Look & Pedal: Hands-free Navigation in Zoomable Information Spaces through Gaze-supported Foot Input

Konstantin Klamka[1], Andreas Siegel[1], Stefan Vogt[1], Fabian Göbel[1], Sophie Stellmach[2], Raimund Dachselt[1]

[1]Technische Universität Dresden
Dresden, Germany
{first.last}@tu-dresden.de

[2]Microsoft Corporation
Redmond, WA, USA
stellmach@acm.org

ABSTRACT

For a desktop computer, we investigate how to enhance conventional mouse and keyboard interaction by combining the input modalities gaze and foot. This multimodal approach offers the potential for fluently performing both manual input (e.g., for precise object selection) and gaze-supported foot input (for pan and zoom) in zoomable information spaces in quick succession or even in parallel. For this, we take advantage of fast gaze input to implicitly indicate where to navigate to and additional explicit foot input for speed control while leaving the hands free for further manual input. This allows for taking advantage of gaze input in a subtle and unobtrusive way. We have carefully elaborated and investigated three variants of foot controls incorporating one-, two- and multidirectional foot pedals in combination with gaze. These were evaluated and compared to mouse-only input in a user study using Google Earth as a geographic information system. The results suggest that gaze-supported foot input is feasible for convenient, user-friendly navigation and comparable to mouse input and encourage further investigations of gaze-supported foot controls.

Categories and Subject Descriptors

H.5.2 [**Information interfaces and presentation**]: User Interfaces: Input devices and strategies.

Keywords

Multimodal interaction; foot input; gaze input; eye tracking; gaze-supported interaction; navigation; pan; zoom

1. INTRODUCTION

In common computer desktop workspaces, the interaction is usually limited to mouse and keyboard input. Especially complex tasks, which are common in a variety of application areas, e.g., in geographic information systems (GIS), often require to switch between interaction modes (e.g. *navigation* and *manipulation*). If the user wants to drag an object to a remote point of the map, the mouse controls used for dragging can no longer be used for panning, and parallel navigation would be welcome. While the hands are occupied, other modalities such as gaze and foot input remain unused, although they show high potential to support and enhance conventional computer interaction.

Figure 1: Envisioned setup: A single user works in a common computer desktop environment and can benefit from additional gaze and foot input for fluently navigating through zoomable information spaces while leaving the hands free for other tasks.

On the one hand, gaze input is suitable as fast and implicit pointing modality, however, lacks precision and should be used with care to avoid overwhelming the user. On the other hand, foot input allows for parallel explicit input controls, for example, to quickly confirm an action or to gradually adjust movement speeds, while it is less appropriate for precise pointing tasks [14, 18]. Thus, while our eye gaze is ideal to quickly indicate a user's current point of interest, foot interaction is well suited for parallel hands-free input controls [17]. This allows for using gaze input in a subtle and unobtrusive way, while still maintaining a fast and convenient interaction by addressing prevalent challenges associated with gaze interaction such as the *Midas Touch* problem [15] (i.e., involuntarily issuing a command via gaze). Although Pearson and Weiser [19] have already indicated the promising potential for combining gaze and foot input in the mid 1980s, further investigations of this multimodal input combination are still required.

This paper focuses on the exploratory investigation of a novel way of multimodal interaction, namely a combination of gaze, foot and manual input, and contributes to a better understanding of how to exploit the unique benefits of gaze and foot input in concert with manual controls. As an example, users could perform high precision *(primary)* interaction tasks with their hands to precisely select and manipulate objects or for fast text entry. Meanwhile, gaze and foot input could support *(secondary)* navigation tasks in zoomable information spaces for browsing through geographic data in GIS or for panning and zooming in an image editor. For example, while editing or annotating a picture, the user may want to get a closer look at details by further zooming in or moving the currently manipulated graphical content towards the center of the screen (i.e., pan the image) without the need to move the hands from mouse or keyboard or to switch system modes. With *primary* tasks we refer

ICMI 2015, November 9–13, 2015, Seattle, WA, USA.
© 2015 ACM. ISBN 978-1-4503-3912-4/15/11 ...$15.00.
DOI: http://dx.doi.org/10.1145/2818346.2820751.

to interactions that may require high input precision and concentration and which are considered more suitable for manual input. *Secondary* tasks are supporting activities and should be executable with low mental and physical effort to avoid unnecessarily distracting users from their primary task.

Motivated by this, we investigate how pan and zoom actions can be performed through a combination of gaze-supported input [25] and foot controls in a desktop computer setup along with additional manual mouse input for primary interaction tasks (namely interacting with markers on a virtual geographic map) as illustrated in Figure 1. We carefully elaborated and examined several combinations of gaze input with different types of foot pedals ranging from common one-directional car-like pedals, two-directional rocking pedals to multi-dimensional foot joysticks. In our iterative design process, we developed several prototypes and evaluated them in two user studies. To summarize, in this work, we contribute to a better understanding of gaze-supported foot navigation, including:

- **Novel input:** We contribute a novel multimodal combination of implicit gaze and explicit foot input for navigation tasks combined with manual input for further tasks.
- **Feasibility:** We successfully demonstrate that gaze-supported foot input is feasible for navigation and explore particular design choices to improve performance and usability.
- **Usability:** We present a combination of gaze and foot input that was assessed very positively and is interesting for further investigations (e.g. in other application contexts).
- **Challenges:** We discuss several design considerations and potential risks for gaze-supported foot interaction that provide a foundation for on-going work.

2. GAZE-BASED PAN AND ZOOM

Our eye gaze presents an implicit and quick way to indicate interest and provides one of the fastest possible pointing methods, as our gaze reaches a target prior to a manual pointer without even thinking about it [15, 26]. For navigation controls, this can be used to quickly enlarge looked-at content either locally (zoom lenses) (e.g., [3, 16, 25]) or globally across the entire screen space [1, 4, 12, 23].

