Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays

Florian Müller
TU Darmstadt
Darmstadt, Germany
mueller@tk.tu-darmstadt.de

Joshua McManus
Simon Fraser University
Burnaby, British Columbia, Canada
josh@joshuamcmanus.com

Sebastian Günther
TU Darmstadt
Darmstadt, Germany
guenther@tk.tu-darmstadt.de

Martin Schmitz
TU Darmstadt
Darmstadt, Germany
schmitz@tk.tu-darmstadt.de

Max Mühlhäuser
TU Darmstadt
Darmstadt, Germany
max@informatik.tu-darmstadt.de

Markus Funk
TU Darmstadt
Darmstadt, Germany
funk@tk.tu-darmstadt.de

ABSTRACT

From voice commands and air taps to touch gestures on frames: Various techniques for interacting with head-mounted displays (HMDs) have been proposed. While these techniques have both benefits and drawbacks dependent on the current situation of the user, research on interacting with HMDs has not concluded yet. In this paper, we add to the body of research on interacting with HMDs by exploring foot-tapping as an input modality. Through two controlled experiments with a total of 36 participants, we first explore *direct* interaction with interfaces that are displayed on the floor and require the user to look down to interact. Secondly, we investigate *indirect* interaction with interfaces that, although operated by the user's feet, are always visible as they are floating in front of the user. Based on the results of the two experiments, we provide design recommendations for *direct* and *indirect* foot-based user interfaces.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Mixed / augmented reality**; *Interaction techniques*;

KEYWORDS

Human Factors; Foot Interaction; HMD; User Study

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300707

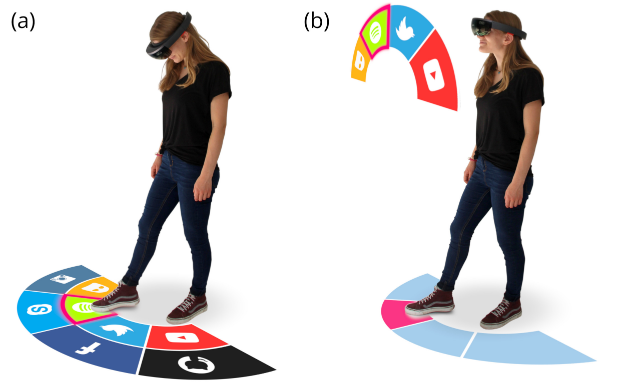


Figure 1: We propose foot-taps as a *direct* (a) and *indirect* (b) input modality for interacting with HMDs.

ACM Reference Format:
Florian Müller, Joshua McManus, Sebastian Günther, Martin Schmitz, Max Mühlhäuser, and Markus Funk. 2019. Mind the Tap: Assessing Foot-Taps for Interacting with Head-Mounted Displays. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300707

1 INTRODUCTION

With increasing progress in display and tracking technology [6], head-mounted displays (HMDs) became untethered [19] and capable of creating realistic Augmented Reality (AR) experiences that can be registered in the physical world. While voice and gesture input are still considered state-of-the-art for interacting with HMDs, they are known to be limited to quiet environments and having social implications [37, 61], or being prone to fatigue [32], respectively.

To overcome these challenges, research has explored many forms of interaction with HMDs: From *direct* interaction such as using hand [10] and finger [46] gestures to *indirect* interaction with body-worn accessories such as belts [14], shirts [54], trousers [15], or rings [2]. While many body parts - from arms [29] and hands [47] to the head [57, 69] - have already been explored, foot-based interaction has not yet been systematically evaluated for interaction with HMDs.

In this paper, we aim to close this gap and add to the body of research on interacting with HMDs by exploring foot-taps as an input modality for HMDs. The contribution of this paper is two-fold: First, we contribute the results of two controlled experiments, assessing the benefits and drawbacks of 1) *direct interaction* with interfaces that are displayed on the floor and require the user to look down to interact and 2) *indirect interaction* with interfaces that, although operated by the user's feet, are displayed as a floating window in front of the user (see fig. 1). Second, based on the results of the two experiments, we provide a set of guidelines for designing the input space for both types of interaction.

2 RELATED WORK

There exists a large body of related work on interacting with HMDs as well as foot-based interaction. Further, our work was strongly inspired by proprioceptive and imaginary user interfaces, which we discuss at the end of the section.

Interaction with HMDs

Research proposed various new approaches and improvements to established techniques for interaction with HMDs.

Focussing on gesture interfaces, Mistry and Maes [46] proposed a wearable interface supporting natural gestures. Continuing this path, Colaço et al. [10] presented a system for capturing and interpreting more elaborate single-handed gestures. Other examples include proximity-based interfaces [48], finger-gestures [7], the use of a haptic glove [33] or combinations with other modalities such as gaze [31, 60] or head-movements [38]. While such techniques have several benefits (e.g., the *direct* manipulation of content), gestures are prone to fatigue, also known as the *gorilla arm syndrome* [32], and, thus, not suited for long-lasting interactions.

Research proposed on-body [28, 65] interfaces to interact with HMDs by touching various body-parts: Beyond arm [29, 66] and hand [13, 47, 62], Serrano et al. [57] proposed hand-to-face interfaces. Other examples for such touch interfaces on the head include the cheek [69] or the ear [39]. While practical and useful, these techniques require at least one, often both hands of the user and, thus, hardly support situations where users are encumbered.

Further, research proposed the use of additional accessories beyond the hand-held devices used with today's HMDs. For example, Dobbelstein et al. [14] proposed an interactive belt for unobtrusive touch input and Ashbrook et al. [2] presented an interactive ring. Further, research proposed to augment the user's pocket [15] or sleeves [54]. However, such accessory interfaces are missing means for *direct* manipulation and may be misplaced or lost.

Foot-based Interaction

In our work, we were strongly inspired by the large body of research on foot-based interaction techniques [64] that have a long history in operating industry machines [4, 5, 11, 36, 50] and have been explored for seated [63], standing [52] and walking [68] users in different scenarios.

Research proposed multiple use cases for such foot-based input modalities. Yin and Pai [70] presented an interactive animation system, controlled using foot gestures. Simeone et al. [59] used foot-based input for 3D interaction tasks, Schöning et al. [55] presented support for navigating spatial data. Further examples include support for the interaction with large displays [20, 35], interactive floors [3] and other public interfaces [22]. More general, Alexander et al. [1] and Felberbaum and Lanir [21] proposed user-defined foot-gestures for typical GUI tasks in different domains.

Further, foot-controls have been used to increase the input space for desktop [58] or mobile [42] games or to operate a smartphone in the pocket of the user [5, 18, 27]. Besides the sole use as an input modality, foot interaction has been used in conjunction with hand-gestures [41, 43, 44] or gaze-input [24, 51]. Pakkanen and Raisamo [49] investigated foot-based interaction as a second input channel for non-accurate spatial tasks and found that foot interaction is appropriate, "maintaining adequate accuracy and execution time". Highly related, Saunders and Vogel [52] explored indirect interaction with ring-shaped foot interfaces. However, the exploration by Saunders et al. was limited to 1) indirect interfaces and 2) two different layouts.

Research also focused on the applicability of foot-based interfaces for HMDs. Matthies et al. [45] presented a technical prototype to provide hands-free interaction for VR applications. Fukahori et al. [23] used the shifting of the user's weight on their foot for subtle gestures to control HMDs interfaces. Furthermore, Fan et al. [17] focused on foot-based interaction techniques for exploring a VR representation of a planet. Highly related, Lv et al. [44] used foot-based interaction techniques for controlling an AR game. However, to the best of our knowledge, there is no systematic investigation of the human ability to interact with HMDs through foot-taps.

Imaginary and Proprioceptive User Interfaces

Gustafson et al. [25, 26] introduced imaginary interfaces as a novel approach to interaction without any visual feedback, leveraging the human's ability to map the spatial memory to (physical) surfaces. Dezfuli et al. [13] extended this idea using proprioception [40, 53], the subconscious knowledge about the relative position and orientation of our body parts, showing that users were able to create a mental mapping between on-screen user interfaces and eyes-free touch on the hand. In this work, we extend these ideas for foot control.

3 METHODOLOGY

We conducted two controlled experiments to assess the accuracy and efficiency of *direct* and *indirect* interfaces for foot-based interactions with HMDs. More specifically, we investigated the following research questions:

- RQ1 How accurately and efficiently can users interact with *direct* interfaces where targets are visualized on the floor within reach of their legs?
- RQ2 How accurately and efficiently can users interact with *indirect* interfaces where the visualization is shown as a floating window in front of the user, thus dividing the location of input and output?
- RQ3 How to design the targets for *direct* and *indirect* interfaces regarding size and convenient boundaries to attain high accuracy and efficiency?

