This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

Multisensory Proximity and Transition Cues for Improving Target Awareness in Narrow Field of View Augmented Reality Displays

Christina Trepkowski*, Alexander Marquardt*, Tom David Eibich, Yusuke Shikanai, Jens Maiero, Kiyoshi Kiyokawa, Ernst Kruijff, Johannes Schöning, and Peter König

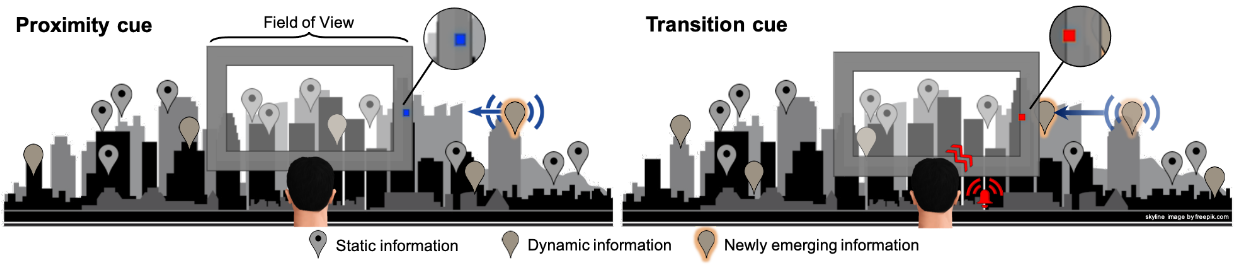


Fig. 1: Demonstration of proximity and transition cues in an augmented reality city information system scenario. Being aware of newly emerging, yet potentially important information can become difficult in dynamic environments. Proximity cues (left side) provide spatial information in either the visual or the auditory modality and help users to become aware of out-of-view objects ("target"). When the target crosses the border of the field of view of the visual output device, the proximity cue is replaced by a transition cue (right side). A short visual, auditory or tactile notification signals the user that the target entered the field of view and is potentially visible.

Abstract—Augmented reality applications allow users to enrich their real surroundings with additional digital content. However, due to the limited field of view of augmented reality devices, it can sometimes be difficult to become aware of newly emerging information inside or outside the field of view. Typical visual conflicts like clutter and occlusion of augmentations occur and can be further aggravated especially in the context of dense information spaces. In this paper, we evaluate how multisensory cue combinations can improve the awareness for moving out-of-view objects in narrow field of view augmented reality displays. We distinguish between proximity and transition cues in either visual, auditory or tactile manner. Proximity cues are intended to enhance spatial awareness of approaching out-of-view objects while transition cues inform the user that the object just entered the field of view. In study 1, user preference was determined for 6 different cue combinations via forced-choice decisions. In study 2, the 3 most preferred modes were then evaluated with respect to performance and awareness measures in a divided attention reaction task. Both studies were conducted under varying noise levels. We show that on average the Visual-Tactile combination leads to 63% and Audio-Tactile to 65% faster reactions to incoming out-of-view augmentations than their Visual-Audio counterpart, indicating a high usefulness of tactile transition cues. We further show a detrimental effect of visual and audio noise on performance when feedback included visual proximity cues. Based on these results, we make recommendations to determine which cue combination is appropriate for which application.

Index Terms—Augmented Reality, view management, guidance, multisensory cues, performance, situation awareness

1 Introduction

A major problem of commonly used optical see-through (OST) Augmented Reality (AR) displays is their relatively narrow field of view (FOV). This means that only a small part of human vision can be augmented. Humans have a binocular FOV exceeding 210° horizontally and 150° vertically [47]. In contrast, the FOV covered by current OST head-worn AR devices is much smaller, e.g. the 52° diagonal FOV of the Microsoft HoloLens (2. generation) which measures 43° horizontally and 29° vertically [18]. Consequently, in large information spaces a number of objects is always outside the user's FOV and can only be brought into view through head or object movement.

To draw attention to objects of interest and to improve the awareness of out-of-view objects, researchers have developed a variety of visual, but also non-visual **guidance techniques**. These techniques provide cues on the spatial location of one or more targets to the users (see the Related Work Section). However, in uncontrolled and dynamic real-world environments the usage of guidance techniques can be impaired. Here, the situation can occur that users perceive guidance cues and ambient sensory input ("noise") in close temporal and spatial proximity. Within a single modality but also across sensory modalities the perception of cues can be suppressed by **environmental noise** due to close interactions of neural responses [40]. In addition, during real-world AR usage users often perform concurrent activities [52] that distract from the cues. As human processing capacities are limited [26, 63], there may no longer be sufficient resources available for the reliable interpretation of guidance cues, especially under conditions of **divided attention**. Mechanisms that impair the perception of guidance cues can also cause the AR target object itself to go unnoticed when it enters the FOV. Especially in dense AR environments newly entering targets can easily be missed as augmentations potentially clutter and occlude each other and/or other objects in the scene [27].

- *Christina Trepkowski, Alexander Marquardt, Tom David Eibich, Jens Maiero, and Ernst Kruijff are with Bonn-Rhein-Sieg University of Applied Sciences. E-mail: {christina.trepkowski, alexander.marquardt, tom-david.eibich@h-brs.de, jens.maiero, ernst.kruijff}@h-brs.de.*
- *Ernst Kruijff is also with Simon Fraser University.*
- *Yusuke Shikanai and Kiyoshi Kiyokawa are with Nara Institute of Science and Technology. E-mail: shikanai327327@gmail.com, kiyo@is.naist.jp.*
- *Johannes Schöning is with the University of St. Gallen. E-mail: johannes.schoening@unisg.ch.*
- *Peter König is with Osnabrück University. E-mail: pkoenig@uos.de.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx.

These authors contributed equally to this work.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

In this work, we introduce a novel feedback technique for raising awareness of an augmented target object that moves towards and enters the FOV. Our method is tailored to work in situations where environmental noise is present and when visual attention has to be divided. To achieve these goals, our technique combines two different types of cues: The first cue is referred to as **proximity cue** and provides spatial feedback on moving out-of-view objects (see Fig. 1 left). It can be used to guide the user to the out-of-view object or, alternatively, to improve awareness of the overall situation without approaching the object. The second cue is referred to by us as **transition cue** and makes the user particularly aware of the entry and exit of AR target objects into and out of the FOV (see Fig. 1 right). We created multiple variants of combined feedback (sequential provision of proximity and transition cues) where different unimodal and also multimodal combinations of visual, auditory, and tactile cues were used: We assume that depending on the usage context it can be beneficial to provide multisensory feedback. In this paper, we will refer to the combination of proximity and transition cue as "mode". Throughout two user studies we compared preference and performance measures between different modes. To mimic real world conditions both studies were performed under two different visual and auditory noise intensity levels. In study 1, via forced-choice decisions a preference score was determined for each implemented mode: Audio-Audio, Audio-Tactile, Audio-Visual, Visual-Audio, Visual-Tactile, and Visual-Visual. Decision criterion was the usefulness to make aware of an augmentation out-of-view and its transition inside the FOV. In study 2, we then evaluated modes with the highest preference scores Visual-Audio, Visual-Tactile and Audio-Tactile in a divided attention task. Users had to react as fast as possible to out-of-view objects that enter the FOV while performing a concurrent visual task in the center visual field.

1.1 Contribution

We extend prior work by the development of a combined feedback specifically targeting the spatial and temporal perception of moving objects in dynamic and information-rich AR environments. In this work, for the first time, subjective and objective performance measures are compared between different modalities of proximity and transition cues presented in seamless succession under real world conditions. In doing so, we included different levels of visual and auditory noise and a divided attention task in our studies. We progress beyond the state of the art (see [91]) by tailoring the composition and presentation of the feedback, the testing scenario and tasks to narrow FOV AR. Throughout two user studies we address two research questions and present associated contributions:

RQ$_1$ (Study 1): *Which combination of visual, audio and vibro-tactile cues is most preferred by users to make them aware of an augmentation out-of-view and its transition inside the field of view? Is there an effect of different visual and auditory noise levels on preferences?*

C$_1$: We found that with reduced noise Audio-Tactile, Visual-Tactile, and Visual-Audio and modes were significantly preferred over combinations Audio-Visual and Visual-Visual, but not Audio-Audio. Under increased noise the Audio-Tactile combination was preferred over all other modes except for Visual-Tactile which was comparably popular.

RQ$_2$ (Study 2): *Which cue combination mode performs best regarding target localization and reaction to targets that enter the field of view when a concurrent visual task is performed? How does the level of visual and auditory noise affect performance in both tasks?*

C$_2$: We showed a high usefulness of feedback combinations that include tactile transition cues by yielding 46% faster responses to incoming out-of-view targets with the Audio-Tactile and 42% with the Visual-Tactile mode compared to their Visual-Audio variant. Adding environmental noise increased the reaction time benefit of Audio-Tactile to 54% and Visual-Tactile to 47% while the latter proved to be most robust against noise.

2 Related Work

In the following we will describe lower-level perceptual mechanisms which are involved when users build up their awareness of surrounding objects. We will illustrate which problems occur when the FOV is restricted and how they can be counteracted, for example through different existing guidance techniques.

2.1 Perceptual foundations

To build up or maintain awareness of (unknown) surroundings, humans usually need to search for information in their environment. When searching for information, basic object features (e.g., colors, size, curvature) can be detected automatically and processed preattentively in the entire visual field [105, 106] in contrast to other aspects (e.g. meaning of words) which have to be focused on [14]. Guiding features in the peripheral visual field can be beneficial [14] as they can significantly reduce visual search times [15, 123]. However, when searching for augmented objects in narrow-FOV AR displays guiding features in parts of the near and, in particular, the far peripheral field of vision are permanently absent. As a consequence, target search can be slowed down considerably, especially when information density increases [14]. This issue is further exacerbated by the fact that also the form-factor of an OST device can occupy a significant amount of the user's visual field, resulting in critical oversights of the surroundings [92]. Another effect of restricted peripheral vision is that more head movement is required to see out-of-view information that would normally be brought into view by a short eye movement [19, 44]. Finally, a change of visual scan patterns [14, 19], the degradation of navigation abilities and development of spatial knowledge [2, 117] can be observed if a narrow FOV is used.

To substitute missing guiding features in the periphery, cues can be provided that trigger orienting attention to a particular location in space. Responses to targets that appeared at a cued location can potentially be facilitated [71]. Cross-modal studies on spatial cueing of attention showed that voluntarily orienting attention to a location facilitated the reaction to subsequent targets regardless of the cue and target modalities (reviewed in e.g. [24]). Furthermore the provision of redundant cues of different modalities can speed up responses [90]. In the following sections higher level visual and non-visual guidance methods are introduced that make use of aforementioned attentional mechanisms.

2.2 Visual Methods

Visual methods that enhance the awareness of out-of-view objects have been studied in the context of display media other than AR displays, yet basic mechanisms likely can also be applied to augmented reality environments. We report found advantages and disadvantages though note further studies may need to be performed to prove their transfer to AR.

Overview + Detail methods provide one or multiple overviews of contextual information (at a usually reduced scale) and a detailed view simultaneously. An overview of these methods can be found in [12].

Contextual cue techniques use screen space more efficiently by showing only the cues for pre-defined objects of interest instead of the whole context. Techniques often locate these cues inside the borders of the focus area, leaving the center vision free. For example, the City Lights method [125] uses a line segment or a point to indicate the direction of an object on the edge of the screen, while Halo and Wedge techniques augment the detail view with partial visualizations [33]. The latter two have the advantage that their exact position can be inferred by mentally completing the visualization such as rings (Halo) or triangles (Wedge). EdgeRadar [37] – an extension of City Lights – reserves a border along the display edge for tracking off-screen moving targets and serves also as inspiration for recent techniques like EyeSee360 [34]. As only the peripheral area of the display is occupied the user is still able to see and process potentially important real world environmental cues through the central display area. Alternatively, the space is available to display other augmented content. Methods like edge-highlighting or the use of external LED indicators placed in the user's peripheral vision can be used to increase awareness that has been impaired due to occlusion caused by OST HMDs [92]. Other techniques use one or multiple virtual arrows to point in the direction of one or more targets while scaled arrows can additionally encode distance [11, 39]. Solutions like 3D Arrows, AroundPlot, 3D Radar, sidebARs, and Mirror Ball

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

have specifically been developed for visualizing out-of-view objects in a 3D environment. An overview of these methods and how they perform against each other in a visual search task can be found in [9].

Other methods trying to extend the limited FOV using a so-called fisheye view. For example, Aroundplot [46] maps 3D spherical coordinates to a 2D orthogonal fisheye, which addresses typical issues of location cue displays such as occlusion of information. Radial distortion [98] and EXMAR [43] are other fisheye distortion techniques that expand the AR field of view and differ from each other in the requirements, including the fisheye style and FOV.

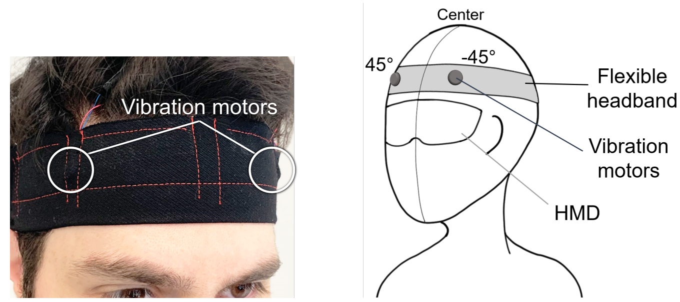
2.3 Non-Visual Methods

As an alternative to visual methods, a variety of non-visual techniques in the form of auditory and tactile cues exist. These techniques can provide spatial information such as direction and distance. In this section, we address both, purely non-visual methods and multisensory techniques that combine visual and non-visual cues. A meta analysis [91] showed, that when added to a baseline task or presented simultaneously with visual cues, vibro-tactile cues enhance task performance. While vibrotactile cues can be an effective replacement for visual alerts, they are not necessarily effective in terms of replacing visual direction cues. Tactile cues have been adopted quite frequently to direct navigation, e.g. using vests [62], gloves [107] or a belt [55]. 3D selection methods using tactile cues are studied for non-directional feedback [3] and directional feedback [68]. Tactile cues also can be used to support visual search tasks [60] and target finding by device-based target scanning techniques [1] for mobile devices. Finally, researchers developed vibrotactile head-mounted displays to provide guidance cues, e.g. in form of a ring-based tactile setup [21, 69] around the user's head or even higher resolution tactor grid resembling an EEG setup [50] to enhance awareness and support guidance.

