Should I Interrupt or Not? Understanding Interruptions in Head-Mounted Display Settings

Ceenu George[1], **Philipp Janssen**[1], **David Heuss**[1], **Florian Alt**[2]
[1]LMU Munich, Germany,{firstname.lastname}@ifi.lmu.de
[2]Bundeswehr University Munich, Germany, florian.alt@unibw.de

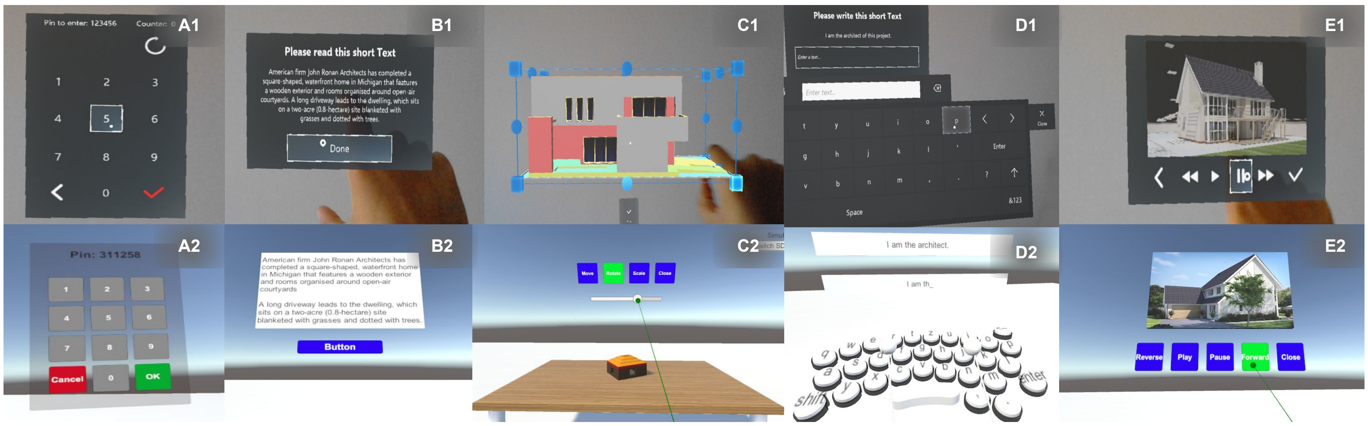

Figure 1. We investigate interruptions of head-mounted display (HMD) users by people not wearing an HMD (bystanders). At the focus of our work we explore whether bystanders can identify task switches of the HMD user and, hence, opportune moments for interruptions. In particular, we compare 5 tasks in AR (top) and VR (bottom) : authentication (A), reading (B), manipulation (C), typing (D), and watching a video (E).

ABSTRACT

Head-mounted displays (HMDs) are being used for VR and AR applications and increasingly permeate our everyday life. At the same time, a detailed understanding of interruptions in settings where people wearing an HMD (HMD user) and people not wearing an HMD (bystander) is missing. We investigate (a) whether bystanders are capable of identifying when HMD users switch tasks by observing their gestures, and hence exploit opportune moments for interruptions, and (b) which strategies bystanders employ. In a lab study (*N*=64) we found that bystanders are able to successfully identify both task switches (83%) and tasks (77%) within only a few seconds of the task switch. Furthermore, we identified interruption strategies of bystanders. From our results we derive implications meant to support designers and practitioners in building HMD applications that are used in a co-located collaborative setting.

Author Keywords

HMD; Gesture; Interruption; Virtual and Augmented Reality

CCS Concepts

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

DIS'19, June 07–12, 2016, San Diego, CA, USA

ISBN 123-4567-24-567/08/06. . . $15.00

DOI: http://dx.doi.org/10.475/123_4

INTRODUCTION

Augmented and Virtual Reality (AVR) is finding its way into many application areas, including but not limited to architecture and automotive development [16, 32, 44]. As users interact in AVR by means of head-mounted displays (HMDs), they become widely unaware of the real world (RW). This creates a challenge in situations, where people in the real-world (bystanders) need to interrupt the user wearing the HMD – similar to situations in a workspace where colleagues engage in brief discussions or ask each other for advice. In particular, the interruption not only impacts on the current task and performance, but it also leads to a loss of immersion and presence[1].

Building on research in the subfield of interruptions within HCI we investigate, whether real world bystanders can recognize what the HMD (AR vs. VR) user is currently doing without the aid of technology. We are interested in whether or not humans can tell from observing the behaviour of an HMD user – such as position, interaction gestures, and head movements – when they are switching between two tasks. This knowledge is valuable, as prior research on interruptions showed that changes in tasks represent an opportune moment for reducing interruption costs [38, 68]. This presents a chance for bystanders to approach an HMD user while minimizing any negative effect of the interruption on their level of immersion and presence [63, 79] and, ultimately, their productivity.

[1]Immersion and presence are established terms for measuring the quality of IVR experiences. Immersion refers to an objective evaluation of the hardware capabilities (e.g., frame rate, field of view) [63, 66], whereas presence describes an individual evaluation of the AVR experience [46, 63, 66].

We conducted a lab study (N=64), revealing that bystanders are indeed able to recognize task switches (83%) and tasks (77%) by observing HMD users' behaviours. On average it took them 5.02 s (SD=2.8) to recognize a task switch and 7 s (SD=1.6) to identify a task. Furthermore, we discovered strategies observers employ to identify task switches / tasks.

To our knowledge, this paper is first to (1) explore users' strategies for interrupting an HMD user and to (2) investigate observations of AR vs VR users as means to inform opportune moments for interruption. We believe the findings of this work to be useful for researchers and practitioners who are designing for the use of HMDs in settings where people in the real world are working together with people in AVR.

BACKGROUND AND RELATED WORK

We draw from several strand of prior work. presence and immersion, observing gestures, and interruptions.

Interaction between HMD Users and Bystanders

Prior work mainly focused on interaction between users in mixed realities with a common collaborative goal [6, 45], rather than evaluating how a bystander in the real world can interrupt an HMD user. Gugenheimer et al. directed the focus onto bystanders by including them in the VR gameplay through a tablet device [24] and a face display [25, 48], mounted onto the HMD. They found this type of involvement to improve social interaction and acceptance between VR user and RW bystander. However, their experiments were centred around a common gaming theme in which bystanders were exposed to technical devices (tablet vs. HMD face-display).

Prior work has reviewed intrinsic interruptions by displaying virtual content to the AVR user, in form of projections of the bystander [51] and through notifications [21]. Although, these are valid approaches, the aim of our work is to empower the bystander to make an informed decision when to interrupt or not, such that they can observe the task and judge it against their own urgency. This is similar to a real world scenario where people carefully decide when to interrupt, based on the intensity of a conversation and their own needs.

In contrast to prior work, we evaluate how a real world bystander can interact with an HMD user, without technological aids and a common collaborative goal. As VR and AR provide different levels of immersion and presence in the real and virtual world, with varying capabilities for gesture interaction [20], it is necessary to understand whether the technology has an effect on how bystanders interrupt in a co-located collaborative setting. Our assumption is that bystanders feel more comfortable to interrupt an AR user only partially immersed in the virtual world but fully aware of the real world and thus, is visually aware of bystanders' location and possible motion.

When and How to Interrupt?

Prior research repeatedly confirmed that interruptions negatively affect productivity and performance [36, 50]. However, they are also an integral part of teamwork and necessary for successful collaborations between the person being interrupted and the initiator [54, 60]. Thus, research in this area has been trying to find balanced solutions for enabling interruptions, whilst minimizing the negative effect on users.

A well-known approach is to support interruptions during task switches, when attention shifts from one task to another. Prior research found this to be an opportune moment in which interruption cost is reduced [38, 52, 68]. Additionally, low levels of mental workload have been used as an identifier for appropriate interruption times [2, 37].

Several solutions have been discussed on how to recognize perfect moments for interruptions, such as task switches. The majority focused on internal, mainly posteriori approaches, such as observing physical activity indices through machine learning (e.g., key strokes, motion detector) [5, 17, 18, 36] and their visualization (e.g., traffic light system [82]).

