# GazeTap: Towards Hands-Free Interaction in the Operating Room

Benjamin Hatscher
Otto von Guericke University
Magdeburg, Germany
benjamin.hatscher@ovgu.de

Maria Luz
Otto von Guericke University
Magdeburg, Germany
maria.luz@ovgu.de

Lennart E. Nacke
University of Waterloo
Waterloo, ON, Canada
lennart.nacke@acm.org

Norbert Elkmann
Fraunhofer Institute for Factory Operation and Automation
Magdeburg, Germany
norbert.elkmann@iff.fraunhofer.de

Veit Müller
Fraunhofer Institute for Factory Operation and Automation
Magdeburg, Germany
veit.mueller@iff.fraunhofer.de

Christian Hansen
Otto von Guericke University
Magdeburg, Germany
christian.hansen@ovgu.de

## ABSTRACT

During minimally-invasive interventions, physicians need to interact with medical image data, which cannot be done while the hands are occupied. To address this challenge, we propose two interaction techniques which use gaze and foot as input modalities for hands-free interaction. To investigate the feasibility of these techniques, we created a setup consisting of a mobile eye-tracking device, a tactile floor, two laptops, and the large screen of an angiography suite. We conducted a user study to evaluate how to navigate medical images without the need for hand interaction. Both multimodal approaches, as well as a foot-only interaction technique, were compared regarding task completion time and subjective workload. The results revealed comparable performance of all methods. Selection is accomplished faster via gaze than with a foot only approach, but gaze and foot easily interfere when used at the same time. This paper contributes to HCI by providing techniques and evaluation results for combined gaze and foot interaction when standing. Our method may enable more effective computer interactions in the operating room, resulting in a more beneficial use of medical information.

## CCS CONCEPTS

• **Human-centered computing** → **Interaction paradigms**; **Interaction techniques**; *Interaction devices*; *Gestural input*; Pointing;

## KEYWORDS

Input techniques, multimodal interaction, foot input, gaze input, eye tracking, gaze-foot interaction, HCI in the operating room

**ACM Reference Format:**
Benjamin Hatscher, Maria Luz, Lennart E. Nacke, Norbert Elkmann, Veit Müller, and Christian Hansen. 2017. GazeTap: Towards Hands-Free Interaction in the Operating Room. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3136755.3136759

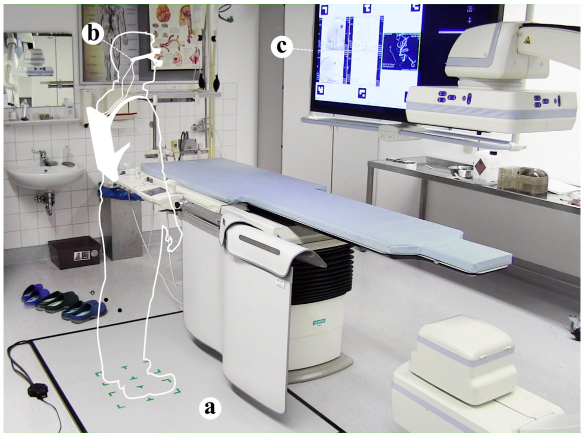


**Figure 1: Technical setup for hands-free interaction consisting of a tactile floor (a), a mobile eye tracking device (b) and the large screen of an angiography suite (c).**

## 1 INTRODUCTION

In medicine, minimally-invasive interventions allow a variety of treatments without the need to open the patient. During such interventions, imaging modalities such as angiography systems enable the physician to navigate medical instruments inside the patient's body. Images from various angles as well as preoperative planning data may be used during the process, which makes intraoperative interaction with medical images important. Thus, a common Human-Computer Interaction (HCI) task in the operating room (OR) is scrolling through image sequences and rotating 3D representations [12].

Controlling a computer in the OR suffers from several restrictions. A sterile environment does not allow for direct manipulation

*ICMI'17, November 13–17, 2017, Glasgow, UK*

ACM ISBN 978-1-4503-5543-8/17/11…$15.00
https://doi.org/10.1145/3136755.3136759

of input modalities such as mouse, joystick, or touchscreen because of the risk of bacterial contamination [7, 32, 36]. This problem is addressed by wrapping interface controls and touchscreens in sterile plastic sheaths. Additionally, physicians have to change their position every time they want to access controls [12]. Adding to the issue of contamination and accessibility, the physician's hands are generally preoccupied with surgical tasks, such as handling medical instruments, which have to be interrupted to manipulate controls [21]. In clinical routine, interaction tasks therefore are often delegated verbally or via gestures to a medical assistant located in the OR or a non-sterile control room nearby. This requires additional personnel, may cause interruption of the workflow and easily leads to misunderstandings [8, 12, 26].

When hands are not available, foot pedals are a conventional input device in medical computing systems. Interacting with feet works well for simple, secondary tasks such as scrolling, but is less suitable for spatial interaction tasks such as pointing [27, 30, 45]. In fact, foot interaction in the OR requires a second modality to cover the whole range of requirements regarding physician-computer interaction. According to Sibert and Jacob, people easily gaze at their surroundings while performing other tasks [37]. The combination of foot with other input modalities may allow rich interaction without occupying the hands. Therefore, we combined gaze and feet as input modalities and investigated how they can be used to control a medical image viewer at an upright stance. We present two interaction techniques based on gaze pointing and short foot steps. Our first approach focuses on continuous gaze interaction; the second one introduces a gaze-and-foot confirmation gesture. To investigate these techniques, we created an input system based on a mobile eye tracker and a tactile flooring. This setup controls a software which resembles the Graphical User Interface (GUI) of an angiography suite. We conducted a study with 13 participants in an angiography suite to determine the performance of our input techniques and a foot-only approach when interacting with medical images. We compared the input setups by measuring task completion time and workload. Via video analysis, we gathered how much time was required to complete individual subtask.

We contribute to Human-Computer Interaction (HCI) research by providing approaches for hands-free gaze and foot interaction as well as a user study in which our techniques are applied to control a medical image viewer. The insights from our study reveal how gaze and foot influence each other in an upright stance, and provide suggestions for researchers and practitioners on how to create hands-free interaction methods that might enable physicians to work more efficiently during minimally-invasive interventions in the future.

