Who is in the Room? Notification for Intrusions While in Virtual Reality

Aaron Crowson, Zachary H. Pugh, Michael Wilkinson, Christopher B. Mayhorn
North Carolina State University

The development of head-mounted display virtual reality systems (e.g., Oculus Rift, HTC Vive) has resulted in an increasing need to represent the physical world while immersed in the virtual. Current research has focused on representing static objects in the physical room, but there has been little research into notifying VR users of changes in the environment. This study investigates how different sensory modalities affect noticeability and comprehension of notifications designed to alert head-mounted display users when a person enters his/her area of use. In addition, this study investigates how the use of an orientation type notification aids in perception of alerts that manifest outside a virtual reality users' visual field. Results of a survey indicated that participants perceived the auditory modality as more effective regardless of notification type. An experiment corroborated these findings for the person notifications; however, the visual modality was in practice more effective for orientation notifications.

INTRODUCTION

Head-mounted displays (HMDs) offer a better virtual reality (VR) experience by using two of the most important senses, vision and audition, shown to be important for immersion and presence in the virtual world (Dinh, Walker, Hodges, Song, & Kobayashi 1999). However, by capturing some of our arguably most important senses, individuals lose their perception of events happening in the real world. This can lead to accidents such as running into furniture and hitting walls (Simeone, Velloso, & Gellersen 2015).

Ghosh et al. (2018) investigated common real-world interruptions experienced by VR users and asked them which interpretations should receive a notification. The survey revealed that most users were concerned with notifications about their physical environment. In addition to the survey, Ghosh et al. (2018) tested prototype notification systems. They used visual, auditory, haptic and paired combinations (haptic + visual, audio + visual, audio + haptic) for notifying users that a person was entering the room or talking to them. They found that visual notifications suffered in noticeability. This was potentially due to a lack of cues that indicated where to locate the visual notifcations. Thus, if a notification was not already in the visual field it was likely to be missed. Moreover, haptic notifications, while noticeable, were hard to interpret. Auditory notifications elicited the strongest feelings of urgency. The auditory-visual pairing was rated the most popular for a person entering the play area.

Based on their results, Ghosh et al. (2018) presented a few recommendations for the design of notifications. First, create distinction such that external event notifications can be distinguished from a VR environment/game. Second, visual search should be reduced such that participants do not have to continually scan their visual field to find a notification. Third, precautions should be taken to avoid jump scares that may happen when a notification is instantiated too close to the user's body, specifically for the person-intrusion scenario. Finally, familiar images and sounds should be used, such as footsteps to indicate a person's movement in the play area.

While these recommendations may be important for VR environments involving few distractors (i.e. non-game VR programs), some are harder to implement in the context of gaming. For example, footstep sounds that are meant to indicate a person-intrusion may be misinterpreted as noise generated by elements within the game itself.

VR provides unique notification development challenges because events can happen anywhere around the user such that typical guidelines for 2D displays become harder to implement (Milgram & Kishino 1994). However, standard first person 2D display video games have been successful in alerting users to hazards behind them through various elements (Rogers 2014).

One such element is an orientation notification. Orientation notifications are notifiers that indicate the direction of important information that is occurring outside the user's visual field. Typical orientation notifiers in video games include taking damage from another player off screen or navigating the user's character towards an in-game objective.

Based on the success these notifiers have on alerting players in video games (Burigat & Chittaro 2007), it is theorized that orientation notifications may help inform VR users of the locations of real-world intrusions. It is understood that if a VR game is already using such a notifier for in game content, the notification for external stimulus must be made distinct. With these design recommendations in mind, the present study extends the work of Ghosh et al. (2018) in developing and testing external notification systems in VR.

The Present Study

The present study expands the literature on notification effectiveness for representing external stimuli in a virtual environment. Specifically, this study investigates how the visual and auditory modalities, along with their combination, influence a VR user's perception of a notification's effectiveness in alerting the user to an external stimulus and identification of the stimulus. Additionally, this study explores how visual notifications, auditory notifications, and a combination of those influence the response to and orientation towards an external stimulus entering a VR user's area in a simulated game environment. Finally, this study investigates if users' perceptions of a notification's effectiveness align with their performance towards it.

