Danny Schott*, Florian Heinrich, Dominic Labsch, Bennet Hensen, and Christian Hansen

Towards multimodal interaction for needle-based procedures in a virtual radiology suite

https://doi.org/10.1515/cdbme-2022-0018

Abstract: Touchless interaction is popular in the medical domain because it maintains sterility and ensures physicians' autonomy. Evaluating these technologies, however, proves difficult due to technical and human hurdles. Virtual reality leaves these limitations behind and allows for the exploration of promising concepts by simulating an environment and the interactions that takes place within it. We present a virtual radiology suite in the context of needle-based MR-interventions to evaluate touchless interactions. Hand and foot inputs were implemented on a custom interface and evaluated in a user study (n= 16). Results show that activating the system and manipulating values was faster with foot input. However, multimodal interaction is preferable because it is less demanding.

Keywords: Mulitmodal Interaction, Virtual Reality, Gestural Interface, MRI, Needle-based Intervention

1 Introduction

Minimally invasive procedures allow for a variety of patient-friendly treatments. However, the direct view of instruments, risk structures and target tissue in the patient's body is impeded. Physicians must therefore rely on radiological imaging, e.g., magnetic resonance imaging (MRI). Current systems offer many ways of intraoperative interaction on monitors to manipulate generated images. During MRI-assisted percutaneous interventions, radiologists have to shift and rotate the image plane several times, start or stop sequences and change contrast and brightness values. This is often done by assistants receiving verbal control commands [4]. However, this delegation can be frustrating [6], is error-prone and may take a long time [7]. Touchless inputs simultaneously grant sterility and can assist the surgeon in operating the system autonomously. Thus, touchless interaction by foot [8] or hand [1] offers great potential in the medical domain. The combination of different input modalities are also promising and are becoming increasingly popular in research[3]. Usability research in MR-guided interventions is restricted because of limited access during real procedures caused by ethical concerns and hurdles in evaluating technical devices due to electromagnetic radiation. It is especially important in the early development phases of prototypes to include the real workflow and to draw conclusions about ergonomics and usability. A combination of physical mock-up and virtual reality (VR) hardware to simulate MR interventions in virtual radiology suites can close this gap [5]. In VR, early prototypes for touchless interaction can be tested and conclusions can be drawn that can thereafter be applied in reality [9, 10]. In this work, hand and foot input for needle-based MR interventions were investigated and evaluated in a simulation of virtual treatment on a physical MR mock-up. Basic multimodal operating concepts were compared and a custom graphical user interface was developed.

2 Material and Methods

2.1 Apparatus

The study was conducted in a laboratory for medical usability evaluations. To increase immersion, tangible furnishings of the laboratory, e.g., a physical MRI model, a table and a patient mannequin, were integrated into the corresponding virtual environment (see Fig. 1). It was then overlaid with virtual objects in VR and supplemented with various furniture, medical instruments and devices to increase the liveliness (see Fig. 2). The virtual and actual dimensions of the rooms were $4 \times 6m$. We used *Valve Index* VR glasses with a *Leap Motion* controller for hand tracking. Foot tracking was done with *HTC Vive* trackers attached to the top of surgical shoes. To simulate holding a needle, the left VR controller was used.

2.2 Graphical user interface

A custom graphical user interface (GUI) for control of the MRI was placed on a virtual monitor corresponding to real interventional positioning (see Fig. 2). An indirect interface concept was created to help users concentrate on the image data and not on their hands or feet. Five individual buttons were aligned on a horizontal axis. The current sequence, the current quality and contrast value, the target contrast value and DI-

***Corresponding author: Danny Schott,** Otto-von-Guericke University Magdeburg, Universitätsplatz 2, Magdeburg, Germany, e-mail: danny.schott@ovgu.de
Florian Heinrich, Dominic Labsch, Christian Hansen, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
Bennet Hensen, Hanover Medical School , Hannover, Germany

Fig. 1: User equipped with VR hardware standing next to physical mock-up of a MRI while performing hand gestures.

Fig. 2: Virtual operating room with custom MRI-User interface.

COM image data were visible at all times. Values were shown in orange if they deviated from the target values, otherwise they were shown in turquoise. Target and actual values were abstracted to focus more on the evaluation of interaction methods. . The GUI is divided into a main menu, a sequence menu and a contrast menu, which also consists of five menu-specific buttons. Each button has three visual states: pressed, selected and inactive. Slightly translated DICOM data represented real time image data of a breathing patient.