To avoid learning effects, we addressed RQ1 and RQ2 in two separate experiments, although the basic design and procedure show large overlaps. To keep the results of both experiments comparable, we only changed the visualization technique between *direct* and *indirect* interfaces between the experiments. No participant took part in both experiments. In the analysis, we used the results of both experiments to address RQ3. The following description of the methodology applies to both experiments unless stated otherwise.

Design and Task

We defined a semicircular interaction grid that is anchored to the user's standing position (see fig. 2). We varied the number of *rows* and *columns* that divide the grid into several targets as independent variables in a repeated measures design. For the independent variables, we used three levels for the number of rows (1,2,3) and three levels for the number of columns (2,4,6). Therefore, we tested grids from $1 * 2 = 2$ to $3 * 6 = 18$ targets. We considered these variables to assess their impact on participants' performance regarding accuracy and efficiency. We opted for at least three repetitions of each target (i.e., based on the most complex condition 3-row, 6-column: $3*6*3 = 54$). To prevent the influence of fatigue, we designed the experiment with an equal number of trials in each condition. This resulted in a total of $3*3*54 = 486$ trials per participant. We counterbalanced the order of conditions using a Balanced Latin Square design. For each condition, the series of targets was randomized while maintaining an equal number for each target.

Experiment I: Direct Visualization. We visualized the semicircular grid within leg reach on the floor in front of the participant. Therefore, there was a *direct* connection between the location of input and output. Depending on the condition, we divided the semicircle to a grid with 2-6 horizontal columns of equal size and 1-3 rows of equal size. We divided the columns over the complete semicircle (see fig. 3c). Based on the average human leg length [16], we used a fixed height of 8.5 cm for each row. We chose this size to allow all participants to reach the goals within the 3-row conditions comfortably. The participants' task was to look at the floor in front of them and to tap highlighted targets.

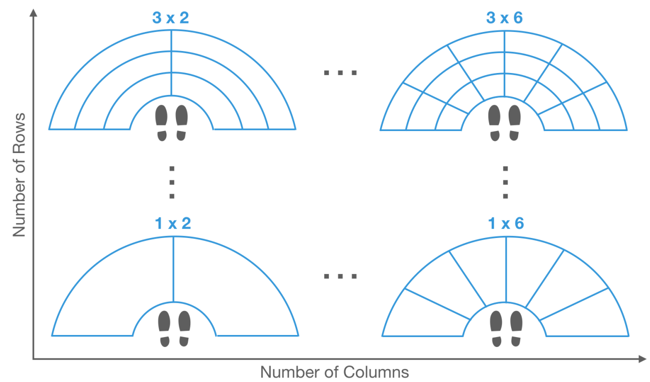


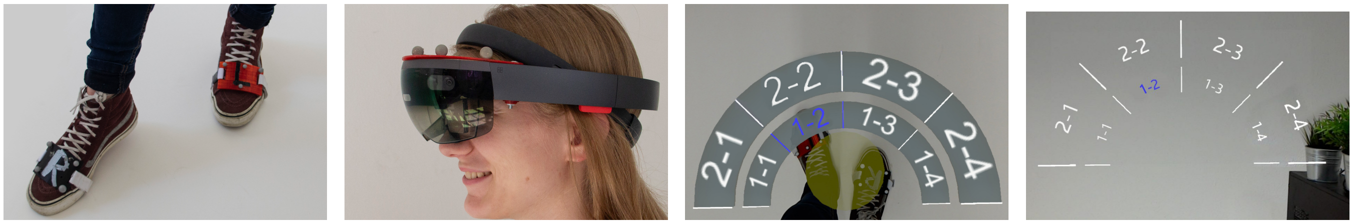
Figure 2: The independent variables (number of rows and number of columns) tested in the two experiments.

Experiment II: Indirect Visualization. For the second experiment, we chose an *indirect* head-up display (HUD) visualization, floating in front of the eyes of the user (see fig. 3d). Our goal was to understand how the participants would naturally map the presented target areas to the ground in front of them. Therefore, we decided not to give the participants feedback about the position of their feet. Such feedback would have given the participants an indication of the size of the target areas, thereby distorting the results. The participants' task was to tap targets highlighted in the floating visualization.

Study Setup and Apparatus

We used an optical tracking system (OptiTrack) to measure the position of the participant's feet. For this, we attached 3D-printed parts, each augmented with a set of retro-reflective markers, to both feet of the participants (see fig. 3a). Furthermore, the participants wore a Microsoft Hololens (also with retro-reflective markers, see fig. 3b) which displayed the respective visualization.

We implemented a study client application that allowed us to set the task from a desktop located next to the participant. For each trial, we logged the trace of the participants' *feet movements* and *head (HoloLens) movements* to establish a matching between the visual feedback and the foot-taps. Furthermore, we measured the time between displaying the task and touching the floor with the foot as the *task-completion time (TCT)* and logged it together with the *foot used for interaction*, the *tap position* (relative to the participant), the *target* and the *condition* for later analysis.



(a) Foot Tracking **(b) HoloLens Tracking** **(c) Direct Visualization** **(d) Indirect Visualization**

Figure 3: We tracked the position and orientation of the feet (a) and the hololens (b). During the first experiment, we used a *direct* visualization on the floor (c). In the second experiment, we used an *indirect* floating visualization (d).

Procedure

After welcoming the participants, we introduced them to the concept and the setup of the study. During this, we used the method proposed by Chapman et al. [8] to measure the foot preference. Further, we measured the height and leg length of the participants as we expected them to have an impact on the performance. Then, we mounted the trackable apparatuses on their feet and put the Hololens on their head. To avoid learning effects, we gave them five minutes to get accustomed to the hardware and the interfaces.

We calibrated the system with the participants standing relaxed and looking straight ahead. After starting the condition, the participants saw the respective visualization. Once ready and in starting position (both feet together), we started the current condition. The system then colored the current target to be reached in blue (see fig. 3c, 3d) and informed the participant about the start of the trial with an additional audio signal. Then, the participant moved the foot and tapped the floor on the target position. We did not enforce the usage of a specific foot but told the participants to use the foot that seemed most comfortable for each trial. After tapping the target, the system changed the target color to green to inform the participant that the measurement was recorded and that the participant should move the foot back to the starting position. Once reached, the system waited 2 seconds before proceeding to the next target.

We instructed the participants to focus on the accuracy (tapping the center of the target) instead of the speed. Participants did not receive any feedback regarding their performance during the study. After each condition, participants completed a NASA TLX [30] questionnaire and answered questions regarding their experiences on a 5-point Likert-scale (1: strongly disagree, 5: strongly agree). We further enforced a 5-minute break between the conditions during which we asked the participants for qualitative feedback in a semi-structured interview. Each experiment took about 60 minutes per participant.

Analysis

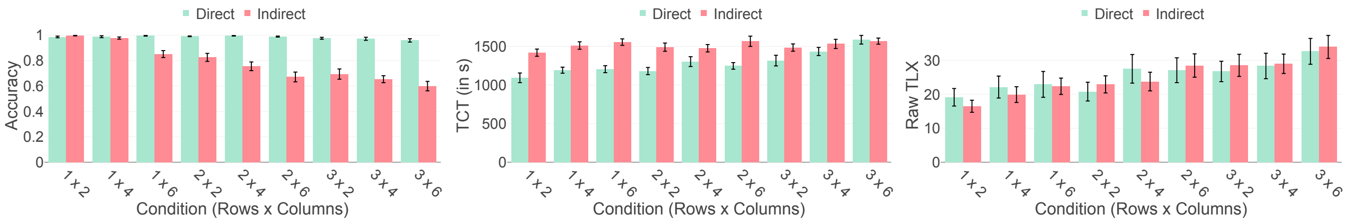
We analyzed the recorded data using two-way repeated measures ANOVA. For the Likert questionnaires, we performed an Aligned Rank Transformation as proposed by Wobbrock et al. [67]. We tested the data for normality with Shapiro-Wilk's test and found no significant deviations. Where Mauchly's test indicated a violation of the assumption of sphericity, we corrected the tests using the Greenhouse-Geisser method and report the ϵ. When significant effects were revealed, we used Bonferroni corrected pairwise t-tests for post-hoc analysis. We further report the eta-squared η^2 as an estimate of the effect size and use Cohen's suggestions to classify the effect size [9]. As an estimate of the influence of the individual factors, we report the estimated marginal mean (EMM) as proposed by Searle et al. [56]. For the analysis of the NASA TLX questionnaires, we used the raw method, indicating an overall workload as described by Hart [30].

