Non-Visual Cues for View Management in Narrow Field of View Augmented Reality Displays

Alexander Marquardt, Christina Trepkowski, Tom David Eibich, Jens Maiero, and Ernst Kruijff

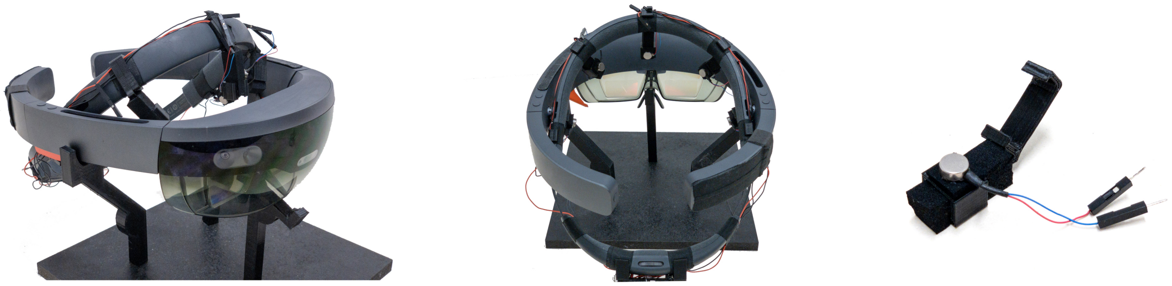


Fig. 1: The novel head-worn vibration feedback mechanism attached to the Microsoft HoloLens. In our studies, we used five vibration motors touching the forehead and temples. Three further vibration elements can be attached to the back of the head. The left and middle image show the arrangement of the used vibration elements, the right image a close up of the mount that was optimized for vibration feedback through a custom-build flexible mechanism that comfortably presses the vibration motor to the head for optimal skin contact.

Abstract—Head-worn devices with a narrow field of view are common commodity for Augmented Reality. However, their limited screen space makes view management difficult. Especially in dense information spaces this potentially leads to visual conflicts such as overlapping labels (occlusion) and visual clutter. In this paper, we look into the potential of using audio and vibrotactile feedback to guide search and information localization. Our results indicate users can be guided with high accuracy using audio-tactile feedback with maximum median deviations of only 2° on longitude, 3.6° on latitude and 0.07 meter in depth. Regarding the encoding of latitude we found a superior performance when using audio, resulting in an improvement of 61% and fastest search times. When interpreting localization cues the maximum median deviation was 9.9° on longitude and 18% of a selected distance to be encoded which could be reduced to 14% when using audio.

Index Terms—Augmented Reality, audio-tactile feedback, guidance, depth perception

1 INTRODUCTION

When increasing the density of information displayed in Augmented Reality (AR) applications, view management – the presentation and layout of augmentations – becomes challenging [80]. Conflicting visual cues and high density information can eventually lead to various degrees of sensory overload [51]. While visual attention is not necessarily affected by the number of distractors in visual search [86], the abundance of labels with potential visual conflicting cues can be difficult to process [43]. As human processing capacities are limited, once this capacity (or tolerance level) is exceeded by the stimulus input, overload occurs: a person will not be able to cope with all information within a fixed period of time, thus affecting user performance [51]. In AR, this predominantly occurs in the visual sensory channel, due to the prevalent nature of most view management systems being visual-only. View management becomes increasingly difficult when information needs to be compressed inside a narrow FOV, resulting in a highly dense and potentially confusing view on the information space. AR see-through head mounted displays typically provide a horizontal FOV of about 20-60 degrees [14], whereas the current Microsoft HoloLens offers a horizontal FOV of about 30 degrees. Once human capacities are reached, it may cause behavioral changes that may depend on individual differences, and the nature of the stimuli itself, including level, diversity, patterning, instability and meaningfulness [85]. Generally, human reactions such as performance fluctuations or frustration are preceded by increasing cognitive load [51].

To exemplify some of the challenges for narrow FOV displays, consider a domain that can exhibit dense information, namely location based services (LBS). These systems are used frequently, for example for interactive city guides. While view management systems have existed for a long time [21], they still have limitations. Though view management systems are improving, most systems likely will produce visual clutter when information density is increasing in LBS. The usage of in-view labelling [44] can further exacerbate this problem as the view management system may try to place additional labels inside the limited FOV that refer to objects outside the FOV. A typical problem is overlapping labels, where labels occlude each other and potentially the reference object in the scene [20]. This may cause visual conflicts such as those related to visibility, legibility, depth ordering, scene distortion and object relationship issues [43]. For example, consider finding a particular restaurant among many others in a downtown area. Here, labels will refer to objects in the scene at different distances. This abundance of cues can be difficult to entangle, as labels likely are cluttered due to limited screen space, and may overlap. Yet, users will still need to process all cues until the searched restaurant, or an alternative in its surroundings, is found.

An approach to reduce overload (and conflicts) is to minimize the number of stimuli in one sensory channel. This can be achieved by

---

All authors are with Bonn-Rhein-Sieg University of Applied Sciences. E-mail: {alexander.marquardt, christina.trepkowski, tom.eibich, jens.maiero, ernst.kruijff}@h-brs.de.

Ernst Kruijff is also with Simon Fraser University.

Manuscript received 22.03.2019; accepted 28.05.2019. Date of Publication 14.10.2019; date of current version 02.07.2019. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

transferring some information towards another sensory channel [52]. This process, called sensory substitution, has often been deployed in assistive technologies, to overcome limitations of blocked sensory channels (e.g., for the visually disabled). However, transferring information between modalities has also been achieved for other purposes: data sonification is one example [54]. We assume multisensory view management can have a positive effect on performance in narrow FOV displays as visual information density (complexity) would be reduced: some information would be transferred and thus reside in an other perceptual channel. The usage of a non-visual sensory channel could be particularly useful for higher density environments. As an example, this transfer could take the form of sonified label details, but also the provision of additional cues that support guidance or localization, being the main focus of this paper.

Multisensory view management is still an open field for exploration in AR. Not surprisingly, the potential and implications of multisensory view management in relation to information density is not well understood. Even though multisensory interfaces exist [50], they are used infrequently and with few exceptions (e.g., the audio notes presented in [45]) for other purposes than view management. To shed light into this area, we will present the results of multiple studies that compare different audio-tactile methods on their ability to convey not only longitude and latitude, but also depth information. We do so by looking into the usage of audio and vibrotactile cues for (a) guided search performance, where users are guided towards a target (study 1 and 2) and (b) information location provision, where users are informed about the location of additional information inside (e.g., further away) and outside their FOV (study 3). Both directions are of high relevance, as the search for information is a common task in AR applications [73]. In this paper, we regard our methods as an integral part of the view management system. However, applications can be envisioned where it can also be used independently, e.g. in navigation systems.

1.1 Contributions

We present the following contributions that provide more insights into the usefulness of multisensory view management for in particular narrow FOV displays. To provide non-visual cues, we make use of a novel tactile interface extension for the Microsoft HoloLens.

- We explore audio and tactile cues for encoding longitudinal, latitudinal and distance information guidance without visual cues, showing that the mode that encoded latitude with audio and depth with vibrotactile pulse exhibits the highest accuracy in latitude estimation and also highest subjective preference (Study 1).
- We use the same audio and vibrotactile cues for a guided search task with the presence of visual information, where we showed that users could complete the task also quickest with the mode that encoded latitude with audio. Again the aforementioned mode was preferred most (Study 2).
- Finally, we investigated how audio-tactile cues can be used to determine the absolute longitudinal position and depth for localizing information (instead of the relative feedback used for guidance), showing users can define the position of a cue with relative precision when audio depth feedback is used. Generally, depth can be judged more precisely in the area that is close to the user (Study 3).

Head-based vibrotactile guidance cues have been studied before e.g., in Virtual Reality applications, in wider FOV immersive displays or guidance of the visually impaired. However, these approaches lacked the necessary distance cues [18] or are dependent on a high-resolution grid over the full head, being not feasible for mobile AR setups, while also not focusing on visual search [38]. Furthermore, cues were studied in absence of audio cues. We progress beyond the state of the art by providing non-obtrusive non-visual feedback methods not only for guidance towards a target (directional and distance), but specifically also for information localization in AR information spaces. Thereby, we introduce new mode combinations, by using both vibrotactile and audio cues. We show and discuss performance measurements, extending previous findings that mainly focused on guidance aspects that only in part would cover for view management requirements in AR.

1.2 Related Work

Our studies touch upon several fields of research, namely view management, visual search and guidance methods. View management methods have been developed since long time [7], optimizing the layout and appearance of information. Among others, researchers have looked into label placement for size and position [4, 7] and depth-placed ordering [64, 65]. The appearance of labels has also been focused upon, for example in relation to foreground-background issues [27], or the legibility of text [26, 48]. While view management for wide FOV displays has found some interest [41, 44], with few exceptions (e.g., [68] and [80]) there has hardly been any focus on view management for narrow FOV displays, a gap we address in this paper.

