

Influence of augmented reality interaction on a primary task for the medical domain

Alina Solovjova

alina.solovjova@ovgu.de
Faculty of Computer Science
Research Campus STIMULATE
Otto-von-Guericke University
Magdeburg
Magdeburg, Germany

Benjamin Hatscher

benjamin.hatscher@ovgu.de
Faculty of Computer Science
Research Campus STIMULATE
Otto-von-Guericke University
Magdeburg
Magdeburg, Germany

Christian Hansen

christian.hansen@ovgu.de
Faculty of Computer Science
Research Campus STIMULATE
Otto-von-Guericke University
Magdeburg
Magdeburg, Germany

ABSTRACT

Augmented reality can deliver valuable information right where the user needs it. When used to display complex data, interaction is required to access all the information. For general use-cases, hand gestures are the go-to technique to interact in augmented reality contexts. However, in some use-cases, the hands are not always available for interaction. For instance, in the operating room, it is likely that the surgeon needs to perform a primary task with the hands simultaneously. This work investigates the influence of interaction tasks using head & hand, head & foot and head & speech on a primary task and the suitability of these input modalities for two kinds of interaction tasks. Results show that interaction techniques used in a multitasking environment should always be evaluated together with a primary task as it might cause user preferences to shift.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI); Interaction techniques; Pointing; Gestural input.**

KEYWORDS

multitasking, foot interaction, speech interaction, gesture interaction, operating room, touchless interaction

1 INTRODUCTION

The usage of head-mounted displays (HMDs) has the potential to revolutionize the workflow in various application

fields. In the medical field, for instance, HMDs can lessen the cognitive and visual workload of surgeons while improving their hand-eye coordination [16, 18]. This has the potential to reduce malpractices, to improve patient safety, operation efficiency, and to reduce costs. Similar benefits apply to the maintenance and assembly field, where instructions can be displayed inside the working space [17]. With the introduction of the Microsoft HoloLens in 2016, interactive AR technology is made more available. The build-in hand gesture and speech recognition is a standard input method to interact inside the HoloLens [1]. These interaction techniques make sense when manipulating virtual content is the only task to be performed. However, there are AR applications, which are used to support the user during the performance of a manual task. This raises questions on the efficiency of interaction techniques and if a physician-computer interaction conflicts with the primary task of performing, e.g., a medical procedure, an assembly or maintenance task. To investigate this problem, we conducted a study with an abstracted primary task during which participants had to perform two different secondary interaction tasks with varying interaction techniques (head & hand, head & foot, head & speech). The abstracted primary task is based on a medical manual task. Our results show that it is crucial to evaluate interaction techniques together with a primary task whenever it is used in a multitasking environment and that performing a primary task might change user preference in favor of input methods other than hand gestures.

2 RELATED WORK

Minakata et al. conducted a comparison of pointing by foot, head and gaze in a head-mounted display and showed that pointing by head has a higher throughput than by foot and gaze [14]. They found that although pointing by eye gaze can be faster, head pointing allows for greater targeting accuracy [11]. A wrist-mounted IMU for hand gesture and a capacitive floor sensor were used for touchless interaction with medical images by Jalaliniya et al. [9]. Mentis et al. investigated the uses of voice control versus gestural control in the operating

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

MuC'19 Workshops, Hamburg, Deutschland

© Proceedings of the 6th Mensch und Computer 2019 Workshop on Mensch-Maschine-Interaktion in sicherheitskritischen Systemen. Copyright held by the owner/author(s).

<https://doi.org/10.18420/muc2019-ws-133-07>

room. They found that speech interaction was used easily to initialize the medical system and mode selection. However, once continuous interactions started, the surgeon switched to hand gesture interaction to allow discussions with colleagues [13]. A combination of head pointing and speech interaction was used by Grinshpoon et al. to manipulate 3D AR data on a HoloLens. An AR model is rotated and scaled through small head rotations of a user. With the aid of voice commands different functions can be applied [6].

Depending on the environment and the context, the user has to simultaneously deal with different stressors while interacting. Taking the medical field into account, an accumulation of different stressors during intervention [19] can lead to a high-stress level, which affects cognitive processes and technical skills [2, 10]. The impact of stress-inducing conditions on the performance of a laparoscopic task has been investigated by Moorthy et al. During the study participants had to perform a laparoscopic transfer task under four conditions consisting of: operating with theatre background noise, verbally solving a simple mathematical task, performing as quickly as possible and all three stressors combined [15]. The results show that there is a significantly higher number of errors when all three stressors are combined. Pairing the transfer task with a mathematical task leads to a particularly high number of errors compared to the other two single conditions.

