimate task materials, such as suturing or tying a knot.</p> <p>You may be able to issue voice commands to perform particular functions. Your view of the task materials may also be augmented with virtual content such as images, text, or 2D/3D elements; the virtual content will change according to the task and the system feature that is under analysis. You may hear recorded audio. Further, you may also feel real-time or recorded vibrations of the robotic instruments in the hand controllers.</p> <p>During the study, the experimenter will instruct you how to perform each specific task. Please ask any questions that you have at any point.</p> <p>We may record your voice, the vibrations of the robotic instruments, the force you apply with the robotic instruments, and the left and right da Vinci camera views while you do each task.</p>	<p><b>Was ist meine Aufgabe?</b> Zuerst werden Sie gebeten, einen kurzen Fragebogen auszufüllen, der Ihr Alter, Geschlecht, Händigkeit, Erfahrungen mit roboterassistierten chirurgischen Systemen und chirurgische Erfahrungen dokumentiert. Diese Informationen zeigen uns, wie der Teilnehmerkreis im Vergleich zur allgemeinen Bevölkerung aussieht. Nach Abschluss des demografischen Fragebogens erhalten Sie eine kurze Einführung in die experimentelle Aufgabe.</p> <p>In jeder Versuchsbedingung dieser Studie werden Sie aufgefordert, den da Vinci-Roboter zu bedienen, um eine bestimmte Aufgabe an nicht-lebendigen Arbeitsmaterialien auszuführen, wie z.B. chirurgisches Nähen oder Binden eines Knotens.</p> <p>Sie können möglicherweise Sprachbefehle erteilen, um bestimmte Funktionen auszuführen. Ihre Sicht auf die Arbeitsmaterialien kann möglicherweise auch durch virtuelle Inhalte wie Bilder, Texte, oder 2D/3D-Elemente erweitert werden; die virtuellen Inhalte werden sich je nach Aufgabe und Systemfunktion, die gerade analysiert wird, verändern. Sie können möglicherweise Audio-Aufnahmen hören. Außerdem können Sie möglicherweise Echtzeit- oder aufgenommene Vibrationen der Roboterinstrumente in den Handreglern spüren.</p> <p>Während der Studie wird die Person, die den Versuch durchführt, Sie anweisen, wie Sie die einzelnen Aufgaben ausführen sollen. Bitte stellen Sie zu jedem Zeitpunkt jede Frage, die Sie haben.</p> <p>Wir werden möglicherweise Ihre Stimme, die Vibrationen der Roboterinstrumente, die Kraft, die Sie mit den Roboterinstrumenten ausüben, und die Ansichten der</p>