Horowitz et al. [33] inferred interruptibility from recording external factors (e.g., phone being used, door shut), shifting the focus from only enabling computers to interrupt humans (e.g., PC notifications) to humans interrupting humans.

Without the aid of technology, humans rely on social conventions to judge whether interrupting another person is appropriate by observing body orientation, head direction, or gaze [28, 55]. Avrahami et al. reviewed human estimation of interruptibility, investigating self- and bystander estimation [4]. Estimation was mainly based on contextual information (e.g., social engagement, phone, door closed/open). Rivera et al. found that for a human to judge when to best interrupt, they have to know the task and person beforehand [59].

An HMD user can only partially see the real world and, therefore, cannot intentionally communicate their availability. Furthermore, the work setup, whereby the AVR user is wearing an HMD, may be compared to the notion of having the doors closed in an office setting, indicating that you are shutting yourself out from the outside world [35], thus, creating a difficult environment for a bystander to infer the appropriateness of interruptions based on social clues.

Although observing physical activity, such as key strokes, to indicate task switches may be applicable to AVR [71], we argue that it is crucial for the successful interaction between HMD user and bystander if this can be achieved without technology aid but by human judgement only. On one hand, HMDs are becoming self-contained and opportunities to visualize appropriate moments for interruption on a screen nearby may be unavailable. On the other hand, it would enable natural interaction between HMD user and bystander, reducing the need for training on computer-aided visualizations showing interruptibility (traffic light system).

Observability of Gestures

Gestures are an integral part of human communication and have early on been adopted by HCI. Of particular interest is the classification of gestures by Kendon and Wexelblatt who suggest to differentiate between 'conversational gestures' and 'gesture languages' [40, 78]. The latter describes gestures not natural to our technology-independent conversation but rather movements, enforced to interact with certain interfaces. Gesture interaction in AVR mainly falls into this category unless explicitly used for social interactions.

Kurtenbach et al. [43] define gestures as communication tool. In contrast, we investigate the idea that controller clicks and movements in a 360 degree AVR interface are also forms of communication. Rather than focusing on the intentional gesture communication with an AVR user [72], we investigate the involuntary communication to real-world bystanders.

Prior work pointed out that bystanders may indeed act as unintentional communication partners and that they are able to understand gestures by observation. Kendon et al. investigated what type of data can be gathered from gesture observation (e.g., watching someone give a talk, replaying a video of an interaction) and found that, even without speech, bystanders are able to understand the context of an action, emotional state of the speaker, and individual tasks [23, 41]. Heath and Luff [29] go a step further and suggest that controlled gestures may support co-present collaboration. Also, 'gesture languages' are specifically designed to be easily learned or discovered [10].

We argue that for collaboration with HMDs, gestures can never be for the system only, as real world bystanders can be assumed to watch, whether it is intentionally or unintentionally. Intentional bystanders, who want to collaborate with the HMD user, can benefit from gesture cues for the purpose of interruptions. Unintentional bystanders, may benefit by creating awareness around the meaning of the system gestures to support a co-located collaborative environment. Montero et al. [53] reviewed social acceptability of gestures and found that bystanders who cannot interpret user gestures, associate negative feelings with the user/device.

Mid-air gestures are widely used for interaction in other domains, such as public displays and televisions [74, 77]. Yet, research on bystanders observing mid-air gestures as well as the advantages and disadvantages of this is still scarce, [3, 8, 73, 75], particularly in the context of AVR. George et al. [22] reviewed how bystanders observed and interpreted gestures executed in the context of authentication in VR with the goal to identify passwords. Denning et al. [14] investigated the interest of bystanders in AR interactions and found that they were concerned about being filmed by the user's device rather than concerned with the interaction itself.

We build upon prior research by investigating to which extent HMD interaction gestures can be observed by bystanders and whether it is possible to understand the HMD 'gesture language' to a point where it is possible to identify task switches. We believe this knowledge to be a useful starting point for researchers designing approaches to minimize negative influences resulting from interruptions in AVR.

Summary

From related work we learn that interruptions during task switches can reduce the cost for interruptions. However, it is unclear whether observing HMD users' gestures to identify task switches, leads to the same results. Understanding (1) whether users can identify task switches (2) strategies bystanders employ to do so and (3) the effect technology (AR vs. VR) has on these results, is valuable. This knowledge can be taken into account when designing novel UIs in AVR in a way such that they support identifying task switches.

IDENTIFYING COMMON TASKS IN AVR

To generalize our results, we identified five common tasks in AVR with varying interactions (see below). They are based on two streams of analyses: Firstly, we completed a literature review of AVR papers and summarized the most common tasks and their implementation. Secondly, we reviewed existing AVR use cases from the industry. We then transformed the tasks from the use cases to general tasks for our study [2].

Watch

A frequently occurring, passive AVR interaction is the replay of simulations, prominent in automotive [76] and architecture [67] settings. This may be compared to the general task of watching a movie (e.g. fast forward simulation).

Type

Although typing may be performed better in the real word, the task switching cost and loss of presence and immersion such a change introduces, motivated us to include a writing task. Note taking, e.g., during collaborative mechanical prototyping [81], is already a typical AVR application.

Read

Reading instructions, for example, to understand client requirements while creating landscapes/buildings, is already a task that is partially being completed in AVR [12, 7].

Authenticate

The wearer of an HMD has to authenticate to use the HMD or individual apps in AVR. Although currently not common, we included this task as it is probably the most frequently used task on protected mobile devices and HMDs are envisioned to be detached ubiquitous [22, 80].

Manipulate

HMD users manipulate 3D objects in AVR (scaling, rotating and dragging) during prototyping [19, 32].

An overview and detailed description of the developed designs for the above mentioned tasks is provided in Table 1. We designed them so as to require a similar amount of time to be solved across tasks and HMDs. Due to the interactive capabilities of the two HMDs, the designs differ slightly.

STUDY

We conducted a lab study to explore how bystanders in the real world interrupt an HMD user and how easily they could observe their interaction gestures. Studies adhered to ethical research standards within our institution.

Our research was guided by the following hypotheses:

RQ1 How do bystanders interrupt HMD users?

RQ2 Are bystanders capable of identifying opportune moments for interrupting an HMD user?

[H1] Bystanders are able to identify switches between tasks when observing the behaviour of an HMD user.

[H2] Bystanders can identify the type of individual tasks by observing gestures of the HMD user.

[2] We do not claim that these are all the available tasks in AVR. However, they cover a vast amount of possible gestures.

Table 1. We identified five common tasks from prior work and industry. The designs of the tasks for our study were completed in close alignment with these examples. Figure 1 shows screenshots of the virtual tasks.

Task	Design of Tasks for Study	Examples for VR	Examples for AR
Watch	Users watched a short movie on a virtual canvas (4:3, 1.5 m distance to screen). Interactive elements change colour upon activation. Task was completed as the movie ended.	Traffic simulation for driving [76], video for visualizing travelling destination [26].	Replay of demos [42], video conference [47], demos of architecture prototypes [67]
Type	Users had to write a short pre-defined text using a xylophone -like virtual keyboard, where buttons are pressed by means of 'drumsticks' attached to the controllers. The text was displayed on top of the keyboard	Keyboard is presented in different levels of realism [51], Xylophone-like keyboard that is operated with a laserbeam from the controller [1].	Envisioning typing with point and click on virtual keybaord [27, 58],
Read	The user had to read a short text, presented on a pane floating in mid-air. After reading the text, users press 'OK'.	Varying reading in VR between fixed position and always in field of view [7, 12, 61],	hide and show annotations in AR glasses [11, 39, 62, 65]
Authenticate	User had to enter a given PIN correctly and submit it.	PIN entry with laserbeam from controller [22, 80] and from HMD [70], randomized PIN entry with touch from controller [64]	Entering PIN through point and click on PIN pad [34, 58]
Manipulate	The user had to rotate, scale and move the 3D object of a house. Each interaction was performed once.	Inclusion of RW 3D objects manipulated with controller [51], Selection of 3D objects through controller or gaze laser beam[49]	3D production assembly [19, 32, 57].