## 2 RELATED WORK

Several solutions to the problem of sterile interaction in the operating room such as touchless gesture and voice control have been proposed [22]. However, these methods bear some disadvantages. Sensors for gesture recognition such as the Microsoft Kinect, the Leap Motion Controller or the Myo Armband [4, 11, 25] allow for direct control but may be constrained by holding medical instruments [26]. Voice control does not occupy the hands, but was deemed sensitive to background noise, pronunciation, accent, and choice of commands [1, 4]. Thus, an input method which combines gaze and the feet might be a suitable alternative to overcome the main disadvantages of current approaches by allowing sterile, direct interaction while keeping the hands free at the same time. Therefore, existing approaches from the field of HCI regarding both modalities are described in the following.

### 2.1 Gaze Interaction

Gaze as input method is fast [48] and indicates the user's coarse area of attention [37, 41, 49]. However, when used for selection tasks it is not ideal due to inaccuracies [48] and an uncomfortable sensation for most users when using a dwell-time approach, since looking at the same spot for a long time is unnatural [39]. Additionally, gaze interaction suffers from the so-called *Midas Touch* problem (i.e., triggering actions involuntarily by looking at controls) [14]. Therefore, gaze should better be used in combination with more explicit input methods than as a singular input channel [13].

### 2.2 Gaze in Multimodal Setups

Gaze has been combined with additional input modalities for various reasons. Accuracy issues of eye tracking devices have been overcome by adding head tracking [15], mobile touch devices [40] or hand gestures [3]. For multitouch-tables, utilizing gaze allows selection of distant targets, but some subjects reported looking at their hands on the touchscreen instead of the (gaze) target while performing a task [20].

### 2.3 Foot Interaction in an Upright Stance

Guidelines for foot interaction while standing suggest choosing toe taps over kicks and sizes for angular targets of 90° when not providing cursor feedback, as well as a 20cm radius for tapping interactions [34]. Sets of discrete interaction areas for the feet, placed around the user, proved to be suitable for mail- and photo-sorting applications [23] as well as for controlling various desktop applications [35]. Manipulation of a value via a discrete vocabulary can be realized by continuous interaction as long as a certain condition is met. Applied to foot interaction, this could be holding the foot or leg in defined positions or standing on a specific spot. Holding a posture allows for faster stopping and therefore reduces overshooting of target positions, but kicks were preferred over holding a posture since the user can put the foot at rest between subtasks [2].

### 2.4 Feet in Multimodal Setups

Foot interaction has already been combined with various hand-operated input modalities including multi-touch tables, tangible interfaces, and mice [45]. Interaction with hand and feet via pressure-sensitive surfaces has been investigated [33]. The authors report weight distribution gestures as more comfortable than foot rotation or transition gestures but point out that it might cause balancing issues. For interaction with medical image sequences by using foot and hand gestures, double tap gestures were found more suitable than swipe gestures for foot interaction due to balancing issues [16]. In a virtual-reality OR, paging through medical images via rotation around the heel and toe tapping performed comparably to hand gestures and verbal task delegation using "up" and "down" as commands [30].

### 2.5 Gaze and Foot Setups

As an alternative to the mouse, Engelbart had already suggested combinations of input modalities including eye pointing, knee and foot movement in 1984 [5]. Gaze and feet as additional input channels to the traditional mouse and keyboard have been investigated for zoom-and-pan tasks [10, 18]. Using gaze for indicating the zoom direction and a two-directional pedal for the zoom speed was preferred in contrast to gaze panning, which was reported as distracting and tiring when used for fast or long movements [18]. Implicit gaze input in combination with continuous foot input is stated as promising and the use of alternatives such as multi-touch floors is suggested [10]. An approach where combined gaze and foot interaction was used as a substitute for mouse (i.e. hand) input on a desktop workplace was found comparable to the mouse as long as the dimensions of the interactive elements stay over a certain threshold (0.60" x 0.51") [29].

Our work differs from literature in combining gaze and the feet in an upright stance as input channels for hands-free human-computer interaction. The area of application we had in mind is the operating room, which means the hands have to be sterile and might even be occupied with holding medical instruments. Since this might cause additional cognitive workload in the long term, we focus on intuitive interaction approaches for each modality. Therefore, gaze is used to implicitly provide the coarse area of attention combined with a natural mapping of foot gestures to spatial tasks.

## 3 IDENTIFIED INTERACTION TASKS

In this work, we focus on the task of image interaction, since it is a common one which appears regularly during minimally-invasive interventions [12]. To create interaction techniques that fit the needs, we analyzed the required tasks and describe the outcome in the following. During interventions, planning data, preoperative images and recently acquired medical images are used to support the physician. Most of these data sets consist of time- or spatial-encoded series of images. In operating rooms, multiple displays and/or viewports on large screen displays are used to show this data. Even though size and position provide a good view for the physician, only one image of such a data set can be shown at a time. To access the full range of information, two tasks have to be fulfilled: first, the desired viewport has to be selected, and second, the medical image data set has to be manipulated.

### 3.1 Viewport Selection

For interaction with a data set, the appropriate viewport has to be selected first. Currently, this has to be done by mouse or joystick, which causes sterility issues, or via task delegation which may be affected by misunderstandings. In terms of interactions, this selection phase can be split up into the following subtasks:

- Choose a viewport
- Confirm the viewport

In clinical routine, physicians simply look at the viewport containing the information required and return to the current task at hand. Therefore, we used gaze to choose a viewport in a natural way.

