



Supporting Collaborative Discussions In Surgical Teleconsulting Through Augmented Reality Head Mounted Displays

Sophie Maria, Helena Mentis, Geoffroy Canlorbe, Ignacio Avellino

► To cite this version:

Sophie Maria, Helena Mentis, Geoffroy Canlorbe, Ignacio Avellino. Supporting Collaborative Discussions In Surgical Teleconsulting Through Augmented Reality Head Mounted Displays. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), ACM, Apr 2023, Hambourg, Germany. 10.1145/3544548.3580714 . hal-04017042

HAL Id: hal-04017042

<https://hal.science/hal-04017042>

Submitted on 6 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Supporting Collaborative Discussions In Surgical Teleconsulting Through Augmented Reality Head Mounted Displays

Sophie Maria

sophie.maria@sorbonne-universite.fr
Sorbonne Université, ISIR. AP-HP, Pitié-Salpêtrière
University Hospital
Paris, France

Geoffroy Canlorbe

geoffroy.canlorbe@aphp.fr
Sorbonne Université, INSERM UMRS 938, CRSA. AP-HP,
Pitié-Salpêtrière University Hospital
Paris, France

Helena M. Mentis

mentis@acm.org
University of Maryland, Baltimore County
Baltimore, United States

Ignacio Avellino

ignacio.avellino@sorbonne-universite.fr
Sorbonne Université, CNRS, INSERM, ISIR
Paris, France

ABSTRACT

Although Augmented Reality (AR) has been touted as the future of surgery, its contribution to distributed collaboration such as in surgical teleconsulting has not been articulated. We propose AR-Head Mounted Displays (AR-HMD) to tackle two previously-identified challenges: operating surgeons needing to view and interact with imaging systems that reside away from the operative field, and, their lack of gesturing tools to point and annotate on the shared images and physical environment. We report on a controlled lab experiment where 12 expert gynecology surgeons perform a tumor localisation task guided by a remote radiologist (confederate) via an AR-HMD. We find that bringing the shared images to the place of work reduces the need for clarifications and provides opportunistic access to information when required, and, that pointing and annotating provides opportunities to further support verbal instruction in deictic communication. Our results inform the design of intraoperative AR-HMD systems for surgical telecollaboration.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing.**

KEYWORDS

teleconsulting, augmented reality, remote collaboration

ACM Reference Format:

Sophie Maria, Helena M. Mentis, Geoffroy Canlorbe, and Ignacio Avellino. 2023. Supporting Collaborative Discussions In Surgical Teleconsulting Through Augmented Reality Head Mounted Displays. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3544548.3580714>

1 INTRODUCTION

Surgery is a complex activity that oftentimes requires different experts to collaborate towards solving a problem. For instance, a surgeon can face difficulties during an operation and require the advice of a remote radiologist, to collaboratively interpret the patient's preoperative images such as MRI (Magnetic Resonance Imaging) or CT-Scan (Computed Tomography Scan). Using collaborative technologies has been proposed to support this form of *teleconsulting*, when a consulting expert is remote, especially as surgeons are becoming more open to the possibility of receiving and providing remote help given the shift in work after the COVID-19 pandemic. One of these systems is the VisitOR1 (Karl Storz, Germany), a commercial system that captures the view of the Operating Room (OR), and lets remote experts present preoperative images and draw annotations. This system exemplifies the mainstream design choices of systems for remote collaborative work in surgery today, which bring with them two challenges. First, the remote expert may present information on a monitor placed outside of the operative field in what is called a *secondary* imaging system, which complements *primary* imaging systems that show a real-time view of the patient during surgery [29]. As the monitors of secondary imaging systems are distant, this forces teams to reconfigure (a) spatially, as looking at the monitor requires direct line of sight [33]; and, (b) cognitively, as they create the need for additional articulation work and attention shift of the team for decision making [29]. The second challenge is that these systems predominantly support single-user input, as only the remote expert can virtually point and annotate. Previous work has shown how single-user input tools lead to imbalanced communication [15], reducing participation of novices at the receiving end of instructions, which in turn can lead to poor decision making and negatively impact learning [15, 39].

In this work, we address these two challenges by using Augmented Reality Head Mounted Displays (AR-HMDs) for surgical teleconsulting, as these systems let the operating surgeon reposition and annotate virtual images that a remote expert presents. We investigate two research questions: (RQ1) what is the impact of the ability to position images freely on collaborative work during surgical teleconsultation, and, (RQ2) what is the impact of dual-user input tools on communication when discussing secondary images

CHI '23, April 23–28, 2023, Hamburg, Germany

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany, <https://doi.org/10.1145/3544548.3580714>.

during surgical teleconsultation. AR-HMDs are increasingly being used for surgery as devices become smaller, more portable, and most important: wearable. The advantage of wearable devices is that surgeons can interact with information through mid-air hand gestures without breaking sterility, as opposed to AR systems through tablets [2]. We conduct an experiment where we use the Microsoft HoloLens2 to enable the operating surgeon to (a) freely reposition remotely presented images through mid-air gestures, and (b) annotate these images. Through the analysis of verbal communication structure and content, non-deictic communication, and use of the secondary display, our study shows that 1) repositioning the secondary display has communication benefits, and provides on-demand access to information when it is needed; 2) the ability for the operating surgeon to perform virtual gestures provides opportunities to further support verbal instruction in deictic communication. We conclude with perspectives for future work on the study and improvement of AR for surgical teleconsulting.

2 BACKGROUND

Complex activities that require numerous skills oftentimes need several experts coming together during their execution, each bringing complementary knowledge to a collaborative discussion. For example in surgery, a surgeon oftentimes faces technical or decisional difficulties while operating on a patient that requires the advice of a skilled radiologist, to read together a preoperative image such as an MRI or CT-Scan. This type of situation can occur in everyday surgery, and confronts the operating surgeons with two challenges. First, accessing images on a monitor distant from the operative field. This poses an interaction problem as surgeons need to interact with a non-sterile input device (mouse) within a sterile environment. As reviewing images requires scrolling through layers, surgeons have to remove their gloves, interact with the device, and then disinfect themselves again before returning to the surgical site, while preserving the asepsis of their gowns. They can also delegate this interaction to a nurse, which makes the task more complex. Secondly, confronting the operating surgeons with the challenge of finding an expert radiologist that is available. Typically, these remote consultations take place as a telephone call, where the expert can only explain verbally their interpretation of the image and the different locations on the body, which limits the operating surgeon's understanding of the provided information.

3 RELATED WORK

For better framing the problems raised by the collaborative discussions in surgical teleconsulting, we first explore the actual implementation of AR for collaborative work in surgery, then, on this basis, we explore the use of imaging systems, annotation tools and their limitations.

3.1 AR for Surgical Telecollaboration

The development of AR has led researchers to study its possible contribution to the surgical environment. The introduction of a new technology in a field as sensitive as surgery has mostly involved the realization of feasibility studies to ensure the safety and ease of use of this new tool [1, 13, 43]. In a study where surgeons could take pictures, record video and have access to preoperative

data through an AR-HMD, Borgmann et al. [7] demonstrates that AR-HMDs are safe and appreciated by physicians. Their results motivated further studies to investigate its contribution, especially in teleconsulting. This was undertaken by Rojas Munoz in 2019 [35], adding the AR-HMDs to their System for Telementoring with Augmented Reality (STAR) [2]. The initial system enabled collaboration between remote surgeons using a tablet placed directly between the operating surgeon (novice) and the operating field, which the remote mentor could then anchor annotations on the novice's screen. Its main limitation was the lack of depth perception and incorrect transmission of novice's perspective. Adding AR-HMDs overcame these limitations and enabled a decrease in the rate of errors and increase confidence of novice surgeons performing leg fasciotomies (IPS Mackenzie) [35]. More recently, and with the aim of better immersion, increasing the feeling of co-presence, Gasques has created a collaborative Mixed Reality system, ARTEMIS (Augmented Reality Technology-Enabled reMote Integrated Surgery) [19], enabling collaboration through a hybrid interface where the remote expert uses virtual reality (with 3D reconstruction of the patient's body) and guides the novice through an AR interface. Focused on the expert point of view and needs, this feasibility study based on a participatory design process with surgeons, does not study communication dynamics and operating surgeon point of view. Lastly, Seo et al. [42] proposed an interaction technique for remote experts to point and annotate in mobile views captured from AR-HMDs, that is fluid as it does not require freezing the transmitted video. Most studies in the literature focuses on this kind of collaboration, where a remote expert provides knowledge and instructions to a local novice. We propose to study another type of remote interaction, between peers, which does not follow the same dynamics and is more akin to teleconsulting.

