robotics

MDPI

Article

Evaluation of a Voice-Enabled Autonomous Camera Control System for the da Vinci Surgical Robot

Reenu Arikkat Paul [1], Luay Jawad [2], Abhishek Shankar [2], Maitreyee Majumdar [1], Troy Herrick-Thomason [3,4] and Abhilash Pandya [1,*]

[1] Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202, USA; hm5958@wayne.edu (R.A.P.); maitreyee@wayne.edu (M.M.)
[2] Department of Computer Science, Wayne State University, Detroit, MI 48202, USA; ljawad@wayne.edu (L.J.); abhishek.shankar@wayne.edu (A.S.)
[3] Department of Psychology, Wayne State University, Detroit, MI 48202, USA; gh4657@wayne.edu
[4] Department of Chemistry, Wayne State University, Detroit, MI 48202, USA
* Correspondence: apandya@wayne.edu

Abstract: Robotic surgery involves significant task switching between tool control and camera control, which can be a source of distraction and error. This study evaluated the performance of a voice-enabled autonomous camera control system compared to a human-operated camera for the da Vinci surgical robot. Twenty subjects performed a series of tasks that required them to instruct the camera to move to specific locations to complete the tasks. The subjects performed the tasks (1) using an automated camera system that could be tailored based on keywords; and (2) directing a human camera operator using voice commands. The data were analyzed using task completion measures and the NASA Task Load Index (TLX) human performance metrics. The human-operated camera control method was able to outperform an automated algorithm in terms of task completion (6.96 vs. 7.71 correct insertions; p-value = 0.044). However, subjective feedback suggests that a voice-enabled autonomous camera control system is comparable to a human-operated camera control system. Based on the subjects' feedback, thirteen out of the twenty subjects preferred the voice-enabled autonomous camera control system including the surgeon. This study is a step towards a more natural language interface for surgical robotics as these systems become better partners during surgery.

Keywords: robotic surgery; automated camera; da Vinci system; natural language processing

Citation: Paul, R.A.; Jawad, L.; Shankar, A.; Majumdar, M.; Herrick-Thomason, T.; Pandya, A. Evaluation of a Voice-Enabled Autonomous Camera Control System for the da Vinci Surgical Robot. *Robotics* **2024**, *13*, 10. https://doi.org/10.3390/robotics13010010

Academic Editors: Naira Hovakimyan and Kerstin Thurow

Received: 21 October 2023
Revised: 24 December 2023
Accepted: 27 December 2023
Published: 1 January 2024

1. Introduction

Robotic surgery has revolutionized the field of minimally invasive procedures, offering improved precision and flexibility to surgeons. The motivation behind this research is to address the unique challenges of teleoperating a surgical robot using remote video views. This paper aims to address these challenges, particularly in camera control, and compares our current implementation of two camera control systems: (1) voice-enabled autonomous camera control (VACC); and (2) human-operated camera control (HOCC). To evaluate the performance and user interface of VACC as compared to HOCC, a comprehensive human factors study was conducted involving twenty subjects. In addition, the paper discusses the implications of these findings and explores potential avenues for future research such as extended language capabilities.

Medical errors are the third leading cause of death in the US [1], with many errors during surgery being caused by poor visualization, decision-making, and inadvertent tool movements. The Centers for Disease Control and Prevention (CDC) estimates that more than 2 million laparoscopic surgeries (which could potentially be undertaken using surgical robots as the technology matures and becomes more affordable) are performed each year just in the US [2]. The World Health Organization states that surgical safety should now be

a substantial global public health concern because of the high death and complication rates of major surgical procedures [3].

Robot-assisted minimally invasive surgery (RAMIS) has achieved widespread global clinical adoption, primarily due to reduced trauma to the patient as well as improved precision, dexterity, and visualization for the surgeon [4]. RAMIS heavily relies on the use of real-time imaging through an endoscopic camera. The camera provides a high-resolution video stream of the surgical site and offers wide-angle visibility. The surgeon manipulates robotically articulated instruments through a surgical console, also known as a human–machine interface (HMI), which relies on the video stream for guidance. The combination of minimally invasive techniques and robotic assistance has significantly contributed to the rapid growth of RAMIS over the past decade [5]. The most widely adopted RAMIS system to date is the da Vinci surgical system developed by Intuitive Surgical Inc. in Sunnyvale, CA, USA [6].

During robotic surgery, the surgeon lacks the sensory feedback present in traditional surgical operations, resulting in reduced situational awareness and a high mental workload. Since the surgeon is expected to control most aspects of the remote interaction, reacting effectively to irregular events in the surgical environment can be taxing and error prone. To ensure safe operations, a well-designed user interface for robotic surgical tasks is essential. Managing the camera viewpoint is one key element of a user interface. During both robotic and traditional laparoscopic surgery, the surgeon must continually control the camera's position to ensure that they have sufficient visibility of the tools while performing precise procedures. However, camera-related tasks are sources of considerable distraction and divert the operator's attention from performing the surgery [4]. Achieving optimal visualization in a dynamic environment is challenging, and robotic tool movements initiated without an appropriate view of the operating environment or tool can have unintended and potentially dangerous consequences. The current state of the art in robotic surgery provides only manual solutions (e.g., clutch to move the camera) for assisting with the challenging task of dynamic camera positioning.

To perform successful surgery, uninterrupted and seamless visualization of the operating field and the surgical tools is essential. This enables surgeons to monitor instrument–tissue interactions while treating patients. The surgical interface must minimize task switching; for example, switching between moving the camera and tools or using the clutch and keypad interface. The surgeon's time within the console observing the surgical site must be maximized, not infringed upon by constantly adjusting robot parameters using the keypad or clutch. Poor visualization and continuous task switching can reduce surgical precision and compromise patient safety by enabling risky tool movements [5]. Therefore, advanced user interface techniques with safe low-level automation that seamlessly integrates with the surgeon's tool gestures and voice commands are needed in order to improve control issues in robotic surgery. Additionally, a voice interface that allows for snapshots of pictures, videos, and recordings of surgical tool movements would be ideal for clear documentation of surgical procedures for the patient and for the potential training of surgical residents.

