Effects of Visual, Auditory, and Haptic Feedback on Gesture Interaction in the Operating Room

Nicole Steinke†[1], Andre Muehlenbrock†[1], Merle Schlender[2], Tim Schneider[2], Verena Uslar[2], Dirk Weyhe[2], Rene Weller[1], and Gabriel Zachmann[1]

[1] Computer Graphics and Virtual Reality Lab, University of Bremen, Germany
{steinke,muehlenb,rweller,zach}@uni-bremen.de
[2] University Clinic for Visceral Surgery, University of Oldenburg, Germany

Abstract. An increasing number of studies explore gesture control as a sterile and intuitive input modality for medical devices to increase surgeons' autonomy and reduce errors caused by miscommunication with assisting staff. In this context, gesture feedback is critical, as surgeons must quickly and unambiguously perceive whether an input has been correctly registered. Prior work has compared visual, auditory, and haptic feedback modalities for gesture-based interaction, but often neglects the specific environmental constraints of the operating room (OR), such as high cognitive load and dense acoustic noise from medical equipment. In this paper, we investigate visual, auditory, and haptic gesture feedback under simulated OR conditions. We report on two controlled studies conducted in an OR context with non-surgeons (n = 25) and practicing surgeons (n = 10). Across visual, auditory, and haptic feedback modalities, perceived workload was comparable. However, participants' qualitative feedback revealed that haptic cues were often perceived as advantageous in terms of unambiguity and reliability in noisy OR environments, although preferences varied substantially across individuals. While performance differences did not reach statistical significance, an observable trend toward higher performance with haptic feedback warrants further investigation and may have practical implications for its use in the OR.

† These authors contributed equally to this work.

Keywords: Gesture Control · Haptic Feedback · Operating Room.

1 Introduction

In recent years, research on interaction techniques in the operating room (OR) has increasingly focused on touchless input modalities [1]. Among these, mid-air gesture control has emerged as a promising approach for enabling sterile interaction with operating room infrastructure, such as surgical lights, medical devices, or imaging systems [4, 11, 13–15, 18, 19]. By allowing surgeons to directly manipulate systems without physical contact, gesture-based interfaces aim to support intraoperative workflows and reduce dependencies on assisting staff, where verbal coordination and handovers are known sources of errors [9].

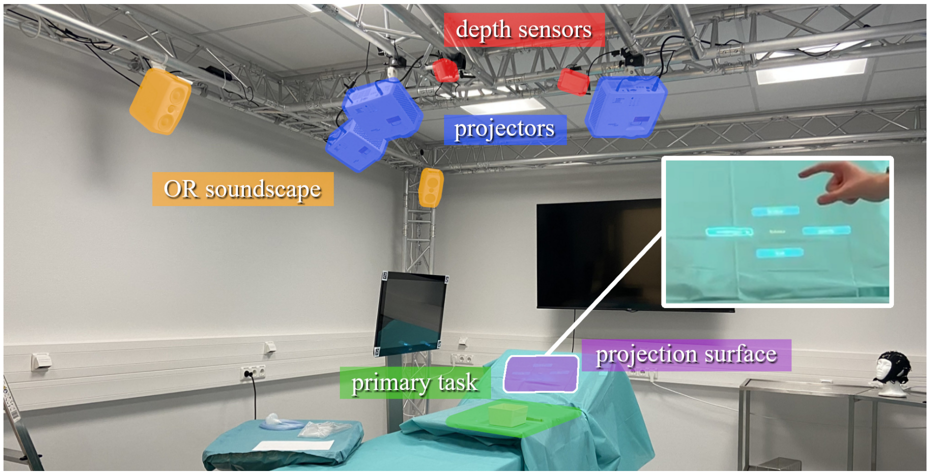

Fig. 1: Research operating-room setup with RGB-D–tracked projection system, six-speaker audio for realistic OR acoustics, and an operating table used as projection surface and primary-task workspace.

For gesture-based interaction to be viable in the operating room, appropriate feedback is essential. Surgeons must be able to quickly and confidently determine whether a gesture has been detected and interpreted correctly, often under time pressure and while maintaining visual and cognitive focus on the surgical task. Feedback therefore plays a critical role in ensuring both usability and safety.

Previous research has investigated visual, auditory, and haptic feedback modalities for gesture-based interaction across a range of application domains and has examined perceptual differences between different types of feedback [3, 5–7, 16]. However, much of this work has been conducted in controlled laboratory settings or everyday environments and does not sufficiently account for the specific constraints of the operating room. In particular, surgeons operate under high cognitive load, limited visual attention, and a dense acoustic environment created by medical equipment and alarms. These factors may substantially influence how different feedback modalities are perceived and how reliably they can be used in practice.

