# Intelligent Interruptions for IVR: Investigating the Interplay between Presence, Workload and Attention

**Ceenu George**
LMU
Munich, Germany
ceenu.george@ifi.lmu.de

**Manuel Demmler**
LMU
Munich, Germany
demmler@cip.ifi.lmu.de

**Heinrich Hussmann**
LMU
Munich, Germany
hussmann@ifi.lmu.de


*CHI'18 Extended Abstracts*, April 21–26, 2018, Montreal, QC, Canada
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188686

## Abstract

Whereas interruptions is a very active subfield of research within HCI, as of today interruptions in immersive virtual reality (IVR) have received only little attention. We conducted a lab study ($N$=20) with a head mounted display (HMD) to understand the relationship between presence, workload and attention in IVR when measuring three virtual interruption designs. The answer to this question is interesting because prior research has revealed a positive effect on performance when providing intelligent interruptions, for example based on users' level of attention. Our work launches research on interruptibility in IVR by investigating (1) the relationship between attention, presence and workload, and the (2) methods for measuring them in IVR. Our analysis suggests that a trade-off between presence and attention is required when designing interruptions for IVR. Our findings are valuable for researchers and practitioners who want to collect data on attention, presence and workload in IVR to inform interruptibility.

## Author Keywords

VR; DRT Task; Interruptions

## ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: Multimedia Information Systems - Artificial, augmented, and virtual realities


**Figure 1:** We measure attention, presence and workload on three interruption designs within two different scenes.

## Introduction

Interruptions have been found to disrupt the task at hand, leading to a loss of overall performance [1, 12]. In the context of IVR there is the added negative effect on the users' perception of presence and immersion. Although there exist varying definitions, both terms measure the quality of any virtual reality (VR) experience. Presence describes a subjective evaluation of how the virtual world is perceived [9] and immersion is an objective evaluation of the technological capabilities (e.g. display size, tracking accuracy).

Past work has revealed that there exist opportune moments, in which the cost of an interruption is reduced [18]. They provide evidence that there are advantages of designing intelligent interruptions by targeting moments with low workload and attention [4]. Although there exist well established methods to measure these human variables in other domains, such as NASA TLX [6] and iGroupPresence questionnaire [16], they have not been sufficiently analysed in the context of IVR with HMD's. Additionally, correlations and causalities between these variables, although highly discussed in real world (RW) literature [7], often remain unexplored for IVR. Prior work on the interplay between presence, workload and attention has focused on virtual environments [19, 16], however there exists no current evidence for their appropriateness in IVR with HMDs.

We iteratively designed and developed a prototype that allows the measurement of attention (through detection response tasks (DRT)), presence and workload (through questionnaires) in IVR, while exposed to three virtual interruption designs across two different scenes (Fig. 1).

We conducted an exploratory lab study ($N = 20$) with our prototype, which revealed that a trade-off between presence and attention (acceptance of VR vs. fast reaction time) is required when designing interruptions for IVR.

## Related Work

*Presence and Immersion in IVR*

Witmer and Singer [19] define presence as a tool to measure the "level of realism" compared to the RW. If presence is high, their so called 'looming effect' becomes apparent, whereby users react in the physical world to actions that are perceived in the virtual world. For example, placing physical controllers on a virtual table or avoiding moving virtual cars by stepping aside in the RW [17]. The original list of factors affecting presence also includes 'immeadiacy of control', 'multimodal presentation' [17] and 'broad focus' [5]. These factors were used as requirements for the creation of the scenes for the study prototype.

The most adopted way to measure presence is through subjective evaluations. In prior work presence was found to be "something that people can be consciously aware of" and therefore report on in a post-study questionnaire or through verbal feedback [10]. Oftentimes, these questionnaires probe the awareness of the RW [19], acknowledging that RW interruptions decrease presence. Contrary to prior work, we focus on virtual interruptions, displayed in VR. A summary of alternative measurements can be found in [10]. The most relevant alternative measure for the purpose of this paper include findings from [2] who found a negative correlation between secondary reaction time and immersion. We don't explicitly measure immersion, as the latest HMDs (e.g. HTC Vive) already offer a high level of immersion, but rather focus on the subjective measure of presence by using a post-study (IPQ) questionnaire [16].