3 INTERACTION TASKS

Once an AR environment is composed of more than an overlay in the form of a simple annotation, the user needs to interact with the presented data. Basic interactions to control such virtual data are pointing, selection and manipulation of a specific value. Head-mounted displays can leverage the users gaze for pointing [6, 8], which leaves confirmation and manipulation as tasks to be performed by other means.

In medical software objects, modes or settings have to be selected inside patient data. Patient images have to be continuously manipulated e.g. to scroll image layers, zoom to specific structures or to change the image contrast. These functionalities should also be supported during an intervention. The mentioned interaction tasks exist not only inside medical software. Similar tasks are also present in other AR applications as soon as an interactive environment is involved [3–5]. To simulate such scenarios, abstract tasks representing the primary, manual task as well as the two types of interaction tasks *confirmation* and *continuous interaction* have to be designed.

Primary Task

The primary task replicates a medical manual task, as in a needle intervention or ultrasound examination, where the surgeon has to hold an instrument and change its position.

During such tasks, the location of the needle or the ultrasound scanner has to be corrected whenever the patient's body moves due to respiration. The abstracted task consists of a slider with a back and forth moving green area (see Fig. 1). The user has to hold the marker inside the green area using an Xbox One controller. This design was chosen so that the participants could effortlessly perform the task without having any prior knowledge, but still have to pay attention to the task.

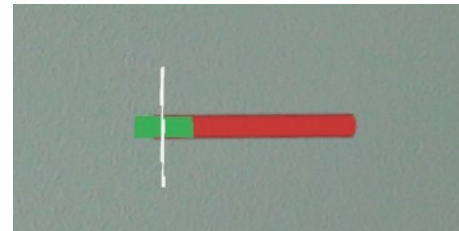


Figure 1: Primary task. Red slider with a moving green area and a white marker which is operated by the user.

Confirmation Task

The first secondary task (further referred to as *confirmation* task) is a selection task modeled after EN ISO 9241-420:2011 where the user has to select highlighted spheres. The cursor is moved via head movement, which is tracked by the HoloLens. When the highlighted sphere is focused on by head movement, the color changes, and the selection can be confirmed via hand gesture, foot gesture, or voice command. Selection via hand is done by the HoloLens', "air tap" gesture. For the selection, via foot, a toe tap gesture is used. To select via voice command, the user can choose one of four voice commands: "choose item", "okay", "pick out" or "select".

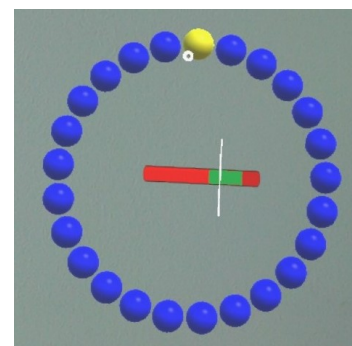


Figure 2: Confirmation task together with the manual task. The sphere in yellow has to be targeted with head movement and then selected via hand, foot or voice.

Continuous Interaction Task

The second secondary task (further referred to as *continuous interaction* task) targets manipulation of a continuous value. This is realized in an image gallery which can be scrolled back and forth. The gallery is positioned in a horizontal, circular manner and the scrolling is animated fluently. Users have to scroll a predefined number of times in a particular direction displayed under the corresponding arrow. For scrolling, via hand, the HoloLens "drag" gesture is used. A heel rotation gesture is used to scroll the images via foot. In order to scroll the images via voice, two different kinds of commands are used. The first group, consisting of "begin", "move", "slide", or "start moving", activates the scroll mode. The user has to rotate his head in the direction in which the scrolling should take place. The second group of voice commands, consisting of "break", "exit" or "stop action", deactivates the scroll mode.

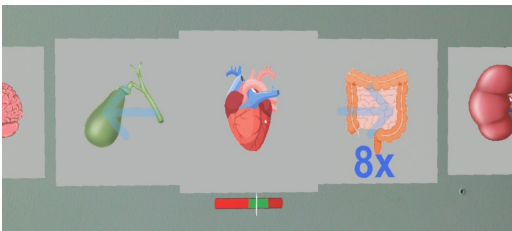


Figure 3: Continuous interaction task together with the manual task. The current task is shown as the number of images the participant has to scroll in the corresponding direction.