Auditory cues also have been used to support visual search [78] and navigation [49]. Studies have looked at the effects of motion, location and practice on visual search performance with 3D auditory cues, showing an improvement of search performance by 22% and 25% in static and dynamic environments respectively [73]. Another approach of using auditory cues is called sonification in which sound properties are used (predominantly pitch, but also loudness, duration/tempo, and timbre) [25] with respect to the presence of the auditory reference. Sonification metaphors are commonly used for speeding up visual search tasks (e.g. [64]) but can be also found in other application scenarios like car parking systems, which provide distance information through a decreasing time interval between impulse tones [87]. This metaphor can also be applied for spatial data exploration and guidance for visually impaired people in AR [8, 94] or for improving accuracy in high precision tasks [95] without occluding the visual field. Furthermore cross-modal effects have been studied, including audio-tactile effects [42, 82] and conflicts between audio and visual cues [54]. In general, multisensory stimulation has been shown to have benefits over unimodal feedback with regard to improved task performance [13, 45, 91] and higher user satisfaction [59]. Considering real world usage conditions, multisensory feedback can also improve noticeability under noise [5] and reduce cognitive load [86, 100], which can be useful in divided attention situations. Especially in narrow FOV AR displays the integration of non-visual cues has some advantages. Given that the majority of tasks in AR generally require primarily visual information processing, they enable the user to use AR glasses without increasing the visual load. Moreover, non-visual cues support a free field of vision, promoting a faster and more reliable detection of background targets [70]. However, in comparison visual feedback is generally highly trained, easy to learn and can be interpreted fast [70]. Thus, choosing the appropriate method can be challenging and depends on the context of application.

3 System and Study Setup

In our studies we deployed the widely used Microsoft HoloLens (1st gen). The built-in speakers provided a spatialized HRTF sound. With its 35° (diagonal) field of view, the Microsoft HoloLens represents a typical FOV for AR OST head worn devices and reflects the aforementioned problem of limited screen size for augmentations.



(a) User wearing the tactile interface.

(b) Scheme of the tactile headband with the according tactor locations.

Fig. 2: Custom-made tactile headband with 2 Vibration motors are placed on $-45°$ and $45°$ from the user's point of view.

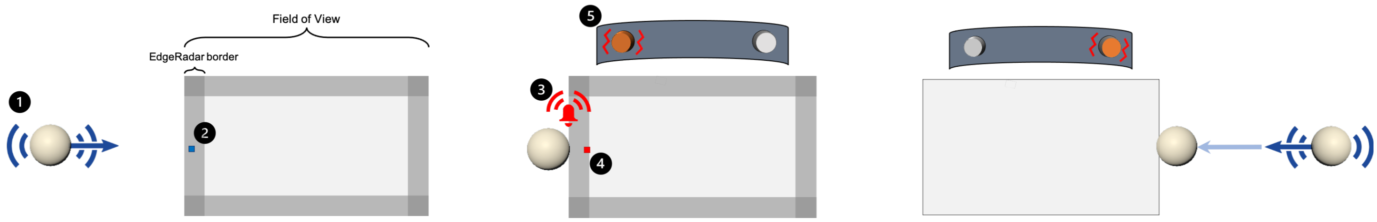
Although devices with a slightly wider FOV have become available (e.g. HoloLens 2), we assume that problems will also occur in these devices (see the Discussion section). The system was implemented using Unity 2019.2. In the laboratory the study participants used a Microsoft Xbox One wireless controller as input device while sitting. To ensure the safety of the participants and preventing the transmission of Covid-19 we followed a deep cleaning protocol. The equipment and workspace was cleaned and disinfected thoroughly after each test subject and adhered general safety guidelines [65].

A modified version of [70] with 2 vibration motors (Precision Microdrives 308-107) provided accurate tactile cues on the users forehead. We assume that adding further motors was not necessary to improve the feedback, since we focused on specific object movement trajectories in our studies (see the User Studies Section). Motors are placed into sewn pockets on the headband at $-45°$ and $45°$ from the user's point of view (see Fig. 2a). This motor placement was based on the perceptual acuity for vibration at the frontotemporal region (point on the forehead at about 45°), which is higher compared to the hairy skin on the temple regions [22]. The vibration motors were driven by a Raspberry Pi 3 Model B+ running a Python-based version of Open Sound Control. It allows the microcontroller to communicate with the Unity App running on the Microsoft HoloLens. In this specific setup, the network latency between the individual system components as well as the typical start-up times of the vibration motors used were initially measured and taken into account for the purpose of generating the vibrotactile cues. The network latency between the HoloLens and the Raspberry Pi was about 21 ms on average. The typical rise time of the used vibration motors is 54 ms. The activation of the feedback is generally coupled to the FOV of the used device. Accordingly, transition feedback should be noticed right after the virtual object exceeds the boundaries of the AR FOV. However, in order to take the previously mentioned delay times into account correctly, the vibrotactile cues are triggered slightly before the virtual object touches the FOV´s virtual boundaries, depending on its current velocity.

4 Feedback Modes

In the following subsections we provide the definition and the design implementation of respectively proximity and transition cues in different sensory modalities. In the design and creation of cues, we have been inspired by existing and established visual and non-visual out-of-view guidance methods. If necessary, we adapted and extended these techniques to our needs. These cues were combined into a final mode, starting with the proximity cue (see Fig. 3a) when an augmented information approaches and triggering a short transition cue (see Fig. 3b) as soon the augmentation passes the border of the FOV. Feedback modes are designed to be a subsequent, non-overlapping combination of one consistent proximity cue followed by a short transition cue. The change from proximity to transition is triggered almost seamlessly ($\sim$ 90 ms in between), resulting in a smooth, sequential combination of both cues. Multiple variants of combinations of multisensory proximity and transition feedback were tested to determine personal preferences and effectiveness (see Study 1, Section 6.2). The provision of proximity

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics



(a) **Proximity cues:** For auditory proximity ❶, the virtual object emits a continuous spatial sound. Visual proximity ❷ is represented by a proxy symbol, moving analogous to the virtual object across the border.

(b) **Transition cues:** Auditory transition ❸ triggers a short sound notification as soon as the virtual object touches the border of the FOV. Visual transition ❹ highlights the proxy symbol in red color on the inside of the EdgeRadar border. Tactile transition ❺ provides a short vibration burst at the corresponding side on the tactile headband.

(c) Example for the **Audio-Tactile mode**. The virtual object is approaching from the right side, emitting the continuous spatial sound for increasing awareness and localization. As soon as the object touches the border of the FOV, the vibration motor on the right side of the tactile headband triggers a short but noticeable vibration burst.

Fig. 3: Description of how the different (a) proximity and (b) transition cues work. One mode consists of a combination of proximity and transition cue in a subsequent order. (c) shows one possible example of the combination of an auditory proximity cue with the tactile transition cue (Audio-Tactile mode). Note that the borders of EdgeRadar are not visible when no visual cues are involved. *Best seen in color.*

and transition cues depends on the location of a target object of interest relative to the AR FOV. The target location can dynamically change over time due to movements of the user or by the object moving itself (compare to Fig. 1). As all modes are directly coupled to the FOV of the used device, head rotations are generally possible and do not affect the feedback behavior. However, head rotations were not part of the conducted experiments. In the following, we will further describe our design choices and justifications for each feedback modality.

4.1 Proximity cue

4.1.1 Definition

The proximity cue informs the user about the presence of the object by providing spatiotemporal information like relative distance and speed of the object. It is triggered and applied continuously when an object of interest falls within a radius from the center of the visual field, but has not yet entered the FOV. In a real-world application, this radius (in our experiments on average 180°, see Fig. 5), should be adjustable by the user depending on the application context and spatial distribution of the objects of interest. The cue can be used as a navigation aid and can help to approach the object (e.g., move the head) in a targeted manner. Alternatively, the cue can be used passively to improve the awareness of the overall situation without specifically putting visual attention at the object, see Fig. 1.

4.1.2 Design

For the proximity cue, we used either visual or auditory feedback. The duration of the proximity cue lasted from 1.5-2 seconds depending on the starting position of the target object. In the case of *visual proximity* feedback, we used EdgeRadar [37], a 2D visualization technique that utilizes a border along the screen edges to visualize out-of-view objects. Objects are represented as proxy dots along the screen borders, conveying the relative distance and direction of their off-screen counterparts by compressing them proportionally into the border (see colored proxy dots in Fig. 1 or Figs. 3a and 3b). This method has the advantage that it takes a relatively small portion of the screen space at the borders while the center of the screen remains free for other information (represented by visual stimuli in the concurrent visual task in study 2). EdgeRadar has also shown to be particularly useful for tracking moving out-of-view objects accurately [37]. Although methods like 3D Radar have proven to be very useful in tracking off-screen object trajectories as well [35], the visualization still suffers from a certain overlapping and cluttering with on-screen augmentations, see [37]. In addition, such overview approaches tend to increase the cognitive load as it is required to mentally integrate all views, while context information along the borders is more in line with the human frame of reference [35]. Also, more recent and highly efficient out-of-view guidance approaches such as EyeSee360 (which is based on EdgeRadar) involve a high workload and occupy even more display space, especially for devices with a restricted FOV [33]. Therefore, we decided to choose EdgeRadar as visual reference method, as it has been shown to provide good performance for the intended use as proximity feedback while the FOV still remains clutter-free. In terms of visual proximity, the out-of-view object is visualized as a blue proxy symbol along the border approaching the users FOV (see Fig. 3a) until it reaches the inner edge of the border.

Audio proximity uses a spatial sound at 250 Hz originating from the center of the target sphere. This frequency has been chosen because it was shown that localization accuracy is highest for low-frequency stimuli centered at around 250 Hz [124] and frequency discrimination works sufficiently well [16] in this range. As we used spatial audio, users were able to determine the direction of arrival and distance of a sound source. The audio proximity cue was then played continuously while the target sphere approached the users FOV and stopped when it was replaced by a transition cue.

In terms of the tactile modality, we wanted to ensure comparability across modalities also in follow-up investigations, e.g. for targets approaching from vertical and oblique angles. However, the provision of proximity cues would not be easily possible with the current tactile interface design as tactors are distributed around the head at a fixed height ("headband"). We avoided the adoption of solutions that cover the head with tactors (e.g., [50]), as this ultimately causes restrictions in real-life usage. This is in contrast to audio and visual feedback, where spatial cues can easily be provided. Furthermore, in the prospect of an unimodal cue combination (Tactile-Tactile), it would become difficult to make different vibration patterns effectively distinguishable for the user [97]. As a consequence, we have decided to keep the respective system as a basis with no tactile proximity cues for the time being, focusing mainly on visual and audio proximity cues.

4.2 Transition cue

4.2.1 Definition

A transition cue is a short but clear signal triggered as soon as the AR object of interest contacts the display border. It indicates that the object is potentially visible now (available for interaction or as a source of information) or - in the opposite case - just left the AR FOV and cannot be visually detected at the moment. The transition cue also provides directional information about where the AR object entered or left the FOV. A transition cue can be especially useful in information-rich and dynamic AR environments, where an object of interest can be easily overlooked due to information overload, occlusion or distraction.

4.2.2 Design

The transition cue relied on either visual, auditory or tactile feedback. Regarding the duration of transition cues, values were chosen that are based on the design of warning signals, as their objective is to attract attention. It is suggested that auditory alarm signals should be at least

200 ms to allow the ear enough time to integrate the warning signal [88]. For tactile alerts values between 200 ms for short vibration respectively 600 ms for long vibration are recommended [97]. Based on these values, we chose a fixed duration of 300 ms per transition cue (visual, audio and tactile) to make it well perceptible, but not too intrusive or annoying over time. Furthermore, all transition cues had the same length for comparability.

In terms of *visual transition*, we extended the existing EdgeRadar method by a transition cue as soon as the augmented information enters the FOV. We introduced a short blinking color change from blue to red as this visual change has been shown to work well in the periphery and potentially being able to incite drawing attention in relation to FOV [56]. Furthermore, the red color was chosen in particular because it provides a good color contrast to the proximity cue and was shown to have an impact on attention behaviour [58]. With regards to *auditory transition*, we introduce a short advisory tone as auditory transition cue. It is composed of a fast tonal change from about 390 Hz to 440 Hz to attract the user's attention [77] to an information entering the FOV. Furthermore, the sound has a frequency spectrum that is acoustically well perceptible for humans and also sufficiently different from the auditory proximity cue. The *tactile transition* cue is inspired by the tactile method of [69] that uses an intensity modulation between 50-200 Hz. We chose to use the maximum intensity of 200 Hz for the tactile transition cue as this intensity has shown to provide noticeable but not unpleasant vibration burst at 300 ms during the initial pilot study.

5 Environmental noise

In real-world situations the user is frequently exposed to varying levels of environmental noise. By the term noise we refer to perceptual input which can interfere with the perception of other stimuli. With regard to visual noise, variable light intensity, background colors, and optical flow can typically affect general perception in AR [30, 57]. In particular bright light intensity can limit projection in AR displays [57]. Light intensity changes depending on factors as location (e.g. indoors versus outdoors), time of the day (day versus night) and weather (e.g. sunny versus cloudy) [30, 57]. Background colors and textures of common objects in an urban setting can also vary and affect perception in AR [30]. As for audio noise, it is emitted in many locations and situations with different sources and sound pressure levels [17]. Only few situations in which vibro-tactile noise occurs can affect perception, e.g. low-frequency whole body vibration [5]. The presence of environmental noise can suppress the perception of cues due to close and direct neural interactions among sensory inputs [40]. Research addressed perceptual suppression within the visual [10], auditory [116] and vibrotactile [111] modality but also cross-modal effects [40]. However, cues provided in one modality can also amplify the perception of cues in another modality through additive or facilitatory integration [40, 103]. So, multisensory stimulation can enhance noticeability [5] and improve task performance [13, 45, 91]. A predictive cue can also improve awareness and the perceptual sensitivity for a subsequent masked target [72]. Cues of different modalities may also supplement each other in an environment in which variable levels of noise might mask a signal if only one modality was used [38]. These advantages motivated the integration of multisensory modes for comparison.