3.2 Collaborative work

As defined by Roschelle and Teasley, collaboration is a “*coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem*” [36]. In order to increase the communication efficiency, the stakeholders build *common ground*, mutual knowledge, beliefs and assumptions [9]. Speakers mutually assume that the other has understood well enough what they wanted to express in order to achieve their goal. *Grounding*, the process which bases common ground, consists of two phases: Presenting (the speaker presents the statement to the receiver) and Accepting (the receiver accepts the statement by providing evidence of understanding) [10]. To develop this common ground, partners choose techniques that require the least collaborative effort [10], from the presenting and accepting phases. Collaborators try to minimize their efforts spent in the interaction, which should ideally be short and informative. Communication costs vary according to the medium used and may concern the speaker (production and formulation), the receiver (reception, understanding) and both, in a non-exhaustive way, linked with the display used, misunderstandings or errors made. In collaborative work, speakers do not communicate the same way according to their level of expertise. When making a reference, a novice tends to describe the location of the object and experts are more likely to articulate the task object [22].

3.3 Surgery and Collaborative Work

Surgical collaboration has some specificity, firstly because it is enacted in a sterile environment which limits gestures and contacts between collaborators. Then, because members of the surgical team are exposed to many sources of information – from team members themselves, from the patient or from imaging systems in the OR – they must assimilate all in order to coordinate their action and anticipate their future action through articulation work [29]. Challenges have then emerged such as facilitating interaction with information sources (such as image systems) distant from the surgical working sphere and providing tools for supporting deictic communication.

3.3.1 Interacting with Secondary Displays in Surgery. Surgery relies on several imaging systems, from which Mentis [29] distinguishes primary, necessary to do the work, from secondary, supportive for better work practices such as decision making. These imaging systems are shown through a display, which is usually located on the wall of the OR, outside the sterile zone around the patient. This disposition re-configures the surgical team [29], and imposes frequent focus changes between the operating field and the monitor, which induce a shift in workflow. While, having the images by the tableside allows for tighter integration between the information from the images and the information from the surgical site [34]. Furthermore, going to a remote computer and handling a non-sterile mouse induces a cost, especially for surgeries where these needs are frequent. In addition to a loss of time on task and an overall increase in operating times, this movement can impact patient safety [23]. Solutions to avoid breaching asepsis while handling secondary displays were then studied, particularly touchless interactions [27, 32, 33], which can be illustrated by gesture or voice commands of the secondary display. Nonetheless, the display remains located far from the surgeon. In summary, previous work has identified the shortcomings of the distant positioning between imaging systems and surgeons, but so far addressed interaction techniques with the distant display. The first goal of our work is to provide surgeons with ways to reposition imaging systems, to understand its benefits during remote consultation with colleagues (RQ1).

3.3.2 Deictic Communication in Surgical Work. Many studies aim to find a way for surgeons to communicate efficiently. The multiplication of communication supports aims to make interactions increasingly more efficient, in order to decrease *cognitive load*, i.e., the level of mental energy required to process information [4]. First, by enabling shared vision of a common view which lets collaborators co-construct their understanding of the information [31]. Then with a virtual pointer, which has proved its worth in co-located mentoring by reducing cognitive load and improving communication [38]. In addition, annotations bring information as shape, orientation or direction of motion and enable a more efficient collaboration than pointers cues regarding the communication of object position and orientation [25]. Although research in telementoring has shown that single-user tools can improve task performance with a shorter learning curve [44] and faster completion time [24], when it comes to OR imaging systems specifically, they can lead to an imbalanced communication as trainers using the tool tend to dominate the decision making while local trainees under-contribute to such process [15]. Multi-user input over a shared display has

been shown to support group process and improve performance. Creating shared displays through multi-user input is not straightforward for cognitively and physically demanding tasks such as surgery. The secondary task of giving input for controlling pointing or annotating, impose demands on an already cognitively taxed operator [3]. Surgery has been singled out as a domain that particularly struggles with this dual-task challenge [6, 8], and in response, a different approach has been studied, of shifting the load of a more balanced communication to the remote trainer. Previous work by Semsar et al. [39, 40] has shown that training on the communication skills for appropriate use of telementoring tools could effectively convey remote instruction and mitigate the imbalance, although in a more recent study, the authors show that this has limitations as local trainees still show the need for virtual referencing, as well as ownership of the virtual annotations to manage them according to their needs as the task advances, and not when the trainer decided [41]. In summary, single-user input tools can improve task performance but also can make communication imbalanced. The latter has been addressed from a *social* approach, by training trainers on communication skills, but not from a *technological* approach given the interaction limitations while performing surgery – i.e., the surgeon cannot annotate using a mouse. In this work, we provide annotation tools through mid-air hand gestures, to study the benefits of dual-user input tools (RQ2).

4 METHOD

We explore the use of AR-HMDs for repositioning secondary displays and dual-user gesturing tools in remote surgical collaboration. We address a concrete situation where a surgeon consults with a radiologist when facing difficulties in finding a non-visible structure during surgery, previously identified in a preoperative image. We operationalize this through six scenarios where participants in the role of operating surgeons have to identify the structures.

4.1 Participants

Part.	Age	Gender	Speciality	Surgical Experience	AR Experience
P1	37	M	Gynecology	3 years	0
P2	35	F	Gynecology	6 years	0
P3	38	M	Gynecology	8 years	0
P4	36	F	Gynecology	7 years	1
P5	41	M	Gynecology	10 years	0
P6	31	F	Gynecology	0.5 years	0
P7	34	F	Gynecology	6 years	0
P8	32	F	Gynecology	3 years	0
P9	39	F	Gynecology	9 years	0
P10	29	F	Gynecology	1 years	0
P11	35	F	Gynecology	3 years	0
P12	36	M	Gynecology	8 years	0

Table 1: Participants demographics.

12 attending surgeons participated in this study as the local operating surgeons (Table 1). They were all gynecologic surgeons, had different levels of experience varying from 0.5 to 10 years ($M=5.8y$), and practiced in various hospitals in Paris. Most participants had no previous experience in AR ($n=11/12$), only one had tried an AR-HMD once the previous year. They were recruited on a voluntary basis without financial compensation, after being contacted and informed about the study by members of the research team.

Condition and description	Figure
<p>Condition FD-SU <i>Fixed DISPLAY + Single USER.</i> The image is shown on a screen fixed to the wall of the OR, and only the confederate can make annotations. This emulates the classic approach for telementoring.</p>	
<p>Condition MD-SU <i>Mobile DISPLAY + Single USER.</i> The image is shown next to the patient table through the AR-HMD that the participant wears, he can point but only the confederate can make annotations.</p>	
<p>Condition MD-DU <i>Mobile DISPLAY + Dual USER.</i> The image is shown next to the patient table through the AR-HMD that the participant wears, and both the remote and operating surgeons can make annotations.</p>	

Table 2: Experimental conditions.

4.2 Experimental Design

The experiment follows a [2x2] within-subjects design, with factors:

- **DISPLAY:** the location of the display showing the image, with levels: *fixed* and *mobile*; and,
- **TOOL:** the access to the annotation tool, with levels: *single-user* and *dual-user*.

We selected three conditions out of the possible four, discarding *fixed DISPLAY* with *dual-user* input *TOOL*. This is because this condition would require participants to interact with a non-sterile screen if they wished to produce annotations, use voice control, or use a completely different type of technology that enables touchless interaction such as a Microsoft Kinect [33], which is not the scope of our study. Table 2 details the retained conditions. Each

participant had to play the role of the operating surgeon. The role of the remote radiologist was consistently played by a member of the research team as a confederate (a gynaecological surgeon with 3 years experience), who conducted the communication according to a pre-established script. We chose a single confederate to perform each scenario to maintain consistency in the way the remote radiologist communicated, with the goal of eliminating biases linked to the way each pair interacts during the study, as the main focus is on the operating surgeon. In each condition, participants perform two scenarios; we counterbalanced the condition order using a Latin square, as well as the scenarios. We obtained ethics committee approval for this study, we also pre-registered the study: <https://osf.io/zd7bx>.