<p>You will periodically be asked to answer questions about your experiences in this study. You will do your evaluations by filling out a questionnaire displayed on a computer screen, presented on paper, or presented auditorily. The experimenter will always be physically present to answer any questions and ensure that the study proceeds safely and as expected.</p>	<p>linken und rechten da Vinci-Kamera aufzeichnen, während Sie die einzelnen Aufgaben ausführen.</p> <p>Sie werden regelmäßig dazu aufgefordert, Fragen bezüglich Ihrer Erfahrungen in dieser Studie zu beantworten. Sie führen Ihre Auswertungen durch, indem Sie einen Fragebogen ausfüllen, der auf einem Computerbildschirm angezeigt, auf Papier präsentiert oder akustisch präsentiert wird. Die Person, die den Versuch durchführt, wird immer physisch anwesend sein, um alle Fragen zu beantworten und sicherzustellen, dass die Studie sicher und wie erwartet verläuft.</p>
<p><b>What are the risks from this experiment?</b> The risks associated with participation in this study are not greater than those encountered when using a clinical da Vinci Surgical System, which is FDA- and CE-approved. We try our best to avoid any problems that could arise for participants from this research.</p> <p>The primary risk is mental fatigue: you might become mentally tired from responding to the questions and from performing the different tasks in this study. We try to mitigate the fatigue caused by the experiment by providing a break at the end of each task. You are also welcome to ask for a break at any time during the experiment.</p> <p>While you do the experiment, the investigator will monitor everything to make sure the devices function safely and smoothly. If something malfunctions, the investigator will turn off the system and abort the experiment.</p>	<p><b>Welche Risiken birgt dieses Experiment?</b> Die mit der Teilnahme an dieser Studie verbundenen Risiken sind nicht größer als die Risiken bei der Verwendung eines klinischen da Vinci-Chirurgie Systems, das FDA- und CE-zugelassen ist. Wir versuchen unser Bestes, um jegliche Probleme zu vermeiden, die für die Teilnehmer aus dieser Forschung entstehen könnten.</p> <p>Das Hauptrisiko ist geistige Ermüdung: Sie können geistig müde werden, wenn Sie auf die Fragen antworten und die verschiedenen Aufgaben in dieser Studie ausführen. Wir versuchen, die durch das Experiment verursachte Müdigkeit zu mildern, indem wir am Ende jeder Aufgabe eine Pause einlegen. Gerne können Sie auch während des Experiments um eine Pause bitten.</p> <p>Während Sie das Experiment durchführen, wird die Verantwortliche alles überwachen, um sicherzustellen, dass die Geräte sicher und reibungslos funktionieren. Wenn etwas nicht funktioniert, schaltet die Verantwortliche das System aus und bricht das Experiment ab.</p>
<p><b>Can I leave the study before it ends?</b> The study may be stopped without your consent for the following reasons:</p> <ul style="list-style-type: none"> <li>• The investigator or principal investigator feels it is best for your safety and/or health.</li> <li>• The experimental equipment is not functioning as expected.</li> <li>• You have not followed the study instructions.</li> </ul> <p>You have the right to drop out of the research study at any time; simply contact Fabian Krauthausen using the information provided on the front page of this document.</p>	<p><b>Kann ich die Studie vorzeitig beenden?</b> Die Studie kann ohne Ihre Zustimmung aus den folgenden Gründen abgebrochen werden:</p> <ul style="list-style-type: none"> <li>• Die/der Verantwortliche oder die Hauptverantwortliche ist der Ansicht, dass es das Beste für Ihre Sicherheit und/oder Gesundheit ist.</li> <li>• Die Versuchsausrüstung funktioniert nicht wie erwartet.</li> <li>• Sie haben die Studienanweisungen nicht befolgt.</li> </ul> <p>Sie haben das Recht, die Forschungsstudie jederzeit abzubrechen; teilen Sie dies einfach Fabian Krauthausen mit, dessen Kontaktdaten Sie auf der ersten Seite dieses Dokumentes finden.</p>
<p><b>How will confidentiality be maintained and my privacy be protected?</b> The research team will make every effort to keep all the information we record during the study</p>	<p><b>Wie wird die Vertraulichkeit gewahrt und meine Privatsphäre geschützt?</b> Das Forschungsteam wird alles dafür tun, um alle Informationen, die wir während der Studie erfassen, streng</p>