4 EXPERIMENT I: DIRECT INTERACTION

We conducted a controlled experiment investigating RQ1 and, thus, focusing on *direct* interfaces using the visualization on the floor as described in section 3. For this, we recruited 18 participants (6 female), aged between 21 and 30 years ($\mu = 24.9$, $\sigma = 3.0$), using our University's mailing list. Three of them had prior experience with AR. We excluded 7 out of 8748 trials as outliers due to technical problems.

Accuracy

We used the physical dimensions of the targets (visible on the floor through the HMD) to classify the taps of the participants as hits and errors to obtain an accuracy rate. The analysis revealed that the number of rows had a significant ($F_{1.32,22.51} = 4.068$, $p < .05$, $\epsilon = .662$, $\eta^2 = .099$) influence on the accuracy with a small effect size. Post-hoc tests confirmed significantly higher accuracy rates for the 1-row (EMM $\mu = 98.9\,\%$, $\sigma_{\overline{x}} = 0.6\,\%$) and 2-row (EMM $\mu = 99.1\,\%$, $\sigma_{\overline{x}} = 0.6\,\%$) conditions compared to the 3-row (EMM $\mu = 96.8\,\%$, $\sigma_{\overline{x}} = 0.6\,\%$) conditions.



(a) Accuracy **(b) Task-Completion Time** **(c) Raw-TLX**

Figure 4: Accuracy, Task-Completion Time and Raw TLX in both experiments. All error bars depict the standard error.

We could not find any significant influence of the number of columns ($F_{2,34} = .515$, $p > .05$) or interaction effects between both factors ($F_{2.63,44.70} = 1.699$, $p > .05$, $\epsilon = .657$). Overall, we found high accuracy rates up to the highest (3-row, 6-column) condition ($\mu = 95.9\%$, $\sigma = .5\%$). Figure 4a (green) depicts the measured accuracy rates for all conditions.

Task Completion Time

The analysis unveiled that both, the number of rows ($F_{2,34} = 14.47$, $p < .001$, $\eta^2 = .059$) and the number columns ($F_{2,34} = 43.39$, $p < .001$, $\eta^2 = .203$) had a significant influence on the task-completion time with a medium and large effect size, respectively. We further found interaction effects between the number of rows and the number of columns ($F_{2.16,36.67} = 3.22$, $p < .05$, $\epsilon = .539$, $\eta^2 = .024$) with a medium effect size.

Post-hoc tests confirmed significantly rising TCTs for higher numbers of rows (1-row: EMM $\mu = 1.163$ s, $\sigma_{\overline{x}} = 0.046$ s, 2-row: EMM $\mu = 1.243$ s, $\sigma_{\overline{x}} = 0.046$ s, 3-row: EMM $\mu = 1.445$ s, $\sigma_{\overline{x}} = 0.046$ s) between all levels ($p < .05$ between 1-row and 2-row, $p < .001$ otherwise). For the number of columns, post-hoc tests showed significant differences between the 2-column (EMM $\mu = 1.196$ s, $\sigma_{\overline{x}} = 0.045$ s) and 6-column (EMM $\mu = 1.346$ s, $\sigma_{\overline{x}} = 0.045$ s) conditions ($p < .001$) as well as between the 4-column (EMM $\mu = 1.310$ s, $\sigma_{\overline{x}} = 0.45$ s) and the 6-column conditions. Figure 4b (green) shows the TCTs for all conditions.

Footedness and Foot Used for the Interaction

We could not find any influence of the footedness of the participants on the accuracy ($F_{1,16} = .570$, $p > .05$) nor on the TCT ($F_{1,16} = 1.42$, $p > .05$). Interestingly, although we left it up to the participants to decide which foot they wanted to use, virtually all targets to the left of the participants' line of sight were performed with the left foot and vice versa ($\mu > 96\%$ for all conditions). Matching this, we found no significant influences of the number of rows ($F_{2,32} = .408$, $p > .05$), the number of columns ($F_{1.21,19.28} = .292$, $p > .05$, $\epsilon = .603$) or the footedness ($F_{1,16} = .451$, $p > .05$) on the foot used for interaction.

Size of the Target Areas

We analyzed the influence of the target position (as target row and target column) on the spread of the recorded tapping positions. As a measurement for the spread of data, we calculated individual 95% data probability ellipses (i.e., ellipses containing 95% of the recorded points for this target) per participants and compared the areas of these data ellipses.

The analysis showed a significant influence of the target row on the area of the targets ($F_{2,34} = 13.36$, $p < .001$, $\eta^2 = .04$) with a small effect size. Post-hoc tests confirmed significantly larger areas if the target was located in 3-row (EMM $\mu = 0.0454\,\mathrm{m}^2$, $\sigma_{\overline{x}} = 0.006\,\mathrm{m}^2$) compared to 1-row (EMM $\mu = 0.005\,\mathrm{m}^2$, $\sigma_{\overline{x}} = 0.006\,\mathrm{m}^2$) and 2-row (EMM $\mu = 0.008\,\mathrm{m}^2$, $\sigma_{\overline{x}} = 0.006\,\mathrm{m}^2$), both $p < .001$. Despite rising means, we could not observe significant effects between targets in the 1-row and 2-row conditions.

We could not find a significant influence of the target column ($F_{11,187} = 1.62$, $p > .05$) nor interaction effects between the number of rows and the number of columns ($F_{22,374} = 1.62$, $p > .05$). We did not analyze the overlap between the target areas as the *direct* visualization limited the size of the target areas. Figure 5 depicts the 95% data probability ellipses for the 4-column conditions and illustrates the rising area sizes for targets in outer rows.

TLX and Questionnaire

The Raw Nasa-TLX (RTLX) questionaire showed a significant influence of the number of rows ($F_{2,34} = 16.82$, $p < .001$, $\eta^2 = .047$) with a small effect size. Post-hoc tests confirmed a significant effect for the number of rows between the 1-row (EMM $\mu = 21.4$, $\sigma_{\overline{x}} = 3.28$) and 2-row (EMM $\mu = 25.1$, $\sigma_{\overline{x}} = 3.28$) conditions ($p < .05$), the 1-row and 3-row (EMM $\mu = 29.2$, $\sigma_{\overline{x}} = 3.28$) conditions ($p < .001$) as well as between the 2-row and 3-row conditions ($p < .05$).

We further found a significant influence of the number of columns ($F_{1.34,22.85} = 6.83$, $p < .01$, $\epsilon = .672$, $\eta^2 = .023$) with a small effect size. Post-hoc tests showed significant differences between the 2-column (EMM $\mu = 22.2$, $\sigma_{\overline{x}} = 3.3$) and 4-column (EMM $\mu = 26.0$, $\sigma_{\overline{x}} = 3.3$) conditions as well

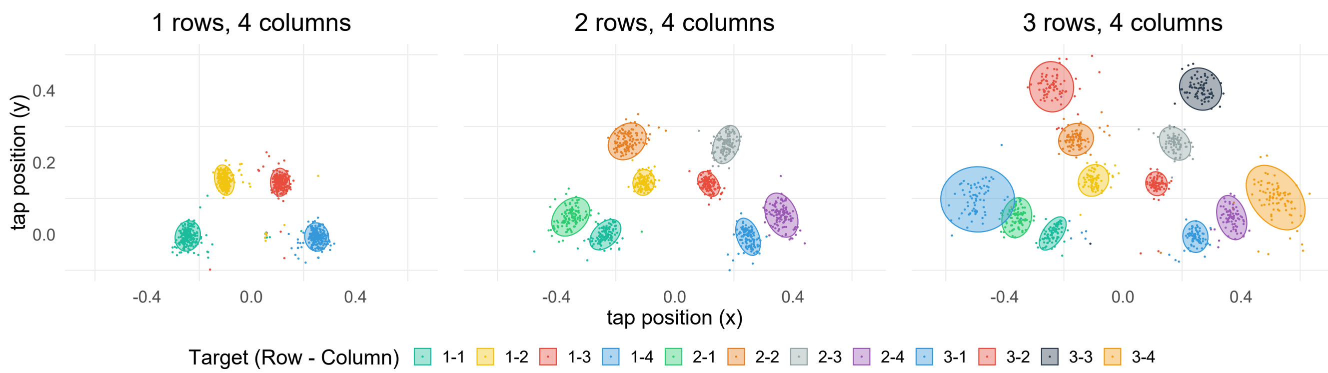


Figure 5: Scatter plots with 95% data probability ellipses for the 4-column conditions with *direct* interfaces in the first experiment. All target areas can be separated. The outer (3-row) target areas are larger than the nearer targets.

as between the 2-column and 6-column (EMM $\mu = 27.5$, $\sigma_{\overline{x}} = 3.3$) conditions. We could not find any interaction effects between the factors ($F_{4,68} = 2.28$, $p > .05$). Figure 4c (green) depicts the measured values for all conditions.