Several works have investigated how to use gaze input for pan and zoom. For instance, Hansen et al. [12] use a scalable rectangular mask superimposed over the screen with an inner zone to *pan & zoom* and an outer zone toward the screen border to *only pan*. So, if looking at the pan area, the currently viewed target will move towards the center of the screen. They use gaze-only input, which makes it difficult to adjust movement speed or to quickly zoom in and back out. Similarly, Adams et al. [1] use distinct pan regions at each screen border that are associated with a particular pan direction and speed. They compare four different pan and zoom input techniques including gaze-based panning in combination with different input techniques for zooming, such as mouse clicks, head movements towards or away from the screen, and gaze-based dwelling. While none of the gaze-based methods proved to be as efficient as conventional mouse input, user feedback for the novel techniques was encouraging.

Zhu et al. [28] present a panning-only gaze user interface for remotely controlling a camera for which the captured camera image is displayed on a desktop screen. The panning direction depends on the vector between screen center and current gaze point, whereby the panning speed refers to the length of this vector. This approach allows for continuously moving the currently viewed target towards the screen center similar to the work from Hansen et al. [12] and Adams et al. [1].

Stellmach and Dachselt [23] investigate five gaze-supported pan and zoom variants in combination with a mouse scroll wheel, single touch gestures, and tilting of a handheld smartphone. In particular gaze-based pivot zooming [25], the implicit use of gaze to indicate where to zoom in, was positively highlighted by participants in their user study. This refers to the *fixed point zoom* proposed by Furnas and Bederson [10] for which a selected pivot (e.g., the point-of-regard) remains at a fixed display location during zooming, in contrast to *central zoom* for which the zoom pivot is always at the screen center. In addition, Stellmach and Dachselt [23] point out that while gaze-directed panning was positively assessed for slow exploration of an area, quick gaze-directed movements sometimes lead to disorientation and motion sickness for the user [23].

3. FOOT-BASED NAVIGATION

Already in the 1980s, Pearson and Weiser [19] have proposed using our feet as supporting input to overcome interruptions of the user's workflow due to shifting the hand between mouse and keyboard. In this context, several studies indicate that feet are rather suitable for low accuracy interaction tasks, such as scrolling [14, 18]. Various works have investigated foot controls for navigating in virtual environments, e.g., using floor pads [5, 17], a Wii Balance Board™ [9, 20], and treadmills [8, 22]. Several works also address *multimodal* foot input, e.g., with multi-touch gestures [7, 20], but none with gaze input. However, Pearson and Weiser have already mentioned the potential of combining gaze and foot input, but did not further investigate this approach due to expensive and cumbersome eye tracking equipment at that time [19].

Two general types of foot input can be distinguished: *discrete* (e.g., taps and kicks) and *continuous* (e.g., gradually pushing down a foot pedal). According to Crossan et al. [6], single and double foot taps can be used for conveniently navigating in a menu (e.g., on a mobile device). Scott et al. [21] examine differences between toe and heel tapping and conclude that users find toe tapping more comfortable. Beckhaus et al. [5] compare discrete foot taps on a dance pad with continuous input from leaning into certain directions on a chair. While both techniques were assessed as easy to understand, the chair variant was preferred, because it was more enjoyable to use and more flexible for quickly performing several movements.

An example for *continuous* foot input is presented by Alexander et al. [2] investigating *distance-based* and *velocity-based* foot movements (kicks) to pan on a mobile map. They conclude that continuous velocity-based foot input requires less physical effort, provides higher control, is faster and more accurate. Further examples for continuous foot navigation are presented by Haan et al. [9] and Schöning et al. [20] using a Wii Balance Board™. Users can steer by shifting their weight in the corresponding direction. Haan et al. point out that users had problems stopping a movement in time, because quick weight shifts were difficult. Schöning et al. [20] combine foot with multi-touch interaction for more precise navigation of spatial data on a large-sized display. For example, a user can zoom in by pointing with the hand to a location on the map and lean forward on the Balance Board™. In addition, it is possible to pan the map towards the direction in which a user shifted his/her weight. While their initial evaluation suggests that the proposed interaction technique allows a smooth navigation, the lacking haptic feedback of the Balance Board™ led to overshooting problems.

In a nutshell, several works have highlighted the promising potential of gaze- and foot-supported navigation. For this, fast and simple ways to activate and deactivate gaze input and to smoothly control movement speeds are required. While continuous rate-based foot input is promising for this, the need for haptic feedback has been pointed out. Foot pedals, as familiar foot controls, provide this feature and have not yet been thoroughly investigated for gaze-supported foot navigation.

4. GAZE-SUPPORTED FOOT NAVIGATION

Gaze-supported foot navigation shows high potential for conveniently performing pan and zoom actions, while leaving the hands free for additional manual input. However, as this multimodal combination has not yet been investigated before, we are faced with the question in which way foot and gaze input may complement each other well to offer convenient and fast pan and zoom control. While gaze input is suitable as a fast, implicit and coarse pointing modality, foot input seems well suited for trigger actions, as it allows parallel explicit input controls. In the following, we first describe several of our design considerations for our investigation of *gaze-supported foot navigation*. Secondly, we describe our three elaborated setups that combine gaze input with different variants of foot pedals.

4.1 Design Considerations

For our design, we assume an ordinary desktop computer setup with a single user who sits on a chair and works with mouse and keyboard in an application that requires both *primary* interaction tasks (e.g., object selection and positioning) and *secondary* navigation tasks (pan and zoom).

Gaze input is used to define the position where to zoom in (or out) and to indicate the pan direction (e.g., also see [23, 28]), while foot input is used to control pan and zoom speeds. Please note that a negative zoom speed is interpreted as zoom *out*. For this, we take advantage of multiple cursors for manual and gaze input. This avoids a conflict between the mouse cursor required for selecting content and a secondary pointer to indicate the steering direction. Hence, the mouse cursor is not influenced by gaze or foot input.