### 3.2 Image Manipulation

Once a viewport is selected, the second part consists of manipulating the content. A variety of functions are available, most of which can be controlled with a single degree of freedom (DOF). A spatial or time encoded series or "stack" of images can be scrolled back and forth, and contrast or brightness can be adjusted. One of the rare occasions where two DOF are required is interaction with 3D data. For situations where 2D image data are not sufficient, a 3D representation of the patient's inner structures can be obtained by a method called 3D digital subtraction angiography. To rotate the resulting dataset, at least two DOF must be available. In this work, we focus on both cases of image interaction tasks, which can be summarized in the following list of subtasks:

- Image stack interaction (1 DOF)
  - scroll up/down
- 3D model interaction (2 DOF)
  - rotate up/down
  - rotate left/right

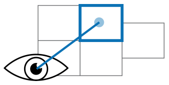
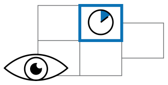
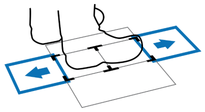
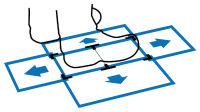
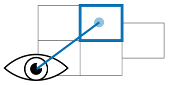
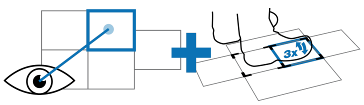
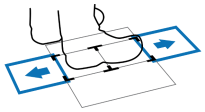
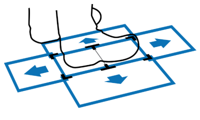
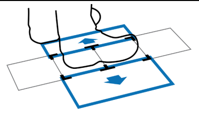
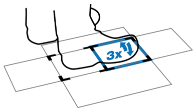
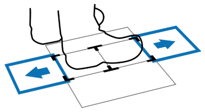
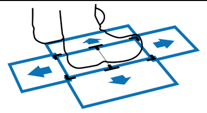
## 4 USER INTERFACE

To provide reference for the following interaction techniques, the interface of our system is explained in the following. In general, our software resembled the GUI of an angiography system (see Figure 2). It was created using MeVisLab [31], a medical prototyping software. The interface consists of five viewports. Four viewports display series of angiography images which can be scrolled back and forth; the rightmost viewport displays a 3D-representation of a blood vessel structure. Visual feedback is given for both interaction tasks. Since we did not want to overlay medical image data with a gaze pointer to prevent misinterpretations or coverage of important details, gaze position is indicated by visual feedback in the form of borders around viewports. An orange border indicates the currently gazed-at viewport, a green border the selected one. During image manipulation, arrows indicate the direction of scrolling or rotation.

## 5 INTERACTION TECHNIQUES

According to our subtask definition, gaze and foot interaction can be applied in a natural way: gaze indicates the viewport on which the user currently focuses. Foot input is suitable for coarse interaction and therefore allows fulfilling discrete input tasks with one or two DOF. Based on this, we propose two techniques called *On-the-Fly Manipulation* and *Dedicated Lock Gesture* to combine gaze and foot input for viewport selection. Foot movements already suggest a spatially consistent mapping to the manipulation of image stacks and 3D representations. Therefore, both approaches apply the same foot interaction technique described in subsection 5.3. Even though image manipulation is realized in the same way, it resembles the working task which needs to be fulfilled with the support of our input techniques and therefore is integrated in each approach for evaluation. An overview of both interaction techniques and a foot-only approach can be found in Table 1.

**Table 1: Overview of the functions provided by the system and evaluated combinations of input setups.**

| Input Setup | Viewport Selection | | Image Manipulation | |
|---|---|---|---|---|
| | choose a viewport | confirm the viewport | Image stack interaction | 3D model interaction |
| On-the-fly Manipulation (OTF) | <br>gaze pointing | <br>gaze dwell 1.5 sec. | <br>foot interaction (up/down) | <br>foot interaction (left/right/up/down) |
| Dedicated Lock Gesture (DLG) | <br>gaze pointing | <br> <br>gaze pointing + triple-tap | <br>foot interaction (up/down) | <br>foot interaction (left/right/up/down) |
| Foot-only Interaction (FI) | <br>foot interaction (left/right) | <br> <br>triple-tap | <br>foot interaction (up/down) | <br>foot interaction (left/right/up/down) |

## 5.1 On-the-Fly Manipulation

The first technique called On-the-fly Manipulation (OTF) continuously tracks the viewport the user is looking at and allows navigation of the currently selected image data via the feet at any time. Viewport selection requires only coarse indication of the user's area of interest, without actual control elements on the screens. Therefore, we assume that accuracy issues and the *Midas Touch* problem influence our system far less than in related research, which makes continuous gaze tracking a reasonable choice. Informal tests showed that the eye-tracker's accuracy only comes into play when gazing near the viewport's borders. We added a visual indicator and a 1.5 second dwell-time when selecting a viewport so that the user can interrupt an imminent, involuntary switch. Since we utilize dwell-time not for a series of successive selection tasks such as gaze-typing but for coarse context identification, we use a much longer dwell-time than that suggested by literature (500 - 900 ms) [19].

## 5.2 Dedicated Lock Gesture

Mauderer et al. investigated distant selection via gaze and flicking touch gestures [20]. Combined gaze and gesture had the lowest rate of correct selections applied to a grid of targets compared to sole gaze or gesture interaction. Additionally, some subjects reported they were looking at their hand on the touchscreen instead of the target. Their findings suggest that gaze might be affected by checking one's own posture when proprioceptic feedback alone is not sufficient. Therefore, the second technique establishes a Dedicated Lock Gesture (DLG) which uses gaze only during the viewport selection phase, similar to Chatterjee et al. [3]. To select the viewport, gaze indicates which viewport to choose while a triple-tap is performed with the ball of the foot to confirm it. After successful performance of the gesture, interaction by feet affects the selected viewport, no matter where the gaze points. In this state, the visual feedback on the currently focused viewport is given to allow switching anytime by performing the triple-tap gesture again. We decided to use a triple-tap gesture since it's easy to distinguish from occasional foot lifts during pose correction.

## 5.3 Natural Mapping of Foot Gestures

Once a viewport is selected by one of the previously described techniques, the images can be manipulated via discrete, sensitive areas on the floor. From the user's initial position, a defined area is set which only responds to triple-taps, since the feet are lifted occasionally to maintain a stable stance [35]. A discrete interaction area extends from each of the edges of this region. These "buttons" in each of the four directions allow a natural mapping of spatial tasks.

Iterating through image stacks by mouse or joystick establishes a mental mapping of the physical directions front and back to these interactions. Consequently, we mapped the buttons in front and behind the user's position to these functions. Additionally, 3D objects can be rotated left and right by stepping on the areas on the side. The elevation angle of the point of view is controlled via the front and back buttons in this case. Simeone et al. proposed a technique which uses both feet for feet-only object rotation around three axis. They applied the method we use for object rotation to camera manipulation instead [38].