It is clear from the research that camera control is a complex and demanding task for the surgeon [7,8]. The studies cited indicate that the surgeon may need to reposition the camera frequently during surgery and that this can result in a higher workload for the operator, both physically and cognitively. This can lead to longer operation times, decreased visual quality, and a higher probability of errors [9–11]. These findings highlight the need for more advanced user interfaces with some automation (e.g., camera control systems) to alleviate the surgeon's workload and enhance surgical outcomes [7,12,13].

Various camera control methods (for example, eye gaze tracking and segmentation algorithms), as well as systems such as AESOP (Automated Endoscopic System for Optimal Positioning) and EndoAssist, have been developed for use during laparoscopic and robotic surgery. Reactive and proactive automated camera control approaches have also been developed as reported in [7].

Several autonomous camera systems have been developed to facilitate minimally invasive surgical procedures, as referenced in the range of publications [14–23]. These systems primarily employ image processing or kinematics to determine the tool tip positions in relation to the camera for tracking purposes. Typically, these systems rely on a set of predefined rules to determine the camera's target position and zoom level. Another study which used an autonomous camera system on an ex vivo animal neobladder reconstruction study showed promising results [24]. The literature on voice control for laparoscopic camera manipulation is scant. No researchers that we have found compare an automated system that uses voice with a human operator in the laparoscopic domain.

In our case, we have integrated a rule-based approach on our da Vinci platform, directing the camera towards the midpoint of two tracked tool-tips and adjusting the zoom level as required to maintain the tool-tips within the camera's frame. These systems are relatively simplistic, as they primarily focus on the tool-tips without considering other factors. We have already shown that these autonomous camera systems far out-perform the current clutch-and-move systems. In addition, user studies have revealed that users often require customization of the autonomous camera's behavior to accomplish their specific tasks, as discussed in [25].

To enable seamless adjustments to autonomous camera behavior, we previously introduced a system that incorporates a voice assistant, allowing users to modify the base behavior of the autocamera system using voice commands, such as tracking the left or right tool with a specific voice command [26]. The current paper assesses the performance of the previously developed VACC system in comparison to the HOCC system within the context of our user study.

2. Materials and Methods

2.1. Test Platform Development

2.1.1. Hardware and Software Implementation Architecture

Our test platform is primarily comprised of a da Vinci standard surgical system that has been modified to work with the da Vinci Research Kit (dVRK) [27]. The instrument cart shown in Figure 1a contains five control boxes, one to interface with each of the master tool manipulators (MTMs) (2×), the patient-side manipulators (PSMs) (2×), and the endoscopic camera manipulator (ECM) (1×). The control boxes also serve as the interface between the hardware and the software. Under normal circumstances, the operator controls the da Vinci from the surgeon console in Figure 1b using the video feed, MTMs, and foot pedals. Figure 1c depicts the patient-side system, which serves as the location for the testbed placement and interaction with the subjects. The primary components of the dVRK, aside from the robot, are hardware control boxes that have field-programmable gate array (FPGA) boards and amplifiers, and open-source software that facilitates computerized control of the arms.

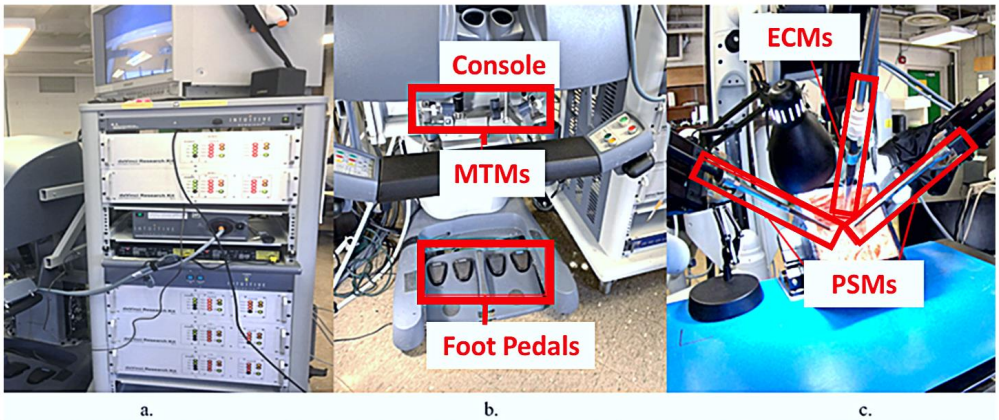

Figure 1. The da Vinci surgical system: (**a**) the instrument cart with dVRK control boxes; (**b**) the surgeon console with foot pedals and MTMs; (**c**) the patient-side system with labeled PSMs and ECM.

The dVRK software employs the open-source robot operating system (ROS) framework, the computer-integrated surgical systems and technology (CISST), and surgical assistant workstation (SAW) libraries created by Johns Hopkins University [27]. The software framework relies heavily on ROS, a system that enhances interprocess communication, enabling seamless interaction among processes on the same or different computers via a network of nodes. Within this software framework, the interface entails the exchange of data among a network of ROS nodes, the da Vinci hardware, the surgeon, and our automated camera software. The fundamental objective of this system is to harness the capabilities of the ROS network to retrieve manipulator joint values from the patient-side hardware, process these data to determine the desired camera orientation, and subsequently configure the joint values for the camera arm, thereby establishing an autonomous camera platform.