5.1 Noise conditions

To address environmental noise we included two different intensity levels of combined visual and auditory noise in out studies. As we expected that a certain amount of noise is present in most application scenarios we did not include a condition without noise. We further assumed that tactile noise occurs less frequently in general and often does not affect perception, so we limited ourselves initially to visual and audio noise. Therefore, we created conditions with reduced and increased level environmental noise, resembling real outdoor conditions.

5.1.1 Reduced-noise condition

The reduced-noise condition contained a uniform visual background that produces a minimum of distraction, dimmed indoor lightning conditions and low levels of auditory ambient noise.

Users performed this part of the study in 1.5 meters distance of a uniformly gray background in form of a large gray pinboard (see Fig. 7a). The background filled the entire FOV of the AR device during use. Users were facing towards the inside of the lab room, which exposed them to lower ambient light. Furthermore, in contrast to the increased-noise condition the loudspeaker was turned off to ensure an acoustically stable, calm environment. Note that in this condition, minor external influences of indoor lighting and soft auditory background noises were not completely eliminated. Although it was ensured that lighting conditions were constant, indoor lighting can affect the representation of augmentations of the used AR device [57]. Consequently, the user was still exposed to a minimum amount of noise in this condition, while its influence can likely be considered as negligible.

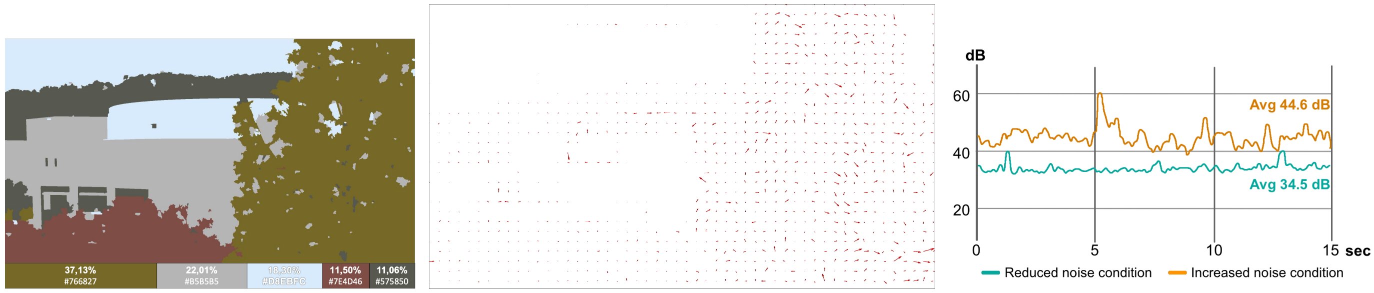
Regarding the lightning exposure in the reduced-noise condition, we measured an average illuminance right in front of the user on eye level of about 210 lux, which is in the range of an average office illumination [84]. In terms of ambient auditory background noise, an average loudness of the laboratory environment of about 34.5dB was measured (that corresponds to the acoustics of a library) on the basis of an audio recording made during this condition (see Fig. 4c). Also, the contrast ratio for the visual proximity cue (proxy symbol) and the background was calculated following WCAG 2.1 guidelines [112] by using the formula $(L1+0.05)$ / $(L2+0.05)$. $L1$ is the relative luminance of the lighter of the colors, and $L2$ of the darker of the colors. The analysis revealed a contrast ratio of about 2.66:1, which can be interpretable as perceivable for user interface components according to WCAG 2.1 guidelines. In terms of auditory ambient noise, the audio proximity cue was measured on average at about 50dB at ear level. This roughly resembles the WCAG 2.1 audio guidelines, which recommends a 20dB difference to background sounds to make audio content clearly perceivable to the user.

5.1.2 Increased-noise condition

The increased-noise condition included a vivid outdoor background (see Fig. 7b) with a heterogeneous color palette (see Fig. 4a). It contained lighting, some optical flow and a higher level of ambient auditory noise. Users were seated about 1.5 meters in front of a window. The outdoor environment viewed from the window made up the entire background of the AR display as can be seen in Fig. 7b. We chose a mixed urban/landscape as background with some natural movements of trees and people. The experiment was carried out between 11 am and 3 pm to ensure that all participants experience similar light conditions for both conditions. We also ensured that we performed the experiment only in comparable weather conditions and that the user was only exposed to indirect, reflected light. To generate a controlled level of auditory noise a loudspeaker in 1 meter distance in front of the user was placed to provide prerecorded background noises. We chose prerecorded sounds which are characteristic for many everyday scenarios (coffee shop ambient sounds including conversations of people and soft traffic sounds).

An average illuminance of about 2000 lux was measured in front of the user, which corresponds to typical lighting on a cloudy day [30]. Furthermore, we analysed the background scenery in the increased-noise condition by using fuzzy c-means algorithms (as proposed in [61]) in order to perform a color segmentation to assess the most dominant colors of scene. The resulting analysis showed that the view was dominated by natural colors (green/brown/reddish) and urban tones (shades of grey), see Fig. 4a. Also, in contrast to the reduced-noise condition, motions do exist in the background of the environment, which potentially affects user performance [52]. On this basis, we performed an optical flow analysis based on a ten-second video recording of the outdoor environment. Dense optical flow analysis (Gunnar Farneback method) shows generally minor motions around the left and the center part of the background environment while the right side shows more motion (see Fig. 4b). The user was exposed to an average auditory noise of about 44.6dB (resembling a sound intensity similar to light traffic noise), which are about 10dB more than the user experiences in the reduced-noise condition – therefore roughly doubling the loudness on a subjective level [17] (see Fig. 4c).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics



(a) Color distribution in the increased-noise condition. Dominant colors are isolated into groups using a fuzzy c-means clustering algorithm.

(b) Optical flow estimation during the increased-noise condition. Smaller motions (light optic flow) were detected in the background, primarily on the right side.

(c) Audio noise comparison between both conditions. The increased-noise condition shows an increase of about 10 dB compared to the reduced-noise condition.

Fig. 4: Noise analysis for the increased-noise condition, containing the examination of the color distribution (a) and the motion analysis (b) of the environmental background, as well as the comparison of the loudness between both conditions (c). Further information can be found in Fig. 7.

Regarding the noise ratios for the increased-noise condition, the contrast analysis showed a decrease of contrast of about 25% of the proxy symbol, which made the visual cue less visible than in the reduced-noise condition. However, it needs to be mentioned that dynamic conditions and illumination changes can abruptly affect contrast and other properties on OST devices. Regarding auditory perception, the proximity audio cue was measured about 5.4dB louder than the background ambient noise at ear level, which makes the cue generally more difficult to perceive. However, since the frequency of the audio cue was chosen to be highly discriminable (see Section 4.1.2), the stimulus should still be reasonably well distinguishable despite the exposure of auditory noise.

6 User Studies

Following, we will use the following notation for the used cue combination/modes, consisting of the proximity cue at the first element and the transition cue at the second element separated by a dash: Audio-Audio (AA), Audio-Tactile (AT), Audio-Visual (AV), Visual-Audio (VA), Visual-Tactile (VT), Visual-Visual (VV). We performed two user studies to compare modes on each noise intensity level in narrow FOV AR. Study 1 addressed RQ_1 of which mode is most preferred by users to make aware of an augmentation out-of-view and its transition inside the field of view. Study 2 addressed RQ_2 to identify the best mode regarding target localization and reaction to targets which enter the FOV under divided attention.

To examine the effect of the feedback mode and noise level, the number of other influential factors had to be kept low to avoid too much variability in data. For this reason, for the time being we did not include other AR objects as distractors in our studies. Furthermore, even though finally all feedback modes are intended to be usable with all possible object trajectories we initially restricted ourselves to **horizontal object movements**. The user and many relevant dynamic objects in the surroundings usually follow trajectories on the horizontal plane, since their motion is ground-based due to gravitational forces. It has further been shown that typical eye movements mostly feature a predominance of horizontal saccades (leftwards or rightwards) compared to other directions (vertical or oblique angle saccades) when viewing scenes [29]. Regarding search behaviour in AR, there also seems to be a tendency for horizontal search patterns [14], which causes targets to enter the FOV more often through horizontal head movements [56].

For data visualization in both studies, boxplots were used to mark the largest and lowest observed data point that falls within 1.5 times the interquartile range from the upper and lower quartile.

6.1 Pilot Studies

To prepare study 1 and fine-tune the designed cues, we did extensive pilot testing. We presented the first design of combined feedback modes (AA, AT, AV, VA, VT and VV) to 5 users under controlled conditions in a simulated AR environment (performed in VR). Virtual environments were used in this case to ensure the same preconditions (e.g. lightning, visual and auditory noise) for all users [93], allowing them to mainly focus on the perceived cues. Afterwards, users filled in a questionnaire including items which addressed usability issues, preferences and improvement suggestions for both, individual cues and combined feedback. We checked items for values tending more to the negative than to the positive end point. Overall, all cues were well received.

In the next step, we performed a second pilot study – again in a simulated AR environment – with 10 users (300 trials) where the new sound was used as proximity cue. Users were presented all modes in the course of the task procedure used in study 1. Subsequently, they filled in a questionnaire that was similar to the pre-test but extended by items on the task procedure. Data showed that users understood the procedure and learned the feedback in a short time. They received it positively and could apply it to enhance their awareness of moving AR target objects in terms of proximity and transition events.

Prior to study 2 we performed pre-tests with 5 users to determine the optimum time for target stimulus presentation for a reaction time task (described in more detail in Section 6.3). The aim was to achieve that users had to consistently keep their eyes in the central display area to avoid missing the target and to ensure an appropriate amount of cognitive load (medium difficulty level). We found that presenting the target for 750ms met the requirements. We then continued pre-testing by running the 5 participants through the entire procedure of study 2 before interviewing them on the usefulness of cues and usability. Explorative data analysis showed tendencies as in study 2. Care was taken to ensure that participants from the pilot testing for study 2 did not participate in the main study 2. However, participants could participate in study 1 and study 2. We did not assume any influence on the results because a significant period of time passed between study 1 and the later study 2 (∼6 months). Also, tasks in study 1 and 2 were of different nature: Study 1 assessed general preferences and study 2 divided attention reaction performance. Finally, sufficient training procedures were ensured before each experimental run.

6.2 Study 1 - Mode Preference

Study 1 addressed the following research question. **RQ_1:** *Which combination of visual, audio and vibro-tactile cues is most preferred by users to make them aware of an augmentation out-of-view and its transition inside the field of view? Is there an effect of different visual and auditory noise levels on preferences?*

As in related work visual and auditory cues have been shown to work well in spatial localization tasks (see the Related Work Section), we did not make specific assumptions about which modality would work better as proximity cue. Regarding **RQ_1** we hypothesized that modes that combine different modalities are in general more often preferred than unimodal combinations. This would be compliant with [59] where the multimodal interface yielded higher user satisfaction than the unimodal variant. We also also expected a higher preference of multimodal modes due to spatial synergies in the attentional processing of information across sensory modalities [24]. Among the multimodal modes we further assumed that modes with tactile and audio transition cues would

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

be preferred over modes with the visual variant as they have a higher temporal sensitivity [109]. With regard to the effect of visual and auditory noise we expected the perception of modes with visual or auditory cues to be impaired while modes that include tactile cues would still be well perceivable. Based on these considerations, the following hypotheses emerge:

H$_1$: *We hypothesize that in the condition with reduced noise bi-modal modes with an audio or tactile transition cue component, namely Visual-Audio, Audio-Tactile and Visual-Tactile would be more often preferred than the other modes (Audio-Audio, Audio-Visual and Visual-Visual).*

H$_2$: *We expect that under the influence of visual and auditory noise bi-modal modes that include tactile cues, namely the Audio-Tactile and Visual-Tactile mode, to be more often preferred than the other modes which are composed of only visual and/ or audio cues.*

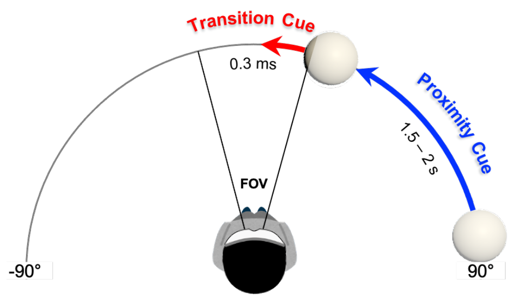


Fig. 5: Top view: During the awareness task, the target sphere follows a circle path around the user on the horizontal plane. The user receives the proximity cue as long as the sphere follows the path outside the FOV (blue line). As soon as the sphere touches the border of the FOV, the transition cue is triggered (red line).

6.2.1 Method

Study 1 included **10 participants** (1 female) aged between 22 and 34 ($M = 25$, $SD = 3.4$), more than half of them wears glasses (6/10). They were recruited at the university (students and employees). The majority of users (7/10) played video games at least weekly. Almost half of them (4/10) was experienced with VR glasses as they use it at least weekly.

For the main part of the preference study, a 6 x 2 **within-subjects design** was employed to explore how the factor mode (cue combinations AA, AT, AV, VA, VT and VV) and noise (reduced-noise, increased-noise) affected the number of times the feedback mode was preferred over other modes (preference score).

For subjective feedback evaluation, a questionnaire was used which included different items on cue specific aspects of spatial and temporal perception all of which were answered by each user. The questionnaire analysis was oriented toward specific findings of the main part of the study. As a result, proximity and transition cues were compared on each noise factor level. For both audio and visual proximity cues (independent variable), ratings were provided on 3 different items (dependent variables), namely the usefulness for the perception of target direction, the estimation of target speed and the interpretation effort. For visual, audio and tactile transition cues (independent variable) users rated 3 different items (dependent variables) on the usefulness for the perception of transition time, the perception of transition direction and interpretation effort. See 3.1 in the appendix for the exact wording of respective items (1-6).

6.2.2 Procedure

Visual, audio and vibro-tactile cues were combined to inform the user on an AR object which is 1) moving outside the AR FOV and 2) entering or leaving the AR FOV. Users were instructed to determine which mode (combination of two cues) is most appropriate to make them aware of these two stages of the process. Users were informed in advance about the intended functions of the proximity and transition cue. At this point it must be noted that although we make specific demands on each proximity and transition cue, we address their interplay in particular. Users were not presented proximity and transition cues in isolation but in combination (sequentially).