With respect to guidance, the usage of visual aids to accelerate search has been studied for long [86], also in relation to more complex search tasks [62]. Visual search is affected by the types of features the search target and distractors elicit, which have been widely discussed in various theories [67], while specific aspects relevant for AR such as target eccentricity, orientation [15] and depth [57] have also been focused at. In general, search behavior has been studied widely, also specifically in AR by using eye tracking [25]. While visual cues such as the pop-out effects [29] have found reasonable wide application to draw the user's direction towards an item [69], also less obtrusive methods have been studied. Examples include subliminal cueing [66] and saliency modulation, also with specific application in AR [82]. Furthermore, the usage of specific pointers to targets, like arrows or attention tunnels have been studied [74]. Another common example of visual aids used for guidance purposes are head-up displays (HUD). HUDs are widely used in the aircraft sector, among others for basic navigation, flight information and combat operations [1, 5, 60], pathway guidance [24] and to increase situation awareness of pilots [22, 23]. Similar to that, windshield HUDs are becoming more common in cars, where navigation [39, 59] and attention factors [72, 79] have been studied. Furthermore, HUDs can be used to guide through assembly tasks and manufacturing, [17, 78] and maintenance processes [31]. Finally, traditional visual overview methods like 3D Arrows and modern approaches such as EyeSee360 and 3D Radar [11, 28] have also been used to speed up search performance.

With respect to non-visual guidance methods, the usage of vibrotactile cues has been adopted quite frequently to direct navigation [49, 81] 3D selection [3, 55], target finding on mobile AR devices [2] and visual search tasks [46]. Of direct influence to our physical setup are the ring-based tactile guidance systems around the user's head [8, 18], and the top head/forehead system with a higher resolution tactor grid resembling an EEG setup [38]. Audio has also been used to guide visual search [58] and navigation [36]. Examples include studies that look specially at the effects of motion, location and practice on visual search performance with 3D auditory cues, e.g., with audio improving search performance by about 22-25% [56]. Audio cues have also been adapted in visual search tasks based on gaze direction [53]. Finally, within the frame of visual search tasks, cross-modal effects have been studied, including audio-tactile effects [33, 61] and conflicts between audio and visual cues [42]. Sonification strategies also use auditory cues to inform or guide the user. These paradigms use the main perceptual attributes of a sound, namely pitch, loudness, duration/tempo, and timbre with respect to the presence of the auditory reference. Pitch is by far the most used auditory dimension in sonification [19]. This metaphor can be also found in modern parking car systems, where the distance information is provided through a decreasing time interval between impulse tones [63]. Furthermore, this method can be applied for spatial data exploration and guidance [37, 75] and to support navigation tasks for visually impaired people in AR [10, 36, 70]. Another application area of sonification is the improvement of accuracy during the performance in high precision tasks [9, 71], e.g., in medical AR without obstructing the visual field with additional information.

With respect to our vibration methods, insights of [18] are of critical importance for this paper. In this work, the authors created a tactile guidance system consisting of seven tactors placed around the user's head to improve spatial awareness. To study performance, virtual spheres were placed systematically in the main experiment on four different elevation angles (45°, 22.5°, 0°, -22.5°). Positions of a 3D target around the person on the horizontal plane were indicated by "pointing" towards the direction of the object using a vibration on the according vibration motor on the users head. If the user turned the head to the direction of the target object, the vibration moved to the center of their forehead. The vertical position of the object on the other hand was indicated by varying the vibration frequency that increased towards to the target elevation angle and peaked at the correct target position on the vertical plane. For that frequency modulation, a quadratic growth function was used since it allowed a more accurate, precise, and faster target localization in an active head pointing task compared to other tested growth functions. Results showed that subjects using the vibrotactile setup could find targets in different positions with higher accuracy, precision and lower reaction times over time as an effect of learning. Overall, the results of [18] indicated that the overall mislocalization of a target was about 7% on the horizontal position and 4.5% on the vertical position.

2 SYSTEM APPROACH AND IMPLEMENTATION

Within this paper, we study the usage of audio and vibrotactile cues in cohesion with visual information. To provide tactile cues, we created a novel tactile interface extension for the Microsoft HoloLens, depicted in Figure 1. The extension consists of a row of 5 vibrotactors along the temples and the forehead in 45° intervals, schematically depicted in Figure 2.

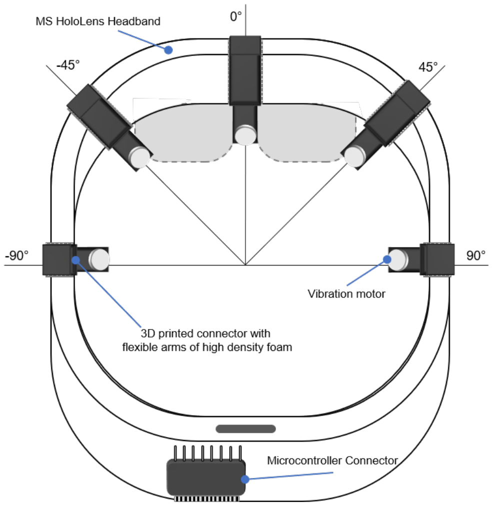


Fig. 2: Custom made tactor attachment on the Microsoft HoloLens headband with 5 vibrotactors placed in 45° intervals. A connector attached at the rear of the headband ensures a flexible an easy connection to the microcontroller.

We used Precision Microdrives pancake vibration motors with a diameter of 8mm (model 308-100). The vibrotactors are attached to the HoloLens headband using custom 3D printed connectors to which flexible arms of high density foam are connected (see Figure 1, Right). This is the result of an extensive iterative design process, as a construction had to be found that would provide good tactor-skin contact without pressing the vibrotactors too light or too hard to the head. Among others, this had to be achieved to avoid too much head vibration due to bone conduction through the skull: the skin touches the skull almost directly, which makes localization of cues difficult as a larger area on the skull may vibrate. Due to the flexibility of the arms, the pressure on the tactors is automatically adjusted for different head shapes, and can be worn comfortably.

The system was implemented using Unity, version 2018.1.0f2 together with the Microsoft Mixed Reality Toolkit v2017.4.3.0. The vibrotactors where connected to a Raspberry Pi 3 Model B+ running a python-based version of Open Sound Control to communicate with the Unity App on the HoloLens. We used the Microsoft HRTF Spatializer plugin in Unity to enable spatial sound.

With respect to the non-visual feedback methods, we distinguish between the categories **longitudinal, latitudinal**, and **depth** cues. For these categories we developed different metaphors to transcode visual cues into audiotactile feedback for multisensory view management. The methods reported in the next sections are the result of pilot testing (see Section 3).

2.1 Longitudinal feedback

For study 1 and 2 we reimplemented the metaphor for longitudinal feedback from [18], see Figure 3, and adapted it to our system. We did so as the target selection performance on the horizontal plane was shown by the authors to be particularly good. In the original implementation the user is informed about the relative position of the target in the horizontal plane by the tactor position in the vibrotactile setup, while the motor frequency is depending on the target elevation. If the target angular position horizontally is located between two tactor positions both motors vibrates. In case of longitudinal absolute feedback (study 3) where all targets are placed on the same latitudinal plane, motor intensity of both motors is set in relation to the angular distance of the target. This is done to achieve an interpolation effect to indicate that a target lies in between the physical motor setup, similar to the phantom effect described in [34].

With respect to audio, considerations about using absolute auditory cues to find targets on the horizontal plane by making use of the HRTF were discarded since in comparison to lateral localization, a generic HRTF itself might be not enough to localize a sound precisely in the frontal area (see for a discussion [12, 40] and specific details on front-to-back confusion in [35]).

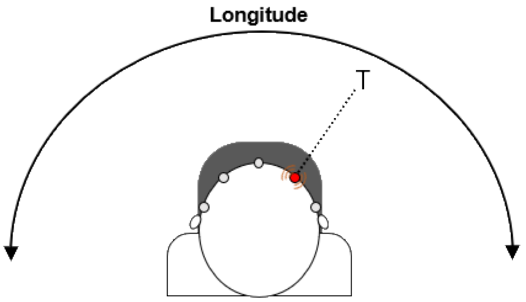


Fig. 3: Longitudinal feedback method by motor position adapted from [18].

2.2 Latitudinal feedback

For latitudinal feedback, we created two different modes, namely vibrotactile (Figure 4A) and auditory (Figure 4B). Both methods use the adapted modulating function with a quadratic growth of [18]:

$$Latitude_{intensity} = Latitude_{Audio} = \frac{100 - 5/6 * \sqrt{-(x-180)*x}}{100}$$

$$\text{where } x = \alpha_{cameraRotation} - \alpha_{targetRotation} \text{ in degrees}$$

In case of vibrotactile feedback by intensity modulation (Figure 4A)), the range is between $[25, 100]$, where 25 is the minimum and 100 the maximum intensity in percent to drive the particular vibrotactor in relation to the elevation distance to the target. 25% is used as minimal frequency as it has been shown that this value (approx. 50 Hz) is sufficient to overcome initial motor inertia and is perceptible as a low vibration for the users [55].

In case of latitudinal audio feedback (4B), the modulating function adjusts the pitch and the volume of the sound source instead of the vibration intensity with its highest frequency and volume on the target elevation level. Unlike [18], we did not discretise the latitudinal intensity calculations into nine frequency levels, but used a continuous form to benefit from the high resolution of the human hearing mechanism. The human auditory cortex is able to discriminate even smallest changes in frequency thresholds (1 to 3 Hz for frequencies up to about 1000 Hz) [30]. In contrast it has been shown that users are able to discriminate a maximum of only 9 levels of frequency on the skin [13]. Therefore we expected it to perform better than the frequency adjustment of the vibrotactile cues on the users forehead.