The study took part in two phases. The first was a survey that assessed participants' perceptions of developed notifications. The notifications were developed based on typical notifiers in first person shooter video games as well as incorporating design considerations postulated for peripheral displays (Matthews 2006), ambient information systems (Tomitsch, Kappel, Lehner, & Grechenig 2007), and auditory displays (Adcock & Barrass 2004). The second phase consisted of an experimental evaluation for effective notifications as indicated by the survey. These notifications took place across auditory only (A), visual only (V), and auditory + visual (AV) modalities combined.

METHOD

Participants

The survey portion recruited 103 participants from Amazon's Mechanical Turk (MTurk) workers. Using Curren's (2016) criterion to address careless responding, 12 participants were excluded as their completion time was more than one standard deviation below mean completion time (*M*= 20.48min, *SD*=11.30min). Ninety-one participants were included in the final analysis. Seven college-aged students were recruited for the experimental phase from a large southeastern university.

Materials

Real-world intrusions notification systems assessment survey. The purpose of the survey was to assess which notifications participants perceived as most effective in capturing their attention and conveying the correct information about orientation to a person entering the room. Thirty notifcations were designed; ten per modality (A, V, AV). Five notifications in each modality were theorized to be effective for representing a person, while the other 5 were anticipated to be effective for helping a user orient to an intrusion. Visual person type notifications included an orb, an outline of a person, an opaque person, footsteps on the ground, and a transparent person wearing a hat, gloves, and shoes. Visual orientation notifications included a textbox alert, an objective marker, a dynamic arrow, a rotating arrow, and a rotating light at the screen's edge. Auditory person type notifications included a door opening, a store door chime, a narrator voice stating a person had entered the room, an alarm stating intruder alert, and phone notification chime. Auditory orientation type notifications included footsteps around the participant, a radar noise that increased in frequency, a radar noise that increase in frequency and amplitude, a heartbeat that increased in amplitude, and an emergency alert tone that increased in amplitude. The combined modality (AV) notifications included a random paring of the notifications from the auditory and visual modalities. Participants received the same questions for all notifications which were Likert items based on a 1-7 scale. The questions were: "How would you rate this notification's effectiveness of grabbing your attention?, Do you believe this notification is effective if it was intended to be an in-game alert (e.g. an enemy or an enemy approaching)?, Do you believe this notification is effective if it was intended to be an alert for something happening in the real world (e.g. your friend or your friend approaching)?". After a notification prototype was presented it was followed by the assessment questions. An example notification can be seen in Figure 1.

Figure 1. *The dynamic arrow is an example of a visual only orientation notification.*

Virtual reality game task. The game was designed according to Sebastian Lague's *Create a Game* tutorial (2015). At the start of the game, the participant began in the center of a 30x30 grid map, with each grid cell either occupied by an obstacle or unoccupied, according to a flood-fill algorithm that ensured complete navigability of the environment. AI "robots" (displayed as floating black capsules) spawned at 1.5-second intervals at random locations on the map and pursued the participant. The robots emitted spatialized sound that allowed participants to localize robots outside their field of view. The participant's task was to destroy the robots using a plasma gun and avoid being attacked.

The participant was able to move around the map using the Oculus system's left joystick. For rotational motion, the participant was instructed to physically turn.

Intruder notifications occurred as varying combinations of modalities: (1) a glowing white sphere with the text 'Person' in front of the sphere (V); (2) a stereo sound auditory notification, comprising a male's voice stating that "a person has entered the room" (A); (3) A combination of 1 and 2 (AV); (4) an emergency alert tone, which played in surround sound from the source of intrusion, allowing for auditory localization (A); (5) a dynamic GUI arrow, which remained pointed toward the intrusion regardless of the participant's orientation (V); (6) the combination of 4 and 5 (AV). As such, nine combinations were assessed; those that included at least one person-classifying notification (1, 2, or 3) and at least one orienting notification (4, 5, or 6).