Study Design

Independent Variables

To evaluate how bystanders interrupt an HMD user and understand how easily different tasks could be observed, we introduced five types of tasks ($task_{\text{types}}$) explained in the previous section. The independent, counterbalanced, within-subjects variable $task_{\text{types}}$ has five conditions: $task_{\text{auth}}$, $task_{\text{read}}$, $task_{\text{watch}}$, $task_{\text{man}}$ and $task_{\text{write}}$ (Figure 1). We also differentiate between two types of *HMDs*, namely AR and VR (between subjects variable).

The study was separated into two parts, whereby the second part had two rounds. In the first part, participants had to interrupt an HMD user. In the second part, participants had to recognize switches between tasks and identify the type of *tasks*. This was done in two rounds ($round_{\text{one}}$ and $round_{\text{two}}$), to analyse learning effects.

Dependent Variables

To investigate interruption behaviour, we captured two types of data:

[Quantitative Data] We measured the time it took participants to regognize a task swith ($rec\_time_{\text{taskswitch}}$) and to identify the task the user switched to (rec_time_{task}). To do so, we measured the time from the start of a new task by the HMD user until participants indicated to have recognized a task switch or identified the task. To measure the time, participants had to tell the experimenter when they identified a task / task switch. The experimenter used a tablet app, on which he pressed a button upon the notification by the participant. We deliberately decided to give the tablet to the experimenter rather than the bystander to avoid disctraction.

In addition, we counted all correct ($guess_{\text{correct}}$) and incorrect guesses ($guess_{\text{wrong}}$), that is the total number of times they said to have noticed the identity of a task (i.e. the total number of button presses on the tablet app). A guess is correct when the task was correctly identified while the HMD user was performing it. Incorrect guesses are cases in which participants indicated to have observed a task which was not the one perfomed by the HMD user. This also includes missed guesses, i.e. when a task was performed by the HMD user but nothing was voiced by the bystander.

Note, that the number of guesses can be unlimited. To calculate the success rate, we only consider the first button press per task ($task_{\text{correct}}$). In addition, we also calculate the error rate ($task_{\text{wrong}}$), which is the sum of incorrectly identified tasks, including tasks were nothing was voiced.

[Qualitative Data] Participants completed (a) a demographic questionnaire and (b) a questionnaire at the end of both part I and II that assessed the effects the interruption had on both participants as well as the difficulty and confidence to observe tasks as perceived by the bystander. Feedback was gathered through a semi-structured interview.

Apparatus

We used an HTC Vive headset with accessories (Figure 2) and a Holo Lens on an i7-6700K PC with an Nvidia GTX980 graphics card as hardware. The VR prototype was created using Unity 3D with support from the Valve SteamVR plugin. To capture observer input, a customized web app ran on the experimenter's iPad Air. This was synchronized with the virtual environments. The app records timings.

Procedure

Participants were recruited in pairs of two. They knew each other and were of equal hierarchy - same level of status rather than supervisor and junior. We refer to the technology under the general term HMD, as the procedure was the same for AR and VR.

Figure 2. Real world view of the HMD user with an HTC Vive, controllers and headphones. The bystander, interrupts the HMD user, identifies tasks and task switches.

Two experimenters conducted the study. One (E1) always stayed with the VR user while the second one (E2) attended to the bystander. We used two rooms, as the individual participants had different tasks and thus needed separate introductions. The main part of the study was completed in the main room (MR; Figure 2) where the HMD and cameras to record the study were setup. The second room (R2) was smaller and equipped with a table and a video camera. In general, both participants stayed in the main room and the bystander only left the second room to complete questionnaires.

Introduction

As participants arrived at the lab, we first introduced them to the topic of the study. In particular, they were then told that they had to take on the role of co-located architects, working together on a project, yet were not told about the interruptions. This setting was chosen based on related work that highlighted a trend in the architecture [9, 15, 32] and automotive industry, where HMDs are already in use (cf. Table 1). We showed them the video cameras and asked them to complete a consent form. After that they were given the opportunity to familiarize themselves with the HMD. In particular, they completed a training, which included working on a number of tasks, closely resembling the ones we would later use in the observation study. Participants were taking turns to complete the initial training and the demographics questionnaire in a separate room. We then randomly assigned one participant to be the HMD user and the other one to be the bystander.

The main part of the study was split into two parts, whereby the second part consisted of two rounds.

Part I

The bystander left the main room with E2. In R2, they were told that their task was to observe their partner while interacting with the HMD and to interrupt them whenever they wanted to. In the meanwhile, E1 told the HMD user to complete a set of 5 tasks [1], while their partner would watch. Once E2 returned with the bystander into MR, the session started and the HMD user put on the HMD. The bystander was first guided through the room and was then asked to choose a spot anywhere in the room to interrupt the AVR user. The session was completed as soon as the bystander interrupted the HMD user. Subsequently, E2 left the room with the bystander to complete a follow-up questionnaire to capture subjective feedback on the interruption. Similarly, E1 asked the HMD user to fill out a questionnaire to collect feedback on how the interruption as perceived. The aim of the first part was to understand interruption strategies, why participants chose a specific point in time, and perception of the interruption.

After completing the questionnaires, participants were provided with instructions for Part II. The HMD user was again asked to complete 5 tasks in VR. The bystander was told to observe the HMD user's gestures and communicate when they notice task switches and the identity of the tasks. This is the first time in the study that they were familiarized with the term 'task switch'.

Part II

Once experimenters and participants were united again in MR, the HMD user put on the HMD ($round_{\text{one}}$). E1 recorded bystanders' feedback on task switches and task identities on the tablet. This round finished when all 5 tasks were completed by the HMD user[1]. Then, E2 went to R2 with the bystander for a follow-up structured interview, where they state their confidence in their recognized tasks and task switches and answer additional questions on observation strategies. Afterwards, we repeated this part ($round_{\text{two}}$).

Conclusion

After the main parts, both experimenters and participants sat together in MR to have a semi-structured interview on the study and their experience as a pair. Participants were compensated with an Amazon voucher. Each session lasted approximately 45 minutes.

Participants

Overall, we had 62 participants (29 females, avg. age = 27, SD = 4.4), 30 had corrected to normal eyesight, and one stated a red-green colour blindness. Participants rated their previous experience with VR with MD=3 and for AR it was MD=1 (7 point likert scale: 1=very bad, 7=very good) (cf. Table 6).

Limitations

We recruited via a university mailing list. Hence, the majority of participants were young but had little HMD experience. Yet, they represent the target group we are designing for. The study was completed in the western culture with high-end HMDs. Thus, findings may only be applicable to users with a similar background and setting.

RESULTS

Results are based on log files from the tablet app, questionnaires, a semi-structured interview and video captures. Note, numbers are aggregated across both AR and VR condition, unless there were significant differences that are specifically pointed out.

[1] Note, it was communicated to both participants that tasks are random and could appear once or multiple times.

Figure 3. Total count of recognitions and guesses for AR vs. VR for identifying $task_{\text{types}}$.

Quantitative Findings

Participant provided 378 guesses (total button presses) while they were observing 320 tasks (32×2×5).

Interruptions

First we looked at the time it took participants to interrupt the user and the employed strategies. A Welch ANOVA showed that technology had a significant effect on the time it took observers to interrupt, ($F_{1,27} = 16.33$, $p < 0.001$). The time it took the bystander to interrupt the VR user ($Mean = 57.18$, $SD = 39.57$) was significantly longer than for the AR user ($Mean = 17.13$, $SD = 13.83$). 66% of bystanders naturally interrupted while the AVR user was completing a task rather than in-between tasks.

The vast majority of bystanders decided to *speak* to the AVR user to interrupt (62.5%). The others used the following strategies: 18.8% *speak & touch*, 9.4% *touch*, 3.1% *wave* (AR only). 6.3% did not interrupt (VR only). In the semi-structured interview, we questioned these bystanders on the reason for their behaviour, considering that the instruction they were given for Part I was specifically to interrupt the HMD user. They revealed to have felt uncomfortable about interrupting someone who is deeply engaged with another device.