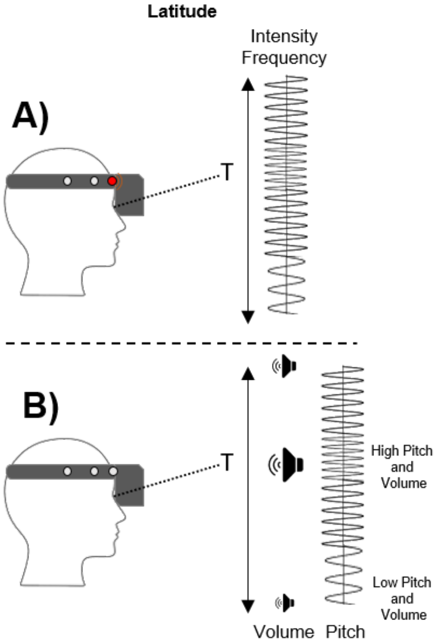


Fig. 4: Two variants of latitudinal feedback: A) Analogous to [18]; B) shows the same feedback but the frequency modulation of the vibration motor is replaced with a sound. Volume and pitch adjusted by the target elevation.

For latitudinal audio feedback we played sounds in the range of 300 Hz to 1300 Hz frequency depending on the current elevation level. A 300 Hz sound is played if the user is very far away located from the target on the elevation plane. The closer the user is getting to the target elevation, the higher the frequency of the sound gets adjusted, reaching its maximum of about 1300 Hz right on the target elevation level. We chose these values as human frequency discrimination works quite well within that range and higher frequencies can be perceived as unpleasant over time [16]. Additionally the volume level of the sound increases in a similar manner, with closer objects sounding stronger.

2.3 Depth feedback

With respect to depth, analogous metaphors to the latitudinal feedback are applied to ease learning and potentially reduce cognitive load. We differentiate between two implemented modes: auditory depth feedback by adjusting volume and pitch (Figure 5A), and using a variable on/off pattern (Figure 5B) of the specific vibration motors dependent on target depth - hereafter referred to as pulse. For depth calculation, the following equation is used:

$$Depth_{Audio} = Depth_{Pulse} = \frac{100 - 25 * \sqrt{-(y-6)*y}}{100}$$

where y is the distance to the target in meter

Target depths are set between 1-3 meters in studies 1 and 2 since this region works well to place augmentations within the HoloLens. The auditory metaphor is the same as for the latitudinal feedback but adapted to the target depth, visualized in Figure 5A. For pulse feedback, results of before-mentioned equation are scaled into values $[0.1, 0.5]$ in seconds and represents the pulse frequency of a vibration motor. If the target is very far away on the depth plane, both the time the motor is turned on t_{on} and turned off t_{off} is set to 500ms. That on/off pattern is noticeable for the user as a slow pulsating vibrational feedback. t_{on} and t_{off} then successively gets faster the closer the user gets to the target. Right on target depth, the pulse frequency is set to 100ms for t_{on} and t_{off} to create a very fast vibrational pattern. This method is adapted from car parking metaphors that are easy to understand for most people. This behavior is illustrated in Figure 5B. 100ms is chosen as maximum pulse speed to comply with the physical restrictions of the used vibration motors, where a faster on/off pattern would lead to interferences where motors do not have enough time to rise up due to the specific motor inertia [55].

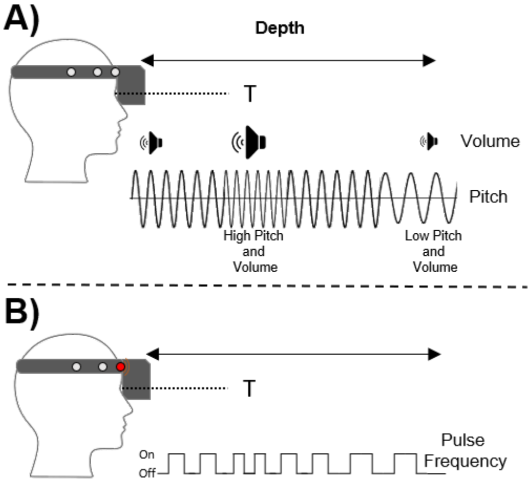


Fig. 5: A) Analogous feedback like in Figure 4A, adapted to depth. B) Pulse frequency adjustment depending on the target depth.

3 PILOT STUDY

In order to generate an integrated non-visual guidance approach, the previously mentioned metaphors of longitudinal, latitudinal and depth cues had to be integrated into a single mode. We use longitudinal feedback as described in Section 2.1 (direction indication by tactor position) since it already delivered good results in [18], is intuitive in its usage to describe a horizontal direction and is easy to learn. Other alternatives like using auditory cues for longitudinal feedback were discarded as many localizing issues exist [47], especially using non-individualized HRTFs [83, 84]. Latitudinal and depth feedback on the other hand could be either indicated by frequency modulation, pulse, or audio adjustment.

To combine all possible metaphors into one mode, it is necessary to ensure that each metaphor (frequency modulation, pulse, audio) only occurs once in each mode. Allowing one metaphor for two geographical indications (e.g. pulse metaphor for both latitude and depth) would make them indistinguishable and lead to confusion for the user. Taking this requirement into account results into 3^2 considerable permutations for feedback modes to be examined for the main study as presented in Table 1.

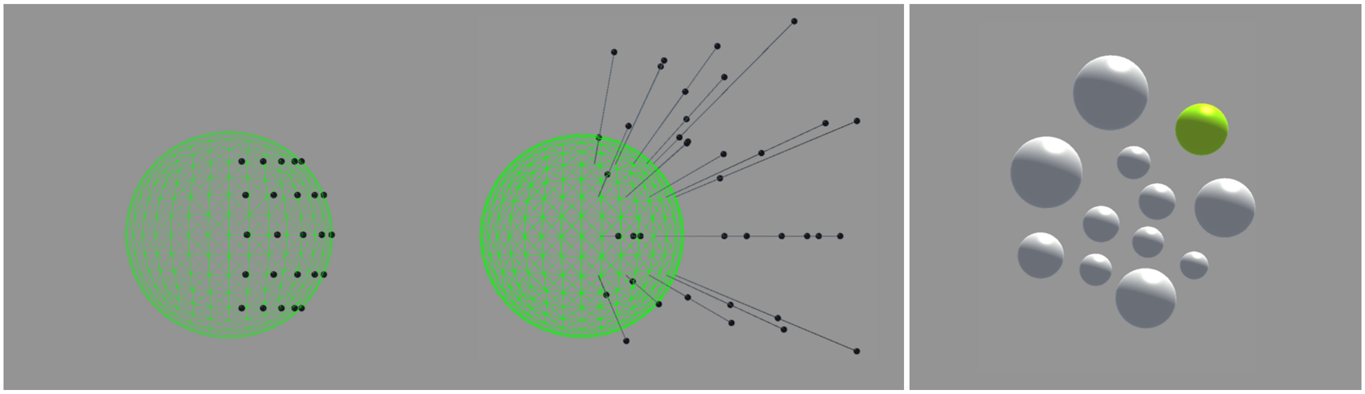

(A) (B) (C)

Fig. 6: A) Item population like in [18] on a hemisphere around the user (side view). B) Adding factor depth by giving the items a random depth position between 1-3 meters for study 1. C) Example item cluster in study 2, used instead of single items in study 1 - the target is highlighted after selection.

Table 1: Considerable permutations for feedback modes to be examined for the main study.

No.	Londitude	Latitude	Depth
1.	TP	V-i	A
2		V-i	V-p
3.		A	V-i
4.		A	V-p
5.		V-p	V-i
6.		V-p	A

TP = Tactor position, V-i = Vibration intensity,
V-p = Vibration pulse, A = Audio (Pitch/Volume)

We tested all possible feedback combinations in a pilot study with 6 users with respect to their usefulness, usability and intuitive usage. The initial idea of mode no. 1 & 2 was to extend the feedback method from [18] with additional depth cues for object localization. These two modes showed already promising results during the pilot phase where depth cues by audio and by vibrotactile pulse patterns were well accepted and understood by the participants. Mode no. 4 (latitude/audio & depth/pulse) revealed a good usability for the purpose of guidance as well. Users stated that they could comprehend the cues well with a relatively high accuracy on the latitudinal plane using audio cues. Modes no. 3, 5, and 6 on the contrary showed slightly worse performance and ratings compared to the before mentioned modes. These combinations were rated as less precise according to depth localization compared to auditory or vibrotactile pulse cues. This behavior might be explained by the fact that audio and pulse cues for distance might be perceived as more intuitive by experience gained from real world metaphors like acoustic parking system in cars. Finally, as a result of the pilot study, we focused just on the most promising feedback modes for the subsequent main study, namely mode 1 (latitude/intensity & depth/audio), mode 2 (latitude/intensity & depth/pulse), and mode 4 (latitude/audio & depth/pulse), see Table 2. This also provided the advantage that users would not get strained or confused by the need to learn too many different feedback modes.

Table 2: Three isolated cue combination modes with longitudinal, latitudinal and depth feedback for study 1 and 2. Each mode uses the longitudinal metaphor presented in Figure 3.