2.3 Interaction techniques

Hand and foot input for standing users was provided. In addition to a separate investigation of these modalities, combinations were also compared, allowing for multimodal interaction with the system. Unintentional inputs using touchless interaction can cause delays and thus lead to user frustration [2]. To prevent unintentional manipulation of parameters, two different activation gestures were implemented. They were designed to avoid unintended execution. Thus, a "gunshot gesture" was chosen as a hand gesture metaphor, in that a quick pull ($0.5s$) of an imaginary trigger with the index finger activates the system (see Fig. 3a). As seen in Figure 3d, activation by feet is done by rapidly pulling the insides of the feet together and apart ($0.5s$). Only when the controller was near the patient were activation gestures accepted, thus simulating an occupied hand holding the needle (see Fig. 1). A warning was shown in case of non-contact. Deactivating the system was performed in a manner analogous to activation.

After recognizing activation gestures, five static, transparent and equally spaced bounding box elements are generated to represent the five GUI buttons. For hand selection, they are placed on a $30cm$ long baseline at thumb level, which is parallel to the ground and to the front vector of the VR headset. The center element is always set to thumb tip at activation. For foot selection, the elements are created around the *Vive tracker* in a similar baseline. In addition, another element used for selection with the same position as the middle element is generated. The additional element always moves in the same distance and orientation to the tracker of the dominant foot.

For hand control, bounding boxes are at thumb level. The tracked position of the index finger is set relative to the distance of the fingertip and below the corresponding baseline. As long as the fingertip is below it, horizontal movement on this axis on the GUI will select the menu button corresponding to the element closest to the fingertip (see Fig. 3b). To press a selected button, the index finger must be moved straight up over the baseline and then moved back below it (see Fig. 3c). For foot control, menu buttons are accessed by rotating the dominant heel. A heel rotation changes the position of the selection element synchronized with the foot position (see Fig. 3e). The menu button closest to the selection sphere is selected in the GUI. To confirm a selected menu button, the toe of the foot must be raised $8cm$ (see Fig. 3f).

2.4 Evaluation

A within-subject design study was conducted to compare the input modalities. The task was to manipulate the system's ac-

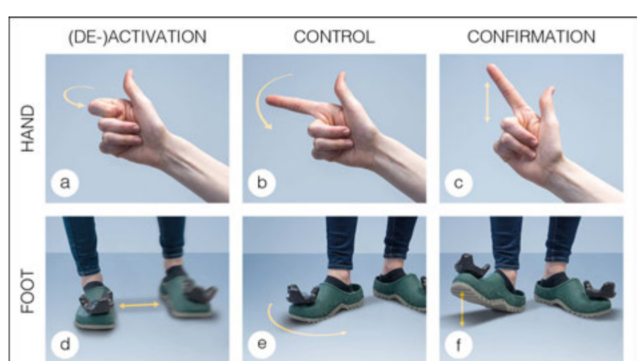

Fig. 3: Input gestures for activation and manipulation of the GUI.

Tab. 1: Summary of the ANOVAs' results on *task completion time* and *subjective workload* ($\alpha < .05$). Degrees of freedom in the nominator (DFn) and in the denominator (DFd), test statistics F-value and p-value, significance, and effect size (η^2) are reported.

Variable / Effect type	Factor	DFn	DFd	F	p	Sig.	η^2	Effect	Figure
Task Completion Time									
Main effects	***Activation***	**1**	**15**	**6.750**	**0.020**	*	**0.062**	**Medium**	**Fig. 4a**
	Selection	**1**	**15**	**6.412**	**0.023**	*	**0.112**	**Medium**	**Fig. 4b**
Interaction effect	***Activation * Selection***	**1**	**15**	**2.384**	**0.143**		**0.026**	**-**	**Fig. 4c**
Subjective Workload									
Main effects	***Activation***	**1**	**15**	**0.048**	**0.830**		**0.000**	**-**	**Fig. 4d**
	Selection	**1**	**15**	**0.761**	**0.397**		**0.026**	**-**	**Fig. 4e**
Interaction effect	***Activation * Selection***	**1**	**15**	**14.540**	**0.002**	*	**0.044**	**Small**	**Fig. 4f**

