Audio-visual AR to Improve Awareness of Hazard Zones Around Robots

Ane San Martín
IK4-TEKNIKER
Eibar 20600, Spain
anesanmartin3@gmail.com

Johan Kildal
IK4-TEKNIKER
Eibar 20600, Spain
johan.kildal@tekniker.es

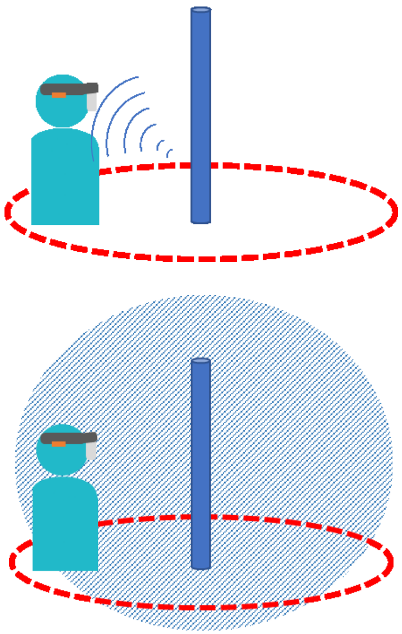

Figure 1: A user wearing a HoloLens device walks towards the source of a hazard (seen by the user as a stationary hologram in the shape of a pole). When entering the hazard zone around the pole (the red dotted circle, invisible for the user) she is informed by the auditory augmentation (a) or visual augmentation (b) of the environment.

ABSTRACT

Navigating a space populated by fenceless industrial robots while carrying out other tasks can be stressful, as the worker is unsure about when she is invading the area of influence of a robot, which is a hazard zone. Such areas are difficult to estimate and standing in one may have consequences for worker safety and for the productivity of the robot. We investigate the use of multimodal (auditory and/or visual) head-mounted AR displays to warn about entering hazard zones while performing an independent navigation task. As a first step in this research, we report a design-research study (including a user study), conducted to obtain a visual and an auditory AR display subjectively judged to approach equivalence. The goal is that these designs can serve as the basis for a future modality comparison study.

1 INTRODUCTION

Scenarios in which robots and humans share spaces and activities are becoming ever more frequent. In particular, the presence of collaborative robots (cobots) is rapidly increasing in industrial facilities, where human workers and cobots work alongside each other. In such scenarios, fenceless cobots are deployed to behave in intrinsically safe ways.

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312996

KEYWORDS

Collaborative robots; hazard warning; pedestrian navigation; multimodal; auditory; visual; display; audio-visual; augmented reality; HoloLens; head-mounted device

Table 1: Requirements to design the information displays, either auditory or visual

To be perceived as hazard warnings
To signal when a hazard zone was entered or exited
To inform about the relative location (orientation and distance) of the origin of the hazard
To minimize AR clutter, and avoid interfering with the perception of reality (e.g., objects, voice conversations)
To minimize introduction of additional cognitive demands
To provide a good user experience

Table 2: Core design patterns to generate displays in each modality

Boolean	When inside the hazard zone, the user perceives a constant stimulus, independently of the distance to the source of the hazard
Progressive	When inside the hazard zone, the user perceives a stimulus that increases gradually in salience with proximity to the source of the hazard
Stepped Progressive	Same as for Boolean but with an abrupt increment in salience when crossing threshold distance inside the hazard zone

For instance, as part of their safe behavior, they react to the nearby presence of humans by moving more slowly or even stopping altogether. Other aspects of robots' safe behavior include collision detection through the monitoring of the forces that they exert, which are rapidly reduced to levels that cannot harm a person [3].

Such safe behavior, however, comes at the expense of reducing the robots' productivity, which is undesirable for the human worker. In addition, despite safe-behavior measures, the perception of residual potential danger may remain for workers (because of e.g., objects that robots manipulate and move around, fears of malfunctions in the safety mechanisms), affecting their trust in robots and leading to poor UX. As a result, such poor situation awareness and uncertainty lead to heightened sustained levels of stress (chronic stress), which can over time be very damaging for the health of workers [15].

To address this problem, the user has to know when he/she is entering the area of influence of a robot (a potential hazard zone), so that the user can decide how to act accordingly. Robots themselves can provide the additional information for situation awareness that the worker is lacking: robots know where the worker is in relation to them and whether such presence of the worker is triggering any safety mechanisms. We propose making such information available through head-worn AR displays just for the worker entering a robot's area of influence. Head-mounted displays (HMDs) permit presenting such information to target individual in real time.