Recognition Accuracy and Time for Task Switches

First, VR participants *correctly* identified 107/125 task switches (16×2×4, upon review of the data in a box-plot diagram, we removed 3 outliers), resulting in a 86% success rate ($round_{\text{one}}$: 85% & $round_{\text{two}}$ 86%). AR results were similar: 100/125 switches identified, 80% accuracy ($round_{\text{one}}$: 79.7% & $round_{\text{two}}$ 80.3%).

Note, that in 143 cases, participants did not communicate a task switch (84 in VR and 59 in AR) but later correctly identified the task. As we noticed this phenomenon during the study, we asked participants during the final interview. They stated to have forgotten about stating the task swich and primarily focused on identifying tasks. We had to exclude these cases for calculating the time required to identify a task switch and only considered them for calculating the recognition accuracy. We found that in 83% of the cases, participants were able to correctly identify a task switch.

Second, we looked at *how quickly* participants could detect task switches. This took participants on average 5 s. They were quicker in the AR ($Mean = 4.94$, $SD = 4.01$) condition compared to VR ($Mean = 5.11$, $SD = 3.71$).

Summary.

From this we learn, that upon directing to look out for task switches (a) participants are indeed able to correctly identify switches of tasks (83% correctly identified $task_{\text{switches}}$) and (b) that they can do so before AVR users deeply start engaging with the new task.

Recognition Accuracy and Time for Task Recognitions

Again, we first looked at *how accurately* participants could identify tasks (Figure 3). Participants correctly identified 77% of all tasks: 75 % for AR ($round_{\text{one}}$: 71% & $round_{\text{two}}$ 79%) and 80% for VR ($round_{\text{one}}$: 80% & $round_{\text{two}}$ 80%). At the same time, participants were wrong about 71 tasks (23%). This includes 33 cases where tasks were missed. Further analysis revealed that the type of task had a significant influence on whether or not it was identified correctly, $\chi^2(4) = 14.267, p = 0.05$. Cramer-V shows a strong relationship ($\geq$ 0.3).

Authentication was often mistaken for tasks that had a very similar observable behaviour. For example in AR, users were typing characters onto a virtual canvas, which had the same size in both tasks (PIN pad vs. keyboard). For an observer, this results in the same mid-air motion, within the same size of 3D space. In comparison, the xylophone-like writing task in VR seemed to have a very unique observable behaviour, resulting in a 100% recognition accuracy. This phenomenon is described in more detail in Table 2. It shows a matrix of tasks completed by the HMD user (lines) and the corresponding guesses made by bystanders (columns).

Second, we looked at *how quickly* participants could identify tasks (Table 3). Overall, bystanders took 7 s (SD=1.6), whereby they were quicker when observing AR ($Mean = 6.63$, $SD = 1.0$) than VR ($Mean = 7.81$, $SD = 1.92$) gestures.

For AR, bystanders were able to identify $task_{\text{read}}$ and $task_{\text{watch}}$ quicker than the other tasks in $round_{\text{one}}$ (Table 3). However, they show a significantly lower success rate than the other

Table 2. A matrix of tasks (lines) and their corresponding guesses (columns). $guess_{\text{correct}}$ **is highlighted in green. For example,** $task_{\text{auth}}$ **was often mistaken for** $task_{\text{man}}$ **(10) in VR and for** $task_{\text{write}}$ **(9) in AR.**

	VR						AR						
	auth	*read*	*man*	*type*	*watch*	sum	*auth*	*read*	*man*	*type*	*watch*	oth	sum
auth	24	0	10	1	4	39	20	1	0	9	0	1	31
read	0	24	0	0	2	26	0	20	0	0	7	0	27
manip	0	0	33	2	4	39	1	0	60	2	0	1	64
type	0	0	0	40	0	40	5	2	2	33	1	1	44
watch	0	2	0	0	25	27	5	5	0	1	27	3	41
sum	24	26	43	43	35	171	31	28	62	45	35	6	207

tasks and comparably it took participants longer to recognize these task in $round_{\text{two}}$, whereas for all other tasks, bystanders show a learning effect. Reading and watching a video require consistent eye movements that can be tracked quickly through AR glasses, however, the eye movements are the same for both tasks, which result in a nearly 50% chance of guessing correctly between both tasks (Table 2).

Analysis of the same tasks for the VR condition reveals that bystanders had similar difficulties with the reading and watching task. Although there was a learning effect visible for both tasks, they were also mistaken for each other. Yet, the success rates for recognition are above 70% and better than for AR. Qualitative data – mentioned in the next section – suggests that, in comparison with AR, this may be due to the larger virtual canvas that was provided in VR, which resulted in more prominent observable gestures (e.g. head movements for reading lines of text) for the bystander.

Summary

In general it is possible to determine task types AVR users switch to. However, the time required strongly depends on the task type: if there are tasks leading to similar observable behaviour or require subtle interactions, as in our case the AR reading and watching tasks, this increases recognition times.

Qualitative Results

Qualitative data was gathered pre-, mid- and post-study through questionnaires and an unstructured interview. Data from the interview was coded by two experimenters.

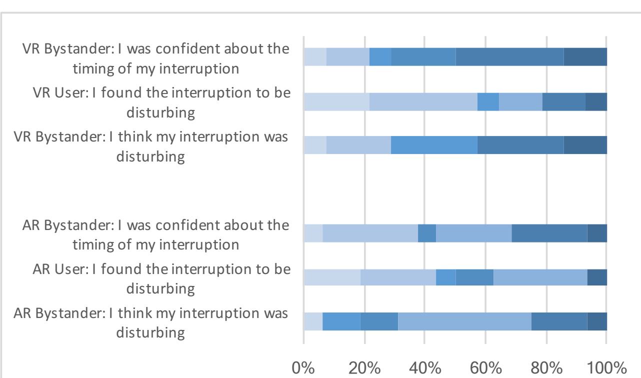

Figure 4. On a 7-point Likert-scale (1=not - 7 = very much), for AR (MD=5), the bystander perceived their interruptions to be more disturbing than for VR (MD=3).

Table 3. The left column shows the overall success rate per task type. The other columns summarize recognition times for each round (mean in s).

	rate	$round_{\text{one}}$	$round_{\text{two}}$	overall
AR				
auth	55%	8.50 (SD=5.2)	7.09 (SD=3.5)	7.67 (SD=4.3)
read	56%	5.19 (SD=2.9)	6.42 (SD=4.4)	5.80 (SD=3.8)
manip	97%	7.58 (SD=6.0)	3.52 (SD=2.6)	5.69 (SD=5.2)
write	90%	7.36 (SD=5.1)	5.16 (SD=3.4)	6.26 (SD=4.4)
watch	75%	6.04 (SD=3.1)	9.21 (SD=7.4)	7.76 (SD=6.0)
VR				
auth	69%	4.94 (SD=1.3)	5.04 (SD=1.6)	4.99 (SD=1.5)
read	71%	9.95 (SD=3.5)	8.14 (SD=2.2)	9.05 (SD=3.0)
manip	87%	10.2 (SD=4.7)	9.2 (SD=5.8)	9.68 (SD=5.3)
write	100%	7.62 (SD=3.5)	5.88 (SD=2.6)	6.75 (SD=3.2)
watch	72%	8.94 (SD=5.8)	8.3 (SD=4.3)	8.62 (SD=5.2)

Interruptions

Participants rated their perceptions of how disturbing the interruptions were on a 7-point Likert scale (1=not disturbing at all; 7= very disturbing). In AR, users did not feel disturbed when they were interrupted (MD=2). However, observers thought their interruptions was most likely disturbing (MD=5). In VR, users also did not feel disturbed when interrupted (MD=2), but contrary to AR, observers rated their interruptions as less disturbing (MD=3) - see Figure 4.