In this paper, we address this research gap by systematically investigating visual, auditory, and haptic feedback for mid-air gesture interaction under simulated operating room conditions. We focus on how different feedback modalities affect perceived workload and user preference in a simulated OR environment that incorporates key characteristics of real surgical settings, including acoustic conditions. To this end, we conducted two controlled studies: (1) one with non-surgeons to capture general usability effects (n=25), and (2) one with practicing surgeons to account for expert perspectives (n=10).

(a) Interior hardware (b) Vibration motor (c) Worn device

Fig. 2: Custom-built wristband for haptic feedback, containing an nRF52840+, a CMOS battery, a vibration motor, and a Velcro strap for attachment.

2 Experimental Setup

To investigate feedback modalities in high-stakes environments such as ORs, participants performed a primary task representing a surgical procedure while completing a Stroop test as a secondary task via gesture-based interaction with a projected display. Due to the limited availability and time constraints of practicing surgeons, the study was conducted with both non-surgical participants and surgeons under slightly different conditions.

2.1 Environment and Apparatus

The experiment was conducted in a medical living lab designed to simulate key characteristics of an operating room, see Figure 1. To simulate a realistic OR soundscape, a six-speaker system was used to reproduce authentic audio recordings obtained from actual surgical environments. The acoustic background included ventilation noise, staff movement, and sounds generated by medical equipment. Sound pressure levels at the participants' position were standardized to 70 dB SPL, reflecting noise levels commonly reported in a clinical OR [12]. Auditory feedback consisted of a short confirmation sound designed for medical environments, delivered via a speaker integrated into the projection system, positioned approximately 1–1.5 m from the participant's head, and thus spatially separated from the OR soundscape speakers.

The graphical user interface was projected onto an inclined surface mimicking a suspended surgical drape using a projection system for uneven, deformable surfaces [14]. Hand movements were captured via ceiling-mounted RGB-D sensors (Orbbec Femto Bolt) that observed a predefined interaction volume adjacent to the participant.

2.2 Wristband for Haptic Feedback

During all conditions, participants wore a custom-built wristband on their dominant hand. The wristband included a Bluetooth receiver and provided haptic

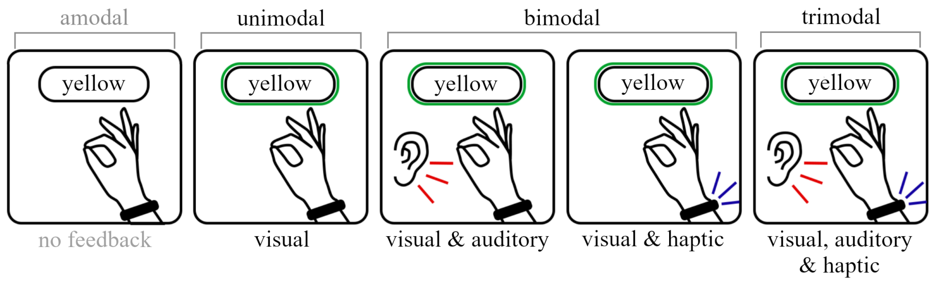

Fig. 3: Five experimental conditions comprising no feedback, visual feedback, and combined feedback modalities. Owing to surgeons' time constraints, the no-feedback condition was tested only with non-surgical participants.

feedback by vibrating upon a registered click in the respective conditions. It consisted of a 3D-printed enclosure with a battery holder, a low-energy XIAO nRF52840+ mainboard connected to the host computer via Bluetooth, and a single 3 V DC vibration motor (10,000 rpm, 80 mA), which was precisely controlled using a MOSFET, see Figure 2. Bluetooth parameters were configured to achieve a balance between responsiveness and energy efficiency, resulting in a maximum latency of approximately 40 ms and an average latency of about 20 ms. Preliminary testing indicated no perceptible delay between click registration and vibration onset. Each registered click triggered a 100 ms vibration.

2.3 Gesture-Based Interaction

Hand motion within the interaction volume was mapped in real time to a cursor on the projected interface. Selection was triggered via a deliberate double pinch gesture, defined as two rapid successive contacts between thumb and index finger. This gesture was chosen to reduce accidental activations and to increase intentionality of input, which is particularly relevant in sterile, touchless interaction scenarios.

