How can the Augmented Reality Experience for Surgeons be Improved During Total Knee Arthroplasty?

Ánxela Pérez Costa, Anna De Liddo, Fridolin Wild, Nieves Pedreira Souto, and Florian Michaud

Laboratory of Mechanical Engineering, Centro de Investigación en Tecnologías Navales e Industriales (CITENI), Campus Industrial de Ferrol, Universidade da Coruña, Ferrol, Spain
Knowledge Media Institute (KMi), The Open University, UK
Institute of Educational Technology, The Open University, UK
Department of Computer Science and Information Technologies, Faculty of Computer Science, CITIC-Research Center of Information and Communication Technologies, Universidade da Coruna, Biomedical Research Institute of A Coruña (INIBIC), A Coruña, Spain
Centro de Investigación CITIC, Universidade da Coruña, 15071 A Coruña, Spain
Correspondence: anxela.perez@udc.es

DOI: *https://doi.org/10.17979/spudc.9788497498913.39*

Abstract: Technological advances in Total Knee Arthroplasty (TKA) have been shown to enhance surgical precision, patient outcomes, and efficiency. New computer-assisted surgery solutions using Augmented Reality (AR) support TKA by locating objects in 3D space with high precision, helping surgeons perform accurate treatments while avoiding time-consuming or inaccurate traditional methods. These advances, however, also present novel challenges, such as the complexity of integrating new systems with existing surgical tools and workflows. In this study, the authors propose a strategy to address surgeons' concerns by leveraging feedback from 50 surgeons and 41 sales representatives. Based on this, a research plan was developed to further enhance the AR experience for surgeons.

1 Introduction

Technological advances in Total Knee Arthroplasty (TKA) have revolutionized the field, enhancing surgical precision, patient outcomes, and overall efficiency. Innovations such as robotic-assisted surgery, computer navigation, and Augmented Reality (AR) are at the forefront. New computer-assisted orthopaedic surgery solutions using Augmented Reality support TKA by locating objects in a 3D space with good precision, helping surgeons perform accurate treatments while avoiding time-consuming or inaccurate traditional methods (Bennett et al., 2023). Nevertheless, while these advances offer significant benefits, they also present several challenges. One key challenge is that integrating new systems with existing surgical tools and workflows can be complex and time-consuming (Lex et al., 2024). Moreover, surgeons and medical staff may require extensive training to effectively use these new technologies, potentially slowing down adoption. For these reasons, it is important to consider the user's interaction experience, aiming to both make the application as user-friendly as possible, while reducing its learning curve. In addition, human factors must be considered when proposing new surgical tools. Not only for ergonomic reasons, but also because poor interaction can lead to severe complications.

Several studies analysed personal and organizational factors during surgery, focusing on surgeons' perceptions and cognitive skills using various traditional techniques (Kohls-Gatzoulis et al., 2004; Lowndes and Hallbeck, 2014). Recent publications also examined the use of robot-assisted surgery (Luko et al., 2017; Talamini et al., 2021). The results of these studies highlight that considering human factors can improve surgical outcomes, prevent errors, reduce fatigue, and enhance communication. On the other hand, numerous investigations have sought to deepen understanding of factors influencing user experience in virtual environments (Liberatore and Wagner, 2021). In Li et al. (2023), the authors present a comparative study systematically investigating the impact of 2D and 3D user interfaces on repetitive tasks in AR. They developed prototypes of these interfaces and conducted empirical evaluations focusing on cognitive load, perceived usability, and learnability. Merino et al. (2020) conducted a systematic review of technology-centric performance evaluations and human-centric studies of mixed and augmented reality (MR/AR) to offer insights for future evaluations of MR/AR approaches. Finally, Birlo et al. (2022) provided a summary of current research on head-mounted display usage in Augmented Reality for surgery and examined potential barriers to clinical adoption of such devices. These findings underscore the challenges posed by this emerging technology, with each application requiring tailored attention.