We want to investigate how to best design the AR information presented through HMDs, with displays that address the user's visual or auditory sensory channels, or the combination of both (multimodal display). For this work, we have selected the Hololens (Microsoft) device, which provides visual and auditory AR displays representative of the state of the art. Visual AR is already being used to inform workers about relevant aspects of their collaborative work with robots (e.g., [9]). In a related application, it has also been proposed to help pedestrians with awareness autonomous car traffic [12]. Examples of the use of auditory-only AR do not abound (e.g., [4, 5]). As for audio-visual multimodal AR interfaces, relevant prior work includes [1, 14]. To investigate the relative roles of audio and visual AR in providing hazard zone information, we first had to design auditory and visual representations of that information that (i) were effective as warning displays, (ii) could be rendered through the Hololens device, and (iii) were as closely equivalent to each other as possible in terms of the information they provided. This third condition was essential in order to design a future study in which both single-modality displays can be compared, together with the combination of both. In this paper, we report the user-centered design process that we followed to obtain each single-modality display (visual and auditory), fulfilling the conditions (i) to (iii) just stated. The comparative study of both will inform our immediate future work.

2 DESIGNING COMPARABLE AUDIO AND VISUAL AR DISPLAYS

For either display modality (auditory or visual), we set a list of design requirements, shown in Table 1 (left). These requirements were obtained from an analysis of the problem in the context described in the introduction.

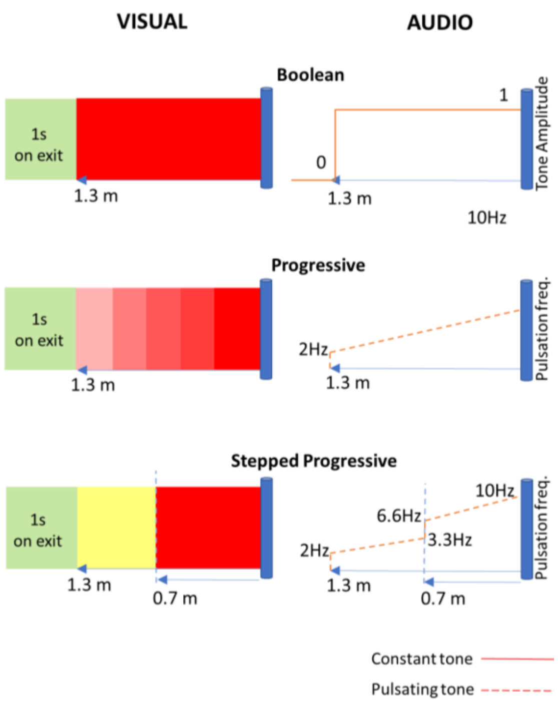

Figure 2: Auditory and visual display designs. Horizontal axes (leftwards) show distance to hazard. Visual displays (left): color cast rendered inside hazard zone; green cast (one second long) only on exiting zone. Auditory displays (right): sound volume same in all cases; for *Progressive* and *Stepped Progressive*, vertical axis indicates frequency of pulsation of sound.

Based on these requirements, we first designed three eligible display variants for each modality. Then, we conducted a user study to support the final selection of the two best-matching display designs (one auditory and one visual).

We began the design process by conducting an internal workshop with a small group of interaction designers. The goal was to obtain designs of auditory and visual augmentations for a user wearing a Hololens device when walking into a hazard zone. To structure the process, a simplified description of a hazard zone was created. The robot supposedly generating the hazard in its influence area was synthesized as a vertical blue pole (Ø0.1m), emerging from the ground (see Figure 1). This was rendered as a holographic stationary object, always visible to the user in both the audio-based and the vision-based hazard display designs. The hazard zone around the pole was a concentric cylindrical volume (Ø2.6m), which was invisible for the user from a distance (only the pole could be seen). The projection of the hazard zone on the ground is represented as a red dotted circle in Figure 1. The challenge was to design the interactive representation of the danger zone (visual or auditory) that the user would perceive when physically entering the hazard zone.

From the design generated at the workshop, we pre-selected six display designs (three auditory and three visual) that instantiated three design patterns, as devised by the designers during their discussions. Such patterns were called *Boolean*, *Progressive*, and *Stepped Progressive* (see Table 2).