The game consisted of 162, 30-second rounds. For every set of six rounds, a notification would appear during a randomly selected round. Each participant received each of the nine notification combinations three times. The combination's order of presentation was counterbalanced across participants.

Procedure

During the first phase of the experiment, Mturk participants took the notification systems assessment survey. Results were analyzed and stimuli were developed from the prototype stage for testing.

For the testing phase, participants were instructed to don the Oculus Rift VR HMD and controllers. The participants were presented with a tutorial video describing the controls of the game, their task, and the proper response to intruder notifications. After the tutorial, participants began the game. Participants were instructed that the developed notification systems may attempt to notify them in some manner and to be vigilant. In responding to notifications, participants were told to first pause the game as quickly as possible and then turn to where the notifications indicated. The program recorded the times of several events: intrusion onset, when the player first pressed pause in response to the intrusion, and when the participant located the intrusion. A break was given after completing trials 54 and 108. Participants were reminded of the proper response to a notification before resuming the experiment.

RESULTS

Pre-Test Survey.

Notifications were first assessed on their comprehension. Notifications that had at least 70% agreement among participants as either a person or orientation notification were reclassified. As such, the text box, objective marker, footstep audio, and heartbeat audio were reclassified as person type notifications. The rotating light notification, and rotating arrow notification were rated as ambiguous and were thus included in both person and orientation type notifications. Notifications were then assessed for their perceived effectiveness by summing the three Likert items mentioned previously. One item was reverse coded. It is important to note that for visual notifications only, a score of 0 could be given by a participant due to the question “Is there a notification present in this scene?” If a participant answered “no”, they did not receive the follow-up questions and were advanced to the next item.

Person Notifications.

A repeated measures ANOVA revealed a significant difference in effectiveness score between the visual notifications [$F(8,720) = 47.32$, $p < .001$, $\eta^2 = .35$]. Paired t-tests with Bonferroni corrections revealed that the text box ($M = 13.66$, $SD = 4.27$) and orb ($M = 13.36$, $SD = 3.29$) notifications were perceived to be the most effective but were not significantly different from one another.

A repeated measures ANOVA revealed a significant difference in effectiveness scores between the auditory notifications [$F(6,540) = 18.11$, $p < .001$, $\eta^2 = .17$]. Paired t-tests with Bonferroni corrections revealed that the narrator voice ($M = 15.48$, $SD = 3.35$) was rated as most effective.

A repeated measures ANOVA revealed a significant difference in effectiveness scores between the combined modality notifications [$F(6,540) = 11.83$, $p < .001$, $\eta^2 = .11$]. Specifically, it was revealed through paired t-tests with Bonferroni corrections that the combination of visual footsteps and the audio narrator voice ($M = 15.49$, $SD = 3.20$) was perceived to be the most effective, while the visual objective maker and heartbeat audio ($M = 12.95$, $SD = 2.88$) were rated as the least effective. It is interesting to note the pattern of most and least effective notifications for the combined visual and auditory modalities follows a similar pattern to the auditory modality on its own.

Orientation Notifications.

A repeated measures ANOVA revealed a significant difference in effectiveness scores between the visual orientation notifications [$F(2,180) = 56.53$, $p < .001$, $\eta^2 = .39$]. Specifically, it was revealed through paired t-tests with Bonferroni corrections that the dynamic arrow notification ($M = 11.21$, $SD = 4.55$) was perceived as significantly more effective as an orientation device.

A repeated measures ANOVA revealed a significant difference in effectiveness scores between the auditory modality orientation notifications [$F(3,270) = 21.87$, $p < .001$, $\eta^2 = .20$]. In this case, paired t-test with Bonferroni corrections indicated that participants rated the emergency alert tone ($M = 14.65$, $SD = 2.53$) notification as being significantly more effective for the auditory modality.