<p>strictly confidential, as required by law. Any documents you sign, where you can be identified by name, will be kept in a locked archive in Dr. Kuchenbecker's department. These documents will be kept confidential. All signed documents will be destroyed when the study is over. The surveys you fill out and the electronic data we record will be marked with a unique subject identification number, but we will not keep any data that will connect this number with your name. The surveys and electronic data will be destroyed when the investigators decide that they are no longer needed for active research projects.</p> <p>The data recorded during the sessions (your name, demographic information, answers to survey questions, voice, images, videos acquired by the stereo-endoscope, robotic instrument vibrations, and robotic instrument contact forces) will be stored on secure computers and servers of the Max Planck Institute for Intelligent Systems. To protect your privacy, your name will be stored only on this paper consent form. The rest of the data will be stored by your unique subject identification number, and there will be no occurrence of your name in the digitally stored data. Unauthorized people will not have access to the stored data.</p>	<p>vertraulich zu behandeln, wie es das Gesetz vorschreibt. Jegliche von Ihnen unterschriebenen Dokumente, bei denen Sie namentlich identifiziert werden können, werden in einem verschlossenen Archiv in der Abteilung von Dr. Kuchenbecker aufbewahrt. Diese Dokumente werden vertraulich behandelt. Alle unterschriebenen Dokumente werden nach Abschluss der Studie vernichtet. Die von Ihnen ausgefüllten Fragebögen und die von uns erfassten elektronischen Daten werden mit einer eindeutigen Teilnehmer-Identifikationsnummer gekennzeichnet, aber wir speichern keine Daten, die diese Nummer mit Ihrem Namen verbinden. Die Erhebungen und elektronischen Daten werden vernichtet, wenn die Verantwortlichen entscheiden, dass sie für aktive Forschungsprojekte nicht mehr benötigt werden.</p> <p>Die während der Sitzungen aufgezeichneten Daten (Ihr Name, demografische Informationen, Antworten auf die Fragen der Fragebögen, Stimme, Bilder, Videos erworben durch das Stereo-Endoskop, Vibrationen der Roboterinstrumente, und Kontaktkräfte der Roboterinstrumente werden auf sicheren Computern und Servern des Max-Planck-Instituts für Intelligente Systeme gespeichert. Um Ihre Privatsphäre zu schützen, wird Ihr Name nur auf dieser Einverständniserklärung gespeichert. Der Rest der Daten wird unter Ihrer eindeutigen Teilnehmer-Identifikationsnummer gespeichert, und Ihr Name wird nicht in den digital gespeicherten Daten vorkommen. Unbefugte Personen haben keinen Zugang zu den gespeicherten Daten.</p>
<p><b>What will I receive after the experiment?</b> Participants who are not employed by the Max Planck Society will be compensated 8 euros per hour for their participation.</p>	<p><b>Was erhalte ich nach dem Experiment?</b> Teilnehmer, die nicht bei der Max-Planck-Gesellschaft angestellt sind, erhalten für ihre Teilnahme eine Vergütung von 8 Euro pro Stunde.</p>
<p><b>Your Rights</b> You have the right to obtain information regarding your stored data, the right to correct inaccurate data, and the right to demand the deletion of data in cases of inadmissible data storage and portability within the scope of the legal possibilities. You also have the right to submit an objection to the supervisory authority, which is the Bavarian State Office for Data Protection Supervision, Postfach 606, 91511 Ansbach, Germany.</p>	<p><b>Ihre Rechte</b> Sie haben im Rahmen der gesetzlichen Möglichkeiten ein Recht auf Auskunft über die bei uns gespeicherten Daten, ein Recht auf Berichtigung von fehlerhaften Daten sowie das Recht, die Löschung von Daten im Falle einer unzulässigen Datenspeicherung und Portabilität zu verlangen. Außerdem steht Ihnen das Recht zu, sich an die Aufsichtsbehörde zu wenden: Die zuständige Behörde ist das Bayerische Landesamt für Datenschutzaufsicht, Postfach 606, 91511 Ansbach, Deutschland.</p>

Printed Name

Name in Druckbuchstaben \_\_\_\_\_

Signature

Unterschrift \_\_\_\_\_

Place, Date

Ort, Datum \_\_\_\_\_

I give my consent for images, video recordings, audio recordings, and haptic recordings of my participation in the study to be shown in the following ways. *Please write **your initials** next to the uses that you agree to, if any:*

Ich bin damit einverstanden, dass Bilder, Video-, Audio- und Haptik-Aufzeichnungen von meiner Teilnahme an der Studie auf folgende Weise gezeigt werden dürfen. *Bitte schreiben Sie **Ihre Initialen** neben die Verwendungszwecke, denen Sie zustimmen (falls Sie zustimmen):*

\_\_\_\_\_ As part of talks at research conferences and other scientific venues.

Im Rahmen von Vorträgen auf Forschungskonferenzen und anderen wissenschaftlichen Veranstaltungen.

\_\_\_\_\_ Publicly, as images associated with this research (e.g., an image in a published paper).