Altough subjective feedback from the questionnaire suggests that participants were not disturbed, 68% of HMD users mentioned in the unstructured interview that they would have preferred to be interrupted during the change of a task. P6 states that it would be good if bystanders "[...] wait until I am done with a task" (P6). 18% mentioned "uncritical tasks" as an opportune moment for interruptions. 25% of HMD users highlighted the low cognitive load that was required to complete the tasks, which they also stated as a reason for why they did not perceive the interruption to be disturbing. "[completion of the tasks] was not so bad, I could easily have a conversation [while I am going through them]" (P23). 31% said that the interruptions were found to be less disturbing as they knew the bystander. "I trusted the voice" (P11). In AR, 25% of HMD users said the interruption could have been less disturbing if the bystander was in their field of view during interruption rather than behind them.

Task Switches and Tasks

On a 7-point Likert scale (1=very bad, 7=very good), participants rated their overall recognition accuracy as good (MD=5) and they perceived observing the task switch and recognizing the type of task to be rather difficult (MD=3). Self-assessed recognition accuracy and perceived difficulty significantly correlated in the second round, $rs(32) = 0.6$ $(p < 0.05)$. Although recognition accuracy stayed high (Med=5), perceived difficulty increased, indicating that participants were better able to assess the difficulty of the task type whilst maintaining a good understanding of their own recognition skills (Table 4). Similarly, although not significant, the data suggests a training effect between the rounds (i.e. recognitions were higher $round_{\text{two}}$ compared to $round_{\text{one}}$.)

A Pearson product-moment correlation coefficient also revealed a negative correlation $rs(32) = -0.59$, $(p < 0.05)$ between total success rate and perceived difficulty (Table 5).

Table 4. **Results (Median) from post-round questionnaires showing an overview of perceived difficulty and self-assessed confidence for recognizing task switches and task types.**

	Round 1		Round 2	
	difficulty	confidence	difficulty	confidence
$task_{types}$	3.0	5.0	3.5	5.0
$task_{switches}$	2.0	5.0	3.0	5.0

Participants are able to judge themselves whether they are capable of identifying a task and a task switch correctly, by considering the difficulty of the observation. This was only apparent in $round_{two}$.

In $round_{one}$, 36% of participants indicated that the best position for observing task switches is from the diagonal front or side. This number increased in $round_{two}$ to 81%. In particular, 90% of participants reported to have focused on the HMD users' hand movements, 34% also looked at head movements and 5 participants mentioned looking at the eyes in AR (Figure 2). Only one participant paid attention to possible sounds from the headphones.

Participants also commented on what they thought would improve their recognition success:

Emphasizing gestures. 50% of participants pointed out that there should be unique movements to point out individual tasks. "[...] larger differences between the gestures would have helped [with the gesture recognition]" (P4). As an example, the VR writing task was mentioned by three of these participants. "[...] as soon as I recognized a unique gesture, I was sure [...]" (P6).

Differentiating tasks. A quarter of participants pointed out that the reading and watching task were particularly difficult to observe and differentiate. They mentioned that it would have helped if these tasks ended/started differently, such that they could at least know when the task switch was happening. "[...] special gestures at the beginning of a task [...]" (P2). This is also reflected in the quantitative data (Table 2) where this confusion of tasks became apparent.

Sound. Two participants acting as the bystander mentioned that the sound from the headphones could be made louder, such that bystanders can also listen in.

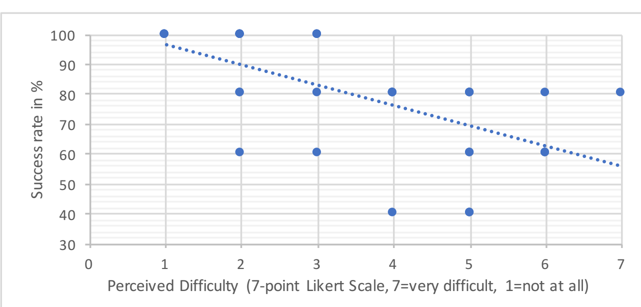

Figure 5. A Pearson product-moment correlation coefficient revealed a negative correlation rs(32) = -0.59, $(p < 0.05)$ **between total success rate and perceived difficulty in** $round_{two}$**.**

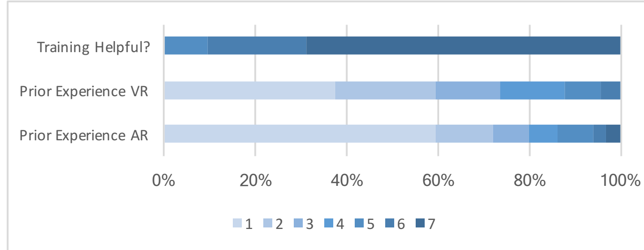

Figure 6. On a 7-point Likert scale (7=very, 1=no) the majority of participants agreed (Med=7) post-study that the training helped in successfully recognizing tasks and task switches. More than half of the participants had very little or no prior experience with AVR technology (Med=1).

Change of head/body position. 21% of the bystanders mentioned that a change in head direction and/or body position might support the recognition of a task switch.

A Pearson product-moment correlation coefficient showed no relationship between prior AVR experience and success rate for recognizing switches, rs(32) = -0.2, $(p > 0.05)$ (Figure 6).

DISCUSSION AND IMPLICATIONS

In the following we discuss interesting observations and point out design opportunities as well as directions for future research, where appropriate.

Interruptions are (Not) Technology Dependent

At first glance, our data suggests that interruption behaviour is technology dependent. VR observers interrupted after 57 s while AR observers already interrupted after 17 s. This is accentuated by the behaviour of 2 VR observers, who refused to interrupt in order not to disturb an HMD user engaged in their task, while not seeing the observer. Yet, the picture is more diverse.

"Feeling Skilled" Reduces Technology Effects

Our analysis shows that people more quickly interrupt as they become more confident in their 'interruption skills', leading to smaller differences concerning observation behaviour between the two technologies (6.6 s in AR vs. 7.8 s in VR). Overall, it seems that two aspects contribute to feeling skilled: becoming more experienced (this effect was apparent in the second round) and becoming more confident in identifying the correct moment in time an interruption occurs. One factor that influences how quickly people interrupt and that we did not investigate, is the urgency of the interruption. It would be interesting to see whether findings from prior work in related domains apply in the context of VR [30, 31], especially as bystanders wanting to interrupt do not have visual cues to support their decision making process, whre they evaluate theirtheir own urgency against the impact of the interruption on the user.

Repetition Improved Interruption Behaviour

Bystanders' self-assessed ability to judge the difficulty of the observed tasks and task switches improved in the second round for both technologies, resulting in more accurate and quicker guesses for the majority of tasks that were uniquely identifiable. Bystanders also became more skilled with their strategies, such that they chose better positions to improve their observability and focused on further gesture cues (e.g. not just hands but also head movements and body position).

How the Technology Affected Observations

Overall, participants were quicker to identify tasks and task switches in AR. However, they did this more accurately for the VR condition. This seems to be influenced by the hardware differences. AR affords subtle interactions that are within a limited field of view, whereas VR also enables larger, more prominent gestures. The latter enables bystanders to more easily identify unique gestures, therefore improving accuracy.

The Necessity to Focus on the Bystander

Our findings suggest that future work should direct more attention towards the bystander.

Generating Awareness on Interruption Location

We assumed AR interruptions to be perceived as less disturbing by both bystander and user as AR HMDs allow parallel viewing of the real and virtual world – unlike VR, where participants are secluded from the real-world. Interestingly, only one participant took advantage of this difference and interrupted the AR user by waving and trying to get attention by standing in front of them. We expected, the majority of participants to interrupt by catching AR users' visual attention. Still, they chose to interrupt through speech, independent of the technology. Qualitative results revealed that eye contact was made after interruption rather than before, suggesting that it was possible to obtain social cues through the glasses. Adding this feedback to AR users' preference to see the bystander when being interrupted might create awareness on appropriate locations for interruptions, and could hence be included as part of the training, when introducing HMDs.