2.1.2. Baseline Autocamera Algorithm

The autocamera algorithm developed by Eslamian et al. serves as the base for more refined methods of camera control [25]. Autocamera tracks the midpoint of the two PSMs and ensures that this is always in the field of view via zoom adjustment. The positions of the PSMs are calculated via forward kinematics and the midpoint of the two tools are projected on to the image plane, which is used to calculate the camera arm position so that it points towards the centroid of the tools. The system relies on accurate camera calibration and optimization of the transformations between the three robot arms of the system (the camera arm and the two tool arms).

Another feature of the baseline autocamera algorithm is a zone/time-based gesture. The zoom level of the camera is determined by the projection of the tools into the 2D view. To facilitate this, the camera image is partitioned into three areas: the inner zone, dead zone, and outer zone. The algorithm zooms in when the tools are in the inner zone and zooms out when they are in the outer zone, until both tools are situated in the dead zone. The zooming process will only commence if the tools remain motionless in either the inner or outer zone for at least 100 milliseconds. This simple method ensures that the tools can be zoomed in or out without interfering significantly with the task [25].

2.1.3. Natural Language (Keyword) Interface

A natural language (keyword-based) interface was added to the autocamera algorithm to provide more control of the camera movement to the user [26]. The Vosk speech recognition API, based on the Kaldi toolkit, was used to integrate the interface [28]. Vosk offers an API that works offline and allows for the limiting of the existing vocabulary to certain keywords. The natural language interface allows the user to specify the required camera movements such as tracking the left or the right tool, tracking the midpoint, keeping a point in the field of view, etc. Based on feedback from a robotic surgeon, two additions were made to the existing list of functions—"record video" and "take picture". This allows the user to record a video or take a picture of a procedure (or sub procedure) from the endoscopic camera, which can be used for documenting or teaching purposes. Table 1 shows the list of commands that the da Vinci responds to.

Table 1. This is the list of commands for voice recognition. Note that each command on the left must be preceded by the keyword "da Vinci".

Command	Action
Start/Stop	Start/stop autocamera algorithm
Track middle	Track midpoint of the tools
Track left/right	Track the left/right tool
Keep left/right	Keep the point at which the left/right tool is at currently, in the field of view for future movements
Find my tools	Move camera to have both tools in the field of view
Take picture	Save an image from the endoscopic camera
Begin/end recording	Start/stop the video recording

2.2. Human Subject Testing Protocol

A study involving twenty subjects (including one robotic surgeon) using two different camera control methods—VACC and HOCC—was performed. Each subject underwent a series of trials during which video and kinematic data were systematically documented to enable a comparative analysis of the two camera control methods. Additionally, participants were administered a NASA-TLX questionnaire, and their performance data underwent a comprehensive evaluation. The comprehensive testing protocol is explained in the following sections.

2.3. Test Setup

Our human factors study received approval from Wayne State University's Institutional Review Board (IRB) under the reference number IRB #22-03-4453. The study comprised twenty participants, consisting of nineteen novices in the 20–40 age range, representing the engineering department and medical school communities. It is important to note that novices in this context refer to participants without any prior surgical experience. Additionally, one participant was identified as an experienced robotic surgeon. The trials for each participant lasted approximately 90 min. Throughout the study, each subject repeatedly performed a specific task while the camera control was alternated between the voice-enabled automated camera system and a single trained camera operator utilizing a joystick, as detailed in Figure 2. Both the autonomous camera system and the joystick operator had access to the same movement parameters of the camera. A yaw, pitch, or zoom of the camera were allowed by both (for simplicity, a rotation of the camera was not allowed).

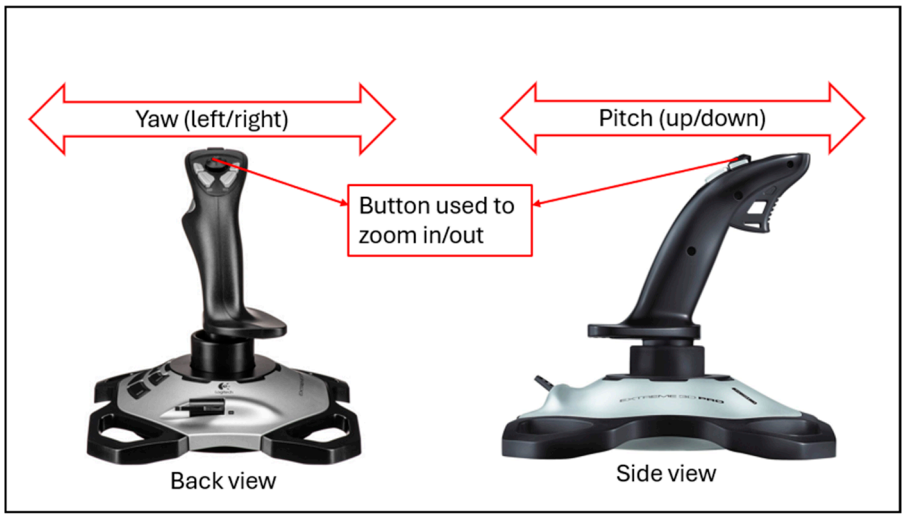

Figure 2. Joystick used to control the camera (HOCC). The joystick operator moves the joystick left and right (from the back view) to yaw the camera and pushes the joystick forwards and backwards (from the side view) to pitch the camera up or down. The toggle button at the top physically moves the camera in or out for a zooming motion.

The subjects were allowed to give freeform commands to the joystick operator. Typical commands included "zoom in/out", "follow my left/right tool", and "start/stop moving the camera". This allowed for more flexibility in camera movements, and this approach also ensured that the operator had the ability to respond to the unique requirements of each task while maintaining consistency in camera operation. The reason for one joystick operator was to not confound the results with multiple operators. The joystick operator proficiency was hence the same for all subjects and was not a variable for the study. In addition, the operator was instructed to just follow the voice commands of the participant and not add any extra movements from knowledge of the scene.