In the beginning of each trial the user was instructed to look straight. After the trial started, the user was free to choose a focus point at any time since study 1 did not involve a secondary task. We assume that in visual conditions, the user focuses on the proxy symbols of EdgeRadar during the proximity or transition phase as the feedback can be interpreted with less effort and more accurately in the focal area. An augmented sphere appeared outside the FOV where it was not visible to the user. This starting position was fixed on $+90°$ or $-90°$ to cover a maximum radius of $180°$ on the horizontal axis between trials (see Fig. 5). Next, the sphere approached the AR FOV from either the left or right side. Then it entered the FOV (and thus became visible), passed through and left it on the opposite side. Accordingly, cues of a mode were always presented to the user successively merging into one another (cue sequence: proximity, transition, no cue, transition, proximity). The sphere always entered and left the AR FOV in the center of the left or right display border. Then the process was repeated, starting from the opposite side. As soon as the sphere spawned at either $90°$ or $-90°$, the user was provided with *proximity feedback* indicating the spatial location of the out-of-view object. The feedback continued while the sphere moved along a circle path around the user on the horizontal plane (see Fig. 5) to ensure a constant distance between the user and the object. When the sphere crossed the border of the FOV – thus entering or leaving it – the user was provided with a *transition cue*. The six modes (AA, AT, AV, VA, VT and VV) competed against each other in the course of pairwise comparisons. In each trial, the user was presented two modes in subsequent order. Next, the user had to choose which mode was more preferred with regard to raising awareness for a moving out-of-view object and its transition into and out of the FOV (forced-choice). Using direct comparisons allows to reduce problems such as response distortions [119], reference group effects [20], and non-linear distances between anchors [121]. The resulting dependent measure was a preference score. To make the choice two buttons were shown, each representing one of the presented feedback modes. Modes could be pointed at and selected by using the Xbox controller.

After having made the choice, the next trial started with another pairing. Each mode competed against the other modes two times, resulting in 30 (15 pairings x 2) pairwise comparisons in randomized order. We exposed the user to a particular pairing for two times to change the order in which both modes were presented to avoid order effects. Overall, each subject took about 45 min to complete study 1 (about 15 min for each noise condition including 2 min introduction, 3 min training and 10 min questionnaire). The order of noise conditions was counterbalanced across participants.

A **questionnaire** composed of two parts was used to receive subjective user ratings. Part 1 was presented after each noise condition and part 2 at the very end. Users could indicate their level of agreement on 7-point Likert items with annotated end points 1 = "strongly disagree" and 7 = "strongly agree". In the first part of the questionnaire items addressed the usefulness of different proximity and transition cues with regard to temporal and spatial perception (see section 6.2.1). The second part of questionnaire comprised items on overall comfort and usability (see 3.2 in the appendix).

6.2.3 Results

Mode preference analysis. A total of 300 trials was included in the analysis (30 pairwise comparisons x 10 users). To compute the preference score, each time a mode was preferred over another mode in direct comparison, it scored one point. As each mode competed against 5 other modes for two times, a mode could be preferred by one user maximum 10 times. A repeated-measures ANOVA was used to examine the effect of feedback mode and the interaction of mode and noise on the preference score. Pairwise comparisons with Sidak correction were used for post-hoc testing.

A significant main effect of mode ($F(5,45) = 29.48$, $p < .001$, $\eta_p^2 = .77$) and an interaction of mode and noise was found ($F(5,45) = 2.47$, $p = .03$, $\eta_p^2 = .23$). Preference scores by mode and noise condition are visualized in Fig. 6.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

Simple effects were examined by comparing different modes at each noise factor level. In the reduced-noise condition as expected ($\mathbf{H}_1$), AT, VT and VA were more often preferred than VV (each $p < .001$) and AV ($p_{AT} = .003$, $p_{VA} = .035$, $p_{VT} = .013$). There were no differences between the modes AT, VT, VA. However, AA was also popular and did not differ significantly from AT, VT, and VA either. In the increased-noise condition we expected AT and VT to be more often preferred than other modes ($\mathbf{H}_2$). Accordingly AT had a higher preference score than AA ($p = .001$), AV ($p = .001$), VV ($p < .001$) and VA ($p = .045$). VT on the other hand was not significantly superior to AA ($p = .09$), AV ($p = .064$) and VA ($p = .157$) and showed a higher score compared to only VV ($p < .001$). However, the AT and VT mode did not differ significantly either. We also created final ranking lists across all participants, in total and by noise condition, ordered by the preference score (see Table 1). As 10 participants took part in the study the maximum total score was 100 in each the noise with increased and with reduced noise.

Mode	Reduced- Noise	Increased-Noise	Total
AA	49	49	98
AT	72	88	160
AV	28	31	59
VA	61	49	110
VT	82	75	157
VV	8	8	16

Table 1: Number of times a feedback mode was preferred over another mode in VR and in AR in each noise condition (maximum possible score = 100) and in total. Pairings were presented to 10 users twice in each condition.

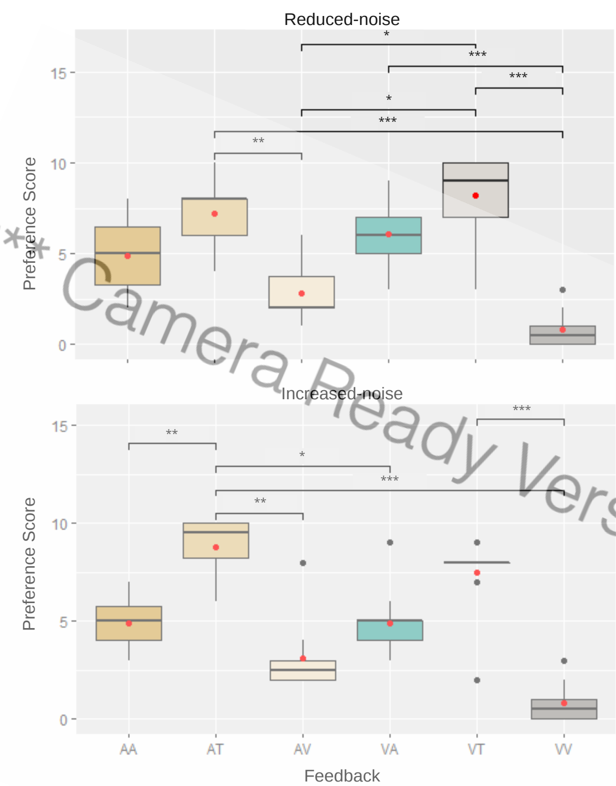


Fig. 6: Study 1, condition with reduced noise (top) with comparisons of Preference scores of AT, VA and VT each versus AA, AV and VV feedback (see $\mathbf{H}_1$) and Noise condition (bottom) with comparison of Preference Scores of AT and VT each versus VA, AA, AV and VV (see $\mathbf{H}_2$). Only significant differences are visualized.*p <.05, **p<.01,***p <.001.

Questionnaire Analysis. In addition to the forced-choice decisions, we analyzed questionnaire data related to individual cues.

Specifically, we compared ratings on Likert type items for the visual and the audio proximity (Wilcoxon signed-rank test), ratings for visual, audio and tactile cues (Friedman test), and pairwise post-hoc comparisons (Wilcoxon test). If necessary, Sidak correction was applied. Fig. 8 shows the distribution of answers for each item and noise condition.

For the **condition with reduced noise** the comparison of ratings on Likert type items for proximity cues showed no significant differences between visual and audio cues with regard to the item on helpfulness of the cue for perceiving the direction of the sphere, helpfulness for estimating the speed of the sphere and interpretation effort. As expected audio and visual proximity cues performed comparably well here. The comparison of transition cue ratings showed significant differences between cue modalities regarding ratings of the item on helpfulness of the cue for perceiving the exact moment when the sphere entered the FOV ($\chi^2(2) = 6.64$, $p = .036$, $W = .684$) and the interpretation effort to interpret cues ($\chi^2(2) = 11.04$, $p = .004$, $W = .744$). However, post-hoc pairwise comparisons were only significant for interpretation effort, indicating that users needed less concentration to interpret audio than visual transition cues ($Z = 2.46$, $p = .041$, $r = 0.778$). Significant Friedman test and visualized data indicate a tendency to our expectation that visual transition cues receive lower ratings. However, post-hoc test were not significant for most items. Thus, differences were not strongly pronounced. Ratings on the item on helpfulness of the cue for perceiving where the sphere entered the FOV did not differ between the transition cue modalities.

For the **condition with increased noise** users indicated significantly higher ratings for the audio than for the visual proximity cue with regard to the helpfulness of the cue to perceive the direction from which the sphere was approaching ($Z = 2.67$, $p = .008$, $r = .844$). The audio proximity cue was also rated as more helpful to estimate the speed of the sphere ($Z = 2.11$, $p = .035$, $r = .667$) and as interpretable with less concentration compared to the visual variant ($Z = 2.45$, $p = .014$, $r = .775$). While low ratings for visual proximity cues confirmed our hypothesis, the audio variant was much less affected than expected. Regarding ratings for transition cues a significant difference was found between modalities for the item that addressed the helpfulness of the cue to perceive the direction where the sphere entered the FOV ($\chi^2(2) = 11.67$, $p = .003$, $W = .05$), helpfulness of the transition cue to perceive the exact moment when the sphere entered the FOV ($\chi^2(2) = 13.69$, $p = .001$, $W = .332$) and interpretation effort ($\chi^2(2) = 14.89$, $p < .001$, $W = .552$). In each case ratings for the audio and tactile transition cues did not differ from each other significantly but received higher ratings than the visual cue. They were rated as more helpful to perceive the sphere's direction (A vs V: $Z = 2.56$, $p = .030$, $r = .81$; T vs V: $Z = 2.55$, $p = .033$, $r = .806$) and more helpful to perceive the exact moment when the sphere entered the FOV (A vs V: $Z = 2.69$, $p = .021$, $r = .851$; T vs V: $Z = 2.56$, $p = .033$, $r = .81$). Users also indicated a higher interpretation effort when interpreting visual transition cues compared to audio ($Z = 2.56$, $p = .033$, $r = .81$) and tactile ($Z = 2.57$, $p = .03$, $r = .813$).

(a) Condition with reduced noise. (b) Condition with noise.

Fig. 7: The experiment conditions from user perspective with **(a)** reduced noise and **(b)** increased noise. Both images illustrate the awareness task (approaching sphere) and focal attention task (digit at the center of FOV) in study 2. *Note: The approaching spheres are illustrated for comprehension purposes for the awareness task and were only visible in study 1, but not in study 2. Also, in the condition with increased noise (b), ambient audio background noise was provided.*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

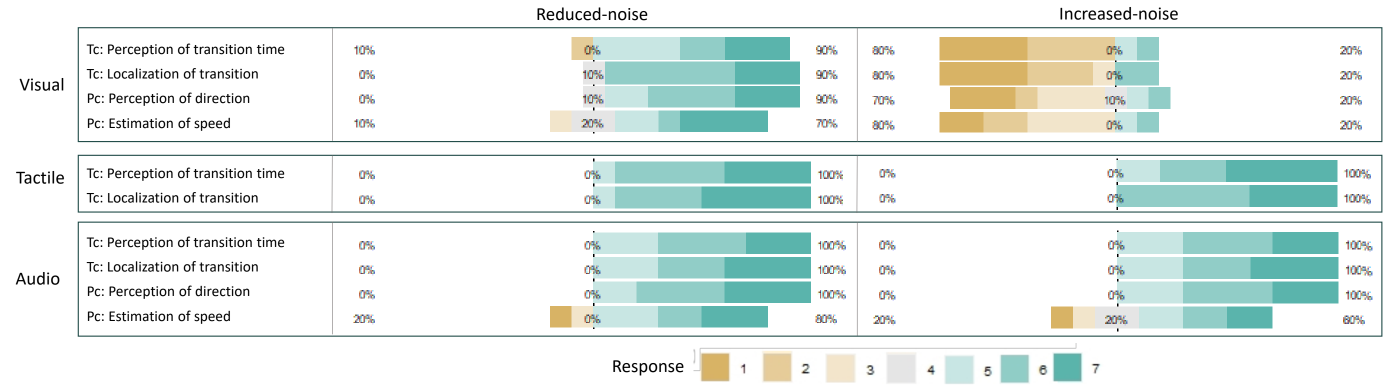


Fig. 8: Study 1. Users indicated their level of agreement on 7-point Likert type items (ranging from 1= "strongly disagree" to 7= "strongly agree") on how the respective cue helped them to perceive temporal and spatial events of an AR object under conditions with reduced noise (left) and with noise (right). Pc = Proximity cue, Tc = Transition cue.

To sum up questionnaire rating results, all variants of proximity and transition cues received positive ratings in the reduced-noise condition. As proximity cue visual and audio worked equally well. Regarding the transition cue, the visual variant showed a slight tendency to be less suited than audio and tactile. In the increased-noise condition, the visual proximity feedback did no longer sufficiently support target awareness and was strongly rejected while audio proved to be more robust here and could still be interpreted well. For the transition feedback audio and tactile were clearly superior to the visual variant when environmental noise was present. Mode preference choices are reflected in questionnaire results. The VV mode was least preferred, followed by AV which are both composed of the lower rated visual transition cue. However, in the condition with increased noise the visual proximity cue which received rather low ratings in the questionnaire was still quite often preferred when being combined with the tactile transition cue to the VT mode. VT was not less often preferred than modes with audio proximity cues (AA, AT, AV).

6.2.4 Implications for study 2

For the performance study we needed to reduce the number of feedback modes to be able to perform a more extensive testing within a reasonable period of time (maximum one hour) to avoid fatigue effects. The decision was made based on the analysis of the preference score (resulting from forced-choice decisions) and not questionnaire ratings as they were expected to be less biased (see Procedure in Section 6.2.1). In addition, users selected the preferred feedback immediately after its presentation in the frame of the forced-choice procedure while the questionnaire was presented after each noise condition. However, people are not good at reporting detailed information about past mental events, even recent ones, as they are tending to overgeneralize and overrationalize [28]. Furthermore, questionnaire ratings addressed proximity and transition cues separately while within the forced-choice procedure combined feedback was provided. The decision which feedback to prefer was therefore also dependent on the interplay of both proximity and transition cue. Based on the preference score analysis we decided to include AT and VT in study 2 as they had a high preference across noise conditions. We also included VA in the second study as a baseline against which we could test as in general, visual-audio techniques are more widely used to provide spatial information than tactile techniques. In addition, based on descriptive data, VA was the third most preferred feedback mode (for a complete list of rankings, see Table 1). We also chose VA as baseline for study 2 as it is more comparable to VT and AT regarding its bi-modal structure (feedback in which cues from two sensory modalities are combined) compared to e.g. AA.