Supporting Quicker, More Accurate Interruptions

For both technologies, initial feedback from participants suggests, that HMD users did not find interruption as disturbing as we expected. However, during the interview, the majority of participants pointed out that they prefer to be interrupted during a task switch or during an "uncritial task". The ambiguity in their perception of the interruption is twofold: Firstly, participants' familiarity with each other seems to have decreased negative effects interruptions may have, which is common in work settings. Secondly, they pointed out the low workload that was required to complete the tasks, which made being interrupted less disturbing. This confirms prior work [2, 13, 37], where the person being interrupted is at the focus of the research. However, our work highlights that situations in which interruptions occur can be further enhanced by supporting bystanders, e.g., to make quicker more accurate interruptions during task switches, rather than during tasks.

Notably, prior work has investigated a vast number of solutions to support quicker, more accurate interruptions, thereby focusing on both sides: (a) the user – for example, the display of interruptions in a 3D space [21], and (b) the bystander – for example, traffic light feature to aid decision making [82]. It is possible to repurpose the HMD device's surface (e.g. on the straps) to display such a traffic light system for the bystander. Yet, this might create additional overhead in a co-located context and be unnecessary due to people's capability to naturally analyse situations based on gestural cues. The effects of such a technological mediation for interruptions in AVR settings needs to be investigated in future studies.

Supporting Collaboration by Addressing Bystanders

Analysis of bystanders' perceptions reveal that they were concerned about the effect the interruption has on the user. Interestingly, this was more prominent in the AR setting than in VR which may be a result of the AR users unintentional feedback through their eye movements, immediately giving away that they stopped the interaction. Bystanders concerns about the negative effect their interruption has, support our motivation to not only focus on the HMD users when designing a collaborative AVR experience but also on the bystander. Awareness on when and how to interrupt and familiarity with the tasks improves gesture recognition and thus the ability to successfully determine opportune moments for interruptions. However, bystanders may not always know the tasks an HMD user performs. This can be the case in environments where HMD users perform very specialized tasks or as new tasks find their way into VR. Our observation that prior experience in AVR did not influence the success rate suggest that bystanders may identify the task switch in such cases, in particular if the new task requires different gestures. An in-depth investigation of such cases should be subject to future work.

Design Implications to Improve Observability

Introducing HMDs to Mixed Collaborative Settings

When introducing an HMD to a collaborative setting, both user and bystander should be familiarized with the device interactions. This empowers the bystander to interrupt effectively by recognizing tasks and task switches (both within 7s), rather than interrupting during task completion - which is what they would naturally do according to our findings.

Distinguishing Different Gestures

Participants stated that unique gestures led to a better recognition of individual tasks (for example, position of hands in relation to each other, typing gesture). This confirms previous work by Rivera et al. [59] who found that prior knowledge of people and tasks being observed provides the best case scenario for observation success. At the same time, gestures that were similar in multiple tasks (for example, authentication and manipulation) allowed participants to identify task switches, but not the particular task.

To support bystanders, designers could, firstly, increase distinguishability by pronouncing subtle cues. For example, in our study reading and watching were very similar, but few participants recognized subtle head movements. Hence, designers could, for example, slightly scale up the text a user is reading. As a result, users are likely to perform more recognizable head movements as they read and switch between the lines. An interesting question here is to strive a balance between being easily observable while still being usable and ergonomic (i.e. excessive head movements may cause fatigue). Second, easy-to-recognize, task-specific gestures could be supported. For example, starting the playback of a video could be realized with a simple swipe gesture instead of a button click if those were used in another similar task.

Adapting Real-World Conventions

In congruity with prior work [33, 35] and well known design principles, such as Nielsens heuristics [56], designers could adapt real-world conventions when designing AVR interactions. Prominent gestures may be used to provide social cues to the bystander, such as opening and walking through a door to join a conference/presentation in AVR. Similarly, head/body position may be used to show deep engagement to the bystander: For example, by displaying high-workload tasks at the bottom of the FOV, AVR users would be guided to look down for interaction. Another approach could be to embed common signs from sign language (e.g., stop sign) into gesture interaction with HMDs, which provide clear clues to the bystander.

Exploiting Spatial Layout of the Environment

It was also interesting to see that participants sometimes based their decision for guessing the task type based on head direction and/or body position in the real-world. This creates two interesting design opportunities. On one hand, systems could be designed in a way, such that virtual tasks in which it is ok to interrupt (for example, reading emails), would always be performed in the same location, such as a virtual desk, mapped to a constant real-world location. On the other hand, it creates an opportunity for HMD users themselves to control, whether or not they would be ok with being interrupted. For example, HMD users could decide to read email displayed in a canvas floating in mid-air in front of them (and thus blend with the current task) – similar to reading email on a smartphone on the go – or change their location and read emails at their desk.

Designing Non-Observable Task Switches

Note that there are cases in which it is not desirable that observers can easily identify task switches. Although the success rate was only 69%, the authentication task was recognized fastest in VR but performed the worst with regards to both recognition accuracy and time in AR. Although the tasks were designed in a very similar way, the latter HMD provides a smaller field of view for interaction, which seems to provoke sublte interactions and decrease bystanders ability to recognize the tasks. This is important from a usable security perspective, since identifying this type of task switch may allows observers to determine the beginning of users entering a password and hence perform a shoulder surfing attack [22]. For developers this means that in order to decrease observability, subtle interaction in a limited FOV should be designed. To address this, future work could investigate, whether AR interaction methods for authentication can be transferred to VR settings and whether this improves security.

Similarly, we believe that the decision to hide information and work in a "private" mode, for example, during authentication, should be an option for the AVR user and/or interactions should consciously be designed to be unobservable by a bystander. For example, in the context of VR, where location and space can be easily tracked, interactions can be designed such that the user is facing a "private real-world wall". This would hide their input from bystanders without changing the experience itself for the AVR User. Tekin and Reeves [69] found that fully or partially hiding user gestures also affects bystanders perception in public spaces. However, this needs to be re-evaluated in the context of co-located collaboration with HMDs.

CONCLUSION AND FUTURE WORK

In this paper, we investigated how bystanders interrupt an HMD user and whether they are able to recognize tasks and task switches solely through observing HMD users' gestures. Task switches were found to be an opportune moment for interruptions in previous research and to our knowledge, this paper takes the first step towards applying these findings into the context of HMDs. Results from our study indicate that observers are able to correctly guess task switches 83% of the time [*H1*] and that they are able to identify the type of tasks 77% of the time [*H2*], both within 7 s.

To our knowledge, this is the first paper on interruption behaviour and observability of gesture interaction in HMD settings. This ground research enables to do further studies on the differences between external (through gesture recongition and naturally) vs internal interruptions - from a user and bystander perspective. Additionally, future work may review whether bystanders' ability to retrieve meaning from AVR users' gesture interactions, has a positive effect on social acceptability [53] of HMD devices.

Dabbish et al. reviewed the workload interrupters require [13]. This may be applied to our work, to determine workload for gesture recognition. We decided to explore interruptions between collaborators that are aware of the tasks at hand, it would be interesting whether tasks can be recognized by bystanders without priming (e.g. training prior to observation). This may allow passive bystanders to join a collaborative group of HMD and non-HMD users, without prior knowledge of the system. Similarly, we are planning to do further studies with more tasks and sub-tasks to understand how the type of task may influence interruption and gesture recognition. Our work benefits designers and practitioners who want to introduce HMDs to co-located collaborative settings.

REFERENCES

[1] 2016. (May 2016). https://developers.googleblog.com/2016/05/daydream-labs-exploring-and-sharing-vrs.html

[2] Piotr D. Adamczyk and Brian P. Bailey. 2004. If Not Now, when?: The Effects of Interruption at Different Moments Within Task Execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 271–278. DOI: http://dx.doi.org/10.1145/985692.985727

[3] Ilhan Aslan, Andreas Uhl, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Mid-air authentication gestures: an exploration of authentication based on palm and finger motions. In *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM, 311–318.

[4] Daniel Avrahami, James Fogarty, and Scott E. Hudson. 2007. Biases in Human Estimation of Interruptibility: Effects and Implications for Practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 50–60. DOI: http://dx.doi.org/10.1145/1240624.1240632

[5] Abraham Bernstein, Peter Vorburger, and Patrice Egger. 2008. A scenario-based approach for direct interruptability prediction on wearable devices. *International Journal of Pervasive Computing and Communications* 3, 4 (2008), 426–438.