4 EVALUATION

We performed a user study to evaluate the influence of AR interaction on a primary, manual task and to assess the suitability of the multimodal interaction techniques head & hand gesture, head & foot gesture and head & voice commands in an AR multitasking scenario. The study consisted of two blocks corresponding to the *confirmation* task and *continuous interaction* task. Both blocks used the same technical setup and measures. Due to technical issues, two participants had to be excluded from the *continuous interaction* results. Two additional participants were recruited for this block, resulting in 12 participants each. Therefore, the group demographics differ slightly between the study blocks. However, this does not influence the validity of our findings as no comparison between both blocks is performed.

Participants

Twelve right-footed students (2 females) between 20 and 29 years ($M=25$, $SD=3.28$) took part in the *confirmation* block with majors in computer science (4), water management (3), mechanical engineering (2), computational visualistics (2) and environmental and energy process technology (1). 50%

of them stated previous experience with Augmented or Virtual Reality, 25% reported using the HoloLens before. Twelve right-footed students (2 females) between 20 and 29 ($M=24.83$, $SD= 3.21$) took part in the *continuous interaction* block with majors in computer science (4), water management (3), mechanical engineering (2), computational visualistics (1), biomedical engineering (1) and environmental and energy process technology (1). In terms of previous experience, 42% used Augmented and/or Virtual Reality Systems before, 17% stated experience with the HoloLens.

Measures

To assess subjective workload, the NASA-TLX questionnaire was used. This tool has been adopted for a wide range of activities and task types such as data entry, visual and auditory monitoring and decision making [7]. The subjective physical strain was assessed using a questionnaire based on [12]. Subjective ratings were gathered by asking the participants to rank the interaction approaches. As an indicator for performance, task completion times and the number of overshoots were recorded.

Apparatus

For augmented reality visualization, hand gesture recognition and speech recognition, the HoloLens Development Edition (Microsoft, Redmond, Washington, USA) was employed. Data on foot movement was gathered using a MTw Awinda Wireless Motion Tracker (xsens, Enschede, Netherlands). Manual input for the primary task was performed using the right joystick of an Xbox wireless controller (Microsoft, Redmond, Washington, USA).

Procedure

The user study took part in a computer laboratory. It consisted of two independent parts corresponding to the *confirmation* task and *continuous interaction* task. After welcoming the participant, a general introduction was given, a demographic questionnaire was filled out, and the system was calibrated. Next, the participant executed the manual task for 60 seconds to gather baseline data without the influence of an interaction task. First, the *confirmation* block was introduced as well as the first of the three input modalities *hand*, *foot* and *speech* selected. Before the recorded runs with an input modality, the participants absolved a not recorded training run. The training run ended when the participant felt comfortable with the interaction technique, which took no longer than 4 minutes. The order of the tested modality and the condition (with and without manual task) was counterbalanced to minimize learning effects, which might be caused by the limited experience of the participants with the system. A NASA-TLX questionnaire was filled out after

each run and the subjective physical strain was assessed after each finished modality. After the completion of the first part, the participant ranked the interaction approaches. The second part (*continuous interaction* block) was carried out analogously.

5 RESULTS

The following results are evaluated descriptively with no statistical analysis has been conducted due to the small sample size of 12 people.

For the *confirmation task*, hand gesture interaction without the additional manual task had the fastest task completion time ($M = 83.58$ s, $SD = 31.40$) compared to foot gesture ($M = 92.83$ s, $SD = 21.37$) and voice commands ($M = 95.58$ s, $SD = 24.30$). On average hand gesture interaction had the fewest overshoots ($M = 2.58$, $SD = 2.07$), followed by speech interaction ($M = 2.94$, $SD = 1.32$) and foot interaction ($M = 3.98$, $SD = 1.46$). In the runs with the additional manual task, all per-

Table 1: Task completion times for the *confirmation* block

	without primary task		primary task present	
	Mean	SD	Mean	SD
foot	92.83 s	21.37 s	117.25 s	26.72 s
hand	83.58 s	31.40 s	123.08 s	41.08 s
voice	95.58 s	24.30 s	132.42 s	38.83 s

formances deteriorated regarding task completion time and overshoots (see Tab. 1). The number of overshoots in these runs are close to each other (see Fig. 4). The best manual

