Communication Cost of Single-user Gesturing Tool in Laparoscopic Surgical Training

Yuanyuan Feng
University of Maryland, Baltimore County
Baltimore, Maryland
fengy1@umbc.edu

Katie Li
Pomona College
Claremont, California
katiesli16@gmail.com

Azin Semsar
University of Maryland, Baltimore County
Baltimore, Maryland
asemsar1@umbc.edu

Hannah McGowan
University of Maryland, Baltimore County
Baltimore, Maryland
ha24@umbc.edu

Jacqueline Mun
Vassar College
Poughkeepsie, New York

H. Reza Zahiri
Anne Arundel Medical Center
Annapolis, Maryland
hzahiri@aahs.org

Ivan George
Johns Hopkins Medicine
Baltimore, Maryland
igeorge2@jhmi.edu

Adrian Park
Anne Arundel Medical Center
Annapolis, Maryland
apark@aahs.org

Andrea Kleinsmith
University of Maryland, Baltimore County
Baltimore, Maryland
andreak@umbc.edu

Helena M. Mentis
University of Maryland, Baltimore County
Baltimore, Maryland
mentis@umbc.edu

ABSTRACT

Multi-user input over a shared display has been shown to support group process and improve performance. However, current gesturing systems for instructional collaborative tasks limit the input to experts and overlook the needs of novices in making references on a shared display. In this paper, we investigate the effects of a single-user gesturing tool on the communication between trainer and trainees in a laparoscopic surgical training. By comparing the communication structure and content between the trainings with and without the gesturing tool, we show that the communication becomes more imbalanced and the trainees become less active when using the single-user gesturing tool. Our findings highlight the needs to grant all parties the same level of access to a shared display and suggest further directions in designing a shared display for instructional collaborative tasks.

CCS CONCEPTS

• **Human-centered computing → Computer supported cooperative work**.

KEYWORDS

Shared display, instructional collaborative tasks, team communication, common ground, surgical training, turn-taking, communication content

ACM Reference Format:
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 *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300841

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300841

1 INTRODUCTION

There is long-standing interest in the human-computer interaction community in designing shared displays to support group processes in different task settings [3, 28, 34, 36, 47]. Previous research, both on vertical wall-sized displays and tabletops, demonstrates the importance of referencing gestures over the shared displays for group communication and suggests multi-user input as a key feature for the design of shared displays [26, 41, 48]. These studies often focused on peer collaboration, where all users make similar amount of physical contributions to the tasks. Yet, we have limited understanding on gesturing over a shared display in instructional collaborative tasks, where a novice is primarily performing the task on a display, monitored and guided by an expert, such as in simulation training, remote repairing, and minimally invasive surgery [11, 22]. In these tasks, experts mainly supply expertise, or in-situ knowledge, for the novices to accomplish the task, without physically contributing to the task itself.

Most current gesturing tools for instructional collaborative tasks are single-user input, enabling only the experts to interact with a display [12, 21, 23, 37]. These tools have been demonstrated to improve task performance and communication efficiency [10, 11, 21]. However, much less discussion has been on the ability of novices in gesturing over a shared display. Although novices may use task-related actions to substitute their speech, these actions are mainly for acknowledging the acceptance of a statement [13]. In contrast, novices prefer to use gestures to present a piece of information and make contributions to the group decision making [1]. In addition, compared to experts, who tend to articulate the task objects, novices are more likely to describe the location of the objects when making a reference [19].

When a single-user gesturing tool was provided in an instructional task for experts to use, the novices' language use was found to decrease [21]. On one hand, it indicates improved communication efficiency [5]. On the other, it may reflect the reduced participation of novices in the group process [25]. To build tools to support the instructional collaborative tasks over shared displays, we need to elucidate the process in which the current single-user gesturing tools affect the language use of the novices. Although previous research has demonstrated that novices raised fewer questions with more directive instructions [21], we have limited information on these questions themselves. For example, do these questions contain any new information? Is there any embedded knowledge elicited by these questions? Why are these questions reduced? A more detailed understanding of the communication process allows us to comprehensively evaluate the impact of single-user gesturing tools on group communication in instructional collaborative tasks, as well as make informed design decisions on designing the interactive shared display.

In this study, we investigate the potential communication costs - the efforts for speakers to take over the floor by formulating and producing their communicative acts [5] - incurred by the use of a single-user gesturing tool in an instructional collaborative task. Previous studies have shown that experts and novices co-constructed the knowledge through hand gestures over the display [29, 30, 32]. In our study, we provide a gesturing tool to the experts on top of the standard interactions to identify any changes in the structure and content of the team communication. With a thorough examination in the communication process, our study reveals that the communication became less balanced with reduced active participation from the novices when using the single-user gesturing tool in the instructional collaborative task. Based on our findings, we discuss the design directions on supporting equal communicative access to the shared display.

2 RELATED WORKS

In this section, we first present the theories that lay the ground for the study. We then identify the knowledge gap by comparing current studies in shared displays for peer collaboration with collaboration between experts and novices. Finally, we present the context of our study and state our hypotheses.

Common Ground

Communication is a two-way process, where both parties coordinate to contribute. The success of coordination is based on the development of common ground - mutual knowledge, beliefs, goals, and assumptions [4]. In group collaboration, a group engages in joint activities, which can be partitioned into two sets of actions: the basic joint activity, which is the work the team is trying to do, and the coordinating joint actions, which consist of the communicative acts required to establish and maintain the common ground [6].

Grounding, a collective process of establishing common ground, generally consists of two phases - presentation phase and acceptance phase [5]. When the addressee's acceptance of the speaker's presentation is registered, the common ground is achieved [5, 33]. The grounding process follows the Principle of Least Collaborative Effort, which states that participants try to minimize their collaborative efforts in communication [5].