In this study, to address surgeons' concerns regarding the use of AR head-mounted displays in TKA, the authors propose a strategy that leverages both quantitative and qualitative data provided by a company offering Augmented Reality solutions for TKA. This approach aims to gain insight into potential improvement of user experience and minimization of error risks. Building on these findings, a plan for further research is elaborated against the backdrop of an extensive literature review. This plan is focused on optimizing the Augmented Reality experience for surgeons, enhancing the system's usability and effectiveness in surgical practice.

2 Material and methods

2.1 Experimental data collection

Participants

A total of 50 surgeons and 41 sales representatives (responsible for managing product demonstrations, addressing customer inquiries, and facilitating the integration of the AR solution into surgical practices) were evaluated, including 36 surgeons and 31 sales representatives from the European Union. The sample included both novice and experienced users.

Experiment and survey

Participants used a commercial computer-assisted orthopedic surgery solution during total knee arthroplasty. By wearing smart glasses, the surgeon could view a virtual model aligned with the patient's bone, enabling precise positioning of the instrumentation according to the operative plan. The company's evaluations involved a series of predefined tasks designed to cover various scenarios, including not only surgical performance but also device settings, communication, and data access. Evaluators observed and reported on the users' performance, and after each session, participants completed a questionnaire, generating both quantitative and qualitative data. Most evaluations took place in the surgeons' operating rooms, except for the initial 7 surgeons and 5 sales representatives outside the EU, who were assessed in an office or exhibition room using a training setup.

2.2 Data analysis

Quantitative data

Quantitative data focused on task performance, providing numerical measures of human error frequency. Each step of the task was scored using a 4-point Likert scale (Nagata et al., 1996) with the following values: 0 (correct use), 1 (close to error), 2 (erroneous use), and 3 (help required). These numerical assessments were used to derive quantitative data on the tasks. The percentage of users who encountered errors was calculated in relation to the total number of users, with the errors categorized according to the severity scale. Additionally, the percentage of task steps where errors occurred was calculated in relation to the total number of steps. Evaluations were conducted both during the device's settings configuration and the surgical protocol. For this study, however, we focused exclusively on results from the 22 tasks related to the surgical procedure, as it represents the most critical and relevant phase.

Qualitative data

Qualitative data provided rich, detailed feedback that uncovered the underlying reasons for user difficulties, preferences, and areas for improvement. This approach allowed for a more nuanced and flexible interpretation of user feedback, identifying specific themes and insights that quantitative metrics might overlook. The evaluation was conducted using the think-aloud technique, followed by post-test questionnaires, which captured immediate feedback on each step of the process. A total of 140 statements were extracted from these evaluations. Given the non-standardized nature of the feedback, a specific strategy was necessary. A thematic analysis was performed using iterative deductive coding and sub-coding (Butler et al., 2018; Willson, 2019) to categorize the 140 statements. In the first phase, statements were grouped by topic similarity, with each assigned a code and differentiated into positive or negative comments. In the second phase, these codes were further grouped into broader categories, allowing for a deeper review of the statements within each category.

Research plan

Data from the evaluations revealed surgeons' concerns and uncovered interface issues with the current AR head-mounted display solution in TKA, highlighting critical areas for further improvement. In the medical field, more than in any other, reducing the potential for interface errors is paramount, as even minor mistakes can have significant consequences for patient safety and even lead to complications. As technological advances enable the development of new interfaces, it is essential to create a research plan that introduces solutions aimed at minimizing the risk of error without increasing the learning curve, ensuring both safety and efficiency in the surgical process.

To develop a comprehensive research plan, a review of the relevant literature was conducted, identifying existing research gaps. Subsequently, insights are synthesized from both literature review and the evaluation. This iterative process led to the formulation of specific, clear, and feasible research questions.

3 Results

3.1 Quantitative outcomes

The quantitative data analysis revealed that the majority of users (73.2%) successfully completed all 22 surgery-related tasks without making any errors. Additionally, 3.6% of users came close to making an error (classified as an "almost-mistake"), and 1.8% made at least one actual error for selecting or validating certain actions. Furthermore, 23.2% of users encountered some difficulties during the process. Notably, all users were able to complete 15 of the tasks (68.2%) without errors. Three tasks (13.6%) saw at least one user nearly make a mistake, while two tasks (9.1%) had at least one user who made an actual error. Lastly, eight tasks (36.4%) presented difficulties for at least one user.