In the following, we discuss how to address certain challenges associated with gaze input and several design aspects for convenient gaze- and foot-supported interaction.

4.1.1 Gaze Interaction

One major challenge is the *Midas Touch* problem [15]: the unintentional triggering of an action by merely looking at a control element. By combining implicit gaze with explicit foot input, we can provide the user with the ability to clearly and conveniently communicate when to perform a certain action. Also, if no foot input is performed, the user can freely look around without accidentally triggering an action.

Another major challenge is the *double role* of our eye gaze for orientation and control [24]. This is particularly critical for *fast* gaze-controlled panning movements [23], as the user needs to orientate him-/herself in the scene, while also trying to control the pan direction and speed with his/her gaze. To address this issue, we investigate different combinations of gaze and foot input to release the eyes for free observation. On the one hand, this includes the above mentioned approach: The user can quickly switch between gaze control and pure observation mode if not performing any foot input. This gives the user more control, as s/he can quickly stop if feeling lost. On the other hand, both *pan direction* and *speed* can be controlled via foot input to completely free the eyes for observation and orientation.

Another challenge is how to cope with *inaccurate* and *unreliable* gaze data. Jittery gaze data can be counteracted by smoothing tracked data. Offsets are more difficult to address and thus, low-precision pointing tasks are recommend for convenient and fast gaze input. Therefore, only actions that do not require high precision should be controlled via eye gaze. If the tracked gaze point is completely lost (e.g., due to tracking problems), we use the last valid gaze point.

4.1.2 Foot Interaction

For the design of *convenient* yet effective foot navigation in our envisioned desktop scenario (cf. Figure 1), we take several design goals into account. Firstly, the foot controls should require *low effort* for prolonged use. Thus, excessive foot movements should be avoided. Secondly, the foot controls should offer *high control* to the user for smoothly steering through a virtual scene, for example, to easily perform fast movements or to quickly stop to avoid overshooting. Thirdly, the foot controls should be *easy to use*, yet *fast* and (sufficiently) *precise*.

Based on these design goals, we consider foot pedals as promising, because many people are already familiar with them (e.g. from driving a car). They offer high control for fast and precise navigation by allowing *continuous rate-based* input to gradually adjust movement speeds and/or directions and by offering haptic feedback when stepping on a pedal. In addition, high-resolution input devices allow a precise and smooth mapping of foot movements to navigation tasks. Furthermore, foot pedals are *unobtrusive*, as they can be installed in the desktop environment, instead of attaching any equipment to the user's body. Thus, similar to a car scenario, a user can simply sit down and start interacting with the system without any further preparations.

4.2 Foot Input Devices

Based on the described design considerations, we selected pedals that match the pan+zoom techniques best in terms of degrees of freedom, resting position and possible mappings. We elaborated a range of stationary foot pedals, ranging from one-, two- to multi-directional foot pedals which are described in the following.

4.2.1 One-directional Foot Pedals

First, we consider commercial *foot pedals*, building on people's familiarity with common car pedals (see Figure 3a). They offer the most constrained interaction by only allowing a one-directional movement: Pushing the pedal forward or releasing it. Haptic feedback is provided by its increasing resistance. The aluminium pedals are robust and offer various adjustment possibilities to adapt to different user preferences, e.g., the tilt angle, order and orientation of pedals. The pedals provide an input range of 1024 distinct values (10-bit resolution) allowing a continuous and fine-grained speed control.

4.2.2 Two-directional Foot Pedal: Foot-Rocker

To allow for more freedom of foot movement compared to commercial car foot pedals, we designed the *Foot-Rocker* (see Figure 3b). It extends the one-directional movement of the commercial pedals by adding a second direction, allowing a back and forth tilting mo-

Figure 3: a) One-directional Pedals, b) Two-directional Foot-Rocker and c) Multi-directional Foot-Joystick.

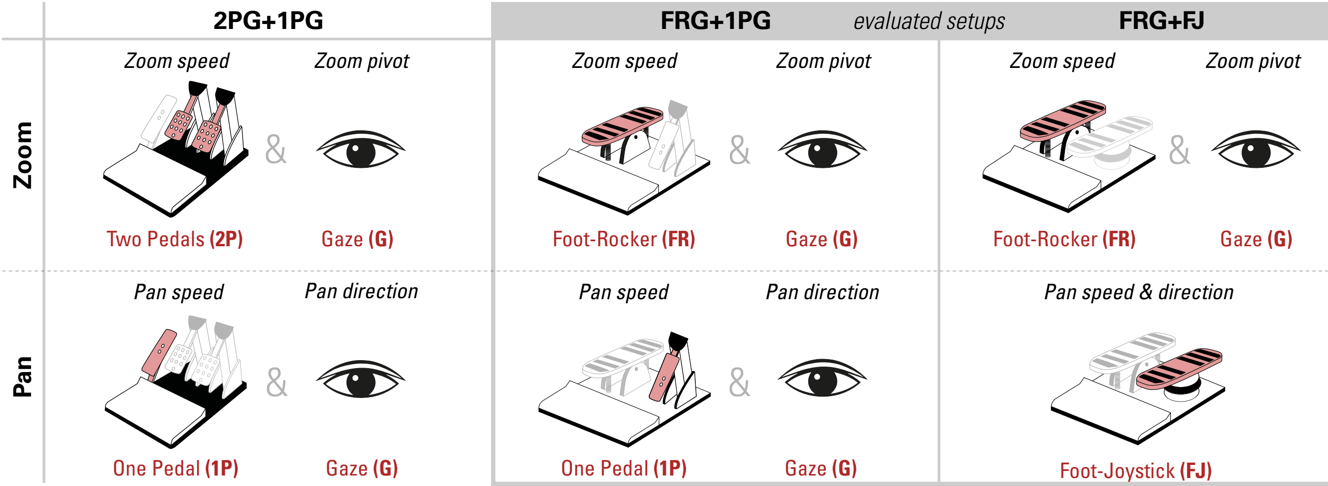

Figure 2: Investigated gaze- & foot-supported pan-and-zoom combinations.

tion. We constructed the *Foot-Rocker* with a fixed centered axis that can be tilted by about 20°, returning to the neutral position with the help of springs. Due to its prototypical character, the joystick sometimes suffers from little stiffness and slightly incomplete returns to the neutral position.