Studies in understanding the grounding process in a cooperative work examined the verbal conversations in a team, who collaborate and communicate in cooperative work [7, 9]. The researches mainly focused on the communication structure and the communication content. The communication structure was measured based on turn-taking [9, 40]. The

changes in the coordination of turn-taking, on one hand, indicate the efficiency of communication [40], and on the other, relate to the grounding costs, such as the costs in language processes, i.e., the construal of meaning, and the costs in signaling and accepting [14]. For example, more turns, fewer words, and more synchronicity manifest in teams with an increased amount of common ground, and thus more efficient communication [9, 40]. The communication content was examined through a content analysis of task-related dialogue acts between team members [8, 9]. The changes in dialogue patterns elucidate how individual team members contribute to the grounding process. For instance, team members use query-reply dialogue acts to explicitly build the understanding of when content is needed, as well as management acts to share rules on how to run the task [9].

Shared Displays for Peer Collaboration

The broad direction in designing interactive shared displays for collaborative work is stemmed from the concept of single display groupware (SDG), which supports co-located users using multiple input devices to share knowledge in accomplishing a task over a single display [46]. Compared to traditional computer workstations with a single input and output channel, SDG emphasizes collective contributions from individual team members through a shared user interface [46].

Previous researches have evaluated the impacts of multi-user input of SDG on peer collaboration, where team members shared similar levels of expertise [18, 38, 39, 42]. Compared to single-user input, multi-user input allows for more concurrent and sequential interactions among team members and supports efficient and rapid information sharing, manifested by an increase in the frequency of team verbal activities [42]. Besides, multi-user input provides team members equal access to the task, which not only encourages individual team members' contribution to the group problem solving process [18], but also facilitates team members requests for help from each other [39]. The improved communication process and team participation by multi-user interaction leads to increased team and individual performance - the task completion time is reduced while the individual team members' task-related ability and skills are enhanced [38, 42].

Shared Displays for Instructional Collaborative Tasks

Although multi-user input has been a key feature of shared displays in peer collaboration, it is scarcely adopted in instructional collaborative tasks [46]. In contrast, general attention around instructional collaborative tasks has been on designing single-user input to support experts in providing instructions to novices [12, 21, 23, 37]. These studies aimed at improving team performance through providing more explicit instructions to support the grounding process between experts and novices [1, 11, 17, 20].

However, communication is a collective process, where common ground is developed by contributions from both experts and novices [5]. Previous studies have demonstrated that it is seeing the novices' actions and relating them to the task context that enhances the development of common ground [13, 21, 22]. These actions are often deliberately designed to convey information, judgments and understandings [15, 31]. In instructional collaborative tasks, novices prefer using actions instead of speech to communicate [13]. Thus, we argue that the use of single-user gesturing tool is not sufficient for efficient team communication. Kirk et al. has shown that with a gesturing tool, experts obtained more control of the communication. In this paper, we investigate how single-user gesturing tool affect novices' contributions to the grounding process.

Hypotheses

The overall goal of this study is to expand the understanding of the impacts of a single-user gesturing tool on the team communication process, through identifying the costs for novices in making the contributions. In this, we conducted an experimental study comparing instructional collaboration with and without a single-user gesturing tool in co-located laparoscopic surgical training.

In laparoscopic surgical training, the trainers and trainees are required to coordinate around shared view of the work via laparoscopic video. The trainees manipulate the surgical instruments and perform the task, assisted by the trainers holding the camera to capture the view of the task object. The trainers monitor the process and provide guidance and feedbacks. Thus, the trainers and trainees have similar access to the shared display - they both can point and gesture over the display. A unique access for the trainees is that they can manipulate the task object through the surgical instruments. A unique access for the trainers is that they interact with the operative field through the control of the camera. Yet, these two unique accesses are interdependent - the trainers' camera should chase the instrument movements for the trainees to see their actions; and the range of trainees' instrument movements depends on the field of view the trainers captured.

In this study, we provided the single-user gesturing tool to the trainers on top of all the other possible interactions the trainers can make with the display or the trainees. We scrutinized into the speech and actions of both parties at the communication structure level and the dialogue act level in their grounding process, as well as examine the communication changes in the training context. Our hypotheses are:

H1: Novices in the single-user gesturing tool supplemented trainings will take less turns than those in the trainings without the tool.
H2: Novices in the single-user gesturing tool supplemented trainings will exhibit less judgment and decision statements and more acknowledgements than those in the trainings without the tool.

3 TECHNOLOGY IN USE

In the study, we used a lab-developed video pointing and annotation tool, Virtual Pointer (VP), as the single-user gesturing device. This tool was specially designed to enable trainers to point or draw on laparoscopic videos for the trainees to see [10]. The Microsoft Kinect sensor version 2 (Microsoft Corporation, USA) was used as a mechanism of touchless interaction - enabling the system to be used in the sterile operating field. The application is a transparent window that can be overlaid on any screen or other application. It uses a combination of audio keywords and hand movements to call upon the different functionalities, such as a pointer for referencing or a freehand drawing tool.

Figure 1 shows the interface of the system. The collection of the verbal commands is showing in the upper left corner as a reminder. The present function is presented in the center above the laparoscopic view. The lower left corner shows the user's skeleton to provide timely feedback of the user's movement. To awaken the Kinect, the first command is verbally saying "Kinect ready". When this is said, the Kinect starts detecting other verbal cues and gestures. There are two verbal cues the Kinect is looking for, either "Kinect draw" or "Kinect point", to switch between the drawing mode and the pointing mode. In the pointing mode, the user moves the hand to control a small green circle which acts as a pointer. In the drawing mode, the user makes a fist to sketch over the video. The position of the pointer and the drawing responds to the position of the user's hand. To clear the screen of all annotations, the verbal command "Kinect clear" should be said to the Kinect. When the program is finished being used, the voice command "Kinect close" can be used at any time to set the program to sleep and stop the Kinect from detecting.