[6] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2001. The MagicBook: a transitional AR interface. *Computers & Graphics* 25, 5 (2001), 745–753.

[7] Tobias Blum, Valerie Kleeberger, Christoph Bichlmeier, and Nassir Navab. 2012. mirracle: An augmented reality magic mirror system for anatomy education. In *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*. IEEE, 115–116.

[8] Miss Bhavana Borkar, Miss Shiba Sheikh, and PD Kaware. 2016. 4d password mechanism. *Imperial Journal of Interdisciplinary Research* 2, 5 (2016).

[9] Daniel BROSCHART and Peter ZEILE. ARchitecture–Augmented Reality in Architecture and Urban Planning. (????).

[10] Erica A Cartmill, Özlem Ece Demir, and Susan Goldin-Meadow. 2012. 14 Studying Gesture. (2012).

[11] Yun Suk Chang, Benjamin Nuernberger, Bo Luan, Tobias Höllerer, and John O'Donovan. 2017. Gesture-based augmented reality annotation. In *Virtual Reality (VR), 2017 IEEE*. IEEE, 469–470.

[12] Jian Chen, Pardha S Pyla, and Doug A Bowman. 2004. Testbed evaluation of navigation and text display techniques in an information-rich virtual environment. In *Virtual Reality, 2004. Proceedings. IEEE*. IEEE, 181–289.

[13] Laura Dabbish and Robert E. Kraut. 2004. Controlling Interruptions: Awareness Displays and Social Motivation for Coordination. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, New York, NY, USA, 182–191. DOI:http://dx.doi.org/10.1145/1031607.1031638

[14] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2377–2386.

[15] Isabelle Doll, Karsten Rexroth, and Petra Von Both. 2017. Augmented Reality in der Stadt: Neue Potenziale durch die Entwicklung einer Lichtmarker. *Wireless Commun.* 6 (2017), 65–75.

[16] Joshua A Fisher, Amit Garg, Karan Pratap Singh, and Wesley Wang. 2017. Designing intentional impossible spaces in virtual reality narratives: A case study. In *Virtual Reality (VR), 2017 IEEE*. IEEE, 379–380.

[17] James Fogarty and Scott E. Hudson. 2007. Toolkit Support for Developing and Deploying Sensor-based Statistical Models of Human Situations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 135–144. DOI: http://dx.doi.org/10.1145/1240624.1240645

[18] James Fogarty, Scott E. Hudson, and Jennifer Lai. 2004. Examining the Robustness of Sensor-based Statistical Models of Human Interruptibility. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 207–214. DOI: http://dx.doi.org/10.1145/985692.985719

[19] Markus Funk, Mareike Kritzler, and Florian Michahelles. 2017. HoloCollab: a shared virtual platform for physical assembly training using spatially-aware head-mounted displays. In *Proceedings of the Seventh International Conference on the Internet of Things*. ACM, 19.

[20] Nirit Gavish, Teresa Gutiérrez, Sabine Webel, Jorge Rodríguez, Matteo Peveri, Uli Bockholt, and Franco Tecchia. 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments* 23, 6 (2015), 778–798.

[21] Ceenu George, Manuel Demmler, and Heinrich Hussmann. 2018. Intelligent Interruptions for IVR: Investigating the Interplay between Presence, Workload and Attention. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW511.

[22] Ceenu Goerge, Mohamed Khamis, Emanuel von Zezschwitz, Marinus Burger, Henri Schmidt, Florian Alt, and Heinrich Hussmann. 2017. Seamless and Secure VR: Adapting and Evaluating Established Authentication Systems for Virtual Reality. In *Proceedings of the Network and Distributed System Security Symposium (NDSS 2017) (USEC '17)*. Internet Society. DOI: http://dx.doi.org/10.14722/usec.2017.23028

[23] Jean Ann Graham and Michael Argyle. 1975. A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology* 10, 1 (1975), 57–67.

[24] Jan Gugenheimer, Evgeny Stemasov, Julian Frommel, and Enrico Rukzio. 2017. ShareVR: Enabling Co-Located Experiences for Virtual Reality Between HMD and Non-HMD Users. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4021–4033. DOI: http://dx.doi.org/10.1145/3025453.3025683

[25] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. FaceDisplay: Towards Asymmetric Multi-User Interaction for Nomadic Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 54, 13 pages. DOI: http://dx.doi.org/10.1145/3173574.3173628

[26] Daniel A Guttentag. 2010. Virtual reality: Applications and implications for tourism. *Tourism Management* 31, 5 (2010), 637–651.

[27] Mohammad S Hasan and Hongnian Yu. 2017. Innovative developments in HCI and future trends. *International Journal of Automation and Computing* 14, 1 (2017), 10–20.

[28] Mary Jo Hatch. 1987. Physical barriers, task characteristics, and interaction activity in research and development firms. *Administrative Science Quarterly* (1987), 387–399.

[29] Christian Heath and Paul Luff. 1993. Disembodied conduct: Interactional asymmetries in video-mediated communication. *Technology in working order: Studies of work, interaction, and technology* (1993), 35–54.

[30] Chih-Yuan Ho, Mark I Nikolic, Molly J Waters, and Nadine B Sarter. 2004. Not now! Supporting interruption management by indicating the modality and urgency of pending tasks. *Human Factors* 46, 3 (2004), 399–409.

[31] Joyce Ho and Stephen S Intille. 2005. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 909–918.

[32] Paul Hockett and Tim Ingleby. 2016. Augmented reality with HoloLens: Experiential architectures embedded in the real world. *arXiv preprint arXiv:1610.04281* (2016).

[33] Eric Horvitz and Johnson Apacible. 2003. Learning and Reasoning About Interruption. In *Proceedings of the 5th International Conference on Multimodal Interfaces (ICMI '03)*. ACM, New York, NY, USA, 20–27. DOI: http://dx.doi.org/10.1145/958432.958440

[34] Yichao Huang, Xiaorui Liu, Xin Zhang, and Lianwen Jin. 2016. A pointing gesture based egocentric interaction system: Dataset, approach and application. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 16–23.

[35] James M. Hudson, Jim Christensen, Wendy A. Kellogg, and Thomas Erickson. 2002. "I'D Be Overwhelmed, but It's Just One More Thing to Do": Availability and Interruption in Research Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. ACM, New York, NY, USA, 97–104. DOI: http://dx.doi.org/10.1145/503376.503394

[36] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 257–264. DOI: http://dx.doi.org/10.1145/642611.642657

[37] Shamsi T. Iqbal and Brian P. Bailey. 2005. Investigating the Effectiveness of Mental Workload As a Predictor of Opportune Moments for Interruption. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1489–1492. DOI: http://dx.doi.org/10.1145/1056808.1056948

[38] Shamsi T. Iqbal and Brian P. Bailey. 2006. Leveraging Characteristics of Task Structure to Predict the Cost of Interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 741–750. DOI: http://dx.doi.org/10.1145/1124772.1124882

[39] J. Adam Jones, J. Edward Swan, II, Gurjot Singh, and Stephen R. Ellis. 2011. Peripheral Visual Information and Its Effect on Distance Judgments in Virtual and Augmented Environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization (APGV '11)*. ACM, New York, NY, USA, 29–36.

[40] Adam Kendon. 1986. Current issues in the study of gesture. *The biological foundations of gestures: Motor and semiotic aspects* 1 (1986), 23–47.

[41] Adam Kendon. 1994. Introduction to the special issue: gesture and understanding in interaction. *Research on Language and Social Interaction* 27, 3 (1994), 171–173.

[42] Andrew L Kun and Christian P Janssen. 2017. Calling While Driving : An Initial Experiment with HoloLens. *Driving Assessment Conference University* (2017), 200–206.

[43] Gordon Kurtenbach. 1990. Gestures in human-computer communication. *The Art of Human-Computer Interface Design, B.* (1990).

