WatchVR: Exploring the Usage of a Smartwatch for Interaction in Mobile Virtual Reality

Teresa Hirzle
Institute of Media Informatics
Ulm University, Ulm, Germany
teresa.hirzle@uni-ulm.de

Jan Gugenheimer
Institute of Media Informatics
Ulm University, Ulm, Germany
jan.gugenheimer@uni-ulm.de

Jan Rixen
Institute of Media Informatics
Ulm University, Ulm, Germany
jan.rixen@uni-ulm.de

Enrico Rukzio
Institute of Media Informatics
Ulm University, Ulm, Germany
enrico.rukzio@uni-ulm.de

Submitted to CHI 2018

Abstract
Mobile virtual reality (VR) head-mounted displays (HMDs) are steadily becoming part of people's everyday life. Most current interaction approaches rely either on additional hardware (e.g. Daydream Controller) or offer only a limited interaction concept (e.g. Google Cardboard). We explore a solution where a conventional smartwatch, a device users already carry around with them, is used to enable short interactions but also allows for longer complex interactions with mobile VR. To explore the possibilities of a smartwatch for interaction, we conducted a user study in which we compared two variables with regard to user performance: *interaction method* (touchscreen vs inertial sensors) and *wearing method* (hand-held vs wrist-worn). We found that selection time and error rate were lowest when holding the smartwatch in one hand using its inertial sensors for interaction (hand-held).

Author Keywords
3D pointing; smartwatch; nomadic virtual reality; mobile virtual reality

ACM Classification Keywords
H.5.2 [Information interfaces and presentation (e.g., HCI)]: User Interfaces

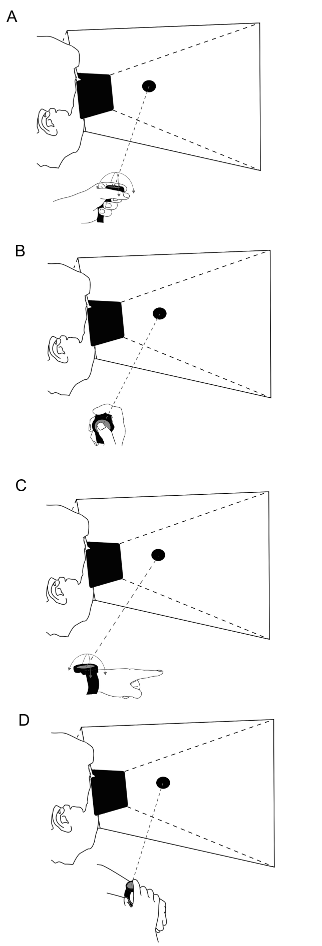


Figure 1: The four systems that are evaluated in the user study: A/B hand-held and C/D wrist-worn, each with both interaction methods: sensor-based and touchscreen-based.

Introduction

Mobile VR devices have the potential to make virtual reality accessible to a bigger population. In 2020 there will be expectedly 135 million mobile VR users worldwide [3]. Compared to stationary VR systems, mobile VR devices do not require permanent tethering and can be carried around effortless. Multiple smartphone-based VR systems are emerging by delegating imaging, computing and tracking capabilities to the smartphone, which allows for a more affordable VR experience [5].

Due to apparent physical limitations, the ways to interact with such mobile headsets differ from the usual interaction with a smartphone. As the case covers most of the device, buttons and touchscreen cannot be operated properly. Thus, other concepts to interact in VR have been implemented [12, 9, 6]. Consumer devices, such as the Samsung GearVR and Google Daydream, provide additional controllers for interacting in mobile VR scenarios. These controllers provide three degrees of freedom (DoF), a touchpad and several buttons for interaction. Whereas this approach allows for more manifold interactions than using gaze direction only, users are still limited as they have to carry around an additional device.