Figure 2 shows graphically the instantiations of each design pattern, for an audio and a visual display. In the visual displays (Figure 2, left), the stimulus perceived by the user was the color cast of the whole scene rendered by the HMD device. In every case, color casts were created using spheres of light (Figure 1), with their center at the approximate adjustable height of the user's eyes. In the *Visual Boolean* design, the scene turned see-through red for as long as the user remained anywhere inside the hazard zone. With the *Visual Progressive* design, the intensity of the red cast gained in saturation with proximity to the center of the hazard zone, with a maximum level of intensity equal to that in the Boolean design. The *Visual Stepped Progressive* design utilized two distinct ranges of color cast; yellow in the region further from the center, and red in the more interior region. In all three visual displays, when leaving the hazard zone, the scene appeared cast in see-through green for a duration of one second, after which any color cast disappeared.

In the case of the auditory displays (Figure 2, right), all stimuli perceived were based on a 440Hz pure tone rendered at an amplitude that was common across auditory display designs, defined arbitrarily as similar in salience to the most saturated red color cast used in the visual displays (although no psychophysical study was conducted to confirm the accuracy of this equivalence). The *Auditory Boolean* display produced a continuous audible tone for as long as the user was inside the hazard zone. The *Auditory Progressive* display produced discrete grains consisting of one period of a 440Hz tone, with spacing between grains shortening linearly as the user approached the center of the hazard: tone grains produced at a frequency of 2Hz in the periphery of the hazard zone, and at 10Hz in its center. Finally, the *Auditory Stepped Progressive* design was similar to Auditory Progressive but with an abrupt step in the linear path of shortening of temporal space between grains. Frequencies in the outer region: 2Hz to 3.3Hz; frequencies in the inner region: 6.6Hz to 10Hz.

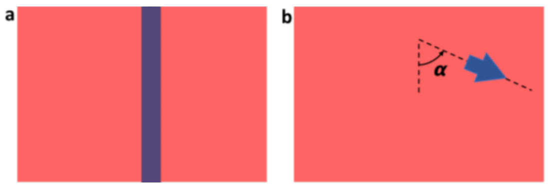

Figure 3: visual augmentation when a user is inside a hazard zone, thus seeing a color cast (this figure is an illustration, not a display capture). (a)-The origin of the hazard (blue column can be seen, and no other cue is shown. (b)-View from inside the same hazard zone when the source of the hazard in not visible in the FoV of the HMD device. In that case, an arrow cue swivels an angle *α*, meaning that the user should turn their head that same angle in the direction of the arrow (right in this case) to be able to see the source of hazard within the FoV of the device. As soon as the pole (origin of the hazard) appears on view, the arrow disappears. The angle *α* maps the rotation of the person's head around her vertical axis with respect to the column hologram causing the hazard (e.g., in Figure 3-b, *α*=90 would mean that the column hologram is located exactly 90 degrees to the right of the user; similarly, *α*=90 would mean that the column is located directly behind the user).

Despite our efforts to create comparable auditory and visual display designs, the resulting implementations on an off-the-shelf HoloLens device were subject to limitations of the hardware itself. The main limiting factor was the narrow field-of-view (FoV) offered by the device. Conversely, the spatial sound of Hololens is a strength, as it can assist with the localization of the origin of the hazard [13, 16]. To partially compensate for the narrow FoV limitation, we designed a visual cue that assisted with the localization of the off-screen pole. Literature describes several visual cue designs to indicate the presence and location of multiple off-screen points of interest simultaneously, such as halos [2], wedges [10] and arrows [6, 11], also tested on an HMD device [7]. For single out-of-view objects on HoloLens, a moving arrow was proposed [8]. We developed a new variant of a swiveling arrow cue optimized for searches anywhere on a 360-degree circle around the user. Figure 3 shows and describes how this cue operated.

3 USER STUDY

In a user study (6 participants, 3 female), we wanted to investigate experimentally which pair of displays (one visual and one auditory) was a best match from the perspective of our design requirements. Although we had implemented pairs of displays based on three patterns, participants were granted freedom to pair designs, based on five proposed criteria: *pleasantness*, *capacity to capture attention*, *capacity to provide sense of safety*, *capacity to convey sense of distance to origin of hazard*, and *preference*. The 5 criteria were assessed in a post-task questionnaire consisting of 10 questions, arranged in alternate order of "*which is highest*" and "*which is lowest*". The last question in the list was selected to be "*which is the one you like best?*", to benefit from the insight gained while filling out the questionnaire and before making their choice.