Confidence. Matching the quantitative results, the participants felt very confident that they hit the correct targets across all conditions (see fig. 6). The analysis showed a significant effect for the number of columns ($F_{2,34} = 5.259$, $p < .05$). Post-hoc tests confirmed a significantly higher confidence for 4-column conditions compared to 2-column and 6-column conditions (both $p < .05$). We could not find effects for the number of rows ($F_{2,34} = .831$, $p > .05$) but interaction effects between the two factors ($F_{4,68} = 5.057$, $p < .01$).

Convenience. We asked the participants how convenient they felt with the layout to interact with information. The analysis showed significant effects for both, the number of rows ($F_{2,34} = 23.984$, $p < .001$) as well as the number of columns ($F_{2,34} = 7.891$, $p < .01$). Post-hoc tests confirmed significantly lower ratings for the 3-row conditions compared to the other levels (both $p < .001$). Regarding the number of columns, we found a significant difference between the 2-column and 6-column conditions ($p < .01$). We could not find interaction effects ($F_{4,68} = 1.065$, $p > .05$).

A closer look at the answers supports the statistical results and, thus, the strong influence of the number of rows: All but the 3-row conditions are rated predominantly positively. Figure 6 depicts all answers from the participants.

Willingness to Use. Further, we asked the participants if they would like to use this arrangement for interacting with HMDs. The analysis showed a significant effect for both, the number of rows ($F_{2,34} = 8.938$, $p < .001$) as well as the number of columns ($F_{2,34} = 6.087$, $p < .01$). We could not find interaction effects ($F_{2,68} = 1.370$, $p > .05$). Post-hoc tests confirmed significantly lower ratings for the 3-row conditions compared to 1-row ($p < .001$) and 2-row ($p < .05$) conditions. For the number of columns, we found a significant higher rating for the 4-column conditions compared to the 6-column conditions ($p < .01$).

Again, the participants' ratings for all but the 3-row conditions were predominantly positive (see fig. 6 for all results).

Qualitative Feedback

In general, all participants appreciated the idea of foot-based interactions with HMDs because it is "*easy to use*" (P6, P8, P11, P12), and "*not tiring [compared to the standard air-tap interface of the Hololens]*" (P8, P17).

Participants commented that the limitations of the used hardware - "*weight*" (P1, P3, P4, P9), "*field of view*" (P5, P6, P7, P8, P15) - had a strong influence on their comfort because it because it forced them into an "*unnatural*" (P14) posture during the study. P17 summarized: "Looking down all the time is a bit exhausting for the neck. So I wouldn't use it for longer-lasting [interactions], but I would love this for quick and short [interactions]".

5 EXPERIMENT II: INDIRECT INTERACTION

We conducted a second experiment focussing on RQ2 and, thus, on *indirect* interfaces using the visualization in front of the participant as described in section 3. For this, we recruited 18 participants (5 female), aged between 21 and 31 years ($\mu = 23.3$, $\sigma = 2.8$), 3 left-footed, using our University's mailing list. None of them had prior experience with AR. During the analysis, we excluded 16 out of 8748 trials as outliers due to technical problems during recording.

Classification

In the first experiment, we used the physical dimensions of the targets (visible as *direct* feedback on the floor) to calculate the accuracy rates. However, we could not transfer this

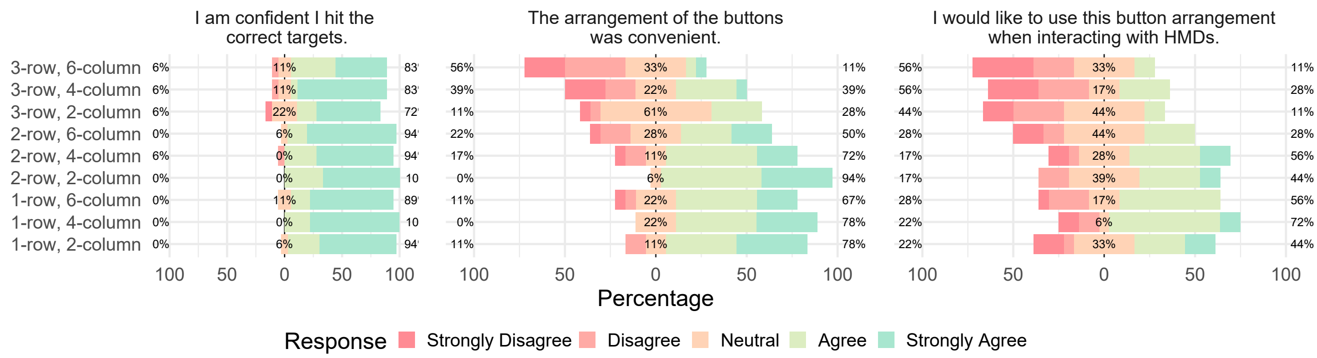


Figure 6: The participant's answers to our questions for direct interfaces on a 5-point Likert-scale.

approach directly to the second experiment, as the participants interacted with an *indirect* visualization. There was, therefore, no direct definition of the accuracy of the participants' hits and misses. As a result, we started the analysis with the construction of suitable classifiers.

We classified our data using support vector machines (SVMs) and trained nine SVM classifiers according to our nine conditions. For this, we divided each corresponding partial data set into an 80% training set and a 20% test set. We used the training sets to train per-condition SVMs with radial kernels. To avoid over-fitting to the data, we used a 10-fold cross-validation with 3 repetitions and used predictions on the 20% test sets to assess the quality of the SVMs. Furthermore, we trained per-participant SVMs and compared the results to the models we trained with the data of all participants. However, as there were only minor differences in the accuracy rates (+/- 2%, depending on the condition), we used the generalized models for further analysis.

Accuracy

We found that both independent variables, the number of rows ($F_{2,30} = 60.87$, $p < .001$, $\eta^2 = .460$) and the number of columns ($F_{2,30} = 11.61$, $p < .001$, $\eta^2 = .082$) had a significant influence on the accuracy with a large and small effect size, respectively. Post-hoc tests confirmed significantly lower accuracy rates for a higher number of rows between all groups (all $p < .001$) and between 2 and 6 ($p < .001$) as well as 4 and 6 columns ($p < .05$). We could not find any interaction effects between the variables ($F_{4,60} = 2.37$, $p > .05$).

Interestingly, a closer look revealed that, as the number of rows and columns increases, the falling accuracy is not directly dependent on the number of resulting targets: In both, the 1-row, 4-column condition as well as the 2-row, 2-column condition, the participants had to hit 4 different targets. However, we found a significantly higher accuracy rate for the 1-row, 4-column ($\mu = 98\%$, $\sigma = 1.1\%$) condition compared to the 2-row, 2-column ($\mu = 83.6\%$, $\sigma = 13.4\%$) condition ($p < .01$). We found the same effect for the 1-row, 6-column ($\mu = 85.1\%$, $\sigma = 12.3\%$) condition compared to the 3-row, 2-column ($\mu = 69.7\%$, $\sigma = 17.2\%$) condition ($p < .01$). This indicates that the number of rows has a greater influence on the accuracy than the number of columns.

Considering the EMM for the individual conditions, we found a overall high accuracy rate for the 1-row (EMM $\mu = 94.3\,\%$, $\sigma_{\overline{x}} = 3\,\%$) conditions. Figure 4a (red) depicts the measured accuracy rates for all conditions.

Task Completion Time

The analysis unveiled that the number columns of the condition had a significant ($F_{2,30} = 7.698$, $p < .01$, $\eta^2 = .032$) effect on the task-completion time with a small effect size. Post-hoc tests confirmed a significantly lower TCT for the 2-column conditions (EMM $\mu = 1.495\,\text{s}$, $\sigma_{\overline{x}} = 0.044\,\text{s}$) compared to the 6-column conditions (EMM $\mu = 1.585\,\text{s}$, $\sigma_{\overline{x}} = 0.044\,\text{s}$), $p < .001$. With regard to the number of rows (EMM μ between 1.52 s and 1.54 s), we could not find any significant influence ($F_{2,30} = .307$, $p > .05$). Also, we could not find any interaction effects between the factors ($F_{4,60} = 1.314$, $p > .05$). Figure 4b (red) depicts the TCTs for all conditions.