Study	Mode	Longitude	Latitude	Depth
1+2	1	TP	V-i	V-p
	2		V-i	A
	3		A	V-p

TP = Tactor position, V-i = Vibration intensity,
V-p = Vibration pulse, A = Audio (Pitch/Volume)

Furthermore we tested how robust these modes are in AR applications. Hereby we wanted to know about the limitations of the approach described in [18], especially regarding resolution to find a specific object in dense scenes. For this purpose, we manually created 10 different clusters of items to populate the scene. Clusters were generated by placing 12 spheres into a fixed radius and giving each sphere a random depth position (see Figure 6C). Positions were then manually adjusted to avoid occlusion of items. Interdistances between the spheres were gradually tested and reduced until targets within a cluster could not be differentiated precisely by longitudinal and latitudinal cues anymore. To ensure that all generated clusters were indistinguishable, the entire cluster received a random rotation and the option to be mirrored horizontally and/or vertically. Finally performance comparable to [18] could not be achieved anymore when replacing the single targets with our generated clusters. Yet, we assumed we could overcome this problem by the usage of additional depth cues next to the longitudinal and latitudinal feedback to facilitate the identification of the correct target in a cluster. We assessed this assumption in study 2.

4 USER STUDIES

We performed three user studies to assess different aspects of non-visual view management, each addressing a different research question (RQ). 12 participants (1 female) aged from 20 to 31 took part in the studies. Prior to the experiment, participants were informed about the study, and signed an informed consent form. They were recruited via a university mailing list (employees and students) and received an Amazon voucher for their participation. Post-experiment questionnaire assessed user preference, cognitive load and usability on an 11-point Likert scale. All studies were performed by the same users. The order in which studies were performed was partly balanced. Half of the users performed study 1 and 2 first, the other half started with study 3 followed by studies 1 and 2. Study 1 was always followed by study 2 as study 2 was based on, and extended study 1. In studies 1 and 2 three different guidance feedback modes were tested that encoded information on the relative target location in the 3D-space. In study 3, we compared two feedback modes that encoded the absolute target location on ground level. Accuracy measures were 1) the directional error on each axis (longitude, latitude and depth) which was calculated as difference between the selected and correct target position, 2) the absolute error on each axis and 3) the euclidean distance of the selected and correct target position. Completion time was also recorded and especially focused in study 2.

4.1 Study 1 - Guidance Accuracy

RQ1: What is the guidance accuracy towards a spatial target of each audio or tactile mode?

In this study, users were asked to place a virtual sphere at the location they were guided towards. They were told to perform the task as precisely as possible without a time limit. A one factorial within-design was used to examine the effect of the guidance feedback mode (modes 1-3, see Table 2 on accuracy performance). All possible target items were placed analogous to [18] around the user, in our case within the grid cells of a unit spheres surface with a radius of 1 meter on five elevation angles (45°, 22.5°, 0°, -22.5°, -45°), see Figure 6A. However, the grid with spheres was in the actual experiment not visible to the user. Additionally to that procedure, the items were set to a random distance of 1-3 meters (Figure 6B). As in the outcome of [18], we used for our final experiments only targets on four elevation angles (45°, 22.5°, 0°, -22.5°), since searching on $-45°$ levels was stated there as physically too demanding over time.

Different feedback modes were tested blockwise with 11 trials per mode/block. The order of blocks was balanced across participants. Each block started with 2 training trials in which correct target position was always shown, followed by a third training trial that followed the same procedure as the following 8 performance trials. At the beginning of each trial the user was shown the current mode for guidance feedback. After pressing a confirmation button, a sphere appeared in front of the participant. The sphere was always in the viewing direction of the user (based on head tracking) and could be moved along longitude and latitude by turning the head. Depth/distance of the sphere could be increased/reduced by pushing/pulling the right analog stick on a gamepad. Using the feedback the user could move the sphere to the location where he/she thought the feedback referred to and press a confirmation button on the gamepad. Afterwards the user was shown the correct target position before the next trial started. We assumed this should facilitate improvement over time.

4.2 Study 2 - Guidance Completion Time

RQ2: How fast can users perform with each audio or tactile mode?

In study 2, users had to find a target object as fast as possible. The study employed a one factorial within-subjects design to examine effect of modes (the same as in study 1, see Table 2) on search time performance. Users were guided towards a visible cluster of spheres (see Figure 6C and Section 3) where a single target could not be matched solely based on feedback for the horizontal and/or vertical position alone since all possible targets were positioned very close to each other. Again analogous to [18], the clusters were set in the same manner like in study 1. Yet now we populated the scene with visible clusters instead of (not visible) single objects. Users were guided towards the object using the feedback of study 1 (see Table 2). Users had to select the target sphere among other spheres as quickly as possible by placing a head tracked cursor in the center of the field-of-view on the sphere. As we avoided an occlusion of more than 50% each sphere could be selected in that way. Unlike in study 1 where depth cues were adjusted by moving a virtual sphere, depth feedback was triggered by focusing a possible target of the cluster with the cursor. The feedback was on the highest level at target depth. If finally longitudinal, latitudinal and depth cues were all on highest level, the user could be certain to have found the right target. The distance of the indicated sphere to the target was recorded.

4.3 Study 3 - Information Localization

RQ3: How well can absolute audio or tactile feedback be used to provide information on target locations?

In contrast to the relative feedback in studies 1+2 we used what we call ”absolute” feedback in study 3. The user always looked straight ahead (hence, without moving the head) while getting feedback on the target location that always remained the same. This was in contrast to the relative guidance cues that would change based on e.g. head direction or closeness to the target. Absolute feedback provided information on the longitudinal position and depth of the target location that was always at the same elevation level in this study. We only used the longitudinal location in absence of the latitudinal position as we assumed that information localization based on general direction and distance would be sufficient for most AR applications, e.g. city information systems where information mostly resides on a plane. During the experiment, users were facing a display that showed a semi circle that looked like Figure 7 without annotations, colors and data points. The semi circle was divided in three different depth areas (near, middle and far) and four angular zones (from 0° to 180° in 45° steps). The area was subdivided to ensure target positions were rather evenly distributed across different zones. The user was instructed to imagine being located in the center of the semi circle (small black triangle in Fig. 7), showing a top-view of the scene. We did not use a specific depth unit. When logging performance data we set depth range from 0 to 1, with 0 being the closest and 1 the most distant point. A one factorial within-subjects design was applied to study the effect of the encoding mode of depth on performance measures (angular distance, directional and absolute difference of indicated and and target depth, distance between indicated and target position). Modes were tested blockwise, while the order was balanced across participants. In each block users got 15 training trials for targets placed on angle directions (0°, 45°, 90°, 135°, 180°) and 12 training trials on different (interpolated) positions between the angles with the respective feedback mode. They provided feedback while the corresponding target position was shown on the display in the semi circle at the same time. Training target positions were chosen to let the user understand the feedback range in depth, as well as the interpolated feedback on longitude between two tactors. After the training the user completed 48 performance trials. In each trial the user had to click a position in the semi circle where he/she thought the feedback referred to.

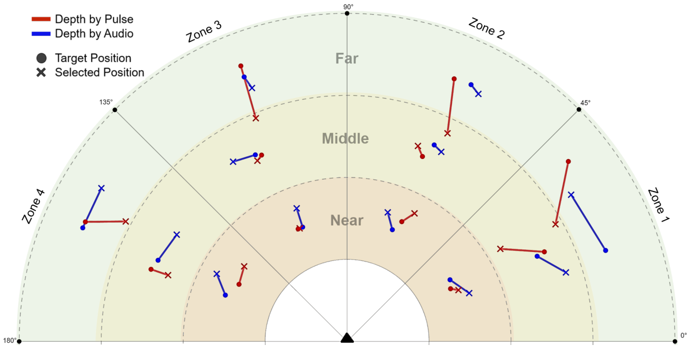


Fig. 7: Zones and depth areas in study 3. The lines depict the average offset of all ratings per zone which each method. For illustration purposes we used the average target points per each zone as the start point of the lines. Interestingly, the ratings are not completely symmetrical. The near depth area corresponds to the range of 0 to 33% signal strength, the middle area to 33% - 66% and the far area 66% to 100%. Zones correspond to angles in a polar coordinate grid.

5 RESULTS

Friedman test was used to to analyze the effect of feedback mode on accuracy performance and completion time. Performance was computed as difference between indicated and correct target position for longitude and latitude (degrees) and for depth (meters). The absolute error was also computed and compared between conditions. The euclidean distance was used as additional measure that considered both errors. Wilcoxon signed-ranks tests were applied for post-hoc pairwise comparisons and to compare questionnaire ratings. Pearson's correlation coefficient *r* was used to measure effect size. Spearman's rank correlation was computed to assess the relationship between trial number and performance to study training effects. We only report on the salient results.

5.1 Guidance Accuracy

There was no significant effect of mode on directional and absolute error in longitude, or depth but on absolute latitude error and on euclidean distance. Absolute latitude error and euclidean distance were lower with the latitude/audio & depth/pulse mode compared to mode latitude/intensity & depth/audio ($r_{Lat} = 0.57, r_{Euc} = 0.47,$) and latitude/intensity & depth/pulse ($r_{Lat} = 0.42, r_{Euc} = 0.39$), see Table 3 and Fig. 8.

Table 3: Absolute errors in longitude, latitude, depth and the euclidean distance of the indicated and target position of the sphere in study 1.