Gesture-based interaction was implemented using MediaPipe [10], which provided a 2D hand skeleton representation. Landmarks around the palm were projected into the depth image to compute an averaged 3D hand position, which was then mapped to a corresponding cursor location on the projected display. A click was identified by measuring the distance between the thumb and index fingertip and detecting two threshold crossings within a 1.35-second time window, as determined as an appropriate interval in preliminary testing. To ensure reliable detection, participants were instructed to orient their hand appropriately toward the depth camera.

2.4 Feedback Conditions

Five feedback conditions were evaluated using a within-subject design (see Figure 3): (1) *no feedback* (none), (2) *visual feedback* only (visual highlight of the selected item), (3) *visual & auditory feedback* (highlight plus confirmation sound), (4) *visual & haptic feedback* (highlight plus vibrotactile pulse), and (5) *visual, auditory & haptic* feedback. Condition order was counterbalanced using a Latin square to mitigate potential learning effects, and a familiarization phase was included during which all feedback modalities were tested twice.

Because visual feedback provides precise and unambiguous spatial information and plays an inherent role in practical OR interfaces, we treated it as a baseline modality in all non-baseline conditions. Consequently, the study focuses on the additive effects of additional auditory and haptic feedback on top of visual feedback, rather than isolating these modalities independently.

The no-feedback condition served as a baseline control in the non-surgeon study to validate that feedback generally improves performance and reduces workload – a question conclusively answered in that group (all $p < 0.01$). As the primary aim of the surgeon study was the comparison among feedback modalities rather than establishing the general necessity of feedback, this condition was omitted for surgeons, also in consideration of their strictly limited availability.

2.5 Task Design and Cognitive Load

To approximate surgical workload, participants sorted 1×2 LEGO® bricks from a mixed container using tweezers (primary task). This activity was intermittently interrupted by a Stroop test displayed on a projected interface (secondary task). The Stroop test required identifying the font color of a color word while ignoring its semantic meaning, thereby inducing cognitive interference and attentional demand [2, 17]. Responses were provided via gesture-based input. The dual-task paradigm was designed to reflect two aspects of intraoperative work: continuous fine-motor activity that limits visual attention, and parallel cognitive demands requiring selective attention.

The primary task was performed for 25 seconds by non-surgical participants and 20 seconds by practicing surgeons before each Stroop test was presented automatically. After responding, participants immediately resumed the primary task. Per condition, the Stroop test was administered seven times for non-surgical participants and five times for surgeons due to time constraints.

2.6 Data Collection and Error Induction

Each condition lasted approximately four to five minutes. To ensure the relevance of the feedback, a recognized gesture-based click was intentionally ignored by the implemented system in two of the seven Stroop tests for non-surgical participants and in one of the five Stroop tests for practicing surgeons, thereby requiring a repeated selection.

After each condition, subjective workload was assessed using the NASA-TLX (Raw) test [8], complemented by additional ratings addressing perceived input confidence and convenience of the feedback. After all conditions were completed, participants reported their preferences and provided qualitative feedback.

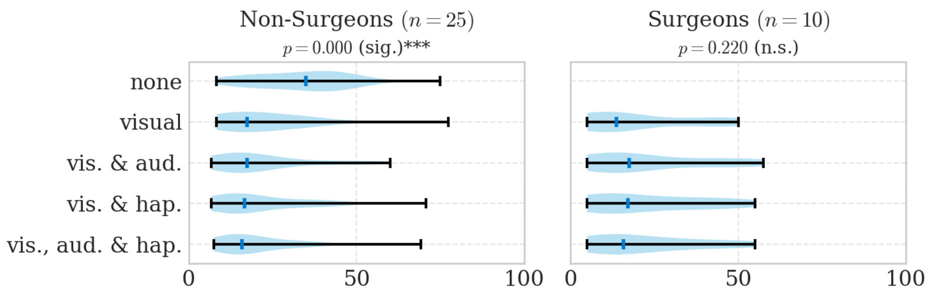

Fig. 4: Raw NASA-TLX scores. p-values were obtained using the Friedman test.