Footedness and Foot Used for the Interaction

We could not find any influence of the footedness of the participants on the accuracy ($F_{1,14} = .145$, $p > .05$) nor on the TCT ($F_{1,14} = 2.08$, $p > .05$).

As in the first experiment, almost all targets to the left of the participants' line of sight were performed with the left foot and vice versa ($\mu > .97$ for all conditions). Again, we found no significant influences of the number of rows ($F_{1.27,17.74} = .044$, $p > .05$, $\epsilon = .633$), the number of columns ($F_{1.33,18.61} = .344$, $p > .05$, $\epsilon = .665$) or the footedness of the participant ($F_{1,14} = .048$, $p > .05$) on the foot used for interaction.

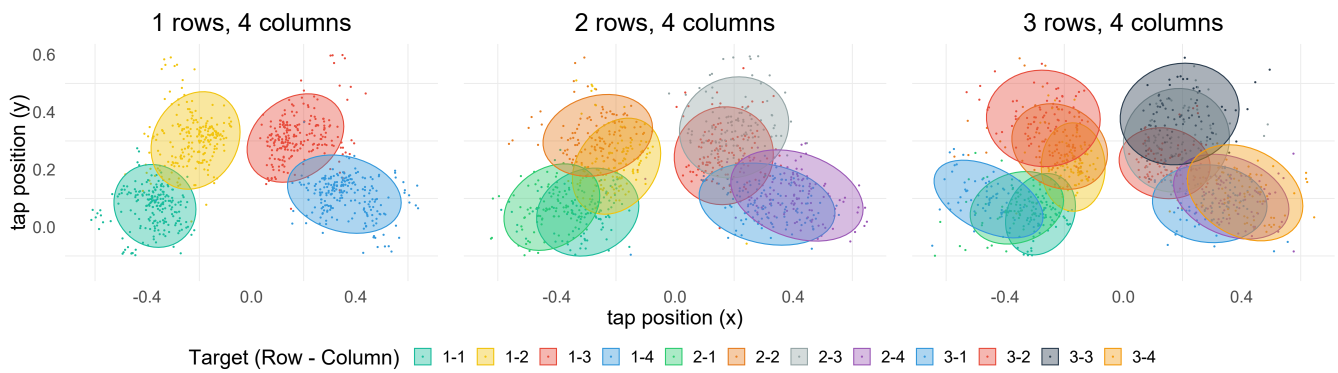


Figure 7: Scatter plots with 95% data probability ellipses for the 4-column conditions with *indirect* interfaces in the second experiment. While the data points for four columns can be separated, this is not possible for more than one row.

Size of the Target Areas

Again, the analysis showed a siginificant influence of the target row on the area of the targets ($F_{2,32} = 8.90$, $p < .001$, $\eta^2 = .027$) with a small effect size. Post-hoc tests confirmed significantly smaller areas if the target was in 1-row (EMM $\mu = 0.042\,\text{m}^2$, $\sigma_{\overline{x}} = 0.009\,\text{m}^2$) compared to 3-row (EMM $\mu = 0.074\,\text{m}^2$, $\sigma_{\overline{x}} = 0.009\,\text{m}^2$), $p < .001$. We could not find a significant influence of the target column ($F_{11,176} = 1.58$, $p > .05$) nor interaction effects ($F_{22,352} = 1.02$, $p > .05$).

Overlap

For conditions with multiple rows, there were noticeable overlaps in the distribution of the tapping points (see fig. 7 for the 4 column conditions). As a measure for these overlaps, we compared the number of points from adjacent targets in the row direction and in the column direction that fell into the 95% data ellipse of each target.

The analysis showed a significant difference between the overlap in row and column direction ($F_{1,17} = 324$, $p < .001$, $\epsilon = .890$, $\eta^2 =$) with a large effect size. Post-hoc tests confirmed a significantly lower overlap in row direction ($\mu = 4.0\,\%$, $\sigma = 3.7\,\%$) compared to the column direction ($\mu = 55.0\,\%$, $\sigma = 12.7\,\%$), $p < .001$.

TLX and Questionnaire

The analysis showed a significant influence of the number of rows ($F_{2,34} = 31.02$, $p < .001$, $\eta^2 = .125$) with a medium effect size. Post-hoc tests confirmed a significantly higher perceived cognitive load for higher numbers of rows ($p < .001$ comparing 1-row and 3-row, $p < .01$ otherwise) from EMM $\mu = 19.6$, $\sigma_{\overline{x}} = 2.55$ (1-row) over EMM $\mu = 25.0$, $\sigma_{\overline{x}} = 2.55$ (2-row) to EMM $\mu = 30.5$, $\sigma_{\overline{x}} = 2.55$ (3-row).

We further found a significant influence of the number of columns ($F_{2,34} = 10.481$, $p < .001$, $\eta^2 = .035$) with a small effect size. The post-hoc analysis showed rising estimated marginal means (2-column: EMM $\mu = 22.6$, $\sigma_{\overline{x}} = 2.53$, 4-column: EMM $\mu = 24.2$, $\sigma_{\overline{x}} = 2.53$, 6-column: EMM $\mu = 28.3$, $\sigma_{\overline{x}} = 2.53$) with significant differences between 2 and 6 columns ($p < .001$) as well as between 4 and 6 columns ($p < .05$). We could not observe interaction effects between the number of rows and the number of columns ($F_{4,68} = .447$, $p > .05$). Figure 4c (red) depicts the measured values.

Confidence. We asked the participants how confident they felt to have hit the correct targets. We found significant effects for both, the number of rows ($F_{2,34} = 22.711$, $p < .001$) as well as the number of columns ($F_{2,34} = 35.345$, $p < .001$). Post-hoc tests confirmed significantly higher confidence ratings for 1-row conditions compared to 2-row and 3-row conditions (both $p < .001$). For the number of columns, we found signifigantly rising ratings between all levels (all $p < .001$). We could not find interaction effects ($F_{4,68} = .185$, $p > .05$).

The absolute numbers (see fig. 8) show a high agreement for all 1-row conditions with decreasing confidence for higher numbers. Interestingly, the majority of the participants were convinced that they could keep the targets apart for all conditions (except 3-row, 6-column).

Convenience. We further asked the participants how convenient the layout felt to interact with information. The analysis showed significant effects for both, the number of rows ($F_{2,34} = 56.462$, $p < .001$) and the number of columns ($F_{2,34} = 8.203$, $p < .01$). Post-hoc tests confirmed significantly falling ratings for higher numbers of rows between all levels (all $p < .001$). Regarding the number of columns, we found significantly lower ratings for the 6-column conditions compared to the 2-column ($p < .01$) and 4-column ($p < .05$) conditions. We could not find interaction effects ($F_{4,68} = 1.947$, $p > .05$).

All but the 3-row, 6-column condition were rated predominantly positive.

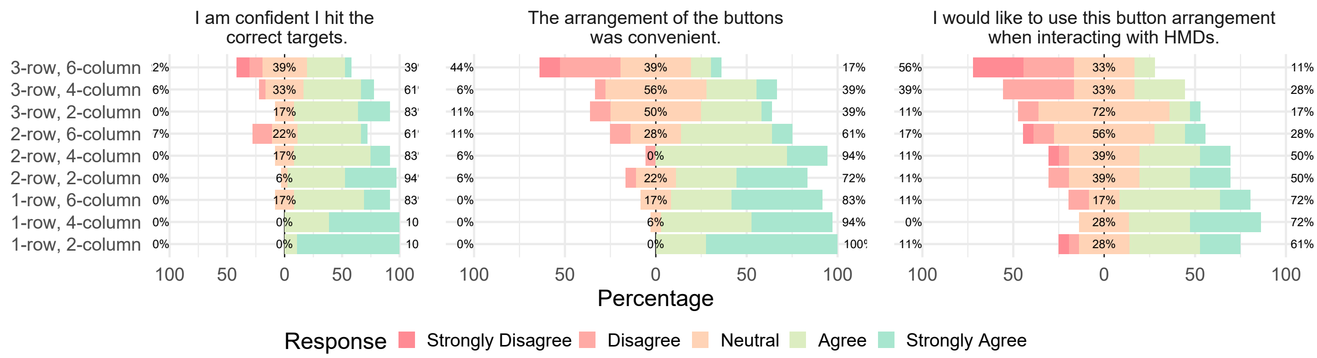


Figure 8: The participant's answers to our questions for indirect interfaces on a 5-point Likert-scale.