To allow for fast, continuous manipulation via a discrete foot input vocabulary, we used an approach with two different rates. Instead of using separate foot positions for each rate such as Reinschluessel et al. [30], we decided to switch rates automatically after a certain amount of time. While standing on a sensitive area, the corresponding interaction (scrolling up or down one image or rotating five degrees) is executed once every second to enable fine-grained selection. After three cycles of continuous pressure on the corresponding area, the action is triggered every 0.2 seconds to speed up navigation to distant targets.

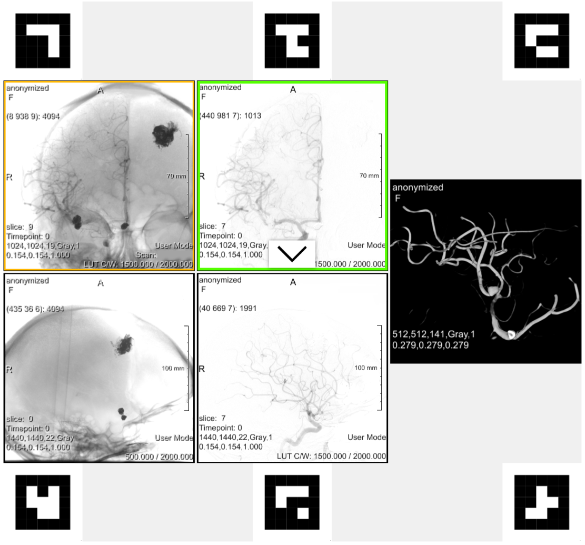


**Figure 2: Prototypical GUI consisting of four viewports showing medical image stacks (left and center) and one 3D-viewport (right). Borders indicate the currently selected viewport (green) and the gazed-at viewport (orange). Image manipulation was indicated by little arrows at the viewport edges.**

## 6 EVALUATION

A user study in a lab setting was conducted to investigate the proposed interaction techniques. We aimed to pinpoint HCI-related problems in our hands-free interaction techniques and did not expect to beat established input methods such as mouse or joystick at this early stage. To reduce uncontrolled factors, we left out aspects which comes into play in real OR situations such as manipulating medical instruments, interpersonal communications, distracting warning signals, handing over control or changing places. An often used method in clinical routine is the verbal delegation of interaction tasks to a medical assistant [12, 26]. Unfortunately, the efficiency depends on the assistant's knowledge, experience and familiarity with a system [42]. Therefore we refrained from using this method as baseline. Instead, we added a Foot-only Interaction (FI) method to gather richer data regarding the role of eye tracking in this setting. In the FI method, a triple-tap gesture switches between a viewport selection and an image manipulation mode. In viewport selection mode, using the left and right sensor areas cycles through all five of the viewports. When switched to image manipulation, the currently chosen viewport is confirmed and the gesture set described in subsection 5.3 can be used.

In informal studies, we compared gaze-only, foot-only and an early versions of gaze and foot interaction. Sensitive areas located at the edges of each viewport were used to manipulate image data by gaze. Caused by the *Midas Touch* problem, gaze-only interaction was outperformed by both, foot-only and gaze and foot interaction and therefore was discarded in the following study. Overall, three setups were evaluated: OTF, DLG and FI (see Table 1).

### 6.1 Participants

Thirteen paid participants (six female, seven male) between 22 and 31 years old (M = 25.5, SD = 3.1) with normal or corrected-to-normal vision took part in the study. Seven of them were students of human medicine; the remaining four participants were recruited from technical courses. The participants were acquired via mailing lists of our university and rewarded with 20 €. Four participants reported medium prior experience with foot interaction and eye tracking as an input method (i.e., rating it 3 on a 5-point Likert scale). All other participants stated no prior experience.

### 6.2 Apparatus

To investigate the proposed interaction techniques, we created a setup consisting of a mobile eye-tracking device, a tactile floor, two laptops, and a 56" display. To account for realistic dimensions, the system was set up in an angiography suite. A low-cost mobile eye tracker from Pupil Labs [17] (accuracy of 0.6° under ideal conditions) was used instead of a stationary one to account for the distance to the screen and movements of the participants. The eye-tracking headset acts as a mount for two cameras: one for tracking the user's pupil, and another one forward-facing to get a first-person view. They are connected to a Thinkpad E320 (4x Intel i5 Cores @ 2.5 GHz, 8 GB RAM) running Ubuntu 15.10, which was stored in a holster to be worn like a backpack (see Figure 1). The corresponding open-source capturing software allows calibration and calculates the gaze position. Fiducial markers on the screen enabled the Pupil Capture markers plugin to calculate on-screen 2D gaze coordinates. Gaze data was sent to a second computer via ad-hoc WiFi using UDP.

A tactile floor from the *Fraunhofer Institute for Factory Operation and Automation IFF* was used, which is currently in operation within the field of human-robot interaction for industrial applications [6]. The sensor system is an array of sensing elements with piezoresistive composite material between two electrodes, embedded in a plastic casing [24]. Additionally, it was covered with linoleum to match the friction of typical OR floorings. The resolution of the floor is 32 x 19 pressure-sensing cells (5 x 5 cm each) at a size of 160 x 95 cm. A microcontroller processes the measured data and provides pressure values at about 50 Hz via USB. A Thinkpad T540p (Intel i5 Core @ 2.6 GHz, 8 GB RAM) running Microsoft Windows 10 received the gaze position, pressure data from the floor and run the main application including a GUI described in section 4. The system was not connected to the angiography system except for the display and therefore did not receive live images. Instead, anonymized datasets were used during the study.

### 6.3 Tasks

Each participant fulfilled a set of six tasks for each setup (within-subjects design) (see Table 2). The tasks were grouped in two blocks of three tasks. The first block required interaction with image stacks only, while the second block consists mainly of 3D interaction tasks. For 3D interaction, rotation around only one axis per task was required. Each task required the participant to change the viewport and manipulate the medical image data. The start position of each task corresponded to the target position of the previous one. In all three setups, image manipulation was done by foot but the method for viewport selection varied.