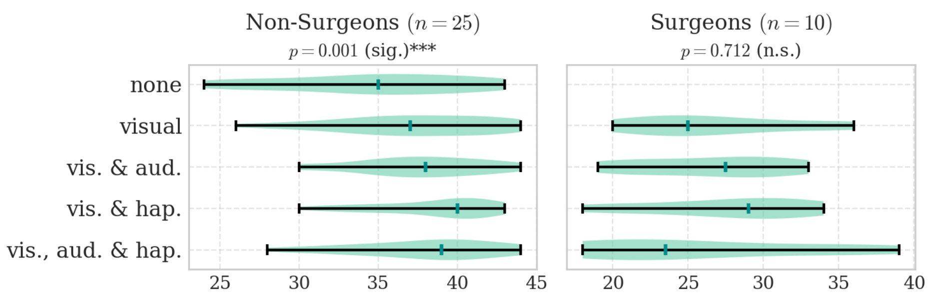

Fig. 5: Number of found bricks (primary task). Due to differences in experimental conditions (e.g., task duration), the Non-Surgeon and Surgeon groups are not directly comparable; interpretation is therefore restricted to within-group differences between conditions. p-values were obtained using the Friedman test.

2.7 Participants

Non-surgical participants (n = 25; 13 male, 12 female; mean age: 29 years) were recruited via posted advertisements and direct inquiry. Surgical participants (n = 10; 6 male, 4 female; mean age: 35.8 years) were recruited at the Pius Hospital, Oldenburg, Germany. To increase sample size, two participants in their final year of training with prior clinical experience were included.

3 Results

For the analysis of the raw NASA-TLX scores, the six subscales were equally weighted and aggregated per condition. In all significance tests, Wilcoxon signed-rank tests with Holm correction were applied. The resulting distributions across participants are shown in Figure 4. The 'none' condition (i.e., no feedback) resulted in a significantly higher perceived workload compared to all other conditions (all $p < 0.01$, $r \in [0.72, 0.86]$). No significant differences were observed among the remaining feedback conditions; however, effect sizes comparing the visual-only condition to the three augmented visual conditions ranged from

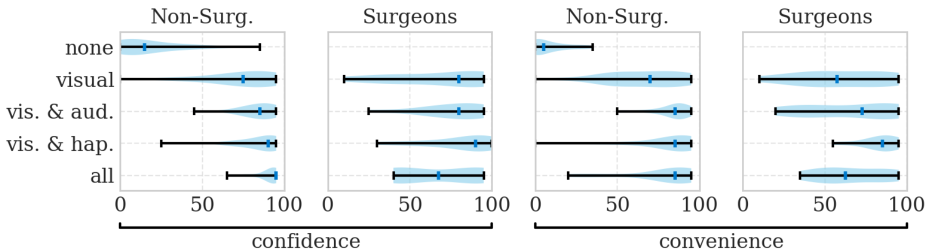

Fig. 6: Scores of 'confidence' and 'convenience'.

$r \in [0.23, 0.50]$, and effect sizes among the augmented visual conditions ranged from $r \in [0.04, 0.27]$ for non-surgeons. No significant differences were observed in the surgeon group, though effect sizes of $r \in [0.05, 0.60]$ were observed.

A comparable pattern was observed for the number of correctly identified LEGO® bricks. All feedback conditions differed significantly from the 'none' condition ($p < 0.05$, $r \in [0.58, 0.67]$). Descriptive non-significant trends suggest potential performance improvements from visual-only to augmented visual conditions ($r \in [0.11, 0.35]$) with the largest effect observed for visual & haptic feedback (see Figure 5), though effect sizes among the augmented visual conditions remained small ($r \in [0.06, 0.26]$). No significant effects were found in the surgeon group ($r \in [0.04, 0.34]$).

With respect to perceived confidence (see Figure 6), all feedback conditions differed significantly from the 'none' condition in the non-surgeon group ($p < 0.01$, $r \in [0.84, 0.88]$). Additionally, the comparison between all and visual feedback reached significance ($p = 0.004$, $r = 0.82$, after Holm correction). While no further comparisons survived correction, effect sizes of $r \in [0.25, 0.58]$ suggest potential trends that may reach significance with larger sample sizes, which is of particular relevance in an operating room context. Effects for convenience ratings (see Figure 6) were generally weaker but again showed significant differences between the 'none' condition and all feedback conditions ($r \in [0.87, 0.88]$), as well as between visual & auditory and visual feedback for non-surgeons ($p = 0.01$, $r = 0.67$). No significant differences were observed in the surgeon group after Holm correction, though individual comparisons for confidence and convenience showed larger effect sizes ($r \in [0.07, 0.8]$).