Willingness to Use. As the last question, we asked the participants if they would like to use this arrangement for interacting with HMDs. The analysis showed a significant effect for the number of rows ($F_{2,34} = 26.849$, $p < .001$) and the number of columns ($F_{2,34} = 3.600$, $p < .05$) as well as interaction effects between the factors ($F_{4,68} = 3.286$, $p < .05$). Post-hoc tests confirmed significantly lower ratings for the 3-row conditions compared to the 1-row and 2-row conditions (both $p < .001$). For the number of columns, post-hoc tests did not confirm significant differences.

Qualitative Feedback

In general, all participants appreciated the idea of being able to interact with HMDs using their feet without looking at the floor. When asked for the reasons, participants told us that this interaction modality felt "*novel*" (P1), "*fun to use*" (P12) and "*very easy to perform in addition to other tasks*" (P15) as the "*hands are not needed*" (P18) and "*it's a low effort extension [...] to interact*" (P9). Participants found the "*radial placement*" (P11) of the targets "*nice*" (P11) and had the feeling that different columns were "*relatively easy to discern*" (P7). Four of the participants felt "*unsure*" (P2, P5, P11, P12) about their performance with multiple rows. P11 even perceived more than one row as "*inconvenient*". P18 summarized: This "*feels quite naturally in comparison to the strange in-air gestures that are used for the Hololens*".

6 COMPARISON OF INTERACTION TECHNIQUES

We compared the two techniques using two-way RM ANOVA with the interaction method as a between-subjects factor.

Accuracy

We found a significant effect of the interaction method on the accuracy with a large effect size ($F_{1,32} = 133.00$, $p < .001$, $\eta^2 = .386$). Post-hoc tests confirmed significantly higher accuracy rates for *direct* (EMM $\mu = 98.9\,\%$, $\sigma_{\overline{x}} = 1.4\,\%$) compared to *indirect* (EMM $\mu = 78.0\,\%$, $\sigma_{\overline{x}} = 1.4\,\%$) ($p < .001$).

Task Completion Time

The analysis unveiled a significant effect of the interaction method on the TCT with a large effect size ($F_{1,32} = 17.8$, $p < .001$, $\eta^2 = .220$). Post-hoc tests confirmed significantly lower TCTs for *direct* (EMM $\mu = 1.277\,\text{s}$, $\sigma_{\overline{x}} = 0.042\,\text{s}$) compared to *indirect* (EMM $\mu = 1.529\,\text{s}$, $\sigma_{\overline{x}} = 0.042\,\text{s}$) ($p < .001$).

Size of Target Areas

We found a significant effect of the interaction method on the size of the target areas with a small effect size ($F_{1,33} = 19.7$, $p < .001$, $\eta^2 = .042$). Post-hoc tests confirmed significantly smaller areas for *direct* (EMM $\mu = 0.019\,\text{m}^2$, $\sigma_{\overline{x}} = 0.006\,\text{m}^2$) compared to *indirect* (EMM $\mu = 0.055\,\text{m}^2$, $\sigma_{\overline{x}} = 0.006\,\text{m}^2$) interactions ($p < .001$).

TLX and Questionaire

The analysis did not show significant effect of the interaction method on the raw TLX ($F_{1,34} = .002$, $p > .05$). We found a significant influence of the interaction method on the confidence ($F_{1,34} = 14.05$, $p < .001$). Post-hoc tests confirmed significantly higher ratings for *direct* compared to *indirect* ($p < .001$). We could not find significant effects of the interaction method on the convenience ($F_{1,34} = 1.83$, $p > .05$) or on the willingness to use ($F_{1,34} = 1.53$, $p > .05$).

7 DISCUSSION AND GUIDELINES

The results of our controlled experiments suggest that foot-taps provide a viable interaction technique for HMDs. In both experiments, the evaluation showed TCTs suitable for fast interactions. While we found significantly increasing TCTs for finer subdivisions of *direct* interfaces, the TCTs of *indirect* interfaces were stable across all conditions with only slight differences (see fig. 4b). Interestingly, for higher subdivisions, the TCT seem to converge between both styles.

Based on the analysis of the two interaction styles, we developed the following guidelines.

Favour the Division into Columns over Rows

Our results suggest that finer subdivision through higher numbers of rows have a greater impact on the accuracy than finer subdivisions through the addition of columns. This impression was further supported for *indirect* interfaces by investigating the overlap of the individual target areas: We found a significantly larger overlap within a column (i.e., between several rows) compared with the overlap within a row (i.e., between several columns). Also, in both experiments, we found a significantly growing spread of the tapping points for targets in more distant target rows (see fig. 7).

Therefore, we propose to favor the division into columns over rows when designing such interfaces.

Use indirect interfaces for longer-term interactions that require less accuracy

As expected, the accuracy rates for *indirect* interactions were significantly lower compared to *direct* interactions. However, the difference was very low for the 1-row conditions, or even negligible for 2 and 4 targets (see fig. 4a). Together with the differing overlaps in the row and column directions discussed above, this leads us to the conclusion that the participants - despite opposite self-perception - had great difficulties in distinguishing between different rows and, thus, the use of multiple rows for *indirect* interfaces is not feasible. Regarding the Likert-questionnaires and the qualitative feedback, we found greater popularity of the *indirect* interfaces.

Taken together the greater enthusiasm, as well as the lower TLX scores (for 1-row subdivisions), we recommend the use of *indirect* interfaces for most situations. In particular, this applies to situations where 1) a lower number of options is sufficient and 2) a restricted view (as in the *direct* interfaces, where the head is directed to the floor) could be problematic. Based on the analysis, we propose a 1-row, 4-column layout for *indirect* interfaces.

Use direct interfaces for short-term and fine-grained interactions

Direct interfaces delivered significantly higher accuracy rates compared to indirect interfaces. However, the analysis of qualitative feedback and answers in the Likert questionnaires showed a clear preference of participants for *indirect* interfaces. We assume that the limitations of the hardware used in the experiment (e.g., weight, field of view) have a considerable influence on the opinion. However, in particular the downward head posture seems to be rejected by the participants for longer-term interactions in general.

Therefore, we suggest the use of *direct* interfaces for short-term interactions requiring high accuracy and a large number of input options. For such interfaces, a high degree of accuracy is still achieved with 3-row, 6-column layouts.

8 LIMITATIONS AND FUTURE WORK

The design and results of our experiments impose some limitations and directions for future work.

Layout of the Targets

We used a fixed semicircular grid of targets. We chose this layout because of the natural reachability of targets from a fixed standing position. However, other shapes (e.g., rectangular, oval) and arrangements (e.g., not equally sized targets) could also be considered for future work. This is of particular interest as our experiment showed a larger spread for targets further away from the participant.

Feedback for Indirect Interaction

We did not show the participants any feedback about the position of their feet during the *indirect* experiment. Such additional feedback could strongly influence the performance of the participants. We chose this approach to investigate the ability of users to use *indirect* interfaces without visual feedback and, thus, create a baseline for future work.

Other Styles of Interaction

We concentrated on interfaces, which, as an analogy to the traditional point-and-click interfaces, are operated with foot-taps. Other interaction styles, such as gestures for fine-granular control or taps with different parts of the foot (e.g., heel) may be beneficial for the future use of HMDs.

The Midas Tap Problem

Similar to the Midas Touch Problem [34] in eye gaze tracking, it is challenging to separate intentional input from natural motion when using foot-based input. A possible solution could be a special foot input mode, activated using a secondary input modality such as a toggle on the HMD or gaze interaction in the user interface. For *direct* interfaces, just looking at the ground may be sufficient to activate this mode, as actions are only triggered after a subsequent tap. Further, sensor-based gait detection [12, 34] allows to only enable foot input while standing and, thus, help to prevent erroneous activation. Further work in this field is necessary to conclude on the Midas Tap problem.

9 CONCLUSION

We explored foot-taps as a *direct* and *indirect* input modality for HMDs. The results confirmed the viability of foot-taps for accurate and pleasant interaction. Based on the results, we derived guidelines for the design of such user interfaces.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) SWC “ARkTiP” (01IS17050).