**Table 2: Overview of the task sequence used in the evaluation.**

| *Task* | *Task Description* |
|---|---|
| Training phase | - |
| Test phase | Select lower left viewport, go to slice 4<br>Select upper center viewport, go to slice 15<br>Select right viewport, rotate three steps right |
| 1.1 | Select lower center viewport, go to slice 16 |
| 1.2 | Select upper left viewport, go to slice 7 |
| 1.3 | Select lower center viewport, go to slice 3 |
| 2.1 | Select right viewport, rotate eight steps left |
| 2.2 | Select upper center viewport, go to slice 7 |
| 2.3 | Select right viewport, rotate six steps down |

## 6.4 Measures

Task completion times (TCT) were gathered as performance measure. After a verbal query if the participant is ready to perform the task, time measurement and system activation were triggered simultaneously by the investigator. Time was logged until the participant signaled task completion. To include the amount of time required for overshooting and correction, time measurement was stopped manually by the investigator and corrected afterwards using video logs. The last foot movement before the participant conveyed task completion was defined as stop cue. Additionally, the amounts of time required to fulfill the subtasks described in section 3 were identified by analyzing the video logs. Since we were interested in the time required to complete the subtasks as workflow steps, we did not measure single interactions technically, but the time required to achieve each subtask. This means, even when a user involuntarily selected another viewport during image manipulation and had to reselect the correct one, the time was counted towards image manipulation instead of choosing and confirming a viewport.

We used a Raw TLX (RTLX) [9] questionnaire to assess subjective workload. One questionnaire was filled out for each task block.

## 6.5 Procedure

The study took place in an angiography suite. Initially, a demographic questionnaire was filled out by the participants. Additionally, experience with gaze or feet as input modality were assessed, as was shoe size. The sequence of the three input setups OTF, DLG and FI was counterbalanced over all participants to reduce learning effects. The tasks for each setup remained the same and were not randomized because Saunders and Vogel showed that kicking forwards is more effective than backwards [34], which means that the direction of foot movements could influence the performance. The participants wore OR shoes with a hard rubber outer sole during the procedure to avoid different recognition accuracy caused by different types of shoes. A position to stand on was marked on the floor to maintain the same distance and orientation to the screen for all participants. Regardless of the setup sequence, participants wore the eye tracker during the whole study. At setups OTF and DLG, a 16-point eye tracking calibration was performed at the beginning and repeated between tasks when participants experienced inaccurate results. For each input setup, the investigator explained the system to the participants, followed by a free training phase. To ensure a minimal level of confidence with the system, all participants had to perform a test before the measured tasks were performed. Three tasks needed to be fulfilled in under 1:30 min to proceed. The test would have been repeated until the user was able to finish it, but all participants passed on the first try. During measured tasks 1.1 to 2.3 (see Table 3), each instruction was read out by the investigator beforehand to separate comprehension times from task completion times. After each block, participants filled out an RTLX questionnaire to assess subjective workload.

**Table 3: Summary of the test statistics for task completion times and subjective workload.**

| | df | F | p | $\eta^2_{part}$ | Effect |
|---|---|---|---|---|---|
| **Task completion time** | | | | | |
| setup | 2, 24 | 2.27 | .13 | .16 | large |
| subtasks | 1.16, 13.95 | 207.90 | <.01 | .95 | large |
| interaction | 1.89, 22.69 | 47.92 | <.01 | .80 | large |
| **Subjective workload** | | | | | |
| setup | 2, 24 | 1.10 | .35 | .08 | medium |

## 6.6 Results

Task completion times were analyzed by a 3 x 3 ANOVA with the factors setup (OTF, DLG, FI) and step (choose viewport, confirm viewport, manipulate data). If the sphericity assumption was violated, we used a Greenhouse-Geisser correction of degrees of freedom. The analysis revealed a significant main effect for subtasks and a significant interaction effect between setup and subtask (see Table 3). For choosing a viewport, considerable shorter times were achieved by gaze (OTF and DLG) than by foot interaction FI (see Figure 4). Confirming a viewport was accomplished the fastest by continuous gaze (OTF), followed by FI, and took the longest with DLG. Extended task completion times were observed for subtasks which utilize gaze pointing and simultaneous foot interaction. OTF was affected the strongest, but also DLG during the short gaze-and-foot gesture for viewport confirmation. Although the same modality (foot input) was used for image manipulation in all setups, completion times for this subtask differ: participants needed considerably more time for image manipulation using OTF compared to FI and DLG.

RTLX values for both task blocks were averaged for each setup due to insignificant differences. Subjective workload was analyzed with a one-way ANOVA for repeated measures with the input setup (OTF, DLG and FI) as the only factor. This analysis revealed no significant result. A detailed analysis of RTLX dimensions revealed similar values for all input setups (see Figure 3).

# 7 DISCUSSION

We developed and evaluated concepts for hands-free interaction with image data in the OR, utilizing eye gaze and the feet as input