Furthermore, the intention of mobile VR is not only to provide long, refined VR-experiences where one might constantly interact by using a controller, but also "bite-sized" experiences [2]. These describe the idea that mobile VR (in contrast to stationary systems) will be mainly used for short experiences, in which the focus lies on short and simple interactions, such as exploring a 360° selfie or playing/stopping a video. For these kinds of interactions however, a single purpose hand-held device might be too effortful to carry around compared to the purpose of the interaction. This was previously already discussed by Daniel Ashbrook [1] where he argues that the access time for mobile interactions should be appropriate for the actual interaction time. Therefore, we argue for the usage of a device that provides a solution for both scenarios: a smartwatch can be used wrist-worn for short and simple interaction in "bite-sized" experiences, but also as a mobile VR controller for long VR-experiences by holding it in one hand and using its inertial sensors for interaction. This should potentially allow the user to choose the appropriate form of interaction based on the upcoming task.

In a first step we identified and explored the different degrees of freedom a smartwach has in terms of usage for interaction. We conducted a first user study (n = 15), in which we compared two variables: holding the smartwatch ("controller-like") in one hand vs wearing the smartwatch and using the touchscreen of the smartwatch vs using its inertial sensors (accelerometer and gyroscope) for interaction. For this we measured selection time, error rate, level of immersion and mental workload. We found that holding the smartwatch in one hand using its inertial sensors for pointing, resulted in lower selection time and error rate as using the smartwatch wrist-worn. In a next step we plan to conduct a second user study, focusing more on the "bite-sized" experiences. We not only aim to explore the interaction inside of VR but also the access time and how different task durations influence users' preferences on wearing methods (hand-held vs wrist-worn).

Related Work

The field of mobile and nomadic VR [5] is only recently being explored by HCI, since the technology only lately got mature enough to allow for a mobile VR experience. Several interaction techniques were recently proposed to allow users to interact in VR inside an unknown and uninstrumentend environment [12, 9, 6]. Smus et al. presented

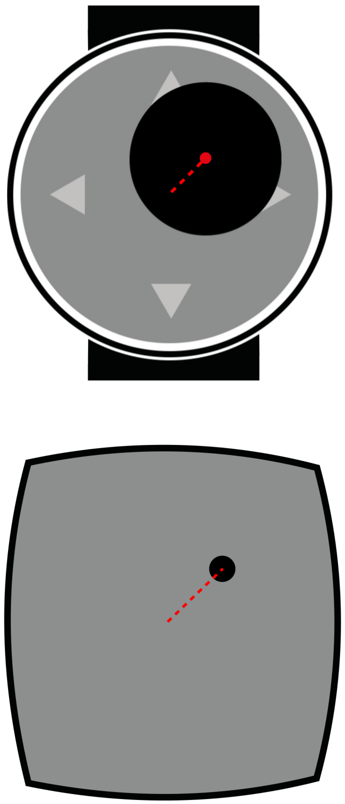


Figure 2: For the touchscreen-based interaction method the pointing ray was displayed relative to the origin on both FOV and touchscreen. Pointing was achieved with one tap, selection with two taps of the black button.

the initial concept of the magnet-based input of the original google cardboard [12]. This was further explored by Lyons et al. by extending the binary selection to a 2D input [9]. Both these works focused on the interaction at the users temple, whereas Gugenheimer et al. further explored how good users can interact with the back of an VR HMD [6]. Our work concentrates on egocentric interaction techniques, especially on virtual pointer methods [4], where a ray is emitted from the user's hand and directed towards an interaction object, which can then be selected and manipulated.

The closest to our work are *Watchcasting* and *TickTockRay* [11, 7], which both use a smartwatch to interact with virtual content. The former describes a 3D interaction technique using an off-the-shelf smartwatch, which enables target selection and translation by mapping the z-coordinate position to forearm rotation. The work shows that a conventional smartwatch is a practical alternative for 3D interaction. Whereas *Watchcasting* provides an adequate way of interacting with screens and large displays, we in contrast propose to use a smartwatch for interactions in VR, similar to *TickTockRay*, which presents one concept using the smartwatch as an input device for mobile VR. However, no user studies or formal evaluations were conducted with *TickTockRay*, whereas our goal was to explore and evaluate a variety of different smartphone interaction concepts (see Fig. 1) and their impact on user performance.