2.4. Data Collection Methodology

For the task mentioned previously, we used two printed images displaying circuit breadboards. The task entailed the precise placement of cable ends into designated holes on these two boards, as illustrated in Figure 3. The procedure required participants to transition the wire from their left hand to their right hand and then insert it into a predefined hole on the breadboard indicated in the task sheet (e.g., G17) using specialized tools. This process necessitated the navigation of the wire around the breadboard, guided by the coordinates provided on the task sheet. This particular task was chosen due to its significant camera movement requirements and its similarity to suturing motions, while being more accessible in terms of the learning curve. Participants were instructed to conduct a preliminary five-minute practice run followed by ten separate three-minute trials, each featuring a random arrangement of ten distinct patterns. The camera control methods were counterbalanced and rotated between the two test conditions—VACC and HOCC.

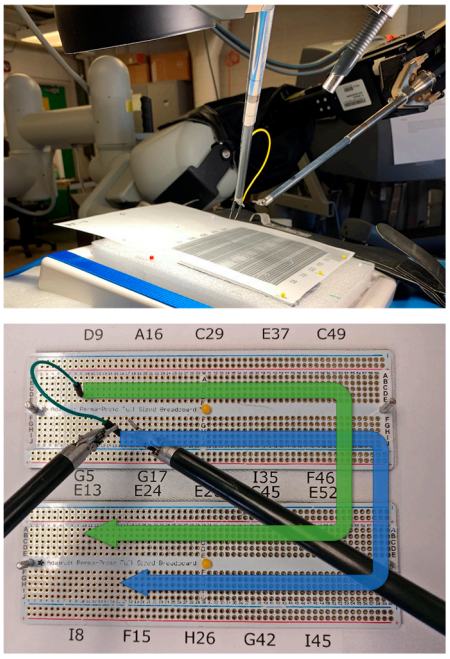

Figure 3. Setup of user study breadboards with permission from [25].

2.4.1. Score

The created punctures in the paper (with the printed images of the circuit breadboards) provided a means to assess the advancements made during the task. We devised a progress score, hereafter referred to as the "score", to provide an objective measurement of a subject's performance in a specific task. Calculating the score involved awarding one point for each puncture in the paper that fell within a distance of one hole from the specified location on the breadboard. Furthermore, errors were monitored in each test using paper sheets. Any puncture located more than one hole away from the intended target was considered an error. It is important to note that the unit of measurement corresponds to the diameter of the individual holes on the board (2 mm). Additionally, there were instances where test subjects unintentionally tore the paper due to a lack of control. This aspect was significant, as mitigating such occurrences is vital for enhancing safety during surgical procedures. Consequently, the number of tears in each trial was also recorded.

2.4.2. Questionnaire

In this study, subjects were administered a NASA-TLX questionnaire aimed at evaluating their experience level with robotic surgery, as well as measuring their level of frustration, effort, and performance during the study. The questionnaire aimed to provide

insight into the subjects' preferences of use between VACC and HOCC. We also discussed their levels of physical, mental, and temporal demand during the different camera control methods. The subjects were asked to assign an input value ranging from 1 (low demand) to 20 (high demand) in each category. In addition to the questionnaire, a short survey was administered for qualitative feedback.

2.5. Evaluation Methods

The data collected from participants, including survey results and scores, were organized and compared for the two different camera control methods. We calculated the average score for each method and its *p*-value (from a paired *t*-test). We also examined the NASA-TLX questionnaire results, which assessed the task's workload (demand) in six areas: mental, physical, temporal, performance, effort, and frustration. The detailed definitions of these categories are available in [29]. These results were compared for each camera method, and we determined average values and *p*-values (from multinomial logistic regression [30]). Besides participant data, we recorded voice commands, laparoscopic videos, kinematics, and joint angles in the ROS bag format for later analysis.

3. Results

Task completion was quantified via the scoring metric described in Section 2.4.1. As previously mentioned, the scores for each method were compared and are visually presented in Figure 4. The average score for the voice-enabled camera control algorithm was found to be 6.96, while the average score for the human-operated camera control system was found to be 7.71. The *p*-value for this analysis was found to be 0.044, suggesting that there is a statistically significant difference in task completion between the two camera control methods. Based on these data, HOCC performed slightly better on average, compared to VACC. The number of tears/rips in the paper used by the subjects as part of the study was also recorded as a measure of the performance of the two camera control methods: 87 for VACC and 71 for HOCC. The *p*-value was found to be 0.29, which suggests that the difference (although higher for HOCC) in the number of tears between the two camera control methods is not statistically significant.

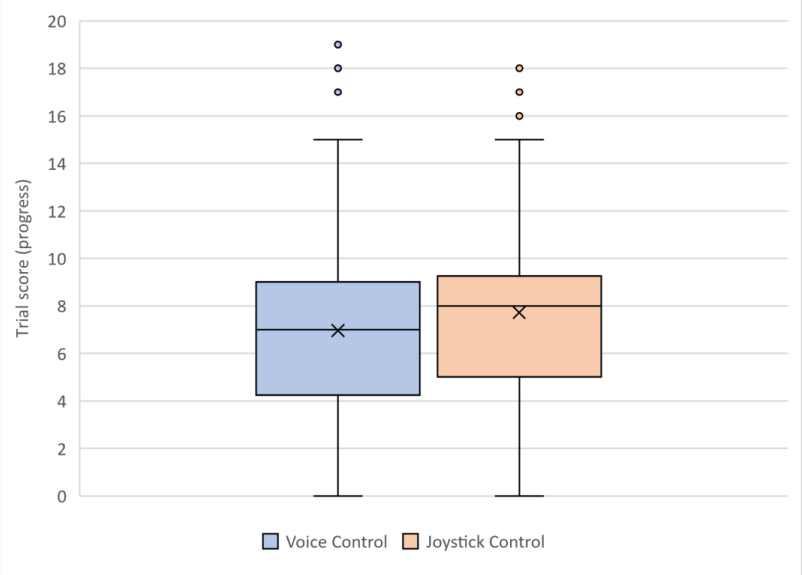

Figure 4. Comparison of the score achieved by the subjects between VACC (blue) and HOCC (orange).