4 STUDY METHODS

Experimental Design

The experimental design is a 2×4 (training conditions and tasks) counter-balanced, within-subject design. We performed a controlled experiment with two training conditions - a Standard training condition as the control and a Virtual Pointer-supplemented (VP) training condition as the intervention. In the Standard condition, trainer instruction was conducted as it would be normally, through verbal or hand gestures. In the VP condition, the Virtual Pointer application was used by the

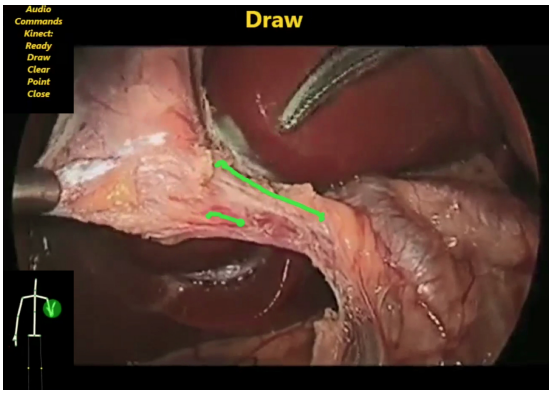

Figure 1: The interface of the Virtual Pointer.

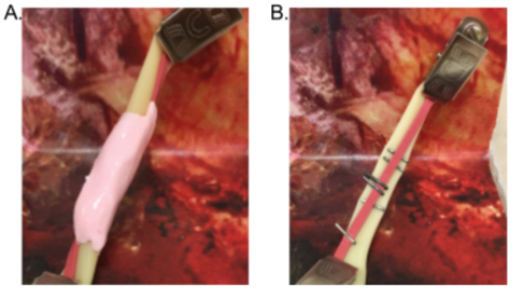

Figure 2: Examples of simulated laparoscopic tasks. (A: mobilizing cystic duct and cystic artery; B: cutting cystic duct and cystic artery)

trainers as an addition to standard guidance. The trainees worked on four simulated laparoscopic tasks under the trainers' guidance. The tasks were selected based on a hierarchical task analysis of the laparoscopic cholecystectomy procedure [35] and confirmed by an attending surgeon that they were of similar difficulty levels and required both skills of anatomical structure identification and instrument manipulation. The tasks were performed on a validated laparoscopic training physical model [43], including (1) mobilizing the cystic duct and the cystic artery, (2) clipping the cystic duct, (3) clipping the cystic artery, and (4) cutting the cystic artery and the cystic duct (Figure 2). The orders of the mentoring approaches and tasks are counterbalanced by constructing a Latin square[27]. The experiment was video recorded and the operative field was screen recorded.

Study Setup

The study was conducted at a simulation center of the Department of Surgery in a private hospital in Mid-Atlantic US. The study setup is shown in Figure 3. The tasks were performed on the Stryker Corporation's Park Trainer^{TM}, a surgical simulation training system comprised of a cart for components of the system, a training module that simulates

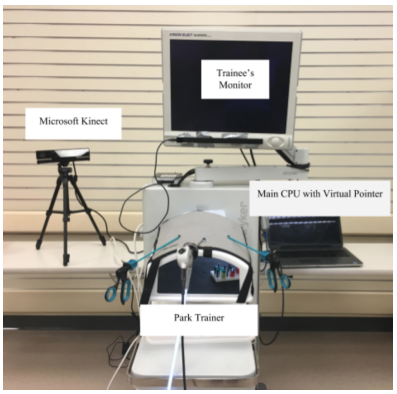

Figure 3: Study setup with the Virtual Pointer system and Park Trainer.

a body cavity, and a skin or cover for the training module for use in laparoscopic surgery. The trainee stood in the center performing the tasks. The Microsoft Kinect sensor was placed on the left of the Park Trainer, where the trainer would be standing, controlling the laparoscope with his right hand and using gestures to point or draw with his left hand.

The laparoscopic video was streamed out of the Park Trainer into the laptop that ran the VP program. The annotated video would then be feed into the monitor facing the trainee. The lag of the video was minimized to less than 1 second.

Participants

Six residents and one fellow in general surgery participated in this study as trainees. The same fellow and an attending surgeon participated as trainers. The fellow was trained by the attending surgeon, and he provided training to the six residents. The pairs' demographics, total number of communicative acts and task durations are shown in Table 1. The subjects consented to participate in the study on a voluntary basis without any monetary compensation.

Communication Structure

We examined communication structure through turn-taking analysis. We applied the adapted analysis scheme used by Sellen [44], which breaks a dialogue into turns and pauses. Since speech and actions interdependently contribute to the grounding process [2, 13, 16], we identified turns based on utterances, actions and VP use transcribed from the trials.

We focused on two major measures in turn-taking analysis: turn frequency and turn distribution. Previous studies showed that an increase in common ground led to more rapid turns [9]. The turn distribution depends on the difficulty of participants in taking the floor - the more difficult for one speaker in taking the floor, the more skewed the turns will be distributed [44]. Given that we considered both speech and actions, the duration would be unscalable for each turn. So, we did not analyze the turn duration.

Turn frequency is calculated as the number of turns per second for each trial. Turn distribution is calculated as the proportion of trainees' turns per trial.

Communication Content

We used the dialogue act coding scheme (Table 2) to examine the content of the utterances, actions and VP use between the trainers and trainees. This coding scheme was developed to understand the development of common ground among interdependent team members in managing emergent complex tasks, which required efficient information-sharing, problem-solving and decision-making[9]. It emphasizes on individual team members' communicative intentions of the dialogues acts in the grounding process, as opposed to linguistic or semantic meaning [7, 8]. In this study, we focused on the dialogue acts, Manage (MN), Judge (J) and Confirm (CO) to represent trainees' contributions in team decision-making process, and Acknowledge (AC) to represent trainees' acceptance of instructions, in order to test H2. We further explored any other possible dialogue pattern changes with this coding scheme.

We coded the communication content with the transcripts and video recordings. The first two authors viewed the video and coded the transcript independently. After the first independent coding session, the inter-rater reliability for the coders was found be to Kappa = 0.62. They negotiated for any conflicting codes and then coded again and achieved an agreement of Kappa = 0.84. This was deemed high agreement [24] and so the remainder of the cases were coded. For any disagreement between the codes in the remainder of the cases, the two coders viewed the video and discussed to achieve the agreement.