	V-i, A	1.98	5.99	0.03	0.24
1	V-i, V-p	1.93	3.60 **	0.07	0.17*
	A, V-p	2.14	1.37*,**	0.05	0.11*
	X^2(2)	ns	16.76**	ns	13.5**

V-i= Vibration intensity, V-p=Vibration pulse,
A= Audio (Pitch/Volume), * = p <.05, ** = p <.01, *** = p <.001

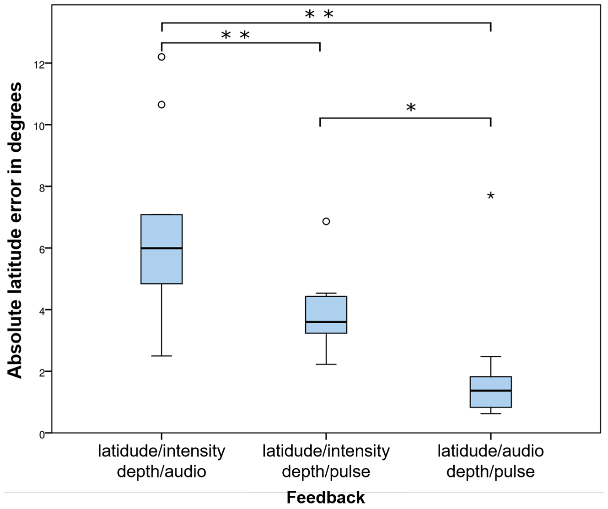


Fig. 8: Absolute latitude error in degrees by mode in study 1.

Modes that encoded latitude by vibration intensity also differed significantly from each other regarding latitude error and euclidean distance. Users performed better when depth feedback was encoded with pulsed vibration compared to audio ($r_{Lat} = 0.54, r_{Euc} = 0.47,$). Furthermore there was a small significant negative correlation between trial number and absolute latitude error only for the latitude by audio encoding mode ($r_{rho} = -.21, p = .03$) which indicates there was an improvement over time (see Fig. 9). The latter indicates that showing the correct position of the target after each trial positively affected learning.

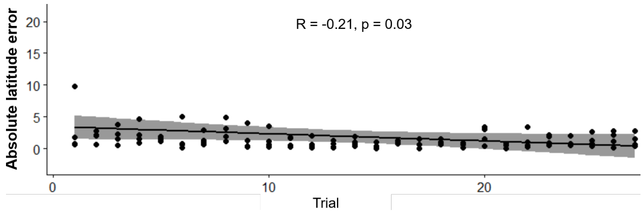


Fig. 9: Absolute latitude error in degrees by trial number in study 1.

5.2 Guidance Completion Time

There was no effect of mode on absolute and directional error in longitude and depth, directional latitude error and on euclidean distance. Generally, the correct sphere was identified with each mode. There was a significant effect of feedback mode on completion time ($X^2(2) = 16.67, p < .001$, see Fig. 10). Users were faster with the mode that encoded latitude with audio ($M = 14.2, IQR = 12 - 17.4$) compared to the mode latitude/intensity & depth/audio ($M = 15.9, IQR = 14.6 - 24.8, Z = 2.04, p = .041, r = 0.42$) and latitude/intensity & depth/pulse ($M = 21.2, IQR = 16.9 - 24.9, Z = 3.06, p = .002, r = 0.62$).

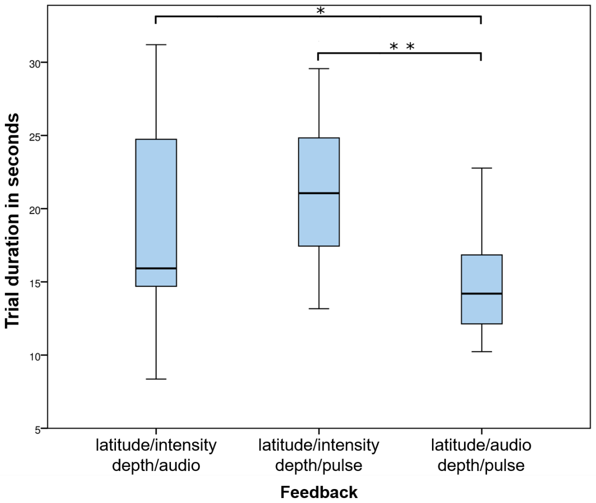


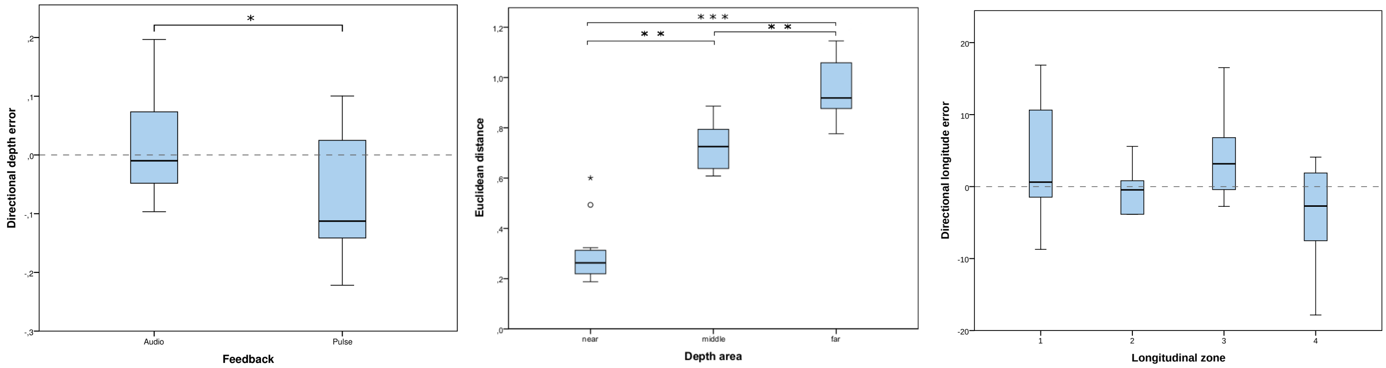
Fig. 10: Time to complete a trial in seconds by mode in study 2.

There was a marginal effect of the feedback mode on the rate of the correctly chosen cluster ($X^2(2) = 5.82, p = .055$ and absolute latitude error ($X^2(2) = 5.91, p = .052$). Post-hoc pairwise comparisons showed no significant differences regarding absolute latitude error. Descriptive values indicate that latitude error was lower with the latitude/audio & depth/pulse mode ($M = 0, IQR = 0 - 0.23$) compared to latitude/intensity & depth/pulse ($M = 0.49, IQR = 0 - 3.08$) and latitude/intensity & depth/audio ($M = 0.43, IQR = 0 - 2.61$). The correct cluster was chosen more often with the latitude/audio & depth/pulse mode ($M = 1, IQR = 1 - 1$) than with the mode latitude/intensity & depth/pulse ($M = 1, IQR = 0.88 - 1$), $Z = 2.07, p = .038, r = 0.42$). Furthermore there was a correlation between trial duration and trial number only for the mode latitude/intensity & depth/audio ($r_{rho} = -.49, p < .001$), see Fig. 12.

5.3 Information Localization

There was no difference between depth/audio and depth/pulse coding regarding the absolute longitude error, the euclidean distance, trial duration (see Table 4) and directional errors. Performance was better with audio than with pulse depth feedback regarding absolute depth error ($Z = 2.04, p = .041, r = 0.59$, see Table 4) and the directional depth error ($Z = 2.59, p = .01, r = 0.75$, see Fig.11A). With pulse depth feedback target depth was significantly underestimated ($Z = 2.28, p = .023$) in contrast to audio feedback ($p = .774$). Consequently, the correct depth area was also chosen more often with audio (hit rate: $M = 0.63, IQR = 0.47 - 0.74$) than with vibration pulse (hit rate: $M = 0.54, IQR = 0.38 - 0.57$), $Z = 2.0, p = .045, r = 0.29$.

Furthermore depth zone (see Fig. 7) of the target had an influence on the absolute depth error ($X^2(2) = 6.5, p = .039$), directional depth error ($X^2(2) = 7.17, p = .028$) and euclidean distance ($X^2(2) = 22.17, p < .001$.). Post-hoc Wilcoxon pairwise comparisons showed significant differences between depth areas only regarding euclidean distance. Users performed the better the closer the area was where the target was located (see Fig. Fig. 11B and 7). The same pattern occurred with both feedback modes. Euclidean distance was significantly lower in the near area than in the middle ($Z = 3.06, p = .002, r = 0.44$) and lower in the middle compared to the far area ($Z = 2.98, p = .003, r = 0.43$). As no effect on longitude error was found, the differences in euclidean



(A) (B) (C)

Fig. 11: Study 3: A) Directional depth error by feedback mode. B) Euclidean distance by depth area. C) Directional longitude error by longitudinal zone. See Fig 7 for depth areas and zones.

distance between depth zones mainly based on the error in estimation of depth.

Furthermore there was an effect of longitude zone on directional longitude error ($X^2(2) = 8.9, p = .031$,). However, Wilcoxon post-hoc pairwise comparisons were not significant. Descriptive data showed that in zones 2 and 4 the indicated angle was slightly underestimated and slightly overestimated in zones 1 and 3 (see Fig. 11C).

Correlation analysis showed there was a negative correlation between trial number and the correctly chosen longitudinal zone for the audio depth feedback ($r_{rho} = -.195, p = .001$) and a positive correlation with the absolute longitude error ($r_{rho} = .148, p = .012$), indicating users performed slightly worse over time. For the pulse depth feedback trial number correlated negatively with trial duration ($r_{rho} = -.15, p = .011$) and directional depth error ($r_{rho} = -.165, p = .005$): The underestimation of the target position increased over time, indicating there was also a slight performance decrease with pulse depth feedback.