tual values to its target values. After collecting demographic information and informed consent, participants were given instructions. This was followed by an introduction of the VR environment, an explanation of the GUI and then testing of all input modalities. If the participants felt confident, the combinations of the respective input modalities were randomly assigned, after which three training runs and three trials were measured. One trial consisted of the participant holding the index controller in the left hand and moving it to the patient's liver, activating the system, opening the sequence menu, selecting the correct sequence, returning to the main menu, pausing the system, opening the contrast menu, setting the correct contrast value and finally deactivating the system. The manipulation of the target and actual values was done in a constant interval of six interaction steps (e.g., taps). The given values in the GUI were randomized. We used a two-factor design, where the independent variables are activation (1) and selection (2). Two dependent variables were considered: Task Completion Time (TCT) and subjective workload. TCT indicates the time between activation and deactivation of the system. Subjective workload was assessed per combination using the NASA-TLX questionnaire after each of the three measured trials.

3 Results and discussion

We recruited 16 subjects (9 female, 7 male) aged between 22 and 31 years (median: 26.57) for the study. Eight participants were medical students and the other half had a technical or computer science background. No effects with respect to participant background were observed. Repeated measures two-way analyses of variance (ANOVAs) were conducted on *task completion time* and *subjective workload*. These results are summarized in Tab. 1. Observed main and interaction effects are illustrated in Fig. 4. Statistically significant main effects were identified for the *task completion time* variable. Both *activation* and *selection* were faster using the foot concepts. No interaction was observed between the independent variables, thus the main effects are considered valid. Regarding *subjective workload*, no statistically significant main effects were detected. However, we found a significant interaction effect on this variable. When using foot activation, hand selection was advantageous compared to foot selection. In contrast, when using hand activation, hand selection resulted in higher workload compared to foot selection. Hand interaction is performed more slowly and induces higher workload. This may be due to the fact that the interaction area was freely selectable and not visible, which meant that the users did not have an exact idea of where the area was located in relation to the index finger. Furthermore, the hand tracking area is restricted by the *Leap Motion Controller*, which limits the performance of gestures by the users. Foot interaction seems to be a faster input modality for the selected tasks. The interaction area is spatially limited by the floor level and allows more precise manipulation and foot placement, which in turn has a positive effect on physical effort. With foot input, we see problems with the necessary weight transfer in standing positions, which can lead to unergonomic postures. Constant movement with the arms can in turn lead to strain, which is why pure activation is better suited here. A comparison to seated procedures would be of interest, as foot input would benefit from this [8]. In general, hand gestures often perform better than foot input in related work [9], which is contradicted by our findings and should be investigated further. A comparison of our techniques with current clinical practice in terms of duration and effort would also be interesting.

4 Conclusion

In this work we investigated the use of VR to evaluate touchless interaction of controlling an MR system in the context of percutaneous interventions. The inclusion of a physical MR mock-up, patient dummy and interior was used to increase immersion in a virtual radiology suite and promote usability investigation. Foot inputs performed better overall, but a combi-