Öffentlich, als Bilder, die mit dieser Forschung verbunden sind (z.B. ein Bild in einer veröffentlichten Arbeit).

\_\_\_\_\_ Publicly, as part of videos associated with this research (e.g., a video associated with a published paper, a YouTube clip from our lab, a feature prepared by journalists).

Öffentlich, als Teil von Videos, die mit dieser Forschung verbunden sind (z.B. ein Video, das mit einer veröffentlichten Arbeit verbunden ist, ein YouTube-Clip aus unserem Labor, ein von Journalisten vorbereiteter Beitrag).

Only for surgeons with more than three years of experience in robot-assisted surgery. I give my consent for images, video recordings, audio recordings, and haptic recordings of my experiment to be shown in the following ways. *Please write **your initials** next to the uses that you agree to, if any:*

Nur für Chirurgen mit mehr als drei Jahren Erfahrung in der roboterassistierten Chirurgie. Ich bin damit einverstanden, dass Bilder, Video-, Audio- und Haptik-Aufzeichnungen von meiner Teilnahme an der Studie auf folgende Weise gezeigt werden dürfen. *Bitte schreiben Sie **Ihre Initialen** neben die Verwendungszwecke, denen Sie zustimmen (falls Sie zustimmen):*

\_\_\_\_\_ To other individuals taking part in this study.

Anderen Personen, die an dieser Studie teilnehmen.

\_\_\_\_\_ To other people who want to test this system in our laboratory.

Anderen Personen, die dieses System in unserem Labor testen wollen.

## A.2 Demographic Questionnaire

### Demographic Questionnaire

Subject number					
Gender	<input type="checkbox"/> Male	<input type="checkbox"/> Female		<input type="checkbox"/> Do not wish to answer	
Handedness	<input type="checkbox"/> Left	<input type="checkbox"/> Right		<input type="checkbox"/> Ambidextrous	
Age	<input type="checkbox"/> ≤ 29	<input type="checkbox"/> 30 - 39	<input type="checkbox"/> 40 - 49	<input type="checkbox"/> 50 - 59	<input type="checkbox"/> ≥ 60
Surgical specialty, if applicable					
Years of clinical experience in <b>open surgery</b>					
Years of teaching experience in <b>open surgery</b>					
Approximate number of cases performed in the last year in <b>open surgery</b>					
Years of clinical experience in <b>non-robotic laparoscopic surgery</b>					
Years of teaching experience in <b>non-robotic laparoscopic surgery</b>					
Approximate number of cases performed in the last year in <b>non-robotic laparoscopic surgery</b>					
Years of clinical experience in <b>robotic surgery</b>					
Years of teaching experience in <b>robotic surgery</b>					
Approximate number of cases performed in the last year in <b>robotic surgery</b>					

## Demografischer Fragebogen

Probandennummer					
Geschlecht	<input type="checkbox"/> Männlich	<input type="checkbox"/> Weiblich	<input type="checkbox"/> Keine Angabe		
Händigkeit	<input type="checkbox"/> Links	<input type="checkbox"/> Rechts	<input type="checkbox"/> Beidhändig		
Alter	<input type="checkbox"/> ≤ 29	<input type="checkbox"/> 30 - 39	<input type="checkbox"/> 40 - 49	<input type="checkbox"/> 50 - 59	<input type="checkbox"/> ≥ 60
Chirurgisches Fachgebiet, falls zutreffend					
Jahre klinischer Erfahrung in <b>offener Chirurgie</b>					
Jahre Lehrerfahrung in <b>offener Chirurgie</b>					
Ungefähre Anzahl an durchgeführten Eingriffen im letzten Jahr in <b>offener Chirurgie</b>					
Jahre klinischer Erfahrung in <b>laparoskopischer Chirurgie ohne Roboter</b>					
Jahre Lehrerfahrung in <b>laparoskopischer Chirurgie ohne Roboter</b>					
Ungefähre Anzahl an durchgeführten Eingriffen im letzten Jahr in <b>laparoskopischer Chirurgie ohne Roboter</b>					
Jahre klinischer Erfahrung in <b>Roboterchirurgie</b>					
Jahre Lehrerfahrung in <b>Roboterchirurgie</b>					
Ungefähre Anzahl an durchgeführten Eingriffen im letzten Jahr in <b>Roboterchirurgie</b>					