Finally, it is important to highlight that the application is fundamentally safe, and the mistakes observed during the evaluations occurred in the selection or validation of certain actions. These errors did not compromise the reliability or safety of the system. Instead, they only resulted in minor delays, as users had to revisit certain steps, but they did not affect the overall accuracy or functionality of the application.

3.2 Qualitative outcomes

Through thematic analysis, the authors extracted detailed insights from the data, identifying key areas of concern by analyzing and breaking down smaller, underlying issues.

This process led to the identification of 34 topics of interest to surgeons, derived from 140 qualitative statements, which were then organized into eight overarching themes (see figure 1): Features and Data; General; Hardware; Technique; Usability; User Interface; Relationship with Real Environment (RE); and Reliability. Some topics differentiate between positive and negative statements, offering insights into both challenges and solutions concerning surgeons.

Overall, most statements were positive, reflecting a high level of user satisfaction with the current solution but also highlight the topics of interest to surgeons for future developments.

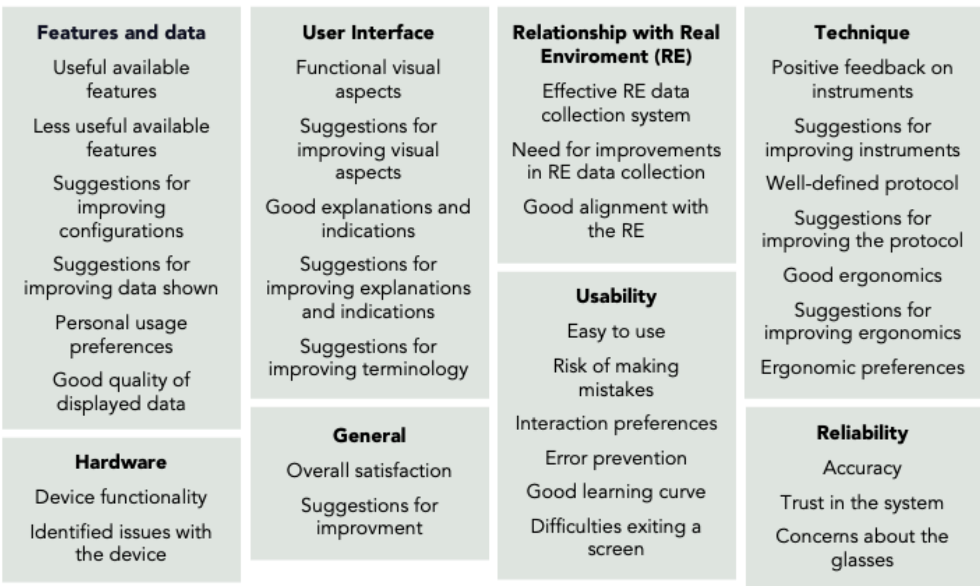

Figure 1: List of the 34 topics of interest to surgeons organized in 8 overarching themes

3.3 Research questions

Based on both the quantitative and qualitative analyses, research questions (RQs) were developed to address the key findings. The quantitative results revealed that 23.2% of users encountered difficulties, and 1.8% made actual errors, emphasizing the first RQ:

RQ1: What human factors affect the use of AR HMD for TKA, and how can these be systematically evaluated? The thematic analysis further highlighted usability and user interface challenges as critical areas of concern, justifying the exploration of how these human factors can be optimized for surgical performance (Bolam et al., 2022; Gestel et al., 2021; Jenny et al., 2008), leading to the second RQ:

RQ2: What is the learning curve for AR HMD in TKA, and how can it be optimized for surgical performance using advanced interaction methods? Additionally, the qualitative insights identified 34 topics, including the need for improved user interaction methods(Murthy, 2020). This highlights the importance of enhancing user experience (UX) with AR HMD to reduce error rates and improve outcomes, resulting in the third RQ:

RQ3: How does user experience (UX) with AR HMD and new interaction methods impact surgical outcomes and error rates? Finally, with 36.4% of tasks presenting difficulties for at least one user, there is a clear need for improved interface design and refined training strategies (Lowndes and Hallbeck, 2014; Theelen et al., 2018). This need is addressed in the fourth RQ:

RQ4: How can interface design, interaction methods, and training strategies be improved to enhance AR HMD usability and reduce the learning curve in TKA?