4.3 Scenarios

We created six scenarios on the basis of current open gynecological surgical practice. They operationalize a common situation in surgery, where a surgeon has trouble identifying structures to be removed (e.g., tumors or nodes), either because they are too small, or because they are particularly difficult to find given anatomical variations (e.g., non palpable adenopathy). The surgeon requires the help of a radiologist to locate these non-visible, non-palpable structures. In our scenario, the surgeon consults a remote radiologist, soliciting a joint and in real-time analysis of the image, to find the lesion on the patient and to guide them in the resection. Two domain experts designed the six scenarios. First, one of the authors (gynecologic surgeon, with three years of surgical experience) reflected on situations actually encountered in the OR, and then tested and iterated on the scenarios with a co-author (gynecologic surgeon, with ten years of experience). The scenarios were designed to have the same level of difficulty. For this, we worked with an expert radiologist to interpret a series of preoperative images and find six where finding the target nodule location requires a comparable level of difficulty. All preoperative images came from a hospital imaging software and were anonymous. Half of the scenarios involve breast surgery and half pelvic surgery.

4.3.1 Breast Surgery Scenarios. The first type of scenarios concerns a small tumor, which requires a radiologists to insert a harpoon-shaped metallic thread under mammographic control preoperatively (before surgery), to guide the procedure during surgery. Unfortunately, it can happen that this harpoon is torn off by accident, for example during the skin disinfection. The surgeon is then left without a guide to perform a resection on a tumor too small to be found through palpation. In this case, the surgeon in consultation with a radiologist relies on the preoperative image to estimate the predictable location of the tumor. The scenarios rely on an MRI, and only differ in the tumor location.

4.3.2 Abdominopelvic Surgery Scenarios. The three abdominopelvic scenarios are based on single node recurrence of cervical cancer, where the goal is to find and remove a lymph node. The small size of the diseased node makes it difficult to identify it during surgery. The scenarios rely on a CT-Scan, and only differ in the node location.

4.4 Apparatus and Setup

The study took place in a simulated OR at the site BOPA (*Bloc Opératoire Augmenté*). It consists of a real patient table and surgical equipment, including endoscopic screens and scialitic lighting. The participant stands next to the table assuming the role of operating surgeon, while a confederate is in a contiguous room assuming the role of radiologist, as shown in Figure 1. In all three conditions, the participant wears a head-mounted camera to capture the video that the confederate sees. In *Fixed DISPLAY + Single USER* we achieve this by having the participant wear the HoloLens2 as they do in conditions with AR, *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER*. This might seem artificial, but we envision a scenario where an operating surgeon would wear a head-mounted camera while performing surgery. The alternative of using a fixed camera to capture the OR would result in a confounding variable for *Fixed DISPLAY + Single USER*, making the video capture different from the

other two, and thus creating behavioral data that is potentially not comparable. In *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER*, the participants can use a virtual pointer to target, select and manipulate the virtual screen. The annotation tool is only available for *Mobile DISPLAY + Dual USER* with which the participant can freely draw, place arrows on the images or on the patient body, with the color of their choice. Participants select the annotation feature and start annotating by making a pinch gesture between thumb and index finger. There is no obligation to use the annotation tool in *Mobile DISPLAY + Dual USER*, participant can either choose pointing or simple hand gestures. Note that the virtual screen and annotations are world-stabilized by the HoloLens2, staying fixed to the real world as participants move their head.



Figure 1: Simulated OR. In the foreground: a participant wears the HoloLens2 and indicates the tumor location on the patient model. In the top-left corner, the confederate watches the transmitted video.

For all conditions, the participant initiates an audio-video call using Microsoft's Dynamics 365 Remote Assist¹ running on the HoloLens2. The confederate answers this call on a Dell tablet through Microsoft Teams². For *Fixed DISPLAY + Single USER* specifically, we run a video call in parallel (no audio) using Microsoft Teams between a laptop computer on the expert side and a screen fixed to the wall of the OR, to show only the preoperative image. The confederate follows the same script for all experimental conditions, and produces annotations only on the preoperative image, never on the real world. We take this precaution to avoid an imbalance between conditions *Fixed DISPLAY + Single USER*, where there is no

¹<https://dynamics.microsoft.com/en-us/mixed-reality/remote-assist/>

²<https://www.microsoft.com/en-us/microsoft-teams/>

Sb1	Sb2	Sb3	Sp1	Sp2	Sp3
Hello, what is your problem?					
I will review the preoperative image. I've just opened it, can you see it?					
Can you see the tumour on the MRI?			Can you see the node on the CT Scan?		
Can you tell me where you see it?			Can you tell me where you see it?		
If participant shows the location: correctly => confirm by annotating on the image incorrectly => show the correct location on the image by annotating and ask for confirmation again					
Let's see the patient, where is her head?			Lets see the patient, where is her head?		
Which breast are you operating on?					
Is there any way to see on the patient the four quadrants of the breast?			Is there any way to see on the patient the vascular structures?		
Can you tell me now where do you locate the tumour?			Can you tell me now where do you locate the node?		
Let's go back to the MRI, I will show you more layers.			Let's go back to the CT Scan, I will show you more layers.		
I will draw the four quadrants on the image			I will draw the different vascular structures.		
Does this help you?			Does this help you?		
Now that it is more clear, let's go back to the patient					
Is there any way to show me the union of the internal quadrants, at 2,3 cm from the nipple?	Is there any way to show me the union of the inferior quadrants at 52 mm from the nipple?	Is there any way to show me the 2 o'clock ray, 3,5 cm from the nipple?	Is there any way to show me between the aorta and the vena cava, 3 cm below the left renal vein?	Is there any way to show me the latero aortic level at 1 cm above the bifurcation?	Is there any way to show me the location next to the vena cava just below the left renal vein?
Okay, this is the location.					

Table 3: Confederate script for the six different scenarios.

AR and thus real world annotations are not possible, and *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER* where real world annotations are technically possible. The confederate uses the application DemoPro³ in *Fixed DISPLAY + Single USER* to annotate on a window showing the preoperative image, and Remote Assist in conditions *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER* where the confederate annotates, which requires making a screen capture. For the breast scenarios, we used a mannequin (model of breast palpation 3B scientific) which had the advantage of recreating the 3D and haptic conditions of real practice. For the abdominopelvic scenarios, we used an A3 printout of a lymph node surgery as it was not possible to obtain a 3D model that faithfully reproduces the large vessels of the abdominal cavity.

4.5 Procedure

We welcomed participants and provided written instructions with a description of the experiment including the tasks, training instructions, and an overview of the three conditions, presented in the particular order for each participant. They filled a pre-task questionnaire (Appendix A) concerning demographics information and experience with AR-HMD.

4.5.1 Training. Participants were taught how to use Microsoft HoloLens2, Dynamics 365 Remote Assist and Teams. First, they were invited to read an instruction manual on how to start Remote Assist, and on the annotation features for drawing and making arrows. Training consists of a series of four simple tasks, carried out both in the real world and on a virtual image (Figure 2). The first training task tests communication, (1) the participant has to look at four objects that the confederate indicates verbally. Then, three training tasks test comprehension of remote annotation, (2) the participant has to identify among four possible objects: one real-world object, one on a monitor, and one virtual object displayed in

an AR screen, all which the remote confederate circles or points to using arrows. Then a task to test the production of annotations, (3) the participant has to circle among four possible objects: one object once on a printed sheet in the real world, and one on a virtual sheet, all which the confederate indicates verbally. The final task tests accuracy of producing annotations, (4) the participants has to first draw three shapes staying inside the two lines that represent the shape's contour, without stepping over, and secondly indicate the correct path to the exit of a labyrinth by using arrows. Both of these in the real world and in a virtual image. We considered training finished when participants performed all tasks successfully.

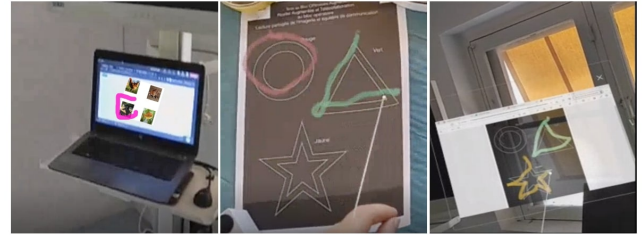


Figure 2: Training tasks. Remote on-screen annotation (left), and, accuracy of annotation both on the real world (middle) and on the virtual window (right).