[44] Glyn Lawson, Davide Salanitri, and Brian Waterfield. 2016. Future directions for the development of virtual reality within an automotive manufacturer. *Applied ergonomics* 53 (2016), 323–330.

[45] Morgan Le Chénéchal, Thierry Duval, Valérie Gouranton, Jérôme Royan, and Bruno Arnaldi. 2016. Vishnu: Virtual immersive support for helping users-an interaction paradigm for collaborative remote guiding in mixed reality. In *3DCVE 2016: International Workshop on Collaborative Virtual Environments*. IEEE, 1–5.

[46] Matthew Lombard, Frank Biocca, Jonathan Freeman, Wijnand IJsselsteijn, and Rachel J. Schaevitz. 2015. *Immersed in Media: Telepresence Theory, Measurement & Technology*. Springer Publishing Company, Incorporated.

[47] Aidong Lu, Jian Huang, Shaoting Zhang, Chuang Wang, and Weichao Wang. 2016. Towards mobile immersive analysis: A study of applications. In *Immersive Analytics (IA), 2016 Workshop on*. IEEE, 25–30.

[48] Christian Mai, Lukas Rambold, and Mohamed Khamis. 2017. TransparentHMD: Revealing the HMD User's Face to Bystanders. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (MUM '17)*. ACM, New York, NY, USA, 6. DOI:http://dx.doi.org/10.1145/3152832.3157813

[49] Tomasz Mazuryk and Michael Gervautz. 1996. Virtual reality-history, applications, technology and future. (1996).

[50] Mark A McDaniel, Gilles O Einstein, Thomas Graham, and Erica Rall. 2004. Delaying execution of intentions: Overcoming the costs of interruptions. *Applied Cognitive Psychology* 18, 5 (2004), 533–547.

[51] Mark McGill, Daniel Boland, Roderick Murray-Smith, and Stephen Brewster. 2015. A Dose of Reality: Overcoming Usability Challenges in VR Head-Mounted Displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2143–2152. DOI:http://dx.doi.org/10.1145/2702123.2702382

[52] Yoshiro Miyata and Donald A Norman. 1986. Psychological issues in support of multiple activities. *User centered system design: New perspectives on human-computer interaction* (1986), 265–284.

[53] Calkin S Montero, Jason Alexander, Mark T Marshall, and Sriram Subramanian. 2010. Would you do that?: understanding social acceptance of gestural interfaces. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*. ACM, 275–278.

[54] Martin Muhlenbrock, Oliver Brdiczka, Dave Snowdon, and J-L Meunier. 2004. Learning to detect user activity and availability from a variety of sensor data. In *Pervasive Computing and Communications, 2004. PerCom 2004. Proceedings of the Second IEEE Annual Conference on*. IEEE, 13–22.

[55] Darren Newtson. 1973. Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology* 28, 1 (1973), 28.

[56] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 249–256.

[57] Kim O'Connell. 16.11.2016. 4 Tipps fÃijr erste Anwendungen von Virtual Reality in der Architektur. https://www.autodesk.de/redshift/virtual-reality-in-architecture/. (16.11.2016).

[58] Noushin Radnia, Alireza Karduni, Isaac Cho, Zachary Wartell, and Eric Sauda. 2017. Augmented : Reality : : Architecture : Interface.

[59] A Joy Rivera. 2014. A socio-technical systems approach to studying interruptions: Understanding the interrupter's perspective. *Applied ergonomics* 45, 3 (2014), 747–756.

[60] AJ Rivera-Rodriguez and Ben-Tzion Karsh. 2010. Interruptions and distractions in healthcare: review and reappraisal. *Quality and Safety in Health Care* 19, 4 (2010), 304–312.

[61] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. 2018a. Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 45, 9 pages. DOI: http://dx.doi.org/10.1145/3173574.3173619

[62] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. 2018b. *Reading on Smart Glasses: The Effect of Text Position, Presentation Type and Walking*. ACM, New York, NY, USA. 45:1–45:9 pages.

[63] Maria V Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nat Rev Neurosci* 6, 4 (2005), 332–339.

[64] Paul Sawers. 2017, Lastchecked: 2018-09-10. Worldpay demos system for making payments inside virtual reality worlds. (May 2017, Lastchecked: 2018-09-10). https://venturebeat.com/2017/05/25/ worldpay-demos-system-for-making-payments-inside-virtual-reality

[65] Jeffrey H. Shuhaiber. 2004. Augmented reality in surgery. *Archives of Surgery* 139, 2 (2004), 170–174.

[66] Mel Slater, Vasilis Linakis, Martin Usoh, Rob Kooper, and Gower Street. 1996. Immersion, presence, and performance in virtual environments: An experiment with tri-dimensional chess. In *ACM virtual reality software and technology (VRST)*, Vol. 163. ACM Press New York, NY, 72.

[67] Søren S. Sørensen. 2006. The development of Augmented Reality as a tool in Architectural and Urban design. *Nordic Association for Architectural Research* 19, 4 (2006), 8.

[68] Takahiro Tanaka and Kinya Fujita. 2011. Study of User Interruptibility Estimation Based on Focused Application Switching. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*. ACM, New York, NY, USA, 721–724. DOI:http://dx.doi.org/10.1145/1958824.1958954

[69] Burak S Tekin and Stuart Reeves. 2017. Ways of spectating: Unravelling spectator participation in Kinect play. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 1558–1570.

[70] Welderufael B. Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, and Jetzabel Serna. 2018. PrivacyGuide: Towards an Implementation of the EU GDPR on Internet Privacy Policy Evaluation. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics (IWSPA '18)*. ACM, New York, NY, USA, 15–21. DOI: http://dx.doi.org/10.1145/3180445.3180447

[71] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruptibility Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 801–812. DOI: http://dx.doi.org/10.1145/2750858.2807514

[72] Kaisa Väänänen and Klaus Böhm. 1993. Gesture driven interaction as a human factor in virtual environments-an approach with neural networks. *Virtual reality systems* (1993), 93–106.

[73] Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: Sparking Civic Discourse by a Public Interactive Poll Visualization. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 1323–1332. DOI: http://dx.doi.org/10.1145/2531602.2531639

[74] Radu-Daniel Vatavu. 2012. User-defined gestures for free-hand TV control. In *Proceedings of the 10th European conference on Interactive tv and video*. ACM, 45–48.

[75] Radu-Daniel Vatavu. 2015. Audience silhouettes: peripheral awareness of synchronous audience kinesics for social television. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 13–22.

[76] Marcel Walch, Julian Frommel, Katja Rogers, Felix Schüssel, Philipp Hock, David Dobbelstein, and Michael Weber. 2017. Evaluating VR Driving Simulation from a Player Experience Perspective. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2982–2989. DOI: http://dx.doi.org/10.1145/3027063.3053202

[77] Robert Walter, Gilles Bailly, Nina Valkanova, and Jörg Müller. 2014. Cuenesics: using mid-air gestures to select items on interactive public displays. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*. ACM, 299–308.

[78] Alan Wexelblat. 1997. Research challenges in gesture: Open issues and unsolved problems. In *International Gesture Workshop*. Springer, 1–11.

[79] Bob G Witmer and Michael J Singer. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments* 7, 3 (1998), 225–240.

[80] Z. Yu, H. Liang, C. Fleming, and K. L. Man. 2016. An exploration of usable authentication mechanisms for virtual reality systems. In *2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. 458–460. DOI: http://dx.doi.org/10.1109/APCCAS.2016.7804002

[81] Farbod Zorriassatine, Catherine Wykes, Robert Parkin, and Nabil Gindy. 2003. A survey of virtual prototyping techniques for mechanical product development. *Proceedings of the institution of mechanical engineers, Part B: Journal of engineering manufacture* 217, 4 (2003), 513–530.

[82] Manuela Züger, Christopher Corley, André N. Meyer, Boyang Li, Thomas Fritz, David Shepherd, Vinay Augustine, Patrick Francis, Nicholas Kraft, and Will Snipes. 2017. Reducing Interruptions at Work: A Large-Scale Field Study of FlowLight. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 61–72. DOI: http://dx.doi.org/10.1145/3025453.3025662