## A.3 Interview Structure

### Procedure

1. Warm up:
  - Let the surgeon **configure da Vinci console** as desired, including finger loops, arm rest height, and display height.
  - Explain that we will sometimes **activate the haptic actuators** that are attached to the da Vinci handles, so that they can feel the vibrations of the instruments. We configured the wires to minimize pulling, but they can still feel the weight of the actuators.
  - Explain that we want them to **keep the camera in a constant location** for all trials, as the visual feedback and the analysis we perform require a consistent viewpoint. If they accidentally move the camera, we should stop the trial, move the camera back to its original position, and start again.
  - Practice **voice commands**.
2. Record (camera views, voice, vibrations) – **Switch on Record**:
  - Knot tying
    - Explain the objective of the knot-tying task
    - Record at least one trial with **only vision** (no live haptic feedback, no verbal explanations)
    - Record at least one trial with **vision and live haptic feedback** (no verbal explanations)
    - Record at least one trial with **vision and verbal explanations** (no live haptic feedback)
    - Choose favorite recording of knot-tying task
  - Suturing
    - Explain the objective of the suturing task
    - Record at least one trial with **only vision** (no live haptic feedback, no verbal explanations)
    - Record at least one trial with **vision and live haptic feedback** (no verbal explanations)
    - Record at least one trial with **vision and verbal explanations** (no live haptic feedback)
    - Choose favorite recording of suturing task
3. Questions about **recording tasks** with multimodal feedback:
  - [WRITTEN EXPLANATION] Our system can currently **record** 3D videos for **visual feedback**, instrument vibrations for **haptic feedback**, and verbal explanations for **audio feedback**. You can see your own real instruments in the console. If haptic feedback is turned on, you can also feel the vibrations of your own instruments.
  - [WRITTEN] What is your opinion of the **knot-tying task**? Do you recommend any modifications?
  - [WRITTEN] What is your opinion of the **suturing task**? Do you recommend any modifications?
  - [WRITTEN] Did you notice any differences in how you performed the tasks **with and without the haptic feedback**? Explain.
  - [WRITTEN] What other comments or suggestions do you have about **these two tasks** for surgeons to learn to use the da Vinci?
  - [ORAL EXPLANATION] We adapted these two tasks from the Fundamentals of Robotic Surgery (FRS).