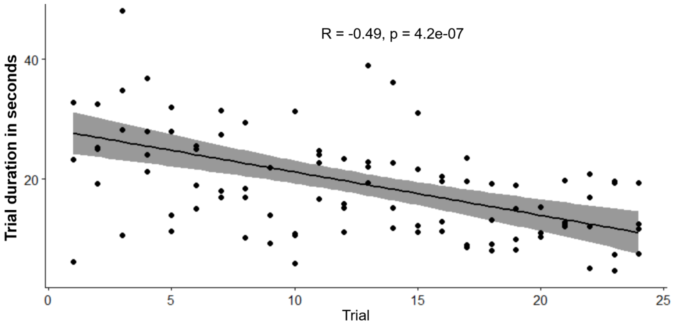


Fig. 12: Time to complete a trial over time with mode latitude/intensity & depth/audio in study 2. The Figure indicates the significant performance improvements caused by a learning effect that was only found in this mode combination.

5.4 Training Effects

The estimation performance of the longitude and depth in study 1 was not affected by the order in which studies have been performed. That is, participants who finished study 3 first did not perform better than the group that started with study 1, which indicates there was no training effect regarding the estimation of longitude and depth. With respect to latitude there was also no difference between the medians of the groups but a small reduction of the interquantile range when users had performed study 3 before. That is, users generated less extreme values and showed a more stable (but not better) performance in study 1 if they had finished study 3 before. Regarding performance in study 3, median errors were rather similar for the group that performed studies 1 and 2 first and the group that did study 3 first, which indicates there was no training effect on median performance. Furthermore we could observe smaller interquantile ranges in errors for longitude and depth estimation in study 3 when users had performed studies 1 and 2 before. That is, as in study 1 training slightly affected only the variability of performance but not the median.

Table 4: Euclidean distance and absolute errors in longitude and depth of the indicated and target position with interquartile ranges in study 3 and trial duration. * = p <.05.

	Longitude error	Depth error	Euclidean distance	Trial duration
A	9.92 (7.7-13,8)	0.14 (0.12-0.20)*	0.7 (0.6-0.8)	3.1
V-p	9.34 (7.7-15.2)	0.18 (0.16-0.18)	0.62 (0.6-0.7)	3.1

5.5 Questionnaire

Users indicated that the augmented image was not disturbed by vibration ($M = 10, IQR = 3.25$) – an important issue as the vibration elements were directly attached to the headset – and that the headset was rather comfortable to wear ($M = 8, IQR = 2.5$). For each feedback mode in each study users rated overall task easiness, ease of learning the feedback and feeling of accuracy (see Table 5). Users preferred the latitude/audio combined with depth/pulse encoding feedback mode in both guidance studies (1 and 2) as ratings for overall task easiness, ease of learning the mode and feeling of accuracy were significantly higher for this mode compared to the latitude/intensity encoding modes with depth/audio and depth/pulse. Ratings were generally very positive for the guidance feedback: Median ratings for the most preferred mode latitude/audio were always 10 and higher and above 7.5 for all modes. In contrast to the guidance studies users rated their feeling of accuracy lower for the absolute target localization task: The feeling of accuracy got median ratings of 6.5 and a wider scattering of measured values. Users further rated the pulse depth feedback as easier to learn than audio although both modes received high ratings here (median above 8).

6 DISCUSSION

The results of our studies indicate the usefulness of both audio and vibrotactile cues to guide towards or inform the user about a location of further information that is not displayed visually. Here, we will discuss our findings and state the relevance of our results for view management systems. We should note that we cannot always directly compare our results to those reported in [18]. In their study potential targets were always visible and hit rate was used as performance measure. Instead,

Table 5: Median questionnaire ratings and interquartile ranges for different modes in studies 1-3. Significant differences between modes that resulted from Wilcoxon pairwise comparisons are marked. In case p-values varied between mode comparisons, different colors were used.

Median ratings and IQR

	Mode: Latitude, Depth	Overall task easiness	Ease of learning	Feeling of accuracy
Study 1	V-i, A	7.5 (2.5)	8.5 (2)	8 (1)
	V-i, V-p	8.5 (2.75)	9 (2.75)	8.5 (2.75)
	A, V-p	10 (1)**	10 (2)*	10 (2)**,*
Study 2	V-i, A	9 (2.75)	9.5 (3)	9 (2)
	V-i, V-p	9.5 (1)	9.5 (2.75)	8.5 (3)
	A, V-p	10.5 (1.75)*	10 (1.75)*	11 (0.75)**
Study 3	- , A	7.5 (2)	8 (2)	6.5 (3.75)
	- , V-p	7.5 (1)	9 (2)*	6.5 (3.75)

V-i=Vibration intensity, V-p=Vibration pulse, A=Audio (Pitch/volume)
* = $p < .05$, **= $p < .01$.

we used clusters instead of single sphere grids in study 2. Nonetheless, as we will show, our results indicate significant performance improvements when the vibrotactile feedback modality is extended by audio. Furthermore, we refrain from directly comparing our results to [38] as their setup is considerably different from ours by ways of resolution.

6.1 Guidance Accuracy

With regards to accuracy, we showed that the latitudinal accuracy can be significantly improved by using auditory cues, in comparison to the vibrotactile frequency modulation presented in [18]. By adjusting the pitch and volume depending on the target elevation (taking values between -22.5° to 45°, 0° corresponding to the eye level) users could be guided towards the target with a deviation of only 1.4°, which was 2.2° better in total compared to the best vibrotactile encoding of latitude with an deviation of 3.6°. Such an improvement of 61% makes a significant difference, especially when guiding towards a target in a dense AR space. Interestingly, latitudinal accuracy also differed significantly between modes that encoded latitude with the same vibrotactile frequency modulation, indicating performance on latitude was probably affected by the simultaneously provided depth cue: Users performed better in total with the vibrotactile mode when depth feedback was also vibrotactile (pulse vibration, accuracy error of 3.6°) instead of auditory (accuracy error of 5.7°), a performance improvement of 63%. The interaction between latitude and depth mode may indicate a crossmodal effect, which warrants further study. Furthermore, as the mode with the highest accuracy encoded latitude with audio and depth with pulse it may be concluded that specific metaphors are most suited to reach highest accuracy (auditory feedback for latitude in our case) and that feedback on different dimensions can potentially interactively affect performance on one dimension.

Regarding longitudinal performance, we could replicate findings of [18] that developed the encoding of longitude that we used for all modes. We found a high accuracy with an error of only 2° in all modes. We could further show that in contrast to performance on latitude, accuracy on longitude was not significantly affected by the feedback on other dimensions. This may be due to different nature of the feedback.

In case of latitude and depth users searched the position that emitted maximum feedback strength, whereas the correct longitudinal orientation could be found by moving the feedback to a certain location on the head. Thus, this kind of feedback can potentially have a higher resolution as smaller differences could be detected. Both in [18] and our study the head had to be turned till feedback was perceived at the forehead to find the correct orientation.

With respect to depth performance, users performed precisely with all modes. Errors ranged from 0.03m with latitude/intensity & depth/audio over 0.05m with latitude/audio & depth/pulse to 0.07m with latitude/intensity & depth/pulse, the best mode performing 57% better than the worst. However, differences were not significant.

6.2 Guidance Completion Time

Our results indicated that users could find the targets fastest while using the latitude/audio & depth/pulse mode, reaching a median trial duration of 14.2 seconds which was an improvement of 33% compared to the median search time with the latitude/intensity & depth/pulse mode (21.2 seconds) and 11% faster than the latitude/intensity & depth/audio mode (15.9 seconds). Variability of values was also lower, indicating users could reach shorter search times quite consistently. Although the latitude/audio & depth/pulse mode was slightly superior regarding the choice of the correct cluster, the latitude/intensity & depth/audio mode also performed very well. Over time participants significantly improved only with this mode, reaching shorter search times more consistently which indicates that with sufficient training this mode may potentially reach a similar search time performance as the latitude/audio & depth/pulse mode when searching targets with additional visual cues.

6.3 Information Localization

With respect to target localization through audio-tactile feedback, we showed that the longitudinal position in a 180-degree range could be perceived reasonably well by the participants through the tactor position with a deviation 9.9° when combined with audio and 9.3° when combined with the pulse condition. As expected, the difference was not significant as the same encoding of longitude was used in both modes. Regarding depth perception, users performed better with audio feedback with an accuracy error of 0.14 compared to 0.18 (22% improvement) with pulsed vibration which was in line with our expectations with respect to auditory perception [13, 30]. Generally, these results indicate that it is more difficult to locate absolute cues in depth than in longitude. In relation, users also subjectively noted a good but not excellent ability to judge location. It remains to be seen what depth accuracy is ideal for view management systems - often it may suffice to understand the approximate depth, to navigate and get closer over time to that point (e.g., reaching a restaurant a couple of blocks away). It has to be noted that the cue would turn from absolute to relative in this case, of course.

The improvement of depth estimation performance with increasing closeness of the target that we found may at first seem surprising. We used a frequency range from 300 Hz to 1300 Hz to encode the target position (the closer the target the higher the frequency). Frequency discrimination performance of the human ear is rather similar in this frequency range and even slightly better for lower frequencies. The superior performance for targets in closer areas that were encoded with higher frequencies may have occurred as we also modulated audio intensity: The closer the target the higher the volume intensity of the cue. The better human frequency discrimination performance for tones of higher compared to lower audio intensity [30] would explain the superior performance for targets in closer areas in our study. Furthermore the sensitivity of the ear increases as frequency increases from 300 Hz to 1300 Hz which could also have facilitated target localization with audio cues of higher frequency. Thus, human discrimination performance and sensitivity for different frequency ranges must necessarily be considered when providing absolute localization feedback as designers can make conscious decisions in which areas a high resolution is needed.