REFERENCES

[1] 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 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, New York, New York, USA, 1229. https://doi.org/10.1145/2207676.2208575

[2] Daniel Ashbrook, Patrick Baudisch, and Sean White. 2011. Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, New York, New York, USA, 2043. https://doi.org/10.1145/1978942.1979238

[3] Thomas Augsten, Konstantin Kaefer, René Meusel, Caroline Fetzer, Dorian Kanitz, Thomas Stoff, Torsten Becker, Christian Holz, and Patrick Baudisch. 2010. Multitoe: High-Precision Interaction with Back-Projected Floors Based on High-Resolution Multi-Touch Input. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*. ACM Press, New York, New York, USA, 209. https://doi.org/10.1145/1866029.1866064

[4] Ralph M Barnes, Henry Hardaway, and Odif Podolsky. 1942. Which pedal is best. *Factory Management and Maintenance* 100, 98 (1942).

[5] Ralph L Barnett. 2009. *Foot Controls: Riding the Pedal.* Technical Report. 13–16 pages. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.843.7223{&}rep=rep1{&}type=pdf

[6] Mark Billinghurst, Adrian Clark, and Gun Lee. 2015. A Survey of Augmented Reality. *Foundations and Trends® in Human–Computer Interaction* 8, 2-3 (2015), 73–272. https://doi.org/10.1561/1100000049

[7] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. 2004. FingARtips: gesture based direct manipulation in Augmented Reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Austalasia and Southe East Asia - GRAPHITE '04*. ACM Press, New York, New York, USA, 212. https://doi.org/10.1145/988834.988871

[8] Jean P. Chapman, Loren J. Chapman, and John J. Allen. 1987. The measurement of foot preference. *Neuropsychologia* 25, 3 (jan 1987), 579–584. https://doi.org/10.1016/0028-3932(87)90082-0

[9] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences.* Routledge. https://doi.org/10.4324/9780203771587

[10] Andrea Colaço, Ahmed Kirmani, Hye Soo Yang, Nan-Wei Gong, Chris Schmandt, and Vivek K. Goyal. 2013. Mime: compact, low power 3D gesture sensing for interaction with head mounted displays. In *Proceedings of the 26th annual ACM symposium on User interface software and technology - UIST '13*. ACM Press, New York, New York, USA, 227–236. https://doi.org/10.1145/2501988.2502042

[11] E. N. Corlett and R. P. Bishop. 1975. Foot Pedal Forces for Seated Operators. *Ergonomics* 18, 6 (nov 1975), 687–692. https://doi.org/10.1080/00140137508931504

[12] Mohammad Derawi. 2010. Accelerometer-Based Gait Analysis , A survey. *Norwegian Information Security Conference* Mv (2010), 33–45. https://dl.acm.org/citation.cfm?doid=846276.846323

[13] Niloofar Dezfuli, Mohammadreza Khalilbeigi, Jochen Huber, Florian Müller, and Max Mühlhäuser. 2012. PalmRC: imaginary palm-based remote control for eyes-free television interaction. In *Proceedings of the 10th European conference on Interactive tv and video - EuroiTV '12*. ACM Press, New York, New York, USA, 27. https://doi.org/10.1145/2325616.2325623

[14] David Dobbelstein, Philipp Hock, and Enrico Rukzio. 2015. Belt: An Unobtrusive Touch Input Device for Head-worn Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 2135–2138. https://doi.org/10.1145/2702123.2702450

[15] David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2017. PocketThumb: a Wearable Dual-Sided Touch Interface for Cursor-based Control of Smart-Eyewear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (jun 2017), 1–17. https://doi.org/10.1145/3090055

[16] Phyllis B. Eveleth and J. M. (James Mourilyan) Tanner. 1990. *Worldwide variation in human growth.* Cambridge University Press. 397 pages.

[17] Kevin Fan, Liwei Chan, Daiya Kato, Kouta Minamizawa, and Masahiko Inami. 2016. VR Planet: Interface for Meta-View and Feet Interaction of VR Contents. In *ACM SIGGRAPH 2016 VR Village on - SIGGRAPH '16*. ACM Press, New York, New York, USA, 1–2. https://doi.org/10.1145/2929490.2931001

[18] Mingming Fan, Yizheng Ding, Fang Shen, Yuhui You, and Zhi Yu. 2017. An empirical study of foot gestures for hands-occupied mobile interaction. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers - ISWC '17*. ACM Press, New York, New York, USA, 172–173. https://doi.org/10.1145/3123021.3123043

[19] Steven Feiner, Blair MacIntyre, Tobias Höllerer, and Anthony Webster. 1997. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Personal Technologies* 1, 4 (dec 1997), 208–217. https://doi.org/10.1007/BF01682023

[20] Yasmin Felberbaum and Joel Lanir. 2016. Step by Step: Investigating Foot Gesture Interaction. In *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*. ACM Press, New York, New York, USA, 306–307. https://doi.org/10.1145/2909132.2926057

[21] Yasmin Felberbaum and Joel Lanir. 2018. Better Understanding of Foot Gestures: An Elicitation Study. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–12. https://doi.org/10.1145/3173574.3173908

[22] Patrick Tobias Fischer, Franziska Gerlach, Jenny Gonzalez Acuna, Daniel Pollack, Ingo Schäfer, Josephine Trautmann, and Eva Hornecker. 2014. Movable, Kick-/Flickable Light Fragments Eliciting Ad-hoc Interaction in Public Space. In *Proceedings of The International Symposium on Pervasive Displays - PerDis '14*. ACM Press, New York, New York, USA, 50–55. https://doi.org/10.1145/2611009.2611027

[23] Koumei Fukahori, Daisuke Sakamoto, and Takeo Igarashi. 2015. Exploring Subtle Foot Plantar-based Gestures with Sock-placed Pressure Sensors. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 3019–3028. https://doi.org/10.1145/2702123.2702308

[24] Fabian Göbel, Konstantin Klamka, Andreas Siegel, Stefan Vogt, Sophie Stellmach, and Raimund Dachselt. 2013. Gaze-supported foot interaction in zoomable information spaces. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. ACM Press, New York, New York, USA, 3059. https://doi.org/10.1145/2468356.2479610

[25] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*. ACM Press, New York, New York, USA, 3. https://doi.org/10.1145/1866029.1866033

[26] Sean Gustafson, Christian Holz, and Patrick Baudisch. 2011. Imaginary phone: learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*. ACM Press, New York, New York, USA, 283. https://doi.org/10.1145/2047196.2047233

[27] Teng Han, Jason Alexander, Abhijit Karnik, Pourang Irani, and Sriram Subramanian. 2011. Kick: investigating the use of kick gestures for mobile interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. ACM Press, New York, New York, USA, 29. https://doi.org/10.1145/2037373.2037379

[28] Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. 2012. On-Body Interaction: Armed and Dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction - TEI '12*. ACM Press, New York, New York, USA, 69. https://doi.org/10.1145/2148131.2148148

[29] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*. ACM Press, New York, New York, USA, 453. https://doi.org/10.1145/1753326.1753394

[30] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (oct 2006), 904–908. https://doi.org/10.1177/154193120605000909

[31] Hwan Heo, Eui Lee, Kang Park, Chi Kim, and Mincheol Whang. 2010. A realistic game system using multi-modal user interfaces. *IEEE Transactions on Consumer Electronics* 56, 3 (aug 2010), 1364–1372. https://doi.org/10.1109/TCE.2010.5606271

[32] Juan David Hincapié-Ramos, Xiang Guo, Paymahn Moghadasian, and Pourang Irani. 2014. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, New York, New York, USA, 1063–1072. https://doi.org/10.1145/2556288.2557130

[33] Yi-Ta Hsieh, Antti Jylhä, Valeria Orso, Luciano Gamberini, and Giulio Jacucci. 2016. Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, New York, New York, USA, 4203–4215. https://doi.org/10.1145/2858036.2858436

[34] Robert J K Jacob. 1995. Eye Tracking in Advanced Interface Design. *Virtual Environments and Advanced Interface Design* (1995), 258–290.