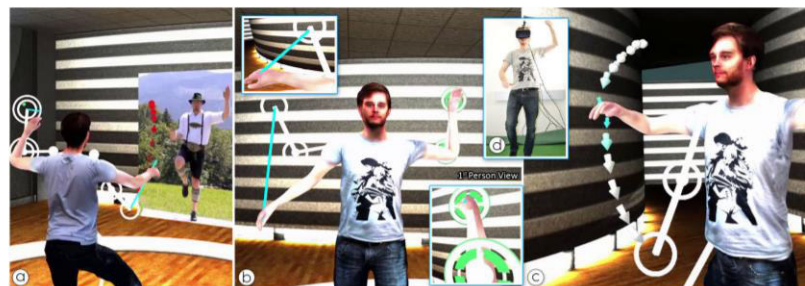
- [ORAL] What are the **tasks you recommend** for a surgeon learning how to use the da Vinci? For example, do you suggest using *FRS tasks*?
  - [ORAL] What is the **approach you recommend** for a surgeon learning how to use the da Vinci? For example, do you suggest using the *second surgeon console* or the *da Vinci VR simulator*?
  - [ORAL] What is the **procedure you recommend** for a surgeon learning how to use the da Vinci? For example, what are the steps your surgeons do before being ready to perform surgeries with the da Vinci?
  - [ORAL] What other comments or suggestions do you have about the **task recordings you made, training tasks, haptic feedback, training approaches and training procedures**?
4. Replay – **Switch on Replay**:
- Knot tying
    - Replay favorite recording at least once with only recorded **visual feedback** (no haptic feedback, no audio feedback)
    - Replay favorite recording at least once with recorded **visual and haptic feedback** (no audio feedback)
    - Replay favorite recording at least once with recorded **visual and audio feedback** (no haptic feedback)
    - Replay favorite recording at least once with recorded **visual and haptic and audio feedback**
  - Suturing
    - Replay favorite recording at least once with only recorded **visual feedback** (no haptic feedback, no audio feedback)
    - Replay favorite recording at least once with recorded **visual and haptic feedback** (no audio feedback)
    - Replay favorite recording at least once with recorded **visual and audio feedback** (no haptic feedback)
    - Replay favorite recording at least once with recorded **visual and haptic and audio feedback**
5. Questions about **replaying tasks** with multimodal feedback:
- [WRITTEN EXPLANATION] Our system can currently **replay** recorded 3D videos for **visual feedback**, recorded instrument vibrations for **haptic feedback**, and recorded verbal explanations for **audio feedback**. We envision that surgeons could experience these expert recordings inside the console to help them learn to use the da Vinci.
  - [WRITTEN] Do you think that **replaying expert performances** inside the console could be beneficial for trainees? Explain.
  - [WRITTEN] What do you think about the replayed **visual feedback** (replay of the expert's procedure from the stereo endoscope)?
  - [WRITTEN] What do you think about the replayed **haptic feedback** (replay of the expert's instrument vibrations)?
  - [WRITTEN] What do you think about the replayed **audio feedback** (replay of the expert's verbal explanations)?
  - [WRITTEN] Do you prefer replaying tasks with only **visual** feedback, **visual and haptic** feedback, **visual and audio** feedback, or with **all**? Rank the options in the provided list (1 as highest priority, 4 as lowest priority). Explain your ranking.

- [WRITTEN EXPLANATION] In the future, we are considering **trainees recording and replaying their own performances** inside the console. They would be able to see and feel how they just performed the task.
  - [WRITTEN] Do you think that trainees would benefit from recording and replaying their **own performances** inside the console? Why or why not?
  - [WRITTEN] What other comments or suggestions do you have about **replaying tasks**?
6. Train – **Switch on Record**:
- Knot tying
    - Train favorite recording at least once with **uninterrupted training mode** (no live haptic feedback)
    - Train favorite recording at least once with **piecewise training mode** (no live haptic feedback)
    - Train favorite recording with favorite training mode at least once with **live haptic feedback**
  - Suturing
    - Train favorite recording at least once with **uninterrupted training mode** (no live haptic feedback)
    - Train favorite recording at least once with **piecewise training mode** (no live haptic feedback)
    - Train favorite recording with favorite training mode at least once with **live haptic feedback**
7. Questions about **training tasks** with multimodal feedback:
- [WRITTEN EXPLANATION] Our system can currently let a surgeon **train** with recorded 3D videos for **visual feedback**. It allows a surgeon to try to follow the movements of a task recorded by an expert surgeon, either **uninterrupted** or **piecewise**. You can see both your own real instruments and the expert's instruments in the console. If haptic feedback is turned on, you can also feel the vibrations of your own instruments.
  - [WRITTEN] Do you think that **training expert performances** inside the console could be beneficial for trainees? Explain.
  - [WRITTEN] What do you think about the **uninterrupted training mode** (seeing what the expert did while also performing the same task with your instruments without interruptions)?
  - [WRITTEN] What do you think about the **piecewise training mode** (seeing what the expert did while also performing the same task with your instruments in pieces)?
  - [WRITTEN] Do you prefer training tasks with **uninterrupted** training mode, or with **piecewise** training mode? Rank the options in the provided list (1 as highest priority, 2 as lowest priority). Explain your ranking.
  - [WRITTEN] What do you think about the **visual feedback** during training (seeing what the expert did while also performing the same task with your instruments)?
  - [WRITTEN] What do you think about the **live haptic feedback** during training (feeling the instrument vibrations you are producing)?
  - [WRITTEN] Do you prefer training tasks with only **visual** feedback, or with **visual** and **live haptic** feedback? Rank the options in the provided list (1 as highest priority, 2 as lowest priority). Explain your ranking.