4.5.2 Tasks and Confederate Script. The confederate followed the same script for each scenario, described in Table 3. After answering the call from the participant, the confederate opens the preoperative image, showing the structures sought (tumor or lymph node) via an annotation on the image. They then ask the participant to confirm if they understand where the structure is located on the image; if the participant is incorrect, the confederate shows it again. Then, the confederate asks the participant to show the patient, asking

³<http://www.demoproapp.com/>

to identify anatomical landmarks that serve as guidance, and to identify the structure location based on those landmarks. Next, the confederate goes back to the image and annotates such landmarks, asking if this will help locate the structure more precisely on the patient. Back on the patient, the confederate asks to identify the location, given one last instruction based on the landmarks, and asks to place a marker on the location. The confederate keeps correcting the participant until they identify an accurate location.

One trial corresponds to performing one scenario, for one given condition. Each participant performed six trials in total, one for breast and one for pelvis in each of the three conditions.

4.6 Data Collection and Analysis

We recorded audio and video of the sessions for later analysis of communication structure and content of each condition. They came from two sources, the HoloLens2 recordings and from an external camera placed in the simulated OR, directed at the operating surgeon. For all measures, we perform pairwise comparisons between conditions *Fixed DISPLAY + Single USER* and *Mobile DISPLAY + Single USER* for answering RQ1, and between *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER* for answering RQ2, as our two research questions are independent. For analyzing time we perform a Kruskal-Wallis non-parametric test as the distribution departs from normality. For measures of verbal communication, non-verbal communication and use of secondary display, we perform a generalized Poisson mixed model as these count measures follow a Poisson distribution. For the recordings analysis we did not take into account interactions between confederates and participants that were not related to the performance of the task. Plots show means with 95% confidence intervals.

4.6.1 Task Completion Time. We recorded TCT (Task Completion Time) in seconds, starting when the confederate answers the call (i.e., “hello”), and ending when the confederate judges that the position of the target indicated by the participant on the patient is correct (i.e., “perfect”).

4.6.2 Communication Balance and Purpose. We use the recorded videos and transcripts to systematically evaluate communication. To analyze the structure of communication, we code speech and turns, to perform a turn-taking analysis. This measure reflects balance, as distribution of turns is skewed according to the difficulty in taking the floor. We compute (a) turn frequency as the number of turns per second, (b) turn duration as the duration of turns in milliseconds, and, (c) turn distribution as the proportion of turns for the trainee. To analyze the content of communication, we code the content of utterances using Feng’s [15] adaptation of the original dialogue act coding scheme by Sellen [37] (Table 4). This scheme, inspired by Convertino et al. [12], consists of: Transfer Info, Check Understanding and Manage Process & Decision. Authors 1 and 4 coded the transcripts from one participant independently from each other while watching the recordings. Using Cohen’s kappa [11], we computed the inter-rater reliability, which yielded $\kappa = 0.278$. Authors 1, 2, and 4 then discussed conflicting codes and authors 1 and 4 re-coded the same two participants achieving an agreement of $\kappa = 0.796$. Authors 1 and 4 then each coded half of the participants

and finished with a joint review of the coding of all twelve participants. At the same time, the confederate dialogues were analysed to check that they were consistent across participants.

Class	Dialogue Act	Description
Transfer Info	Add Info (AI)	Provides new information, not elicited.
	Query (Q)	Question used to elicit new information.
	Reply (R)	Reply to query to provide new information.
Check Understanding	Check (CH)	Verify own understanding of information previously presented by others.
	Align (AL)	Verify partner’s understanding of information previously presented to others.
	Clarify (CL)	Clarifies or restates information already presented.
	Acknowledge (AC)	Signals receipt of information, understanding.
Manage Process & Decision	Manage (MN)	Instruction, command, direct or indirect request for action; orchestrating strategy, how to do the work.
	Summarize (SA)	Summarizes information previously presented.
	Judge (J)	Individual judgment, opinion, or preference.
	Confirm (CO)	Requests partners’ agreement on a proposed decision.
	Agree (AG)	Indicates approval for a prior judgment or decision.

Table 4: Dialogue act coding scheme (Feng et al. [15]).

4.6.3 Communication Through Deictic Referencing. To investigate the use of the deictic referencing tools, we first compare the total amount of actions supporting deictic communication between *single-user* vs. *dual-user* TOOL. In both conditions participants could use their Physical Fingers (PF) or a Virtual Pointer (VP) to show structures both in the real world and images. In the *dual-user* TOOL condition, participants could additionally use Virtual Annotations (VAn) and Virtual Arrows (VAr) to support communication. We investigate the communicative purpose of VAn and VAr, merging them into one category (VA), since we are interested in the use of virtual annotations in general. We associate these events with the corresponding dialogue act referenced in Section 4.6.2.

4.6.4 Use Of Secondary Display. Using video recordings from the HoloLens2 and the OR camera we investigate the use of secondary displays between *Fixed DISPLAY + Single USER* and *Mobile DISPLAY + Single USER*. We have primarily explored the participants need for information, by comparing their looks towards the secondary display, spontaneous or ordered by the confederate, and their effectiveness to understand the information broadcast. Then, we evaluate the participant access to the information displayed by counting the times that the participant “zooms” into the information, either by moving closer to the screen in *Fixed DISPLAY + Single USER* or by placing the screen closer in *Mobile DISPLAY + Single USER*. Finally, in order to understand when these events happen, we note their temporal occurrence.

4.6.5 Perceived Usefulness. To assess the image repositioning and annotation features, participants completed a post-task questionnaire answering in a likert scale from 1 – 7 (Appendix B).

5 RESULTS

5.1 Task Completion Time

The distribution of time departs from normality (Shapiro-Wilk test $W = 0.927$, $p = 0.021$). The Kruskal-Wallis test does not show a difference between *Fixed DISPLAY + Single USER* ($M = 196.62 \pm 31.50$) and *Mobile DISPLAY + Single USER* ($M = 229.96 \pm 55.10$) ($p = 0.16$), neither between *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER* ($M = 257.54 \pm 44.75$) ($p = 0.065$). Thus suggesting no difference in time related to the ability to position images freely, and from the mechanism for deictic referencing and annotating.

5.2 Communication Balance and Purpose

Communication Balance. We do not observe a significant difference in frequency (Figure 3a), duration (Figure 3b), or distribution (Figure 3c) of turn taking. We thus cannot make inferences about the communication balance when the operating surgeon could reposition the secondary images or when they had deictic referencing tools.

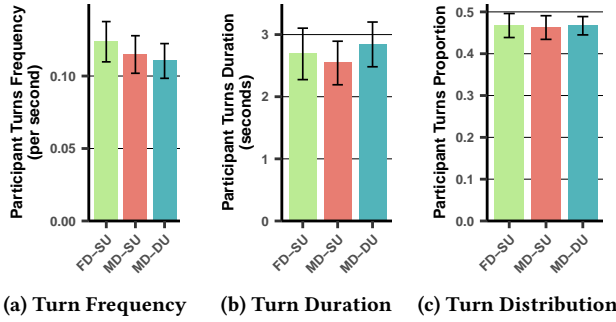


Figure 3: Participant turns.

Communication Purpose. Comparing the participants' dialogue acts between condition *Fixed DISPLAY + Single USER* and *Mobile DISPLAY + Single USER* (Figure 4), we observe a significant increase in management (MN) ($\beta = 0.259$, $p = 0.0286$) when participants can manipulate the virtual window at will, which in all cases they set in the immediate vicinity of the operating table. Additionally, being able to manipulate the virtual window decreases the need for clarifications (CL) ($\beta = -0.305$, $p = 0.0077$). However, we

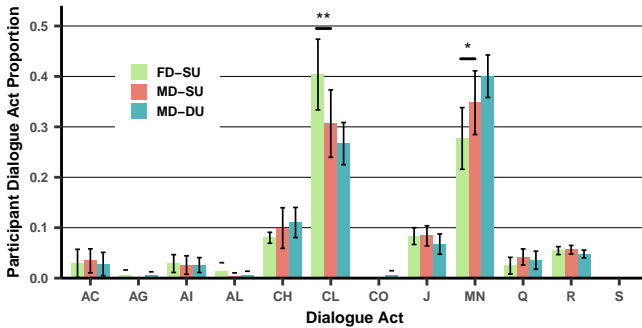


Figure 4: Participant dialogue acts.

do not observe significant differences between conditions *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER* regarding communication content. Therefore, we cannot make inferences about the effect of the operating surgeon's ability to annotate on communication. We note that we performed the same analyses by using the confederate dialogue acts data to understand if the confederate biased communication, which did not yield statistical differences (all p 's > 0.05). This suggests that the confederate did not influence the observed differences across conditions.