Similarly, the results from the NASA-TLX questionnaire are presented in Figure 5. The following metrics were compared between the two camera control methods: mental demand, physical demand, temporal demand, performance, effort, and frustration. Based on the averages, we observed that VACC received slightly lower (better) scores in

a majority of the metrics compared to HOCC. The average scores for temporal demand were observed to be significantly higher (worse) for VACC, and the average scores for the mental demand, physical demand metrics were higher (worse) for HOCC. Table 2 presents the findings of the multinomial logistic regression analysis examining the impact of the two camera control methods (VACC vs. HOCC) on various NASA-TLX categories, where frustration was set as the control factor. Frustration was selected for this purpose due to its relatively consistent average values across both camera control methods, suggesting minimal differences in frustration levels between VACC and HOCC. Effort was found to be the only statistically significant NASA-TLX category. Compared to VACC, HOCC was associated with higher reported effort (p-value = 0.042). This suggests that participants using HOCC perceived the task as requiring more effort than those using VACC. No statistically significant relationships were observed between the two camera control methods and the remaining NASA-TLX categories (mental demand, performance, physical demand, temporal demand). While no statistically significant differences (p-value > 0.05) were observed between VACC and HOCC for mental demand, performance, and physical demand, the raw score comparisons suggest a trend towards lower (better) workload scores for the voice-enabled camera control system.

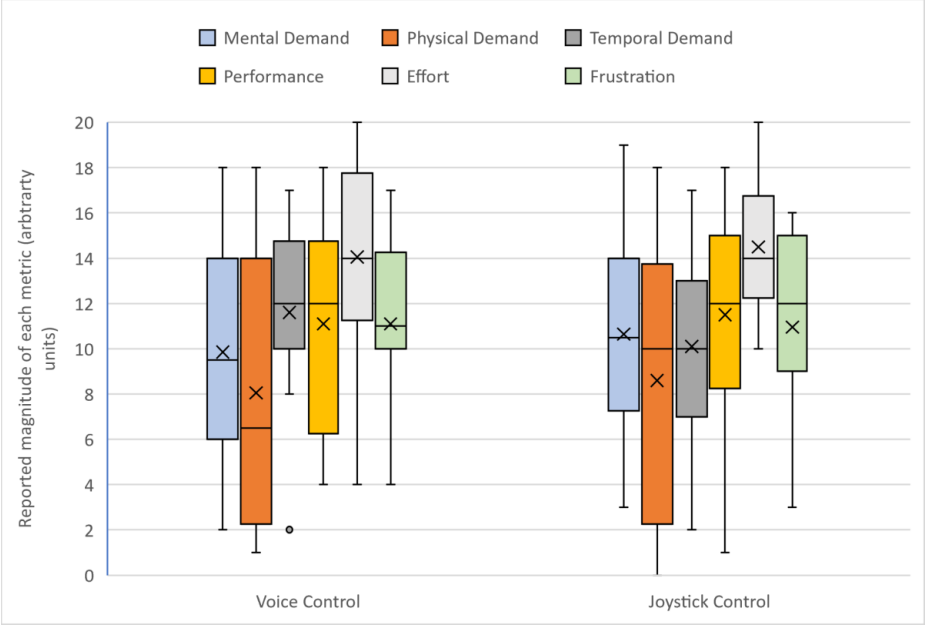

Figure 5. Comparison of the NASA-TLX data recorded by the subjects between VACC and HOCC.

Table 2. Results from multinomial logistic regression with "Frustration" set as the control factor.

NASA-TLX Category	Source (Camera Control Method)	Value	p-Value
Effort	VACC	0.127	0.603
	HOCC	0.570	0.042
Mental Demand	VACC	−0.187	0.392
	HOCC	0.061	0.784
Performance	VACC	−0.061	0.785
	HOCC	0.114	0.617
Physical Demand	VACC	−0.303	0.178
	HOCC	−0.125	0.568
Temporal Demand	VACC	0.214	0.373
	HOCC	−0.238	0.303

Based on the survey feedback, thirteen out of the twenty subjects (including the robotic surgeon) preferred the natural language interface and the other seven preferred the human-operated joystick camera control. The participants' comments are categorized based on how well each mode allowed the control of the camera, the mental demand, and physical demand. A further classification of whether the subject comments were positive or negative is displayed in Tables 3 and 4. Additionally, the subject data (score and NASA-TLX results) from the robotic surgeon are highlighted in Table 5.

Table 3. Subject comments for human-operated camera control method.

Features	Subject Comments for HOCC	
	Positive	**Negative**
Camera Control	Zooming was quicker and smoother	Camera was moved to an unwanted position
	I can tell the operator to stop or zoom out which led to a more granular level of control	Method doesn't automatically track tools
	Minimal lag for zoom in and out	The joystick had more staggered results
	The adjustment period allowed me to work on hand placement/coordination while the vision was set up	Control was inconsistent and I cannot directly control it
	Rapid zooming	It wasn't as quick since speed is also tested
	This was better, fast amount of reaction was enough to complete the task	Sometimes I had a different direction in mind than the one the operator moved the camera to
	Easier for zooming in/out	I couldn't control how much I wanted to move in certain directions
	Didn't have to say da Vinci before everything	N/A **
	Easier to zoom	N/A **
Mental Demand	It was more intuitive to move with human operator	Having to say to follow the tools made me lose focus on the task
	Less work to use	Can be difficult to communicate
	Commands were easy to think of and reaction was great	Takes focus away from task
	It was easier to communicate with a person	Having to ask operator to move camera was time consuming
	N/A **	This method felt like there was more pressure to tell the person operating the joystick where to go; it seemed that I had to focus on the tasks and communicate directions to the operator
		I dislike someone else controlling the camera; it was hard to say how much to zoom sometimes
		I had to communicate which slowed me down
		Having to direct operator took mental effort; it was hard to get perfect zoom in/out
Physical Demand	Joystick has less physical strain because I did not have to maneuver my fingers to zoom	N/A **

** No relevant comments were provided in this category.

Table 4. Subject comments for voice-enabled autonomous camera control method.