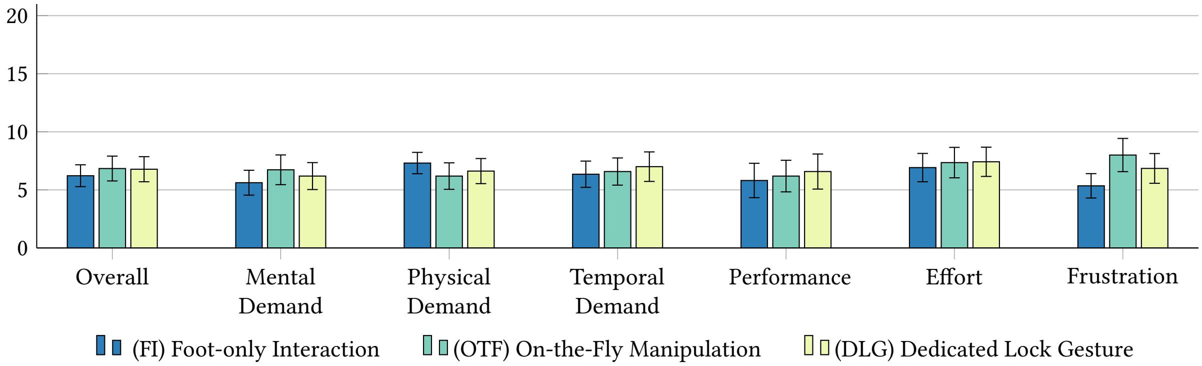


**Figure 3: Mean subjective workload for the user study with standard error bars. (0 = low/good, 20 = high/poor).**

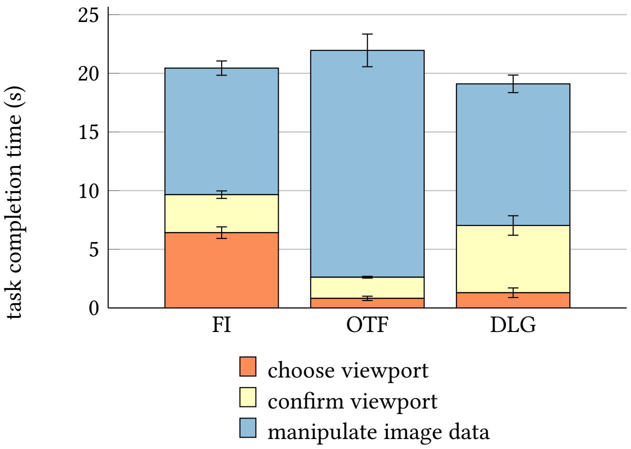


**Figure 4: Mean task completion times in seconds divided into subtasks for *Foot-only Interaction* (F), *Dedicated Lock Gesture* (DLG) and *On-the-Fly Manipulation* (OM) with standard error bars for each subtask.**

modalities. The evaluation shows no significant difference between gaze and foot approaches and foot-only interaction. However, investigation of task completion times for individual subtasks revealed considerable differences. Using gaze allows fast determination of the coarse area of attention and therefore is more suitable for tasks such as viewport selection than foot interaction (see *choose viewport* in Figure 4). Subtasks which involve simultaneous gaze and foot interaction took longer than ones which required solely foot input. This means more time was needed for image manipulation in OTF, and viewport confirmation in DLG. Velloso et al. analyzed trial completion times for gaze selection and mid-air gestures in a similar way, separated in three steps (acquisition, confirmation, translation) [46]. Our results align when it comes to fast target acquisition via gaze, but differ in subsequent subtasks/steps. Whereas Velloso et al. reports almost consistent times for target confirmation and translation over all techniques, image data manipulation took longer when gaze selection had to be maintained during that period in our study. The same goes for DLG, which required simultaneous gaze and foot interaction. A possible reason might be the need for visual checks of the feet when interaction is performed outside the field of view, compared to the fast an easy finger-pinch-gesture used by Velloso et al.

Participants seemed to gaze at the feet to maintain a stable stance, to put them back side by side after pressing a button and to confirm postures and positions of the limbs when proprioceptic feedback is not sufficient. Interference caused by the double-role of the eyes for observation and control when using gaze for interaction is well known [13]. Several studies report difficulties regarding interaction outside the field of view or in peripheral vision [20, 43]. In contrast, no difficulties were found when the limbs (i.e., hands) are in close proximity to the screen [3, 28, 46]. An approach to tackle this problem when eye tracking is used might be to detect whether the users look away, to disable additional input devices. Unfortunately, it is hard to distinguish the last intended gaze position from unintentional fixations while moving the gaze away from the screen. A possible solution might be to discard gaze information from a few milliseconds before looking away from the screen retrospectively. When comparing task completion times for DLG and FI, the intention we had in mind when designing DLG seems to work out even though combined completion times of all subtasks do not differ significantly. Gaze interaction allows choosing a viewport faster than using the feet, while image manipulation works well since eye gaze is not used for interaction at this point and therefore cannot interfere. Since the gaze-and-foot gesture suggests room for improvement, DLG seems to be the most promising approach. A possible solution might be a lock on a selection as soon as the beginning of a foot gesture is detected. A faster, easier foot gesture such as a single or double tap might be used, but requires safe discrimination from steps or unintentional movements.

Overall, all participants were able to fulfill the given tasks, which indicates that gaze and foot interaction in general is a suitable approach when the hands are not available and direct control over a system is required.

## 7.1 Limitations

Even though our approaches performed well given the limited experience the participants had with these kinds of input methods, the system design might have some minor flaws which need to be addressed in subsequent studies. Visual feedback was kept to a minimal level to keep the view of the medical data clear, and therefore lacked a gaze point indicator. Slightly off or drifting calibration of the eye tracker, which could occur through movement of the head-mounted part, was not detected immediately. This can be compensated for by calibration-less gaze interaction techniques such as *smooth path pursuit* [47] or a function to temporarily show a gaze point indicator. The need to look at the feet might be influenced by the fixed position of the sensitive areas on the floor. Saunders and Vogel suggested automated adjustments of the center position based on movement patterns typical for correcting the stance [35]. This issue will be resolved when investigating different, mobile foot input methods, which will be necessary because of certain requirements explained in the following.

From a broader view, our work focuses on the mere interaction method and does not take important aspects of a real OR situation into account yet. Simultaneous tasks performed with the hands might be more demanding and might influence the subjective workload. Foot interaction was used for directional input only, which is natural because a spatial mapping is implicitly given. Increasing functionality of the system will demand more abstract foot interaction that might not align with the natural mapping used in this study. Additionally, physicians have to deal with various auditive and visual stimuli such as interpersonal communication, medical instruments or system status notifications. Therefore, we have to account for interruptions in the human-computer communication and resume as seamlessly as possible soon after. When it comes to foot input, rooting the user to a spot with big, fixed interaction areas to step on isn't suitable in the long term since space at the operating table is limited by hardware and supporting staff. Less space-demanding and mobile foot input methods need to be investigated.

## 8 FUTURE WORK

Our findings raise new questions for gaze and foot interaction when the user is standing. The results of our study show fast pointing via gaze but difficulties when the position has to be kept steady for confirmation or interaction with the feet. This problem can be tackled from two directions. The first one is by selection methods which use gaze for pointing but switch to other modalities at an early state in the selection process. Therefore, strategies to determine optimal situations and their indicators to switch modalities have to be investigated. Such an approach might additionally tackle the "leave before click" issue described by Jacob and Stellmach [13], as both problems require a more reliable gaze position during a multimodal selection process.

A second approach might be to deliver more information about the user's feet because we believe an issue arises from the need to check the feet while interacting with gaze. Status information about the feet might be displayed right at the gaze position on the screen. Alternatively, foot interaction might benefit from wearable sensors which allow vibrotactile feedback, user identification and user-specific parametrization.