For the study, six blue holographic poles were created that were visible through the HoloLens device. Each pole was the center of a hazard zone, and each implemented one of the six displays described above. They were arranged by modality in two rows of three columns each, in random order within modality. In a session, the participant was told that each column was the source of some hazard that should normally be avoided. She was then asked to walk closer to each column in a row and experience the behavior of each display within that modality (the row that was first inspected was counterbalanced across participants). The participant was encouraged to enter and leave each hazard zone repeatedly and at any speed chosen, and to understand its behavior. The participant could switch between columns from that row freely, and there was no time limit. Once satisfied with the inspection, the participant filled out the post-condition questionnaire and proceeded to inspect the three columns with displays from the other modality. Once both conditions were completed, the participant was asked to match each display from a modality with a display from the other modality, based on overall similarity with every aspect being considered, and to rate how good each match was, in a scale from 0 (worst) to 10 (best). Finally, the facilitator conducted a semi-structured interview with the participant, starting by requesting justification for each match proposed and rating given. The facilitator then led the interview with references to choices made by the participant in the post-task questionnaires, in order to gain maximum insight.

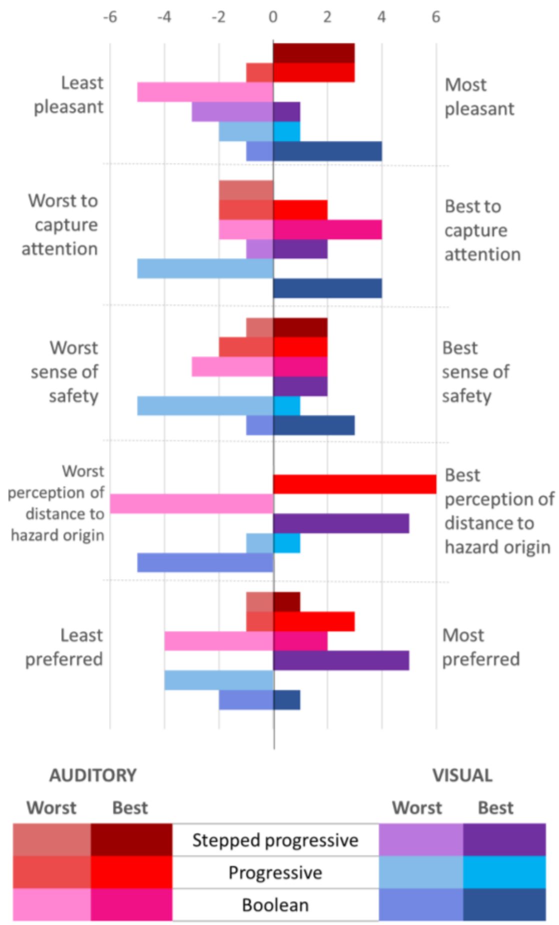

Figure 4: ***Above*****: graph showing number of users selecting each option in the post-task questionnaire Scales shows number of times a stimulus was selected as worst or best in each category.** ***Below*****: legend, showing two shades (worst and best) for each stimulus.**

4 ANALYSIS OF RESULTS

Two pairs of auditory and visual displays were proposed more frequently than others in the user study as best matching. These pairs were: (i) *Auditory Stepped Progressive* with *Visual Stepped Progressive*, and (ii) *Auditory Progressive* with *Visual Stepped Progressive*. Each match was proposed by 4 out of the 6 participants, and both pairs received top score in the ranking of proposed matches by 2 participants. In the post-task questionnaire (see Figure 4), *Visual Stepped Progressive* had been clearly rated as the preferred visual display for the application proposed. It was also clearly better rated than the other visual displays for providing a sense of distance to the source of the hazard. The only possible contending visual display was *Visual Boolean* for being better at capturing attention, and it was also perceived as more pleasant. However, since its preference rating was low, we confirmed our selection of *Visual Stepped Progressive* as the visual display design to build upon in our future work. Among other things, we learned from our interviews that this visual display was preferred because the change in color was indicating progression toward the hazard more clearly than the change in intensity, and some mentioned that additional intermediate colors might be an improvement on the design.

As for the auditory display, two designs appeared in the most-frequently proposed pairs of displays: *Auditory Stepped Progressive* and *Auditory Progressive*. From the post-task questionnaire responses (see again Figure 4) we saw that *Auditory Progressive* was the most preferred design, and it was the clear choice in terms of good indication of distance to the origin. It also performed better in subjective scores for capturing attention. As a result, we selected *Auditory Progressive* as the basis for an auditory display design.