*Attention and Workload in Relation to Presence*

A high level of attention results in a high level of presence [19]. Fontaine [5] also highlights the importance for a 'broad focus', thus increased levels of presence are achieved when the VR scene is perceived as a whole rather than

**Apparatus**
The prototype was created with Unity and the study was conducted with a HTC Vive.

**Scenes:** We created two main tasks (Fig. 1), *scene*$_{\text{bow}}$[a] and *scene*$_{\text{museum}}$. *scene*$_{\text{bow}}$ is an archery game with targets that disappear after being hit or 15 s. Target location was random and participants had to walk around the virtual scene to navigate between the targets. *scene*$_{\text{museum}}$ was a virtual room with 4 paintings on the wall. Virtual paintings were random for each iteration and from well-known artists (e.g. Monet)- all having the same size and quality.

**Interruptions:** Designs were informed by Nielsen et al. [13] taxonomy: A non-diegetic *task*$_{\text{text}}$, that appears horizontally centered as an overlay in the virtual scene. A local-diegetic *task*$_{\text{spotlight}}$, designed to be part of the virtual scene, appearing either to the right or left within the field of view (FOV). A global-diegetic *task*$_{\text{ambient}}$, which changed the global light setting gradually for a few seconds. All notifications appeared within the current FOV of the user and were displayed in the color red [3].

[a]extension of multiple archery scenes available online under GNU General Public License v3.0

when attention is focused on single interactions. The relationship between cognitive load and presence was found to correlate, such that an increase in cognitive load also affects presence positively [15]. Laarni et al. [9] write about presence being a "multidimensional construct", similar to cognitive load and [7] refer to cognitive load as a "subset of attention". We argue that the terms under investigation, namely cognitive load, attention and presence are deeply interrelated. Therefore, this paper measures attention with the same priority as presence. The study presented in this paper follows a DRT design [20], based on the secondary-task methodology [7], to investigate the differences between main and secondary task attention in detail.

## Limitations
We used a university mailing list for recruiting participants. Although they are not solely students, they tended to be below 30 years of age. Furthermore, we put importance on playfulness of task rather than variance in cognition. We will review this for future work.

## Study
We conducted a within-subjects lab study ($N$=20), pursuing a DRT design methodology [20]. The study adhered to ethical research standards within our institution.

*Independent Variables*
*main_task* allows participants to experience two virtual scenes that have different workloads. We refer to *scene*$_{\text{bow}}$ as the active scene with a high workload (e.g. stressing), as users are constantly engaged with a bow and arrow shooting exercise. In contrast, *scene*$_{\text{museum}}$ has a lower workload (e.g. relaxing), as participants were merely observing paintings. Workload for both tasks was informed by a pilot study. *secondary_task* explores the impact of three virtual interruption designs on a task level.

*Dependent Variables*
Reaction time (in sec) measured the time it took participants to react to the *secondary_task* (DRT design). The time was tracked from the display of the *secondary_task* until the participant pressed a button on the physical HTC Vive controller to indicate that they saw it. We also measured workload with a NASA TLX [6] and presence with an igroup presence questionnaire (IPQ) [16]. *shortquestionnaire* was an adapted presence questionnaire based on IPQ and Witmer et al. [19]. We rephrased questions to focus on the notification designs rather than the overall scene and excluded scene-only related questions. Table 1 provides an overview of methodologies used.