[35] Ricardo Jota, Pedro Lopes, Daniel Wigdor, and Joaquim Jorge. 2014. Let's Kick It: How to Stop Wasting the Bottom Third of Your Large Scale Display Ricardo. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, New York, New York, USA, 1411–1414. https://doi.org/10.1145/2556288.2557316

[36] K. H. E. Keoemer. 1971. Foot Operation of Controls. *Ergonomics* 14, 3 (may 1971), 333–361. https://doi.org/10.1080/00140137108931255

[37] Marion Koelle, Abdallah El Ali, Vanessa Cobus, Wilko Heuten, and Susanne CJ Boll. 2017. All about Acceptability?: Identifying Factors for the Adoption of Data Glasses. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), 295–300. https://doi.org/10.1145/3025453.3025749

[38] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, New York, New York, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[39] Roman Lissermann, Jochen Huber, Aristotelis Hadjakos, and Max Mühlhäuser. 2013. EarPut: augmenting ear-worn devices for ear-based interaction. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. ACM Press, New York, New York, USA, 1323. https://doi.org/10.1145/2468356.2468592

[40] Pedro Lopes, Alexandra Ion, Willi Mueller, Daniel Hoffmann, Patrik Jonell, and Patrick Baudisch. 2015. Proprioceptive Interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, New York, New York, USA, 939–948. https://doi.org/10.1145/2702123.2702461

[41] Zhihan Lv. 2013. Wearable Smartphone: Wearable Hybrid Framework for Hand and Foot Gesture Interaction on Smartphone. In *2013 IEEE International Conference on Computer Vision Workshops*. IEEE, 436–443. https://doi.org/10.1109/ICCVW.2013.64

[42] Zhihan Lv, Shengzhong Feng, Muhammad Sikandar Lal Khan, Shafiq Ur Réhman, and Haibo Li. 2014. Foot Motion Sensing: Augmented Game Interface Based on Foot Interaction for Smartphone Zhihan. In *Proceedings of the extended abstracts of the 32nd annual ACM conference on Human factors in computing systems - CHI EA '14*. ACM Press, New York, New York, USA, 293–296. https://doi.org/10.1145/2559206.2580096

[43] Zhihan Lv, Alaa Halawani, Shengzhong Feng, Haibo Li, and Shafiq Ur Réhman. 2014. Multimodal Hand and Foot Gesture Interaction for Handheld Devices. *ACM Transactions on Multimedia Computing, Communications, and Applications* 11, 1s (oct 2014), 1–19. https://doi.org/10.1145/2645860

[44] Zhihan Lv, Alaa Halawani, Shengzhong Feng, Shafiq ur Réhman, and Haibo Li. 2015. Touch-less interactive augmented reality game on vision-based wearable device. *Personal and Ubiquitous Computing* 19, 3-4 (jul 2015), 551–567. https://doi.org/10.1007/s00779-015-0844-1

[45] Denys J. C. Matthies, Franz Müller, Christoph Anthes, and Dieter Kranzlmüller. 2013. ShoeSoleSense: proof of concept for a wearable foot interface for virtual and real environments. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology - VRST '13*. ACM Press, New York, New York, USA, 93. https://doi.org/10.1145/2503713.2503740

[46] Pranav Mistry and Pattie Maes. 2009. SixthSense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches on - SIGGRAPH ASIA '09*. ACM Press, New York, New York, USA, 1. https://doi.org/10.1145/1667146.1667160

[47] Florian Müller, Niloofar Dezfuli, Max Mühlhäuser, Martin Schmitz, and Mohammadreza Khalilbeigi. 2015. Palm-based Interaction with Head-mounted Displays. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '15*. ACM Press, New York, New York, USA, 963–965. https://doi.org/10.1145/2786567.2794314

[48] Florian Müller, Mohammadreza Khalilbeigi, Niloofar Dezfuli, Alireza Sahami Shirazi, Sebastian Günther, and Max Mühlhäuser. 2015. A Study on Proximity-based Hand Input for One-handed Mobile Interaction. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction - SUI '15*. ACM Press, New York, New York, USA, 53–56. https://doi.org/10.1145/2788940.2788955

[49] Toni Pakkanen and Roope Raisamo. 2004. Appropriateness of foot interaction for non-accurate spatial tasks. In *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*. ACM Press, New York, New York, USA, 1123. https://doi.org/10.1145/985921.986004

[50] G. Pearson and M. Weiser. 1986. Of moles and men: the design of foot controls for workstations. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '86*, Vol. 17. ACM Press, New York, New York, USA, 333–339. https://doi.org/10.1145/22627.22392

[51] Vijay Dandur Rajanna. 2016. Gaze and Foot Input: Toward a Rich and Assistive Interaction Modality. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces - IUI '16 Companion*. ACM Press, New York, New York, USA, 126–129. https://doi.org/10.1145/2876456.2876462

[52] 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 - DIS '16*. ACM Press, New York, New York, USA, 323–333. https://doi.org/10.1145/2901790.2901815

[53] Daniel Schmicking and Shaun Gallagher (Eds.). 2010. *Handbook of Phenomenology and Cognitive Science*. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-90-481-2646-0

[54] Stefan Schneegass and Alexandra Voit. 2016. GestureSleeve: using touch sensitive fabrics for gestural input on the forearm for controlling smartwatches. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers - ISWC '16*. ACM Press, New York, New York, USA, 108–115. https://doi.org/10.1145/2971763.2971797

[55] Johannes Schöning, Florian Daiber, Antonio Krüger, and Michael Rohs. 2009. Using hands and feet to navigate and manipulate spatial data. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*. ACM Press, New York, New York, USA, 4663. https://doi.org/10.1145/1520340.1520717

[56] S. R. Searle, F. M. Speed, and G. A. Milliken. 1980. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. *The American Statistician* 34, 4 (nov 1980), 216–221. https://doi.org/10.1080/00031305.1980.10483031

[57] Marcos Serrano, Barrett M. Ens, and Pourang P. Irani. 2014. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, New York, New York, USA, 3181–3190. https://doi.org/10.1145/2556288.2556984

[58] Mara G. Silva and Doug A. Bowman. 2009. Body-based interaction for desktop games. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*. ACM Press, New York, New York, USA, 4249. https://doi.org/10.1145/1520340.1520648

[59] 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)*. IEEE, 71–74. https://doi.org/10.1109/3DUI.2014.6798845

[60] Oleg Špakov and Päivi Majaranta. 2012. Enhanced gaze interaction using simple head gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. ACM Press, New York, New York, USA, 705. https://doi.org/10.1145/2370216.2370369

[61] T.E. Starner. 2002. The role of speech input in wearable computing. *IEEE Pervasive Computing* 1, 3 (jul 2002), 89–93. https://doi.org/10.1109/MPRV.2002.1037727

[62] Emi Tamaki, Takashi Miyak, and Jun Rekimoto. 2010. Brainy hand: an ear-worn hand gesture interaction device. In *Proceedings of the International Conference on Advanced Visual Interfaces - AVI '10*. ACM Press, New York, New York, USA, 387. https://doi.org/10.1145/1842993.1843070

[63] 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 *Lecture Notes in Computer Science*, J Abascal, S Barbosa, M Fetter, T Gross, P Palanque, and M Winckler (Eds.). Springer, Cham, Chapter Human-Comp, 384–401. https://doi.org/10.1007/978-3-319-22701-6_29

[64] Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The Feet in Human–Computer Interaction. *Comput. Surveys* 48, 2 (sep 2015), 1–35. https://doi.org/10.1145/2816455

[65] Julie Wagner, Mathieu Nancel, Sean G. Gustafson, Stephane Huot, and Wendy E. Mackay. 2013. Body-centric design space for multi-surface interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*. ACM Press, New York, New York, USA, 1299. https://doi.org/10.1145/2470654.2466170

[66] Martin Weigel, Vikram Mehta, and Jürgen Steimle. 2014. More Than Touch: Understanding How People Use Skin as an Input Surface for Mobile Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 179–188. http://dl.acm.org/citation.cfm?doid=2556288.2557239

[67] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, New York, New York, USA, 143. https://doi.org/10.1145/1978942.1978963

[68] Tetsuya Yamamoto, Masahiko Tsukamoto, and Tomoki Yoshihisa. 2008. Foot-Step Input Method for Operating Information Devices While Jogging. In *2008 International Symposium on Applications and the Internet*. IEEE, 173–176. https://doi.org/10.1109/SAINT.2008.97

[69] Koki Yamashita, Takashi Kikuchi, Katsutoshi Masai, Maki Sugimoto, Bruce H. Thomas, and Yuta Sugiura. 2017. CheekInput: turning your cheek into an input surface by embedded optical sensors on a head-mounted display. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology - VRST '17*. ACM Press, New York, New York, USA, 1–8. https://doi.org/10.1145/3139131.3139146

[70] Kangkang Yin and Dinesh K Pai. 2003. FootSee: an Interactive Animation System. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation*. Association for Computing Machinery, 385. https://dl.acm.org/citation.cfm?id=846323