Data Analysis

The focus of the data analysis is on identifying significant changes in the communication structure and content between the trainers and trainees. Since we conducted a within-subject experiment, we used mixed models to control the task order, task difficulty level, and repeated measures. Since the turn frequency and dialogue act proportion are count data, following the Poisson distribution, we used generalized Poisson mixed model to compare between the two training conditions. The turn distribution follows binomial distribution and we fitted turn distribution in mixed effects logistic regression. Given the temporal nature in the grounding process, we first analyzed the task order as a fixed effect. If the task order has an insignificant effect on the model, we moved it to the random effect to increase the power. The analyses

Table 1: Demographics and Data Description of Training Pairs

Pair	Trainer	Trainee	Trainee Gender	Trainee Laparoscopic Experience	Total Number of Communicative Acts	Mean Task Duration (s)	Standard Deviation
1	Fellow	Resident 1	Male	3 years	111	64.24	29.30
2	Attending	Fellow	Male	5 years	71	60.46	25.11
3	Fellow	Resident 2	Male	0	139	89.63	26.84
4	Fellow	Resident 3	Male	0	111	61.13	20.30
5	Fellow	Resident 4	Male	0	125	64.85	9.29
6	Fellow	Resident 5	Male	4-5 cases	166	73.53	54.68
7	Fellow	Resident 6	Male	0	224	90.04	51.19

Table 2: Dialogue act coding scheme[9].

Class	Dialogue Act	Description
Transfer Info	Add Info (AI)	Provides new information, not elicited.
	Query (Q)	Question used to elicit new information.
	Replay (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.

were conducted using R version 3.2.2 (R Foundation for Statistical Computing, Austria). The results are shown in the graphs as mean with standard error.

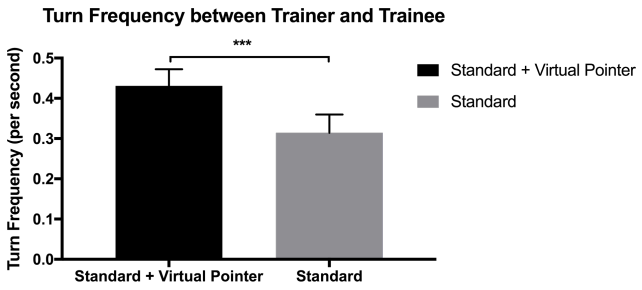

Figure 4: Comparison of Turn Frequency between Virtual Pointer condition and Standard condition. (n = 7, ***: p < .0001)

5 RESULTS

Turn-Taking

We first analyzed the turn frequency to obtain an overall evaluation of the communication process. As shown in Figure 4, with the use of VP, the turn frequency significantly increased (β= -.427, p < .0001). This result is corresponding to the findings from Kirk et al. [21] and Fussell et al. [11], indicating the use of single-user gesturing tool increases overall knowledge sharing and leads to more efficient communication [11, 21].

We further split the number of turns between the trainers and trainees and calculated each group's turn frequency. As shown in Figure 5, the use of VP significantly increased the number of turns by both trainers and trainees. The increase in the trainer's turn frequency is greater than the trainee's. This trend confirms that the VP facilitates the trainers in providing guidance[10], and indicates that the use of VP may influence the turn distribution.

To quantify the extent to which the VP influence the turn distribution, we compared the proportion of trainees' turns between the two training conditions. As shown in Figure 6, there is a significant decrease in trainees' turn proportion in the VP condition, compared to the Standard condition (β= -.670, p = .032). Ideally, the trainees' turn proportion should be 0.5, representing that they have taken the floor

Figure 5: Comparison of Turn Frequency between Virtual Pointer condition and Standard condition in trainer and trainee groups. (n = 7, *: p < .05, ***: p < .0001)

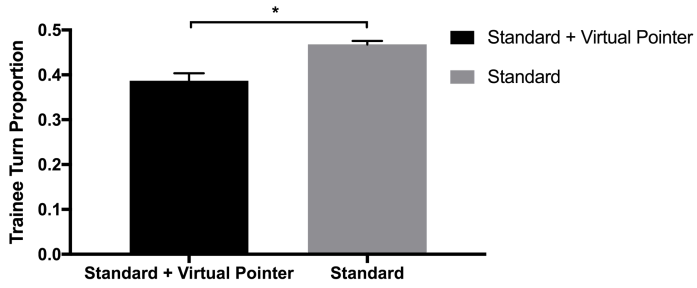

Figure 6: Comparison of trainee turn distribution between Virtual Pointer condition and Standard condition. (n = 7, *: p < .05)

at a equal rate with the trainers. In Standard condition, the trainees' turn proportion has reached 0.468, while in the VP condition, it decreased to 0.387. This decrease reveals that the communication becomes more imbalanced with the use of VP and confirmed our first hypothesis.

Communication Content

The imbalanced communication when using the single-user gesturing tool indicates that there are potential costs for trainees in making their contributions to the communication. In examining the communication content, we aimed to elucidate the types of contributions that the single-user gesturing tool hurdled.

First, we looked at the changes in dialogue act proportions among the trainers and trainees between the two conditions, to obtain an overall view of the impact of VP on the communication content. As shown in Table 3, with the use of VP, the dialogue act for clarification (CL) significantly increased (β= -.532, p = .016), while the requests for confirmation (CO) significantly decreased (β= .826, p = .007). The increase in CL is corresponding to the use of gesturing tools in making the instructions more directive. The decrease in CO indicates that the pairs made less proposal in completing the task.