Interestingly, we found an asymmetric (as per comparison of longitudinal zones) under- and overestimation of longitude (directional error). While the overall test indicated an effect of longitudinal zone on directional error, pairwise group comparisons were not significant. Further study is required to clearly identify performance differences between longitudinal zones. Descriptive data indicate that the directional error could potentially be higher in more peripheral zones.

In comparison to guidance studies we can see that the interpretation of absolute feedback is more difficult than approaching a maximum value of relative feedback (lower accuracy performance and subjective ratings). Despite the higher demands we showed that absolute feedback can be used to provide information on target depth locations if median errors are acceptable.

7 IMPACT ON VIEW MANAGEMENT

In view management systems, visual-only techniques still dominate. Yet, especially in dense scenes problems like visual clutter, overlapping or occluding information and other visual conflicts arise that may lead to performance issues [43]. Using a multisensory view management system that transcodes visual information into audio-tactile cues may reduce visual complexity and potentially the number of distractors. To this respect, depth cues can also help to untangle visually cluttered scenes, something which can be very hard with longitudinal or latitudinal cues alone. Overall, we showed that the feedback mode latitude/audio & depth/pulse works best for non-visual guidance cue, making it an interesting option for interface designers to consider when developing guidance systems, potentially also in cohesion with other visual methods such as [28]. Doing so, attention [32] and crossmodal issues [33] should be regarded. Furthermore, layout methods likely need to be found that balance switching between visual and non-visual cues, especially when localizing multiple sources of non-visual information. Here, situation awareness will be an important factor to assess.

Generally, it is important to note that current research is not conclusive to when visual complexity may lead to sensory overload in narrow FOV displays and what effects it has on performance. While it was not the main focus in this paper, assessing sensory overload is a relevant topic for study. First results indicate that processing dense information spaces in narrow FOV affects search performance negatively [80], however more research is needed. Furthermore, while previous work has not shown significant negative effects of narrow in comparison to wider FOV on cognitive load [6], tasks have been usually of lower information density.

A further relevant issue to consider is how users can actually process multiple non-visual cues – for either guidance or localization – at once, as it can be expected that various sources of information outside the FOV (or e.g., at further depths) may be pointed towards. The processing of multiple stimuli – both single or across multiple modalities – is governed by attention mechanisms and affected by processing resources [77]. Hence again attention is of high relevance: it is a cognitive function that allows humans to continually and dynamically select particularly relevant stimuli from all the available information, in order to allocate neural resources. Thereby, providing information over multiple sensory channels may accommodate sensory stimulus integration [76]. However, in the case of view management, such sensory integration does not necessarily have to take place, as two processes may occur that are not spatially or temporally aligned or connected, hence are interpreted independently. For example – and related to our experiments reported in this paper – an auditory guidance signal may provide directional cues, while the user reads through visual labels to search for a particular target.

8 CONCLUSION AND OUTLOOK

In this paper, we presented a novel approach to improve guidance and information localization in augmented reality applications through non-visual cues provided in a Microsoft HoloLens. By providing audio and vibrotactile feedback along the temples and forehead, we are able to guide or inform the user on the longitudinal and latitudinal plane as well as in depth of targets. We expect this guidance can be particularly useful in AR environments with a high information density by transcoding information into audio-tactile cues. Hereby we extend current methods for vibrotactile guidance that could only used for guidance on longitudinal and latitudinal plane in context of VR setups. We showed that latitudinal precision and performance time can be significantly improved by using auditory cues (on latitude 61%, $p < .05$; 11% in time, $p < .001$), which contrasts vibration-only findings reported in [18, 38]. Furthermore, target localization by providing absolute cues worked precisely for both auditory and vibrotactile pulse feedback (error of 9° on longitude; 14-18% error on selected depth range).

Future work includes the integration of the audio-tactile guidance cues into a multisensory view management system. Considerations must be made about how and when (visual) information will be transcoded into audio-tactile cues – or alternatively, in other visual representation such as EyeSee360 [28] – depending on relative angular or depth location. Further hardware improvements contain the extension of the vibrotactile array on the forehead with more vibration motors and/or including the other parts of the head (comparing [38]). We will also look closely into multiple target guidance in complex situations. This requires follow-up studies that also look into which existing visual cues can be reasonably combined with audio-tactile feedback. Additionally it needs to be investigated which non-visual cues work best in combination without distracting or overloading the user with information. Finally, our methods can also have a positive impact on other domains, like navigation for visually impaired people, while we also expect VR guidance systems can be further improved through addition of audio.

ACKNOWLEDGMENTS

We would like to thank Johannes Schöning for comments that greatly improved the quality of this paper. This work was partially supported by the Deutsche Forschungsgemeinschaft (KR 4521/2-1).

REFERENCES

[1] E. C. Adam. Fighter cockpits of the future. In *[1993 Proceedings] AIAA/IEEE Digital Avionics Systems Conference*, pp. 318–323, Oct 1993. doi: 10.1109/DASC.1993.283529

[2] 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*, ICMI-MLMI '09, pp. 335–342. ACM, New York, NY, USA, 2009. doi: 10.1145/1647314.1647383

[3] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In *Proceedings of IEEE Virtual Reality (VR)*, p. (accepted), 2018.

[4] R. Azuma and C. Furmanski. Evaluating label placement for augmented reality view management. In *Proceedings of the 2Nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, pp. 66–. IEEE Computer Society, Washington, DC, USA, 2003.

[5] R. T. Azuma. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4):355–385, 1997.

[6] J. Baumeister, S. Y. Ssin, N. A. M. ElSayed, J. Dorrian, D. P. Webb, J. A. Walsh, T. M. Simon, A. Irlitti, R. T. Smith, M. Kohler, and B. H. Thomas. Cognitive cost of using augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2378–2388, Nov 2017.

[7] B. Bell, S. Feiner, and T. Höllerer. View management for virtual and augmented reality. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pp. 101–110. ACM, New York, NY, USA, 2001. doi: 10.1145/502348.502363

[8] M. Berning, F. Braun, T. Riedel, and M. Beigl. Proximityhat: A head-worn system for subtle sensory augmentation with tactile stimulation. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ISWC '15, pp. 31–38. ACM, New York, NY, USA, 2015. doi: 10.1145/2802083.2802088

[9] D. Black, J. Hettig, M. Luz, C. Hansen, R. Kikinis, and H. Hahn. Auditory feedback to support image-guided medical needle placement. *International Journal of Computer Assisted Radiology and Surgery*, 12(9):1655–1663, Sep 2017.

[10] 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.

[11] 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 (To appear)*, 2018.

[12] J. Borwick. *Loudspeaker and Headphone Handbook*. Taylor & Francis, 2012.

[13] S. Brewster and L. M. Brown. Tactons: structured tactile messages for non-visual information display. In *Proceedings of the fifth conference on Australasian user interface-Volume 28*, pp. 15–23. Australian Computer Society, Inc., 2004.

[14] O. Cakmakci and J. Rolland. Head-worn displays: a review. *Journal of display technology*, 2(3):199–216, 2006.

[15] M. Carrasco, T. L. McLean, S. M. Katz, and K. S. Frieder. Feature asymmetries in visual search: Effects of display duration, target eccentricity,

orientation and spatial frequency. *Vision Research*, 38(3):347 – 374, 1998. doi: 10.1016/S0042-6989(97)00152-1

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

[17] T. P. Caudell and D. W. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Twenty-Fifth Hawaii International Conference on System Sciences*, vol. ii, pp. 659–669 vol.2, Jan 1992. doi: 10.1109/HICSS.1992.183317

[18] 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, April 2017.

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

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

[21] S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. In *Proceedings of the 1st IEEE International Symposium on Wearable Computers*, ISWC '97, pp. 74–81. IEEE Computer Society, Washington, DC, USA, 1997.

[22] D. C. Foyle, A. D. Andre, and B. L. Hooey. Situation awareness in an augmented reality cockpit: Design, viewpoints and cognitive glue. In *Proceedings of the 11th International Conference on Human Computer Interaction*, vol. 1, pp. 3–9, 2005.

[23] D. C. Foyle, A. D. Andre, R. S. McCann, E. M. Wenzel, D. R. Begault, and V. Battiste. Taxiway navigation and situation awareness (t-nasa) system: Problem, design philosophy, and description of an integrated display suite for low-visibility airport surface operations. Technical report, SAE Technical Paper, 1996.

[24] G. A. French and T. Schnell. Terrain awareness & pathway guidance for head-up displays (tapguide); a simulator study of pilot performance. In *Digital Avionics Systems Conference, 2003. DASC'03. The 22nd*, vol. 2, pp. 9–C. IEEE, 2003.