In the long term, further development of hands-free interaction for minimally-invasive interventions needs to take domain-specific factors into account. In the OR, many visual feedback systems, personnel, and additional medical equipment require visual checks and additionally interrupt gaze interaction. Foot interaction has to be robust when it comes to multiple individuals staying near the physician. Additional functions such as zoom, pan, and changing contrast and brightness need to be implemented to fulfill all needs physicians may have in the OR. Furthermore, the system control must be handed over often during interventions when specialists are involved. Since interventions can last for several hours, gestures which can be performed for a long time without fatiguing the user have to be found and evaluated.

## 9 CONCLUSION

In this work, we presented two approaches to allow hands-free interaction by utilizing gaze and foot as input channels. Therefore, we utilized input modalities mostly used for lab studies, applied them to a real problem in the medical domain and evaluated our interaction techniques in a realistic space. Our proposed interaction techniques performed comparably and allowed successful completion of all given tasks during the evaluation. We could show the potential and challenges of gaze and foot input for hands-free interaction. We confirmed gaze as an excellent modality for pointing tasks but found that it can easily interfere when other body parts are used for interaction. The need for visual checks, especially when controls without feedback are used outside the field of view, voids gaze data and has to be taken into account when creating combined interaction methods. Our results suggest that gaze should be used for pointing tasks, but needs to be confirmed by other modalities. We believe that this work can inform the design of hands-free user interfaces for many domains where delicate, non-interruptible motor tasks can be supported by accessing information directly.

Our findings gives insights when using gaze and foot at an upright stance, which will be relevant for future research when using mobile multimodal interaction. Applied to the medical field, our system has the potential to minimize communication errors by reducing required personnel and therefore make minimally-invasive interventions safer, faster, and more affordable in the long run.

## ACKNOWLEDGMENTS

This work is partially funded by the Federal Ministry of Education and Research (BMBF) within the STIMULATE research campus (grant number 13GW0095A) and the European Regional Development Fund under the operation number 'ZS /2016/04/78123' as part of the initiative "Sachsen-Anhalt WISSENSCHAFT Schwerpunkte". This research was supported by the mobility grant of the German Society for Computer- and Robot-Assisted Surgery (CURAC).

## REFERENCES

[1] Alexandre Alapetite. 2008. Impact of noise and other factors on speech recognition in anaesthesia. *International journal of medical informatics* 77, 1 (2008), 68–77.

[2] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting your best foot forward: investigating real-world mappings for foot-based gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1229–1238.

[3] Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+Gesture. In *the 2015 ACM*, Zhengyou Zhang, Phil Cohen, Dan Bohus, Radu Horaud, and Helen Meng (Eds.). 131–138. https://doi.org/10.1145/2818346.2820752

[4] Lars C. Ebert, Gary Hatch, Garyfalia Ampanozi, Michael J. Thali, and Steffen Ross. 2012. You can't touch this touch-free navigation through radiological images. *Surgical innovation* 19, 3 (2012), 301–307.

[5] Douglas Engelbart. 1984. Doug Engelbart Discusses Mouse Alternatives.(May 1984). *Retrieved March* 3 (1984), 2014.

[6] Markus Fritzsche, José Saenz, and Felix Penzlin. 2016. A large scale tactile sensor for safe mobile robot manipulation. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction.* 427–428.

[7] Charles P. Gerba, Adam L. Wuollet, Peter Raisanen, and Gerardo U. Lopez. 2016. Bacterial contamination of computer touch screens. *American journal of infection control* 44, 3 (2016), 358–360.

[8] Sébastien Grange, Terrence Fong, and Charles Baur. 2004. M/ORIS: a medical/operating room interaction system. In *Proceedings of the 6th international conference on Multimodal interfaces.* 159–166.

[9] 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.

[10] J. Hempel, A. Brychtova, Ioannis Giannopoulos, Sophie Stellmach, Raimund Dachselt, and others. 2016. Gaze and feet as additional input modalities for interacting with geospatial interfaces. (2016).

[11] Julian Hettig, Patrick Saalfeld, Maria Luz, Mathias Becker, Martin Skalej, and Christian Hansen. 2017. Comparison of gesture and conventional interaction techniques for interventional neuroradiology. *International Journal of Computer Assisted Radiology and Surgery* (2017). https://doi.org/10.1007/s11548-017-1523-7

[12] A. Hübler, C. Hansen, O. Beuing, M. Skalej, and B. Preim. 2014. Workflow Analysis for Interventional Neuroradiology using Frequent Pattern Mining. In *Proceedings of the Annual Meeting of the German Society of Computer- and Robot-Assisted Surgery.* Munich, 165–168.

[13] Rob Jacob and Sophie Stellmach. 2016. What you look at is what you get: gaze-based user interfaces. *interactions* 23, 5 (2016), 62–65.

[14] Robert J. K. Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* 9, 2 (1991), 152–169.

[15] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. 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

[16] Shahram Jalaliniya, Jeremiah Smith, Miguel Sousa, Lars Büthe, 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.

[17] Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '14 Adjunct).* ACM, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

[18] Konstantin Klamka, Andreas Siegel, Stefan Vogt, Fabian Göbel, Sophie Stellmach, and Raimund Dachselt. 2015. Look & Pedal: Hands-free Navigation in Zoomable Information Spaces through Gaze-supported Foot Input. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction.* 123–130.

[19] Päivi Majaranta and Kari-Jouko Räihä. 2007. Text entry by gaze: Utilizing eye-tracking. *Text entry systems: Mobility, accessibility, universality* (2007), 175–187.

[20] Michael Mauderer, Florian Daiber, and Antonio Krueger. 2013. Combining Touch and Gaze for Distant Selection in a Tabletop Setting. (2013).

[21] Helena M. Mentis, Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Robert Corish, Antonio Criminisi, Rikin Trivedi, and Pierre Theodore. 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. https://doi.org/10.1145/2702613.2702963

[22] Andre Mewes, Bennet Hensen, Frank Wacker, and Christian Hansen. 2017. Touchless interaction with software in interventional radiology and surgery: a systematic literature review. *International Journal of Computer Assisted Radiology and Surgery* 12, 2 (2017), 291–305. https://doi.org/10.1007/s11548-016-1480-6

[23] Brian Meyers, A. J. Brush, Steven Drucker, Marc A. Smith, and Mary Czerwinski. 2006. Dance your work away: exploring step user interfaces. In *CHI'06 extended abstracts on Human factors in computing systems.* 387–392.