4.2.3 Multi-directional Foot Pedal: Foot-Joystick

We also designed and built *Foot-Joystick*, which extends input capabilities even further. Its ball joint mounted pedal construction provides the user with two degrees of freedom and a deviation angle of 20° in all directions from a predefined neutral position (see Figure 3c) which is also maintained by springs.

4.3 Foot Input Setups

Based on the described foot input variants, we elaborated several gaze and foot input combinations to control pan and zoom directions and speeds in an iterative design process. In the following, the three most promising gaze-supported foot navigation techniques are described in detail. In the remaining paper, the zoom-and-pan techniques are abbreviated with a two-part notation [zoom]+[pan] as listed in Figure 2.

4.3.1 Two Pedals & Gaze + One Pedal & Gaze (2PG+1PG)

The first setup consists of three commercial foot pedals. The user can zoom towards/away from a looked-at area (see gaze-pivot zooming [25]) by pushing the right/middle pedal. Simultaneously, the user can pan towards the current gaze position by pressing the left pedal, which is used for activation and increasing panning speed. It is noteworthy that this setup uses an asymmetric mapping of zoom directions.

4.3.2 Foot-Rocker & Gaze + One Pedal & Gaze (FRG+1PG)

To avoid the potential disadvantage of the first setup and to allow a symmetric mapping of zoom directions, we combine a single pedal with the two-directional foot pedal (the *Foot-Rocker*) in the second setup. Zoom in/out can be performed by tilting the *Foot-Rocker* forward/backward: The current gaze position indicates the zoom target while the tilt angle defines the zooming speed. Panning is performed in the same way as in the first setup using a combination of gaze-directed movement accelerated with the single pedal.

4.3.3 Foot-Rocker & Gaze + Foot-Joystick (FRG+FJ)

Analogue to the second setup, the *Foot-Rocker* is used for zooming towards/away from looked-at content. For panning, we use the *Foot-Joystick*: Pushing the Foot-Joystick to a certain direction results in pan movements in the same direction. The stronger a user tilts the Foot-Joystick, the faster is the panning speed. Thus, in this setup, panning direction and speed are both controlled without additional gaze input.

5. USER STUDY

As a first evaluation step in our iterative design process, we gathered early user feedback. For this, we presented the described prototypes in a hands-on-demonstration at CHI 2013 (Interactivity and workshop) and at several public science events at our university. Visitors could test the gaze-supported foot interaction prototypes and freely explore the virtual world presented in Google Earth without a particular task [11]. To further improve the designs, participants were asked to rate the interaction regarding six usability statements on a 5-point Likert scale (1 - strong disagreement; 5 - strong agreement): Convenience, ease of pan, ease of zoom and ease of navigation, as well as feasibility of additional hand input and the ability to achieve intended goals. We also collected supplementary user comments.

54 people (8 females, 46 males; including 42 HCI experts) provided feedback after testing either **2PG+1PG** or **FRG+FJ**, whereby 20 users tested both. The **FRG+1PG** setup was not tested, as the setup rearrangement was too time-consuming at the public demonstrations.

The overall results from the early user feedback affirm the promising potential for our proposed gaze-supported foot navigation techniques. The participants could easily understand the underlying concepts and were able to use them without the need of additional practice. However, the asymmetrical zooming with two individual pedals (one for zoom in, one for zoom out) was criticized for **2PG+1PG**. Instead users preferred the *Foot-Rocker* for zooming, allowing to tilt the single pedal forward and backward. Therefore, we decided to refrain from the **2PG+1PG** setup for further investigations and to rather concentrate on further enhancing and evaluating our own custom-made foot input prototypes, as both setups **2PG+1PG** and **FRG+1PG** follow the same principles. In the following, we describe a second, more structured user study to investigate **FRG+1PG** and **FRG+FJ** in more depth and compare them to mouse-based navigation (**M**).

5.1 Participants

Twelve unpaid participants (6 female) between 21 and 29 ($M = 25$, $SD = 2.52$) with normal or corrected-to-normal vision volunteered in our study. Participants were interested in the topic and were recruited via a mailing list, 10 of them were computer science students.

5.2 Apparatus

We used a Tobii TX300 binocular eye tracker (at 120 Hz, accuracy of 0.4° under ideal conditions) with an integrated 23" LCD screen (1920x1080 resolution) and a standard office mouse in MS Windows 7 with default cursor speed. To stabilize the gaze cursor position and thus to address jittery gaze data, we used the speed reduction technique [27]. The *Foot-Rocker* and *Foot-Joystick* were installed on the left and right side at floor level in front of the participant's chair. For this, the foot devices were integrated in a floor panel on which the chair was standing (see Figure 1, right). The standard office desk was lifted the same way, maintaining a common table height of approximately 72 cm. The single pedal device was screw-mounted on the platform, 8 cm right from the *Foot-Joystick* device at the platform's far end.

The basic task in this study was to navigate to certain locations in Google Earth and to click on predefined markers (see Figure 4). For this, we integrated the Google Earth Plugin in our software to test the three navigation conditions *Foot Rocker Gaze + One Pedal Gaze* (**FRG+1PG**), *Foot Rocker Gaze + Foot Joystick* (**FRG+FJ**) and *mouse-only* (**M**). Markers were set as map overlays using the Google Earth Plugin API (see Figure 4). In the **M** condition, the default Google Earth pan and zoom methods were used: The user could directly zoom towards the current mouse cursor position (*pivot zoom*) by scrolling the mouse wheel. Panning was realized by clicking the left mouse button and moving the mouse to drag or flick the view (*flick gesture*). Use of the right button was not allowed to prevent unwanted rotation movements, which are not possible with the *gaze-supported foot navigation* (**GF**) setups.