5 CONCLUSIONS AND FUTURE WORK

In this paper we report the first stage in our research on the use of visual and/or auditory AR information to improve awareness of potential hazards around robots and thus improve user safety and experience when walking in a space shared with them. We are interested in investigating the relative roles that auditory and visual information can play in this endeavor, and in learning how to best design them for easy and non-intrusive applications in industrial contexts and beyond.

In the design-research exercise reported here, and through a small qualitative user study, we have concluded that the *Visual Stepped Progressive* and the *Auditory Progressive* display design each provide a good starting point for optimum designs of AR displays in each of these modalities and with current AR HMD technology. What is more, we have also learned that users find an overall degree of equivalence between these two display designs that is higher than what was found for other pairs of designs. This means that a modality comparison study could be designed based on these two displays.

Our immediate next steps in this research will be to develop the auditory and visual displays selected in the first stage that we have reported here. Then, we will conduct a modality comparison study between an auditory, a visual and a combined audio-visual display. We will assess the study with both quantitative and qualitative metrics.

REFERENCES

[1] Baldwin, C.L., Spence, C., Bliss, J.P., Brill, J.C., Wogalter, M.S., Mayhorn, C.B. and Ferris, T.K. 2012. Multimodal cueing: The relative benefits of the auditory, visual, and tactile channels in complex environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (2012), 1431–1435.

[2] Baudisch, P. and Rosenholtz, R. 2003. Halo: a technique for visualizing off-screen objects. *Proceedings of the SIGCHI conference on Human factors in computing systems* (2003), 481–488.

[3] Behrens, R., Saenz, J., Vogel, C. and Elkmann, N. 2015. Upcoming Technologies and Fundamentals for Safeguarding All Forms of Human-Robot Collaboration. *8th International Conference Safety of Industrial Automated Systems (SIAS 2015), Königswinter, Germany* (2015), 18–20.

[4] Black, D., Hahn, H.K., Kikinis, R., W\a ardell, K. and Haj-Hosseini, N. 2018. Auditory display for fluorescence-guided open brain tumor surgery. *International journal of computer assisted radiology and surgery*. 13, 1 (2018), 25–35.

[5] Blum, J.R., Bouchard, M. and Cooperstock, J.R. 2011. What's around me? Spatialized audio augmented reality for blind users with a smartphone. *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services* (2011), 49–62.

[6] Burigat, S., Chittaro, L. and Gabrielli, S. 2006. Visualizing locations of off-screen objects on mobile devices: a comparative evaluation of three approaches. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (2006), 239–246.

[7] Gruenefeld, U., Ali, A.E., Heuten, W. and Boll, S. 2017. Visualizing Out-of-view Objects in Head-mounted Augmented Reality. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, NY, USA, 2017), 81:1–81:7.

[8] Gruenefeld, U., Lange, D., Hammer, L., Boll, S. and Heuten, W. 2018. FlyingARrow: Pointing Towards Out-of-View Objects on Augmented Reality Devices. *Proceedings of the 7th ACM International Symposium on Pervasive Displays - PerDis '18* (Munich, Germany, 2018), 1–6.

[9] Guhl, J., Tung, S. and Kruger, J. 2017. Concept and architecture for programming industrial robots using augmented reality with mobile devices like microsoft HoloLens. *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)* (Sep. 2017), 1–4.

[10] Gustafson, S., Baudisch, P., Gutwin, C. and Irani, P. 2008. Wedge: clutter-free visualization of off-screen locations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2008), 787–796.

[11] Henze, N., Poppinga, B. and Boll, S. 2010. Experiments in the wild: public evaluation of off-screen visualizations in the Android market. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (2010), 675–678.

[12] Hesenius, M., Börsting, I., Meyer, O. and Gruhn, V. 2018. Don'T Panic!: Guiding Pedestrians in Autonomous Traffic with Augmented Reality. *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (New York, NY, USA, 2018), 261–268.

[13] Hong, J.Y., He, J., Lam, B., Gupta, R. and Gan, W.-S. 2017. Spatial audio for soundscape design: Recording and reproduction. *Applied Sciences*. 7, 6 (2017), 627.

[14] Montuwy, A., Cahour, B. and Dommes, A. 2017. Visual, auditory and haptic navigation feedbacks among older pedestrians. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (2017), 83.

[15] Oken, B.S., Chamine, I. and Wakeland, W. 2015. A systems approach to stress, stressors and resilience in humans. *Behavioural brain research*. 282, (2015), 144–154.

[16] Zijlstra, A.T. 2017. *Using the HoloLens' Spatial Sound System to aid the Visually Impaired when Navigating Indoors*. University of Groningen.