5.3 Communication Through Deictic Referencing

We first report that $2/3$ of participants ($N=8$) used annotations and virtual arrows when these were available (condition *Mobile DISPLAY + Dual USER*), showing that some participants actively chose not to use virtual annotations and arrows. Second, we compare the total amount of actions supporting deictic communication between *single-user* vs. *dual-user* TOOL (Figure 5a), regardless of how they were produced – Virtual Annotation (VAn), Virtual Arrow (VAr), Virtual Pointer (VP) or Physical Finger (PF). The comparison shows that when participants have access to annotating tools, they perform significantly more non-verbal actions ($\beta = 0.2624$, $p = 0.0386$). Then, we further investigate if this increase is due to a simple replacement of digital and virtual pointer or if these are new actions. When comparing *single-user* vs. *dual-user* TOOL, we do not find differences, neither for the use of finger (PF) ($p = 0.276$), nor for the use of the pointer (VP) ($p = 0.306$), suggesting that annotations and arrows are being used for new actions.

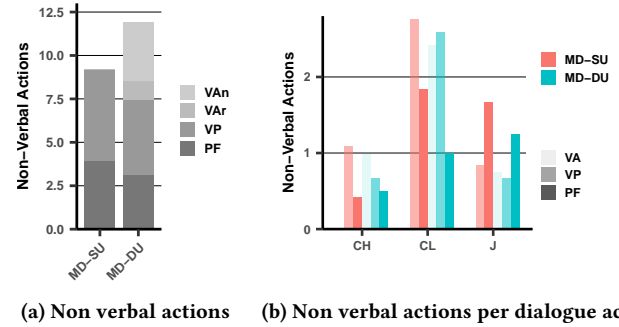


Figure 5: Participant non verbal communication.

Lastly, we investigate the communicative purpose of the Virtual Annotations (VAn) and Virtual Arrows (VAr) when these functions were available (condition *Mobile DISPLAY + Dual USER*). We compute the mean count of these events (Figure 5b), excluding communicative acts that happened rarely with non-verbal actions (AI, AL, CO, Q and MN), which are almost zero in each conditions (below 0.3), from which the following categories remain: CH, CL, and J. Of these three dialogue acts, the majority of the annotations used in the *Mobile DISPLAY + Dual USER* condition are in service of clarification (CL). These are such actions as reiterating the location of a quadrant of the breast. Clarification through annotations was equally as high for clarification through the use of the virtual pointer. However, both of these mechanisms were conducted twice

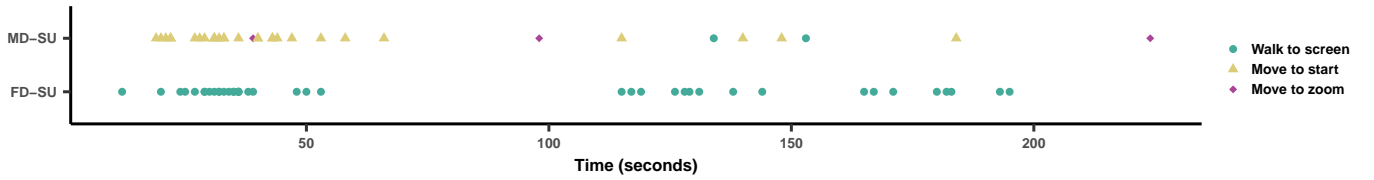


Figure 6: Participant temporal access to the secondary display.

as often than clarification through using one’s finger in the video stream. And in fact, with the ability to utilize the annotations and virtual pointer in the *Mobile DISPLAY + Dual USER* condition, we see a drop in the use of the finger for clarification acts. The second most prevalent use of annotations in the *Mobile DISPLAY + Dual USER* condition was for checking understanding (CH) — e.g., verifying that ‘this point’ is the location the radiologist is referring to. And finally, the third most prevalent use of the annotations in the *Mobile DISPLAY + Dual USER* condition was for judgement (J) acts — e.g., stating the tumor location the body as a final definitive decision. With the relative reduction in use of the virtual pointer and finger for these acts in the *Mobile DISPLAY + Dual USER* condition, the ability to annotate seems to be fulfilling the operating surgeon’s need for exactness in the discussions with the remote radiologist, moreso than can be achieved by simply using an ephemeral pointer, virtual or not.

5.4 Use Of Secondary Display

Regarding the assessment of the participant’s need for information, we do not find a significant difference ($p = 0.3877$) when comparing the amount of spontaneous interactions with the secondary displays between *Fixed DISPLAY + Single USER* ($M = 0.875 \pm 1.261$) and *Mobile DISPLAY + Single USER* ($M = 1.125 \pm 1.295$). The interactions with the secondary display ordered by the confederate also did not differ significantly ($p = 0.906$) between *Fixed DISPLAY + Single USER* ($M = 1.500 \pm 0.589$ and *Mobile DISPLAY + Single USER* ($M = 1.458 \pm 0.588$), which can be explained by using a script. As for accessing the secondary display, we do not find a difference between *Fixed DISPLAY + Single USER* and *Mobile DISPLAY + Single USER* ($p > 0.05$). Although there is no significant improvement, we believe that in our short simulated interactions (<4 minutes) there were not enough instances in accessing information that could yield a significant effect. Nonetheless, the number of walking movements over the course of a longer surgery will add up to further information access events. Finally, as Figure 6 shows, participants “zoom” into the view differently: when the screen is mobile, participants place the view at the beginning of the intervention comfortably and thus this information is always available, reducing (or almost eliminating) the need to “zoom” at the moment when the information is needed. Contrarily, in *Fixed DISPLAY + Single USER*, participants need to walk over (“zoom”) to see the information throughout the task.

5.5 Perceived Usefulness

In their responses to the post-task questionnaire (Figure 7), the participants favoured the mobilisation and positioning of the secondary display near the operating table, which they considered

to be useful both for the transmission of information (83% of responses) and for the performance of the task (75% of responses). The results concerning the virtual supports to the participant’s deictic communication are more contrasted. The virtual pointer was mostly found useful for transmitting information (92% of responses) and for performing the intervention in general (83% of responses), unlike the annotations, which 42% of participants found neither useful for transmitting information nor for carrying out the intervention (50%). According to them annotations nevertheless prove their usefulness for transmitting information (92% of responses) and performing the intervention when they are produced by the remote expert.

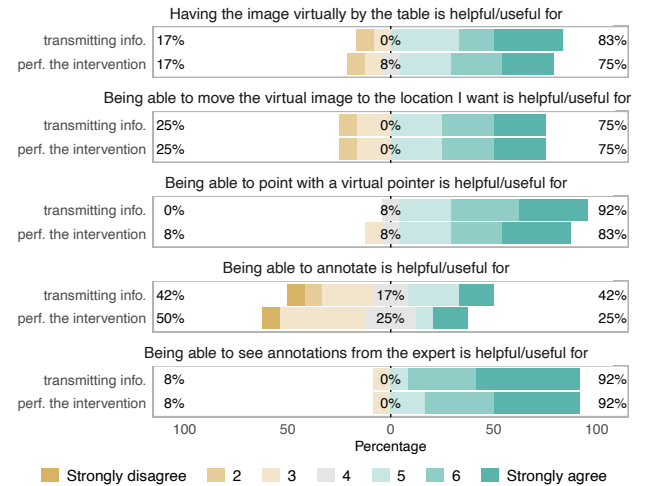


Figure 7: Participant perceived usefulness.

6 DISCUSSION

Over a decade ago, Mentis and colleagues [34] began a series of studies of intraoperative imaging systems in surgery to better understand collaborative image interaction practices and the design of interactive mechanisms for co-constructing knowledge in such expert practices. First, studying touchless input modalities to interact with images (e.g., navigate layers of an MRI), such as voice and hand gestures [32]. Then, exploring gesturing tools for collaborative work and training, showing that pointers and annotations support an expert’s need for referencing [15] and can guide a novice’s gaze as intended [16] in collocated settings. In remote settings, Mentis and colleagues showed that gesturing tools can support the practice of crafting of the view [31], and can even lead to higher

quality of instruction than in collocated settings [40], although as virtual annotations are used as main sources of information, they can also be detrimental to communication [39]. Lastly, Mentis and colleagues explored new interaction techniques for remote instruction in surgery through AR, that do not require freezing the live video to produce annotations [30, 42]. We insert our work in this line of research, investigating AR-HMDs as a means for collaborative discussions in surgical teleconsulting through shared displays. Our investigation focused on communication efficiency and quality including how the structure and content of communication may be enhanced when using AR-HMDs to overcome two identified problems in surgical telemedicine: separation of secondary imaging systems from the operative field [23, 29, 33] and single-user annotation tools [15, 38]. In our analysis, we observe a few significant improvements in communication practices that lend evidence to the benefits gained from the application of AR-HMD, mostly from the ability to reposition images and render information available for discussion at times of need. However, we also see little to no effects in other areas, notably communication balance through turn taking, that prior research findings would have led us to believe there would have been a much stronger effect [15]. We now discuss the factors that may have influenced the study results on communication, then, the benefits on remote collaborative work of both being able to reposition images and annotate, and, finally, future directions on understanding further AR-HMDs and designing augmented imaging systems.