The ranking of feedback modalities (see Figure 7) revealed substantial inter-individual variability in preferences among visual & auditory, visual & haptic, and visual, auditory & haptic feedback. This heterogeneity was also reflected in participants' qualitative comments. While some participants strongly preferred haptic feedback—often citing prior experience with wearable devices and the reduced audibility of audio cues in noisy environments—others perceived haptic feedback as unfamiliar or uncomfortable and instead favored auditory feedback. Notably, whereas non-surgeons frequently preferred multimodal feedback, surgeons more often described the combination of all three modalities as overly intrusive.

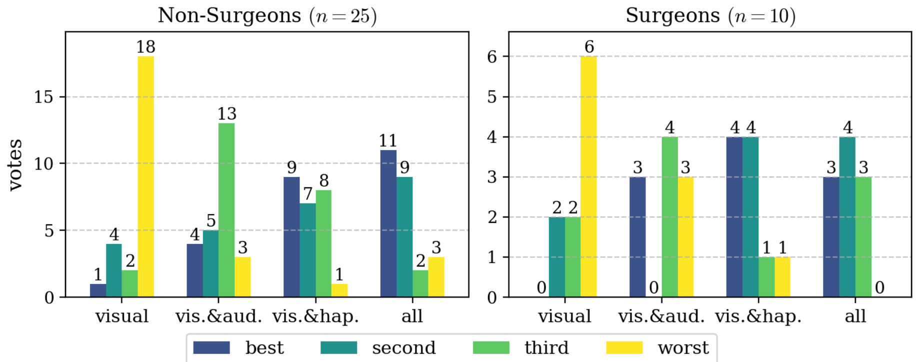

Fig. 7: Rankings of the conditions by the participants from Best to Worst. The fifth 'none' condition in the Non-Surgeons study was omitted because all participants ranked it as the worst category.

Overall, participants emphasized the benefit of augmenting visual feedback with either auditory or haptic cues, as this allowed them to resume the primary task immediately without actively checking the visual display. Although preferences between auditory and haptic feedback differed across participants, haptic feedback was commonly regarded as more unequivocal, given that it cannot be masked by or mistaken for ambient sounds in the operating room.

4 Discussion & Conclusion

While preferences derived from both qualitative and quantitative feedback appear to be highly individual and influenced by prior experience with wearable devices – may be an interesting finding in and by itself – the performance-based results reveal consistent, albeit not statistically significant, trends suggesting potential advantages of haptic feedback when combined with visual feedback. One possible explanation for these observations are differences in attentional demands: whereas visual feedback requires explicit visual attention, auditory and haptic feedback may allow users to resume the primary task more quickly. In direct comparison, haptic feedback provides a clearly distinguishable signal, while auditory feedback may require additional cognitive effort in noisy environments.

A limitation of our design is that auditory and haptic feedback were only evaluated in combination with visual feedback. Consequently, our results do not allow conclusions about their independent effectiveness, but rather about their added value when augmenting visual feedback. While this reflects realistic operating room scenarios, it limits the interpretability of modality-specific effects and prevents direct comparisons between unimodal conditions.

A further limitation is the omission of the 'no feedback' condition in the surgeon group, which prevents direct conclusions about expert-specific reliance on feedback. In our design, this condition primarily served as a control to establish the general importance of feedback. Moreover, the absence of feedback represents a less realistic interaction scenario in practice, as it would hinder the detection of missed or incorrectly registered inputs. Future work should explicitly investigate whether expert users can better compensate for missing feedback.

Haptic feedback introduces a distinct and currently underutilized communication channel in the operating room, whereas the auditory channel is often heavily occupied. At the same time, strong individual preference effects highlight the need for adaptable feedback configurations. Future work should investigate unimodal feedback conditions, include expert baselines, employ tasks with lower variability to better isolate performance effects, and validate findings in real surgical environments.

Acknowledgments. This study was partially funded by BMFTR grant 16SV9239.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article. The study protocol was approved by the responsible ethics committee prior to data collection.