Table 3: Comparison of dialogue act proportion between Virtual Pointer condition and Standard condition. (Bold: significant changes, *: p < 0.05, **: p < 0.01)

Dialogue Act	Standard + VP	Standard
Add Info (AI)	3.52%	2.22%
Query (Q)	0.6%	0.69%
Replay (R)	1.02%	1.87%
Check (CH)	1.05%	0.62%
Align (AL)	0.84%	0.76%
Clarify (CL) *	**9.98%**	**5.4%**
Acknowledge (AC)	28.11%	28.85%
Manage (MN)	41.39%	37.25%
Summarize (SA)	0	0
Judge (J)	6.96%	10.39%
Confirm (CO) **	**3.54%**	**6.79%**
Agree (AG)	3%	5.16%

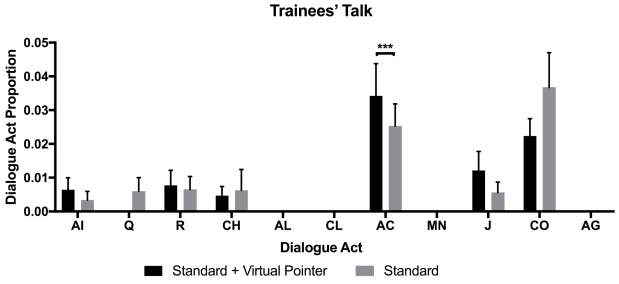

Figure 7: Comparison of dialogue act proportion of trainees' talk between Virtual Pointer condition and Standard condition. (*: p < .0001)**

We then analyzed the dialogue act proportions for trainers and trainees separately. In the following, we used the communication content of trainees to identify the costs, and that of trainers to explain the costs.

Figure 7 shows the comparison of dialogue act proportions for trainees' talks between the two conditions. With the use of VP, the trainees showed their acknowledgement more frequently through verbal utterances (β= -.299, p < .0001). The increase in trainees' explicit acknowledgement indicates information sharing becomes more efficient and accurate. All other dialogue acts remain similar.

In contrast, the significant decreases in the judgment act (J) (β= 2.260, p = .026) and the confirmation act (CO) (β= 1.241, p = .007) are found in the trainees' actions when using the VP (Figure 8). The J and CO are related to the intentions of presenting one's proposal on performing a task. The reduced proportions on the J and CO acts reflect that the

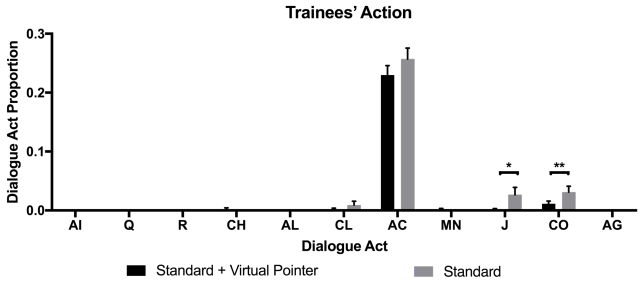

Figure 8: Comparison of dialogue act proportion of trainees' action between Virtual Pointer condition and Standard condition. (*: p < .05, **: p < .01)

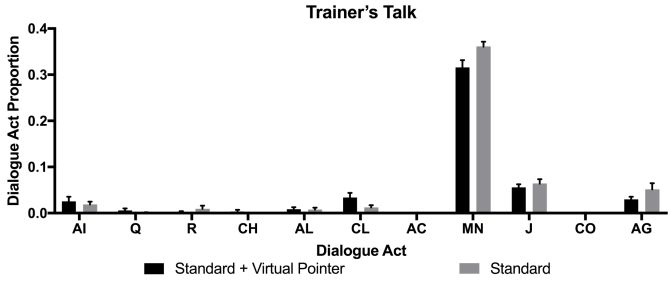

Figure 9: Comparison of dialogue act proportion of trainers' talk between Virtual Pointer condition and Standard condition.

trainees made fewer proposals in the VP condition. This result confirmed our hypothesis 2 that the use of VP hurdles trainees in contributing to the group decision making.

Since communicative acts are interdependent, we further examine the dialogue act changes in trainers' utterances and actions, in order to elucidate the causes of the decreased contributions that the trainees made to the grounding process.

Figure 9 shows the comparison of dialogue act proportions for trainers' talks between the VP condition and the Standard condition. Interestingly, although speech is the main way for the trainers to share their knowledge, there is no significant change of dialogue act found in trainers' utterances when using the VP. The trainers made equal contributions to the grounding process via utterances.

The use of VP is included in the trainers' actions for the comparison. Thus, we expect that the main changes takes place in trainers' actions. As shown in Figure 10, the VP significantly increased the trainers' actions in providing instructions (MN) (β= -1.170, p = .0001), indicating that the trainers were tended to provide more direct instructions when using the VP. This trend is associated with the increase of trainees' explicit acknowledgements (AC) in Figure 7, indicating that the VP provides an efficient way in directly presenting information that is easily accepted. According to the Principle of

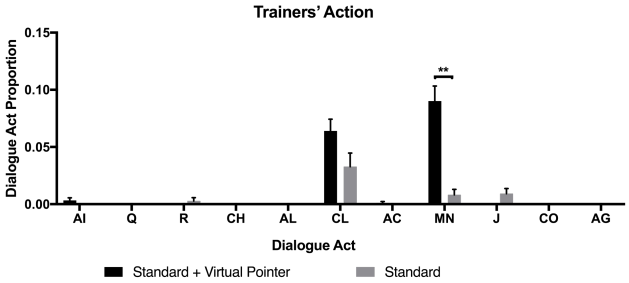

Figure 10: Comparison of dialogue act proportion of trainers' action between Virtual Pointer condition and Standard condition. (: p < .01)**

Least Collaborative Effort[5], the trainees are prone to adopt this efficient way in accomplishing the task, compared to making proposals (J and CO in Figure 8) that may lead to more discussions on repairing the understanding.

Impacts of Trainees' Reduced Contribution

In order to investigate the impacts of trainees' reduced contribution, we compared the training context around trainees' J and CO in the Standard condition with that for the same subtask in the VP condition. In the excerpts, the dialogue acts are tagged in square baskets.