**Table 1:** Methodologies used for scene vs. interruption.

| | Scene | Interruption |
|---|---|---|
| **Presence** | IPQ | *shortquestionnaire* |
| **Attention** | | DRT (Reaction time) |
| **Workload** | NASA TLX | |

*Procedure*
After an initial introduction, participants had to provide written consent before being introduced to the basics of the HTC Vive. The study was split into two consecutive sessions based on the independent variable *main_task*, whereby each one followed the same pattern according to the DRT methodology:

First, participants had to carry out a training with the aim to adjust with the handling of the controllers and the task. To complete the training, all participants had to hit the target three times in a row in *scene*$_{\text{bow}}$ and mark all paintings as "viewed" in *scene*$_{\text{museum}}$. The latter was achieved by pointing at a painting and confirming via button click on the controller. Then, the baseline measurements were captured by asking participants to complete the primary task for each *main_task* without an interruption being shown. Next, the primary task was completed three more times, to capture data on all three interruptions. Each interruption was repeated randomly within three pre-defined intervals to reduce guessing effects from participants. Every repetition of the primary task lasted for 90s and was followed by *shortquestionnaire*, such that participants had to complete both three times in total for each *main_task*. Questions from *shortquestionnaire* were asked out lout by the experimenter while the participant was immersed in VR. This was motivated by the need to maintain presence in VR without extending the length of the study.

**Participants**
Likert scale results (1=not at all, 7=very) showed that participants (N=20, female=12, age=23.3 (SD=4.4)) perceived themselves to be moderately adept to technology (Mdn=5). 13 participants wore glasses but had no other issues with eye sight. They did not wear them in the HMD and communicated no drawbacks when inquired. Previous experience with RW archery and VR was low (Mdn=2) and none of the participants reported any cyber sickness.

Each session, $scene_{\text{bow}}$ and $scene_{\text{museum}}$, was concluded with a NASA TLX and an IPQ. These were filled out by the participant without wearing the HMD. To exclude order effects both *main_task* and *secondary_task* were counterbalanced for all participants. Finally, the study was concluded with demographics questions and a standardized Simulation Sickness Questionnaire [8] questionnaire.

## Results

Data was normally distributed and there were two outliers.

*Presence*

Results from the IPQ revealed a significant difference $(p < 0.05)$ depending on scene, such that a lower level of presence was perceived for $scene_{\text{bow}}(M = 3.7, SD = 0.54)$ compared to $scene_{\text{museum}}(M = 3.2, SD = 0.67)$. A repeated measures ANOVA on *shortquestionnaire* evaluating presence for each *secondary_task* showed significant effects ($F_{2,112} = 7.39, p < 0.001$), such that a lower level of presence was perceived for $task_{\text{text}}(M = 3.4, SD = 0.18)$, compared to $task_{\text{spotlight}}(M = 3.7, SD = 0.14)$ and $task_{\text{ambient}}(M = 3.9, SD = 0.14)$.

*Attention*

A repeated measures ANOVA on reaction time revealed a significant effect on *secondary_task* design ($F_{2,36} = 3.48$, $p < 0.05$). Missed notifications ($task_{\text{spotlight}}$ =4, $task_{\text{text}}$ =1) were captured with a max reaction time of 5s. To exclude these outliers, analysis was based on the best out of three reaction times. Participants reacted to $task_{\text{text}}(M = 0.72, SD = 0.056)$ significantly quicker than to $task_{\text{spotlight}}(M = 0.79, SD = 0.048)$ and $task_{\text{ambient}}(M = 0.78, SD = 0.046)$. We could not confirm any dual-task costs when comparing the baseline condition (without interruptions) with any of the other conditions (with interruptions).

*Workload*

$scene_{\text{bow}}$ scored an average of 56.46 on the NASA-TLX scale, while $scene_{\text{museum}}$ was rated significantly lower with 29.58. A Spearman's rank-order correlation found a moderate positive correlation between presence & workload in $scene_{\text{bow}}$, which was statistically significant ($r_s$ = .396, p = .047, n=18).

## Discussion and Future Work

*Presence Questionnaire*

The IPQ questionnaire revealed significant differences on a scene and task level. We argue that there is a lack of standards for evaluating individual changes (e.g. interruptions), rather than evaluating the whole virtual scene. Although our adapted presence questionnaire and its placement in the study design is only a first step, we believe that it provides a possible approach for differentiating between the overall scene and individual changes within the same scene.

*DRT Design*

Reaction times were found to be significantly different between the virtual interruptions. DRT design was found to be an appropriate methodology for measuring attention in IVR, as it is embedded into the virtual experience rather than disruptive, such as a questionnaire. This allows the experimenter to maintain a continuous level of presence, rather than the time-consuming task of waiting for presence to be achieved again before continuing with the study.