WatchVR

WatchVR is an interaction concept for mobile VR based on a smartwatch. It explores possible interaction capabilities a smartwatch has to offer for VR, aiming at overcoming restrictions current interaction concepts have, such as the usage of additional hardware or limited interaction possibilities (e.g. Google Cardboard). To explore the usage of a smartwatch for interaction in mobile VR we identified two variables: holding the smartwatch in one hand using it "controller-like" vs using the smartwatch wrist-worn (*wearing method*) and using the smartwatch's inertial sensors (gyroscope and accelerometer) vs using the smartwatch's touchscreen (*interaction method*) for interaction. This resulted in four systems (A-D), which are displayed in Fig. 1.

Implementation

For both interaction methods we implemented an absolute pointing task based on the ray casting metaphor. For the systems based on accelerometer and gyroscope data, we implemented the technique proposed by Pietrozek et al. [11], which uses the device's yaw and tilt data. As the smartwatch's position is not traceable, we used the position of the user's head as a reference point to calculate the ray's origin. For the touchscreen-based interaction metaphor, we mapped the smartwatch's touchscreen to the FOV of the HMD, such that the range of action was limited to the FOV (see Fig. 2). For selection in both cases a simple button implemented on the touchscreen was used. We implemented our system using the Google Cardboard, a Nexus 5 and the LG Watch R.

Evaluation

We conducted a first user study, comparing our four proposed systems and measuring the impact of the two identified independent variables *interaction method* (touchscreen vs inertial sensors) and *wearing method* (hand-held vs wrist-worn) on the systems' performance. The goal of our first study was to measure the individual impact each factor (*interaction method* and *wearing method*) has on the users' performance and thereby better understand the full capabilities of the smartwatch for interacting in virtual reality.

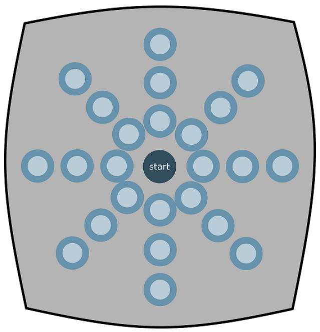


Figure 3: Position and size of targets for the Fitt's Law task. Target sizes (small/big) are indicated through the two different shades of blue.

Figure 4: A participant seated on a desk chair with fold up armrests wearing a Google Cardboard using the hand-held touchscreen system.

Study Design

We implemented a Fitt's Law task, where each iteration consisted of two target selections. Whereas the first target was placed in the middle, the second varied in position and size. We applied 3 different distances (3, 6, 9 Unity Units (UU)), 2 different sizes (2 and 4 UU) and 8 different angles, which resulted in 48 different targets (see Fig. 3) and 96 selections per system, as every combination occurred twice. Conditions were counterbalanced using a Latin square. We measured the following dependent variables: *selection time*, *error rate*, *distance* of the selected point *from the center* of the target, *level of immersion* (E^2I [8]) and *level of mental workload* (Nasa-Tlx [10]).

Participants

We recruited 15 participants (two female) with an average age of 22.7 years (range: 19 to 26). Seven of them reported to have had contact with VR, while four stated to have experienced sporadic interactions. Furthermore, seven participants had already used smartwatches.

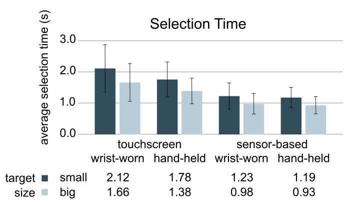
Quantitative Results

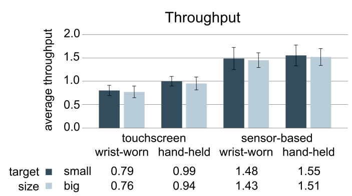
Scores from the Nasa-Tlx and E^2I, *error rate*, *selection time* and *throughput* were analyzed using a repeated measures ANOVA. Differences were examined regarding the *interaction method* and the *wearing method*.