A final repeated measures ANOVA was conducted on the combined orientation notifications that revealed a significant difference in effectiveness scores [$F(2,180) = 11.39$, $p < .001$, $\eta^2 = .11$]. Paired t-tests with Bonferroni corrections revealed that the rotating light paired with the audio emergency alert tone was perceived to be the most effective ($M = 13.76$, $SD = 2.62$). It is again interesting to note the pattern of most and least effective notifications for the combined visual and auditory modalities follows the exact pattern to the auditory modality on its own.

For the experimental phase, the orb, narrator voice and their combination were used for person type notifications. The dynamic arrow, emergency alert tone, and their combination was used for the orientation type notifications. Each person notification was paired with every orientation notification. This was done to test the performance of each notification type individually as well as the interaction between person and orientation type notifications.

Accuracy rates were calculated to assess which combinations of person and orientation notifications were the most likely to be missed. Figure 2 shows accuracy rates for all person and orientation combinations utilized in testing.

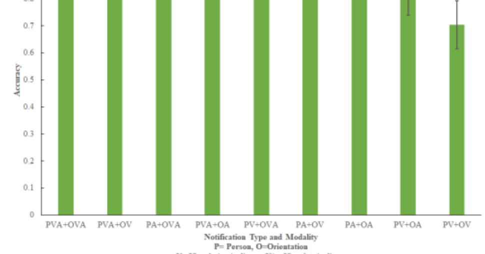

Figure 2. *Accuracy in responding to the various modality combinations of person and orientation notifications.*

Reaction time in responding to notifications was collected for all notification pairs. A two-way repeated measure ANOVA revealed a significant difference in reaction time for the person type notifications [$F(2,12) = 9.83$, $p = .003$, $\eta^2 = .62$]. Specifically, the visual orb notification ($M = 2.24$, $SD = 0.71$) had significantly slower reaction time than the auditory narrator voice ($M = 1.48$, $SD = 0.23$). The visual orb notification also had significantly slower reaction time when compared to the combined modality condition ($M = 1.65$, $SD = 0.57$). These results can be seen in Figure 3 below.

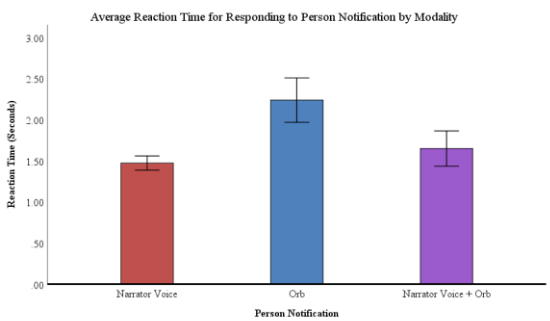

Figure 3. *Reaction time by modality for person notifications.*

A two-way repeated measures ANOVA was also conducted for the orientation type notifications to determine if there was a difference in reaction time between modalities. The ANOVA did reveal a significant difference in reaction time among the visual, auditory and combined modalities for the orientation type notifications [$F(2,12) = 4.87$, $p = .03$, $\eta^2 = .45$]. Interestingly, the pairwise comparisons with Bonferroni corrections did not show any significant difference between the three notifications (narrator voice ($M = 1.74$, $SD = 0.51$), orb ($M = 1.99$, $SD = 0.52$), combined ($M = 1.64$, $SD = 0.46$)).

A final two-way repeated measures ANOVA was conducted to determine if there were any interaction effects between person and orientation type notifications on participant reaction time. Results of the analysis indicated that there was a significant interaction [$F(4,24) = 5.77$, $p = .002$, $\eta^2 = .49$]. When investigating the visual modality for a person type notification (orb) we do see a difference in reaction time between when it is paired with the three orientation notifications. Specifically, when the visual person notification (orb) is paired with the visual orientation notification (dynamic arrow) participants have the slowest reaction time. By contrast, when the visual person notification is paired with the combined visual and auditory orientation notification it has the fastest reaction time. These results are shown in Figure 4.