6.3 Study 2 - Mode Performance

Study 2 addressed the following research question: **RQ$_2$:** *Which cue combination mode performs best regarding target localization and reaction to targets that enter the field of view when a concurrent visual task is performed? How does the level of visual and auditory noise affect performance in both tasks?*

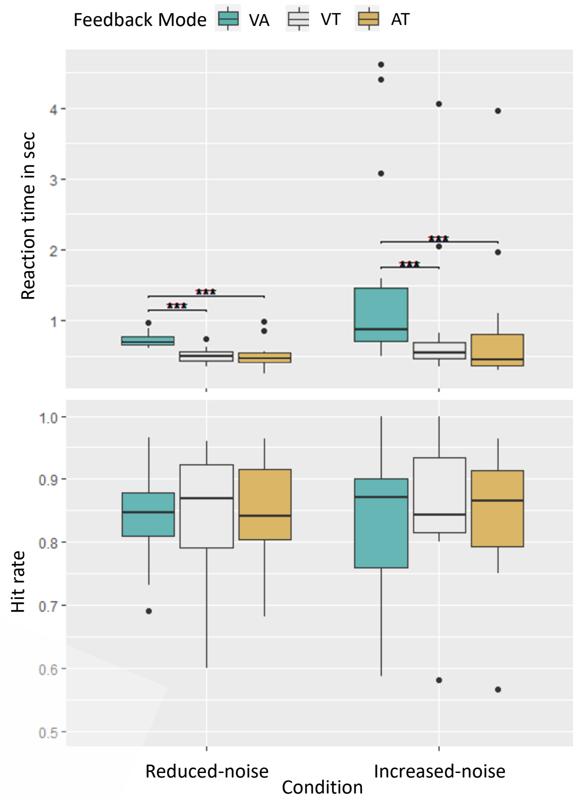


Fig. 9: Average reaction time in the awareness task (top) and hit rate in the focal attention task (bottom) for VA, VT and AT modes by noise condition in study 2. ***p <.001.

Regarding **RQ$_2$** we expected that with reduced noise AT would yield faster reactions than VA and VT as for the latter visual attention has to be divided when interpreting visual cues and performing a concurrent visual task. We further assumed that VT would perform better than VA due to faster reactions to tactile cues [81]. Under increased noise, we expected a weaker detrimental effect on the auditory than on the visual proximity cue based on results in 6.2.3. Thus, we assumed AT would perform best, followed by VT and finally VA in which both sensory channels are expected to be affected by noise. Regarding the concurrent visual task performance we assumed a lower hit rate for VA and VT than for AT as visual attention has to be divided between two tasks. We expected that, when noise is increased, the perception of visual cues which support the reaction to AR targets would take up even more visual resources resulting in less free capacities to perform a visual task simultaneously. Accordingly, we assumed that noise would reduce hit rate in the concurrent visual task with VA and VT but not with the non-visual mode AT. The associated hypothesis summarizes expectations in the following:

H_3: *When localizing and reacting to AR targets (noise and condition with reduced noise), we expected a main effect of mode, assuming that reaction time to incoming out-of-view AR objects is fastest for AT followed by VT and slowest for VA.*

H_4: *In the concurrent visual task (noise and condition with reduced noise) we expected a main effect of mode, assuming a lower hit rate for VA and VT than for AT.*

H_5: *In the concurrent visual task, we expected an interaction effect of noise and feedback mode, assuming that under noise compared to the condition with reduced noise hit rate decreases only for VA and VT and not for AT.*

6.3.1 Method

Study 2 included 14 **participants** (13 male and 1 female) aged between 22 and 33 ($M = 24.3$, $SD = 2.9$). The majority played games at least weekly (64%) and half of the users also uses VR devices weekly or more frequently. Furthermore most of the users (64%) had already tried out AR devices at least a couple of times or more. Users performed two tasks simultaneously: 1) the reaction to incoming AR targets with the help of cues (awareness task) and 2) a concurrent visual reaction respond task (focal attention task).

The main study employed a 3 x 2 **within-subjects design** to examine the effect of factors mode (VA, VT and AT) and noise condition (reduced-noise, increased-noise) on the time to react to transitioning out-of-view AR objects (awareness task) and the hit rate (visual focal attention task). Users completed one block with reduced noise and another block with increased noise. The order of blocks was counterbalanced across participants. In each block modes were presented in random order.

Furthermore, for subjective feedback evaluation, users answered items on spatial and temporal perception related to the awareness task. As in study 1, we examined whether specific results found in the main study part were reflected in the subjective feedback. For this reason, modes were compared on each level of the noise factor. For audio and visual proximity cues (independent variable), users rated 2 items (dependent variables), namely the usefulness for the perception of target direction and the prediction of target transition time. For audio and tactile transition cues (independent variable) users rated the usefulness for reacting fast to the transition event (dependent variable). See 3.1 in the appendix for the exact wording of respective items (1, 7, 8). Also, within the frame of the questionnaire users assigned a rank (1,2 or 3) to each mode in the condition with reduced and with increased noise. In each noise condition the assigned rank (dependent variable) was compared between modes VA, VT and AT). Finally, NASA-TLX ratings (dependent variables) were collected after each noise condition and compared between reduced and increased noise (independent variable).

6.3.2 Tasks and Procedure

Users were required to notice and react to an approaching AR target by using VA, VT and AT (see previous Subsection 6.2.4). The AR target object was again represented by a sphere. The starting position of the approaching sphere varied between trials and took values of either $+80°$ or $-80°$ on the horizontal axis, with a random offset between $\pm 0°$ to $10°$ to cover a maximum range of $180°$ (see Fig. 5). Doing so, it was not possible for the user to predict the time of entering by simply counting seconds. The AR sphere could enter the AR FOV from the left or right center display side and always moved at a constant speed level.

For the awareness task, the user was told to press the left or right shoulder button of the Xbox controller as soon as the sphere reached the respective border of the AR FOV, which was signalized by the transition cue. Before starting with the performance trials in which 50 trials of each mode were presented in random order, users completed 18 training trials for each mode. The first three training trials per mode, which were also presented in random order, included a feedback in which the scene was frozen and the user was displayed the AR target with its moving path after having pressed a button. In this manner the user got to know if the correct side was chosen and if the button was pressed too early or too late. The subsequent 15 training trials per mode were identical to the performance phase to get used to the procedure. During performance trials the AR target was not shown (not even after having entered the FOV) to (a) not provide the user with feedback on his performance (a visualization inside the FOV would inform the user that a target was present) and (b) to avoid occlusion regarding the focal attention task and distraction. Although the latter ones are important issues in this field of research, it falls outside the scope of this study.

Next to the awareness task, a second visual task has to be performed, the focal attention task. The task followed a go/no-go procedure [23] in which participants are generally required to respond to one stimulus but must not react to an alternative. This particular task was chosen as it required monitoring an area, mimicking a scenario in which the user is consistently visually occupied. Furthermore stimuli were presented in the central display as in a real-life AR scenario users would naturally turn to the area of interest and move it to the foveal regions. Stimuli were composed of random digits (1-9) presented in the center one after another for 750ms. Each time a 7 was presented, users had to press the button "A" of the Xbox controller before the next digit appeared to score a hit (see Fig. 7), otherwise a miss was registered. After each noise condition the user filled in the first part of the **questionnaire** on cue specific aspects and the second part on usability and comfort after having completed both conditions. The basic structure of the questionnaire and scaling of items in study 2 corresponds to the material used in study 1. As in study 2 modes VA, VT, and AT were tested, ratings were provided for visual and audio proximity cues and for audio and tactile transition cues. Study 2 took overall about 1 hour (about 20 min for each noise condition in the main experiment, 2 min for the introduction, 8 min training, 10 min for the questionnaire). The order of noise conditions was counterbalanced across participants.

As in study 2 cues were supposed to assist the user in reacting fast to incoming target objects, items also addressed helpfulness for predicting the time of target transition and the support of fast reactions. Respective items were presented for each of the tested modalities. In addition, we included the NASA Task Load Index (NASA-TLX) in the first part of the questionnaire to measure perceived task workload in each noise condition.

6.3.3 Results

The Aligned Rank Transform procedure was used [122] to analyze the effect of the feedback mode (VA, VT, AT) and noise level condition (reduced-noise, increased-noise) on performance measures in the awareness task (reaction time) and the focal attention task (hit rate). The analysis comprised 4200 trials (50 trials x 3 modes x 2 noise levels x 14 participants). The t-Test for dependent variables and Wilcoxon signed-rank test were used for post-hoc pairwise comparisons. The family-wise error rate was controlled by the Holm method. We will also report on workload measures which showed that the experiment was relatively demanding over time.

We found significant main effects of noise ($F(1,13) = 529.66$, $p < .001$, $\eta_p^2 = .976$), feedback mode ($F(2,26) = 272.00$, $p < .001$, $\eta_p^2 = .954$) and an interaction effect of both ($F(2,26) = 66.30$, $p < .001$, $\eta_p^2 = .836$) on reaction time in the awareness task.

The comparison of feedback modes showed a faster reaction with modes VT and AT compared to VA in both noise conditions (each $p < .001$, see Fig. 9). Furthermore, increasing the noise level led to slower reaction times only for the modes which included visual proximity feedback, namely VA ($p < .001$) and VT ($p = .003$) while performance was not affected when the AT mode was used ($p = .358$). Regarding the focal attention task neither an effect of noise nor feedback mode on hit rate was found (see Fig. 9). We also descriptively compared the ratio of missed targets between feedback modes for each noise condition. With reduced noise, 3.4% of targets were missed with VA, 1.3% with VT and 1.9% with AT.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

In the condition with increased noise 21.5% were missed with VA feedback, 9.6% with VT and 8.6% with AT. Due to differences in color and optical flow between the left and right display area in the increased noise condition (see Fig. 4b) and potential effect of handedness (faster reaction with the dominant hand) we also compared reaction time between trials when the sphere approached from the left compared to trials where it approached from the ride side. We differentiated between the condition with reduced and increased noise. There was neither a significant difference between left and right for the reduced noise condition where the background was uniformly gray with no optical flow ($p = 0.217$) nor for the increased-noise condition ($p = 0.153$).

For the **analysis of the questionnaire**, we applied statistical tests as appropriate. Specifically, we compared different ratings on Likert-type items between proximity cues (visual versus audio) and transition cues (audio versus tactile) in each noise condition (t-test for dependent samples), the NASA TLX-ratings between the condition with reduced and increased noise (the t-test for dependent samples), in cases that violated the normality assumption (Wilcoxon signed-rank tests), and the rank list of modes (Friedman test). For post hoc analysis Wilcoxon signed-rank tests were used. If necessary, Bonferroni correction was applied.

The comparison of different **proximity cues** in the condition with reduced noise showed that other than expected, the visual proximity cue was marginally more helpful to estimate direction than the audio variant ($t(13) = 2.03$, $p = .064$, $d = 0.54$). On the other hand, in the noise condition, as expected the audio proximity cue was rated higher than the visual variant regarding the estimation of the target direction ($t(13) = 3.43$, $p = .005$, $d = 0.917$) and prediction of the time of transition ($t(13) = 3.07$, $p = .009$, $d = 0.820$). These ratings are not in line with the actual performance as AT was not significantly better than VT in the awareness task. However, the superior performance of AT and VT compared to VA was reflected in questionnaire ratings for **transition cues**. Users stated to have reacted faster to the tactile variant in both conditions, with reduced noise ($t(13) = 3.45$, $p = .004$, $d = 0.922$) and with noise ($t(13) = 3.14$, $p = .008$, $d = 0.839$).

Users further provided a **rank list of the modes** ordered by which they liked most to perform well for the condition with reduced noise, with noise and overall. Median ranks differed significantly between modes in the condition with reduced noise ($\chi^2(2) = 7$, $p = 0.03$, $W = 0.25$), with noise ($\chi^2(2) = 19$, $p < 0.001$, $W = 0.679$) and overall ($\chi^2(2) = 16.71$, $p < 0.001$, $W = 0.606$). Post-hoc pairwise comparisons showed that with reduced noise the first ranked mode VT differed significantly from the last ranked VA mode ($p = .006$). In the noise condition and overall, both the VT (each $p < .001$) and the AT mode (each $p = .001$) were ranked significantly higher than VA. See Fig. 11 for the distribution of ranks by noise condition.

After each noise condition users provided **NASA-TLX ratings** on each subscale. While the subscales temporal demand, effort and frustration were not affected by noise, mental demand was rated significantly higher in the condition with increased noise ($t(13) = 2.3$, $p = .039$, $d = 0.614$) and performance was rated lower ($t(13) = 2.35$, $p = .035$, $d = 0.628$). The overall NASA-TLX score did not differ significantly between noise conditions but reached a value of 58.6 with reduced and 60.83 with increased noise. Contrasting the score with the quartiles of a global NASA-TLX analysis [31] showed that the mental workload of the task in our study is rather high as the value is around the 75% percentil (= NASA score of 60.00).

7 Discussion

In this paper we evaluated the acceptance and effectiveness of novel multisensory cue combinations for improving the awareness of moving out-of-view objects for narrow FOV augmented reality displays. We addressed one research question for each user study, which are discussed in detail below.

7.1 Mode Preference

RQ$_1$: *Which combination of visual, audio and vibro-tactile cues is most preferred by users to make them aware of an augmentation? Is there an effect of audio and visual noise on preferences?*

We hypothesized that with reduced noise bi-modal modes with audio and tactile transition cues (VA, VT, AT) are more preferred than other modes. While respective modes were chosen more often than VV and AV, somewhat unexpected, the AA mode was comparably popular. This is different from findings that revealed a high usability for combined AT cues compared to auditory cues alone [53]. To further enhance the effectiveness of the AA mode, it might be helpful to choose more distinguishable timbres for auditory proximity and transition cues [101]. Partly conform with **H2**, when visual and audio noise was present, the AT but not VT mode was preferred over all the other modes. Findings indicate that when choosing an appropriate sound as audio proximity cue it can still be perceived sufficiently well while visual noise clearly impairs the perception of visual cues. Since corresponding indications also occur in study 2 in this context, we will discuss this in the respective part in the discussion in more detail.