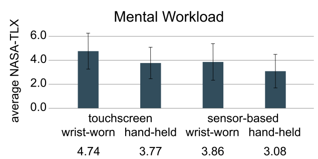
Regarding *selection time* a significant main effect between the two *wearing methods* could be found ($F(1, 14) = 57.722$, $p < .001$). Bonferroni corrected pairwise comparison showed that participants were 13% faster using the hand-held concept ($M = 1.315$ s, $SD = .04$) for pointing than using the wrist-worn concept ($M = 1.498$ s, $SD = .094$). Regarding *throughput* a significant difference between the *wearing methods* could be measured ($F(1, 14) = 131.84$, $p < .001$). The pairwise comparison showed that participants had a significantly higher ($p < .005$) throughput rate with the hand-held concepts ($M = 1.428$, $SD = .043$) than with the wrist-worn ones ($M = 1.326$, $SD = .062$). Regarding the *interaction method* pointing using inertial sensors ($M = 1.74$, $SD = .072$) reached an about 74% greater throughput ($p < .001$) than using the touchscreen ($M = 1.014$, $SD = .046$). For *error rate* no significant difference could be found neither regarding the *interaction method* nor the *wearing method*. Values for *mental workload* were also not different for the *interaction methods*. The pairwise comparison would however show that the hand-held methods ($M = 3.422$, $SD = .29$) produced a significantly lower level of *mental workload* ($p < .001$) than the wrist-worn ones ($M = 4.3$, $SD = .301$). Pairwise comparison showed further that participants felt significantly ($F(1, 14) = 8.067$, $p < .05$) more immersed using the inertial sensors ($M = 5.27$, $SD = .56$) for pointing in contrast to using the touchscreen of the smartwatch ($M = 4.53$, $SD = .37$).

User Feedback

After the study participants were further advised to order the concepts regarding liking and task efficiency. Twelve out of 15 ranked the "controller-like" concept A (wearing the smartwatch in the hand and using inertial sensors for pointing) on the first position regarding their liking, mostly explained through the high intuitiveness and precision of the concept. Furthermore, ten out of 15 participants stated that using the touchscreen for pointing felt slow and inaccurate compared to the inertial sensors. Although two participants mentioned that they would probably have liked it more if they had more practice. Results for the ranking for effectiveness looked similar to the prior one, justified mostly with the same reasons. When asked if the advantage of the "controller-like" concept would justify the additional effort of taking off the smartwatch, most participants answered with "yes". However, one participant stated that it would depend on the duration he intended to use it. For short periods he







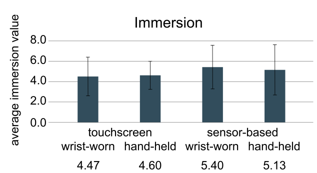


Figure 5: The quantitative results for *selection time*, *throughput*, *mental workload* and *immersion*.

would not justify it but for a case of long usage it would be worth the effort. Based on findings of Daniel Ashbrook [1] we expect similar statements from participants if we would shorten the duration of the task significantly (e.g. instead of one selection task taking 10 minutes use 10 selection tasks each taking 1 minute).

Discussion

With this first study we aimed to explore and identify appropriate (in terms of performance and usability) interaction capabilities a smartwatch has to offer for virtual reality. We compared the usage of the inertial sensors with the touch input capabilities of a smartwatch (*interaction method*) and looked at how the factor of wearing the watch vs holding the watch in the hand (*wearing method*) influenced the performance.

Regarding the *interaction method* using the inertial sensors of the smartwatch for pointing did outperform the touchscreen in almost all points, particularly regarding speed of input and throughput. These findings confirm the emergence of current controller-based interaction methods for mobile VR, such as the Google Daydream controller. Since the three DoF concept relies on applying direct interaction, which is known to result in lower interaction times [13], it outperforms the touchscreen-based one, which relies on indirect interaction.

For the *wearing method* participants preferred the hand-held concept to the wrist-worn one. This also resulted in the best performance values in terms of accuracy and timing. These findings substantiate the importance of choosing an appropriate interaction concept based on the task condition. For long interactions (e.g. one task with 10 min duration), which we evaluated in our first study, the hand-held concept seems to be more appropriate, as the influence of access time (only once) can be neglected. However, when considering 10 tasks with each 1 minute (remove VR HMD between tasks), we expect the wrist-worn concept to be more appropriate as now access time is crucial for task efficiency. We aim to examine this in a second user study.

Conclusion and Future Work

In this work we explored the capabilities of a smartwatch to be used as an input method for mobile virtual reality. We identified two variables (*interaction method* and *wearing method*) and explored their impact on user performance. We found that holding the smartwatch in the hand and using the inertial sensor to cast a ray resulted in the best performance and highest user preference. This also justifies the current usage and distribution of controller-based interaction methods for mobile VR devices (e.g. Google Daydream, Samsung Gear VR).