3.4 Research plan

From these RQ, a research plan (see Figure 2) was developed to optimize the surgical process, aiming to reduce procedure time and minimize the risk of potential errors, without compromising the learning curve or user experience.

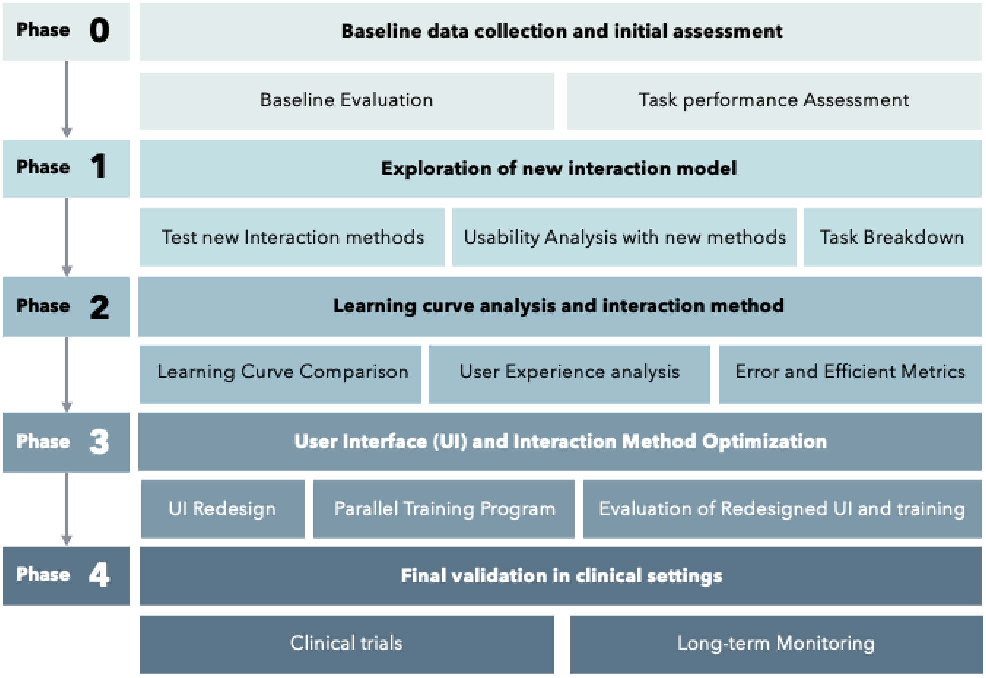

Figure 2: Research plan to optimize the surgical process and minimize the risk of potential errors

4 Limitations

Both quantitative and qualitative data were provided by a company offering Augmented Reality solutions for TKA. While the quantitative data helped identify surgeons' difficulties and potential error risks. On the other hand, outcomes from the qualitative evaluations provided were more limited due to the non-standardized nature of the evaluations, highlighting the need for a more structured methodology. In addition, since this is the first computer-assisted orthopedic surgery solution using AR to support TKA, and often users' first AR experience, it remains challenging for users to provide outcomes or comparisons with other existing solutions. Testing novel advanced interaction methods with users will undoubtedly yield more valuable feedback.

5 Conclusions

This study provides a baseline evaluation of an existing AR HMD solution for TKA and proposes a research plan aimed at optimizing the learning curve, enhancing user experience, and reducing surgical errors by integrating novel advanced interaction methods. Through this, the authors aim to ensure that surgeons can adopt new technology more seamlessly, with minimal risk to patient outcomes and enhanced overall efficiency.

Acknowledgements

Ánxela Pérez Costa and Florian Michaud would like to acknowledge the support of the Galician Government and the Ferrol Industrial Campus by means of her predoctoral research contract 2023/CP/209 and his postdoctoral research contract 2022/CP/048.