*Interplay between Presence, Workload and Attention*
Although previous research indicates relationships between these human variables, our results only found a significant correlation between presence and workload for $scene_{\text{bow}}$, confirming previous work by Ma & Kaber [11]. Subjective feedback revealed that participants perceived virtual text interruptions to be affecting their level of presence more negatively, compared to ambient and spotlight. However, they reacted to text interruptions significantly quicker than the other two virtual designs, indicating that less attention is needed to respond to embedded text rather than for designs within the virtual scene. Similar results were previously confirmed with RW interruptions [14, 19], however future work may evaluate whether virtual interruptions affect presence less than RW interruptions.

Based on the results, we believe it is necessary to track both variables separately to inform interruption designs in IVR. Depending on the use case, one variable may be more important than the other. For example, alerting the immersed IVR user of someone approaching in the RW may be seen as important enough to use a virtual text notification, thus provoking a faster reaction from the VR user whilst also trading it off for a decrease in presence.

## Conclusion
Our preliminary findings provide insights into the effect of three different virtual interruption designs on attention and presence. Results suggest that designers need to compromise on a trade-off between both variables, such that important interruptions may require high attention designs (e.g. text), leading to a loss of presence and vice versa.

## REFERENCES

1. Ernesto Arroyo and Ted Selker. 2003. Self-adaptive Multimodal-interruption Interfaces. In *Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI '03)*. ACM, New York, NY, USA, 6. DOI: **http://dx.doi.org/10.1145/604045.604051**

2. Cheryl Campanella Bracken, Gary Pettey, and Mu Wu. 2014. Revisiting the use of secondary task reaction time measures in telepresence research: exploring the role of immersion and attention. *AI & SOCIETY* 29, 4 (2014), 533–538. DOI: **http://dx.doi.org/10.1007/s00146-013-0494-7**

3. Nilgün Camgöz, Cengiz Yener, and Dilek Güvenç. 2004. Effects of hue, saturation, and brightness: Part 2: Attention. *Color Research & Application* 29, 1 (2004).

4. Daniel Chen, Jamie Hart, and Roel Vertegaal. 2007. Towards a physiological model of user interruptability. In *IFIP Conference on Human-Computer Interaction*. Springer, 439–451.

5. Gary Fontaine. 1992. The experience of a sense of presence in intercultural and international encounters. *Presence: Teleoperators & Virtual Environments* 1, 4 (1992), 482–490.

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

7. Barry H Kantowitz. 2000. Attention and mental workload. In *Proceedings of the Human Factors and*

*Ergonomics Society Annual Meeting*, Vol. 44. SAGE Publications, 3–456.

8. Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

9. Jari Laarni, Niklas Ravaja, Timo Saari, Saskia Böcking, Tilo Hartmann, and Holger Schramm. 2015. Ways to Measure Spatial Presence: Review and Future Directions. In *Immersed in Media*. Springer, 139–185.

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

11. Ruiqi Ma and David B Kaber. 2006. Presence, workload and performance effects of synthetic environment design factors. *International Journal of Human-Computer Studies* 64, 6 (2006), 541–552.

12. 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.

13. Lasse T Nielsen, Matias B Møller, Sune D Hartmeyer, Troels Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 229–232.

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

15. Corina Sas and Gregory MP O'Hare. 2003. Presence equation: An investigation into cognitive factors underlying presence. *Presence: Teleoperators and Virtual Environments* 12, 5 (2003), 523–537.

16. Thomas W Schubert. 2003. The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness. *Zeitschrift für Medienpsychologie* 15, 2 (2003), 69–71.

17. 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 VR software and technology (VRST)*, Vol. 163. ACM Press New York.

18. 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, 4. DOI: **http://dx.doi.org/10.1145/1958824.1958954**

19. 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.

20. Richard A Young, Li Hsieh, and Sean Seaman. 2013. The tactile detection response task: preliminary validation for measuring the attentional effects of cognitive load. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. 71–77.