Features	Subject Comments for VACC	
	Positive	**Negative**
Camera Control	Automatic tracking used less commands	Some of the commands were redundant
	Follows work area; makes work quicker in some circumstances	Slow zoom
	More intuitive, zooming in and out was good	Zoom feature is slow/has delay
	Camera follows instruments *	Not as precise as a regular joystick
Camera Control	The voice-controlled camera was faster and allowed me to move quicker since it was following all my movements	Disliked waiting for zoom in/out
	Camera control was consistent and ability to track continuously made the task easier to complete	The zoom felt slower
	Camera followed the tools without thinking about it	Zoom in done automatically when holding still, would prefer saying zoom to hold still
	It is much easier so all I have to do is zoom in; voice was easier especially with how quick it follows the camera; it also zooms in exactly where I want it to go and quicker	Zoom in/out had difficulties
	Appreciated the camera moving on its own to follow me	Zoom is not the best, takes time; auto track/follow is great
	Automatic zoom was very helpful	Harder to control
	Camera control was fluid	It was difficult to put it in place
Mental Demand	With proper training and as trials progress, it was nice not to allow another person	Need specific commands
	This method was easier, and I was more focused on the tasks; I don't speak very loudly, but it worked fine	Track left and right were a little confusing to use when the camera just follows the tools on its own
	I was in charge of exactly where I wanted to be	N/A **
Physical Demand	N/A **	N/A **

* Specific feedback that was provided by the robotic surgeon. ** No relevant comments were provided in this category.

Table 5. Comparison of the score and NASA-TLX data for the robotic surgeon.

	VACC	**HOCC**
Score	13	15
Mental Demand	2	13
Physical Demand	2	14
Temporal Demand	2	12
Performance	5	11
Effort	4	14
Frustration	4	12

The NASA-TLX data and subjective feedback from the robotic surgeon (Table 5) indicate his preference for the voice-enabled autonomous camera control. He overwhelmingly scored higher in all NASA-TLX measures (mental demand, physical demand, temporal demand, performance, effort, and frustration) for HOCC, signifying his preference for VACC.

Both camera control systems, voice-enabled autonomous camera control and human-operated camera control, exhibit distinct characteristics based on participant feedback.

VACC received positive remarks for its automatic tracking, intuitive controls, and the ability to swiftly follow instruments and user movements, enhancing overall fluidity and speed of operation. The participants appreciated the convenience of voice commands, automatic zoom features, and the system's responsiveness. On the other hand, HOCC demonstrated strengths in providing users with granular control and responsiveness to specific commands. Participants commended the quick and smooth zooming capabilities, though some expressed challenges related to consistency and occasional unwanted movements. While VACC offered an autonomous and hands-free experience, HOCC allowed for more direct and controlled manipulation by the user. These general characteristics highlight the unique advantages and considerations associated with each camera control method.

4. Discussion

Overall, the results from the human factors study indicated a statistically significant difference in task completion between the two camera control methods, with the human-operated camera control slightly outperforming voice-enabled autonomous camera control on average. However, when considering the NASA-TLX questionnaire and the subjective survey results, there was no statistically significant difference except for higher (worse) scores in effort for HOCC. The NASA-TLX findings imply that our implemented voice-enabled autonomous camera control algorithm performs as well as a human-operated camera control system. According to his subjective comments and the NASA-TLX data, the lone surgeon (expert) in the study much preferred the autonomous camera system. While the results may not have yielded statistically significant differences across all metrics, they do provide valuable insights into the user experience with both the camera control systems. Future research should include larger sample sizes to develop more definitive statistical results. Additionally, the subjective feedback from the participants, including the robotic surgeon, highlights the practical advantages of the voice-enabled autonomous camera control system.

The incorporation of a keyword-triggered language interface into the da Vinci surgical robot holds potential in enhancing surgical outcomes and expanding the accessibility of this technology. Our research presents the evaluation of VACC through a user study, demonstrating that, in many respects, it is comparable to a person operating the camera via joystick. Nevertheless, we envision the prospect of a more intuitive interface in the future which allows for more freeform commands to the system. While our current implementation is rule-based, we anticipate that a well-trained neural network could potentially outperform the human-operated camera control system. The advantage of the human joystick-operated camera system was that free-form complex instructions could be given to the operator. The commands were not limited like they were for the autonomous system. Despite this substantial advantage, the HOCC performance was still on-par with VACC. It is encouraging that automated system performance is approaching that of human-controlled systems.

This study sheds light on the current capabilities of surgical voice-enabled camera systems. Here, we provide some guidelines for future research. A key insight from our research is that automated voice-enabled systems' performance is similar to a human operator under direct instruction. Notably, the preference of the surgeon for the automated voice-enabled system underscores its potential in clinical applications. The limitation of the current study is that currently, autonomous systems will likely fall short of truly expert camera operators. If the operator can anticipate and perform a case without any instruction, this would be ideal. This points to future work where we can use more tools and techniques to learn expert behavior, which would involve much more data collection for predictive analysis. An extension of the current study would be to analyze data from a truly expert operator that knows the surgery (or scene) and can anticipate movement. Future studies should aim to involve more highly trained operators. Such inclusion would provide a more rigorous benchmark, pushing the boundaries for the development of even more sophisticated AI voice-enabled systems. For an autonomous camera system, this would require more advanced AI techniques.