Surprisingly, although the perception of both cues is expected to be impaired by visual and audio noise for the mode VA, there was no significant difference in the preference score compared to VT. The audio cue could probably be perceived sufficiently well. In general, the combination of visual and audio cues has already been shown as beneficial before (e.g. [48, 74]) and has been considered to distribute content over visual and auditory modalities as vision can be highly loaded during certain (visual-only) based applications [101].

Modes which included the visual transition cue, namely VV and AV, were overall less preferred in the course of forced-choices than modes with other transition cues. This indicates that the single blink of the visual transition cue during its color change attracted less attention in relation to the auditory and tactile variants. The expected alerting effect of the color red [58] was also not that strong in comparison. This could be reasoned by the small size of the transition cue and the general poorer perception of colors in the periphery [41]. However, adjusting the characteristics of the transition cue, e.g. increasing its size, might increase the noticeability of peripheral augmentations [56]. Nevertheless, this would mean that also more display space would be occupied, aggravating clutter and occlusion issues. However, positive questionnaire ratings on visual transition cues under reduced noise (see Fig. 8) show the cue also performs well in an absolute subjective evaluation. Motion effects like blinking have been shown before as useful to attract attention in the peripheral area [56].

In regard to different cue combinations we did not include a TT mode (tactile proximity and transition cue) due to confusion issues and other ambiguities [97]. However, we assume that modes which involve tactile proximity would be received well for supporting spatial and temporal perception if the single cues are sufficiently distinguishable designed. The potential of tactile proximity cues was already demonstrated for single targets in AR in previous work, e.g. [50]. Such sensory substitution devices are commonly used e.g. for people with visual disabilities [66] and in terms of tactile cues considered to be a fruitful approach to improve user performance [69]. However, it remains difficult to render multiple targets at the same time with tactile proximity cues. Finally, continuous vibration on the head for a longer time is potentially uncomfortable for the user.

In contrast to existing techniques our feedback is composed of subsequent cues which both encode different parameters. Questionnaire results show that, **proximity and transition cue variants support the perception of target parameters important for spatial and temporal perception**: Under reduced noise, taken proximity and transition cues together, all variants facilitate the estimation of target direction and speed, as well as the perception of the exact time and location of the target transition. When the noise level increases, all but the visual cues can be further used to support the perception of target parameters.

7.2 Mode Performance

RQ$_2$: *Which cue combination mode performs best regarding target localization and the reaction to targets that enter the field of view when a concurrent visual task is performed? How does the level of visual and auditory noise affect performance in both tasks?*

Partly conform to our expectations referred to in **H$_3$** we found that noise and feedback mode affected the response time to the target enter-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

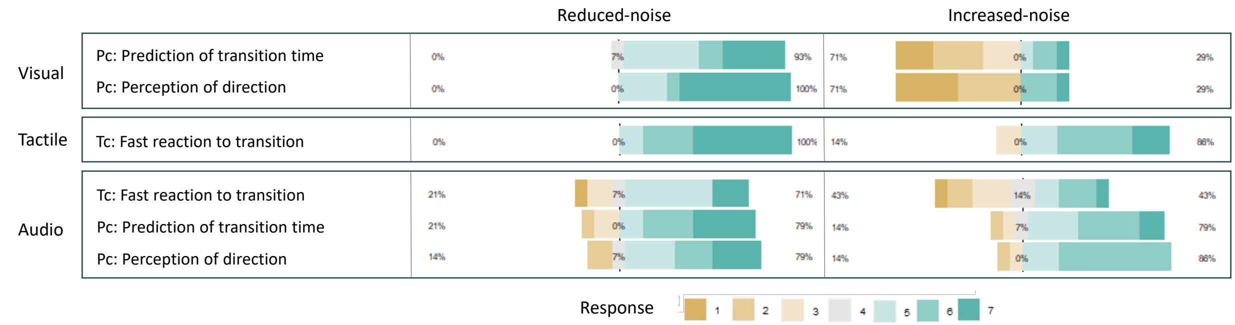


Fig. 10: Study2: Users indicated their level of agreement on 7-point Likert type items (ranging from 1= "strongly disagree" to 7= "strongly agree") on how the respective cue helped them to perceive temporal events and the spatial location (listed on the left) of an AR object under conditions with reduced noise (upper part) and with noise (lower part). Pc = Proximity cue, Tc = Transition cue.

ing the AR field of view in the **divided attention task**. As expected, the combined VA feedback yielded slower reaction times than the VT and AT mode in the reduced- and increased-noise condition. We further assumed that, when using VA and VT feedback modes, visual attention would be divided into the focal attention task and the awareness task which would lead to performance decrements in both tasks. We expected AT to be superior as limitations in visuospatial attention were shown to be circumvented by distributing attentional processing across sensory modalities in comparable dual task scenarios [114].

However, different from our expectations in $\mathbf{H}_3$, reaction time for VT and AT did not differ significantly in the awareness task. $\mathbf{H}_4$ was also not confirmed as in the focal attention task no differences between all three modes were found. Findings differ from [70], which showed that AT feedback facilitated a better visual concurrent task performance than visual feedback. Our findings indicate that sufficient visual attention capacities were available to perform both tasks well even when visual proximity feedback was provided. Subjective ratings on the usefulness of visual proximity cues further indicate that although users had to monitor the central FOV, they still made use of the more peripherally presented visual cues. This is possible as peripheral vision is still relatively good for motion detection [75] although the vision of text, shape and color degrade towards the periphery [41]. It should also be noted that proximity cues have a supportive role in the awareness task by spatially cueing attention [90] and preparing the user for the upcoming transition. Thereby, attention can also be oriented covertly, thus without eye or head movement (e.g., [6, 89]).

The superiority of VT and AT over VA can be attributed to the tactile transition cue indicating that **tactile feedback facilitates a faster response** than audio. Findings indicate that the seamless succession of proximity and transition cues (as is the case with our combined feedback) is likely to produce effects similar to prior investigations on multisensory stimulation. The results are in line with multiple studies that found an advantage of tactile over audio feedback in reaction time performance e.g., when reacting to stimulation [81] and in divided attention situations e.g., when indicating the direction to walk [96], or warning users [76, 79]. This underlines the noticeable yet unintrusive character of tactile feedback and the intuitive design of this particular tactile cue (compare to [21, 50, 69]).

Similar to [4], including tactile cues resulted in an improved performance compared to non-tactile cues with regards to reaction time. In addition, it can be assumed that the majority of tasks requires more visual and auditory than tactile processing capacities. Thus, in comparison, there are usually free resources available in the tactile modality. Tactile cues are also difficult to ignore and can be sensed regardless of the orientation of attention [5]. As pointed out in [101], audio-visual feedback is better suited for single tasks under normal workload conditions. It remains open, however, if and how variations of the visual focal attention task in terms of different perceptual and mental load levels or modalities would affect performance of feedback modes.

We further found a significant slowdown in reaction to targets in the awareness task as an **effect of environmental noise** (see Fig. 4) for modes VT and in particular VA. This can be explained by an increased amount of mental demand (reflected in NASA-TLX ratings) required to perceive and interpret visual and audio cues in the noise condition. Missing significant slowdowns for AT feedback indicate a stronger detrimental effect of visual than audio noise in AR on the perception of cues. As summarized in [57], lighting can lead to the incorrect display of color, incorrect augmentation and lens flare while very bright environments generally limit projection. The audio proximity cue can probably still be interpreted sufficiently well under noise due to the clearly perceivable sound signal [124]. Even though it requires more training than for visual cues to interpret it correctly [69]. However, the provision of audio cues can be inappropriate in situations where it is important to perceive environmental auditory input to avoid dangerous situations and degradation in performance [115]. Respective situations could occur especially with mobile users, for example, if the perception of other actors in road traffic or during various outdoor activities is required.

The slow down in reaction time under increased noise for the VT but not AT mode provides further evidence (in addition to subjective measures) that users make use of proximity cues instead of performing the awareness

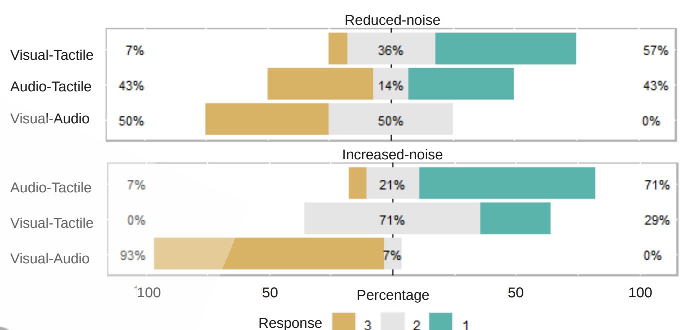


Fig. 11: Study 2, post-hoc questionnaire: Frequencies of the assignment of ranks (3=third, 2=second, 1=best) to a mode in the condition with reduced (upper part) and increased noise (lower part).

task by simply reacting to transition cues. Questionnaire ratings confirm the usefulness of proximity cues to predict the time of the later transition event. These findings underline that the combination of subsequent cues seems to have an added value to enhance spatial and temporal perception for fast responses under divided attention. Furthermore it was found that with reduced noise, less than 2% of targets were missed with VT and AT feedback and less than 10% with increased noise. On the other hand while with reduced noise also with VA few targets were missed (less than 4%), the share of missed targets increased to around 22% when environmental noise was present. Results provide further evidence that, in general, **multisensory signals can effectively capture spatial attention under conditions of concurrent perceptual load** [99, 102] **while tactile cues can be especially useful in noisy environments** [76, 79].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3116673, IEEE Transactions on Visualization and Computer Graphics

Finally we found that performance was not affected by the direction from which the target approached in spite of differences in color and optical flow between left and right display area (see Fig. 4a, 4b). It is conceivable that top-down mechanisms of attention filtered information processing by target facilitation and/or by distractor (background colors and optical flow) inhibition [83].

Overall, study results demonstrate the positive reception and usefulness of particular combined feedback modes for spatial and temporal perception and fast reactions to incoming targets. However, there are some limitations which have to be considered. First, in this work we initially focused on two possible target trajectories. So, it is still unclear how different modes would perform if targets could follow any trajectory especially in terms of accuracy. Prior work provides evidence for a higher localization accuracy for moving audio sources than for visual spatial information that has to be mapped to the real world space (compare [32] to [35]). This may lead to the assumption that the AT mode might also be superior to VA and VT due to the superiority of audio to visual proximity feedback. However, the comparability of modalities is limited across these studies due to differences in measures and procedures. Furthermore, other studies point out some potential issues regarding HRTF audio localization. This includes the mislocalization to the incorrect front/back hemifield [120], an increasing azimuth and elevation localization error behind the head [85] and a high listener specificity [118]. In addition, for combined proximity and transition feedback cross-modal effects could also affect localization performance. In the context of localization, it is also important to note that we only examined the reaction to one moving target at once. It is questionable to what extent simultaneously presented audio or vibration signals can be kept apart, so that a sequential presentation of such proximity cues would be conceivable in multi-target scenarios. However, visual multi-target feedback is likely not as intrusive as the administration of multiple auditory or tactile cues and has already been implemented for 3D spaces (see Table 1 in [35]). Nevertheless, results from studies on multiple object tracking in the visual modality show that if the number of moving targets increases, localization performance systematically decreases and attentional load increases with each additional target [113]. In addition, multisensory research indicates that limitations can likely not be circumvented by distributing attentional processing across sensory modalities for multiple object tracking as only spatial (as opposed to spatial and object-based) attentional processing is required [114]. In this context, the additional transition cue could prove to be very useful as it would allow the user to prioritize by signalizing that specific targets become highly relevant only at a certain point of time. In some use cases, the need to permanently track each target object could be eliminated in this manner. However, proximity feedback on multiple objects could still give the user a general understanding of the structure and dynamics of the environment.

Another aspect which has to be considered is that we always used a constant speed level, so it is still unclear if results could be different for other constant or dynamically changing speed levels. It is conceivable that localization and also reaction to incoming targets would be impaired due to the increasing unpredictability of prospective target positions. A further issue is the abstraction of the target and tasks in our studies for generalization purposes. In practical use, users would typically look for a text label or icon tied to a point of interest. In such a scenario there would likely be many other (non-target) text labels or icons that could act as additional distractors and potentially influence the cue interpretation. Also, in the dual-task scenario in this study, due to the study design, we did not allow the target to become visible when it entered the FOV. Thus, switching from visual in-view augmentations to out-of-view encoded information was only represented in the preference study. Potential confounding influences are still unclear in this context in terms of reaction and localization performance. Another limitation is that we simulated an AR scenario OST HMDs in VR. Although results can likely be applied to other display types (see subsection 7.3.2), the findings of our studies currently refer only to narrow FOV OST devices. Furthermore it also remains to be investigated how the subsequent proximity and transition cue administration compares to other techniques and under different contextual conditions. For example, in several use cases speed might not be that relevant, e.g. during the performance in high precision tasks. Which modes perform best on other dependent measures under different task conditions is still open. Finally, it must be noted that the majority of participants were male and that results might be different with a higher ratio of female users.

7.3 Reflection towards Field of View

7.3.1 Narrow Field of View

Throughout our studies we demonstrate that proximity and transition cues enhance spatial awareness of moving out-of-view target objects. The proximity cue can guide visual search for an out-of-view target to counteract set size effects on search time which are amplified when the FOV is narrowed [14]. Particularly large benefits are expected in search processes where the rejection of distractors is generally time consuming (resulting in strong set size effects), for example with textual labels, as these have to be focused on individually in order to grasp the meaning [14]. In addition, guided search can also counteract varying and unstructured visual search patterns that occur when the field of view narrows [14]. Visual and audio proximity feedback further enable the maintenance of spatial awareness of a target out of view even if it is not searched for actively. EdgeRadar has been proved effective for tracking moving out-of-view objects [36] before, while 3D auditory cues have been shown to be intuitive and easy to use [67].

Regarding transition cues, we showed that the audio and especially the tactile variant can alert the user and yield miss rates below 2%. This is particularly useful in narrow FOV AR where information overload and occlusion are particularly prevalent due to limited space and favor overlooking in-view targets. In this context the proximity cue takes a supportive role by enabling the user to predict the time of transition through information on target distance, direction and speed.