Bibliography

K. M. Bennett, A. Griffith, F. Sasanelli, I. Park, and S. Talbot. Augmented reality navigation can achieve accurate coronal component alignment during total knee arthroplasty. *Cureus*, 15(2), 2023.

M. Birlo, P. E. Edwards, M. Clarkson, and D. Stoyanov. Utility of optical see-through head mounted displays in augmented reality-assisted surgery: A systematic review. *Medical Image Analysis*, 77:102361, 2022.

S. M. Bolam, M. L. Tay, F. Zaidi, R. P. Sidaginamale, M. Hanlon, J. T. Munro, and A. P. Monk. Introduction of rosa robotic-arm system for total knee arthroplasty is associated with a minimal learning curve for operative time. *Journal of Experimental Orthopaedics*, 9, 12 2022.

Z. Butler, I. Bezáková, and K. Fluet. Analyzing rich qualitative data to study pencil-puzzle-based assignments in cs1 and cs2. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, pages 212–217, 2018.

F. V. Gestel, T. Frantz, C. Vannerom, A. Verhellen, A. G. Gallagher, S. A. Elprama, A. Jacobs, R. Buyl, M. Bruneau, B. Jansen, J. Vandemeulebroucke, T. Scheerlinck, and J. Duerinck. The effect of augmented reality on the accuracy and learning curve of external ventricular drain placement. *Neurosurgical Focus*, 51:1–9, 8 2021.

J. Y. Jenny, R. K. Miehlke, and A. Giurea. Learning curve in navigated total knee replacement. a multi-centre study comparing experienced and beginner centres. *Knee*, 15:80–84, 3 2008.

J. A. Kohls-Gatzoulis, G. Regehr, and C. Hutchison. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. *Canadian journal of surgery*, 47(4):277, 2004.

J. R. Lex, J. I. Wolfstadt, A. R. Cohen-Rosenblum, D. C. Landy, and J. A. Bernstein. Implementing new technology in your arthroplasty practice. *The Journal of Arthroplasty*, 39(6):1385–1388, 2024.

X. Li, C. Zheng, Z. Pan, Z. Huang, Y. Niu, P. Wang, and W. Geng. Comparative study on 2d and 3d user interface for eliminating cognitive loads in augmented reality repetitive tasks. *International Journal of Human–Computer Interaction*, pages 1–17, 2023.

M. J. Liberatore and W. P. Wagner. Virtual, mixed, and augmented reality: a systematic review for immersive systems research. *Virtual Reality*, 25(3):773–799, 2021.

B. R. Lowndes and M. S. Hallbeck. Overview of human factors and ergonomics in the or, with an emphasis on minimally invasive surgeries. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 24(3):308–317, 2014.

L. Luko, A. Parush, and L. Lowenstein. Cognitive task analysis of spatial skills in hysterectomy with the da vinci surgical system. In *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, pages 100–107. IEEE, 2017.

L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 438–451. IEEE, 2020.

L. R. Murthy. Multimodal interaction for real and virtual environments. In *International Conference on Intelligent User Interfaces, Proceedings IUI*, pages 29–30. Association for Computing Machinery, 3 2020. ISBN 9781450375139.

C. Nagata, M. Ido, H. Shimizu, A. Misao, and H. Matsuura. Choice of response scale for health measurement: comparison of 4, 5, and 7-point scales and visual analog scale. *Journal of epidemiology*, 6(4):192–197, 1996.

S. Talamini, W. R. Halgrimson, R. W. Dobbs, C. Morana, and S. Crivellaro. Single port robotic radical prostatectomy versus multi-port robotic radical prostatectomy: A human factor analysis during the initial learning curve. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 17(2):e2209, 2021.

L. Theelen, C. Bischoff, B. Grimm, and I. C. Heyligers. Current practice of orthopaedic surgical skills training raises performance of supervised residents in total knee arthroplasty to levels equal to those of orthopaedic surgeons. *Perspectives on Medical Education*, 7:126–132, 4 2018.

R. Willson. Analysing qualitative data: you asked them, now what to do with what they said. In *Proceedings of the 2019 conference on human information interaction and retrieval*, pages 385–387, 2019.