5.3 Tasks

The study includes seven tasks, which participants were asked to carry out (see Table 1). These are divided into three types: *pointing* (requiring only zoom), *combined* (intended for pan and zoom) and *tracking* (basically intended for pan). Each required the participants to navigate to several locations on the geographic map and click on graphical markers. In all three conditions, marker selection was performed with the mouse, but the navigation technique varied. The markers were always visible but only selectable below a predefined altitude, indicated by switching marker colors as shown in Figure 4. The activation altitude varied to enforce certain amounts of pan and zoom actions. For instance, the simple *pointing tasks* started from above the Atlantic Ocean and mainly required zooming in and activating a single marker, while *combined tasks* required several pan and zoom steps, and *tracking tasks* demanded longer pan movements. For *combined* and *tracking* tasks, several marker activations were necessary for which the number and placement of markers varied (see Table 1) and for which the start position was above the target area.

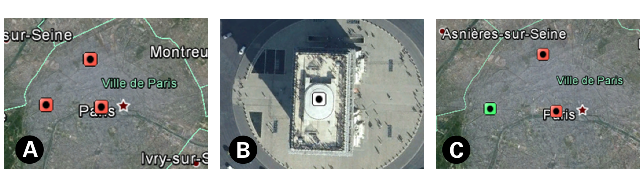

Figure 4: The three different marker color states: a) Not clickable (red), b) clickable (white), c) activated (green).

Type	Task Description
Pointing *(Zoom)*	Navigate to **New York** and select one marker
	Navigate to the city center of **Berlin** and select one marker
Combined *(Zoom + Pan)*	Navigate to 3 different sights in **Paris** and mark them
	Follow the curvy course of a designated **river** by clicking 5 markers
	Mark the border of **Utah** by clicking 6 markers
Tracking *(Pan)*	Follow the border between the US and **Mexico** by clicking 7 markers
	Follow the railroad track between the North Sea island **Sylt** and the German mainland by the help of 12 markers

Table 1: Overview of the seven tasks participants had to carry out.

5.4 Procedure

Each participant was asked to test all three input conditions (within-subject design). The order for testing the conditions was counterbalanced across participants. For each condition the same procedure was followed: In the **GF** conditions, a nine-point eye tracking calibration was performed at the beginning. First, the participants could acquaint themselves with the respective input condition in a free training until they felt sufficiently prepared (on average not more than 3 minutes). Subsequently, they were asked to achieve the described study tasks in the order listed in Table 1 which relates to an increasing complexity (i.e., starting with simple *pointing* and finishing with complex *tracking* tasks). For this, in the **GF** conditions, gaze and feet are used as supporting input modalities in addition to manual mouse input.

A short instruction was presented to the participant before each task on the screen. Each trial concluded with a NASA-TLX usability questionnaire. After testing all three setups, participants answered a final post-study questionnaire. The overall procedure took on average 45 minutes per participant.

5.5 Measures

As performance measure, we gathered task completion times (TCT). The time measurement started with the participant confirming the task instruction on the screen and stopped when the last marker was activated. In addition, user input data was logged, such as mouse clicks, as well as gaze and foot data to later analyze users' interaction strategies. A Raw TLX [13] was conducted after finishing all tasks with a given input condition to rate the perceived task load. Participants also rated their level of visual fatigue, concentration and motivation on 5-point Likert scales before and after testing each of the three input conditions. Finally, the participants were asked to assess several usability aspects in a post-study questionnaire: ease of navigation, ease of panning, ease of zooming, perceived interaction speed, accuracy of task performance and convenience of interaction. The rating was based on a 5-point scale with contrasting input conditions on each end (e.g., 1 - **GF** to 5 - **M**, as shown in Figure 7). This was intended for a relative comparison of **GF** with **M** as well as the two **GF** techniques among each other. In the final questionnaire, participants were also asked for their favorite setup, for the influence of the prototypical character of the setups and whether they can imagine to use gaze-supported foot interaction in everyday work in the context of zoomable information spaces or in different application contexts.

6. RESULTS

We did not expect our novel interaction techniques to outperform conventional mouse-only interaction as it is in place for decades and well established. However, we expected Gaze-supported Foot Navigation to be comparable with traditional interaction for primary task work and to improve user experience and usability due to a more natural interaction for secondary task navigation.

6.1 Task Completion Time

For all input conditions, the mean task completion times ($TCTs$) increased for more complex tasks (see Figure 5). We applied a non-parametric Friedman test and post-hoc analysis (Bonferroni corrected). For the pointing tasks, that mainly require zoom, $TCTs$ in both **GF** conditions were on average shorter by 14% than in the **M** condition (not statistically significant though). An increasing amount of panning actions involved in the *combined* and *tracking tasks* resulted in longer $TCTs$ using **GF** navigation compared to **M**. From the three *combined tasks*, significant differences only occurred for the *river* task. For both *tracking tasks*, we detected significant differences between all conditions, whereby users were fastest with **M**. Further examination of logged input data showed no evidence that this is related to switching between mouse and foot devices, as there was only little discontinuity between mouse clicking and navigation actions. Instead the differences arise from the different panning speeds used in the three input conditions. While **M** allowed fast flicking gestures, users were more careful to pan with their gaze and feet, thus leading to slower panning and longer $TCTs$.