6.1 Influence of Expert–Expert and Structured Communication

As we reflect on our findings, a clear aspect of our study design that stands out in contrast to prior work is the employment of *peers* engaging in collaborative discussions as opposed to, for instance, a mentor–mentee relationship that has been the focus of many of the surgical telemedicine studies in the past [15, 19, 35]. Indeed, contrary to prior work, the goal of discussion between remote and local in our study is not remote instruction, nor learning, but rather on meaning-making of a real-world situation through discussion with an expert. We believe two factors influenced our results. First, the communication we observe was between experts of *different* fields, a surgeon and a radiologist, pulling from multiple sources to make one surgical body as a construction, each bringing different skills to the table. The radiologist is the interpreter of the image, they communicate their ability to see the body. The surgeon is the actor on the body based on interpretations they make with the radiologist. As participants were field experts, they relied on collective common ground, as the field has developed an extensive codified vocabulary designed to precisely communicate structures and locations (e.g., “the breast tumour is not retroareolar but rather on the 2 o’clock ray, 3.5 cm from the nipple”). This made it difficult to study the factors influencing the effectiveness of the communication. The other factor that influenced our study is our choice of a confederate, which structured the communication. Indeed, controlling for variability of communication of a second participant acting as the remote radiologist was a way to obtain results, although it may have forced participants to adopt a dialogue structure and content that is not realistic or spontaneous.

6.2 Impact of Image Repositioning Ability

Our first research question addressed the problem of secondary images being far apart from the place of work, re-configuring teams and imposing frequent focus changes [23, 29, 33]. We observe two effects from participants being able to reposition the images wherever was most suitable through the AR-HMDs. The first effect is a reshaping of communicative interactions with the remote radiologist, as having the image next to the patient significantly reduces the need for clarification of the information presented. Relying on the least collaborative effort theory [10], this can be explained by the fact that to make up for the cost of travel, surgeons will wait to accumulate information needs before walking to the wall-screen and then repeatedly ensure that their understanding is effective before returning to the patient. We can imagine that this cost would be amplified in real surgical situations, where the environment is much stricter as movements are limited by aseptic constraints and space shared with members of the surgical team, with long procedures and with a dynamic operating field where organ exposure must remain constant. Moreover, we observed a significant increase in decision making, as participants performed more management acts. This can be explained by better access to the shared display, which leads to a better individual contribution to the solution of the common problem [21]. The second effect we observe is surgeons redefining the timing of their interactions with the secondary display. By choosing an optimal location for the particular task at hand, surgeons can have easy access to information throughout the operation, avoiding the ebbing from table to image which leads to constant shifting of their focus. The chosen location is influenced by the position of the patient, the size of the structures to be analysed and certain technical constraints such as contrast [33]. Although ceiling-mounted displays could in principle achieve similar benefits, the constant change in surgeon’s needs during hours-long surgery makes them less adequate, as changing their location requires manual positioning and adjustments that comes in conflict with hand sterility. These two benefits did not translate into increased efficiency, however, they are indicative of efficiency gains that are additive in longer and particularly more complex surgical teleconsulting scenarios.

6.3 Impact of Dual-User Input Tools

Our second research question addressed the imbalanced communication given single-user input tools [15]. The AR-HMD provided the ability for the operating surgeon to *also* annotate the image as well as the physical space during discussions. We hypothesized that a dual-user gesturing tool would improve communication as gesturing tools have been shown to support the development of common ground [17], to improve grounding [26], and to increase task performance [18]. Even if we observe a significant increase in deictic communication when surgeons can annotate, we did not find changes in the communication structure nor content. This can be explained by the presence of the virtual pointer in both conditions *Mobile DISPLAY + Single USER* and *Mobile DISPLAY + Dual USER*. Indeed, in the context of our experiment, involving the performance of short duration tasks where participants were mainly asked to show structures on a static surface, the virtual pointer may be easier to use because it does not require special training.

One of the main advantages of annotations is their persistence over time compared to a virtual or physical pointer. When we look at the communicative functions that encompassed the use of the annotations in the *Mobile DISPLAY + Dual USER* condition, the need to have persistence when clarifying, checking, and laying down a judgement is evident. These communicative acts indicate that, in the case of peer-to-peer consultation, ensuring there is no ambiguity, might thus lead one to use an annotation instead of a pointer. This is despite the fact that pointing, particularly with one's figure in the video field, is much for efficient to enact. Thus the benefit of engaging the annotation tool and slowing down one's movements to ensure accuracy outweighs that of simply pointing and adding further verbal clarifications if need be.

6.4 Future Work

We now open perspectives for future work, both towards further studying AR-HMDs in supporting collaborative discussions during surgical teleconsulting, and towards advancing the exploration in HCI on designing imaging systems through using augmented reality.

6.4.1 Further Exploration of AR-HMDs in Surgical Teleconsulting. Future work can complement our findings in several ways. First, expert-expert communication seems to have a different structure and content than between mentors and a mentees. Indeed, as Feng et al. [15] point out, communication structure fundamentally changes when it involves an expert and a novice, with fewer involvement in speaking and decision making for the novice. It would therefore be interesting to conduct comparative studies between peers and mentor-mentee to understand how it impacts communication. Second, although the use of a specific script through a confederate enabled us to control the flow of communication and thus to obtain participant behaviour data that is comparable, this constrained communication more than expected. Therefore, some of our observations regarding communication structure and content may be impacted by the script, and it would be interesting to carry out studies that use two participants as remote and local, to break from this constraint. Finally, we believe that interacting with displays on a head-mounted rather than a wall-mounted device has impacts worth of study, as typically the introduction of new technologies into surgery has unintended consequences to existing social dynamics [5]. One direction for future work is to study proxemics, as AR-HMDs eliminate the need of a clear line of sight and thus potentially the reconfiguration of surgical teams it encompasses [33]. Another direction is to study the shift to single-user *access* displays, as now only one person sees the imaging system. The act of "breaking off" from the patient to interact with a far screen conveys a signal of distributed articulation work occurring [29]. Therefore, as team members are not aware anymore of such interactions, there will most likely be a loss in consequential communication. This may translate for instance in hindering the anticipation of a surgeon's next move, critical in surgery.

6.4.2 Towards Augmented Imaging Systems in Surgery. We conclude our work with three directions to move forward imaging systems through AR, for collaborative purposes but not exclusively. The first direction is further augmenting both input and output.

Typically, imaging systems in operating rooms are displayed on a wall-mounted monitor, with mouse and keyboard interaction. Our study showed benefits of AR-HMDs in augmenting output, by enabling the repositioning of images, as well as input, by supporting mid-air hand interaction without the need for installing new sensing technology into the OR as it is the case of current systems for touchless interaction [27, 32, 34]. Augmenting input can be taken a step further, supporting more advanced mechanisms to place images, for example the definition of an ego-centered frame where the image is fixed to a surgeon as they move, or supporting predefined positions with respect to the patient body. Likewise, augmenting output can also be improved. Our approach has an underlying assumption that both remote expert and local surgeon need symmetric functionalities, both a pointer and annotations, yet their task-related needs are asymmetrical. Future work can study how to best support the local surgeon's needs by implementing different output visualizations (e.g., instruments such as a scalpel when instructing to cut). AR-HMDs are a promising playground in this exploration, as they are a powerful computational media for providing complex and dynamic output visualizations. These future avenues of research can expand beyond telecollaboration (during surgery), for instance supporting visualization of past events in mixed reality during post-operative debriefing [28].