The excerpt 1 and 2 shows that fewer judgments (J) were made by the trainees in the VP condition. As shown in the excerpt 1, every directional instruction was referenced by the VP, based on which the trainee took actions. The VP decomposes the trainer's process knowledge into a protocol, on which, the trainee follows step by step. Although such protocol simplifies the complex and interwoven process and makes the task efficient and accurate, much less knowledge remains in this protocol. For example, the instruction in line 1 provides no information on what the trainee is going to clip, where the clip should be, and why the clip is on the red dot. Sharing such information, however, is the main purpose of the instructional collaboration, through which the trainee acquires expertise.

Whereas, in excerpt 2, the fewer directive instructions encourage the trainee to search for the visual cues that are embedded in the task object. In this, the trainee actively uses the instructions provided and makes their own judgment. The instructor assesses the trainee's judgment and provides timely feedbacks, such as "a little more" and agreement of the actions. This illustrates a process of knowledge co-construction, in which in-situ knowledge is inserted in the course of negotiation. Thus, the reduced trainees' judgment act indicates that trainees passively followed instructions, which leads to less expertise gained.

Excerpt 1 - Pair 6 - Telestration-supplemented guidance - Task 2 clipping the duct.

1. Trainer One there **[MN]**. *(The instructor pointed at the location for the clip.)* **[MN]**
2. Trainee *(The trainee arrived at the dot and applied the clip.)* **[AC]**
3. Trainer Cool, one higher. Like right there. **[MN]** *(The instructor stabilized the dot one the duct.)* **[CL]**
4. Trainee *(The trainee moved the clip applier up a little bit and clipped.)* **[AC]**
5. Trainer Ok, and then move up to like right there. **[MN]** *(The instructor moved the pointer to a higher location.)* **[CL]**
6. Trainee *(The trainee moved up and clipped.)* **[AC]**

Excerpt 2 - Pair 1 - Standard guidance - Task 2 clipping the duct.

1. Trainer So, I usually do distal first. Doesn't really matter, it's just preference. **[J]**
2. Trainee *(The trainee moved the scissors towards the distal of the duct and added the first clip.)* **[J]**
3. Trainer And then, you don't need to take it completely out, just slide it up. **[MN]**
4. Trainee *(The trainee moved the tip up a small amount.)* **[J]**
5. Trainer Yeah, a little more. **[MN]**
6. Trainee *(The trainee moved up a little more and clipped.)* **[J]**
7. Trainer One more there. **[MN]**
8. Trainee *(The trainee slides the clip applier up and added the last one.)* **[J]**

Excerpt 3 - Pair 7 - Telestration-supplemented guidance - Task 4 cutting the artery and the duct.

1. Trainer *(The instructor moved the dot between the clips.)* **[MN]**
2. Trainer So you want to cut the artery right about there, where the dot is. **[MN]**
3. Trainee *(The trainee moved the scissors to the dot and cut the artery.)* **[AC]**

Excerpt 4 - Pair 4 - Standard guidance - Task 4 cutting the artery and the duct.

1. Trainer And just cut in between the two. **[MN]**
2. Trainee Between the two clips? **[CO]** *(The trainee hovered over the artery and the duct.)* **[CO]**
3. Trainer Uh. **[AG]**
4. Trainee This? **[CO]** *(The trainee poked the duct with the tooltip.)* **[CO]**
5. Trainer No not the duct, the artery. **[MN]**
6. Trainee *(The trainee cut the artery.)* **[AC]**

The active learning is also supported by CO acts. For example, in excerpt 4, the trainee uses CO acts to elicit information that the trainer fails to specify. The trainee needs to know where to cut (line 2), and what structure should be cut first (line 4). This information is embedded in the trainer's instruction in line 1 ("between the two"). In requesting the confirmation, the trainee clarifies the instruction and transforms the implicit knowledge into explicit. Whereas, in VP condition (excerpt 3), the instruction is specified by the reference, which the trainee directly follows without further considering what the process could be. Thus, the reduced trainees' CO act results in less knowledge elicited from the trainers, and thus reduces the expertise gained.

6 DISCUSSION

In this paper, we have investigated the communication costs of a single-user gesturing tool and explored the impacts of the costs in an instructional collaborative task. Our results show that the communication becomes more imbalanced and the trainees become less active in team decision-making when using the single-user gesturing tool. These costs shift trainees from actively acquiring expertise into passively accepting information.

We found that the turn frequencies for both the trainers and trainees are increased in the VP condition, yet the turn distribution skews towards the trainers. This result indicates that the trainers become the dominant speaker with the use of the single-user gesturing tool.

The gesturing tool has been shown to facilitate trainers to provide instruction through enabling trainers to make references on the shared display [21], and to improve trainees' understanding of the instructions [11]. Correspondingly, our results show an increase in trainers' management (MN) actions and in trainees' explicit acknowledgement (AC) with the use of VP, suggesting the trainers dominate the floor mainly to provide instructions (MN). With the ease of following directive instructions, the trainees become more reluctant in identifying what knowledge they need, what information the task object contains, and what possible steps they can take. These are the thinking processes that drive the trainees to make their own decisions on how to proceed. In this, we saw a decrease in making procedural proposals in the VP condition.

Along with previous works in understanding the communication structure and content and the grounding process [5, 44], our work suggests a control shift in a collaborative work - if one party's access to the shared display increases, that party may direct the collaboration and is prone to dominate the floor. Simultaneously, the other party's control over the collaboration may decrease, resulting in less contributions, although the access for the other party remains the same. For example, Kirk et al. showed that when the access of

remote helpers to the local task increased, the remote helpers began to direct the communication and the local workers significantly decreased their language use [21].

Implications for Design

The broad implication for the design of shared display in supporting instructional collaboration is to balance the communicative access of all parties to the shared display, in order to support team decision-making and knowledge co-construction. By communicative access, we emphasize the access that leads to contributions to team communication. Specifically, there are two directions for the design - supporting communication coordination and supporting communicative acts.