6.2 Pan and Zoom Strategies

We used input log data to identify different pan and zoom strategies and their potential impact on task completion times. In both **GF** and **M** conditions, all users sequentially performed navigation actions and mouse clicks. Interestingly, participants constantly moved the mouse in both **GF** setups during navigation actions: Users were hovering over the target positions or followed the panning path by dragging the mouse cursor along it. Participants used panning more often in the **FRG+1PG** setup compared to **FRG+FJ**. In the latter condition several participants changed their strategy after a while and rather preferred to use gaze-directed zooming instead of panning with the *Foot-Joystick*. In fact, we noticed that participants took longer to pan using gaze and feet than using gaze-directed zoom. For the *tracking tasks*, this led to changing performances between both **GF** setups. The change of strategy had a bigger influence on the $TCTs$ in the longer Sylt task.

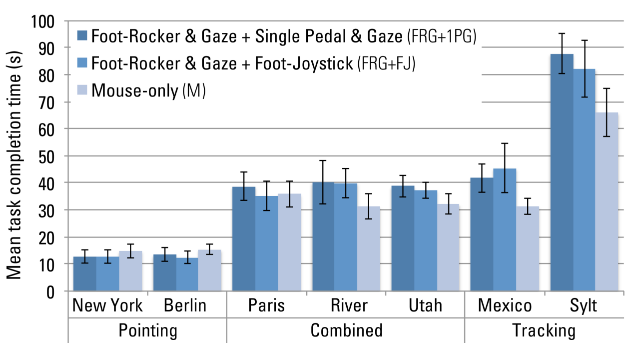

Figure 5: Mean task completion times in seconds with 95% confidence interval bars.

6.3 Workload Rating

As shown in Figure 6, mean TLX ratings did not show large differences between the three interaction techniques although the **M** condition receives lower (better) ratings in most categories. Especially for aspects such as *mental demand* this is not surprising due to the high familiarity with mouse input and the novelty as well as higher complexity of multimodal interaction (manual input + gaze and feet input vs. manual input) or the prototypical character of the **GF** setups. Thus, *physical demand* and *effort* were similarly perceived higher for the novel setups. While the user assessments are least diverged with regard to *performance* for **FRG+1PG**, participants were still more satisfied with their performance using the mouse. *Frustration* with mouse input was perceived similarly low as with **FRG+1PG**. **FRG+FJ** reached much higher ratings, however they were still good. With a deeper analysis of the *frustration scores* for **FRG+FJ** ($M = 8.17, SD = 5.64$), two groups can be distinguished: 7 participants were little frustrated ($M = 3.85$), while the other 5 participants seemed to have difficulties operating the *Foot-Joystick* ($M = 14.2$). In comparison, **M** and **FRG+1PG** did not show such differences: $M_M = 4.75, SD_M = 3.65$ and $M_{FRG+1PG} = 5.50, SD_{FRG+1PG} = 3.09$. Additionally, the evaluation of potential visual fatigue, concentration and motivation showed no effects.

6.4 Quantitative Comparison

After testing all three input conditions, participants were asked to assess several usability aspects (e.g., perceived interaction speed, precision and convenience) on a preference scale ranging between two given input techniques as shown in Figure 7. On the one hand, the two gaze-supported foot navigation techniques were compared to each other (see Figure 7, right scale). On the other hand, gaze-supported foot input in general was compared to mouse-only navigation (see Figure 7, left scale).

For the comparison between gaze-supported foot and mouse input, participants showed an overall preference toward the novel **GF** interaction techniques (cf. Figure 7), in particular with respect to *Ease of Zoom* and *Convenience*. In fact, 11 participants (91.67%) preferred **GF** zooming (7 of them strongly preferred it), while only one participant preferred mouse zooming. Nine participants (75%) stated that **GF** navigation was more *convenient* than mouse interaction. Only two participants (16,7%) indicated to (slightly) prefer **M** in terms of convenience mainly due to their familiarity with mouse input and due to difficulties with panning using the *Foot-Joystick*, which was found cumbersome by several participants. With regard to *Ease of Navigation*, 8 participants (66,67%) tended towards **GF** (2 with a strong preference). Eight users (66,67%) felt being faster with **GF** navigation, which interestingly contradicts the task

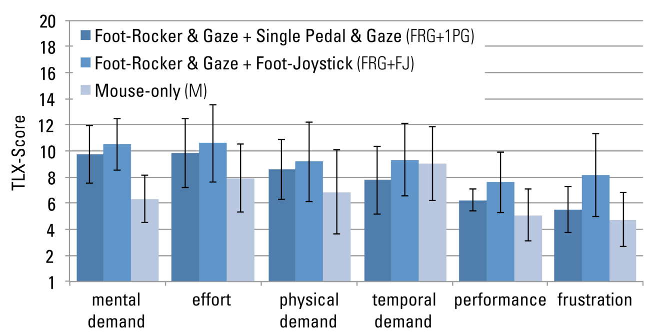

Figure 6: Results from the NASA-TLX questionnaires with 95% confidence interval bars (the lower, the better).

completion time measurements in many cases. The lowest rated usability aspect for **GF** is *accuracy*, for which users interestingly still tended towards **GF** compared to **M**. While gaze- and foot input is in general considered less precise than mouse input, this may indicate that input precision was nevertheless appropriate as pan and zoom tasks do not require high precision after all.

Comparing the two **GF** techniques, participants tended towards **FRG+1PG** for all questioned usability aspects (see Figure 7) which was probably mainly due to the circumstance that panning with the *Foot-Joystick* was sometimes described as cumbersome. While gaze-based panning (for **FRG+1PG**) was also not ideal, as it was described as exhausting to control the pan direction using the eyes over time, it was assessed easier than the foot-based panning: 10 participants (83.3%) preferred **FRG+1PG**. Even though the zoom technique was the same (*Foot-Rocker* & Gaze), four participants (33.3%) rated zooming easier in the **FRG+1PG** setup.