References

1. Alvarez-Lopez, F., Maina, M.F., Saigí-Rubió, F.: Use of commercial Off-The-Shelf devices for the detection of manual gestures in surgery: Systematic literature review. J Med Internet Res **21**(5), e11925 (Apr 2019). https://doi.org/10.2196/11925
2. Cabañero Gómez, L., Hervás, R., González, I., Villarreal, V.: Studying the generalisability of cognitive load measured with eeg. Biomed. Signal Process. Control **70**, 103032 (2021)
3. Cornelio Martinez, P.I., De Pirro, S., Vi, C.T., Subramanian, S.: Agency in mid-air interfaces. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. p. 2426–2439. CHI '17, ACM, New York, NY, USA (2017)
4. Dietz, A., Schröder, S., Pösch, A., Frank, K., Reithmeier, E.: Contactless surgery light control based on 3d gesture recognition. In: Benzmüller, C., Sutcliffe, G., Rojas, R. (eds.) GCAI 2016. 2nd Global Conference on Artificial Intelligence. EPiC Series in Computing, vol. 41, pp. 138–146. EasyChair (2016)
5. Dube, T.J., Ren, Y., Limerick, H., MacKenzie, I.S., Arif, A.S.: Push, tap, dwell, and pinch: Evaluation of four mid-air selection methods augmented with ultrasonic haptic feedback. Proc. ACM Hum.-Comput. Interact. **6**(ISS) (Nov 2022)
6. Evangelou, G., Georgiou, O., Brown, E., Hine, N., Moore, J.: Mid-air haptic feedback improves implicit agency and trust in gesture-based automotive infotainment systems: a driving simulator study. In: Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. p. 116–124. AutomotiveUI '24, ACM, New York, NY, USA (2024)
7. Evangelou, G., Georgiou, O., Brown, E., Moore, J.W.: Sense of agency in gesture-based interactions: modulated by sensory modality but not feedback meaning. Front. Comput. Sci. **7** (2025)
8. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, Advances in Psychology, vol. 52, pp. 139–183. North-Holland (1988)

9. Lingard, L., Espin, S., Whyte, S., Regehr, G., Baker, G.R., Reznick, R., Bohnen, J., Orser, B., Doran, D., Grober, E.: Communication failures in the operating room: an observational classification of recurrent types and effects. Qual. Saf. Health Care **13**(5), 330–334 (Oct 2004)
10. Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.L., Yong, M., Lee, J., Chang, W.T., Hua, W., Georg, M., Grundmann, M.: Mediapipe: A framework for perceiving and processing reality. In: Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR) 2019 (2019)
11. Madapana, N., Chanci, D., Gonzalez, G., Zhang, L., Wachs, J.P.: Touchless interfaces in the operating room: A study in gesture preferences. Int. J. Hum.-Comput. Interact. **39**(3), 438–448 (2023)
12. Mcleod, R., Myint-Wilks, L., Davies, S., Elhassan, H.: The impact of noise in the operating theatre: a review of the evidence. Ann. R. Coll. Surg. Engl. **103**(2), 83–87 (2021), pMID: 33559553
13. Mentis, H.M., O'Hara, K., Gonzalez, G., Sellen, A., Corish, R., Criminisi, A., Trivedi, R., Theodore, P.: Voice or gesture in the operating room. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. p. 773–780. CHI EA '15, ACM, New York, NY, USA (2015)
14. Muehlenbrock, A., Purgin, Y., Steinke, N., Uslar, V., Weyhe, D., Weller, R., Zachmann, G.: Shadow-free projection with blur mitigation on dynamic, deformable surfaces. In: Proceedings of the 2025 31st ACM Symposium on Virtual Reality Software and Technology. VRST '25, ACM, New York, NY, USA (2025)
15. Mühlenbrock, A., Huscher, H., Uslar, V.N., Cetin, T., Weller, R., Weyhe, D., Zachmann, G.: A novel, autonomous, module-based surgical lighting system. ACM Trans. Comput. Healthc. **6**(1) (Jan 2025)
16. Remizova, V., Sand, A., MacKenzie, I.S., Špakov, O., Nyyssönen, K., Rakkolainen, I., Kylliäinen, A., Surakka, V., Gizatdinova, Y.: Mid-air gestural interaction with a large fogscreen. Multimodal Technol. Interact. **7**(7) (2023)
17. Stroop, J.R.: Studies of interference in serial verbal reactions. J. Exp. Psychol. **18**(6), 643 (1935)
18. Wipfli, R., Dubois-Ferrière, V., Budry, S., Hoffmeyer, P., Lovis, C.: Gesture-controlled image management for operating room: A randomized crossover study to compare interaction using gestures, mouse, and third person relaying. PLOS One **11**(4), 1–10 (04 2016)
19. Zargham, N., Reinschluessel, A.V., Mühlenbrock, A., Muender, T., Cetin, T., Uslar, V.N., Weyhe, D., Malaka, R., Döring, T.: Using gesture and speech to control surgical lighting systems: Mixed methods study. JMIR Hum. Factors **12**, e70628 (May 2025)