The second direction is studying how to support a multi-device ecology. Operating rooms contain a multitude of systems displayed on different screens, and aimed at different goals [29]. Future work can take advantage of AR-HMDs capabilities to show multiple displays, and study whether the effects we found hold when surgeons have on-demand access to multiple imaging systems. Lastly, we see ample space for the development of systems that integrate the virtual and real worlds into one coherent space. In our study, it became evident that the current implementation leads to two disjoint worlds which breaks the flow of collaborative work. From the participant side, we observed struggle to recover from loose-coupled integration with the real world. It happened often that annotations produced on the virtual screen remained anchored to the real world and not to the image, thus, when the local surgeon moved the image, the annotations were not aligned anymore. This can lead to misunderstanding of information and a cost in terms of loss of time from having to detect the problem and correct it. From the confederate side, the systems requires taking a screen capture in order to annotate. This lead the confederate to anticipating the need for annotations and "freezing" the view early to keep the conversation fluid. One possible way forward is to show a reconstruction of the environment to the remote expert, where they can navigate the view independently [14], and produce annotations without the need to freeze the view. Technically this approach has been achieved, notably by Gauglitz et al. [20], but the benefits of this approach have not yet been studied.

7 LIMITATIONS

First, the limited number of participants might have prevented us from observing further significant effects, which currently occur as trends in our data. Second, the operationalization of the task resulted in the simplification of many complexities involved in hours-long surgery, that may have limited observable effects. Trials

were focused and thus short, limiting the amount of times participants needed to reposition the display and annotate, therefore, to reflect on their usefulness. The physical support used to simulate the patient certainly enabled participants to contextualize the task, but did not embody the dynamic nature of an intervention: the changing needs for exposures inside the patient change, the unintended movement of tissues as patients breathe, and the changing nature of anatomy as it is manipulated and dissected. We believe that dual-user input tools whose main advantage is their anchoring in the real world can have further impacts on a dynamic surgical field. Lastly, our task did not involve neither strict asepsis nor other actors involved in surgery (e.g., an assistant, nurses and the anaesthetist team), which requires organization and anticipation of their movements in a restricted space. Image repositioning capabilities may have greater benefits under these conditions where traveling must be minimized. Nonetheless, carrying out studies that get closer to real-life surgery conditions and to the stakes of remote communication in the OR is not an easy task. Studying these effects can also be done through clinical trials, which provides ecological validity by is certainly of more complex elaboration.

ACKNOWLEDGMENTS

This work was supported by the French state funds managed by the ANR (Agence Nationale de la Recherche) within the program “Investissements d’Avenir” under references ANR-11-LABX-0004 (Labex CAMI) and ANR-17-CE19-0023 (ROBUST). We thank all the participants for their time and dedication to our study. We also thank the *Chaire d’Innovation BOPA (Bloc OPérateur Augmenté)* for allowing us to conduct this research in their facilities.

REFERENCES

- [1] Rahul Agarwal, Adam W. Levinson, Mohamad Allaf, Danil Makarov, Alex Nason, and Li-Ming Su. 2007. The RoboConsultant: telerobotics and remote presence in the operating room during minimally invasive urologic surgeries using a novel mobile robotic interface. *Urology* 70, 5 (Nov. 2007), 970–974. <https://doi.org/10.1016/j.urolgy.2007.09.053>
- [2] Daniel Andersen, Voicu Popescu, Maria Eugenia Cabrera, Aditya Shanghavi, Gerardo Gomez, Sherri Marley, Brian Mullis, and Juan P. Wachs. 2016. Medical telerobotics using an augmented reality transparent display. *Surgery* 159, 6 (June 2016), 1646–1653. <https://doi.org/10.1016/j.surg.2015.12.016>
- [3] Steven Arild Wuyts Andersen, Peter Trier Mikkelsen, Lars Konge, Per Cayé-Thomasen, and Mads Sølvsten Sørensen. 2016. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *The Laryngoscope* 126, 2 (2016), E74–E79.
- [4] R Anthony Jr et al. 2008. Cognitive load theory and the role of learner experience: An abbreviated review for educational practitioners. *AACE Review (formerly AACE Journal)* 16, 4 (2008), 425–439.
- [5] Ignacio Avellino, Gilles Bailly, Geoffroy Canlorbe, Jérémie Belgiht, Guillaume Morel, and Marie-Aude Vitrani. 2019. Impacts of Telemanipulation in Robotic Assisted Surgery. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 583:1–583:15. <https://doi.org/10.1145/3290605.3300813> event-place: Glasgow, Scotland, UK.
- [6] Ignacio Avellino, Sheida Nozari, Geoffroy Canlorbe, and Yvonne Jansen. 2021. Surgical Video Summarization: Multifarious Uses, Summarization Process and Ad-Hoc Coordination. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 140:1–140:23. <https://doi.org/10.1145/3449214>
- [7] H. Borgmann, M. Rodríguez Socarrás, J. Salem, I. Tsaur, J. Gomez Rivas, E. Barret, and L. Tortolero. 2017. Feasibility and safety of augmented reality-assisted urological surgery using smartglass. *World Journal of Urology* 35, 6 (June 2017), 967–972. <https://doi.org/10.1007/s00345-016-1956-6>
- [8] Caroline GL Cao, Mi Zhou, Daniel B Jones, and Steven D Schwaizberg. 2007. Can surgeons think and operate with haptics at the same time? *Journal of Gastrointestinal Surgery* 11, 11 (2007), 1564–1569.
- [9] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511620539>
- [10] Herbert H. Clark and Susan E. Brennan. 1991. Grounding in communication. In *Perspectives on socially shared cognition*, L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.). American Psychological Association, Washington, DC, US, 127–149.
- [11] Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 1 (April 1960), 37–46. <https://doi.org/10.1177/001316446002000104> Publisher: SAGE Publications.
- [12] Gregorio Convertino, Helena M Mentis, Aleksandra Slavkovic, Mary Beth Rosson, and John M Carroll. 2011. Supporting common ground and awareness in emergency management planning: A design research project. *ACM Transactions on Computer-Human Interaction (TOCHI)* 18, 4 (2011), 1–34.
- [13] Matthew Christopher Davis, Dang D. Can, Jonathan Pindrik, Brandon G. Rocque, and James M. Johnston. 2016. Virtual Interactive Presence in Global Surgical Education: International Collaboration Through Augmented Reality. *World Neurosurgery* 86 (Feb. 2016), 103–111. <https://doi.org/10.1016/j.wneu.2015.08.053>
- [14] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billingham. 2019. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 81–98. <https://doi.org/10.1016/j.ijhcs.2019.05.011>
- [15] Yuanyuan Feng, Katie Li, Azin Semsar, Hannah McGowan, Jacqueline Mun, H. Reza Zahiri, Ivan George, Adrian Park, Andrea Kleinsmith, and Helena M. Mentis. 2019. Communication Cost of Single-user Gesturing Tool in Laparoscopic Surgical Training. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 611:1–611:12. <https://doi.org/10.1145/3290605.3300841> event-place: Glasgow, Scotland UK.
- [16] Yuanyuan Feng, Hannah McGowan, Azin Semsar, Hamid R. Zahiri, Ivan M. George, Adrian Park, Andrea Kleinsmith, and Helena Mentis. 2019. Virtual pointer for gaze guidance in laparoscopic surgery. *Surgical Endoscopy* 34 (Oct. 2019), 3533–3539. <https://doi.org/10.1007/s00464-019-07141-x>
- [17] Yuanyuan Feng and Helena M. Mentis. 2018. Improving Common Ground Development in Surgical Training through Talk and Action. *AMIA Annual Symposium Proceedings* 2017 (April 2018), 696–705. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977723/>
- [18] Susan R. Fussell, Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam D. I. Kramer. 2004. Gestures over Video Streams to Support Remote Collaboration on Physical Tasks. *Hum.-Comput. Interact.* 19, 3 (Sept. 2004), 273–309. https://doi.org/10.1207/s15327051hci1903_3
- [19] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, Yuanyuan Feng, Ru Wang, Zhuoqun Robin Xu, Enrique Zavala, Yifei Zhang, Wanze Xie, Xinming Zhang, Konrad Davis, Michael Yip, and Nadir Weibel. 2021. ARTEMIS: A Collaborative Mixed-Reality System for Immersive Surgical Telerobotics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. <https://doi.org/10.1145/3411764.3445576>
- [20] Steffen Gauslitz, Benjamin Nuernberger, Matthew Turk, and Tobias Höllerer. 2014. World-stabilized annotations and virtual scene navigation for remote collaboration. In *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14*. ACM Press, Honolulu, Hawaii, USA, 449–459. <https://doi.org/10.1145/2642918.2647372>
- [21] Kori M. Inkpen, Wai-ling Ho-Ching, Oliver Kuederle, Stacey D. Scott, and Garth B. D. Shoemaker. 1999. This is fun! we're all best friends and we're all playing: supporting children's synchronous collaboration. In *Proceedings of the 1999 conference on Computer support for collaborative learning (CSCW '99)*. International Society of the Learning Sciences, Palo Alto, California, 31–es.
- [22] Ellen A. Isaacs and Herbert H. Clark. 1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General* 116, 1 (March 1987), 26–37. <https://doi.org/10.1037/0096-3445.116.1.26>
- [23] Rose Johnson, Kenton O'Hara, Abigail Sellen, Claire Cousins, and Antonio Criminisi. 2011. Exploring the potential for touchless interaction in image-guided interventional radiology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Vancouver BC Canada, 3323–3332. <https://doi.org/10.1145/1978942.1979436>
- [24] Seungwon Kim, Gun Lee, Mark Billingham, and Weidong Huang. 2020. The combination of visual communication cues in mixed reality remote collaboration. *Journal on Multimodal User Interfaces* 14, 4 (Dec. 2020), 321–335. <https://doi.org/10.1007/s12193-020-00335-x>
- [25] Seungwon Kim, Gun A. Lee, Nobuchika Sakata, Andreas Dunser, Elina Vartiainen, and Mark Billingham. 2013. Study of augmented gesture communication cues and view sharing in remote collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Adelaide, Australia, 261–262. <https://doi.org/10.1109/ISMAR.2013.6671795>
- [26] David Kirk, Tom Rodden, and Danaë Stanton Fraser. 2007. Turn it this way: grounding collaborative action with remote gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 1039–1048. <https://doi.org/10.1145/1240624.1240782>
- [27] Naveen Madapana, Daniela Chanci, Glebys Gonzalez, Lingsong Zhang, and Juan P. Wachs. 2022. Touchless Interfaces in the Operating Room: A Study in Gesture