[25] K. Fujii, J. Totz, and G.-Z. Yang. Visual search behaviour and analysis of augmented visualisation for minimally invasive surgery. In C. A. Linte, J. T. Moore, E. C. S. Chen, and D. R. Holmes, eds., *Augmented Environments for Computer-Assisted Interventions*, pp. 24–35. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[26] J. L. Gabbard, J. E. Swan, II, and D. Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence: Teleoper. Virtual Environ.*, 15(1):16–32, Feb. 2006. doi: 10.1162/pres.2006.15.1.16

[27] R. Grasset, T. Langlotz, D. Kalkofen, M. Tatzgern, and D. Schmalstieg. Image-driven view management for augmented reality browsers. In *Proceedings of the 2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ISMAR '12, pp. 177–186. IEEE Computer Society, Washington, DC, USA, 2012. doi: 10.1109/ISMAR.2012.6402555

[28] U. Gruenefeld, D. Ennenga, A. E. Ali, W. Heuten, and S. Boll. Eye-see360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, pp. 109–118. ACM, New York, NY, USA, 2017. doi: 10.1145/3131277.3132175

[29] C. Gutwin, A. Cockburn, and A. Coveney. Peripheral popout: The influence of visual angle and stimulus intensity on popout effects. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 208–219. ACM, New York, NY, USA, 2017. doi: 10.1145/3025453.3025984

[30] T. Hamill and L. Price. *The Hearing Sciences, Third Edition*. Plural Publishing, Incorporated, 2017.

[31] S. Henderson and S. Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE transactions on visualization and computer graphics*, 17(10):1355–1368, 2010.

[32] C. Ho, H. Tan, and C. Spence. The differential effect of vibrotactile and auditory cues on visual spatial attention. *Ergonomics*, 49:724–38, 07 2006.

[33] 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.

[34] A. Israr and I. Poupyrev. Tactile brush: Drawing on skin with a tactile grid display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 2019–2028. ACM, New York, NY, USA, 2011. doi: 10.1145/1978942.1979235

[35] S. Jin Cho, A. Ovcharenko, and U.-P. Chong. Front-back confusion resolution in 3d sound localization with hrtf databases. pp. 239 – 243, 11 2006. doi: 10.1109/IFOST.2006.312297

[36] 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.

[37] 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.

[38] 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*, CHI EA '16, pp. 2533–2539. ACM, New York, NY, USA, 2016. doi: 10.1145/2851581.2892355

[39] S. Kim and A. K. Dey. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 133–142. ACM, 2009.

[40] Y. Kim, C. Jun Chun, H. K. Kim, Y. Lee, D. Jang, and K. Kang. An integrated approach of 3d sound rendering techniques for sound externalization. pp. 682–693, 09 2010. doi: 10.1007/978-3-642-15696-0_63

[41] 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 *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 177–186, Sept 2014. doi: 10.1109/ISMAR.2014.6948425

[42] 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.

[43] E. Kruijff, J. S. II, and S. Feiner. Perceptual issues in augmented reality revisited. In *In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 3–12. IEEE Computer Society, 2010.

[44] 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*, pp. 1–1, 2018. doi: 10.1109/TVCG.2018.2854737

[45] T. Langlotz, H. Regenbrecht, S. Zollmann, and D. Schmalstieg. Audio stickies: visually-guided spatial audio annotations on a mobile augmented reality platform. In H. Shen, R. T. Smith, J. Paay, P. R. Calder, and T. G. Wyeld, eds., *Augmentation, Application, Innovation, Collaboration, OzCHI '13, Adelaide, Australia - November 25 - 29, 2013*, pp. 545–554. ACM, 2013.

[46] 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*, UIST '12, pp. 445–452. ACM, New York, NY, USA, 2012. doi: 10.1145/2380116.2380173

[47] T. Letowski and S. Letowski. *Localization error: Accuracy and precision of auditory localization*. 04 2011. doi: 10.5772/15652

[48] A. Leykin and M. Tuceryan. Automatic determination of text readability over textured backgrounds for augmented reality systems. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 224–230, Nov 2004. doi: 10.1109/ISMAR.2004.22

[49] 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.

[50] R. W. Lindeman and H. Noma. A classification scheme for multi-sensory augmented reality. In *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology*, VRST '07, pp. 175–178. ACM, New York, NY, USA, 2007. doi: 10.1145/1315184.1315216

[51] Z. Lipowski. Sensory and information inputs overload: Behavioral effects. *Comprehensive Psychiatry*, 16(3):199 – 221, 1975. doi: 10.1016/0010-440X(75)90047-4

[52] K. R. . G. N. Loomis, J. Sensory substitution of vision: Importance of perceptual and cognitive processing. In R. Manduchi and S. Kurniawan, eds., *Assistive Technology for Blindness and Low Vision*, pp. 162–191. CRC Press, 2012.

[53] 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*, UbiComp '14 Adjunct, pp. 1093–1102.

ACM, New York, NY, USA, 2014. doi: 10.1145/2638728.2641687

[54] T. M. Madhyastha and D. A. Reed. Data sonification: do you see what i hear? *IEEE Software*, 12(2):45–56, March 1995. doi: 10.1109/52.368264

[55] 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.

[56] J. P. McIntire, P. R. Havig, S. N. J. 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.

[57] E. McSorley and J. Findlay. Visual search in depth. *Vision Research*, 41(25):3487 – 3496, 2001. doi: 10.1016/S0042-6989(01)00197-3

[58] S. Mudd and E. McCormick. The use of auditory cues in a visual search task. *Journal of Applied Psychology*, 44(3):184–188, 1960.

[59] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, Mar 2006.

[60] R. L. Newman. *Head-up displays: Designing the way ahead*. Routledge, 2017.

[61] 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. doi: 10.3758/APP.72.6.1654

[62] K. Ogata, Y. Seya, K. Watanabe, and T. Ifukube. Effects of visual cues on the complicated search task. In L. Malmborg and T. Pederson, eds., *Nordic Conference on Human-Computer Interaction, NordiCHI '12, Copenhagen, Denmark, October 14-17, 2012*, pp. 478–485. ACM, 2012. doi: 10.1145/2399016.2399089

[63] 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.

[64] S. Peterson, M. Axholt, and S. R. Ellis. Managing visual clutter: A generalized technique for label segregation using stereoscopic disparity. In *IEEE Virtual Reality Conference 2008 (VR 2008), 8-12 March 2008, Reno, Nevada, USA, Proceedings*, pp. 169–176. IEEE Computer Society, 2008. doi: 10.1109/VR.2008.4480769

[65] S. D. Peterson, M. Axholt, and S. R. Ellis. Label segregation by remapping stereoscopic depth in far-field augmented reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pp. 143–152. IEEE Computer Society, Washington, DC, USA, 2008. doi: 10.1109/ISMAR.2008.4637341

[66] B. Pfleging, N. Henze, A. Schmidt, D. Rau, and B. Reitschuster. Influence of subliminal cueing on visual search tasks. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pp. 1269–1274. ACM, 2013.

[67] P. Quinlan. Visual feature integration theory: Past, present, and future. *Psychological bulletin*, 129:643–73, 10 2003. doi: 10.1037/0033-2909.129.5.643

[68] D. Ren, T. Goldschwendt, Y. Chang, and T. Höllerer. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In *Virtual Reality (VR), 2016 IEEE*, pp. 93–102. IEEE, 2016.

[69] P. Renner and T. Pfeiffer. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 186–194, March 2017.

[70] F. Ribeiro, D. Florłncio, 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.

[71] 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.

[72] M. Rusch, M. Schall, P. Gavin, J. Lee, J. Dawson, S. Vecera, and M. Rizzo. Directing driver attention with augmented reality cues. *Transportation research. Part F, Traffic psychology and behaviour*, 16:127–137, 01 2013.

[73] D. Schmalstieg and T. Hollerer. *Augmented Reality - Principles and Practice*. Addison-Wesley Professional, 2016.

[74] B. Schwerdtfeger and G. Klinker. Supporting order picking with augmented reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pp. 91–94. IEEE Computer Society, Washington, DC, USA, 2008. doi: 10.1109/ISMAR.2008.4637331

[75] S. Shelley, M. Alonso, J. Hollowood, M. Pettitt, S. Sharples, D. Hermes, and A. Kohlrausch. Interactive sonification of curve shape and curvature data. In M. E. Altinsoy, U. Jekosch, and S. Brewster, eds., *Haptic and Audio Interaction Design*, pp. 51–60. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[76] C. Spence and S. Squire. Multisensory integration: maintaining the perception of synchrony. *Current Biology*, 13(13):R519–R521, 2003.

[77] D. Talsma, D. Senkowski, S. Soto-Faraco, and M. G. Woldorff. The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, 14(9):400–410, 2010.

[78] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pp. 73–80. ACM, New York, NY, USA, 2003. doi: 10.1145/642611.642626

[79] M. Tonnis, C. Sandor, G. Klinker, C. Lange, and H. Bubb. Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*, pp. 56–59. IEEE, 2005.

[80] C. Trepkowski, T. Eibich, J. Maiero, A. Marquardt, 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.

[81] 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.

[82] E. E. Veas, E. Mendez, S. K. Feiner, and D. Schmalstieg. Directing attention and influencing memory with visual saliency modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pp. 1471–1480. ACM, 2011.

[83] E. Wenzel, E. M, M. Arruda, K. , D. J, W. , and F. L. Localization using non-individualized head-related transfer functions. *JASA*, 94:111–123, 01 1993.

[84] F. Wightman and D. Kistler. Multidimensional scaling analysis of head-related transfer functions. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 98–101, Oct 1993. doi: 10.1109/ASPAA.1993.379987

[85] J. F. Wohlwill. Human adaptation to levels of environmental stimulation. *Human Ecology*, 2(2):127–147, 1974.

[86] J. M. Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1(2):202–238, Jun 1994. doi: 10.3758/BF03200774