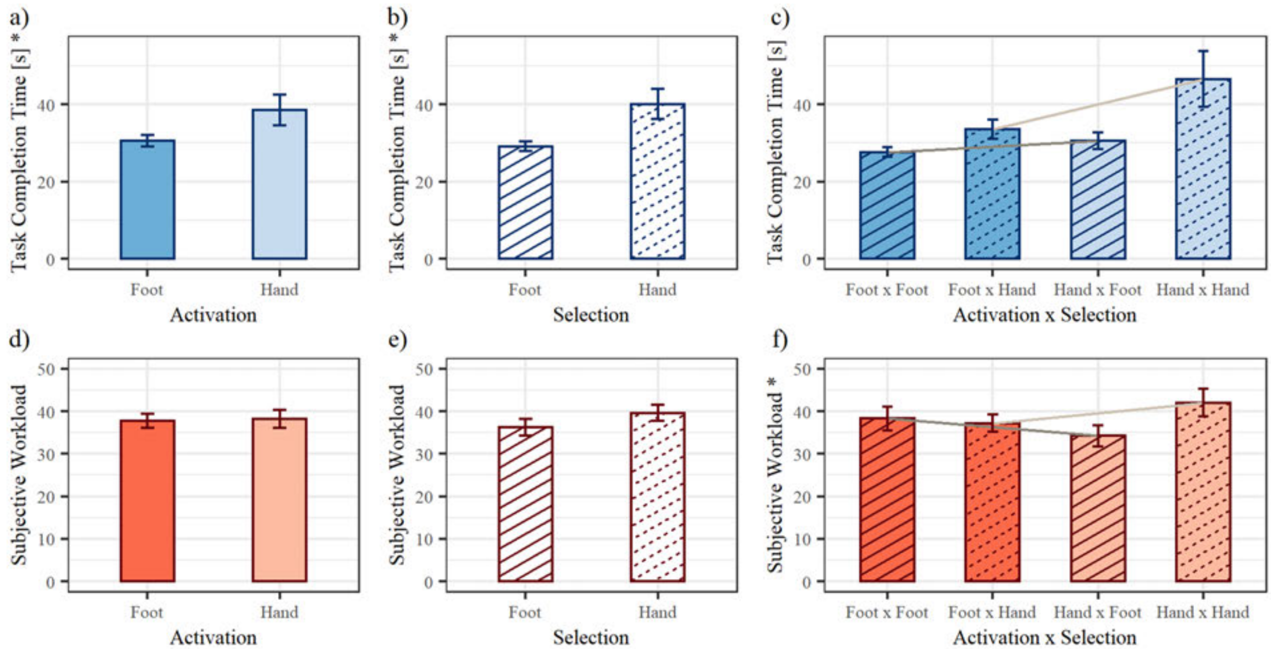

Fig. 4: Main and interaction effects on task completion time (a-c) and subjective workload (d-f). Error bars represent standard errors.

nation of hand activation and foot control seems to be the best compromise in terms of time and effort. Future work could extend the presented approaches with 3D interaction techniques. For example, the possibility of rotation and translation of the image planes could be a challenge to solve.

Author Statement

Research funding: This work was funded by the Federal Ministry of Education and Research within the Forschungscampus *STIMULATE* (grant no. 13GW0473A). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

References

[1] L. Bautista, F. Maradei, and G. Pedraza. Usability test with medical personnel of a hand-gesture control techniques for surgical environment. *Int J Interact Des Manuf*, 14(3):1031–1040, 2020.

[2] S. Cronin, E. Freeman, and G. Doherty. Investigating clutching interactions for touchless medical imaging systems. In *Proc Conf Hum Factor Comput Syst*, pages 1–14, 2022.

[3] B. Hatscher and C. Hansen. Hand, foot or voice: Alternative input modalities for touchless interaction in the medical domain. In *Proc ACM Int Conf Multimodal Interact*, pages 145–153, 2018.

[4] J. Hettig, P. Saalfeld, M. Luz, M. Becker, M. Skalej, and C. Hansen. Comparison of gesture and conventional interaction techniques for interventional neuroradiology. *Int J Comput Assist Radiol Surg*, 12(9):1643–1653, 2017.

[5] R. Kwapik, J. Moritz, B. Hensen, B. Janny, E. Pannicke, D. Schott, et al. Virtual reality-based usability laboratory for interventional mr applications. In *Proc 5th Conf on Image-Guided Interventions*, pages 51–52, 2021.

[6] K. O'Hara, G. Gonzalez, A. Sellen, G. Penney, A. Varnavas, H. Mentis, et al. Touchless interaction in surgery. *Commun ACM*, 57(1):70–77, 2014.

[7] K. O'Hara, G. Gonzalez, G. Penney, A. Sellen, R. Corish, H. Mentis, et al. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. *Comput Support Coop Work*, 23(3):299–337, 2014.

[8] S. F. Paulo, F. Relvas, H. Nicolau, Y. Rekik, V. Machado, J. Botelho, J. J. Mendes, et al. Touchless interaction with medical images based on 3d hand cursors supported by single-foot input: A case study in dentistry. *J Biomed Inform*, 100:103316, 2019.

[9] A. V. Reinschluessel, J. Teuber, M. Herrlich, J. Bissel, M. van Eikeren, J. Ganser, et al. Virtual reality for user-centered design and evaluation of touch-free interaction techniques for navigating medical images in the operating room. In *Ext Abstr Hum Factors Computing Syst*, pages 2001–2009, 2017.

[10] A. Zaman, A. Roy, K. Fatema, N. J. Farin, T. Doring, and R. Malaka. Explore voice and foot-based interaction techniques to navigate 2d radiological images in the virtual reality operation theatre. In *22nd Int Conf Comput Inf Technol*, pages 1–7. IEEE, 2019.