As suggested in our study, the imbalanced communication is a cause for the reduced trainees' process contributions and the passive learning. Thus, one way to support the instructional collaboration is to support team communication coordination, i.e., controlling the turn-taking structure in team communication. Shu and Flowers suggested that the forced turn-taking does not necessarily facilitate team communication and is the least preferred in peer collaboration, where team members shared the similar expertise levels [45]. Yet,in instructional collaborative tasks, where the expertise levels vary between team members, our findings reveal that without no control of turn-taking, experts tend to dominate the floor. In addition, our study suggests that the forced turn-taking should target at human-human interaction level, compared to human-display interaction level. In this direction, open questions may include how to better define a turn in the intersection of utterances, actions, and interaction with a display, how to design the forced turn-taking that can be smoothly integrated in the communication process, and when to implement the tool for the instructional collaboration.

Another way to balance the communicative access is to support communicative acts through providing multi-user input [46]. Most current multi-user input in shared displays provide the same tools to the users and allows for concurrent activities [38, 39, 42]. Yet, these displays are mainly for peer collaboration, where all team members physically work on a task. In the instructional collaboration, the trainees are the primary task performers, while the trainers are monitoring and providing instructions. Thus the ways to present information are different. Instead of providing the same tools, we may target at balancing the types of communicative acts that will be achieved by the tools to accommodate individual users' interaction needs. For example, in a pilot simulation training, the hands of the trainees may be occupied, while the trainers are pointing at the display. In this case, we can support the trainees to make the same types of communicative acts by allowing them to point at the display through eye movements or voice control. In this, different members may interact differently with the display or with each other, but they share the same access to performing communicative acts. This may require specific design that caters individual tasks.

Limitations

It is noteworthy that most trainees recruited in this study were first year general surgery residents who had not performed any laparoscopic surgery before. Although the expertise gap between the trainer and trainees magnify the main feature of the instructional collaborative tasks, we indeed observed the instructions on basic surgical knowledge. The lack of this knowledge may hurdle the trainees in participating in the group decision making. Thus, communication structure and content may be different in the collaboration between more senior residents and expert surgeons.

In addition, our tasks are set to be short to target at the team's initial efforts in developing the common ground. However, we acknowledge that with a longer task, where teams collaborate with more established common ground, the communication may become balanced and the trainees may begin to make proposal. Thus, more studies need to be taken in verifying if the communication costs found in this paper apply to longer tasks.

7 CONCLUSION

In this study, we investigate the communication costs with the use of a single-user gesturing tool in instructional collaborative tasks. We found that communication became more imbalanced and trainees contributed less in the team decision-making, when using the single-user gesturing tool. These costs impede trainees' acquiring expertise by turning active knowledge-seeking into passive information acceptance. Our findings highlight the needs for equal communicative access to the shared display among group members, and suggest two main design directions - supporting the communication coordination and supporting the communicative acts - for achieving this equal access in instructional collaborative tasks.

ACKNOWLEDGMENTS

The authors would like to thank the surgeons and surgical trainees for participating in this study, the Anne Arundel Medical Center for the use of equipment and space in the Simulation to Advanced Innovation and Learning (SAIL) Center, and Mr. Jatin Chhikara and Ms. Jordan Ramsey for their support in developing the system. This work was supported by National Science Foundation IIS #1422671 and #1552837.

REFERENCES

[1] Mathilde M. Bekker, Judith S. Olson, and Gary M. Olson. 1995. Analysis of Gestures in Face-to-face Design Teams Provides Guidance for How to Use Groupware in Design. In *Proceedings of the 1st Conference on Designing Interactive Systems: Processes, Practices, Methods, & Techniques (DIS '95)*. ACM, New York, NY, USA, 157–166. https://doi.org/10.1145/225434.225452

[2] Richard Bentley, John A Hughes, David Randall, Tom Rodden, Peter Sawyer, Dan Shapiro, and Ian Sommerville. 1992. Ethnographically-informed systems design for air traffic control. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM, 123–129.

[3] Jeremy Birnholtz, Abhishek Ranjan, and Ravin Balakrishnan. 2010. Providing Dynamic Visual Information for Collaborative Tasks: Experiments With Automatic Camera Control. *HumanâĂŞComputer Interaction* 25, 3 (2010), 261–287. https://doi.org/10.1080/07370024.2010.500146 arXiv:https://www.tandfonline.com/doi/pdf/10.1080/07370024.2010.500146

[4] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

[5] Herbert H Clark, Susan E Brennan, et al. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991 (1991), 127–149.

[6] Herbert H Clark and Tania Henetz. 2014. Working together. *Oxford handbook of language of social psychology* (2014), 85–97.

[7] Gregorio Convertino, Helena M Mentis, Mary Beth Rosson, John M Carroll, Aleksandra Slavkovic, and Craig H Ganoe. 2008. Articulating common ground in cooperative work: content and process. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1637–1646.

[8] Gregorio Convertino, Helena M Mentis, Mary Beth Rosson, Aleksandra Slavkovic, and John M Carroll. 2009. Supporting content and process common ground in computer-supported teamwork. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2339–2348.

[9] 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), 22.

[10] Yuanyuan Feng, Hannah McGowan, Azin Semsar, Hamid R Zahiri, Ivan M George, Timothy Turner, Adrian Park, Andrea Kleinsmith, and Helena M Mentis. 2018. A virtual pointer to support the adoption of professional vision in laparoscopic training. *International Journal of Computer Assisted Radiology and Surgery* (2018), 1–10.

[11] Susan R Fussell, Leslie D Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. 2004. Gestures over video streams to support remote collaboration on physical tasks. *Human-Computer Interaction* 19, 3 (2004), 273–309.

[12] Steffen Gauglitz, 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*. ACM, 449–459.

[13] Darren Gergle, Robert E Kraut, and Susan R Fussell. 2004. Action as language in a shared visual space. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 487–496.