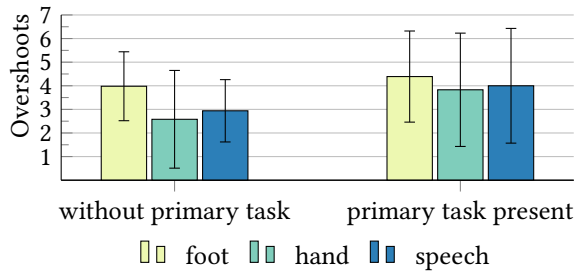


Figure 4: Number of overshoots (i.e. leaving the target sphere with the cursor) during the *confirmation* task.

task performance was achieved while interacting via speech, followed by foot and hand (see Fig. 5). The results of the physical strain questionnaire show that 6 out of 12 participants had a strained shoulder during the interaction with the hand gesture. It was observed that some of them quickly paused the interaction to shake out their arm. 4 out of 12 participants had a strained shin and calf while performing the toe tap. For the selection without additional manual interaction,

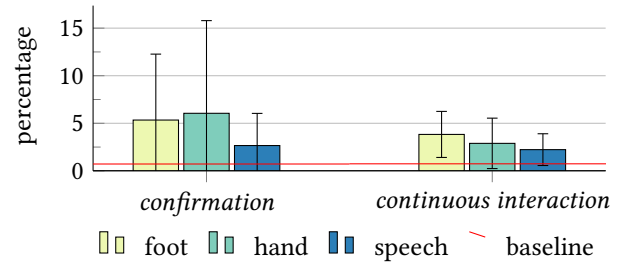


Figure 5: Influence of interaction tasks on primary task performance. The plot shows the mean percentage of the task completion times spent outside the green target area (see Fig. 1). The baseline (primary task without secondary interaction tasks performed for 60 s) is depicted in red.

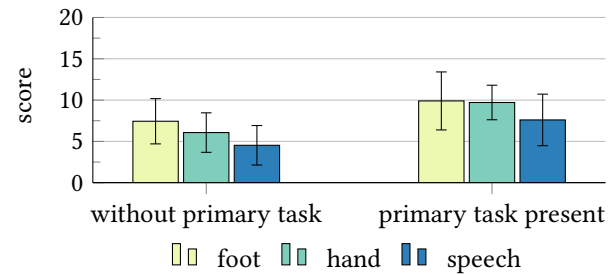


Figure 6: NASA-TLX score for the *confirmation* task. (0 = low perceived workload, 20 = high perceived workload)

42% preferred interaction via voice, 33% via foot and 25% via hand. In the run with manual interaction, the preference shifted to 67% interaction via voice and 33% via foot. These preferences are supported by the overall NASA-TLX score (see Fig. 6) for both runs.

Results of the *continuous interaction* task show that hand gesture interaction without additional manual task had the fastest task completion time ($M = 68.75$ s, $SD = 10.94$) again. Speech interaction was the slowest ($M = 106.50$ s, $SD = 32.62$) while foot interaction is in between ($M = 90.17$ s, $SD = 20.16$) (see Fig. 2). The same result is seen in the average overshoots

Table 2: Task completion times for the *continuous interaction* block

	without primary task		primary task present	
	Mean	SD	Mean	SD
foot	90.17 s	20.16 s	98.58 s	17.86 s
hand	68.75 s	10.94 s	75.92 s	15.99 s
voice	106.50 s	31.62 s	104.83 s	8.41 s

(see Fig. 7). For that, the number of overshoots was split up into left and right direction. Without an additional manual

task, scrolling via speech showed the highest overall number of overshoots while scrolling via hand had the fewest overshoots, closely followed by foot. The manual task per-

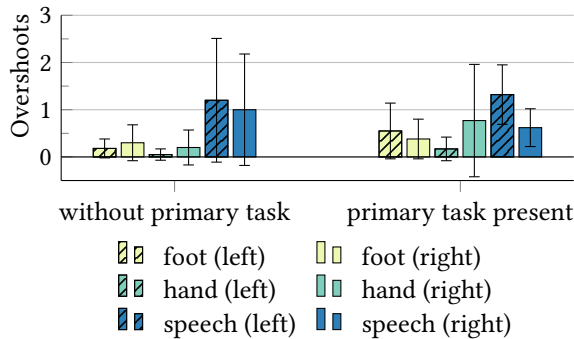


Figure 7: Number of overshoots (i.e. a step in the wrong direction) during the *continuous interaction* task, separated by scrolling direction.