In addition, the interpretation of our combined multisensory feedback leaves sufficient resources free to engage in other activities. Since our cue design keeps the central display free, the already limited space in narrow FOV AR can be further utilized for other (interactive) content. Prior work confirms that multisensory stimulation can reduce cognitive load [86, 100] and capture spatial attention under dual task conditions [99, 102].

7.3.2 Wider Field of View

Although the FOV in AR devices will likely become wider in the future, it remains a challenging goal building displays that are able to fully cover the human visual field [51]. Therefore, typical problems of narrow FOVs associated with visual methods (e.g. cluttering, occlusion, potentially higher workload) are expected to remain for wider FOV devices (e.g. HoloLens 2 or future ultra-wide FOV devices - see [52]). In addition, with an extremely wide FOV in-view information becomes much less noticeable towards the periphery [52]. As a consequence, noticing an in-view target can still be difficult and its localization time-consuming, especially in dense AR environments.

To address the problem, the combined feedback can be adjusted to wide FOV displays to provide in-view guidance. For this, proximity cues can continue to be administered when the target enters the FOV until it falls into the focal field of vision. This design would open up new opportunities to incorporate other successful non-visual guidance methods, for example, combined audio-tactile guidance cues from [69]. However, the continuous administration of proximity feedback carries the risk of overstimulation. In addition, when the FOV is enlarged the length of its edges also increases which can have the consequence that transition cues are also frequently activated. Especially when multiple target objects are involved the user can rapidly be overstimulated ([26, 63]). To limit cue activation, the amount of information can be reduced by filtering and clustering techniques (e.g. by content, distance, amount, etc.) [104]. Furthermore, the user should be given the option to manually turn feedback on and off, as well as configure cue parameters to suit individual needs.

When visual cues are provided at the display borders (as in EdgeRadar and other contextual cueing techniques), further adaptations will likely be necessary for the application in wide FOV displays.

As the FOV is increased, the area in which visual cues are presented moves further away from the center to far peripheral regions. In a task scenario in which it is important to keep attention to the central field of vision (compare to [52]) the user would be forced to frequently switch spatial attention between the center and the far periphery. As an alternative to a redefinition of the area where cues are presented in EdgeRadar, it would also be possible to use a more appropriate guidance technique, for example an adapted version of 3D Radar [9]. This method is useful for tracking moving out-of-view objects while avoiding clutter in the foveated vision of the user [35].

8 Conclusion and Future Work

Throughout two user studies we evaluated novel feedback techniques combining Visual-Audio, Visual-Tactile and Audio-Tactile proximity-transition feedback that were pre-selected based on user preferences. We found that visual proximity cues were clearly more affected by noise than the audio variant. Regarding transition cues, the tactile variant facilitated the fastest reaction to incoming out-of-view AR objects in the course of a divided attention task. The concurrent visual task performance showed no clear advantage of a specific mode.

8.1 Future work

Based on the results and observations to date, the system is to be expanded and further optimized by focusing on specific key issues in the design and subsequent evaluation phase.

In the future development process, we target the encoding of all possible moving trajectories and multiple target objects in the visual, auditory and tactile modality. In this context, limitations in the perception of multiple visual [37] and auditory [126] targets will be considered. To encode target location in the three dimensional space in the tactile modality parameters vibration location, duration, frequency, amplitude and several other (derived) can be used [110]. We aim for a solution that occupies areas of high sensitivity and minimizes the amount of vibration motors required, thus avoiding higher resolution grids (as in [50]). In this manner, we will circumvent perceptual limitations at less sensitive regions of the head [22] and areas with high hair density [80]. The targeted setup is also expected to be easier to maintain and more suitable for everyday use in terms of long-time user comfort and social acceptability compared to an EEG-like solution [21, 50].

In the subsequent evaluation of feedback combinations in the renewed system we will include eye-tracking to enhance the effectiveness of both visual and non-visual guidance metaphors [7], and to use it as an additional indicator for situation awareness [108]. We further intend to vary noise conditions by focusing on uncontrolled outdoor usage of the system to address more dynamically changing conditions. Doing so, we also will address effects of long term usage of the system, to pinpoint when and perhaps when not to provide proximity and transition cues, as we expect that continuous feedback will be burdening over time. Within this context, real-world search tasks instead of abstract attention tasks will be considered. The comparison of individual cues to combined feedback under conditions of varying densities of distractor AR objects represent further interesting extensions to previous work. So, the added value of combined feedback and usefulness in dense information spaces could be objectively quantified.

Acknowledgments

This work was partially funded by Deutsche Forschungsgemeinschaft (DFG) under grant KR 4521/2-1.

References

[1] T. T. Ahmaniemi and V. T. Lantz. Augmented reality target finding based on tactile cues. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pp. 335–342, 2009.

[2] P. Alfano and G. Michel. Restricting the field of view: Perceptual and performance effects. *Perceptual and motor skills*, 70(1):35–45, 1990.

[3] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 327–334. IEEE, 2018.

[4] A. Bajpai, J. C. Powell, A. J. Young, and A. Mazumdar. Enhancing physical human evasion of moving threats using tactile cues. *IEEE Transactions on Haptics*, 13(1):32–37, 2019.

[5] C. L. Baldwin, C. Spence, J. P. Bliss, J. C. Brill, M. S. Wogalter, C. B. Mayhorn, and T. K. Ferris. Multimodal cueing: The relative benefits of the auditory, visual, and tactile channels in complex environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, pp. 1431–1435. SAGE Publications Sage CA: Los Angeles, CA, 2012.

[6] J. Beck and B. Ambler. The effects of concentrated and distributed attention on peripheral acuity. *Perception & Psychophysics*, 14(2):225–230, 1973.

[7] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, pp. 1–9, 2018.

[8] J. R. Blum, M. Bouchard, and J. R. Cooperstock. What's around me? spatialized audio augmented reality for blind users with a smartphone. In A. Puiatti and T. Gu, eds., *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pp. 49–62. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[9] F. Bork, C. Schnelzer, U. Eck, and N. Navab. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2983–2992, 2018.

[10] B. Breitmeyer, H. Ogmen, H. Öğmen, et al. *Visual masking: Time slices through conscious and unconscious vision*. Oxford University Press, 2006.

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

[12] S. Burigat, L. Chittaro, and A. Vianello. Dynamic visualization of large numbers of off-screen objects on mobile devices: an experimental comparison of wedge and overview+ detail. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pp. 93–102, 2012.

[13] J. L. Burke, M. S. Prewett, A. A. Gray, L. Yang, F. R. Stilson, M. D. Coovert, L. R. Elliot, and E. Redden. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis. In *Proceedings of the 8th international conference on Multimodal interfaces*, pp. 108–117, 2006.

[14] C. Trepkowski, and T. Eibich and J. Maiero and A. Marquardt and E. Kruijff and S. Feiner. The effect of narrow field of view and information density on visual search performance in augmented reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019.

[15] M. Carrasco, B. McElree, K. Denisova, and A. M. Giordano. Speed of visual processing increases with eccentricity. *Nature neuroscience*, 6(7):699–700, 2003.

[16] A. Case and A. Day. *Designing with Sound: Fundamentals for Products and Services*. O'Reilly Media, 2018.

[17] D. D. Chiras. *Environmental science: Creating a sustainable future*. Jones & Bartlett Learning, 2004.

[18] M. Corporation. Hololens 2 - overview, features, and specs — microsoft hololens. `https://www.microsoft.com/en-us/hololens/hardware`.

[19] J. Covelli, J. Rolland, M. Proctor, J. Kincaid, and P. Hancock. Field of view effects on pilot performance in flight. *The International Journal of Aviation Psychology*, 20(2):197–219, 2010.

[20] M. Crede, M. Bashshur, and S. Niehorster. Reference group effects in the measurement of personality and attitudes. *Journal of Personality Assessment*, 92(5):390–399, 2010.

[21] V. A. de Jesus Oliveira, L. Brayda, L. Nedel, and A. Maciel. Designing a vibrotactile head-mounted display for spatial awareness in 3d spaces. *IEEE transactions on visualization and computer graphics*, 23(4):1409–1417, 2017.

[22] V. A. de Jesus Oliveira, L. Nedel, A. Maciel, and L. Brayda. Spatial discrimination of vibrotactile stimuli around the head. In *2016 IEEE Haptics Symposium (HAPTICS)*, pp. 1–6. IEEE, 2016.

[23] F. C. Donders. On the speed of mental processes. *Acta psychologica*, 30:412–431, 1969.

[24] J. Driver and C. Spence. Crossmodal spatial attention: Evidence from human performance. *Crossmodal space and crossmodal attention*, pp. 179–220, 2004.

[25] G. Dubus and R. Bresin. A systematic review of mapping strategies for the sonification of physical quantities. *PloS one*, 8(12):e82491, 2013.

[26] P. E. Dux, J. Ivanoff, C. L. Asplund, and R. Marois. Isolation of a central bottleneck of information processing with time-resolved fmri. *Neuron*, 52(6):1109–1120, 2006.

[27] S. Ellis and B. Menges. Localization of virtual objects in the near visual field. *Human Factors*, 40(3):415–431, 1998.

[28] M. R. Endsley. Measurement of situation awareness in dynamic systems. *Human factors*, 37(1):65–84, 1995.

[29] T. Foulsham, A. Kingstone, and G. Underwood. Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision research*, 48(17):1777–1790, 2008.

[30] J. L. Gabbard, J. E. Swan, and D. Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence: Teleoperators & Virtual Environments*, 15(1):16–32, 2006.

[31] R. A. Grier. How high is high? a meta-analysis of nasa-tlx global workload scores. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59, pp. 1727–1731. SAGE Publications Sage CA: Los Angeles, CA, 2015.

[32] M. Grohn, T. Lokki, and T. Takala. Static and dynamic sound source localization in a virtual room. In *Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio*. Audio Engineering Society, 2002.

[33] U. Gruenefeld, A. El Ali, S. Boll, and W. Heuten. Beyond halo and wedge: Visualizing out-of-view objects on head-mounted virtual and augmented reality devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 09 2018.

[34] U. Gruenefeld, D. Ennenga, A. E. Ali, W. Heuten, and S. Boll. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 109–118, 2017.

[35] U. Gruenefeld, I. Koethe, D. Lange, S. WeirB, and W. Heuten. Comparing techniques for visualizing moving out-of-view objects in head-mounted virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 742–746. IEEE, 2019.

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

[37] S. G. Gustafson and P. P. Irani. Comparing visualizations for tracking off-screen moving targets. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems*, pp. 2399–2404, 2007.

[38] E. C. Haas and J. B. Van Erp. Multimodal warnings to enhance risk communication and safety. *Safety science*, 61:29–35, 2014.

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

[40] S. Hidaka and M. Ide. Sound can suppress visual perception. *Scientific reports*, 5(1):1–9, 2015.

[41] J. Hirsch and C. A. Curcio. The spatial resolution capacity of human foveal retina. *Vision research*, 29(9):1095–1101, 1989.

[42] K. Hopkins, S. J. Kass, L. D. Blalock, and J. C. Brill. Effectiveness of auditory and tactile crossmodal cues in a dual-task visual and auditory scenario. *Ergonomics*, 60(5):692–700, 2017.

[43] S. Hwang, H. Jo, and J.-h. Ryu. Exmar: Expanded view of mobile augmented reality. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 235–236. IEEE, 2010.

[44] C. Jay and R. Hubbold. Amplifying head movements with head-mounted displays. *Presence: Teleoperators & Virtual Environments*, 12(3):268–276, 2003.

[45] D. Jia, A. Bhatti, and S. Nahavandi. User-centered design and evaluation of an interactive visual-haptic-auditory interface: a user study on assembly. In *World Conference on Innovative Virtual Reality*, vol. 44328, pp. 263–272, 2011.

[46] H. Jo, S. Hwang, H. Park, and J.-h. Ryu. Aroundplot: Focus+ context interface for off-screen objects in 3d environments. *Computers & Graphics*, 35(4):841–853, 2011.

[47] J. W. Kalat. *Biological psychology*. Nelson Education, 2015.

[48] B. Katz, S. Kammoun, G. Parseihian, O. Gutierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe, and C. Jouffrais. Navig: Augmented reality guidance system for the visually impaired: Combining object localization, gnss, and spatial audio. *Virtual Reality*, 16:17, 01 2012.

[49] B. Katz, E. Rio, L. Picinali, and O. Warusfel. The effect of spatialization in a data sonification exploration task. *14th Meeting of the Intl. Conf. on Auditory Display (ICAD)*, pp. 1–7, 01 2008.

[50] O. B. Kaul and M. Rohs. Haptichead: 3d guidance and target acquisition through a vibrotactile grid. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 2533–2539, 2016.

[51] K. Kim, M. Billinghurst, G. Bruder, H. Duh, and G. Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). *IEEE Transactions on Visualization and Computer Graphics*, PP:1–1, 09 2018.

[52] N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff. Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In *Mixed and Augmented Reality (ISMAR), 2014 IEEE International Symposium on*, pp. 177–186. IEEE, 2014.

[53] S. E. Knobel, N. T. Gyger, T. Nyffeler, D. Cazzoli, R. M. Müri, and T. Nef. Development and evaluation of a new virtual reality-based audio-tactile cueing-system to guide visuo-spatial attention. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3192–3195. IEEE, 2020.

[54] T. Koelewijn, A. Bronkhorst, and J. Theeuwes. Competition between auditory and visual spatial cues during visual task performance. *Experimental Brain Research*, 195(4):593–602, Jun 2009.

[55] S. U. König, F. Schumann, J. Keyser, C. Goeke, C. Krause, S. Wache, A. Lytochkin, M. Ebert, V. Brunsch, B. Wahn, et al. Learning new sensorimotor contingencies: Effects of long-term use of sensory augmentation on the brain and conscious perception. *PloS one*, 11(12):e0166647, 2016.

[56] E. Kruijff, J. Orlosky, N. Kishishita, C. Trepkowski, and K. Kiyokawa. The influence of label design on search performance and noticeability in wide field of view augmented reality displays. *IEEE transactions on visualization and computer graphics*, 25(9):2821–2837, 2018.