6.5 Post-Study User Feedback

Ten participants (83.33%) named one of the **GF** techniques as their favorite setup in the final questionnaire. Six preferred **FRG+1PG**, 4 preferred **FRG+FJ**, and 2 preferred **M**. These results are somewhat surprising, as 75% of participants also stated that their rating was negatively influenced by the prototypical character of the setups. Nevertheless, *Gaze-supported Foot Navigation* shows promising potential for the future: 10 participants can imagine to use **GF** interaction for exploring zoomable information spaces and even 11 participants (91.6%) would use it in different application contexts, e.g., for computer-aided design (CAD).

7. DISCUSSION

Based on our careful design and evaluation, we could highlight the promising potential of gaze-supported foot input for steering in zoomable information spaces. Although our tested **GF** setups still leave room for further improvements (e.g., more fluent control of the *Foot-Joystick*), participants were able to beat mouse input for tasks which primarily required zoom. However, this result is not statistically significant. It nevertheless underlines the interesting possibility to augment conventional mouse & keyboard input with gaze-supported foot controls. The particular advantage of our approach is to decouple the movement control (for pan and zoom) by allowing direction control via continuous gaze input and speed control via continuous rate-based foot input. Several improvements were suggested to further enhance our proposed setups, such as the *Foot-Joystick's* stiffness and problematic return to its neutral position. The *Foot-Rocker* could also be further improved for a more ergonomic design, integrating more sophisticated and robust

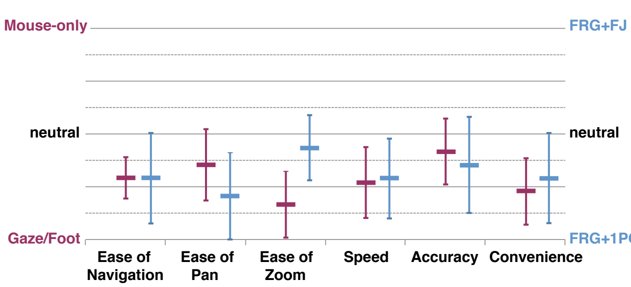

Figure 7: Two diagrams shown in one: User preference towards *Mouse (**M**) vs. Gaze+Foot (**GF**)* (magenta) and Foot-Rocker & Gaze + One Pedal & Gaze ***(FRG+1PG)*** vs. Foot-Rocker & Gaze + Foot-Joystick ***(FRG+FJ)*** (blue) with 95% confidence interval bars.

metal pedals and more persistent springs that reliably move the two-directional pedal (*Foot-Rocker*) back to its neutral position. In the following, we discuss in more detail how well gaze and foot controls are suited for performing pan and zoom actions.

7.1 Zooming

While participants have already been very familiar with mouse-based zoom, most participants preferred the gaze-supported foot zooming. The clear tendency toward this novel way of interaction highlights its suitability for this particular interaction task and encourages further investigations.

The main benefits of this novel input are that the user can smoothly zoom in towards currently looked-at content (*gaze-based zoom direction*) and can continuously and gradually adapt the zooming speed using the foot pedal (*foot-based zoom speed*). Our preliminary study indicated that the one-directional foot pedal is less suited for convenient and quick back and forth zooming, as the user would need to frequently switch pedals or would require an additional mode switch to change between forward and backward movement. The two-directional pedal with it's symmetrical zoom mapping offered a fast and reliable way to address this issue and was the participants clear favorite. The user could remain with his/her foot on the pedal and perform continuous forward or backward tilting to zoom in and out.

7.2 Panning

While both gaze- and foot-based panning were positively assessed, and all participants could achieve the given tasks with the respective input conditions, the user feedback indicates that both input techniques have not yet been ideal for panning. In line with findings from Stellmach and Dachselt [23], gaze-directed panning works well for small pan actions (e.g., for following the curvy course of a river or country border), but can quickly become distracting and tiring for fast or long pan movements due to the *double role* of our eye gaze (for active control and passive observation). The foot-based panning with the *Foot-Joystick* was aimed at addressing this issue. However, users were very careful and hesitant when using their feet for panning or they preferred to zoom out and then back in towards an intended target by looking at it instead. We mainly attribute these findings to the limited technological maturity of our *Foot-Joystick* prototype and the interaction technique's novelty and subsequent unfamiliarity to users. Thus, further improvements and investigations of the *Foot-Joystick* are required to better assess its potential for smooth foot-based pan controls.

7.3 Future Prospects

While we have concentrated on a fundamental investigation of gaze-supported foot controls on supporting navigation tasks in a stationary desktop setup, we see a high potential for further studies of this novel input combination. Since our results yet show users like the new technology – although the mouse is as fast or even faster – it is interesting to investigate how further practice might improve task completion times and workload ratings. Moreover, future work could take other application contexts and domains into account. On the one hand, further interaction tasks could be considered, such as convenient gaze- and foot-supported scrolling of documents or quick object selections. For example, users could quickly switch between chat windows by looking at a window and tapping a foot while leaving the hands on the keyboard for typing. On the other hand, different user contexts could be considered that require a higher mobility, such as the flexible interaction with wall-sized displays.

8. CONCLUSION

In this paper, we have contributed to a better understanding of how gaze and foot input can be conveniently combined to steer in zoomable information spaces, leaving the hands free for other interaction tasks (e.g. precise object selection and manipulation or fast text entry). We initially investigated the design space, carefully elaborated three prototypes and contributed an evaluation in a real-world context.

For our investigation, we have focused on a computer desktop workplace with stationary foot pedal constructions beneath it and an integrated table-mounted eye tracker. We have carefully developed novel ways for convenient and fluent gaze- and foot-supported pan and zoom controls benefiting from implicit gaze input with explicit foot controls. We have designed and examined one-, two- and multi-directional foot pedals. Our main intention was to augment conventional mouse & keyboard interaction by managing hands-free pan and zoom operations only using gaze and feet.