formance is, as in the *confirmations task*, still the best while interacting via voice (see Fig. 5). In this task, however, the gaps between speech, hand and foot interaction were smaller. In the runs with the additional manual task the performances with the interaction via hand and foot deteriorated in task completion time. The performance with speech interaction improved (see Tab. 2). The number of all overshoots increased except for the overshoots for scrolling in the right direction via voice (see Fig. 7). Regarding the physical strain, 6 out of 12 participants still reported a strained shoulder. Only two reported a strained shin and calf and three out of 12 reported a strained neck. Despite the overall higher NASA-TLX score for hand interaction in both runs, scrolling via hand without a manual task was preferred by 67% of the participants (see Fig. 8). Scrolling via foot was preferred from 33% of the participants and had the lowest score in both runs. None of the participants preferred speech interaction, despite the score being in between. The preference shifted as soon as the manual task was present. Here, 50% of the participants preferred scrolling via foot. The other 50% were evenly split between scrolling via hand and voice.

6 DISCUSSION

In this work, three touchless multimodal interaction techniques were compared in a multitasking scenario. The aim was to investigate the influence of different interaction techniques on a simultaneously performed primary task and vice versa. Our evaluation revealed that although interaction via hand for confirmation tasks has the best quantitative results with and without an additional manual task, more participants preferred interacting via voice. The results of the strain questionnaire can explain this. Half of the participants reported a strained shoulder, while speech interaction includes

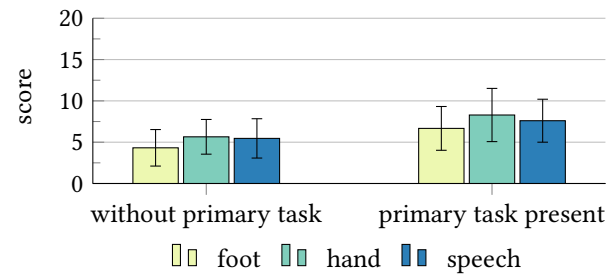


Figure 8: NASA-TLX score for the *continuous interaction* task. (0 = low perceived workload, 20 = high perceived workload)

barely any physical movements. Moreover, simultaneously interacting via voice and performing a manual task demands less mental capabilities than interacting via hand, when the manual task is performed with hands. Observations suggest that the higher number of overshoots for foot interaction resulted from the vibrations of the AR headset caused by performing the toe tap. These vibrations caused the head-controlled cursor to move out of the target and disrupt the selection. Similar head vibrations appeared during speech interaction. The reported strained shin and calf could be due to lack of experience. Many participants performed the toe tap vigorously, even though a soft movement was sufficient enough. This led to stronger vibration and hence to more overshoots and a need to perform more toe taps.

Evaluation of the *continuous interaction* task paints a different picture: Interaction via hand performed fastest and had the highest user preference without a primary task. User preference shifts towards foot interaction when a manual task is present even though it causes the highest workload and deviates most from the baseline in the primary task. Compared to the rankings of the *confirmation task*, none of the participants prefer speech interaction when a manual task is not present. This was caused by a higher level of frustration due to the delayed processing of voice commands, which led to more overshoots. However, when it comes to the primary task, voice commands show the least influence. The preference shifted to some participants preferring interacting via voice commands, despite the problems given above. The next steps include a follow-up study with a bigger sample size and a primary task with different difficulties, which is investigated inside a real setting with surgeons. Furthermore, investigating various manual tasks in different application fields would be worthwhile.

7 CONCLUSION

In general, simultaneously performing a manual task while interacting in AR has a negative effect on the performance of the manual task as well as the speed and accuracy of the

interaction tasks. This work contributes by revealing how performance and subjective ratings might shift when a primary task is added and highlights that there might be no perfect solution for human-computer interaction in multi-tasking scenarios. Giving the choice to the user might lead to the use of more error-prone input methods to avoid interference with the manual task over the fastest or the most accurate input method. Therefore, interaction methods for secondary tasks have to be chosen carefully depending on the priorities of the targeted use case and have to be evaluated with the primary task in mind.