However, we argue that these performances and usability metrics inside a long user task (e.g. Fitt's law task) do not fully represent the mobile VR application scenario. Similar to the concept of 'bite-size VR' by Dobson [2], we argue that mobile VR will have a different usage scenario then VR has at home. Users will probably fluently mix between VR and not VR and will only spend a short interaction cycle inside of virtual reality (e.g. looking at a 360 image of a friend). Therefore, we argue that the access time for the interaction will become more relevant when the task will be adapted to the mobile VR interaction scenario. In our next step we will explore this specific scenario and not only focus on the raw performance but also take task switches and the overall orchestration of the interaction into consideration. We expect that with lower interaction times inside the VR task, users will prefer the wrist-worn *wearing method* to the hand-held one, since the access time of interaction will become a crucial point.

Acknowledgements
This work was conducted within the Emmy Noether research group Mobile Interaction with Pervasive User Interfaces funded by the German Research Foundation (DFG).

REFERENCES

1. Daniel L Ashbrook. 2010. *Enabling mobile microinteractions*. Georgia Institute of Technology.
2. Wolff Dobson. 2015. Lightning Talk: Bite-Sized VR. (2015). Retrieved January 2, 2018 from https://www.youtube.com/watch?v=MueSHMNyGtM.
3. Statista (EMarketer). 2018. Number of mobile virtual reality (VR) users worldwide from 2015 to 2020 (in millions). (2018). Retrieved January 3, 2018 from https://www.statista.com/statistics/650834/mobile-vr-users-worldwide/.
4. Tovi Grossman and Ravin Balakrishnan. 2006. The design and evaluation of selection techniques for 3D volumetric displays. In *Proc. UIST 2006*. ACM, 3–12.
5. Jan Gugenheimer. 2016. Nomadic Virtual Reality: Exploring New Interaction Concepts for Mobile Virtual Reality Head-Mounted Displays. In *Proc. UIST'16 (UIST '16 Adjunct)*. ACM, New York, NY, USA, 9–12. DOI:http://dx.doi.org/10.1145/2984751.2984783
6. Jan Gugenheimer, David Dobbelstein, Christian Winkler, Gabriel Haas, and Enrico Rukzio. 2016. FaceTouch: Enabling Touch Interaction in Display Fixed UIs for Mobile Virtual Reality. In *Proc. UIST'16*. ACM, New York, NY, USA, 49–60. DOI: http://dx.doi.org/10.1145/2984511.2984576
7. Daniel Kharlamov, Brandon Woodard, Liudmila Tahai, and Krzysztof Pietroszek. 2016. TickTockRay: smartwatch-based 3D pointing for smartphone-based virtual reality. In *Proc. of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 363–364.
8. JJ-W Lin, Henry Been-Lirn Duh, Donald E Parker, Habib Abi-Rached, and Thomas A Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Virtual Reality, 2002. Proceedings. IEEE*. IEEE, 164–171.
9. Kent Lyons. 2016. 2D Input for Virtual Reality Enclosures with Magnetic Field Sensing. In *Proc.of the 2016 ACM International Symposium on Wearable Computers (ISWC '16)*. ACM, New York, NY, USA, 176–183. DOI: http://dx.doi.org/10.1145/2971763.2971787
10. Nasa. 2018. Task Load Index (Nasa-TLX). (2018). Retrieved January 5, 2018 from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021487.pdf.
11. Krzysztof Pietroszek, Liudmila Tahai, James R Wallace, and Edward Lank. 2017. Watchcasting: Freehand 3D interaction with off-the-shelf smartwatch. In *3D User Interfaces (3DUI), 2017 IEEE Symposium on*. IEEE, 172–175.
12. Boris Smus and Christopher Riederer. 2015. Magnetic Input for Mobile Virtual Reality. In *Proc.of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 43–44. DOI: http://dx.doi.org/10.1145/2802083.2808395
13. Christian Winkler, Ken Pfeuffer, and Enrico Rukzio. 2012. Investigating Mid-air Pointing Interaction for Projector Phones. In *Proc.of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 85–94. DOI:http://dx.doi.org/10.1145/2396636.2396650