Based on our evaluations, we could successfully demonstrate that gaze-supported foot navigation has not only the potential to beat familiar mouse input (in certain conditions it already did and further improvements have been suggested), but a majority of participants already preferred gaze-supported foot navigation compared to mouse-only input. The positive results for our proposed techniques, in particular the symmetrical pedal combined with gaze-directed zoom, encourage further investigations in this area considering additional interaction tasks and user contexts. While our investigations have been focused on spatial navigation in GIS applications, gaze-supported foot controls could, for example, also be used for document scrolling or for quick target selections merely by looking at a target and stepping on a foot pedal.

9. REFERENCES

[1] N. Adams, M. Witkowski, and R. Spence. The inspection of very large images by eye-gaze control. In *Proc. of AVI '08*, pages 111–118. ACM, 2008.

[2] J. Alexander, T. Han, W. Judd, P. Irani, and S. Subramanian. Putting your best foot forward: investigating real-world mappings for foot-based gestures. In *Proc. of CHI '12*, pages 1229–1238. ACM, 2012.

[3] M. Ashmore, A. T. Duchowski, and G. Shoemaker. Efficient eye pointing with a fisheye lens. In *Proc. of GI '05*, pages 203–210, 2005.

[4] R. Bates, H. Istance, M. Donegan, and L. Oosthuizen. Fly where you look: Enhancing gaze based interaction in 3D environments. In *Proc. of COGAIN '05*, pages 30–32, 2005.

[5] S. Beckhaus, K. Blom, and M. Haringer. Intuitive, hands-free travel interfaces for virtual environments. In *Proc. of IEEE VR '05, Workshop "New directions in 3D User Interfaces"*, pages 57–60. Shaker Verlag, 3 2005.

[6] A. Crossan, S. Brewster, and A. Ng. Foot tapping for mobile interaction. In *Proc. of BCS '10*, pages 418–422. British Computer Society, 2010.

[7] F. Daiber, J. Schöning, and A. Krüger. Whole body interaction with geospatial data. In *Smart Graphics*, volume 5531 of *Lect. Notes in CS*, pages 81–92. Springer, 2009.

[8] R. Darken, W. Cockayne, and D. Carmein. The omni-directional treadmill: a locomotion device for virtual worlds. In *Proc. of UIST '97*, pages 213–221. ACM, 1997.

[9] G. de Haan, E. J. Griffith, and F. H. Post. Using the wii balance board as a low-cost vr interaction device. In *Proc. of VRST '08*, pages 289–290. ACM, 2008.

[10] G. Furnas and B. Bederson. Space-scale diagrams: Understanding multiscale interfaces. In *Proc. of CHI '95*, pages 234–241. ACM Press, 1995.

[11] F. Göbel, K. Klamka, A. Siegel, S. Vogt, S. Stellmach, and R. Dachselt. Gaze-supported foot interaction in zoomable information spaces. In *Proc. of CHI EA '13*, pages 3059–3062, New York, NY, USA, 2013. ACM.

[12] D. W. Hansen, H. H. T. Skovsgaard, J. P. Hansen, and E. Møllenbach. Noise tolerant selection by gaze-controlled pan and zoom in 3d. In *Proc. of ETRA '08*, pages 205–212. ACM, 2008.

[13] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proc. of the HFES Annual Meeting*, volume 50, pages 904–908. Sage Publications, 2006.

[14] E. R. Hoffmann. A comparison of hand and foot movement times. *Ergonomics*, 34(4):397–406, 1991.

[15] R. Jacob. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM TOIS*, 9(2):152–169, 1991.

[16] C. Lankford. Effective eye-gaze input into windows. In *Proc. ETRA '00*, pages 23–27. ACM, 2000.

[17] J. J. LaViola, Jr., D. A. Feliz, D. F. Keefe, and R. C. Zeleznik. Hands-free multi-scale navigation in virtual environments. In *Proc. of I3D '01*, pages 9–15. ACM, 2001.

[18] T. Pakkanen and R. Raisamo. Appropriateness of foot interaction for non-accurate spatial tasks. In *Proc. of CHI EA '04*, pages 1123–1126. ACM, 2004.

[19] G. Pearson and M. Weiser. Of moles and men: the design of foot controls for workstations. *SIGCHI Bull.*, 17(4):333–339, Apr. 1986.

[20] J. Schöning, F. Daiber, A. Krüger, and M. Rohs. Using hands and feet to navigate and manipulate spatial data. In *CHI EA '09*, pages 4663–4668. ACM, 2009.

[21] J. Scott, D. Dearman, K. Yatani, and K. N. Truong. Sensing foot gestures from the pocket. In *Proc. of UIST '10*, pages 199–208. ACM, 2010.

[22] M. Slater, A. Steed, and M. Usoh. The virtual treadmill: A naturalistic metaphor for navigation in immersive virtual environments. In *Virtual Environments '95*, Eurographics, pages 135–148. Springer Vienna, 1995.

[23] S. Stellmach and R. Dachselt. Investigating gaze-supported multimodal pan and zoom. In *Proc. of ETRA '12*, pages 357–360. ACM, 2012.

[24] S. Stellmach and R. Dachselt. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In *Proc. of CHI '13*, pages 285–294. ACM, 2013.

[25] S. Stellmach, S. Stober, A. Nürnberger, and R. Dachselt. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proc. of NGCA '11*, pages 1–8. ACM, 2011.

[26] C. Ware and H. H. Mikaelian. An evaluation of an eye tracker as a device for computer input. In *Proc. of SIGCHI+GI '87*, pages 183–188, 4 1987.

[27] X. Zhang, X. Ren, and H. Zha. Improving eye cursor's stability for eye pointing tasks. In *Proc. of CHI '08*, pages 525–534. ACM, 2008.

[28] D. Zhu, T. Gedeon, and K. Taylor. Moving to the centre: A gaze-driven remote camera control for teleoperation. *Interact. Comput.*, 23:85–95, 1 2011.