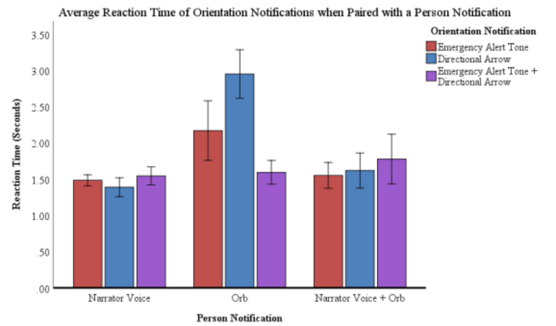

Figure 4. *Reaction time for the interaction between person type notifications and orientation type notifications.*

Dwell time. Dwell time was analyzed for orientation notifications. Dwell time in this case refers to the amount of time it takes for a participant to orient themselves towards the indicated notification. This was operationally defined as the time between when a participant hit the pause button to indicate the presence of a notification and when the headset registered that the participant oriented correctly to the position of the intrusion. A two-way repeated measure ANOVA revealed a significant difference in dwell time among the visual, auditory and combined modalities for the orientation type notifications [$F(2,12) = 7.05$, $p = .009$, $\eta^2 = .54$]. Pairwise comparisons with Bonferroni corrections indicated that the auditory modality ($M = 1.15$, $SD = 0.34$) was significantly slower at orienting a participant to an intrusion when compared to the visual modality ($M = 0.65$, $SD = 0.27$) or the combination ($M = 0.78$, $SD = 0.28$). This suggests that the visual modality conveys the direction to an intrusion better than auditory and that the addition of an auditory component does not increase participant understanding of where to locate that intrusion. These results can be seen in Figure 5.

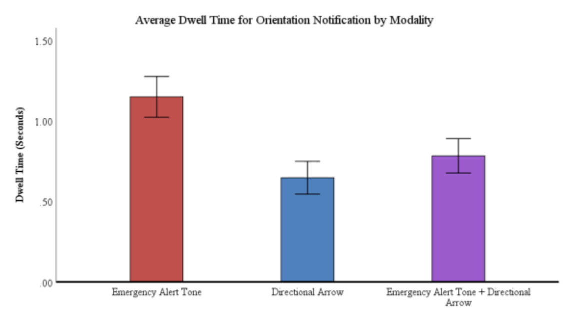

Figure 5. *Dwell time of each modality for orientation type notifications.*

DISCUSSION

Results from the survey phase indicated that the notifications rated as being the most effective for representing a person included the orb for visual notifications and the voice for auditory notifications. Both notifications verbally stated that the source of intrusion was a person. Thus, participants indicated that to correctly interpret a notification, explicit verbalization is preferable to icons and earcons.

Regardless of notification type, the AV modalities combined mirrors the pattern observed within the auditory modality. For example, the notification perceived as most effective within the auditory modality was still perceived as most effective even when paired with a less effective notification from the visual modality. This seems to suggest a bias when evaluating notifications, such that the auditory modality is perceived as more important for correct notification interpretation. The presence of a visual component does not appear to influence a participant's perception of notification effectiveness. This pattern is again observed in the case with the dynamic arrow. The dynamic arrow was rated as being the most effective visual notification but did not receive the same rating when paired with the auditory notifications due to being paired with a perceived less effective auditory notification. This seems to align with other findings of perceived notification effectiveness. Baldwin & Lewis (2014) found that the auditory modality elicited more feelings of urgency compared to the visual modality in participants within a driving context.

In the experimental phase, when both notification types (person and orientation) were visual, participants were more likely to miss the notifications altogether. As such, it appears that an auditory component is necessary for correct identification of an external notification. This suggests that some form of inattentional blindness may be occurring due to the demands of game tasks. Additionally, Multiple Resource Theory may be useful in explaining why the auditory notifications helped with intrusion identification (Wickens, 1984, 2002). It could be the case that there were far more visual demands than auditory and thus the resources for recognizing visual information were depleted allowing for the auditory notification to be more useful since its resources were not yet exhausted. Further research is needed to assess if this is the case within the context of the present study.