- Preferences. *International Journal of Human-Computer Interaction* 0, 0 (April 2022), 1–11. <https://doi.org/10.1080/10447318.2022.2041896>
- [28] Sophie Maria, Solène Lambert, and Ignacio Avellino. 2022. From Déjà vu to Déjà vécu: Reliving Surgery in Post-Operative Debriefing. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE Computer Society, Los Alamitos, CA, USA, 462–465. <https://doi.org/10.1109/VRW55335.2022.00102>
- [29] Helena M. Mentis. 2017. Collocated Use of Imaging Systems in Coordinated Surgical Practice. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW (Dec. 2017), 78:1–78:17. <https://doi.org/10.1145/3134713>
- [30] Helena M. Mentis, Ignacio Avellino, and Jwawon Seo. 2022. AR HMD for Remote Instruction in Healthcare. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE Computer Society, Los Alamitos, CA, USA, 437–440. <https://doi.org/10.1109/VRW55335.2022.00096>
- [31] Helena M. Mentis, Yuanyuan Feng, Azin Semsar, and Todd A. Ponsky. 2020. Remotely Shaping the View in Surgical Telementoring. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–14. <https://doi.org/10.1145/3313831.3376622>
- [32] Helena M. Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. 2015. Voice or Gesture in the Operating Room. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 773–780. <https://doi.org/10.1145/2702613.2702963>
- [33] Helena M. Mentis, Kenton O'Hara, Abigail Sellen, and Rikin Trivedi. 2012. Interaction Proxemics and Image Use in Neurosurgery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 927–936. <https://doi.org/10.1145/2207676.2208536>
- [34] Helena M. Mentis and Alex S. Taylor. 2013. Imaging the body: embodied vision in minimally invasive surgery. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris France, 1479–1488. <https://doi.org/10.1145/2470654.2466197>
- [35] Edgar Rojas-Muñoz, Dan Andersen, Maria Eugenia Cabrera, Voicu Popescu, Sherri Marley, Ben Zarzaur, Brian Mullis, and Juan P. Wachs. 2019. Augmented Reality as a Medium for Improved Telementoring. *Military Medicine* 184, Suppl 1 (March 2019), 57–64. <https://doi.org/10.1093/milmed/usy300>
- [36] Jeremy Roschelle and Stephanie D. Teasley. 1995. The Construction of Shared Knowledge in Collaborative Problem Solving. In *Computer Supported Collaborative Learning*, Claire O'Malley (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 69–97. https://doi.org/10.1007/978-3-642-85098-1_5
- [37] Abigail J. Sellen. 1995. Remote Conversations: The Effects of Mediating Talk with Technology. *Hum.-Comput. Interact.* 10, 4 (Dec. 1995), 401–444. https://doi.org/10.1207/s15327051hci1004_2
- [38] Azin Semsar, Hannah McGowan, Yuanyuan Feng, Hamid R. Zahiri, Ivan M. George, Timothy Turner, Adrian Park, Helena M. Mentis, and Andrea Kleinsmith. 2019. Effects of a Virtual Pointer on Trainees' Cognitive Load and Communication Efficiency in Surgical Training. *AMIA ... Annual Symposium proceedings. AMIA Symposium* 2019 (2019), 1197–1206.
- [39] Azin Semsar, Hannah McGowan, Yuanyuan Feng, H. Reza Zahiri, Adrian Park, Andrea Kleinsmith, and Helena Mentis. 2019. How Trainees Use the Information from Telepointers in Remote Instruction. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW (Nov. 2019), 93:1–93:20. <https://doi.org/10.1145/3359195>
- [40] Azin Semsar, Hannah McGowan, Yuanyuan Feng, H. Reza Zahiri, Adrian Park, Andrea Kleinsmith, and Helena M. Mentis. 2020. Quality of and Attention to Instructions in Telementoring. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 165:1–165:21. <https://doi.org/10.1145/3415236>
- [41] A Semsar, J Ton, N Maharoof, I Avellino, HR Zahiri, F Guckes, and H Mentis. 2021. Effect of Training the Mentor on Quality of Instruction and Trainees' Performance in Laparoscopic Oophorectomy Telementoring. *Journal of Minimally Invasive Gynecology* 28, 11, Supplement (Nov. 2021), S60–S61. <https://doi.org/10.1016/j.jmig.2021.09.452>
- [42] Jwawon Seo, Ignacio Avellino, Damaruka Priya Rajasagi, Anita Komlodi, and Helena M. Mentis. 2022. HoloMentor: Enabling Remote Instruction through Augmented Reality Mobile Views. *Proceedings of the ACM on Human-Computer Interaction* 7, GROUP (Dec. 2022), 11:1–11:29. <https://doi.org/10.1145/3567561>
- [43] Mahesh B. Shenai, Marcus Dillavou, Corey Shum, Douglas Ross, Richard S. Tubbs, Alan Shih, and Barton L. Guthrie. 2011. Virtual interactive presence and augmented reality (VIPAR) for remote surgical assistance. *Neurosurgery* 68, 1 Suppl Operative (March 2011), 200–207; discussion 207. <https://doi.org/10.1227/NEU.0b013e3182077efd>
- [44] Angelina M. Vera, Michael Russo, Adnan Mohsin, and Shawn Tsuda. 2014. Augmented reality telementoring (ART) platform: a randomized controlled trial to assess the efficacy of a new surgical education technology. *Surgical Endoscopy* 28, 12 (Dec. 2014), 3467–3472. <https://doi.org/10.1007/s00464-014-3625-4>

A PRE-TASK QUESTIONNAIRE

Before starting each trial, the participants were asked about their general characteristics (age, gender) and then about their surgical specialty, place of practice and number of years of experience since their first year of residency (open questions). Finally, they reported their past experience in augmented reality with options ranging from never, once a year, once a month, several times a week, once a week to every day.

B POST-TASK QUESTIONNAIRE

Questions are answered in a likert scale from 1–7.

- Having the image virtually by the table is
 - Useful for performing the intervention
 - Helpful for transmitting information
- Being able to move the virtual image to the location I want is
 - Useful for performing the intervention
 - Helpful for transmitting information
- Being able to point with a virtual pointer is
 - Useful for performing the intervention
 - Helpful for transmitting information
- Being able to annotate is
 - Useful for performing the intervention
 - Helpful for transmitting information
- Being able to see annotations from the expert is
 - Useful for performing the intervention
 - Helpful for transmitting information

Then, three open questions were asked, regarding:

- Benefits and challenges of being able to move the image next to the patient?
- Benefits and challenges of making annotations?
- Benefits and challenges of seeing expert annotations?