[14] Jeffrey T Hancock and Philip J Dunham. 2001. Language use in computer-mediated communication: The role of coordination devices. *Discourse Processes* 31, 1 (2001), 91–110.

[15] Christian Heath and Paul Luff. 1991. Disembodied conduct: communication through video in a multi-media office environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 99–103.

[16] Christian Heath, Paul Luff, and Marcus Sanchez Svensson. 2003. Technology and medical practice. *Sociology of Health & illness* 25, 3 (2003), 75–96.

[17] Weidong Huang and Leila Alem. 2013. HandsinAir: a wearable system for remote collaboration on physical tasks. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, 153–156.

[18] Kori M Inkpen, Wai-ling Ho-Ching, Oliver Kuederle, Stacey D Scott, and Garth BD 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*. International Society of the Learning Sciences, 31.

[19] Ellen A Isaacs and Herbert H Clark. 1987. References in conversation between experts and novices. *Journal of experimental psychology: general* 116, 1 (1987), 26.

[20] David Kirk, Andy Crabtree, and Tom Rodden. 2005. Ways of the Hands. In *ECSCW 2005*, Hans Gellersen, Kjeld Schmidt, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). Springer Netherlands, Dordrecht, 1–21.

[21] 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*. ACM, 1039–1048.

[22] Robert E Kraut, Susan R Fussell, and Jane Siegel. 2003. Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interaction* 18, 1-2 (2003), 13–49.

[23] Hideaki Kuzuoka, Toshio Kosuge, and Masatomo Tanaka. 1994. GestureCam: A video communication system for sympathetic remote collaboration. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM, 35–43.

[24] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[25] Linda Lebie, Jonathan A Rhoades, and Joseph E McGrath. 1995. Interaction process in computer-mediated and face-to-face groups. *Computer Supported Cooperative Work (CSCW)* 4, 2-3 (1995), 127–152.

[26] Paul Luff, Karola Pitsch, Christian Heath, Peter Herdman, and Julian Wood. 2010. Swiping paper: the second hand, mundane artifacts, gesture and collaboration. *Personal and Ubiquitous Computing* 14, 3 (2010), 287–299.

[27] Halliday J MacFie, Nicholas Bratchell, KEITH GREENHOFF, and Lloyd V Vallis. 1989. Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of sensory studies* 4, 2 (1989), 129–148.

[28] Roberto Martinez-Maldonado, Peter Goodyear, Lucila Carvalho, Kate Thompson, Davinia Hernandez-Leo, Yannis Dimitriadis, Luis P Prieto, and Dewa Wardak. 2017. Supporting collaborative design activity in a multi-user digital design ecology. *Computers in Human Behavior* 71 (2017), 327–342.

[29] Helena M Mentis. 2017. Collocated Use of Imaging Systems in Coordinated Surgical Practice. *PACMHCI* 1, CSCW (2017), 78–1.

[30] Helena M Mentis, Amine Chellali, and Steven Schwaitzberg. 2014. Learning to see the body: supporting instructional practices in laparoscopic surgical procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2113–2122.

[31] Helena M Mentis, Ahmed Rahim, and Pierre Theodore. 2016. Crafting the Image in Surgical Telemedicine. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 744–755.

[32] 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, 1479–1488.

[33] Andrew Monk. 2003. Common ground in electronically mediated communication: ClarkâĂŹs theory of language use. *HCI models, theories, and frameworks: Toward a multidisciplinary science* (2003), 265–289.

[34] Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. 2006. Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 1201–1210.

[35] AG Nagy. 1999. Hierarchical decomposition of laparoscopic procedures. *Medicine Meets Virtual Reality: The Convergence of Physical & Informational Technologies: Options for a New Era in Healthcare* 62 (1999), 83.

[36] Timothy J Nokes-Malach, Michelle L Meade, and Daniel G Morrow. 2012. The effect of expertise on collaborative problem solving. *Thinking & Reasoning* 18, 1 (2012), 32–58.

[37] Jiazhi Ou, Susan R Fussell, Xilin Chen, Leslie D Setlock, and Jie Yang. 2003. Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks. In *Proceedings of the 5th international conference on Multimodal interfaces*. ACM, 242–249.

[38] Anne Marie Piper and James D Hollan. 2009. Tabletop displays for small group study: affordances of paper and digital materials. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1227–1236.

[39] Yvonne Rogers, William Hazlewood, Eli Blevis, and Youn-Kyung Lim. 2004. Finger talk: collaborative decision-making using talk and fingertip interaction around a tabletop display. In *CHI'04 extended abstracts on Human factors in computing systems*. ACM, 1271–1274.

[40] Alison Sanford, Anne H. Anderson, and Jim Mullin. 2004. Audio channel constraints in video-mediated communication. *Interacting with Computers* 16, 6 (2004), 1069–1094.

[41] Stacey D Scott, Karen D Grant, and Regan L Mandryk. 2003. System guidelines for co-located, collaborative work on a tabletop display. In *ECSCW 2003*. Springer, 159–178.

[42] Stacey D Scott, Garth BD Shoemaker, and Kori Inkpen. 2000. Towards seamless support of natural collaborative interactions. In *Graphics Interface*. 103–110.

[43] F Jacob Seagull, Ivan George, Iman Ghaderi, Marilou Vaillancourt, and Adrian Park. 2009. Surgical abdominal wall (SAW): a novel simulator for training in ventral hernia repair. *Surgical innovation* 16, 4 (2009), 330–336.

[44] Abigail J Sellen. 1995. Remote conversations: The effects of mediating talk with technology. *Human-computer interaction* 10, 4 (1995), 401–444.

[45] Li Shu and Woodie Flowers. 1992. Groupware experiences in three-dimensional computer-aided design. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. ACM, 179–186.

[46] Jason Stewart, Benjamin B Bederson, and Allison Druin. 1999. Single display groupware: a model for co-present collaboration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 286–293.

[47] John C Tang and Scott L Minneman. 1990. VideoDraw: a video interface for collaborative drawing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 313–320.

[48] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*. ACM, 137–146.