- [WRITTEN EXPLANATION] In the future, we are considering **adding audio to the training system**, such as a verbal explanation by the expert surgeon or the sounds from the task.
- [WRITTEN] Do you recommend **adding audio** to the training system? Why or why not? If yes, which sounds?
- [WRITTEN EXPLANATION] The system required you to start every voice command with the **trigger word “SEeAR”**.
- [WRITTEN] What do you think about this trigger word?
- [WRITTEN] What other comments or suggestions do you have about **training tasks**?
- [ORAL] What other comments or suggestions come to your mind about **how trainees could use this technology**?

8. Questions about **motion guidance**:

- [WRITTEN EXPLANATION] The expert’s transparent **ghost tools** you saw in the training part are one of the options we are currently investigating for **guiding the motion of the user**.
- [WRITTEN] What do you think about the **ghost tools** in the training part?
- [WRITTEN] Would you prefer to see the ghost tools **with background** (instruments and body tissue), or **without background** (only instruments) in the training part? Explain.
- [ORAL EXPLANATION] Another possible option for motion guidance is the usage of **trajectories**, which visualize the next few points of the trajectory that the expert surgeon followed while performing the task. Here is a sample showing trajectories in a different application:



[Yu X., et al., 2020]

- [ORAL] Would you find the visualization of **trajectories** useful for training?
- [ORAL] Besides ghost tools and trajectories, what **other motion guidance approaches** could you imagine to be beneficial for trainees, if any?
- [ORAL] What other comments or suggestions do you have about **how motion guidance could be used for trainees**?

9. Questions about **surgical skill evaluation**:

- [WRITTEN EXPLANATION] We are currently investigating different methods that could be used to **quantitatively evaluate the surgical skill of surgeons** during longitudinal training studies. At the moment, we have placed a **force sensor** below the task materials to measure how much the

instruments push and pull, and we have attached **accelerometers** on the robotic instruments to measure their vibrations.

- [WRITTEN] Do you think that **exerted forces** are important when evaluating the skill of a surgeon using the da Vinci?
- [WRITTEN] Do you think that **instrument vibrations** are important when evaluating the skill of a surgeon using the da Vinci?
- [ORAL] How **do you evaluate** a surgeon performing a task? What are the **main things** you consider for your evaluation (in order of importance)?
- [ORAL] What other comments or suggestions do you have about **how to evaluate surgical skill using this technology**?

## A.4 Questionnaire for Recording

Subject number: \_\_\_\_\_

### Recording Tasks with Multimodal Feedback

Our system can currently **record** 3D videos for **visual feedback**, instrument vibrations for **haptic feedback**, and verbal explanations for **audio feedback**. You can see your own real instruments in the console. If haptic feedback is turned on, you can also feel the vibrations of your own instruments.

What is your opinion of the **knot-tying task**? Do you recommend any modifications?

---

---

---

---

---

---

---

---

What is your opinion of the **suturing task**? Do you recommend any modifications?

---

---

---

---

---

---

---

---

Did you notice any differences in how you performed the tasks **with and without the haptic feedback**? Explain.

---

---

---

---

---

---

---

---



## A.5 Questionnaire for Replaying

Subject number: \_\_\_\_\_

### Replaying Tasks with Multimodal Feedback

Our system can currently **replay** recorded 3D videos for **visual feedback**, recorded instrument vibrations for **haptic feedback**, and recorded verbal explanations for **audio feedback**. We envision that surgeons could experience these expert recordings inside the console to help them learn to use the da Vinci.



Do you think that **replaying expert performances** inside the console could be beneficial for trainees? Explain.