Clearly, more research and advancements are needed, and integration of AI systems may provide a good avenue. Attanasio et al. provide a roadmap for future autonomous system development for surgical robotics and give clear levels of autonomy which will be beneficial [31]. An exciting prospect for these future systems is the integration of advanced generative transformers. These would enable the systems not only to respond to voice commands but also to understand the visual context of the surgical scene, thereby offering intelligent, context-aware camera guidance. This approach would mark a significant leap from reactive to proactive camera control, offering a more intuitive and seamless integration of technology into surgical procedures. These advancements would represent a substantial stride towards creating more autonomous, efficient, and surgeon-friendly robotic systems in the operating room.

We envision a robotic surgery system of the future which will have a natural language interface that helps the surgeon at all stages of surgery. This could make surgeries faster, more accurate, and safer. Progress in image and video analysis within the generative transformers framework (e.g., ChatGPT, Bard) presents a promising path for further research and development, leading to more sophisticated and intelligent surgical systems that can truly collaborate with the surgical team [32,33]. To ascertain the effectiveness and usability of the natural language interface in surgical settings, future studies should be conducted utilizing custom datasets tailored to specific surgical procedures. These datasets could provide surgeons with valuable insights.

This group has integrated our da Vinci system to a generative transformer which offers the opportunity to significantly enhance the interface's capabilities, enabling it to assist surgeons in complex free-form ways [34]. Future interfaces could provide alerts, suggestions, alternative options, patient monitoring, fatigue monitoring, and even manipulation of surgical tools. The future steps include further implementation of a local and more secure natural language processing system with a user study evaluation.

Author Contributions: Conceptualization, A.P.; methodology, M.M. and T.H.-T.; software, L.J. and A.S.; execution, R.A.P., L.J., A.S., M.M., T.H.-T. and A.P.; data curation and analysis, R.A.P.; writing—original draft preparation, R.A.P. and A.P.; writing—review and editing, R.A.P., M.M., L.J. and A.P.; supervision, A.P.; funding acquisition, T.H.-T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported with funding from the Michigan Translational Research and Commercialization (MTRAC), grant number 380137/23R343 and 192748/117E37.

Data Availability Statement: The data is available upon request.

Acknowledgments: We would like to thank David Edelman for sharing his decades of experience as a robotic surgeon to help us understand the true needs in robotic surgery.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Makary, M.A.; Daniel, M. Medical Error—The Third Leading Cause of Death in the US. *BMJ* **2016**, *353*, i2139. [CrossRef] [PubMed]
2. Atkinson, T.M.; Giraud, G.D.; Togioka, B.M.; Jones, D.B.; Cigarroa, J.E. Cardiovascular and Ventilatory Consequences of Laparoscopic Surgery. *Circulation* **2017**, *135*, 700–710. [CrossRef] [PubMed]
3. Weiser, T.G.; Regenbogen, S.E.; Thompson, K.D.; Haynes, A.B.; Lipsitz, S.R.; Berry, W.R.; Gawande, A.A. An Estimation of the Global Volume of Surgery: A Modelling Strategy Based on Available Data. *Lancet* **2008**, *372*, 139–144. [CrossRef] [PubMed]
4. Haidegger, T.; Speidel, S.; Stoyanov, D.; Satava, R.M. Robot-Assisted Minimally Invasive Surgery—Surgical Robotics in the Data Age. *Proc. IEEE* **2022**, *110*, 835–846. [CrossRef]
5. Millan, B.; Nagpal, S.; Ding, M.; Lee, J.Y.; Kapoor, A. A Scoping Review of Emerging and Established Surgical Robotic Platforms With Applications in Urologic Surgery. *Soc. Int. D'urologie J.* **2021**, *2*, 300–310. [CrossRef]
6. Koukourikis, P.; Rha, K.H. Robotic Surgical Systems in Urology: What Is Currently Available? *Investig. Clin. Urol.* **2021**, *62*, 14. [CrossRef] [PubMed]
7. Pandya, A.; Reisner, L.; King, B.; Lucas, N.; Composto, A.; Klein, M.; Ellis, R. A Review of Camera Viewpoint Automation in Robotic and Laparoscopic Surgery. *Robotics* **2014**, *3*, 310–329. [CrossRef]