[24] Veit Müller, Markus Fritzsche, and Norbert Elkmann. 2015. Sensor design and calibration of piezoresistive composite material. In *SENSORS, 2015 IEEE.* 1–4.

[25] M. D. Nicola Bizzotto, M. D. Alessandro Costanzo, and M. D. Leonardo Bizzotto. 2014. Leap motion gesture control with OsiriX in the operating room to control imaging: first experiences during live surgery. *Surgical innovation* 1 (2014), 2.

[26] Kenton O'Hara, Gerardo Gonzalez, Abigail Sellen, Graeme Penney, Andreas Varnavas, Helena Mentis, Antonio Criminisi, Robert Corish, Mark Rouncefield, Neville Dastur, and others. 2014. Touchless interaction in surgery. *Commun. ACM* 57, 1 (2014), 70–77.

[27] Toni Pakkanen and Roope Raisamo. 2004. Appropriateness of foot interaction for non-accurate spatial tasks. In *CHI'04 extended abstracts on Human factors in computing systems.* 1123–1126.

[28] Ken Pfeuffer, Jason Alexander, Ming Ki Chong, Yanxia Zhang, and Hans Gellersen. 2015. Gaze-Shifting. In *the 28th Annual ACM Symposium*, Celine Latulipe, Bjoern Hartmann, and Tovi Grossman (Eds.). 373–383. https://doi.org/10.1145/2807442.2807460

[29] Vijay Rajanna and Tracy Hammond. GAWSCHI. In *the Ninth Biennial ACM Symposium*, Pernilla Qvarfordt and Dan Witzner Hansen (Eds.). 233–236. https://doi.org/10.1145/2857491.2857499

[30] Anke Verena Reinschluessel, Joern Teuber, Marc Herrlich, Jeffrey Bissel, Melanie van Eikeren, Johannes Ganser, Felicia Koeller, Fenja Kollasch, Thomas Mildner, Luca Raimondo, and others. 2017. Virtual Reality for User-Centered Design and Evaluation of Touch-free Interaction Techniques for Navigating Medical Images in the Operating Room. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* 2001–2009.

[31] Felix Ritter, Tobias Boskamp, André Homeyer, Hendrik Laue, Michael Schwier, Florian Link, and H-O Peitgen. 2011. Medical image analysis. *IEEE pulse* 2, 6 (2011), 60–70.

[32] William A. Rutala, Matthew S. White, Maria F. Gergen, and David J. Weber. 2006. Bacterial contamination of keyboards: efficacy and functional impact of disinfectants. *Infection Control & Hospital Epidemiology* 27, 04 (2006), 372–377.

[33] Nuttapol Sangsuriyachot and Masanori Sugimoto. 2012. Novel interaction techniques based on a combination of hand and foot gestures in tabletop environments. In *Proceedings of the 10th asia pacific conference on Computer human interaction.* 21–28.

[34] William Saunders and Daniel Vogel. 2015. The performance of indirect foot pointing using discrete taps and kicks while standing. In *Proceedings of the 41st Graphics Interface Conference.* 265–272.

[35] William Saunders and Daniel Vogel. 2016. Tap-Kick-Click: Foot Interaction for a Standing Desk. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems.* 323–333.

[36] Maureen Schultz, Janet Gill, Sabiha Zubairi, Ruth Huber, and Fred Gordin. 2003. Bacterial contamination of computer keyboards in a teaching hospital. *Infection Control & Hospital Epidemiology* 24, 04 (2003), 302–303.

[37] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* 281–288.

[38] Adalberto L. Simeone, Eduardo Velloso, Jason Alexander, and Hans Gellersen. 2014. Feet movement in desktop 3D interaction. In *2014 IEEE Symposium on 3D User Interfaces (3DUI).* 71–74.

[39] Dave M. Stampe and Eyal M. Reingold. 1995. Selection by looking: A novel computer interface and its application to psychological research. *Studies in Visual Information Processing* 6 (1995), 467–478.

[40] Sophie Stellmach and Raimund Dachselt. 2012. Investigating gaze-supported multimodal pan and zoom. In *Proceedings of the Symposium on Eye Tracking Research and Applications.* 357–360.

[41] Sophie Stellmach and Raimund Dachselt. 2013. Still looking: investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 285–294.

[42] Duncan Stevenson, Henry Gardner, Wendell Neilson, Edwin Beenen, Sivakumar Gananadha, James Fergusson, Phillip Jeans, Peter Mews, and Hari Bandi. 2016. Evidence from the surgeons: gesture control of image data displayed during surgery. *Behaviour & Information Technology* 35, 12 (2016), 1063–1079.

[43] Jayson Turner, Andreas Bulling, Jason Alexander, and Hans Gellersen. 2013. Eye drop: an interaction concept for gaze-supported point-to-point content transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia.* 37.

[44] Eduardo Velloso, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. Interactions Under the Desk: A Characterisation of Foot Movements for Input in a Seated Position. In *Human-Computer Interaction – INTERACT 2015 (Lecture Notes in Computer Science)*, Vol. 9296. Springer International Publishing, Cham, 384–401. https://doi.org/10.1007/978-3-319-22701-6

[45] Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The Feet in Human–Computer Interaction: A Survey of Foot-Based Interaction. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 21.

[46] Eduardo Velloso, Jayson Turner, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2015. An empirical investigation of gaze selection in mid-air gestural 3D manipulation. In *Human-Computer Interaction – INTERACT 2015 (Lecture Notes in Computer Science).* Springer International Publishing, Cham, 315–330.

[47] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2013. Pursuits: Spontaneous Interaction with Displays Based on Smooth Pursuit Eye Movement and Moving Targets. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13).* ACM, New York, NY, USA, 439–448.

[48] Colin Ware and Harutune H. Mikaelian. 1987. An evaluation of an eye tracker as a device for computer input2. In *ACM SIGCHI Bulletin*, Vol. 17. 183–188.

[49] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* 246–253.