When participants perceived a notification, modality seemed to be an important factor in how quickly the participant responded. Just as visual notifications were the most often missed, they were also the slowest in participant response, for person type notifications. Also, for person type notifications the auditory component is important for faster reaction time, but when paired with the visual notification it does not gain an increased reaction time. This is likely because the orb may fall outside a participant's visual field, whereas detection of the tone does not depend on the user's position or orientation.

An interesting interaction was observed between person and orientation type notifications. First, reaction time was unchanged between the orientation notification modalities when they were paired with the auditory person type (voice). This suggests that the auditory person notification does such a good job in alerting users to a change in their environment that the orientation notifications do not increase detection to the environmental change. This is important because it seems to corroborate the survey data findings that an explicit auditory verbalization is most effective.

Second, there was a change among the visual person type notifications. Specifically, when the visual person notification was paired with the visual orientation notification participants had the slowest reaction times. Again, suggesting a tendency for inattentional blindness or lack of saliency on the purely visual modality on alerting users of a changing environment state. In contrast, when the visual person notification was paired with the combined visual and auditory orientation notification it had faster reaction times.

Finally, all orientation type notification modalities had relatively similar reaction times when paired with the person combined modalities. All of these results suggest that without the auditory component for alerting users explicitly that another person has entered the room, participants are not as aware a change has taken place.

There was a difference in dwell time between modalities for orientation devices that was expected as their purpose was to help participants locate the intrusion. Interestingly, in contrast to the findings of reaction time for person type notifications, the auditory modality is insufficient, or at least not as effective as the visual modality. The visual modality had the fastest dwell time compared to the auditory modality. This pattern suggests that the visual modality conveys the direction to an intrusion better than auditory. Additionally, there was no difference in dwell time between the visual only modality and the combined modalities suggesting that the addition of an auditory component does not increase a participant understanding of where to locate that intrusion.

REFERENCES

Adcock, M., & Barrass, S. (2004). Cultivating design patterns for auditory displays. *Proceedings of ICAD '04*.

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. *Applied Ergonomics*, *45*(5), 1270–1277.

Burigat, S., & Chittaro, L., (2007). Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies, 65*, 945-958.

Chagué, S., & Charbonnier, C. (2016). Real virtuality: A multi-use immersive platform connecting real and virtual worlds. *Proceedings of the 2016 Virtual Reality International Conference on - VRIC 16*.

Dinh, H., Walker, N., Hodges, L., Song, C., & Kobayashi, A. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*.

Ghosh, S., Winston, L., Panchal, N., Kimura-Thollander, P., Hotnog, J., Cheong, D., . . . Abowd, G. D. (2018). NotifiVR: Exploring Interruptions and Notifications in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*,*24*(4), 1447-1456.

Lague, S. [Sebastian Lague]. 2015, September 17. *Create a Game (Unity 5)*[Video file]. Retrieved from https://www.youtube.com/watch?v=SviIeTt2_Lc&list=PLFt_AvWsXl0ctd4dgE1F8g3uec4zKNRV0

Matthews, T. (2006). Designing and evaluating glanceable peripheral displays. *Proceedings of the 6th ACM Conference on Designing Interactive Systems - DIS 06*.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions of Information and Systems*.

Rogers, S. (2014). *Level up!: The guide to great video game design*. Chichester, West Sussex: Wiley.

Simeone, A. L., Velloso, E., & Gellersen, H. (2015). Substitutional Reality. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI 15*.

Tomitsch, M., Kappel, K., Lehner, A., & Grechenig, T. (2007). Towards a taxonomy for ambient information systems. *Proceedings of Pervasive '07*.

Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Orlando: Academic.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159–177.