8. Daneshgar Rahbar, M.; Ying, H.; Pandya, A. Visual Intelligence: Prediction of Unintentional Surgical-Tool-Induced Bleeding during Robotic and Laparoscopic Surgery. *Robotics* **2021**, *10*, 37. [CrossRef]
9. Berguer, R.; Forkey, D.L.; Smith, W.D. The Effect of Laparoscopic Instrument Working Angle on Surgeons' Upper Extremity Workload. *Surg. Endosc.* **2001**, *15*, 1027–1029. [CrossRef]
10. Keehner, M.M.; Tendick, F.; Meng, M.V.; Anwar, H.P.; Hegarty, M.; Stoller, M.L.; Duh, Q.-Y. Spatial Ability, Experience, and Skill in Laparoscopic Surgery. *Am. J. Surg.* **2004**, *188*, 71–75. [CrossRef]
11. Zheng, B.; Cassera, M.A.; Martinec, D.V.; Spaun, G.O.; Swanström, L.L. Measuring Mental Workload during the Performance of Advanced Laparoscopic Tasks. *Surg. Endosc.* **2010**, *24*, 45–50. [CrossRef] [PubMed]
12. Da Col, T.; Mariani, A.; Deguet, A.; Menciassi, A.; Kazanzides, P.; De Momi, E. Scan: System for camera autonomous navigation in robotic-assisted surgery. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October 2020–24 January 2021; pp. 2996–3002.
13. D'Ettorre, C.; Mariani, A.; Stilli, A.; y Baena, F.R.; Valdastri, P.; Deguet, A.; Kazanzides, P.; Taylor, R.H.; Fischer, G.S.; DiMaio, S.P.; et al. Accelerating surgical robotics research: A review of 10 years with the da vinci research kit. *IEEE Robot. Autom. Mag.* **2021**, *28*, 56–78. [CrossRef]
14. Ali, S.M.; Reisner, L.A.; King, B.; Cao, A.; Auner, G.; Klein, M.; Pandya, A.K. Eye Gaze Tracking for Endoscopic Camera Positioning: An Application of a Hardware/Software Interface Developed to Automate Aesop. *Stud. Health Technol. Inform.* **2008**, *132*, 4–7. [PubMed]
15. Casals, A.; Amat, J.; Laporte, E. Automatic Guidance of an Assistant Robot in Laparoscopic Surgery. In Proceedings of the IEEE International Conference on Robotics and Automation, Minneapolis, MN, USA, 22–28 April 1996; Volume 1, pp. 895–900.
16. Mondal, S.B.; Gao, S.; Zhu, N.; Liang, R.; Gruev, V.; Achilefu, S. Real-Time Fluorescence Image-Guided Oncologic Surgery. In *Advances in Cancer Research*; Elsevier: Amsterdam, The Netherlands, 2014; Volume 124, pp. 171–211, ISBN 978-0-12-411638-2.
17. Ko, S.-Y.; Kim, J.; Lee, W.-J.; Kwon, D.-S. Compact Laparoscopic Assistant Robot Using a Bending Mechanism. *Adv. Robot.* **2007**, *21*, 689–709. [CrossRef]
18. Ko, S.Y.; Kwon, D.S. A Surgical Knowledge Based Interaction Method for a Laparoscopic Assistant Robot. In Proceedings of the RO-MAN 2004, 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759), Kurashiki, Japan, 22–22 September 2004; pp. 313–318.
19. Lee, C.; Wang, Y.F.; Uecker, D.R.; Wang, Y. Image Analysis for Automated Tracking in Robot-Assisted Endoscopic Surgery. In Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 9–13 October 1994; Volume 1, pp. 88–92.
20. Omote, K.; Feussner, H.; Ungeheuer, A.; Arbter, K.; Wei, G.-Q.; Siewert, J.R.; Hirzinger, G. Self-Guided Robotic Camera Control for Laparoscopic Surgery Compared with Human Camera Control. *Am. J. Surg.* **1999**, *177*, 321–324. [CrossRef] [PubMed]
21. Uecker, D.R.; Lee, C.; Wang, Y.F.; Wang, Y. Automated Instrument Tracking in Robotically Assisted Laparoscopic Surgery. *J. Image Guide Surg.* **1995**, *1*, 308–325. [CrossRef]
22. Wei, G.-Q.; Arbter, K.; Hirzinger, G. Real-Time Visual Servoing for Laparoscopic Surgery. Controlling Robot Motion with Color Image Segmentation. *IEEE Eng. Med. Biol. Mag.* **1997**, *16*, 40–45. [CrossRef]
23. Gautier, B.; Tugal, H.; Tang, B.; Nabi, G.; Erden, M.S. Real-Time 3D Tracking of Laparoscopy Training Instruments for Assessment and Feedback. *Front. Robot. AI* **2021**, *8*, 751741. [CrossRef]
24. Da Col, T.; Caccianiga, G.; Catellani, M.; Mariani, A.; Ferro, M.; Cordima, G.; De Momi, E.; Ferrigno, G.; De Cobelli, O. Automating endoscope motion in robotic surgery: A usability study on da vinci-assisted ex vivo neobladder reconstruction. *Front. Robot. AI* **2021**, *8*, 707704. [CrossRef]
25. Eslamian, S.; Reisner, L.A.; Pandya, A.K. Development and Evaluation of an Autonomous Camera Control Algorithm on the Da Vinci Surgical System. *Robot. Comput. Surg.* **2020**, *16*, e2036. [CrossRef]
26. Elazzazi, M.; Jawad, L.; Hilfi, M.; Pandya, A. A Natural Language Interface for an Autonomous Camera Control System on the Da Vinci Surgical Robot. *Robotics* **2022**, *11*, 40. [CrossRef]
27. Kazanzides, P.; Chen, Z.; Deguet, A.; Fischer, G.S.; Taylor, R.H.; DiMaio, S.P. An Open-Source Research Kit for the Da Vinci® Surgical System. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 6434–6439.
28. Povey, D.; Ghoshal, A.; Boulianne, G.; Burget, L.; Glembek, O.; Goel, N.K.; Hannemann, M.; Motlícek, P.; Qian, Y.; Schwarz, P.; et al. The Kaldi Speech Recognition Toolkit. In Proceedings of the IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, HI, USA, 11–15 December 2011.
29. Hart, S.G. NASA-Task Load Index (NASA-TLX); 20 years later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [CrossRef]
30. Applied Logistic Regression. Wiley Series in Probability and Statistics. Available online: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118548387 (accessed on 22 December 2023).
31. Attanasio, A.; Scaglioni, B.; De Momi, E.; Fiorini, P.; Valdastri, P. Autonomy in surgical robotics. *Annu. Rev. Control Robot. Auton. Syst.* **2021**, *4*, 651–679. [CrossRef]
32. Gupta, R.; Park, J.B.; Bisht, C.; Herzog, I.; Weisberger, J.; Chao, J.; Chaiyasate, K.; Lee, E.S. Expanding cosmetic plastic surgery research with chatgpt. *Aesthet. Surg. J.* **2023**, *43*, 930–937. [CrossRef]

33. Samaan, J.S.; Yeo, Y.H.; Rajeev, N.; Hawley, L.; Abel, S.; Ng, W.H.; Srinivasan, N.; Park, J.; Burch, M.; Watson, R.; et al. Assessing the accuracy of responses by the language model chatgpt to questions regarding bariatric surgery. *Obes. Surg.* **2023**, *33*, 1790–1796. [CrossRef]
34. Pandya, A. ChatGPT-Enabled daVinci Surgical Robot Prototype: Advancements and Limitations. *Robotics* **2023**, *12*, 97. [CrossRef]