---

---

---

---

---

---

---

---

What do you think about the replayed **visual feedback** (replay of the expert's procedure from the stereo endoscope)?

---

---

---

---

---

---

---

---

What do you think about the replayed **haptic feedback** (replay of the expert's instrument vibrations)?

---

---

---

---

---

---

---

---

What do you think about the replayed **audio feedback** (replay of the expert's verbal explanations)?

---

---

---

---

---

---

---

---

Do you prefer replaying tasks with only **visual** feedback, **visual** and **haptic** feedback, **visual** and **audio** feedback, or with **all**? Rank the options in the provided list (1 as highest priority, 4 as lowest priority). Explain your ranking.

- Visual** feedback
- Visual** and **haptic** Feedback
- Visual** and **audio** feedback
- Visual, haptic, and audio** feedback
- Other. Please specify \_\_\_\_\_

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---

---



## A.6 Questionnaire for Training

Subject number: \_\_\_\_\_

### Training Tasks with Multimodal Feedback

Our system can currently let a surgeon **train** with recorded 3D videos for **visual feedback**. It allows a surgeon to try to follow the movements of a task recorded by an expert surgeon, either **uninterrupted** or **piecewise**. You can see both your own real instruments and the expert's instruments in the console. If haptic feedback is turned on, you can also feel the vibrations of your own instruments.

Do you think that **training expert performances** inside the console could be beneficial for trainees? Explain.

---

---

---

---

---

---

---

---

What do you think about the **uninterrupted training mode** (seeing what the expert did while also performing the same task with your instruments without interruptions)?

---

---

---

---

---

---

---

---

What do you think about the **piecewise training mode** (seeing what the expert did while also performing the same task with your instruments in pieces)?

---

---

---

---

---

---

---

---





In the future, we are considering **adding audio to the training system**, such as a verbal explanation by the expert surgeon or the sounds from the task. Do you recommend **adding audio** to the training system? Why or why not? If yes, which sounds?

---

---

---

---

---

---

---

---

---

---

The system required you to start every voice command with the **trigger word "SEeAR"**. What do you think about this trigger word?

---

---

---

---

---

---

---

---

---

---

What other comments or suggestions do you have about **training tasks**?

---

---

---

---

---

---

---

---

---

---



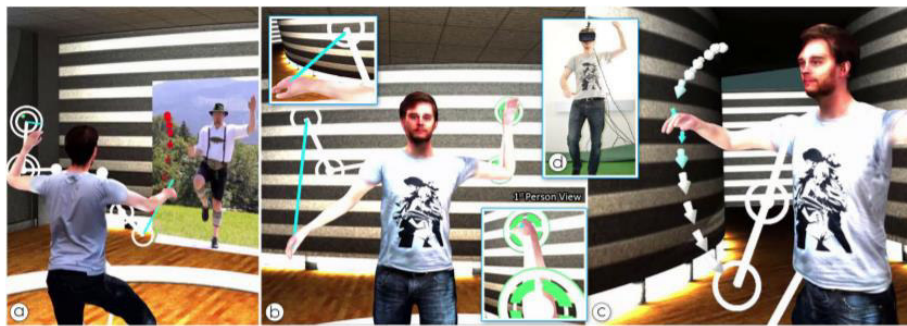
## A.7 Questionnaire for Motion Guidance

Subject number: \_\_\_\_\_

### Motion Guidance

The expert's transparent **ghost tools** you saw in the training part are one of the options we are currently investigating for **guiding the motion of the user**.





[Yu X., et al., 2020]

## A.8 Questionnaire for Surgical Skill Evaluation

Subject number: \_\_\_\_\_

### Surgical Skill Evaluation

We are currently investigating different methods that could be used to **quantitatively evaluate the surgical skill of surgeons** during longitudinal training studies. At the moment, we have placed a **force sensor** below the task materials to measure how much the instruments push and pull, and we have attached **accelerometers** on the robotic instruments to measure their vibrations.





## **Declaration**

I hereby declare that the work presented in this thesis is entirely my own and that I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

---

place, date, signature