ACKNOWLEDGMENTS

The authors would like to thank Maria Luz for her assistance with the study design and study preparation. This work has been funded by the EU and the federal state of Saxony-Anhalt, Germany under grant number ZS/2017/08/87591 and ZS/2016/04/78123 as part of the initiative "Sachsen-Anhalt WISSENSCHAFT Schwerpunkte"

REFERENCES

- [1] 2018. Gestures - Mixed Reality. <https://docs.microsoft.com/en-us/windows/mixed-reality/gestures>. [Online; Zugriff 03.Juli.2019].
- [2] Sonal Arora, Nick Sevdalis, Debra Nestel, Tanya Tierney, Maria Woloshynowych, and Roger Kneebone. 2009. Managing intraoperative stress: what do surgeons want from a crisis training program? *The American Journal of Surgery* 197, 4 (2009), 537–543.
- [3] Steve Beitzel, Josiah Dykstra, Paul Toliver, and Jason Youzwak. 2017. Exploring 3D Cybersecurity Visualization with the Microsoft HoloLens. In *Advances in Human Factors in Cybersecurity: Proceedings of the AHFE 2017 International Conference on Human Factors in Cybersecurity, July 17- 21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA*, Vol. 593. Springer, 197.
- [4] Hitesh Dhiman and Carsten Röcker. 2019. Worker Assistance in Smart Production Environments Using Pervasive Technologies. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 95–100.
- [5] Gabriel Evans, Jack Miller, Mariangely Iglesias Pena, Anastacia MacAlister, and Eliot Winer. 2017. Evaluating the Microsoft HoloLens through an augmented reality assembly application. In *Degraded Environments: Sensing, Processing, and Display 2017*, Vol. 10197. International Society for Optics and Photonics, 101970V.
- [6] Alon Grinshpoon, Shirin Sadri, Gabrielle J Loeb, Carmine Elvezio, and Steven K Feiner. 2018. Hands-free interaction for augmented reality in vascular interventions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 751–752.
- [7] Sandra G. Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. 904–908.
- [8] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. [n.d.]. MAGIC pointing for eyewear computers. In *the 2015 ACM International Symposium*, Kenji Mase, Marc Langheinrich, and Daniel Gatica-Perez (Eds.). 155–158. <https://doi.org/10.1145/2802083.2802094>
- [9] Shahram Jalaliniya, Jeremiah Smith, Miguel Sousa, Lars Bütthe, and Thomas Pederson. 2013. Touch-less interaction with medical images using hand & foot gestures. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. 1265–1274.
- [10] Clemens Kirschbaum, Oliver T Wolf, Mark May, W Wippich, and Dirk H Hellhammer. 1996. Stress-and treatment-induced elevations of cortisol levels associated with impaired declarative memory in healthy adults. *Life sciences* 58, 17 (1996), 1475–1483.
- [11] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A Lee, and Mark Billingham. 2018. Pinpointing: Precise Head-and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 81.
- [12] N J Mansfield and J M Marshall. 2001. Symptoms of musculoskeletal disorders in stage rally drivers and co-drivers. *British Journal of Sports Medicine* 35, 5 (2001), 314–320.
- [13] Helena M. Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. [n.d.]. Voice or Gesture in the Operating Room. In *the 33rd Annual ACM Conference Extended Abstracts*, Bo Begole, Jinwoo Kim, Kori Inkpen, and Woontack Woo (Eds.). 773–780.
- [14] Katsumi Minakata, John Paulin Hansen, Per Bækgaard, I. Scott MacKenzie, and Vijay Rajanna. 2019. Pointing by Gaze, Head, and Foot in a Head-Mounted Display. In *Proceeding of ETRA 2019*.
- [15] K Moorthy, Y Munz, A Dosis, S Bann, and A Darzi. 2003. The effect of stress-inducing conditions on the performance of a laparoscopic task. *Surgical Endoscopy and Other Interventional Techniques* 17, 9 (2003), 1481–1484.
- [16] IM Sauer, M Queisner, P Tang, S Moosburner, O Hoepfner, R Horner, R Lohmann, and J Pratschke. 2017. Mixed Reality in Visceral Surgery: Development of a Suitable Workflow and Evaluation of Intraoperative Use-cases. *Annals of surgery* 266, 5 (2017), 706–712.
- [17] Xiangyu Wang, Soh K Ong, and Andrew YC Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (2016), 1–22.
- [18] Breedveld Wentink. 2001. Eye-hand coordination in laparoscopy-an overview of experiments and supporting aids. *Minimally Invasive Therapy & Allied Technologies* 10, 3 (2001), 155–162.
- [19] Cordula M Wetzel, Roger L Kneebone, Maria Woloshynowych, Debra Nestel, Krishna Moorthy, Jane Kidd, and Ara Darzi. 2006. The effects of stress on surgical performance. *The American Journal of Surgery* 191, 1 (2006), 5–10.