[57] E. Kruijff, J. E. Swan, and S. Feiner. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 3–12. IEEE, 2010.

[58] M. Kuniecki, J. Pilarczyk, and S. Wichary. The color red attracts attention in an emotional context. an erp study. *Frontiers in human neuroscience*, 9:212, 2015.

[59] M. Lee, M. Billinghurst, W. Baek, R. Green, and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Reality*, 17(4):293–305, 2013.

[60] V. Lehtinen, A. Oulasvirta, A. Salovaara, and P. Nurmi. Dynamic tactile guidance for visual search tasks. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pp. 445–452, 2012.

[61] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi. Superpixel-based fast fuzzy c-means clustering for color image segmentation. *IEEE Transactions on Fuzzy Systems*, 27(9):1753–1766, 2018.

[62] R. Lindeman, R. Page, Y. Yanagida, and J. Sibert. Towards full-body haptic feedback: the design and deployment of a spatialized vibrotactile feedback system. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST 2004*, pp. 146–149, 2004.

[63] Z. Lipowski. Sensory and information inputs overload: Behavioral effects. *Comprehensive Psychiatry*, 16(3):199 – 221, 1975.

[64] V. Losing, L. Rottkamp, M. Zeunert, and T. Pfeiffer. Guiding visual search tasks using gaze-contingent auditory feedback. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pp. 1093–1102, 2014.

[65] A. I. magazine. Evaluating immersive experiences during covid-19 and beyond.

[66] S. Maidenbaum, S. Abboud, and A. Amedi. Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews*, 41, 01 2013.

[67] A. Marquardt, E. Kruijff, C. Trepkowski, J. Maiero, A. Schwandt, A. Hinkenjann, W. Stuerzlinger, and J. Schoening. Audio-tactile feedback

for enhancing 3d manipulation. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '18. ACM, 2018.

[68] A. Marquardt, E. Kruijff, C. Trepkowski, J. Maiero, A. Schwandt, A. Hinkenjann, W. Stuerzlinger, and J. Schöning. Audio-tactile proximity feedback for enhancing 3d manipulation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10, 2018.

[69] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, and E. Kruijff. Non-visual cues for view management in narrow field of view augmented reality displays. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 190–201. IEEE, 2019.

[70] A. Marquardt, C. Trepkowski, T. D. Eibich, J. Maiero, E. Kruijff, and J. Schöning. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 2020.

[71] J. J. McDonald, J. J. Green, V. S. Störmer, and S. A. Hillyard. Cross-modal spatial cueing of attention influences visual perception. In *The neural bases of multisensory processes*. CRC Press/Taylor & Francis, 2012.

[72] J. J. McDonald, W. A. Teder-SaÈlejaÈrvi, and S. A. Hillyard. Involuntary orienting to sound improves visual perception. *Nature*, 407(6806):906–908, 2000.

[73] J. McIntire, P. Havig, S. Watamaniuk, and R. Gilkey. Visual search performance with 3-d auditory cues: Effects of motion, target location, and practice. *Human Factors*, 52(1):41–53, 2010.

[74] J. P. McIntire, P. R. Havig, S. N. Watamaniuk, and R. H. Gilkey. Visual search performance with 3-d auditory cues: Effects of motion, target location, and practice. *Human factors*, 52(1):41–53, 2010.

[75] S. P. McKee and K. Nakayama. The detection of motion in the peripheral visual field. *Vision research*, 24(1):25–32, 1984.

[76] R. Mohebbi, R. Gray, and H. Z. Tan. Driver reaction time to tactile and auditory rear-end collision warnings while talking on a cell phone. *Human Factors*, 51(1):102–110, 2009.

[77] J. A. Mossbridge, M. Grabowecky, and S. Suzuki. Changes in auditory frequency guide visual–spatial attention. *Cognition*, 121(1):133–139, 2011.

[78] S. Mudd and E. McCormick. The use of auditory cues in a visual search task. *Journal of Applied Psychology*, 44(3):184, 1960.

[79] A. Murata, T. Kuroda, and W. Karwowski. Effects of auditory and tactile warning on response to visual hazards under a noisy environment. *Applied Ergonomics*, 60:58–67, 2017.

[80] K. Myles, J. Kalb, J. Lowery, and B. P. Kattel. The effect of hair density on the coupling between the tactor and the skin of the human head. *Applied Ergonomics*, 48:177–185, 2015.

[81] A. W. Ng and A. H. Chan. Finger response times to visual, auditory and tactile modality stimuli. In *Proceedings of the international multi-conference of engineers and computer scientists*, vol. 2, pp. 1449–1454, 2012.

[82] M. K. Ngo and C. Spence. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*, 72(6):1654–1665, Aug 2010.

[83] M. P. Noonan, N. Adamian, A. Pike, F. Printzlau, B. M. Crittenden, and M. G. Stokes. Distinct mechanisms for distractor suppression and target facilitation. *Journal of Neuroscience*, 36(6):1797–1807, 2016.

[84] N. O. A. Observatory. Recommended light levels (illuminance) for outdoor and indoor venues. `https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf`.

[85] S. R. Oldfield and S. P. Parker. Acuity of sound localisation: a topography of auditory space. i. normal hearing conditions. *Perception*, 13(5):581–600, 1984.

[86] S. Oviatt, R. Coulston, and R. Lunsford. When do we interact multimodally? cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces*, pp. 129–136, 2004.

[87] G. Parseihian, C. Gondre, M. Aramaki, S. Ystad, and R. Kronland-Martinet. Comparison and evaluation of sonification strategies for guidance tasks. *IEEE Transactions on Multimedia*, 18(4):674–686, April 2016.

[88] R. D. Patterson. *Guidelines for auditory warning systems on civil aircraft*. Civil Aviation Authority, 1982.

[89] M. I. Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980.

[90] M. I. Posner, C. R. Snyder, and B. J. Davidson. Attention and the detection of signals. *Journal of experimental psychology: General*, 109(2):160, 1980.

[91] M. S. Prewett, L. R. Elliott, A. G. Walvoord, and M. D. Coovert. A meta-analysis of vibrotactile and visual information displays for improving task performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(1):123–132, 2011.

[92] L. Qian, A. Plopski, N. Navab, and P. Kazanzides. Restoring the awareness in the occluded visual field for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2936–2946, 2018.

[93] E. Ragan, C. Wilkes, D. A. Bowman, and T. Hollerer. Simulation of augmented reality systems in purely virtual environments. In *2009 IEEE Virtual Reality Conference*, pp. 287–288, 2009.

[94] F. Ribeiro, D. Florêncio, P. A. Chou, and Z. Zhang. Auditory augmented reality: Object sonification for the visually impaired. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, pp. 319–324, Sep. 2012.

[95] H. Roodaki, N. Navab, A. Eslami, C. Stapleton, and N. Navab. Sonifeye: Sonification of visual information using physical modeling sound synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2366–2371, Nov 2017.

[96] D. A. Ross and B. B. Blasch. Wearable interfaces for orientation and wayfinding. In *Proceedings of the fourth international ACM conference on Assistive technologies*, pp. 193–200, 2000.

[97] B. Saket, C. Prasojo, Y. Huang, and S. Zhao. Designing an effective vibration-based notification interface for mobile phones. In *Proceedings of the 2013 conference on Computer supported cooperative work*. Association for Computing Machinery, New York, NY, USA, 2013.

[98] C. Sandor, A. Dey, A. Cunningham, S. Barbier, U. Eck, D. Urquhart, M. R. Marner, G. Jarvis, and S. Rhee. Egocentric space-distorting visualizations for rapid environment exploration in mobile mixed reality. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 47–50. IEEE, 2010.

[99] V. Santangelo and C. Spence. Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6):1311, 2007.

[100] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen. Galvanic skin response (gsr) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems*, pp. 2651–2656, 2007.

[101] R. Sigrist, G. Rauter, R. Riener, and P. Wolf. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review*, 20(1):21–53, 2013.

[102] C. Spence and V. Santangelo. Capturing spatial attention with multisensory cues: A review. *Hearing research*, 258(1-2):134–142, 2009.

[103] B. E. Stein, N. London, L. K. Wilkinson, and D. D. Price. Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of cognitive neuroscience*, 8(6):497–506, 1996.

[104] M. Tatzgern, V. Orso, D. Kalkofen, G. Jacucci, L. Gamberini, and D. Schmalstieg. Adaptive information density for augmented reality displays. In *2016 IEEE Virtual Reality (VR)*, pp. 83–92. IEEE, 2016.

[105] A. Treisman. Preattentive processing in vision. *Computer vision, graphics, and image processing*, 31(2):156–177, 1985.

[106] A. Treisman. Features and objects in visual processing. *Scientific American*, 255(5):114B–125, 1986.

[107] H. Uchiyama, M. Covington, and W. Potter. Vibrotactile glove guidance for semi-autonomous wheelchair operations. In *Proceedings of the 46th Annual Southeast Regional Conference on XX*, pp. 336–339. ACM, 2008.

[108] K. van de Merwe, H. Dijk, and R. Zon. Eye movements as an indicator of situation awareness in a flight simulator experiment. *The International Journal of Aviation Psychology*, 22:78–95, 01 2012.

[109] J. B. Van Erp. Guidelines for the use of vibro-tactile displays in human computer interaction. In *Proceedings of eurohaptics*, vol. 2002, pp. 18–22. Citeseer, 2002.

[110] J. B. Van Erp and H. Van Veen. Vibro-tactile information presentation in automobiles. In *Proceedings of eurohaptics*, vol. 2001, pp. 99–104. Eurohaptics Society Paris, France, 2001.

[111] R. T. Verrillo, G. A. Gescheider, B. G. Calman, and C. L. Van Doren. Vibrotactile masking: Effects of oneand two-site stimulation. *Perception & Psychophysics*, 33(4):379–387, 1983.

[112] W3C. Web content accessibility guidelines (wcag) 2.1. `https://www.w3.org/TR/WCAG21/`.

[113] B. Wahn, D. P. Ferris, W. D. Hairston, and P. König. Pupil sizes scale with attentional load and task experience in a multiple object tracking

task. *PloS one*, 11(12):e0168087, 2016.

[114] B. Wahn and P. König. Can limitations of visuospatial attention be circumvented? a review. *Frontiers in Psychology*, 8:1896, 2017.

[115] S. Wall and S. Brewster. Feeling what you hear: tactile feedback for navigation of audio graphs. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1123–1132, 2006.

[116] R. Wegel and C. Lane. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physical review*, 23(2):266, 1924.

[117] M. Wells, M. Venturino, and R. Osgood. The effect of field-of-view size on performance at a simple simulated air-to-air mission. In *Helmet-Mounted Displays*, vol. 1116, pp. 126–138. International Society for Optics and Photonics, 1989.

[118] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, 1993.

[119] E. Wetzel, J. R. Böhnke, and A. Brown. Response biases. 2016.

[120] F. L. Wightman and D. J. Kistler. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105(5):2841–2853, 1999.

[121] A. R. Wildt and M. B. Mazis. Determinants of scale response: Label versus position. *Journal of Marketing Research*, 15(2):261–267, 1978.

[122] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146, 2011.

[123] J. Wolfe. Guidance of visual search by preattentive information. In L. Itti, G. Rees, and J. Tsotsos, eds., *Neurobiology of Attention*, pp. 101–104. Elsevier, 12 2005.

[124] W. Yost and X. Zhong. Sound source localization identification accuracy: Bandwidth dependencies. *The Journal of the Acoustical Society of America*, 136:2737, 11 2014.

[125] P. T. Zellweger, J. D. Mackinlay, L. Good, M. Stefik, and P. Baudisch. City lights: contextual views in minimal space. In *CHI'03 extended abstracts on Human factors in computing systems*, pp. 838–839, 2003.

[126] X. Zhong and W. A. Yost. How many images are in an auditory scene? *The Journal of the Acoustical Society of America*, 141(4):2882–2892, 2017.

Christina Trepkowski received her Master of Science in Psychology from the Heinrich-Heine-University of Düsseldorf, Germany in 2017. She is currently doing her Ph.D. on the influence of multisensory cues on situation awareness in Augmented Reality at the Bonn-Rhein-Sieg University of Applied Sciences and University of Osnabrück.

Alexander Marquardt obtained his M.Sc in Computer Science at the Bonn-Rhein-Sieg University of Applied Sciences (BRSU). As a Ph.D. student at University of Bremen and BRSU, his research interests are focused on the design and development of experimental multisensory user interfaces for narrow field of view head-worn devices.

Tom David Eibich received his Bachelor in Computer Science from the Bonn-Rhein-Sieg University in 2018. He is currently doing the Master's program and his research interests are augmented reality and interactive environments.

Yusuke Shikanai received his M.E. from Nara Institute of Science and Technology (NAIST) in 2021. He has been working as a software engineer at DeNA Corporation since then. His research interests include augmented reality systems for inducing the Proteus effect in the real environment.

Jens Maiero received the Diploma in mathematics from the Stuttgart Technology University of Applied Sciences, Stuttgart, Germany, in 2006, the M.Sc. degree in Computer Science from the Bonn-Rhein-Sieg University of Applied Sciences (BRSU), Sankt Augustin, Germany, in 2009, and is currently finalizing his Ph.D. degree in Computer Science at Brunel University, London, U.K., and BRSU.

Kiyoshi Kiyokawa is a full professor at Graduate School of Information Science, Nara Institute of Science and Technology from 2017. His research interests include virtual reality, augmented reality, human augmentation, 3D user interface, computer-supported cooperative work, and context awareness.

Ernst Kruijff is professor for human computer interaction at the Institute of Visual Computing, Bonn-Rhein-Sieg University of Applied Sciences. He is also adjunct professor at SFU-SIAT in Canada. For over two decades, his research has focused at the human-factors driven analysis, design and validation of multisensory 3D user interfaces.

Johannes Schöning is a Lichtenberg Professor and Professor of Human-Computer Interaction at the University of Bremen in Germany. He investigates how people interact with digital spatial information and creates new methods and novel interfaces to help people to do so. His research interests lies at the intersection between HCI, geographic information science and ubiquitous interface technologies.

Peter König is professor for NeuroBioPsychologie at the Institute of Cognitive Science, University Osnabrück, and a Guest Professor at the University Hospital Hamburg Eppendorf. His research investigates on the physiological basis of cognitive functions from the perspective of Embodied Cognition.